

# Rotation Invariant Householder Parameterization for Bayesian PCA

Rajbir-Singh Nirwan



# Outline

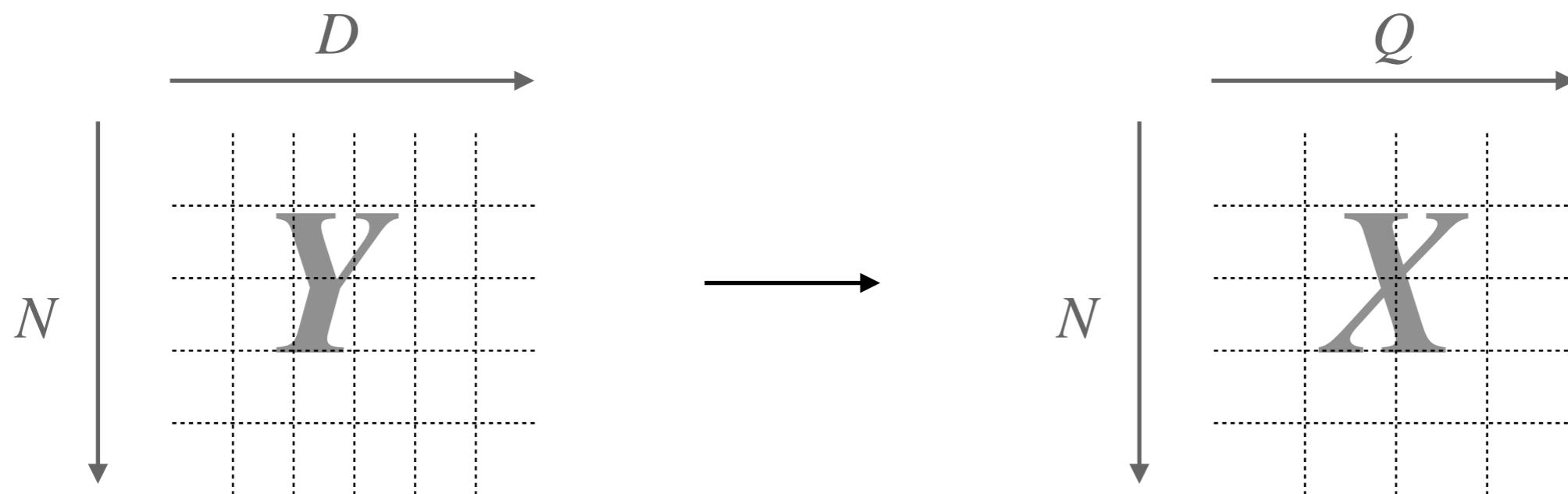
- Probabilistic PCA (PPCA)
- Non-identifiability issue of PPCA
- Conceptual solution to the problem
- Implementation
- Results

# Probabilistic PCA

## Classical PCA

Formulated as a projection from data space  $Y$  to a lower dimensional latent space  $X$

$$Y \in \mathbb{R}^{N \times D} \rightarrow X \in \mathbb{R}^{N \times Q}$$



# Probabilistic PCA

## Probabilistic PCA (PPCA)

Viewed as a generative model, that maps the latent space  $X$  to the data space  $Y$

$$X \in \mathbb{R}^{N \times Q} \quad \rightarrow \quad Y \in \mathbb{R}^{N \times D}$$

$$Y = XW^T + \epsilon$$

$$X \sim \mathcal{N}(\mathbf{0}, I), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$$

# Probabilistic PCA

## Probabilistic PCA (PPCA)

Viewed as a generative model, that maps the latent space  $X$  to the data space  $Y$

$$X \in \mathbb{R}^{N \times Q} \quad \rightarrow \quad Y \in \mathbb{R}^{N \times D}$$

$$Y = XW^T + \epsilon$$

$$X \sim \mathcal{N}(\mathbf{0}, I), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$$

$$p(Y|W) = \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \mathbf{0}, WW^T + \sigma^2 I)$$

$$WRR^TW^T = WW^T \quad \forall RR^T = I$$

# Probabilistic PCA

## Probabilistic PCA (PPCA)

Viewed as a generative model, that maps the latent space  $X$  to the data space  $Y$

$$X \in \mathbb{R}^{N \times Q} \rightarrow Y \in \mathbb{R}^{N \times D}$$

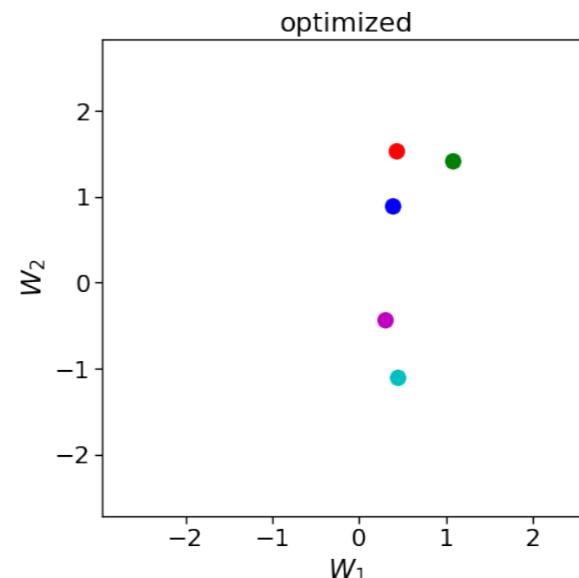
$$Y = XW^T + \epsilon$$

$$X \sim \mathcal{N}(\mathbf{0}, I), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$$

$$p(Y|W) = \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \mathbf{0}, WW^T + \sigma^2 I)$$

$$WRR^TW^T = WW^T \quad \forall RR^T = I$$

- Optimization for  $D=5, Q=2$



$$\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \\ w_{51} & w_{52} \end{pmatrix}$$

# Probabilistic PCA

## Probabilistic PCA (PPCA)

Viewed as a generative model, that maps the latent space  $X$  to the data space  $Y$

$$X \in \mathbb{R}^{N \times Q} \rightarrow Y \in \mathbb{R}^{N \times D}$$

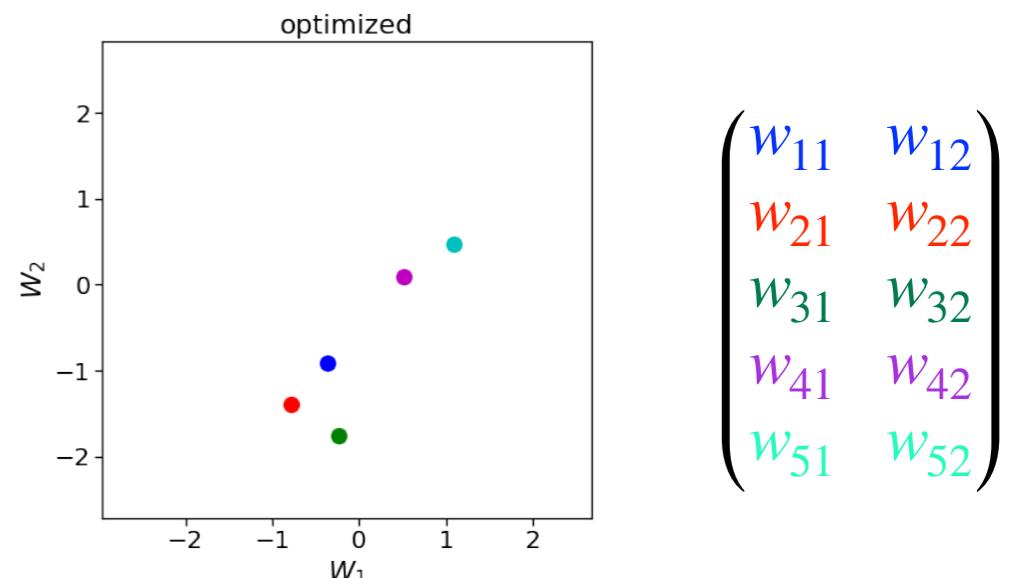
$$Y = XW^T + \epsilon$$

$$X \sim \mathcal{N}(\mathbf{0}, I), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$$

$$p(Y|W) = \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \mathbf{0}, WW^T + \sigma^2 I)$$

$$WRR^TW^T = WW^T \quad \forall RR^T = I$$

- Optimization for  $D=5, Q=2$



# Probabilistic PCA

## Probabilistic PCA (PPCA)

Viewed as a generative model, that maps the latent space  $X$  to the data space  $Y$

$$X \in \mathbb{R}^{N \times Q} \rightarrow Y \in \mathbb{R}^{N \times D}$$

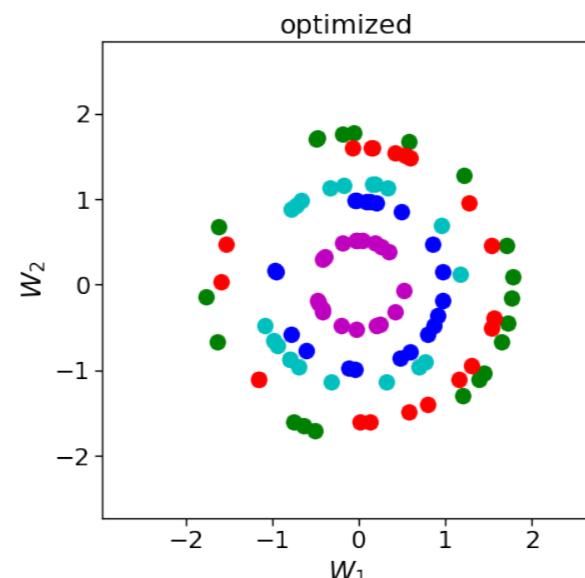
$$Y = XW^T + \epsilon$$

$$X \sim \mathcal{N}(\mathbf{0}, I), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$$

$$p(Y|W) = \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \mathbf{0}, WW^T + \sigma^2 I)$$

$$WRR^TW^T = WW^T \quad \forall RR^T = I$$

- Optimization for  $D=5$ ,  $Q=2$



$$\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \\ w_{51} & w_{52} \end{pmatrix}$$

