# Stiefel Manifolds and their Applications

Pierre-Antoine Absil
(UCLouvain)

Seminar for Applied Mathematics, ETH Zurich

10 September 2009

## Structure

- Definition and visualization
- A glimpse of applications
- Geometry of the Stiefel manifolds
- Applications

# Collaborations

- Chris Baker (Sandia)
- Thomas Cason (UCLouvain)
- Kyle Gallivan (Florida State University)
- Damien Laurent (UCLouvain)
- Rob Mahony (Australian National University)
- Chafik Samir (U Clermont-Ferrand)
- Rodolphe Sepulchre (U of Liège)
- Fabian Theis (TU Munich)
- Paul Van Dooren (UCLouvain)
- ...

# Stiefel manifold: Definition

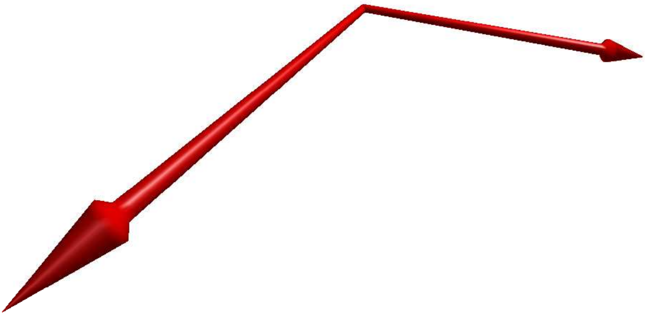
The (compact) Stiefel manifold $V_{n,p}$ is the set of all $p$-tuples $(x_1, \dots, x_p)$ of orthonormal vectors in $\mathbb{R}^n$.
If we turn $p$-tuples into $n \times p$ matrices as follows

$$(x_1, \dots, x_p) \mapsto \begin{bmatrix} x_1 & \cdots & x_p \end{bmatrix},$$

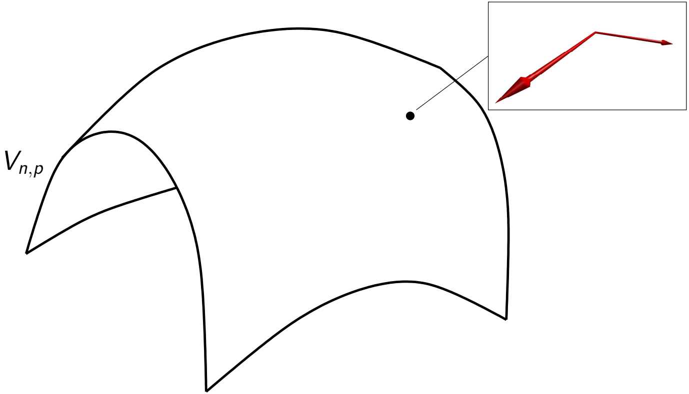
the definition becomes

$$V_{n,p} = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\}.$$

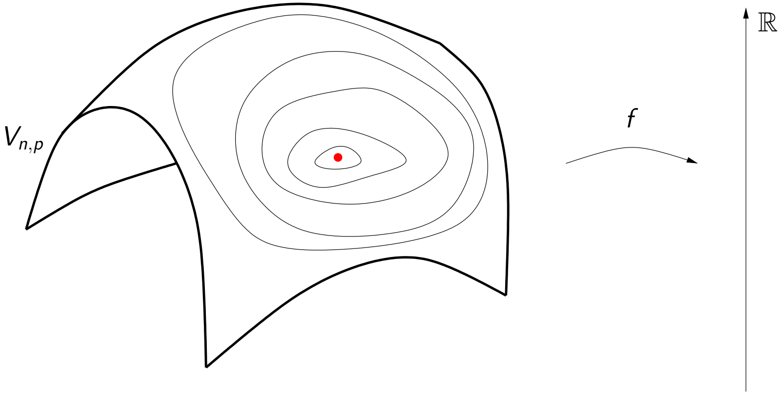
# Visualization: an element of $V_{3,2}$



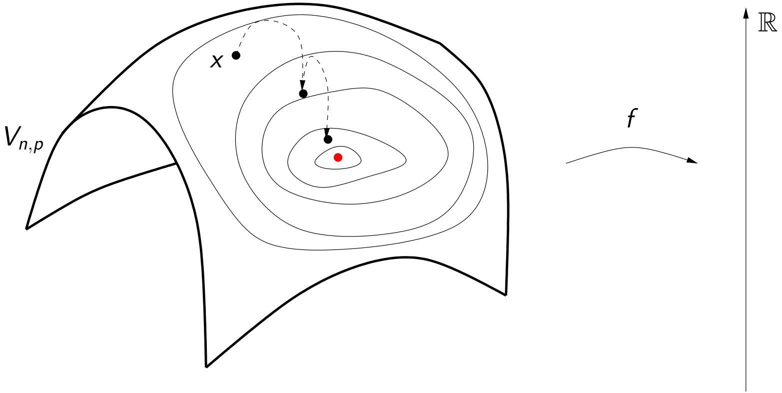
# Stiefel manifold: (very unfaithful) artist view



