Nonlinear Signal Processing
2007-2008

Course Overview

Instituto Superior Técnico, Lisbon, Portugal

João Xavier

jxavier@isr.ist.utl.pt

Introduction

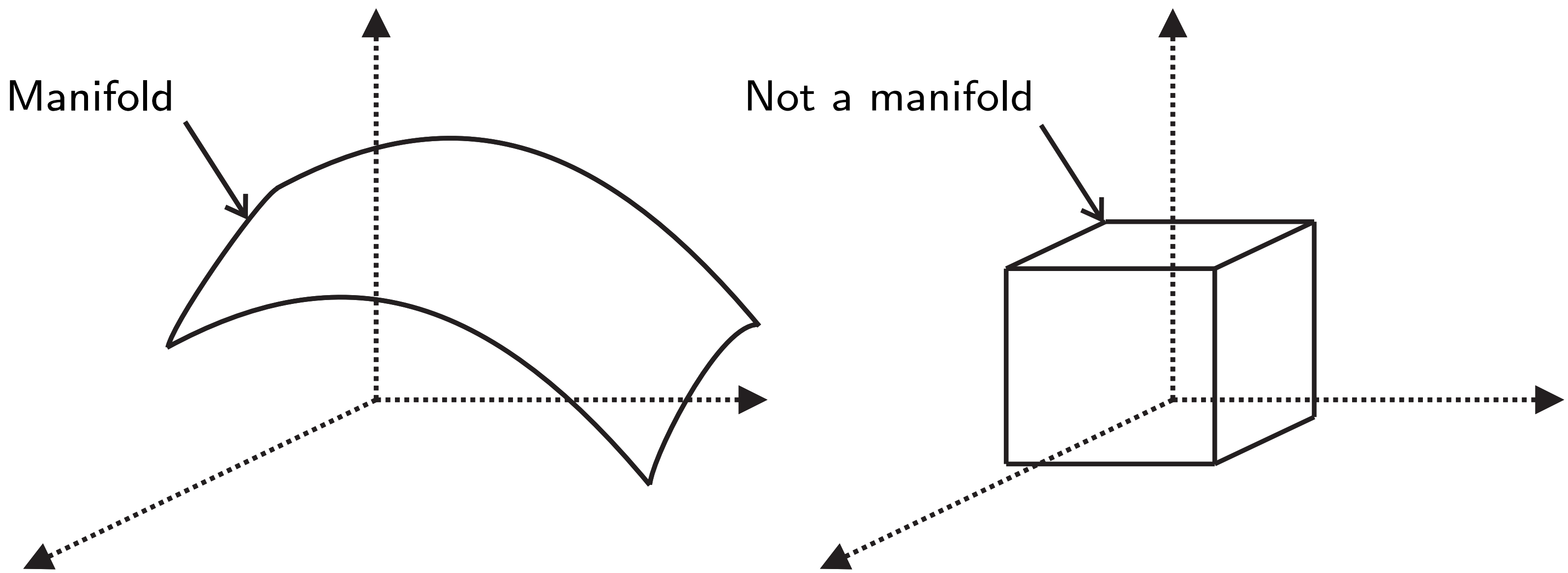
- This course is about applications of differential geometry in signal processing

- **What is differential geometry ?**
  - generalization of differential calculus to manifolds

- **What is a manifold ?**
  - smooth curved set
  - no vector space structure, no canonical coordinate system
  - looks <u>locally</u> like an Euclidean space, but not globally