# Probabilistic PCA

## Probabilistic PCA (PPCA)

Viewed as a generative model, that maps the latent space  $X$  to the data space  $Y$

$$X \in \mathbb{R}^{N \times Q} \rightarrow Y \in \mathbb{R}^{N \times D}$$

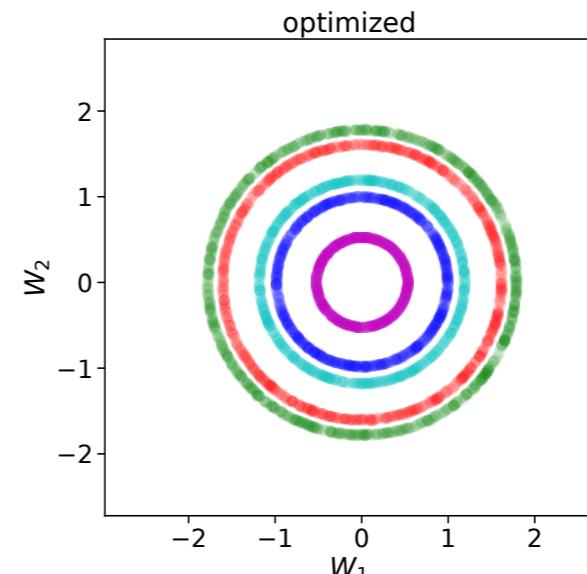
$$Y = XW^T + \epsilon$$

$$X \sim \mathcal{N}(\mathbf{0}, I), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$$

$$p(Y|W) = \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \mathbf{0}, WW^T + \sigma^2 I)$$

$$WRR^TW^T = WW^T \quad \forall RR^T = I$$

- Rotation invariant likelihood



$$\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \\ w_{51} & w_{52} \end{pmatrix}$$

# Probabilistic PCA

## Probabilistic PCA (PPCA)

Viewed as a generative model, that maps the latent space  $X$  to the data space  $Y$

$$X \in \mathbb{R}^{N \times Q} \rightarrow Y \in \mathbb{R}^{N \times D}$$

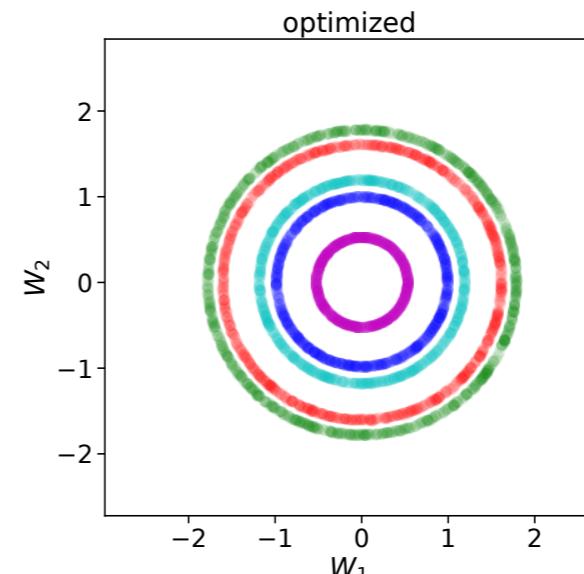
$$Y = XW^T + \epsilon$$

$$X \sim \mathcal{N}(\mathbf{0}, I), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$$

$$p(Y|W) = \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \mathbf{0}, WW^T + \sigma^2 I)$$

$$WRR^TW^T = WW^T \quad \forall RR^T = I$$

- Rotation invariant likelihood



$$\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \\ w_{51} & w_{52} \end{pmatrix}$$

- Solution

$$WW^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma^2 U^T$$

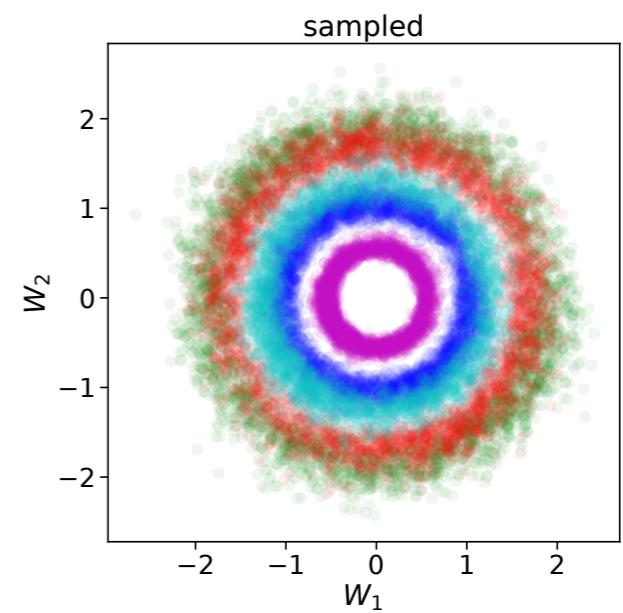
$$p(Y|U, \Sigma) = \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \mathbf{0}, U\Sigma^2 U^T + \sigma^2 I)$$

# Bayesian approach to PPCA

$$p(W|Y) = \frac{p(Y|W)p(W)}{p(Y)}$$

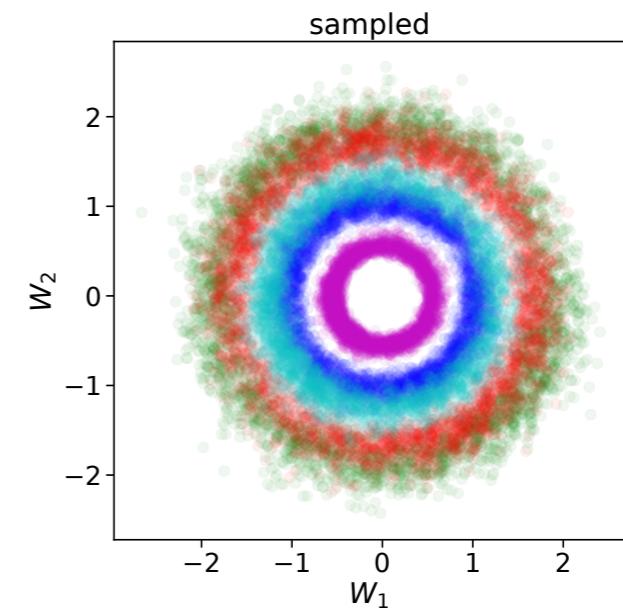
# Bayesian approach to PPCA

$$p(W|Y) = \frac{p(Y|W)p(W)}{p(Y)}$$



# Bayesian approach to PPCA

$$p(W|Y) = \frac{p(Y|W)p(W)}{p(Y)}$$



- If prior does not break the symmetry, posterior will be rotation invariant as well
- Sampling will be challenging, posterior averages are meaningless and the interpretation of the latent space is almost impossible

# Solution

- Find different parameterization of the model, such that the probabilistic model is not changed

## Outline of procedure

- SVD of  $\mathbf{W}$  
$$\mathbf{WW}^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T)^T = \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^T$$
- Fix coordinate system 
$$\mathbf{V} = \mathbf{I}$$
- Specify prior 
$$p(\mathbf{U}, \boldsymbol{\Sigma})$$
- Sample from 
$$p(\mathbf{U}, \boldsymbol{\Sigma} | \mathbf{Y})$$

# Solution

- Find different parameterization of the model, such that the probabilistic model is not changed

## Outline of procedure

- SVD of  $\mathbf{W}$  
$$\mathbf{WW}^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T)^T = \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^T$$
- Fix coordinate system 
$$\mathbf{V} = \mathbf{I}$$
- Specify prior 
$$p(\mathbf{U}, \boldsymbol{\Sigma})$$
- Sample from 
$$p(\mathbf{U}, \boldsymbol{\Sigma} | \mathbf{Y})$$

---

$$p(\mathbf{Y} | \mathbf{W}) = \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \mathbf{0}, \mathbf{WW}^T + \sigma^2 \mathbf{I})$$

$$p(\mathbf{Y} | \mathbf{U}, \boldsymbol{\Sigma}) = \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \mathbf{0}, \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^T + \sigma^2 \mathbf{I})$$

$$\begin{matrix} U \sim ? \\ \Sigma \sim ? \end{matrix} \rightarrow U\Sigma\Sigma^T U^T \text{ Wishart}$$

# Theory

- Since  $\mathbf{U}, \boldsymbol{\Sigma}$  is SVD of  $\mathbf{W} \rightarrow \mathbf{U}$  is a orthogonal matrix

$\mathbf{U} \in \mathcal{V}_{Q,D}$  **Stiefel manifold**

$$\mathcal{V}_{Q,D} = \{\mathbf{U} \in \mathbb{R}^{D \times Q} \mid \mathbf{U}^T \mathbf{U} = \mathbf{I}\}$$

$$\begin{matrix} U \sim ? \\ \Sigma \sim ? \end{matrix} \rightarrow U\Sigma\Sigma^T U^T \text{ Wishart}$$

# Theory

- Since  $U, \Sigma$  is SVD of  $W \rightarrow U$  is a orthogonal matrix

$$U \in \mathcal{V}_{Q,D} \quad \text{Stiefel manifold} \qquad \mathcal{V}_{Q,D} = \{U \in \mathbb{R}^{D \times Q} \mid U^T U = I\}$$

- $U, \Sigma^2$  is eigenvalue decomposition of  $WW^T$

Eigenvectors of Wishart matrix are distributed uniformly in space of orthogonal matrices ( Blai (2007), Uhlig (1994) )

$$\begin{matrix} U \sim ? \\ \Sigma \sim ? \end{matrix} \rightarrow U\Sigma\Sigma^T U^T \text{ Wishart}$$

# Theory

- Since  $\mathbf{U}, \boldsymbol{\Sigma}$  is SVD of  $\mathbf{W} \rightarrow \mathbf{U}$  is a orthogonal matrix

$$U \in \mathcal{V}_{Q,D} \quad \text{Stiefel manifold} \qquad \mathcal{V}_{Q,D} = \{U \in \mathbb{R}^{D \times Q} \mid U^T U = I\}$$