# Stiefel manifold: optimization problems



# Stiefel manifold: optimization algorithms



## Stiefel manifold: Extensions

- Recall: Real case:

$$V_p(\mathbb{R}^n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\} =: V_{n,p}.$$

- Complex case:

$$V_p(\mathbb{C}^n) = \{X \in \mathbb{C}^{n \times p} : X^H X = I_p\}.$$

- Quaternion case:

$$V_p(\mathbb{H}^n) = \{X \in \mathbb{H}^{n \times p} : X^* X = I_p\}.$$

- If $M$ is a Riemannian manifold, one can define

$$V_p(TM) = \{(\xi_1, \dots, \xi_p) | \exists x \in M : \xi_i \in T_x M, \langle \xi_i, \xi_j \rangle = \delta_{ij}\}.$$

# Stiefel manifold: Particular cases

- Recall: Real case:

$$V_p(\mathbb{R}^n) = \{X \in \mathbb{R}^{n\times p} : X^T X = I_p\} =: V_{n,p}.$$

- $p = 1$: the sphere

$$V_{n,1} = \{x \in \mathbb{R}^n : x^T x = 1\}.$$

- $p = n$: the orthogonal group

$$V_{n,n} = O_n = \{X \in \mathbb{R}^{n\times n} : X^T X = I_n\}.$$

## Notation

- E. Stiefel (1935): $V_{n,m}$ (compact), $V^*_{n,m}$ (noncompact).
- I. M. James (1976): $O_{n,k}$ (compact) Stiefel manifold, $O^*_{n,k}$ noncompact Stiefel manifold, $V_{n,k}$ in the real case, $W_{n,k}$ in the complex case, $X_{n,k}$ in the quaternion case.
- Helmke & Moore (1994): $\mathrm{St}(k,n)$ compact Stiefel manifold, $\mathrm{ST}(k,n)$ noncompact Stiefel manifold.
- Edelman, Arias, & Smith (1998): $V_{n,p}$.
- Bridges & Reich (2001): $V_k(\mathbb{R}^n)$.
- Bloch *et al.*(2006): $V(n,N) = \{Q \in \mathbb{R}^{nN}; QQ^T = I_n\}$.

## A glimpse of applications

- Principal component analysis
- Lyapunov exponents of a dynamical system
- Procrustes problem
- Blind Source Separation - soft dimension reduction

# Geometry

- Dimension
- Tangent spaces
- Projection onto tangent spaces
- Geodesics

## Stiefel manifold: dimension

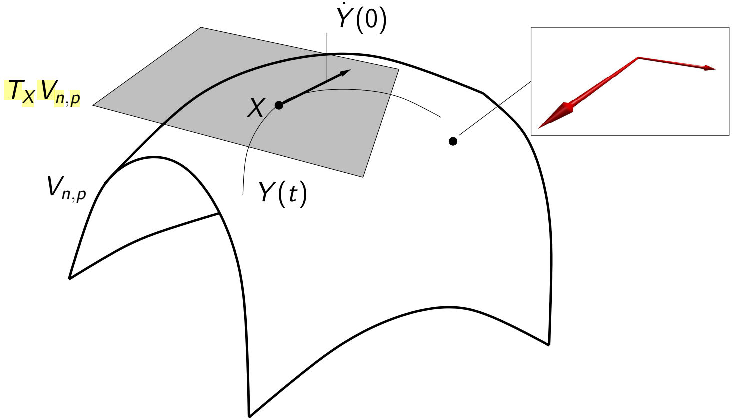
Dimension of $V_{n,p}$:

- 1st vector: one unit-norm constraint: $n-1$ DOF.
- 2nd vector: unit-norm and orthogonal to 1st: $n-2$ DOF.
- ...
- $p$th vector: $n-p$ DOF.

Total:

$$\begin{aligned}\dim(V_{n,p}) &= pn - (1 + 2 + \cdots + p)\\ &= pn - p(p+1)/2\\ &= p(n-p) + p(p-1)/2.\end{aligned}$$

# Stiefel manifold: tangent space



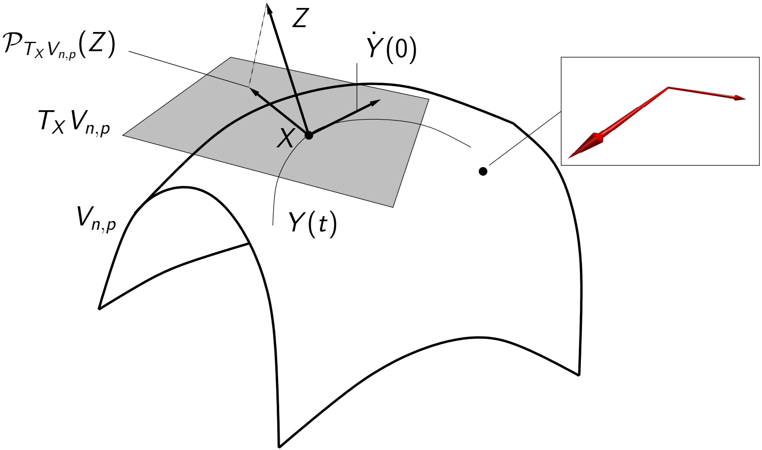
# Stiefel manifold: tangent space

Let $X \in V_{n,p}$ and let $Y(t)$ be a curve on $V_{n,p}$ with $Y(0) = X$.
Then $\dot{Y}(0)$ is a *tangent vector* to $V_{n,p}$ at $X$.
The set of all such vectors is the *tangent space* to $V_{n,p}$ at $X$.
We have