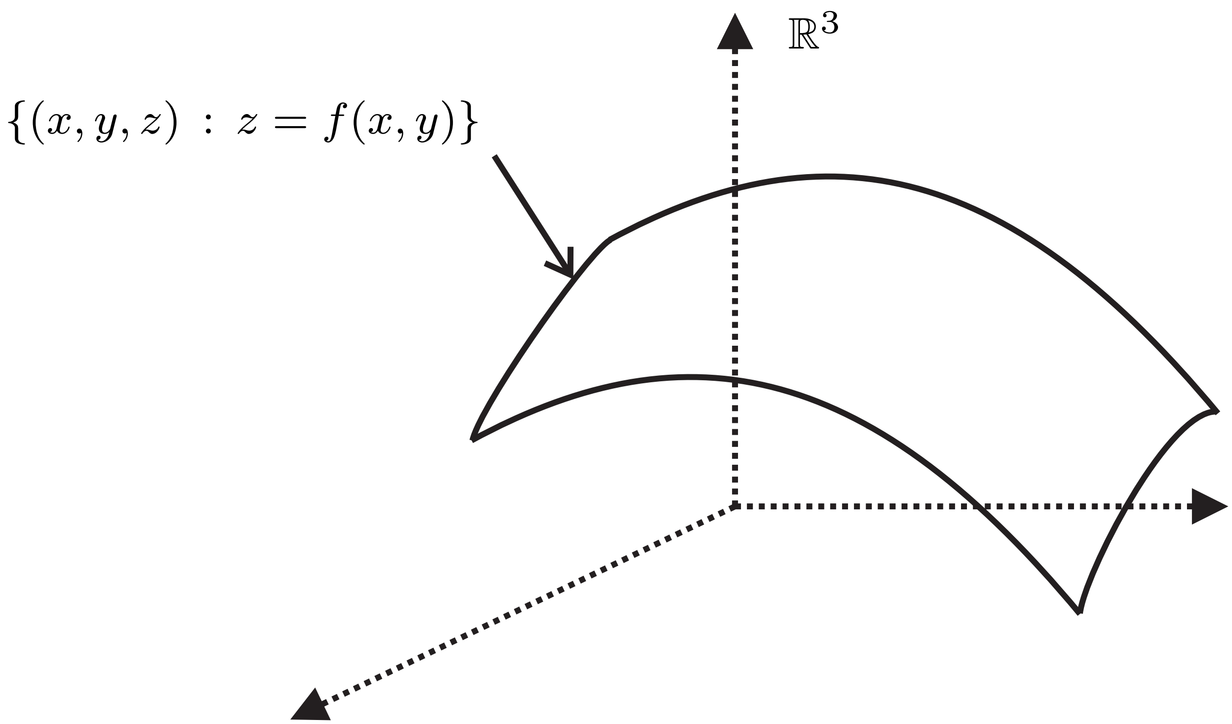
Introduction

- General idea



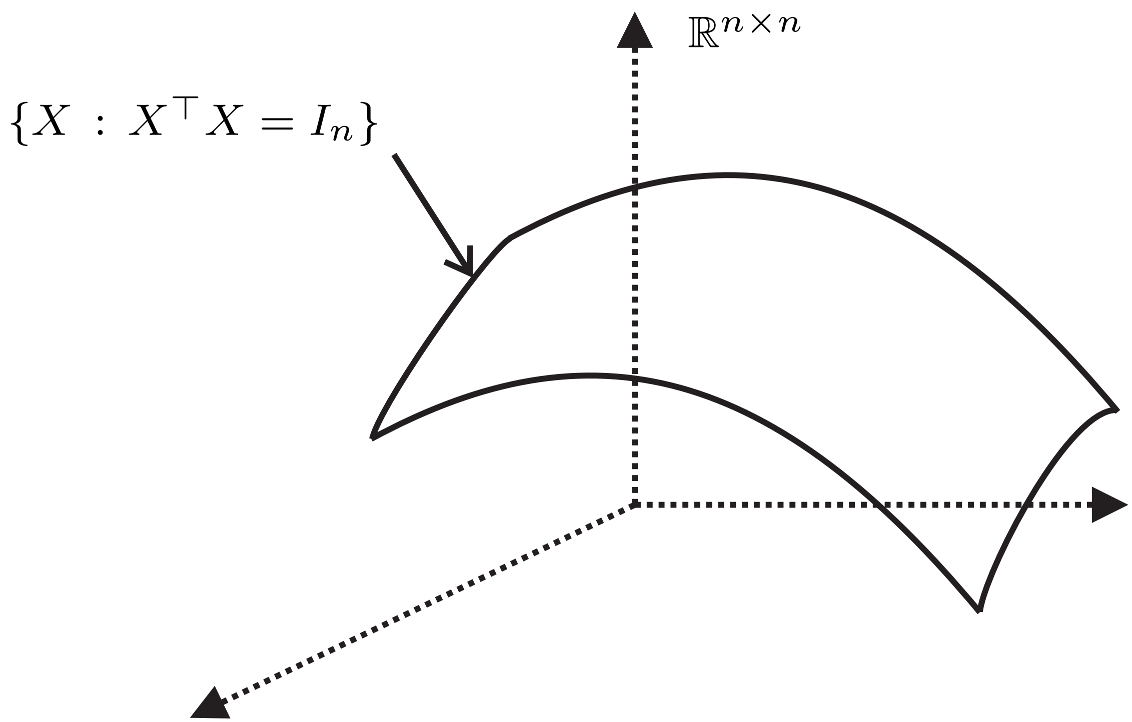
Introduction

- Example: graph of $f(x, y) = 1 - x^2 - y^2$



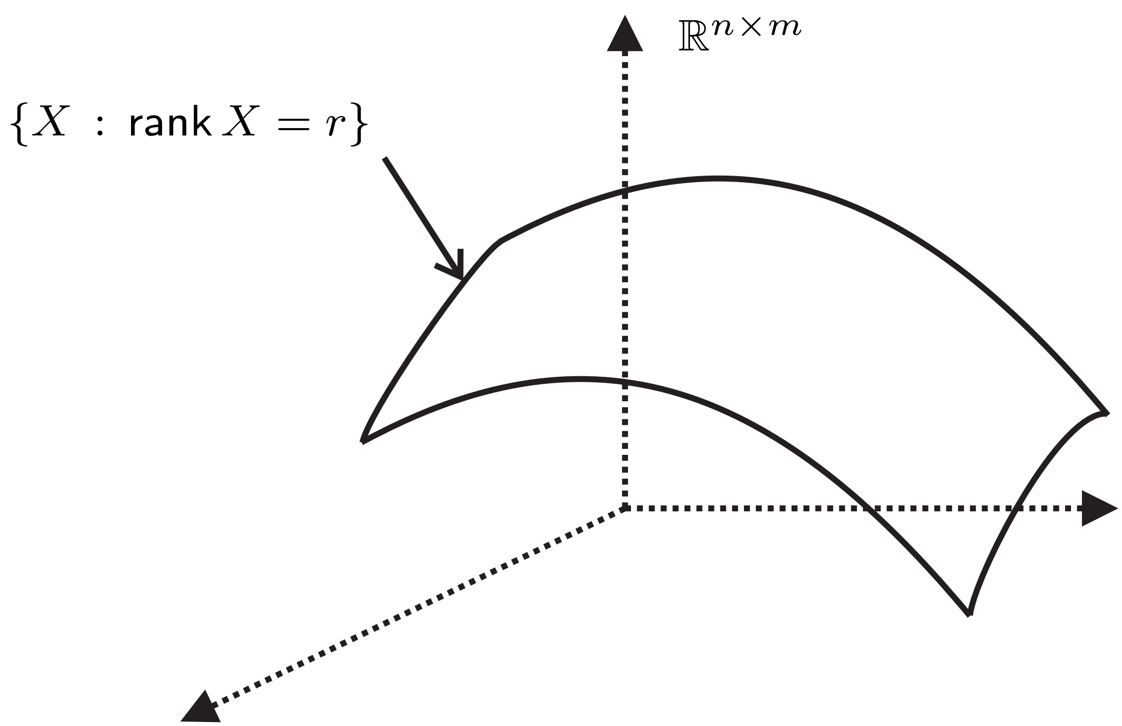
Introduction

- Example: $n \times n$ orthogonal matrices



Introduction

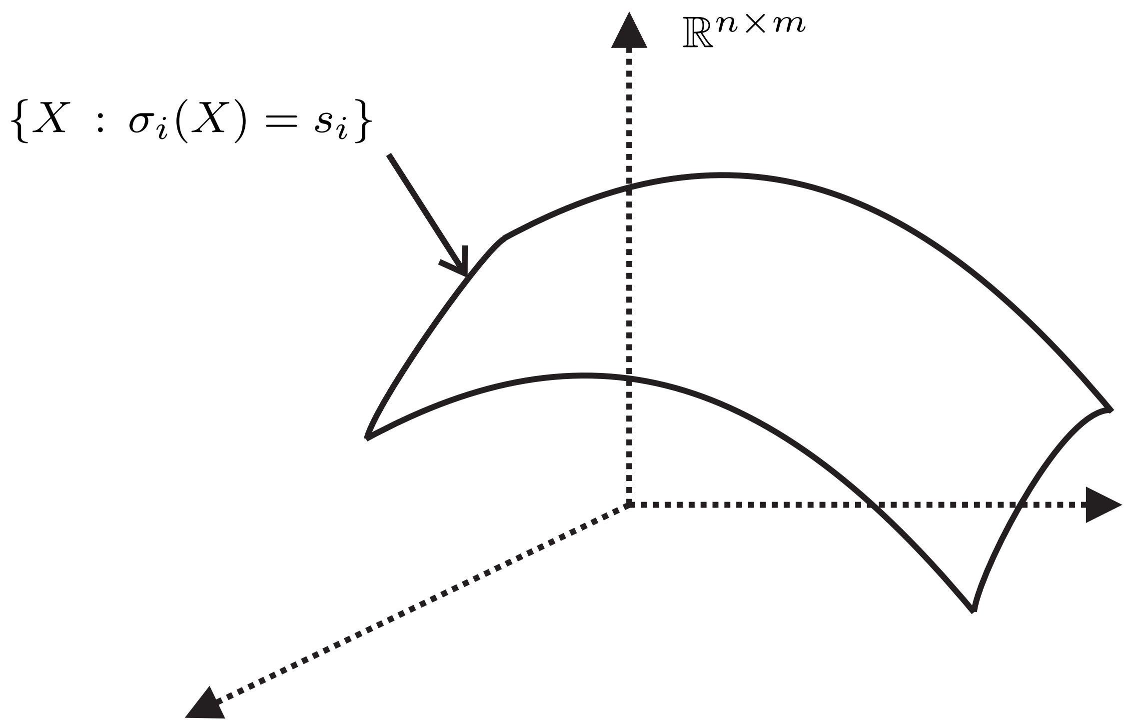
- Example: $n \times m$ matrices with rank r



- Note: $n \times m$ matrices with rank $\leq r$ is not a manifold

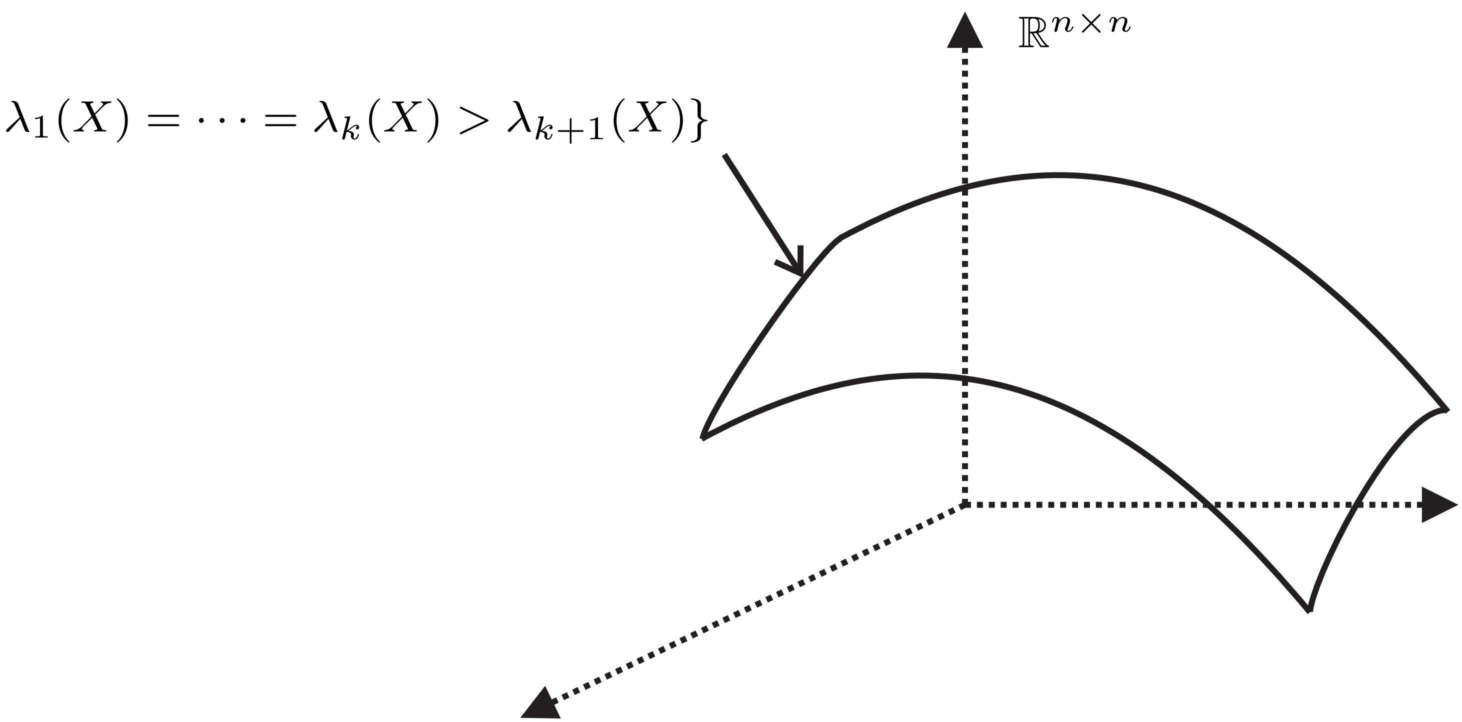
Introduction

- Example: $n \times m$ matrices with prescribed singular values s_i



Introduction

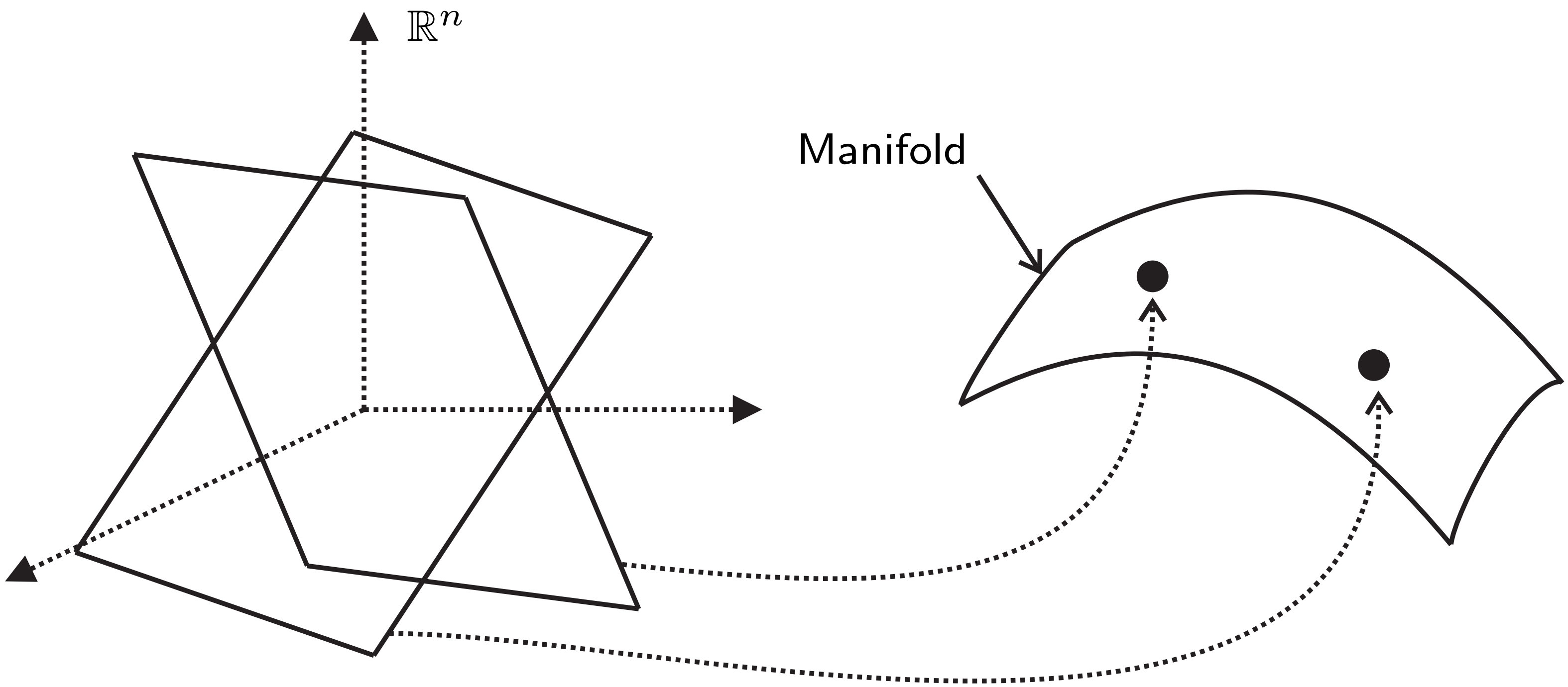
- Example: $n \times n$ symmetric matrices s.t. $\lambda_{\max}$ has multiplicity k