- $\mathbf{U}, \boldsymbol{\Sigma}^2$  is eigenvalue decomposition of  $\mathbf{W}\mathbf{W}^T$

Eigenvectors of Wishart matrix are distributed uniformly in space of orthogonal matrices ( Blai (2007), Uhlig (1994) )

**→  $\mathbf{U}$  is uniformly distributed on the Stiefel manifold**

$$\begin{matrix} U \sim ? \\ \Sigma \sim ? \end{matrix} \rightarrow U \Sigma \Sigma^T U^T \text{ Wishart}$$

# Theory

- Since  $U, \Sigma$  is SVD of  $W \rightarrow U$  is a orthogonal matrix

$$U \in \mathcal{V}_{Q,D} \quad \textbf{Stiefel manifold} \quad \mathcal{V}_{Q,D} = \{U \in \mathbb{R}^{D \times Q} \mid U^T U = I\}$$

- $U, \Sigma^2$  is eigenvalue decomposition of  $WW^T$

Eigenvectors of Wishart matrix are distributed uniformly in space of orthogonal matrices ( Blai (2007), Uhlig (1994) )

**$\rightarrow U$  is uniformly distributed on the Stiefel manifold**

- Distribution of the ordered eigenvalue matrix  $\Sigma^2$  of  $WW^T$  is (James & Lee (2014))

$$p(\lambda) = c e^{-\frac{1}{2} \sum_{q=1}^Q \lambda_q} \prod_{q=1}^Q \left( \lambda_q^{\frac{D-Q-1}{2}} \prod_{q'=q+1}^Q |\lambda_q - \lambda_{q'}| \right)$$

$$p(\sigma_1, \dots, \sigma_Q) = c e^{-\frac{1}{2} \sum_{q=1}^Q \sigma_q^2} \prod_{q=1}^Q \left( \sigma_q^{D-Q-1} \prod_{q'=q+1}^Q |\sigma_q^2 - \sigma_{q'}^2| \right) \prod_{q=1}^Q 2\sigma_q$$

# Implementation

## How to uniformly sample $U$ on $\mathcal{V}_{Q,D}$

**Theorem 2** Let  $\mathbf{v}_D, \mathbf{v}_{D-1}, \dots, \mathbf{v}_1$  be uniformly distributed on the unit spheres  $\mathbb{S}^{D-1}, \dots, \mathbb{S}^0$  respectively, where  $\mathbb{S}^{n-1}$  is the unit sphere in  $\mathbb{R}^n$ . Furthermore, let  $\mathbf{H}_n(\mathbf{v}_n)$  be the  $n$ -th Householder transformation as defined in equation (2.20). The product

$$\mathbf{Q} = \mathbf{H}_D(\mathbf{v}_D) \mathbf{H}_{D-1}(\mathbf{v}_{D-1}) \dots \mathbf{H}_1(\mathbf{v}_1) \quad (2.21)$$

is a random orthogonal matrix with distribution given by the Haar measure on  $O(D)$ .

Mezzadri (2007)

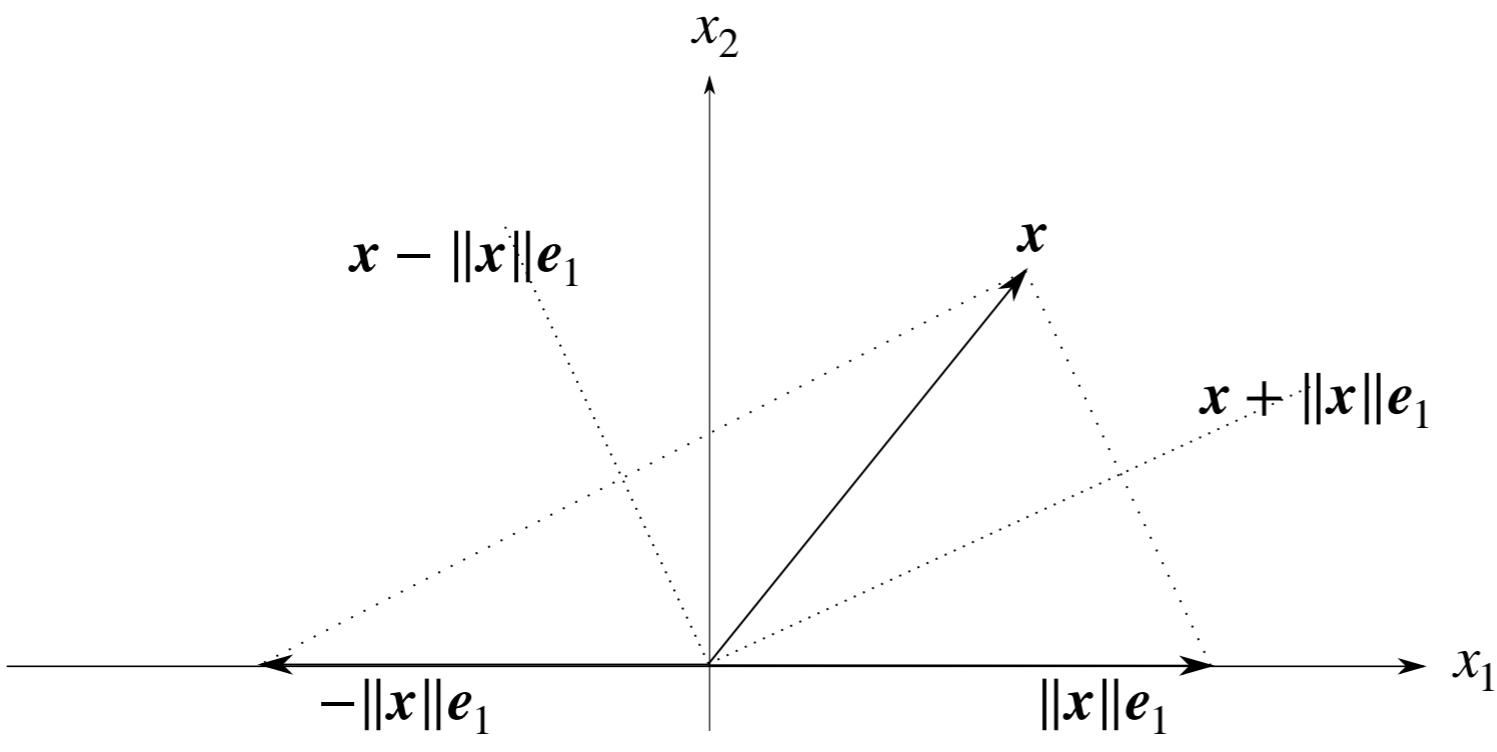
**for**  $n = D : 1$

$$\begin{aligned} \mathbf{v}_n &\sim \text{uniform on } \mathbb{S}^{n-1} \\ \mathbf{u}_n &= \frac{\mathbf{v}_n + \text{sgn}(\mathbf{v}_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1}{\|\mathbf{v}_n + \text{sgn}(\mathbf{v}_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1\|} \\ \tilde{\mathbf{H}}_n(\mathbf{v}_n) &= -\text{sgn}(\mathbf{v}_{n1}) (\mathbf{I} - 2\mathbf{u}_n \mathbf{u}_n^T) \\ \mathbf{H}_n &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{H}}_n \end{pmatrix} \end{aligned}$$

$$U = H_D(\mathbf{v}_D) H_{D-1}(\mathbf{v}_{D-1}) \dots H_1(\mathbf{v}_1)$$

# Householder Transformations

- Used to reflect a vector in such a way that all coordinates but one disappear, e.g.: QR-decomposition



$$u = x \pm \|x\|e_1$$

$$H = \mathbf{1} - 2\hat{u}\hat{u}^T$$

$$Hx = \|x\|e_1$$

# Householder Transformations

Example for  $D = 2$

$$U \in \mathcal{V}_{2,2}$$

$$\boldsymbol{v}_1 = \begin{pmatrix} 0 \\ v_{11} \end{pmatrix} \quad \boldsymbol{v}_2 = \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}$$

for  $n = D : 1$

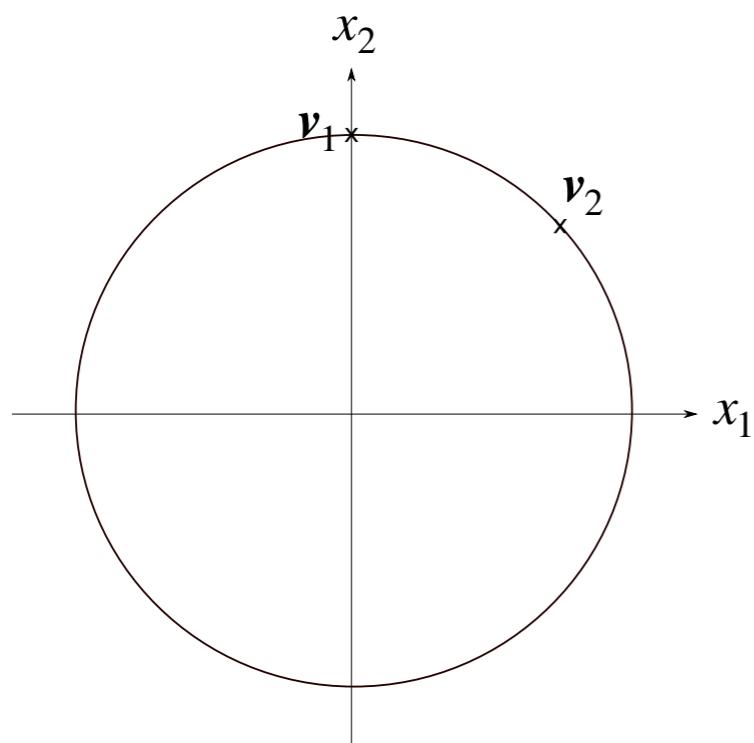
$\boldsymbol{v}_n \sim \text{uniform on } \mathbb{S}^{n-1}$

$$\boldsymbol{u}_n = \frac{\boldsymbol{v}_n + \text{sgn}(\boldsymbol{v}_{n1}) \|\boldsymbol{v}_n\| \boldsymbol{e}_1}{\|\boldsymbol{v}_n + \text{sgn}(\boldsymbol{v}_{n1}) \|\boldsymbol{v}_n\| \boldsymbol{e}_1\|}$$

$$\tilde{\boldsymbol{H}}_n (\boldsymbol{v}_n) = -\text{sgn}(\boldsymbol{v}_{n1}) (\boldsymbol{I} - 2\boldsymbol{u}_n \boldsymbol{u}_n^T)$$

$$\boldsymbol{H}_n = \begin{pmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{H}}_n \end{pmatrix}$$

$$U = \boldsymbol{H}_D (\boldsymbol{v}_D) \boldsymbol{H}_{D-1} (\boldsymbol{v}_{D-1}) \dots \boldsymbol{H}_1 (\boldsymbol{v}_1)$$