$$Y(t)^T Y(t) = I_p \quad \text{for all } t$$
$$\frac{\mathrm{d}}{\mathrm{d}t}(Y(t)^T Y(t)) = 0 \quad \text{for all } t$$
$$\dot{Y}(0)^T Y(0) + Y(0)^T \dot{Y}(0) = 0$$
$$X^T \dot{Y}(0) \text{ is skew}$$
$$\dot{Y}(0) = X\Omega + X_\perp K, \;\; \Omega^T = -\Omega.$$

Hence $T_X V_{n,p} = \{X\Omega + X_\perp K : \Omega^T = -\Omega, K \in \mathbb{R}^{(n-p)\times p}\}$.

# Stiefel manifold: projection onto the tangent space



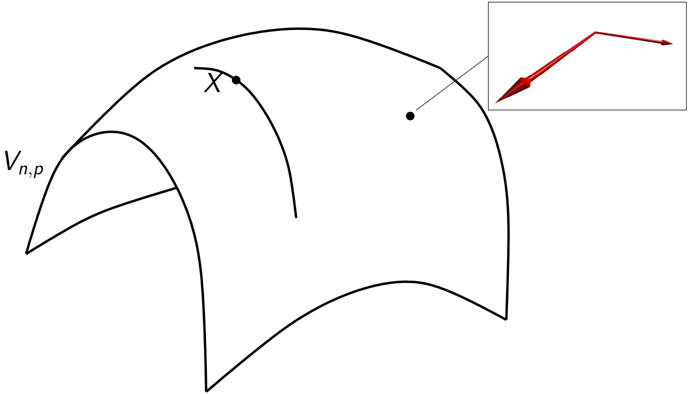
# Stiefel manifold: projection onto the tangent space

- Tangent space: $T_X V_{n,p} = \{X\Omega + X_\perp K : \Omega^T = -\Omega, K \in \mathbb{R}^{(n-p)\times p}\}$.
- Normal space: $N_X V_{n,p} = \{XS : S^T = S\}$.
- Projection onto the tangent space:

$$\begin{aligned}\mathcal{P}_{T_X V_{n,p}}(Z) &= Z - X\,\mathrm{sym}(X^T Z)\\ &= (I - XX^T)Z + X\,\mathrm{skew}(X^T Z),\end{aligned}$$

where $\mathrm{sym}(M) = \frac{1}{2}(M + M^T)$ and $\mathrm{skew}(M) = \frac{1}{2}(M - M^T)$.

# Stiefel manifold: geodesics



# Stiefel manifold: geodesics

A curve $X(t)$ on $V_{n,p}$ is a *geodesic* if, for all $t$,

$$\ddot{X}(t) \in N_{X(t)} V_{n,p}.$$

Ross Lippert showed that

$$X(t) = \begin{bmatrix} X(0) & \dot{X}(0) \end{bmatrix} \exp t \begin{bmatrix} X(0)^T \dot{X}(0) & -\dot{X}(0)^T \dot{X}(0) \\ I & X(0)^T \dot{X}(0) \end{bmatrix} I_{2p,p} e^{-tX(0)^T \dot{X}(0)}.$$

# Stiefel manifold: quotient geodesics

Bijection between $V_{n,p}$ and $O_n/O_{n-p}$:

$$V_{n,p} \ \ni \ X \leftrightarrow \{\overbrace{\begin{bmatrix} X & X_\perp \end{bmatrix}}^{U} : U^T U = I_n\} \ \in \ O_n/O_{n-p}$$

Quotient geodesics: If

$$U(t) = U(0) \exp t \begin{bmatrix} A & -B^T \\ B & 0 \end{bmatrix}.$$

then $U_{:,1:p}(t) \in V_{n,p}$ follows a *quotient geodesic*.

# Applications

- Principal component analysis
- Lyapunov exponents of a dynamical system
- Procrustes problem
- Blind Source Separation - soft dimension reduction

# Principal component analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the $p$ dominant eigenvectors of $A$.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^TX=I_p} \mathrm{tr}(X^TAXN).$$

  The columns of $X$ are the $p$ dominant eigenvectors or $A$.
- A basic method: Steepest-descent on $V_{n,p}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{tr}(X^TAXN)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{V_{n,p}}(X) = \mathcal{P}_{T_X V_{n,p}}(AXN) = AXN - X\mathrm{sym}(X^TAXN)$, where $\mathrm{sym}(Z) := (Z+Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{V_{n,p}}(X)$.

## Principal component analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the $p$ dominant eigenvectors of $A$.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^T X = I_p} \mathrm{tr}(X^T A X N).$$

The columns of $X$ are the $p$ dominant eigenvectors or $A$.

- A basic method: Steepest-descent on $V_{n,p}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{tr}(X^T A X N)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{V_{n,p}}(X) = \mathcal{P}_{T_X V_{n,p}}(AXN) = AXN - X\mathrm{sym}(X^T AXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{V_{n,p}}(X)$.