Introduction

- Not all manifolds are “naturally” embedded in an Euclidean space

- Example: set of k-dimensional subspaces in $\mathbb{R}^n$ (Grassmann manifold)



Introduction

- **How is differential geometry useful ?**
  - systematic framework for nonlinear problems (generalizes linear algebra)
  - elegant geometric re-interpretations of existing solutions
    - Karmakar's algorithm for linear programming
    - Sequential Quadratic Programming methods in optimization
    - Rao distance between pdf's in parametric statistical families
    - Jeffrey's noninformative prior in Bayesian setups
    - Cramér-Rao bound for parametric estimation with ambiguities
    - ... many more
  - suggests new powerful solutions

Introduction

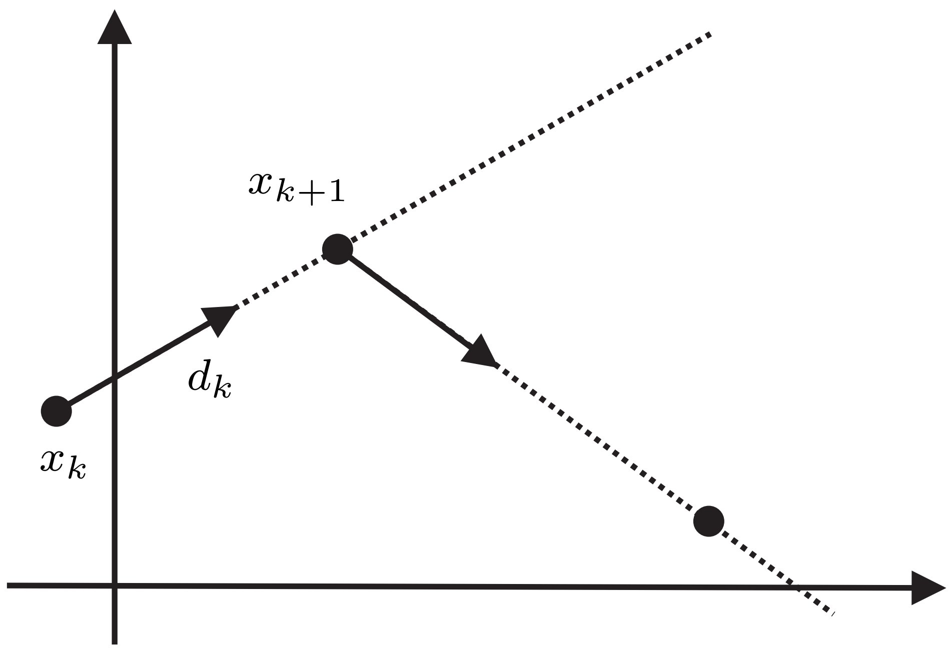
- **Where has differential geometry been applied ?**
  - Optimization on manifolds
  - Kendall's theory of shapes
  - Random matrix theory
  - Information geometry
  - Geometrical interpretation of Jeffreys' prior
  - Performance bounds for estimation problems posed on manifolds
  - Doing statistics on manifolds (generalized PCA)
  - ... a lot more (signal processing, econometrics, control, etc)

Application: optimization on manifolds

- Unconstrained problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

- Line-search algorithm: $x_{k+1} = x_k + \alpha_k d_k$



- $d_k = -\nabla f(x_k)$ [gradient], $d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$ [Newton], others ...

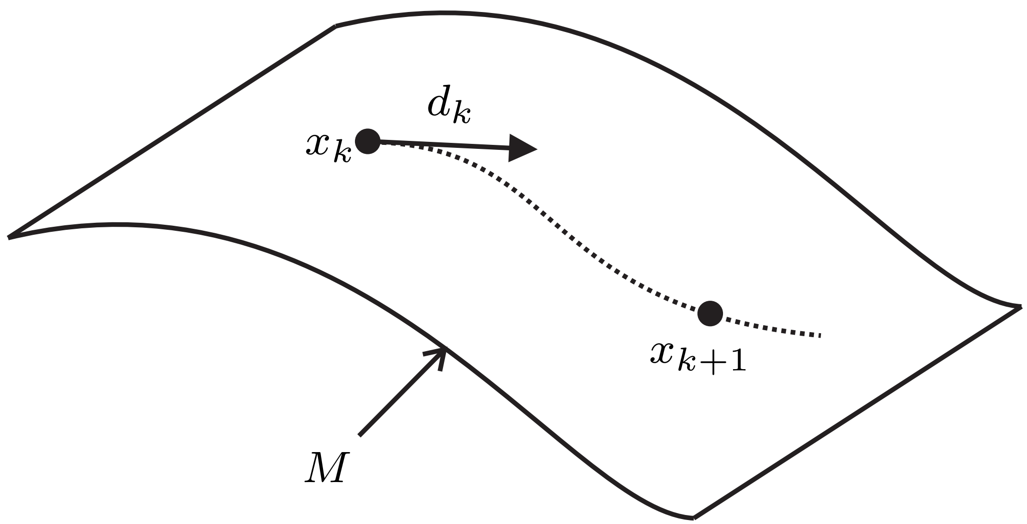
Application: optimization on manifolds

- Constrained problem

$$\min_{x \in M} f(x)$$

- Re-interpreted as an unconstrained problem on manifold M

- Geodesic-search algorithm: $x_{k+1} = \exp_{x_k}(\alpha_k d_k)$



Application: optimization on manifolds

- Works for abstract spaces (e.g. Grassmann manifold)

- Theory provides generalization of gradient, Newton direction (not obvious)

- Closed-form solutions for important manifolds (e.g. orthogonal matrices)

- Geodesic-search is not the only possibility:
  - optimization in local coordinates
  - generalization of trust-region methods

- **Innumerous applications**:
  - blind source separation, image processing, rank-reduced Wiener filter,. . .

Application: optimization on manifolds

- **Example:** Signal model

$$y[t] = Qx[t] + w[t] \quad t = 1, 2, \ldots, T$$

  - Q: unknown orthogonal matrix ($Q^\top Q = I_N$)
  - $x[t]$: known landmarks
  - $w[t] \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma)$

- **Maximum-Likelihood estimate:**

$$Q^* = \arg \max_{Q \in \mathsf{O}(N)} p\,(Y\,;\,Q)$$

  - $\mathsf{O}(N)$= group of $N \times N$ orthogonal matrices
  - $Y = \begin{bmatrix} y[1] & y[2] & \cdots & y[T] \end{bmatrix}$ matrix of observations
  - $X = \begin{bmatrix} x[1] & x[2] & \cdots & x[T] \end{bmatrix}$ matrix of landmarks

Application: optimization on manifolds

- **Optimization problem:** Orthogonal Procrustes rotation

$$
\begin{aligned}
Q^* &= \arg\min_{Q \in \mathsf{O}(N)} \|Y - QX\|^2_{\Sigma^{-1}} \\
&= \arg\min_{Q \in \mathsf{O}(N)} \mathsf{tr}\left\{Q^T \Sigma^{-1} Q \widehat{R}_{xx}\right\} - \mathsf{tr}\left\{Q^T \Sigma^{-1} \widehat{R}_{yx}\right\}
\end{aligned}
$$

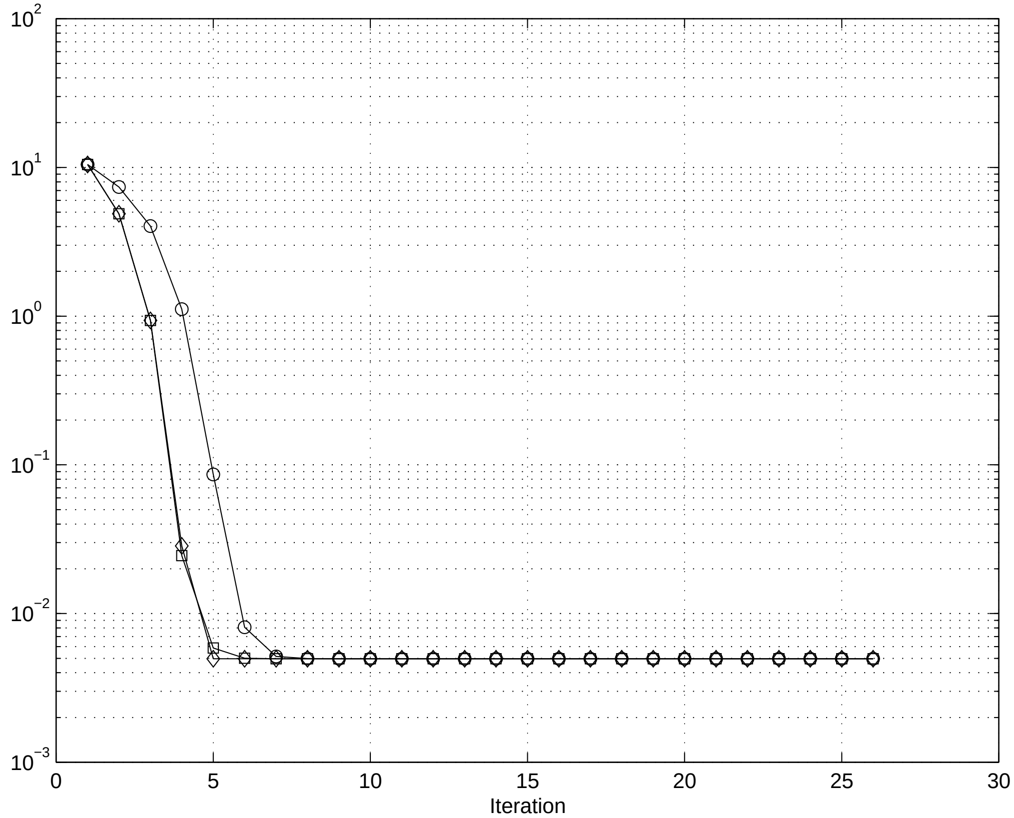
  – $\widehat{R}_{yx} = \frac{1}{T}\sum_{t=1}^{T} y[t]x[t]^\top$ and $\widehat{R}_{xx} = \frac{1}{T}\sum_{t=1}^{T} x[t]x[t]^\top$

- The eigenstructure of Σ controls the Hessian of the objective:

$$
\kappa(\Sigma^{-1}) = \frac{\lambda_{\mathsf{max}}(\Sigma^{-1})}{\lambda_{\mathsf{min}}(\Sigma^{-1})} \text{ is the condition number of } \Sigma^{-1}
$$