# Householder Transformations

Example for  $D = 2$

$$\mathbf{U} \in \mathcal{V}_{2,2}$$

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ v_{11} \end{pmatrix} \quad \mathbf{v}_2 = \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}$$

for  $n = D : 1$

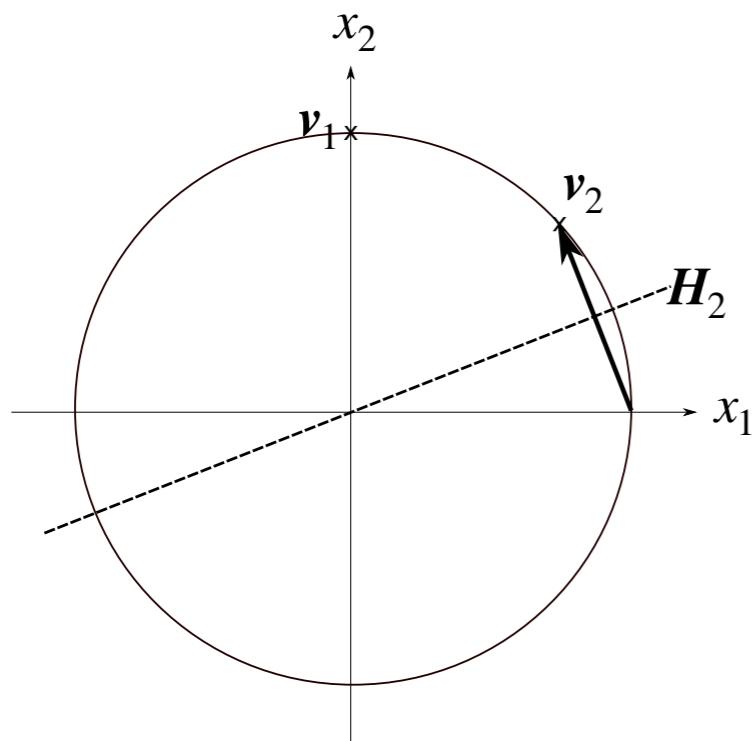
$\mathbf{v}_n \sim \text{uniform on } \mathbb{S}^{n-1}$

$$\mathbf{u}_n = \frac{\mathbf{v}_n + \operatorname{sgn}(v_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1}{\|\mathbf{v}_n + \operatorname{sgn}(v_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1\|}$$

$$\tilde{\mathbf{H}}_n(\mathbf{v}_n) = -\operatorname{sgn}(v_{n1})(\mathbf{I} - 2\mathbf{u}_n\mathbf{u}_n^T)$$

$$\mathbf{H}_n = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{H}}_n \end{pmatrix}$$

$$\mathbf{U} = \mathbf{H}_D(\mathbf{v}_D) \mathbf{H}_{D-1}(\mathbf{v}_{D-1}) \dots \mathbf{H}_1(\mathbf{v}_1)$$



# Householder Transformations

Example for  $D = 2$

$$\mathbf{U} \in \mathcal{V}_{2,2}$$

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ v_{11} \end{pmatrix} \quad \mathbf{v}_2 = \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}$$

for  $n = D : 1$

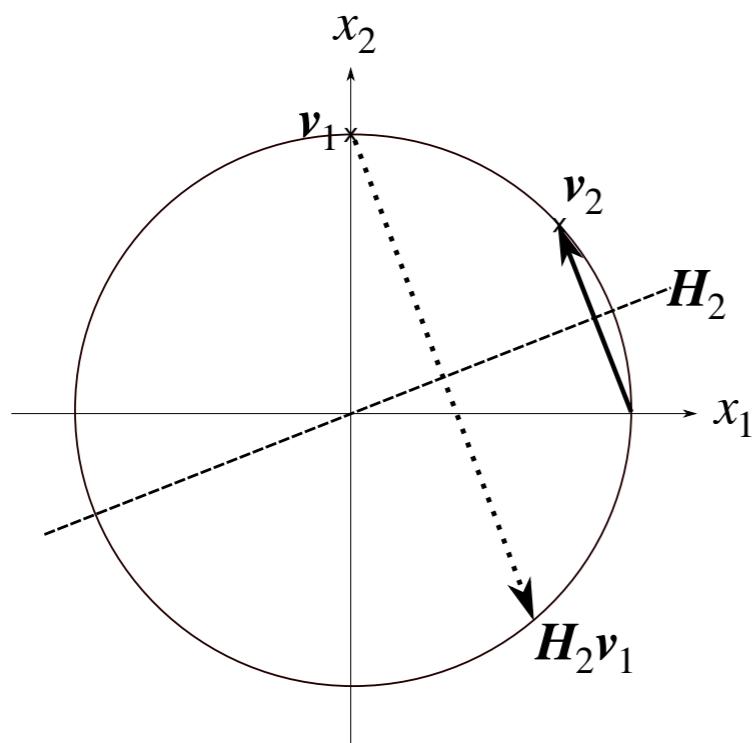
$\mathbf{v}_n \sim \text{uniform on } \mathbb{S}^{n-1}$

$$\mathbf{u}_n = \frac{\mathbf{v}_n + \text{sgn}(\mathbf{v}_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1}{\|\mathbf{v}_n + \text{sgn}(\mathbf{v}_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1\|}$$

$$\tilde{\mathbf{H}}_n (\mathbf{v}_n) = -\text{sgn}(\mathbf{v}_{n1}) (\mathbf{I} - 2\mathbf{u}_n \mathbf{u}_n^T)$$

$$\mathbf{H}_n = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{H}}_n \end{pmatrix}$$

$$\mathbf{U} = \mathbf{H}_D (\mathbf{v}_D) \mathbf{H}_{D-1} (\mathbf{v}_{D-1}) \dots \mathbf{H}_1 (\mathbf{v}_1)$$



# Householder Transformations

Example for  $D = 2$

$$U \in \mathcal{V}_{2,2}$$

$$\boldsymbol{v}_1 = \begin{pmatrix} 0 \\ v_{11} \end{pmatrix} \quad \boldsymbol{v}_2 = \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}$$

for  $n = D : 1$

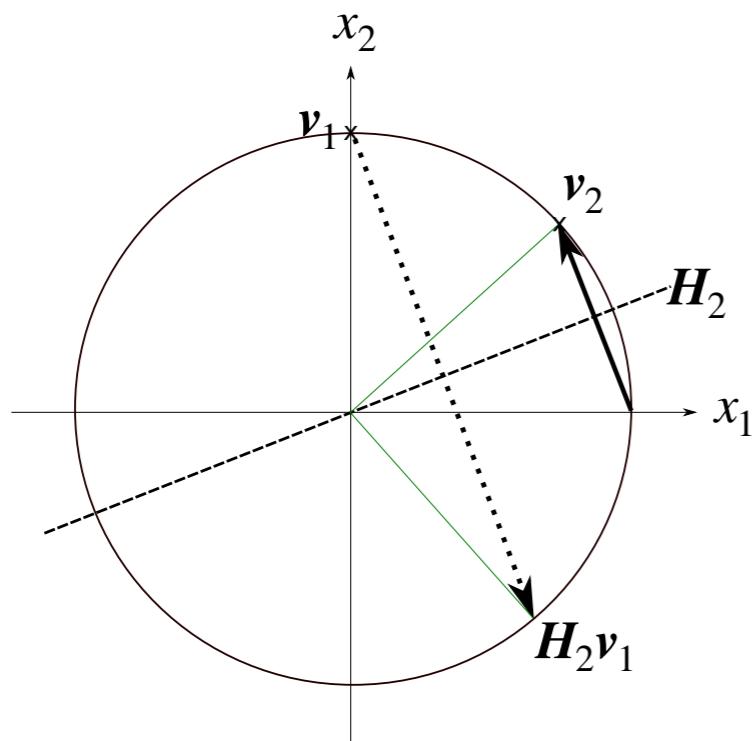
$\boldsymbol{v}_n \sim \text{uniform on } \mathbb{S}^{n-1}$

$$\boldsymbol{u}_n = \frac{\boldsymbol{v}_n + \text{sgn}(\boldsymbol{v}_{n1}) \|\boldsymbol{v}_n\| \boldsymbol{e}_1}{\|\boldsymbol{v}_n + \text{sgn}(\boldsymbol{v}_{n1}) \|\boldsymbol{v}_n\| \boldsymbol{e}_1\|}$$

$$\tilde{\boldsymbol{H}}_n (\boldsymbol{v}_n) = -\text{sgn}(\boldsymbol{v}_{n1}) (\boldsymbol{I} - 2\boldsymbol{u}_n \boldsymbol{u}_n^T)$$

$$\boldsymbol{H}_n = \begin{pmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{H}}_n \end{pmatrix}$$

$$U = \boldsymbol{H}_D (\boldsymbol{v}_D) \boldsymbol{H}_{D-1} (\boldsymbol{v}_{D-1}) \dots \boldsymbol{H}_1 (\boldsymbol{v}_1)$$