## Principal component analysis

- Let $A = A^T \in \mathbb{R}^{n \times n}$.
- Goal: Compute the $p$ dominant eigenvectors of $A$.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^T X = I_p} \mathrm{tr}(X^T A X N).$$

The columns of $X$ are the $p$ dominant eigenvectors or $A$.

- A basic method: Steepest-descent on $V_{n,p}$.
- Let $f : \mathbb{R}^{n \times p} \to \mathbb{R} : X \mapsto \mathrm{tr}(X^T A X N)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{V_{n,p}}(X) = \mathcal{P}_{T_X V_{n,p}}(AXN) = AXN - X\mathrm{sym}(X^T AXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{V_{n,p}}(X)$.

# Principal component analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the $p$ dominant eigenvectors of $A$.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^T X = I_p} \mathrm{tr}(X^T A X N).$$

  The columns of $X$ are the $p$ dominant eigenvectors or $A$.
- A basic method: Steepest-descent on $V_{n,p}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{tr}(X^T A X N)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{V_{n,p}}(X) = \mathcal{P}_{T_X V_{n,p}}(AXN) = AXN - X\mathrm{sym}(X^T AXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{V_{n,p}}(X)$.

# Principal component analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the $p$ dominant eigenvectors of $A$.
- Principle: Let $N = \operatorname{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^T X = I_p} \operatorname{tr}(X^T A X N).$$

  The columns of $X$ are the $p$ dominant eigenvectors or $A$.
- A basic method: Steepest-descent on $V_{n,p}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \operatorname{tr}(X^T A X N)$.
- We have $\frac{1}{2}\operatorname{grad} f(X) = AXN$.
- Thus $\frac{1}{2}\operatorname{grad} f|_{V_{n,p}}(X) = \mathcal{P}_{T_X V_{n,p}}(AXN) = AXN - X\operatorname{sym}(X^T AXN)$, where $\operatorname{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \operatorname{grad} f|_{V_{n,p}}(X)$.

## Principal component analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the $p$ dominant eigenvectors of $A$.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^T X = I_p} \mathrm{tr}(X^T A X N).$$

The columns of $X$ are the $p$ dominant eigenvectors or $A$.

- A basic method: Steepest-descent on $V_{n,p}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{tr}(X^T A X N)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{V_{n,p}}(X) = \mathcal{P}_{T_X V_{n,p}}(AXN) = AXN - X\mathrm{sym}(X^T AXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{V_{n,p}}(X)$.

# Computing Lyapunov exponents: a method on the Stiefel manifold

- Ref: T. Bridges and S. Reich, *Computing Lyapunov exponents on a Stiefel manifold*, Physica D 156, pp. 219–238, 2001.
- Dynamical system: $\dot{x} = f(x)$.
- Nominal trajectory: $x_*(t)$.
- Goal: Describe the behavior of nearby trajectories.

# Computing Lyapunov exponents: a method on the Stiefel manifold

- Dynamical system: $\dot{x} = f(x)$.
- Nominal trajectory: $x_*(t)$.
- Goal: Describe the behavior of nearby trajectories.

## Computing Lyapunov exponents: a method on the Stiefel manifold

- Dynamical system: $\dot{x} = f(x)$.
- Nominal trajectory: $x_*(t)$.
- Goal: Describe the behavior of nearby trajectories.
- Principle: Consider the evolution of an infinitesimal ball of perturbed initial conditions. The ball becomes distorted into an infinitesimal ellipsoid. If the length $\delta_k(t)$ of the $k$th principal axis evolves as

$$\delta_k(t) \approx \delta_k(0) e^{\lambda_k t},$$

then $\lambda_k$ is the $k$th *Lyapunov exponent* of the system along the nominal trajectory.
The mean Lyapunov exponents are given by

$$\lambda_k = \lim_{t \to \infty} \frac{1}{t} \frac{\|\delta_k(t)\|}{\|\delta_k(0)\|}.$$

# Computing Lyapunov exponents: a method on the Stiefel manifold

- Dynamical system: $\dot{x} = f(x)$.
- Nominal trajectory: $x_*(t)$.
- Goal: Describe the behavior of nearby trajectories.
- Principle: $\delta_k(t) \approx \delta_k(0)e^{\lambda_k t}$.
- Challenge 1: Compute just a few Lyapunov exponents $\rightsquigarrow$ work with $p$-frames (noncompact Stiefel manifold).
- Perturbed system:

$$\dot{Z} = A(t)Z, \quad Z \in \mathbb{R}^{n\times p}, \quad A(t) := \mathrm{D}f(x_*(t)). \tag{1}$$

## Computing Lyapunov exponents: a method on the Stiefel manifold

- Dynamical system: $\dot{x} = f(x)$.
- Nominal trajectory: $x_*(t)$.
- Goal: Describe the behavior of nearby trajectories.
- Principle: $\delta_k(t) \approx \delta_k(0)e^{\lambda_k t}$.
- Challenge 1: Compute just a few Lyapunov exponents $\rightsquigarrow$ work with $p$-frames (noncompact Stiefel manifold).
- Perturbed system:

$$\dot{Z} = A(t)Z, \quad Z \in \mathbb{R}^{n \times p}, \quad A(t) := \mathrm{D}f(x_*(t)). \tag{1}$$