Application: optimization on manifolds

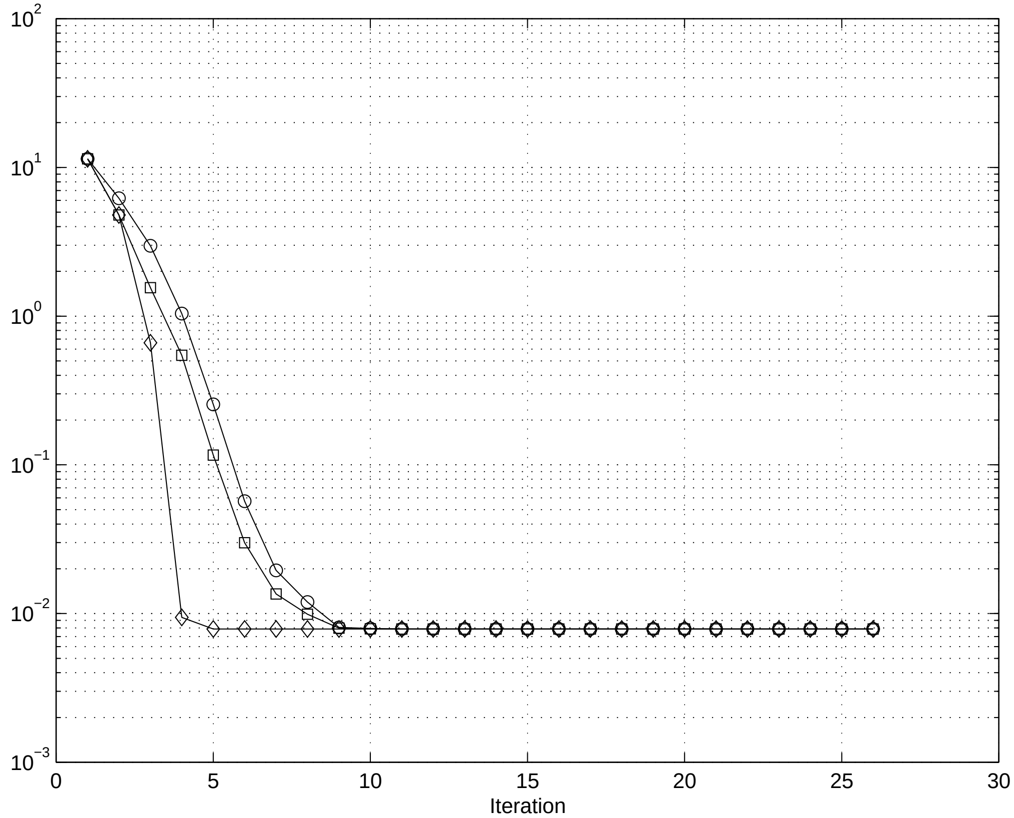
- **Example:** $N = 5$, $T = 100$, $\Sigma = \text{diag}(1, 1, 1, 1, 1)$, $\kappa(\Sigma^{-1}) = 1$



○=projected gradient □=gradient geodesic descent ◇=Newton geodesic descent

Application: optimization on manifolds

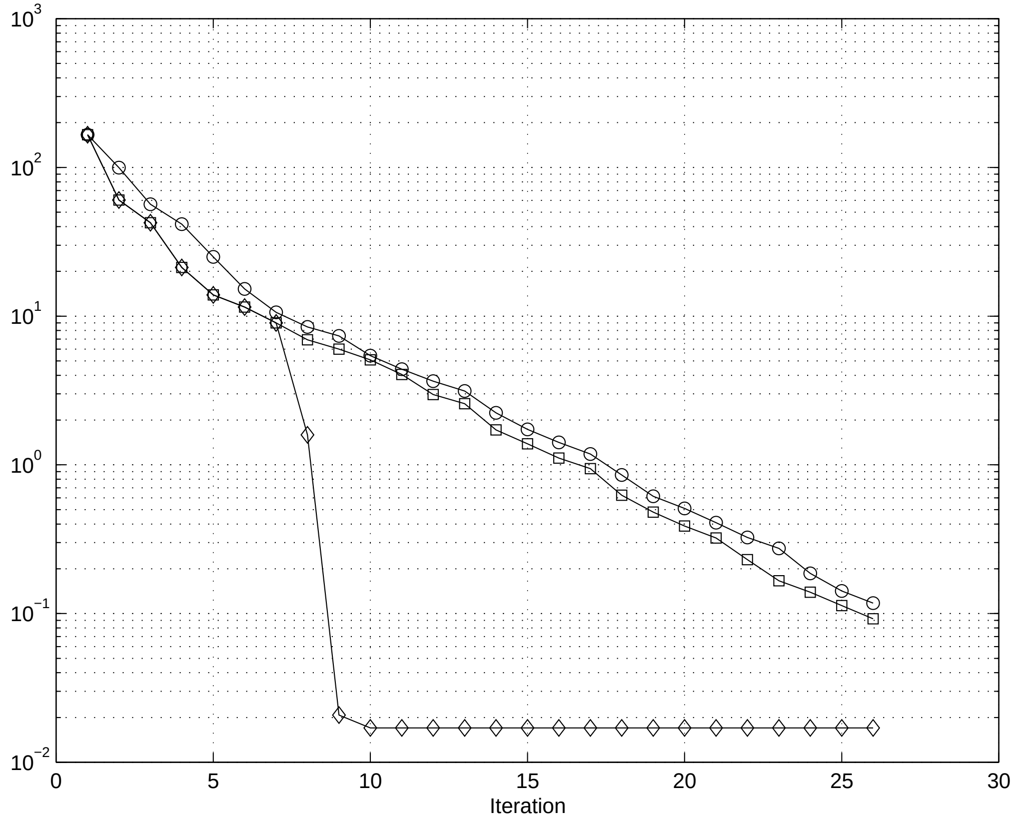
- **Example:** $N = 5$, $T = 100$, $\Sigma = \text{diag}(0.2, 0.4, 0.6, 0.8, 1)$, $\kappa(\Sigma^{-1}) = 5$



$\circ$=projected gradient $\Box$=gradient geodesic descent $\diamond$=Newton geodesic descent

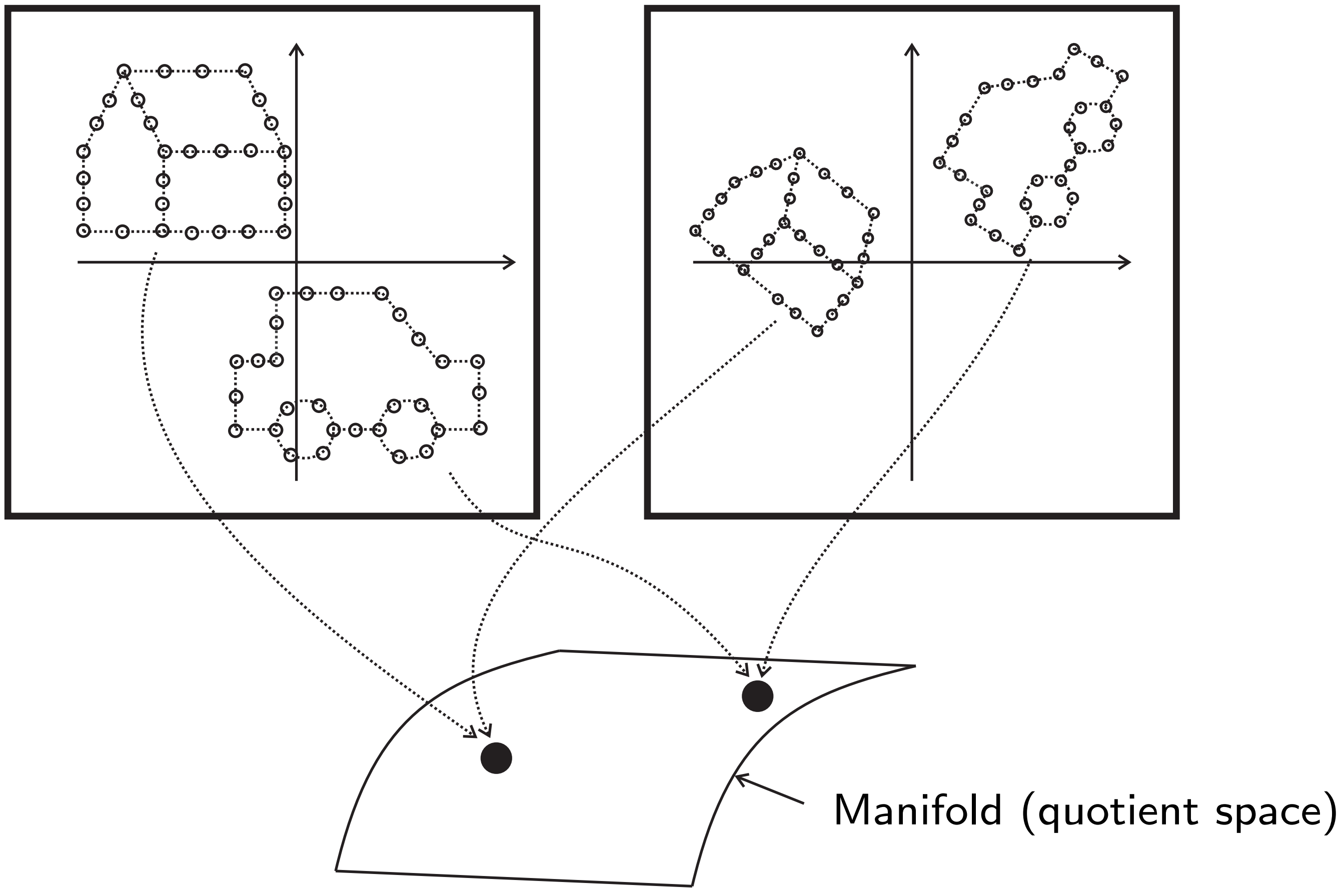
Application: optimization on manifolds

- **Example:** $N = 5$, $T = 100$, $\Sigma = \mathsf{diag}(0.02, 0.05, 0.14, 0.37, 1)$, $\kappa(\Sigma^{-1}) = 50$