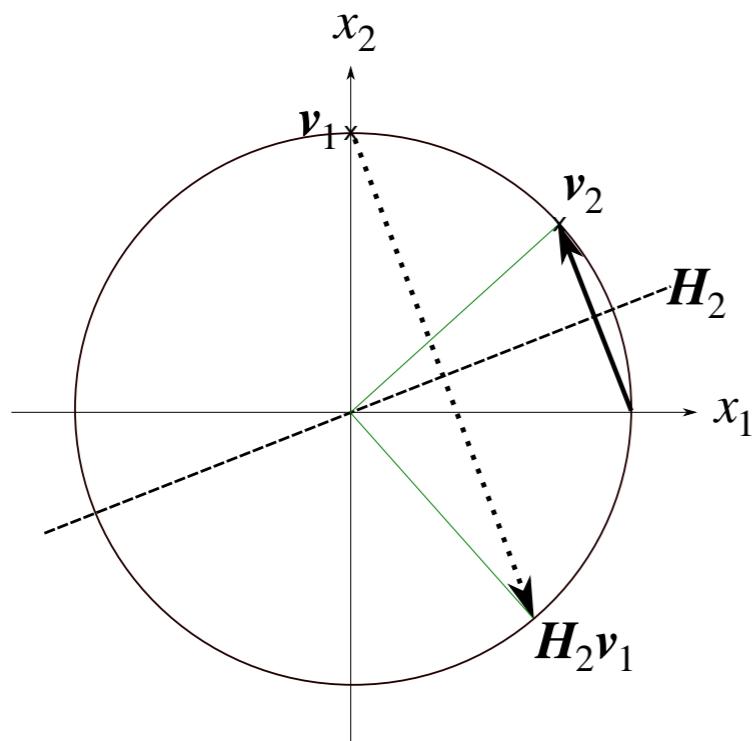


# Householder Transformations

Example for  $D = 2$

$$\mathbf{U} \in \mathcal{V}_{2,2}$$

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ v_{11} \end{pmatrix} \quad \mathbf{v}_2 = \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}$$



for  $n = D : 1$

$\mathbf{v}_n \sim \text{uniform on } \mathbb{S}^{n-1}$

$$\mathbf{u}_n = \frac{\mathbf{v}_n + \operatorname{sgn}(v_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1}{\|\mathbf{v}_n + \operatorname{sgn}(v_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1\|}$$

$$\tilde{\mathbf{H}}_n(\mathbf{v}_n) = -\operatorname{sgn}(v_{n1})(\mathbf{I} - 2\mathbf{u}_n\mathbf{u}_n^T)$$

$$\mathbf{H}_n = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{H}}_n \end{pmatrix}$$

$$\mathbf{U} = \mathbf{H}_D(\mathbf{v}_D) \mathbf{H}_{D-1}(\mathbf{v}_{D-1}) \dots \mathbf{H}_1(\mathbf{v}_1)$$

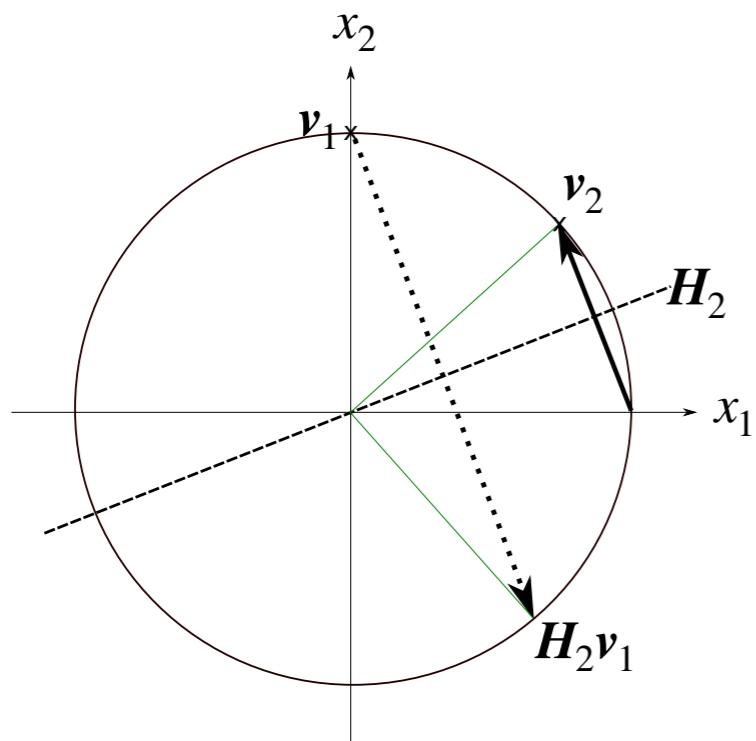
$$\mathbf{U} = \mathbf{H}_2 \mathbf{H}_1 = \mathbf{H}_2 \begin{pmatrix} 1 & 0 \\ 0 & v_{11} \end{pmatrix} = (\mathbf{v}_2 \quad \mathbf{H}_2 \mathbf{v}_1)$$

# Householder Transformations

Example for  $D = 2$

$$\mathbf{U} \in \mathcal{V}_{2,2}$$

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ v_{11} \end{pmatrix} \quad \mathbf{v}_2 = \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}$$



**for**  $n = D : 1$

$\mathbf{v}_n \sim \text{uniform on } \mathbb{S}^{n-1}$

$$\mathbf{u}_n = \frac{\mathbf{v}_n + \text{sgn}(\mathbf{v}_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1}{\|\mathbf{v}_n + \text{sgn}(\mathbf{v}_{n1}) \|\mathbf{v}_n\| \mathbf{e}_1\|}$$

$$\tilde{\mathbf{H}}_n (\mathbf{v}_n) = -\text{sgn}(\mathbf{v}_{n1}) (\mathbf{I} - 2\mathbf{u}_n \mathbf{u}_n^T)$$

$$\mathbf{H}_n = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{H}}_n \end{pmatrix}$$

$$\mathbf{U} = \mathbf{H}_D (\mathbf{v}_D) \mathbf{H}_{D-1} (\mathbf{v}_{D-1}) \dots \mathbf{H}_1 (\mathbf{v}_1)$$

$$\mathbf{U} = \mathbf{H}_2 \mathbf{H}_1 = \mathbf{H}_2 \begin{pmatrix} 1 & 0 \\ 0 & v_{11} \end{pmatrix} = (\mathbf{v}_2 \quad \mathbf{H}_2 \mathbf{v}_1)$$

$$\mathbf{U} = \mathbf{H}_3 \mathbf{H}_2 \mathbf{H}_1 = (\mathbf{v}_3 \quad \mathbf{H}_3 \mathbf{v}_2 \quad \mathbf{H}_3 \mathbf{H}_2 \mathbf{v}_1)$$

# Solution

- Find different parameterization of the model, such that the probabilistic model is not changed

## Outline of procedure

- SVD of  $W$  
$$WW^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma^2 U^T$$
- Fix coordinate system  $V = I$
- Specify prior  $p(U, \Sigma)$
- Sample from  $p(U, \Sigma | Y)$

---

$W \sim \mathcal{N}(\mathbf{0}, I) \rightarrow WW^T$  is **Wishart distributed**

$U \sim ?$      $\Sigma \sim ?$      $\rightarrow U\Sigma\Sigma^T U^T$  is **Wishart distributed**

# Implementation

**The full generative model for Bayesian PPCA:**

$$\boldsymbol{v}_D, \dots, \boldsymbol{v}_{D-Q+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\boldsymbol{\sigma} \sim p(\boldsymbol{\sigma})$$

$$\boldsymbol{\mu} \sim p(\boldsymbol{\mu})$$

$$\mathbf{U} = \prod_{q=1}^Q \mathbf{H}_{D-q+1} \left( \boldsymbol{v}_{D-q+1} \right)$$

$$\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma})$$

$$\mathbf{W} = \mathbf{U}\boldsymbol{\Sigma}$$

$$\sigma_{\text{noise}} \sim p(\sigma_{\text{noise}})$$

$$Y \sim \prod_{n=1}^N \mathcal{N}(Y_{n,:} | \boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T + \sigma_{\text{noise}}^2 \mathbf{I})$$

# Results

## Synthetic Dataset

- Construction  
 $(N, D, Q) = (150, 5, 2)$

$$X \sim \mathcal{N}(\mathbf{0}, I) \in \mathbb{R}^{N \times Q}$$

$$U \sim \text{uniform on Stiefel } \mathcal{V}_{Q,D}$$

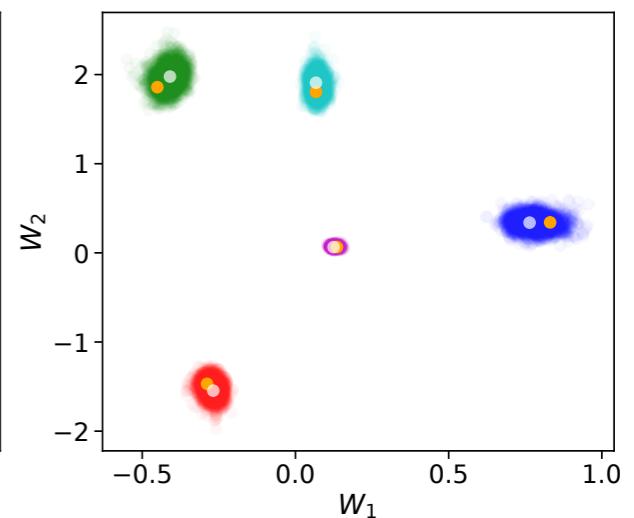
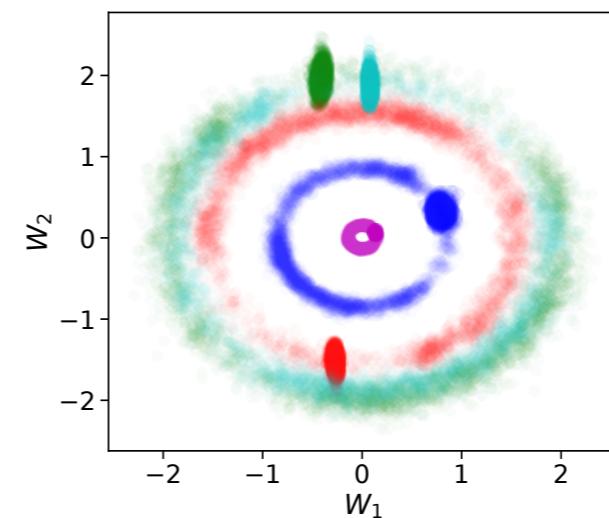
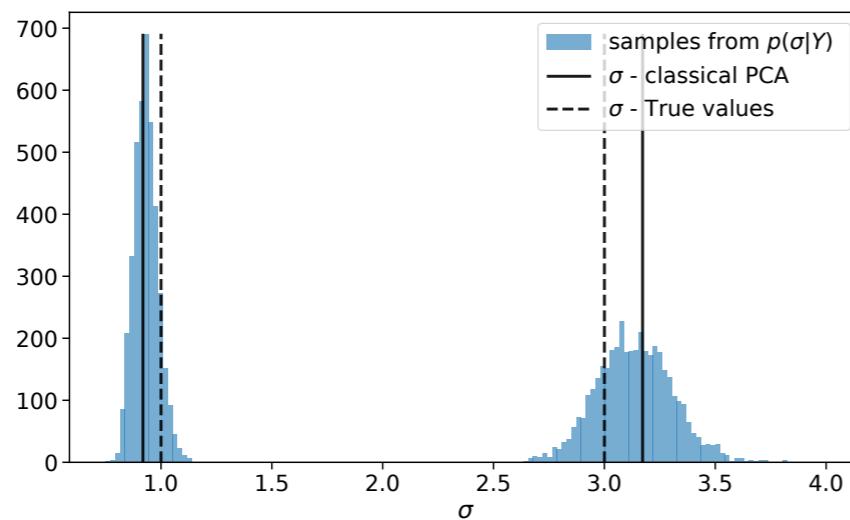
$$\epsilon \sim \mathcal{N}(0, 0.01) \in \mathbb{R}^{N \times D}$$