# Computing Lyapunov exponents: a method on the Stiefel manifold

- Dynamical system: $\dot{x} = f(x)$.
- Nominal trajectory: $x_*(t)$.
- Goal: Describe the behavior of nearby trajectories.
- Principle: $\delta_k(t) \approx \delta_k(0)e^{\lambda_k t}$.
- Challenge 1: Compute just a few Lyapunov exponents $\rightsquigarrow$ work with $p$-frames (noncompact Stiefel manifold).
- Perturbed system:

$$\dot{Z} = A(t)Z, \quad Z \in \mathbb{R}^{n\times p}, \quad A(t) := \mathrm{D}f(x_*(t)). \tag{1}$$

## Computing Lyapunov exponents: a method on the Stiefel manifold

- Dynamical system: $\dot{x} = f(x)$.
- Nominal trajectory: $x_*(t)$.
- Goal: Describe the behavior of nearby trajectories.
- Principle: $\delta_k(t) \approx \delta_k(0)e^{\lambda_k t}$.
- Perturbed system:

$$\dot{Z} = A(t)Z, \quad Z \in \mathbb{R}^{n\times p}, \quad A(t) := \mathrm{D}f(x_*(t)). \tag{1}$$

- Challenge 2: Perform continuous orthogonalization to prevent the columns of $Z$ from converging to 1st Lyapunov vector $\rightsquigarrow V_{n,p}$.
- Method: Follow the evolution of $Q(t)$ in the thin QR decomposition

$$Z(t) = Q(t)R(t).$$

## Computing Lyapunov exponents: a method on the Stiefel manifold

- Dynamical system: $\dot{x} = f(x)$.
- Nominal trajectory: $x_*(t)$.
- Goal: Describe the behavior of nearby trajectories.
- Principle: $\delta_k(t) \approx \delta_k(0)e^{\lambda_k t}$.
- Perturbed system:

$$\dot{Z} = A(t)Z, \quad Z \in \mathbb{R}^{n \times p}, \quad A(t) := \mathrm{D}f(x_*(t)). \tag{1}$$

- Challenge 2: Perform continuous orthogonalization to prevent the columns of $Z$ from converging to 1st Lyapunov vector $\rightsquigarrow V_{n,p}$.
- Method: Follow the evolution of $Q(t)$ in the thin QR decomposition

$$Z(t) = Q(t)R(t).$$

# Computing Lyapunov exponents: details of method on Stiefel

- Perturbed system: $\dot{Z} = A(t)Z$, $Z \in \mathbb{R}^{n\times p}$, $A(t) := \mathrm{D}f(x_*(t))$.
- Principle: track $Q(t)$ in $Z(t) = Q(t)R(t)$.
- We have the formula

$$\dot{Q} = (I - QQ^T)\dot{Z}R^{-1} + QS, \quad S_{i,j} = \begin{cases} (Q^T\dot{Z}R^{-1})_{i,j}, & i > j \\ 0 & i = j \\ -(Q^T\dot{Z}R^{-1})_{j,i} & i < j. \end{cases}$$

- Hence

$$\dot{Q} = (I - QQ^T)A(t)Q + QS, \quad S_{i,j} = \begin{cases} (Q^TA(t)Q)_{i,j}, & i > j \\ 0 & i = j \\ -(Q^TA(t)Q)_{j,i} & i < j. \end{cases}$$

- This is a *dynamical system on the Stiefel manifold* $V_{n,p}$. It can be rewritten as

$$\dot{Q} = A(t)Q - QT,$$

where $T$ is upper triangular.
- For almost all initial $Q(0)$, the $p$ columns of $Q(t)$ converge to the $p$ leading Lyapunov vectors for the given trajectory of the given

## Computing Lyapunov exponents: details of method on Stiefel

- Perturbed system: $\dot{Z} = A(t)Z$, $Z \in \mathbb{R}^{n\times p}$, $A(t) := \mathrm{D}f(x_*(t))$.
- Principle: track $Q(t)$ in $Z(t) = Q(t)R(t)$.
- We have the formula

$$\dot{Q} = (I - QQ^T)\dot{Z}R^{-1} + QS, \quad S_{i,j} = \begin{cases} (Q^T\dot{Z}R^{-1})_{i,j}, & i > j \\ 0 & i = j \\ -(Q^T\dot{Z}R^{-1})_{j,i} & i < j. \end{cases}$$

- Hence

$$\dot{Q} = (I - QQ^T)A(t)Q + QS, \quad S_{i,j} = \begin{cases} (Q^TA(t)Q)_{i,j}, & i > j \\ 0 & i = j \\ -(Q^TA(t)Q)_{j,i} & i < j. \end{cases}$$

- This is a *dynamical system on the Stiefel manifold* $V_{n,p}$. It can be rewritten as

$$\dot{Q} = A(t)Q - QT,$$

where $T$ is upper triangular.
- For almost all initial $Q(0)$, the $p$ columns of $Q(t)$ converge to the $p$ leading Lyapunov vectors for the given trajectory of the given