○=projected gradient □=gradient geodesic descent ◇=Newton geodesic descent

Application: Kendall's theory of shapes

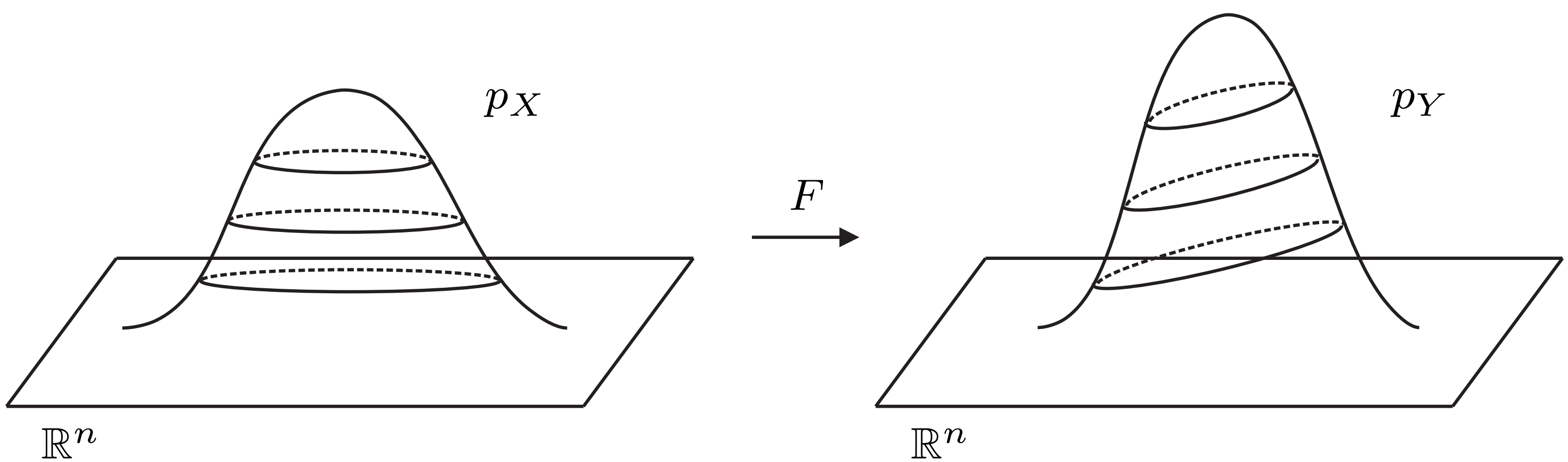


- Applications:
  - Morph one shape into another, statistics ("mean" shape), clustering, ...

Application: random matrix theory

- **Basic statistics:** transformation of random objects in Euclidean spaces

$$\left\{\begin{array}{l} x \text{ is a random vector in } \mathbb{R}^n \\ x \sim p_X(x) \\ F : \mathbb{R}^n \to \mathbb{R}^n \text{ smooth, bijective} \\ y = F(x) \end{array}\right. \quad \Rightarrow \quad \begin{array}{l} y \sim p_Y(y) = p_X(F^{-1}(y))\, J(y) \\ J(y) = \dfrac{1}{\det(DF(F^{-1}(y)))} \end{array}$$

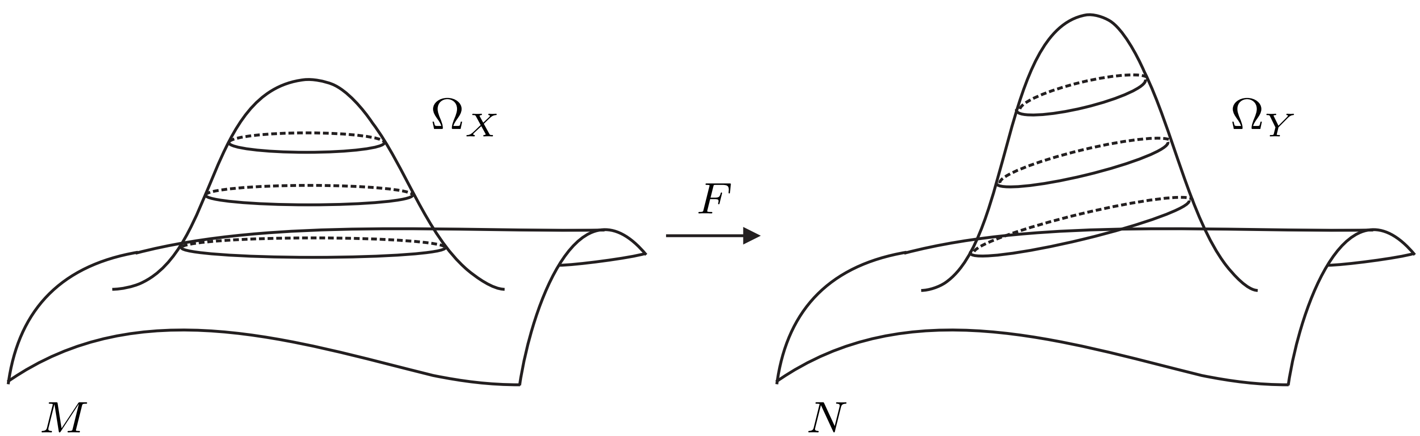


Application: random matrix theory

- **Generalization:** transformation of random objects in manifolds M, N

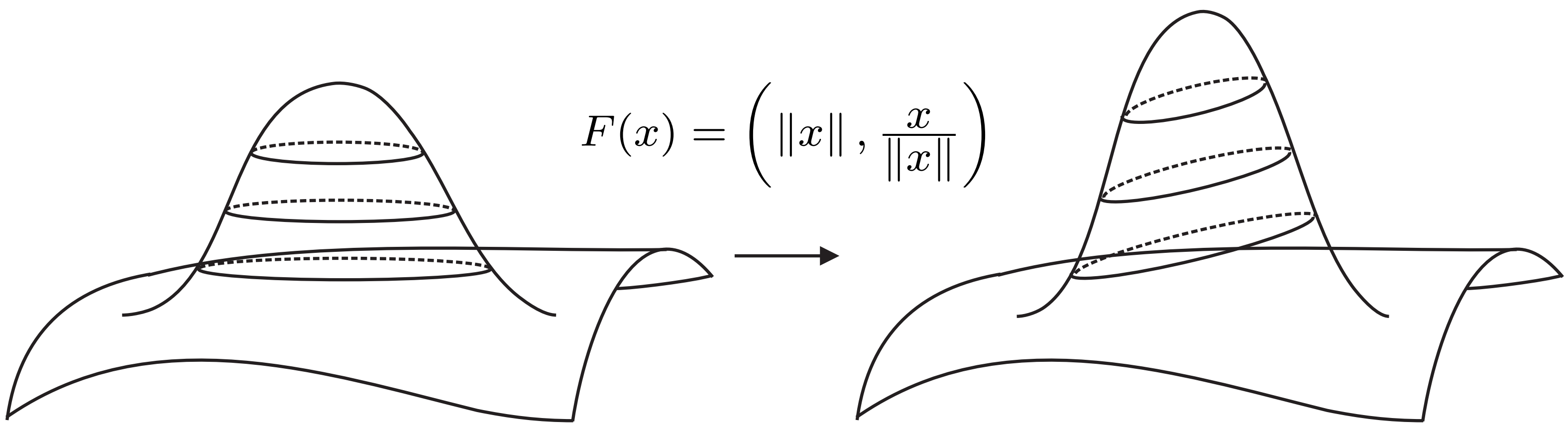
$$\left\{\begin{array}{l} x \text{ is a random point in } M \\ x \sim \Omega_X \text{ (exterior form)} \\ F : M \to N \text{ smooth, bijective} \\ y = F(x) \end{array}\right. \qquad \Rightarrow \qquad y \sim \Omega_Y = \ldots$$

- The answer is provided by the calculus of exterior differential forms



Application: random matrix theory

- Example: decoupling a random vector in amplitude and direction



$$F(x) = \left(\|x\| \, , \frac{x}{\|x\|} \right)$$

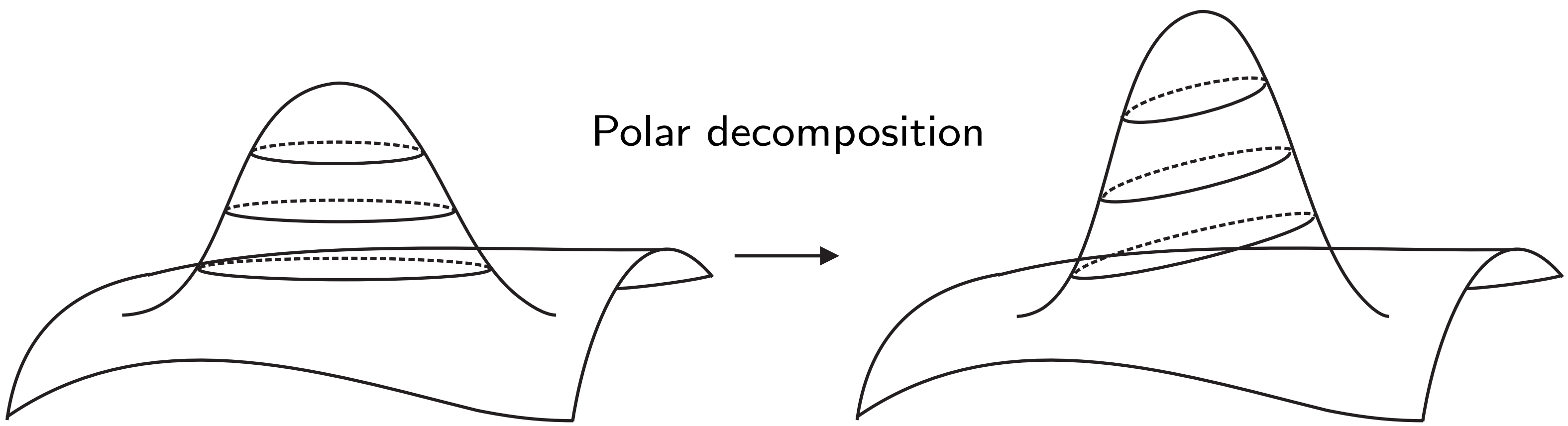
$M = \mathbb{R}^n - \{0\}$

$N = \mathbb{R}_{++} \times \mathsf{S}^{n-1}$
$\quad = \{(R, u) \, : \, R > 0, \|u\| = 1\}$