$$\Sigma = \text{diag} (\sigma_1, \sigma_2) = \text{diag} (3.0, 1.0)$$

$$W = U\Sigma \in \mathbb{R}^{D \times Q}$$

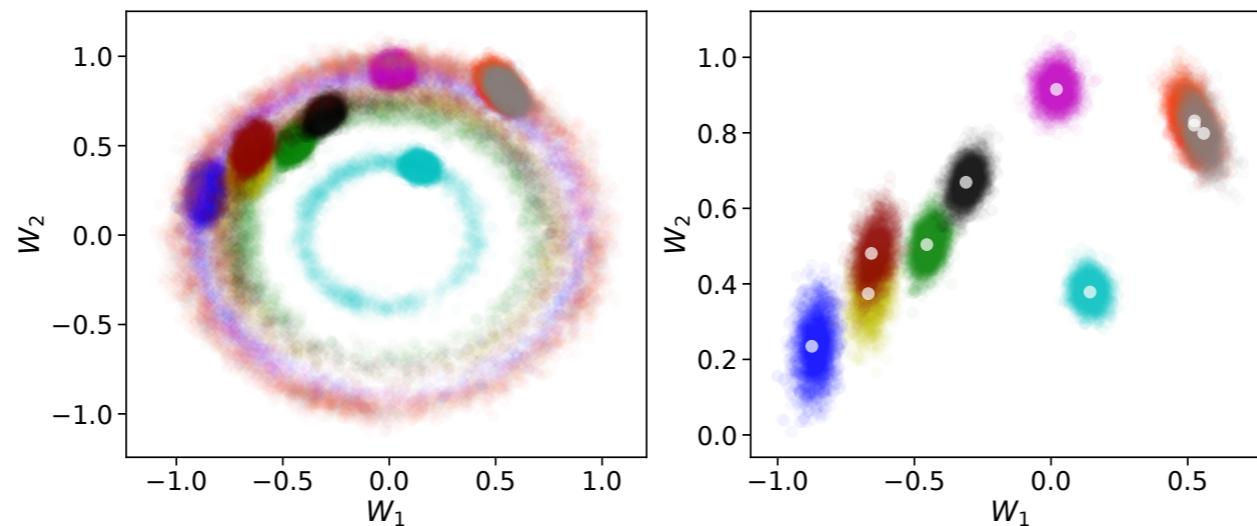
$$Y = XW^T + \epsilon$$

- Inference



# Results

Breast Cancer Wisconsin Dataset  $(N, D) = (569, 30)$



## Advantages

1. Breaks the rotation symmetry without changing the probabilistic model
2. Enrichment of the classical PCA solution with uncertainty estimates
3. Decomposition of  $W$  into rotation  $U$  and principle variances  $\Sigma$ 
  - Allows to construct other priors without issues
  - Sparsity prior on principle variances without a-priori rotation preference
  - If desired a-priori rotation preference without affecting the variances

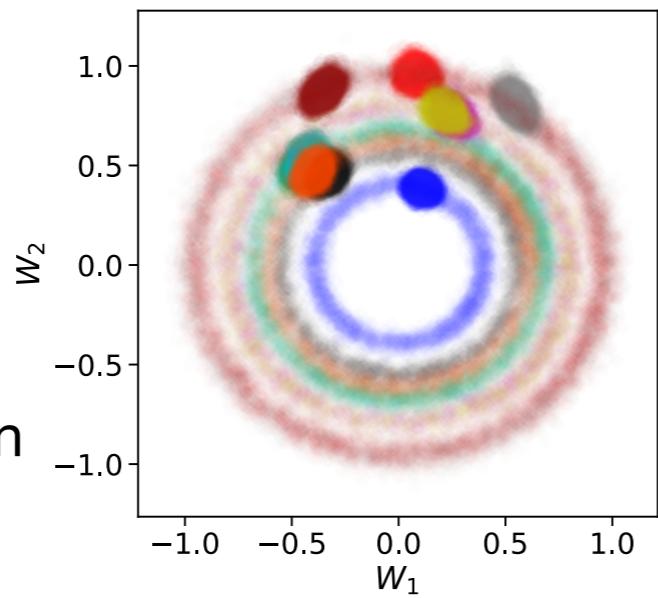
# Results

Breast Cancer

$(N, D) = (569, 30)$

Time House: 9.5 min

Time Standard: 25.6 min

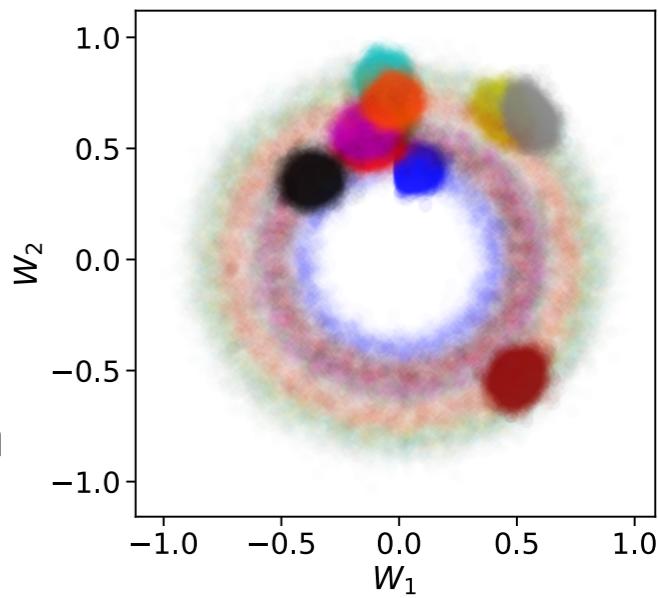


Diabetes

$(N, D) = (442, 10)$

Time House: 8.9 min

Time Standard: 2.4 min

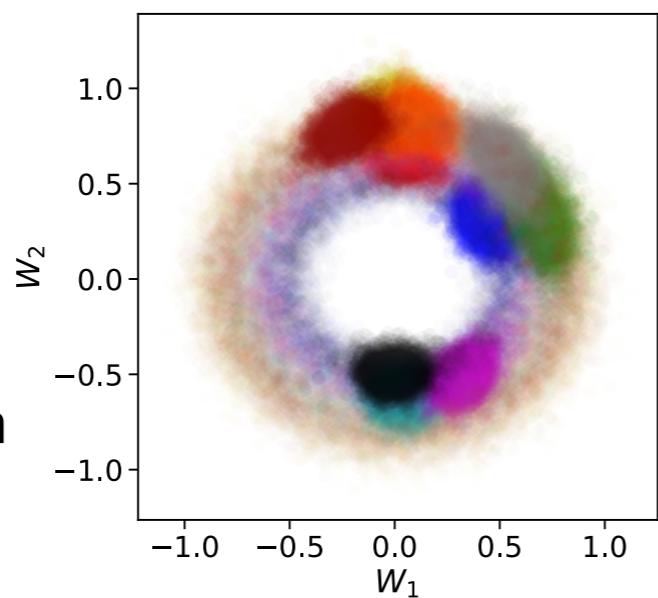


Wine

$(N, D) = (178, 13)$

Time House: 0.4 min

Time Standard: 1.2 min

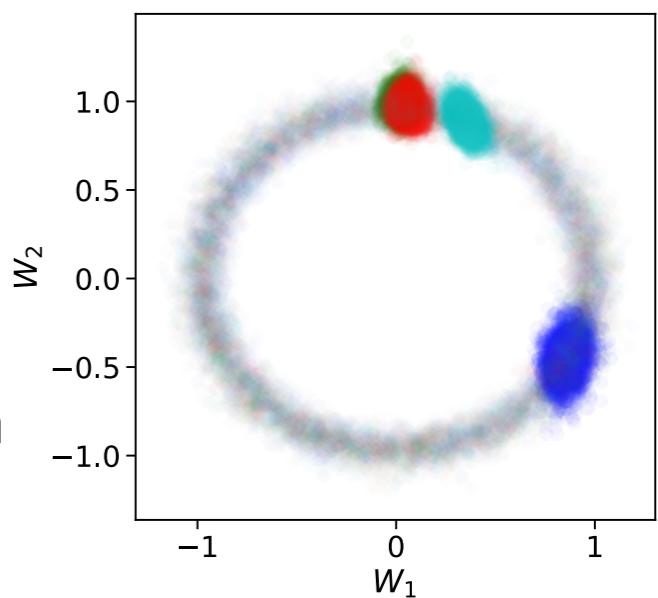


Iris

$(N, D) = (150, 4)$

Time House: 0.2 min

Time Standard: 0.8 min



# Results

**for**  $n = D : 1$

$$\boldsymbol{v}_n \sim \text{uniform on } \mathbb{S}^{n-1}$$

$$\boldsymbol{u}_n = \frac{\boldsymbol{v}_n + \text{sgn}(\boldsymbol{v}_{n1}) \|\boldsymbol{v}_n\| \boldsymbol{e}_1}{\|\boldsymbol{v}_n + \text{sgn}(\boldsymbol{v}_{n1}) \|\boldsymbol{v}_n\| \boldsymbol{e}_1\|}$$

$$\tilde{\boldsymbol{H}}_n (\boldsymbol{v}_n) = -\text{sgn}(\boldsymbol{v}_{n1}) (\boldsymbol{I} - 2\boldsymbol{u}_n \boldsymbol{u}_n^T)$$

$$\boldsymbol{H}_n = \begin{pmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{H}}_n \end{pmatrix}$$

$$\boldsymbol{U} = \boldsymbol{H}_D (\boldsymbol{v}_D) \boldsymbol{H}_{D-1} (\boldsymbol{v}_{D-1}) \dots \boldsymbol{H}_1 (\boldsymbol{v}_1)$$