# Computing Lyapunov exponents: details of method on Stiefel

- Perturbed system: $\dot{Z} = A(t)Z$, $Z \in \mathbb{R}^{n\times p}$, $A(t) := \mathrm{D}f(x_*(t))$.
- Principle: track $Q(t)$ in $Z(t) = Q(t)R(t)$.
- We have the formula

$$\dot{Q} = (I - QQ^T)\dot{Z}R^{-1} + QS, \quad S_{i,j} = \begin{cases} (Q^T\dot{Z}R^{-1})_{i,j}, & i > j \\ 0 & i = j \\ -(Q^T\dot{Z}R^{-1})_{j,i} & i < j. \end{cases}$$

- Hence

$$\dot{Q} = (I - QQ^T)A(t)Q + QS, \quad S_{i,j} = \begin{cases} (Q^TA(t)Q)_{i,j}, & i > j \\ 0 & i = j \\ -(Q^TA(t)Q)_{j,i} & i < j. \end{cases}$$

- This is a *dynamical system on the Stiefel manifold $V_{n,p}$*. It can be rewritten as

$$\dot{Q} = A(t)Q - QT,$$

where $T$ is upper triangular.
- For almost all initial $Q(0)$, the $p$ columns of $Q(t)$ converge to the $p$ leading Lyapunov vectors for the given trajectory of the given

# Computing Lyapunov exponents: details of method on Stiefel

- Perturbed system: $\dot{Z} = A(t)Z$, $Z \in \mathbb{R}^{n \times p}$, $A(t) := \mathrm{D}f(x_*(t))$.
- Principle: track $Q(t)$ in $Z(t) = Q(t)R(t)$.
- Hence

$$\dot{Q} = (I - QQ^T)A(t)Q + QS, \quad S_{i,j} = \begin{cases} (Q^T A(t) Q)_{i,j}, & i > j \\ 0 & i = j \\ -(Q^T A(t) Q)_{j,i} & i < j. \end{cases}$$

- This is a *dynamical system on the Stiefel manifold $V_{n,p}$*. It can be rewritten as

$$\dot{Q} = A(t)Q - QT,$$

where $T$ is upper triangular.
- For almost all initial $Q(0)$, the $p$ columns of $Q(t)$ converge to the $p$ leading Lyapunov vectors for the given trajectory of the given system. The corresponding Lyapunov exponents can be computed as

$$\lambda_j = \lim_{t \to \infty} \frac{1}{t} \int_0^t T_{jj}(s) \mathrm{d}s, \quad j = 1, \ldots, p.$$

## Computing Lyapunov exponents: details of method on Stiefel

- Perturbed system: $\dot{Z} = A(t)Z$, $Z \in \mathbb{R}^{n\times p}$, $A(t) := \mathrm{D}f(x_*(t))$.
- Principle: track $Q(t)$ in $Z(t) = Q(t)R(t)$.
- Hence

$$\dot{Q} = (I - QQ^T)A(t)Q + QS, \quad S_{i,j} = \begin{cases} (Q^T A(t) Q)_{i,j}, & i > j \\ 0 & i = j \\ -(Q^T A(t) Q)_{j,i} & i < j. \end{cases}$$

- This is a *dynamical system on the Stiefel manifold* $V_{n,p}$. It can be rewritten as

$$\dot{Q} = A(t)Q - QT,$$

  where $T$ is upper triangular.
- For almost all initial $Q(0)$, the $p$ columns of $Q(t)$ converge to the $p$ leading Lyapunov vectors for the given trajectory of the given system. The corresponding Lyapunov exponents can be computed as

$$\lambda_j = \lim_{t\to\infty} \frac{1}{t} \int_0^t T_{jj}(s)\mathrm{d}s, \quad j = 1, \ldots, p.$$

# Procrustes problem on the Stiefel manifold

- Ref: Lars Eldén and Haesun Park, *A Procrustes problem on the Stiefel manifold*, Numer. Math. (1999) 82: 599–619.
- *Orthogonal Procrustes problem*: given $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times p}$, find $Q \in \mathbb{R}^{n\times p}$ that solves

$$\min_{Q^TQ=I_p} \|AQ - B\|_F^2.$$

- First-order optimality condition *à la* manifold:
  Consider $f : \mathbb{R}^{n\times p} \to \mathbb{R} : Q \mapsto \|AQ - B\|_F^2$. We have

$$f(Q) = \mathrm{tr}(B^TB) + \mathrm{tr}(Q^TA^TAQ) - 2\mathrm{tr}(Q^TA^TB)$$

$$\mathrm{D}f(Q)[\dot{Q}] = -2\mathrm{tr}(\dot{Q}^TA^TB) + 2\mathrm{tr}(\dot{Q}^TA^TAQ),$$

$$\mathrm{grad}\, f(Q) = -2A^T(B - AQ),$$

$$\mathrm{grad}\, f|_{V_{n,p}}(Q) = \mathrm{grad}\, f(Q) - Q\mathrm{sym}(Q^T\mathrm{grad}\, f(Q)),$$