- Answer: $x \sim p_X(x) \quad \Rightarrow \quad p(R, u) = p_X(Ru) \, R^{n-1}$

Application: random matrix theory

- Example: decoupling a random matrix by the polar decomposition $X = PQ$



$$M = \mathsf{GL}(n) = \{X \in \mathbb{R}^{n\times n} \,:\, |X| \neq 0\}$$

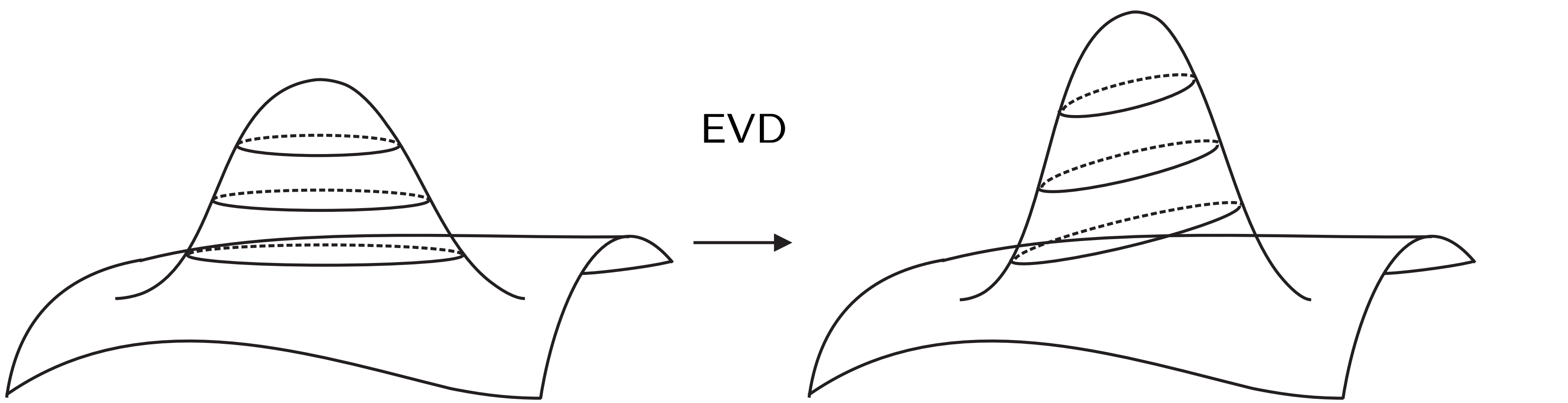
$$N = \mathsf{S}^n_{++} \times \mathsf{O}(n) = \{(P, Q) \,:\, P \succ 0, Q^\top Q = I_n\}$$

- Answer: $X \sim p_X(X) \quad \Rightarrow \quad p(P,Q) = \ldots$ (known)

Application: random matrix theory

- Example: decoupling a random symmetric matrix by eigendecomposition

$$X = Q\Lambda Q^\top$$



$M = \mathsf{S}^n$
$= \{X \in \mathbb{R}^{n\times n} : X = X^\top\}$

$N = \mathsf{O}(n) \times \mathsf{D}(n)$
$= \{(Q, \Lambda) : Q^\top Q = I_n, \Lambda : \text{diag}\}$

- Answer: $X \sim p_X(X) \quad \Rightarrow \quad p(Q, \Lambda) = \ldots$ (known)

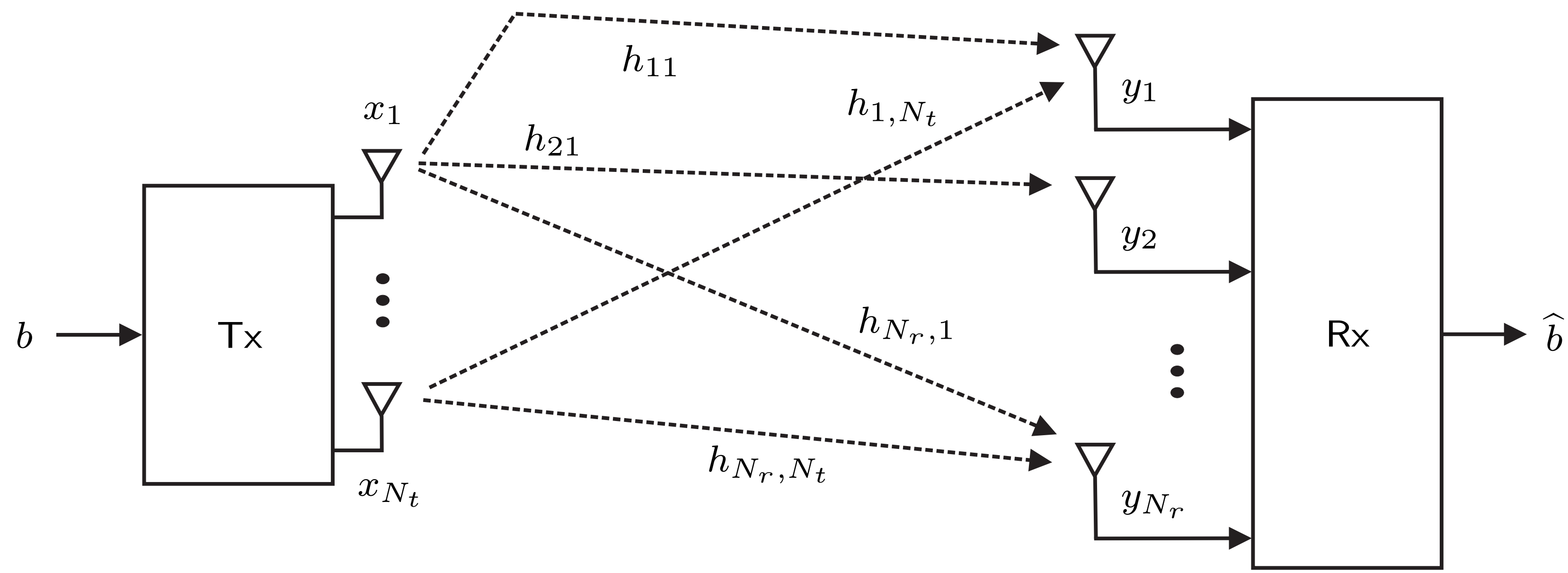
- Technicality: in fact, the range of F is a quotient of an open subset of N

Application: random matrix theory

- Many more examples:
  - Cholesky decomposition (e.g., leads to Wishart distribution)
  - LU
  - QR
  - SVD

Application of RMT: coherent capacity of multi-antenna systems

- Scenario: point-to-point single-user communication with multiple Tx antennas



Application of RMT: coherent capacity of multi-antenna systems

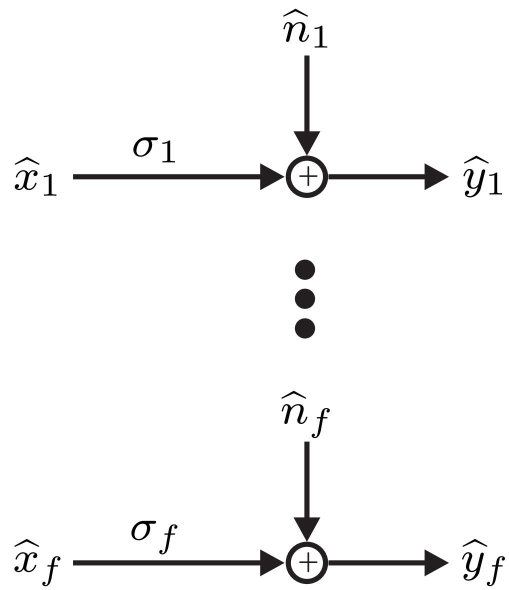
- **Data model:** $y = Hx + n$ with $y, n \in \mathbb{C}^{N_r}$, $H \in \mathbb{C}^{N_r \times N_t}$, $x \in \mathbb{C}^{N_t}$
  - N_t = number of Tx antennas
  - N_r = number of Rx antennas

  Assumption: $n_i \overset{\text{iid}}{\sim} \mathbb{C}\mathcal{N}(0, 1)$

- **Decoupled data model:**
  - SVD: $H = U\Sigma V^H$ with $U \in \mathsf{U}(N_r)$, $V \in \mathsf{U}(N_t)$, $\Sigma = \mathsf{Diag}(\sigma_1, \ldots, \sigma_f, 0)$,
    $(\sigma_1, \ldots, \sigma_f)$ = nonzero singular values of H, $f = \min\{N_r, N_t\}$
  - Transform the data: $\widehat{y} = U^H y$, $\widehat{x} = V^H x$ and $\widehat{n} = U^H n$
  - Equivalent diagonal model: $\widehat{y} = \Sigma \widehat{x} + \widehat{n}$

Application of RMT: coherent capacity of multi-antenna systems

- **Interpretation:** The matrix channel H is equivalent to f parallel scalar channels



Application of RMT: coherent capacity of multi-antenna systems

- **Assumption:** channel matrix H is random and known only at the Rx

- **Channel capacity:**

$$C = \max_{p(x), \mathsf{E}\{\|x\|^2 \le P\}} I(x; (y, H))$$

$I =$ mutual information

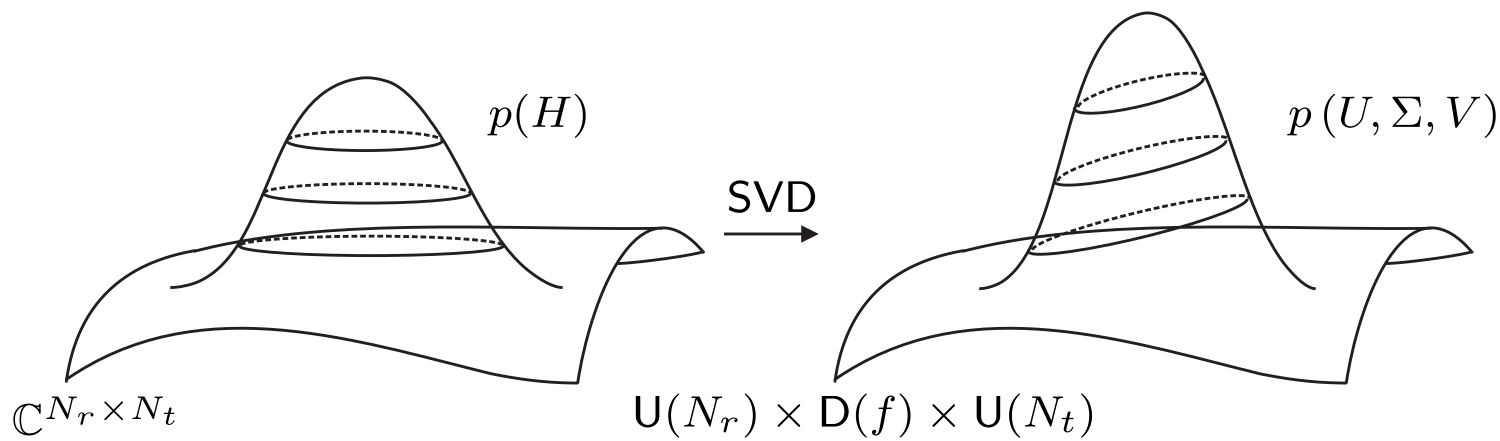
- **Solution:**

$$C = \mathsf{E}_H \left\{ \sum_{i=1}^{f} \log\left(1 + (P/N_t)\sigma_i^2\right) \right\}$$

Recall: $(\sigma_1, \ldots, \sigma_f) =$ random singular values of H, $f = \min\{N_r, N_t\}$

Application of RMT: coherent capacity of multi-antenna systems

- H is random and $H = U\Sigma V^H$ (SVD)