# Results

**for**  $n = D : 1$

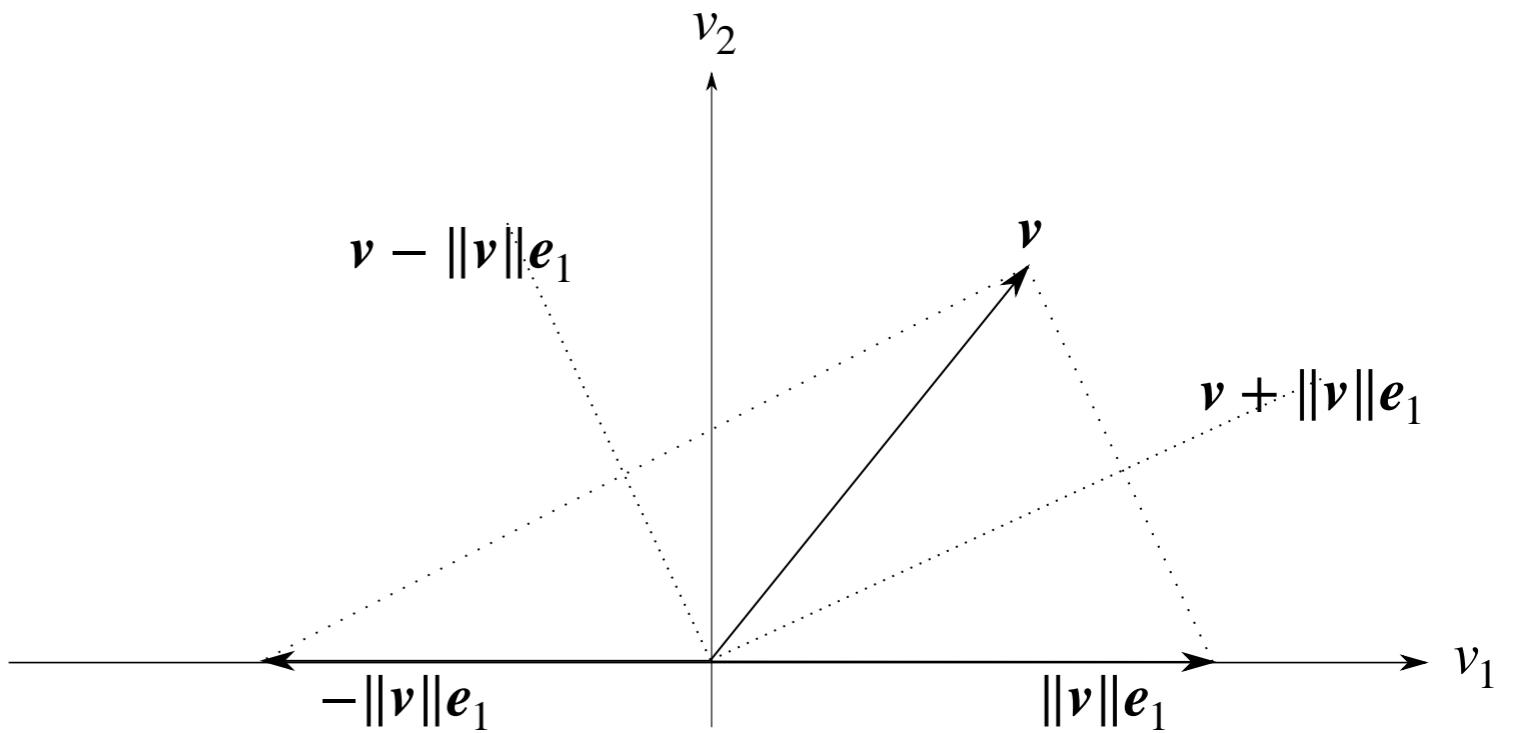
$v_n \sim \text{uniform on } \mathbb{S}^{n-1}$

$$u_n = \frac{v_n + \text{sgn}(v_{n1}) \|v_n\| e_1}{\|v_n + \text{sgn}(v_{n1}) \|v_n\| e_1\|}$$

$$\tilde{H}_n(v_n) = -\text{sgn}(v_{n1}) (I - 2u_n u_n^T)$$

$$H_n = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \tilde{H}_n \end{pmatrix}$$

$$U = H_D(v_D) H_{D-1}(v_{D-1}) \dots H_1(v_1)$$



# Results

**for**  $n = D : 1$

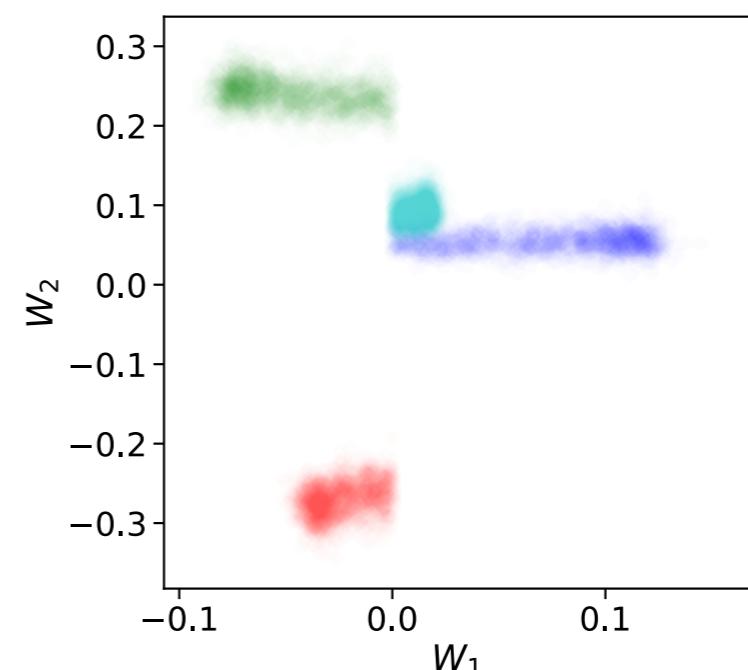
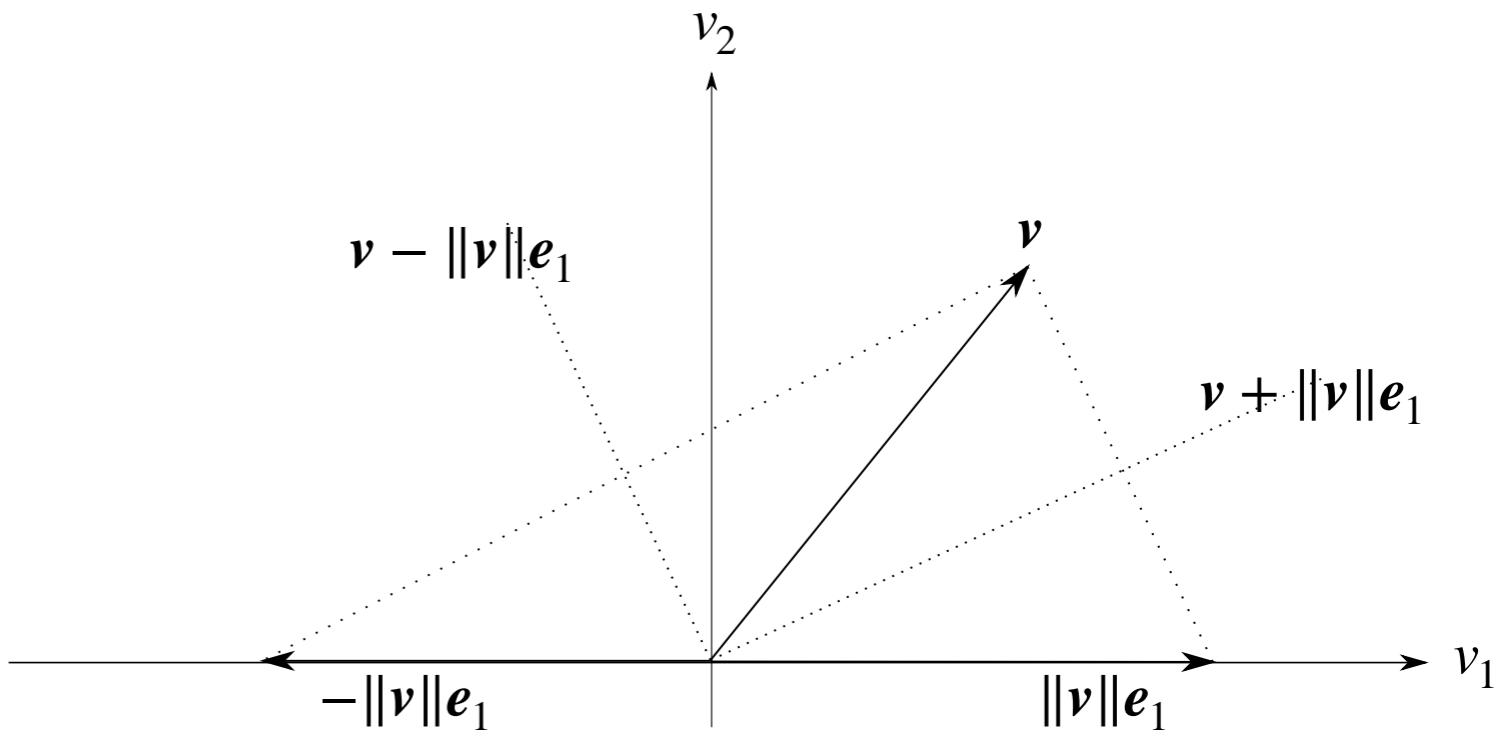
$v_n \sim \text{uniform on } \mathbb{S}^{n-1}$