  where $\mathrm{sym}(A) := \frac{1}{2}(A + A^T)$.
- Case $p = n$: Then $f|_{V_{n,p}}(Q) = cst - 2\mathrm{tr}(Q^TA^TB)$, and the solution is given by

$$A^TB = U\Sigma V^T, \qquad Q = UV^T.$$

# Procrustes problem on the Stiefel manifold

- *Orthogonal Procrustes problem*: given $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times p}$. find $Q \in \mathbb{R}^{n\times p}$ that solves

$$\min_{Q^TQ=I_p} \|AQ - B\|_F^2.$$

- First-order optimality condition *à la* manifold:
  Consider $f : \mathbb{R}^{n\times p} \to \mathbb{R} : Q \mapsto \|AQ - B\|_F^2$. We have

$$\begin{aligned}
f(Q) &= \mathrm{tr}(B^TB) + \mathrm{tr}(Q^TA^TAQ) - 2\mathrm{tr}(Q^TA^TB) \\
\mathrm{D}f(Q)[\dot{Q}] &= -2\mathrm{tr}(\dot{Q}^TA^TB) + 2\mathrm{tr}(\dot{Q}^TA^TAQ), \\
\mathrm{grad}\, f(Q) &= -2A^T(B - AQ), \\
\mathrm{grad}\, f|_{V_{n,p}}(Q) &= \mathrm{grad}\, f(Q) - Q\mathrm{sym}(Q^T\mathrm{grad}\, f(Q)),
\end{aligned}$$

  where $\mathrm{sym}(A) := \frac{1}{2}(A + A^T)$.
- Case $p = n$: Then $f|_{V_{n,p}}(Q) = cst - 2\mathrm{tr}(Q^TA^TB)$, and the solution is given by

$$A^TB = U\Sigma V^T, \qquad Q = UV^T$$

  if $A^TB$ is invertible.

# Procrustes problem on the Stiefel manifold

- *Orthogonal Procrustes problem*: given $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times p}$. find $Q \in \mathbb{R}^{n \times p}$ that solves

$$\min_{Q^T Q = I_p} \|AQ - B\|_F^2.$$

- Applications: factor analysis, used notably in psychometrics.
- First-order optimality condition *à la* manifold:
  Consider $f : \mathbb{R}^{n \times p} \to \mathbb{R} : Q \mapsto \|AQ - B\|_F^2$. We have

$$\begin{aligned} f(Q) &= \mathrm{tr}(B^T B) + \mathrm{tr}(Q^T A^T A Q) - 2\mathrm{tr}(Q^T A^T B) \\ \mathrm{D} f(Q)[\dot{Q}] &= -2\mathrm{tr}(\dot{Q}^T A^T B) + 2\mathrm{tr}(\dot{Q}^T A^T A Q), \\ \mathrm{grad}\, f(Q) &= -2A^T(B - AQ), \\ \mathrm{grad}\, f|_{V_{n,p}}(Q) &= \mathrm{grad}\, f(Q) - Q\mathrm{sym}(Q^T \mathrm{grad}\, f(Q)), \end{aligned}$$

  where $\mathrm{sym}(A) := \frac{1}{2}(A + A^T)$.
- Case $p = n$: Then $f|_{V_{n,p}}(Q) = cst - 2\mathrm{tr}(Q^T A^T B)$, and the solution is given by

$$A^T B = U \Sigma V^T, \qquad Q = UV^T$$

  if $A^T B$ is invertible.

## Procrustes problem on the Stiefel manifold

- *Orthogonal Procrustes problem*:

$$\min_{Q^T Q = I_p} \|AQ - B\|_F^2.$$

- First-order optimality condition *à la* manifold:
  Consider $f : \mathbb{R}^{n\times p} \to \mathbb{R} : Q \mapsto \|AQ - B\|_F^2$. We have

$$\begin{aligned}
f(Q) &= \mathrm{tr}(B^T B) + \mathrm{tr}(Q^T A^T A Q) - 2\mathrm{tr}(Q^T A^T B) \\
\mathrm{D}f(Q)[\dot{Q}] &= -2\mathrm{tr}(\dot{Q}^T A^T B) + 2\mathrm{tr}(\dot{Q}^T A^T A Q), \\
\mathrm{grad}\, f(Q) &= -2A^T(B - AQ), \\
\mathrm{grad}\, f|_{V_{n,p}}(Q) &= \mathrm{grad}\, f(Q) - Q\mathrm{sym}(Q^T \mathrm{grad}\, f(Q)),
\end{aligned}$$

  where $\mathrm{sym}(A) := \frac{1}{2}(A + A^T)$.
- Case $p = n$: Then $f|_{V_{n,p}}(Q) = cst - 2\mathrm{tr}(Q^T A^T B)$, and the solution is given by

$$A^T B = U\Sigma V^T, \qquad Q = UV^T$$

  if $A^T B$ is invertible.