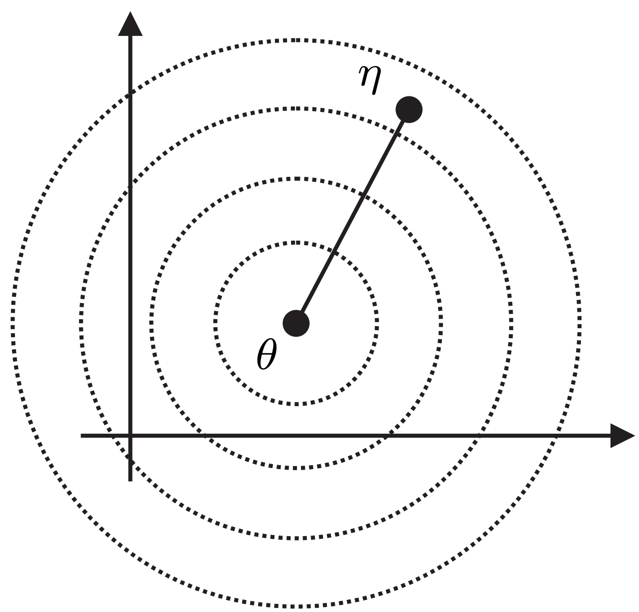
- **Capacity:** when $[H_{ij}] \overset{\text{iid}}{\sim} \mathbb{C}\mathcal{N}(0,1)$

$$C = \int_0^\infty \log(1 + (P/N_t)\lambda) \sum_{k=0}^{f-1} \frac{k!}{(k+g-f)!} (L_k^{g-f}(\lambda))^2 \lambda^{g-f} e^{-\lambda}\, d\lambda$$

$g = \max\{N_r, N_t\}$ and L_j^i=Laguerre polynomials

Application: information geometry

- **Problem:** given a parametric statistical family $\mathcal{F} = \{p(x;\theta) : \theta \in \Theta\}$ assign a distance function $d : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$
- **Example:** $\mathcal{F} = \{\mathcal{N}(\theta, \Sigma) : \theta \in \Theta = \mathbb{R}^n\}$ (covariance Σ is fixed)
- Naive choice: $d : \Theta \times \Theta \to \mathbb{R} \quad d(\theta, \eta) = \|\theta - \eta\|$

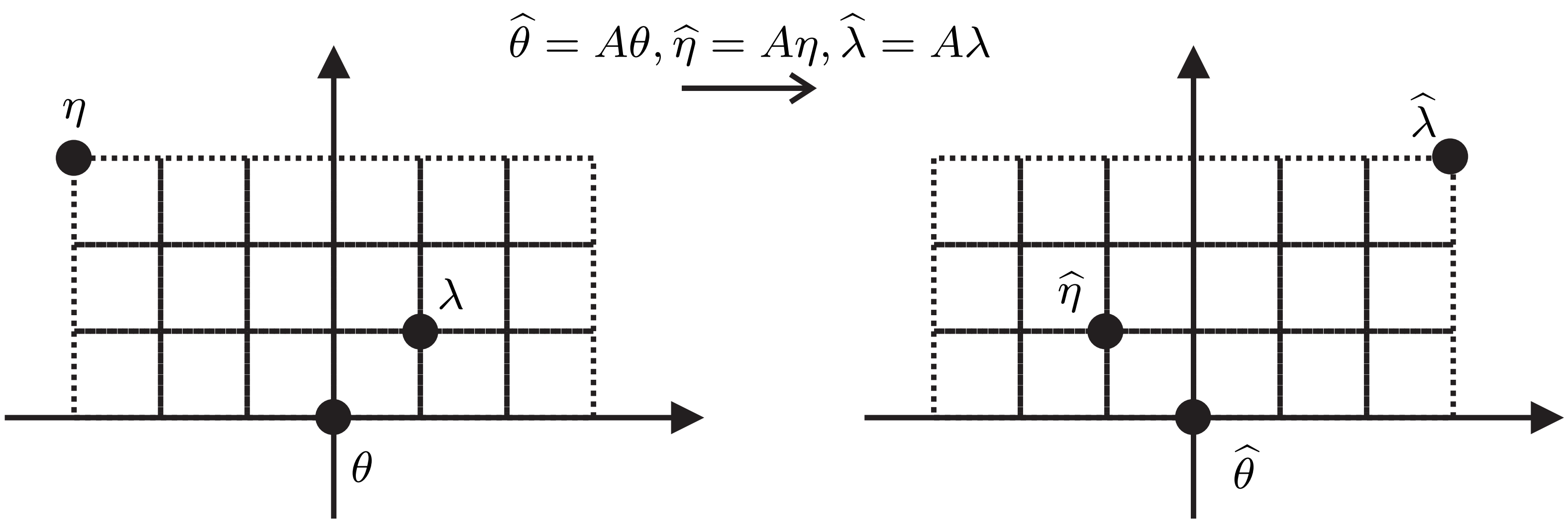


- This method does not produce "intrinsic" distances (parameter invariant)

Application: information geometry

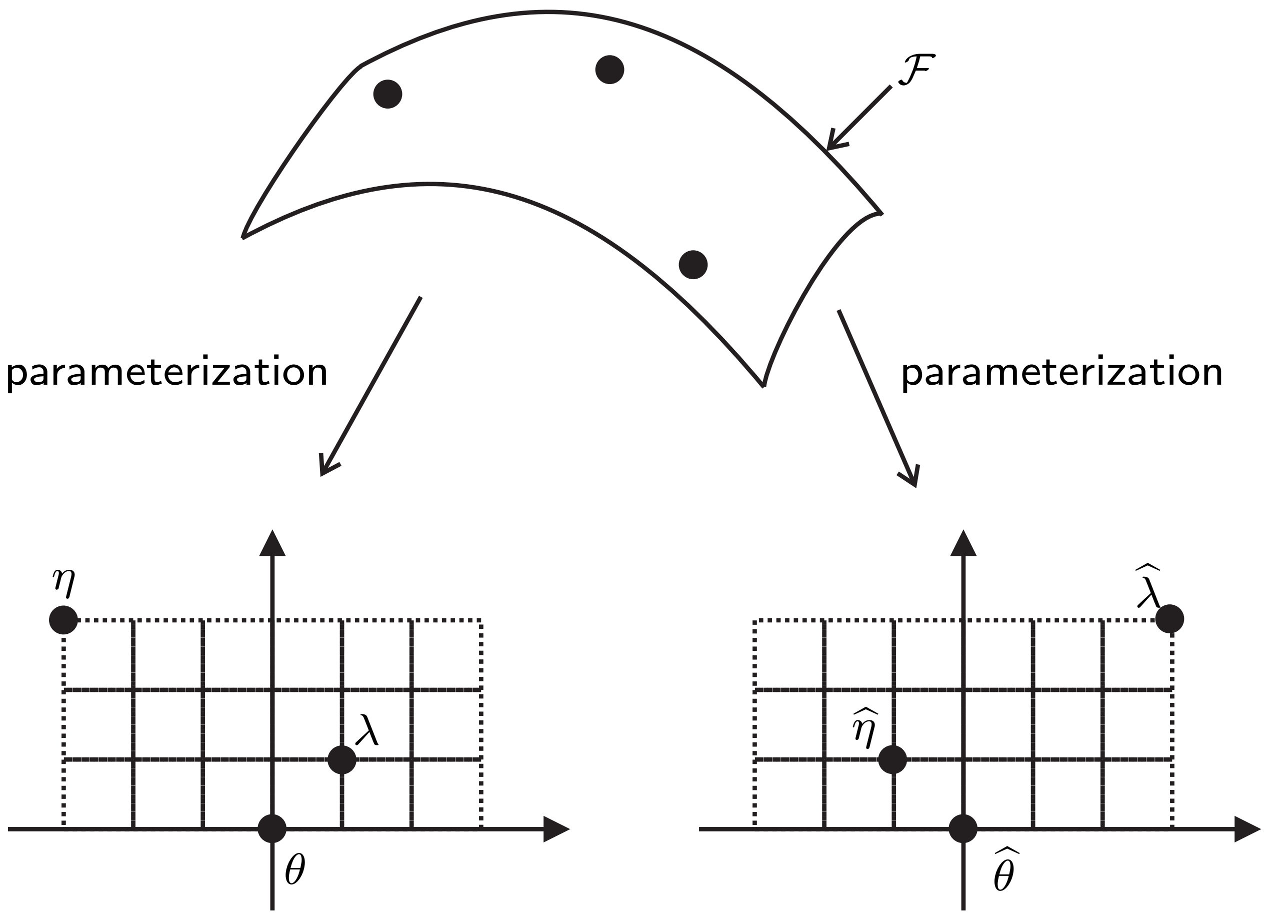
- Re-parameterization $\widehat{\theta} = A\theta$: $\mathcal{F} = \left\{ \mathcal{N}(A^{-1}\widehat{\theta}, \Sigma) \; : \; \widehat{\theta} \in \widehat{\Theta} = \mathbb{R}^n \right\}$

- **Example:** $\theta = (0,0), \eta = (-3,3), \lambda = (1,1), A = \begin{bmatrix} 5/3 & 4/3 \\ 4/3 & 5/3 \end{bmatrix}$



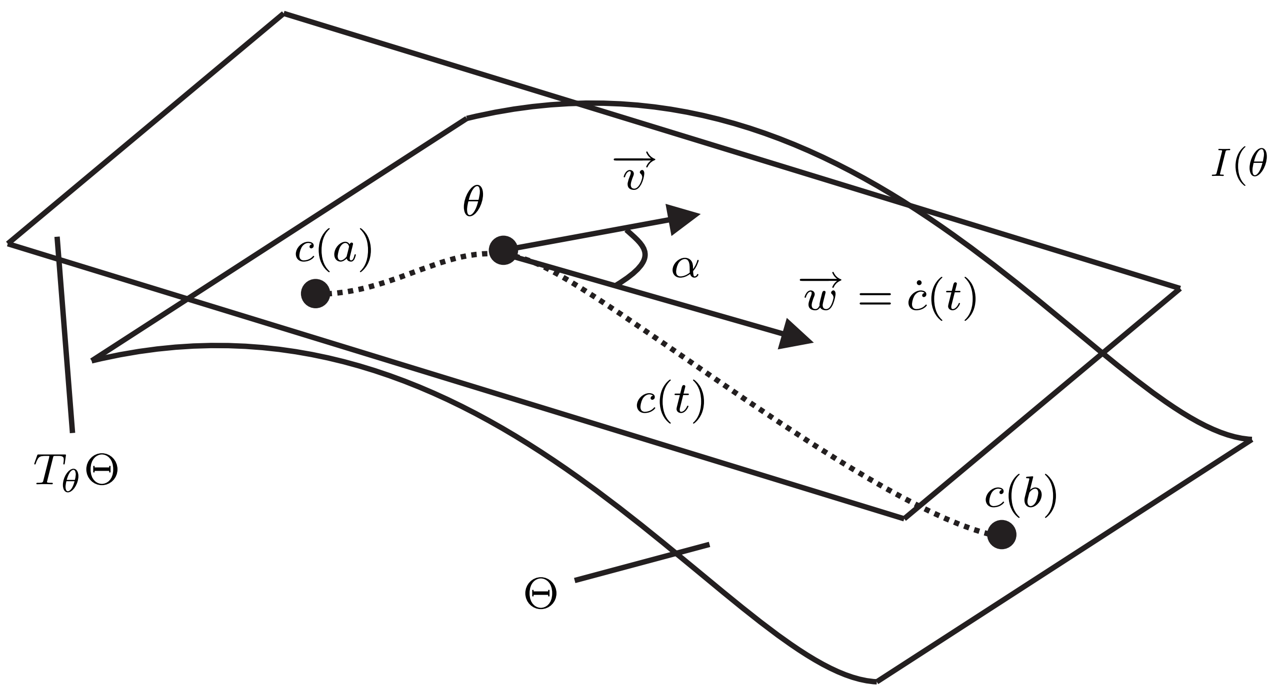
$d(\theta, \lambda) < d(\theta, \eta)$ $\qquad$ $d(\widehat{\theta}, \widehat{\lambda}) > d(\widehat{\theta}, \widehat{\eta})$