$$u_n = \frac{v_n + \text{sgn}(v_{n1}) \|v_n\| e_1}{\|v_n + \text{sgn}(v_{n1}) \|v_n\| e_1\|}$$

$$\tilde{H}_n(v_n) = -\text{sgn}(v_{n1}) (\mathbf{I} - 2u_n u_n^T)$$

$$H_n = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \tilde{H}_n \end{pmatrix}$$

$$U = H_D(v_D) H_{D-1}(v_{D-1}) \dots H_1(v_1)$$



# Extension to non-linear models

- GPLVM with the same rotation invariant problem

$$p(Y|X) = \prod_{d=1}^D \mathcal{N}(Y_{:,d}|\boldsymbol{\mu}, \mathbf{K} + \sigma^2 I)$$

$$\mathbf{K} = \mathbf{X}\mathbf{X}^T, \quad K_{ij} = \mathbf{X}_{i,:}^T \mathbf{X}_{j,:} = k\left(\mathbf{X}_{i,:}, \mathbf{X}_{j,:}\right)$$

$$k_{\text{SE}}(\mathbf{x}, \mathbf{x}') = \sigma_{\text{SE}}^2 \exp\left(-0.5 \left\| \mathbf{x} - \mathbf{x}' \right\|_2^2 / l^2\right)$$

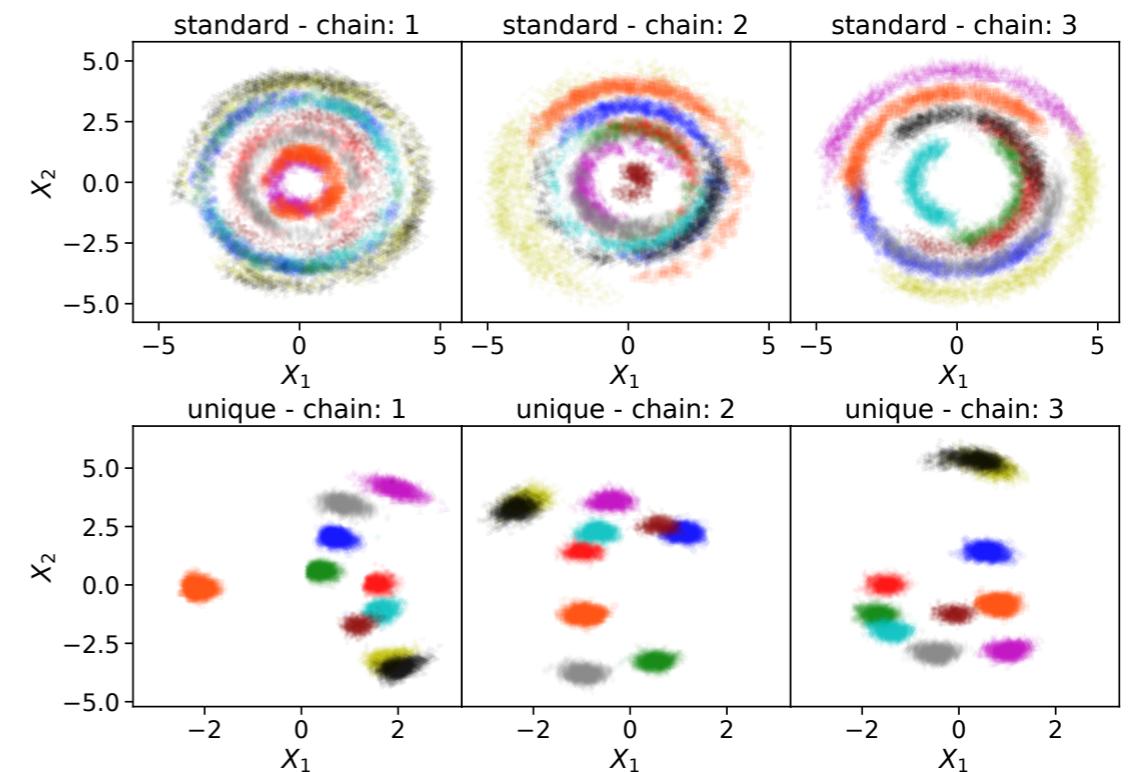
# Extension to non-linear models

- GPLVM with the same rotation invariant problem

$$p(Y|X) = \prod_{d=1}^D \mathcal{N}(Y_{:,d}|\boldsymbol{\mu}, \mathbf{K} + \sigma^2 I)$$

$$\mathbf{K} = \mathbf{X}\mathbf{X}^T, \quad K_{ij} = \mathbf{X}_{i,:}^T \mathbf{X}_{j,:} = k\left(\mathbf{X}_{i,:}, \mathbf{X}_{j,:}\right)$$

$$k_{\text{SE}}(\mathbf{x}, \mathbf{x}') = \sigma_{\text{SE}}^2 \exp\left(-0.5 \left\| \mathbf{x} - \mathbf{x}' \right\|_2^2 / l^2\right)$$



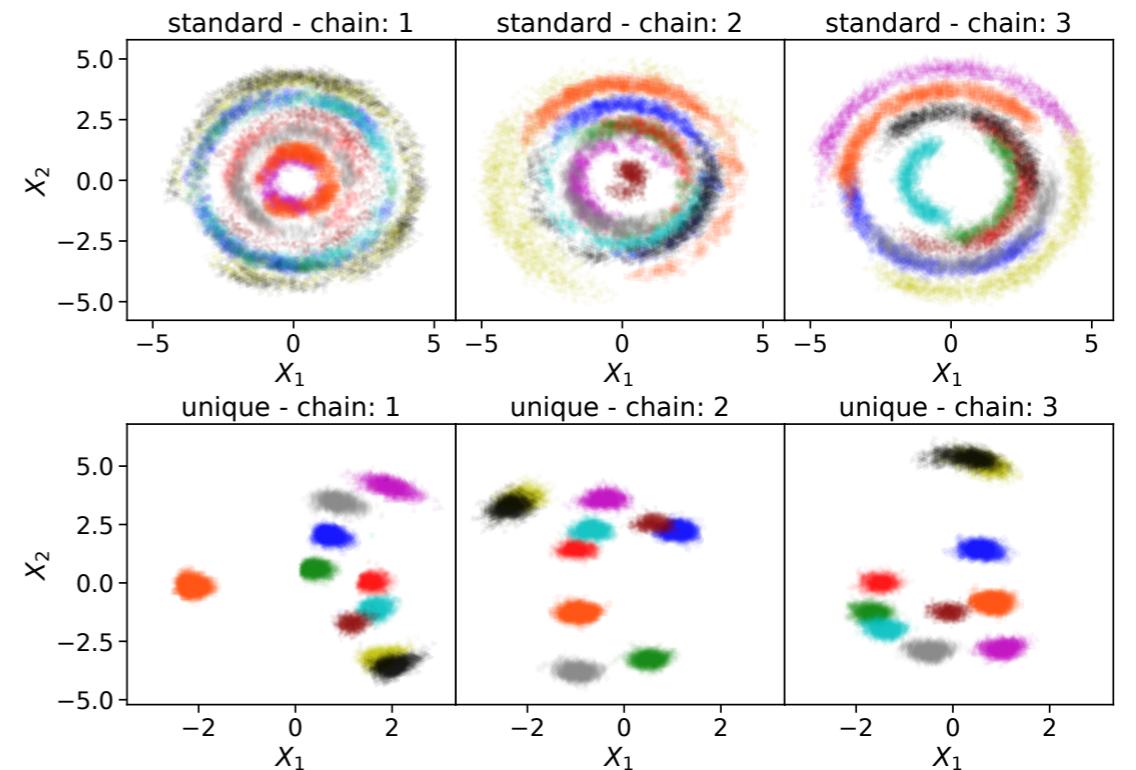
# Extension to non-linear models

- GPLVM with the same rotation invariant problem

$$p(Y|X) = \prod_{d=1}^D \mathcal{N}(Y_{:,d}|\boldsymbol{\mu}, \mathbf{K} + \sigma^2 I)$$

$$\mathbf{K} = \mathbf{X}\mathbf{X}^T, \quad K_{ij} = \mathbf{X}_{i,:}^T \mathbf{X}_{j,:} = k\left(\mathbf{X}_{i,:}, \mathbf{X}_{j,:}\right)$$

$$k_{\text{SE}}(\mathbf{x}, \mathbf{x}') = \sigma_{\text{SE}}^2 \exp\left(-0.5 \left\| \mathbf{x} - \mathbf{x}' \right\|_2^2 / l^2\right)$$



- No rotation symmetry in the posterior for the suggested parameterization
- Different chains converge to different solutions due to increased model complexity

# Conclusion

- Suggested new parameterization for  $\mathbf{W}$  in PPCA, which uniquely identifies principle components even though the likelihood is rotationally symmetric
- Showed how to set the prior on the new parameters such that the model is not changed compared to a standard Gaussian prior on  $\mathbf{W}$
- Provided an efficient implementation via Householder transformations (no Jacobian correction needed)
- New parameterization allows for other interpretable priors on rotation and principle variances
- Extended to non-linear models and successfully solved the rotation problem there as well

**Thanks for your  
attention!**

**Supervisor: Prof. Dr. Nils Bertschinger**