# Procrustes problem on the Stiefel manifold

- *Orthogonal Procrustes problem*:

$$\min_{Q^T Q = I_p} \|AQ - B\|_F^2.$$

- First-order optimality condition *à la* manifold:
  Consider $f : \mathbb{R}^{n\times p} \to \mathbb{R} : Q \mapsto \|AQ - B\|_F^2$. We have

$$\begin{aligned} f(Q) &= \mathrm{tr}(B^T B) + \mathrm{tr}(Q^T A^T A Q) - 2\mathrm{tr}(Q^T A^T B) \\ \mathrm{D}f(Q)[\dot{Q}] &= -2\mathrm{tr}(\dot{Q}^T A^T B) + 2\mathrm{tr}(\dot{Q}^T A^T A Q), \\ \mathrm{grad}\, f(Q) &= -2A^T(B - AQ), \\ \mathrm{grad}\, f|_{V_{n,p}}(Q) &= \mathrm{grad}\, f(Q) - Q\mathrm{sym}(Q^T \mathrm{grad}\, f(Q)), \end{aligned}$$

  where $\mathrm{sym}(A) := \frac{1}{2}(A + A^T)$.
- Case $p = n$: Then $f|_{V_{n,p}}(Q) = cst - 2\mathrm{tr}(Q^T A^T B)$, and the solution is given by

$$A^T B = U\Sigma V^T, \qquad Q = UV^T$$

  if $A^T B$ is invertible.

# Joint diagonalization on the Stiefel manifold

- Measurements $X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_f) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_f) \end{bmatrix}$.
- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of
$$Y = W^T X$$
look as statistically independent as possible.
- Decompose $W = U \Sigma V^T$. We have
$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$
- Whitening: Choose $\Sigma$ and $U$ such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose $V$ to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve

# Joint diagonalization on the Stiefel manifold

- Measurements $X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_f) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_f) \end{bmatrix}$.
- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of

$$Y = W^T X$$

  look as statistically independent as possible.
- Decompose $W = U\Sigma V^T$. We have

$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$

- Whitening: Choose $\Sigma$ and $U$ such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose $V$ to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve

# Joint diagonalization on the Stiefel manifold

- Measurements $X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_f) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_f) \end{bmatrix}$.
- Goal: Find a matrix $W \in \mathbb{R}^{n\times p}$ such that the rows of

$$Y = W^T X$$

  look as statistically independent as possible.
- Decompose $W = U\Sigma V^T$. We have

$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$

- Whitening: Choose $\Sigma$ and $U$ such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T\tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X})V$. Choose $V$ to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve

# Joint diagonalization on the Stiefel manifold

- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of
$$Y = W^T X$$
look as statistically independent as possible.
- Decompose $W = U \Sigma V^T$. We have
$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$
- Whitening: Choose $\Sigma$ and $U$ such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose $V$ to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve
$$\max_{V^T V = I_p} \sum_{i=1}^{N} \|\mathrm{diag}(V^T C_i(\tilde{X}) V)\|_F^2.$$

# Joint diagonalization on the Stiefel manifold

- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of
$$Y = W^T X$$
look as statistically independent as possible.
- Decompose $W = U \Sigma V^T$. We have
$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$
- Whitening: Choose $\Sigma$ and $U$ such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose $V$ to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve
$$\max_{V^T V = I_p} \sum_{i=1}^{N} \|\mathrm{diag}(V^T C_i(\tilde{X}) V)\|_F^2.$$

# Joint diagonalization on the Stiefel manifold

- Goal: Find a matrix $W \in \mathbb{R}^{n\times p}$ such that the rows of
$$Y = W^T X$$
look as statistically independent as possible.
- Decompose $W = U\Sigma V^T$. We have
$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$
- Whitening: Choose $\Sigma$ and $U$ such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T\tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose $V$ to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve
$$\max_{V^T V = I_p} \sum_{i=1}^{N} \|\mathrm{diag}(V^T C_i(\tilde{X}) V)\|_F^2.$$

# Joint diagonalization on the Stiefel manifold: application

The application is blind source separation.
Two mixed pictures are given as input to a blind source separation algorithm based on a trust-region method on $V_{2,2}$.

# Joint diagonalization on the Stiefel manifold: application: input

# Joint diagonalization on the Stiefel manifold: application: output

# Some References


- E. Stiefel, *Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten*, Comment. Math. Helv. 8, no. 1, 305–353, 1935.
- U. Helmke, J. B. Moore, *Optimization and Dynamical Systems*, Springer, 1994.
- A. Edelman, T. Arias, S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. 20(2), pp. 303–353, 1998.
- L. Eldén, H. Park, *A Procrustes problem on the Stiefel manifold*, Numer. Math. 82: 599–619, 1999.
- T. Bridges, S. Reich, *Computing Lyapunov exponents on a Stiefel manifold*, Physica D 156, pp. 219–238, 2001.
- Y. Nishimori, S. Akaho, *Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold*, Neurocomputing 67: 106–135, 2005.
- A. M. Bloch, P. E. Crouch, A. K. Sanyal, *A variational problem on Stiefel manifolds*, Nonlinearity 19, pp. 2247–2276, 2006.
- F. Tompkins, P. J. Wolfe, *Bayesian Filtering on the Stiefel Manifold*, 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007.
- P.-A. Absil, R. Mahony, R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2008.
- F. J. Theis, T. P. Cason, P.-A. Absil, *Soft Dimension Reduction for ICA by Joint Diagonalization on the Stiefel Manifold*, Proc. ICA 2009.