Application: information geometry



Application: information geometry

- Rao suggested the information metric to obtain distances between pdf's
- **Differential geometric interpretation:** The Fisher Information Matrix is adopted as the Riemannian tensor on Θ



$$\langle \overrightarrow{v}, \overrightarrow{w} \rangle = \overrightarrow{v}^\top I(\theta) \overrightarrow{w}$$

$$I(\theta) = -\mathsf{E}_\theta \left\{ \nabla^2_\theta \log p(x;\theta) \right\}$$

$$\left| \overrightarrow{v} \right| = \sqrt{\langle \overrightarrow{v}, \overrightarrow{v} \rangle}$$

$$\alpha = \frac{\langle \overrightarrow{v}, \overrightarrow{w} \rangle}{\left| \overrightarrow{v} \right| \left| \overrightarrow{w} \right|}$$

$$\mathsf{length(c)} = \int_a^b \left| \dot{c}(t) \right| dt$$

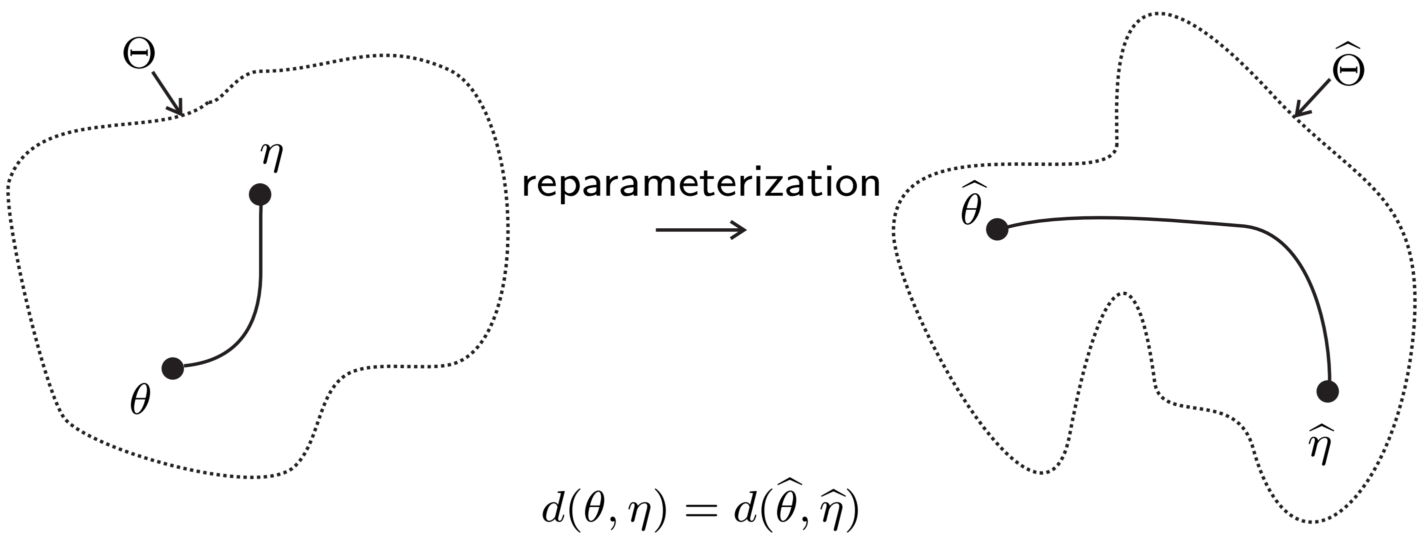
- **Insight:** A parametric statistical family is an autonomous geometrical object

Application: information geometry

- **Information distance:**

$$d(\theta, \eta) = \inf\left\{\text{length}(c) \,:\, c \text{ is a curve on } \Theta \text{ connecting } \theta \text{ to } \eta\right\}$$

- The information distance is invariant to reparameterizations



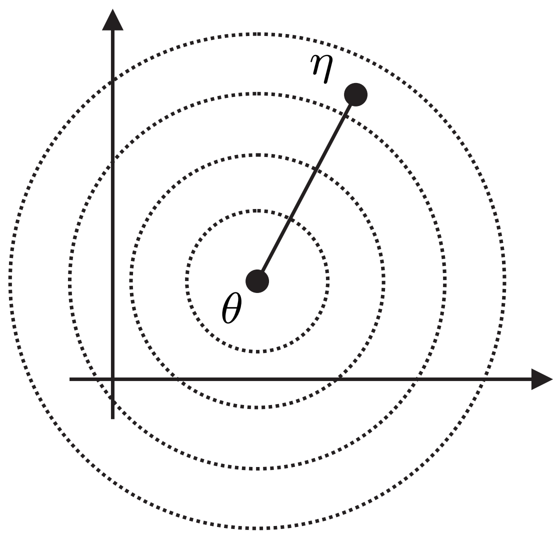
$$d(\theta, \eta) = d(\widehat{\theta}, \widehat{\eta})$$

- **Link with Kullback-Leibler distance:** $d_{\mathsf{KL}}(\theta, \eta) = \frac{1}{2}\, d(\theta, \eta)^2 + O\left(d(\theta, \eta)^3\right)$

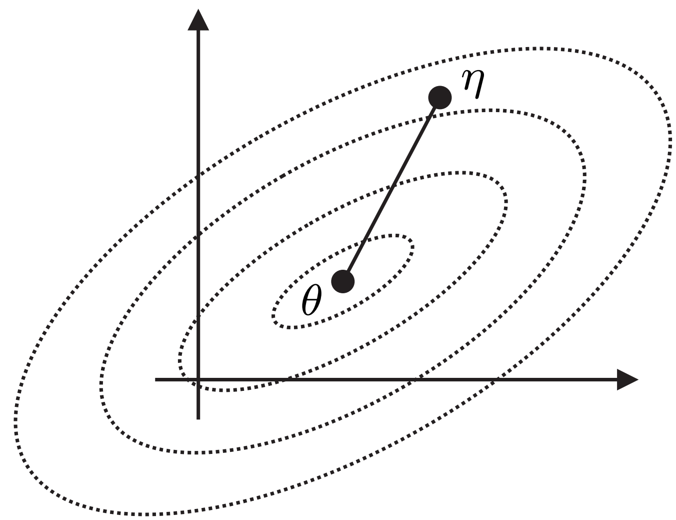
Application: information geometry

- **Example:** $\mathcal{F} = \{\mathcal{N}(\theta, \Sigma) : \theta \in \Theta = \mathbb{R}^n\}$ (covariance Σ is fixed)

$$d(\theta, \eta) = \sqrt{(\theta - \eta)^T \Sigma^{-1} (\theta - \eta)}$$ [Mahalanobis distance]



Euclidean distance

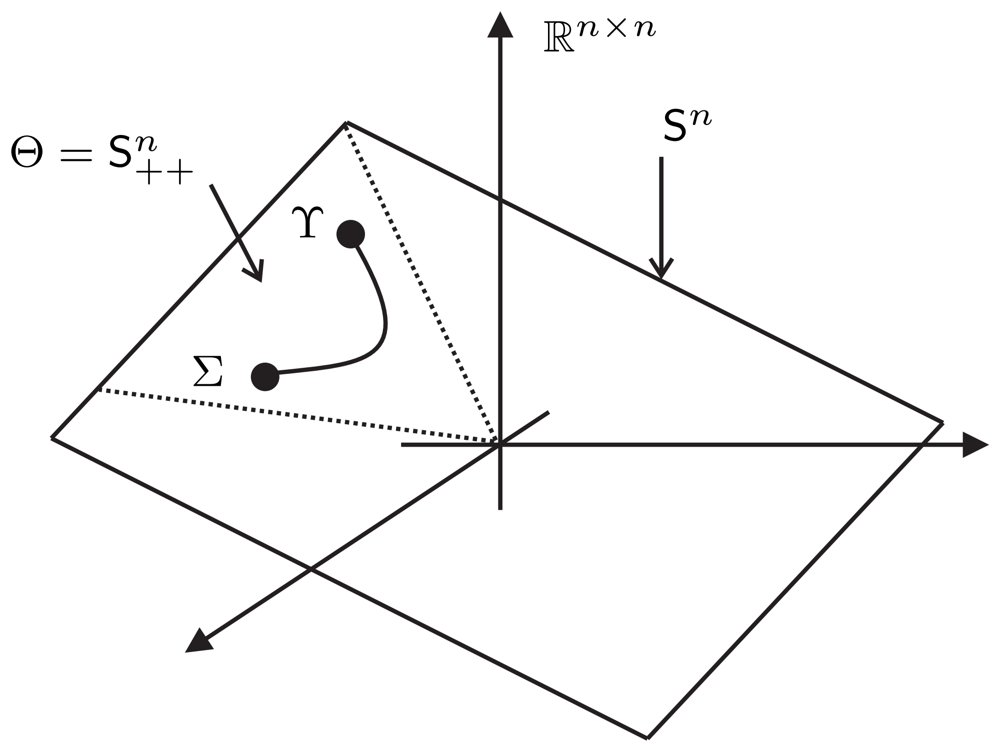


Information distance

Application: information geometry

- **Example:** $\mathcal{F} = \{\mathcal{N}(\mu, \Sigma) \, : \, \Sigma \in \mathsf{S}^n_{++}\}$ (mean-value μ is fixed)

$$d(\Sigma, \Upsilon) = \sqrt{\frac{1}{2}\sum_{i=1}^{n}(\log\,\lambda_i)^2} \quad (\lambda_1, \ldots, \lambda_n) = \text{generalized eigenvalues of } (\Sigma, \Upsilon)$$

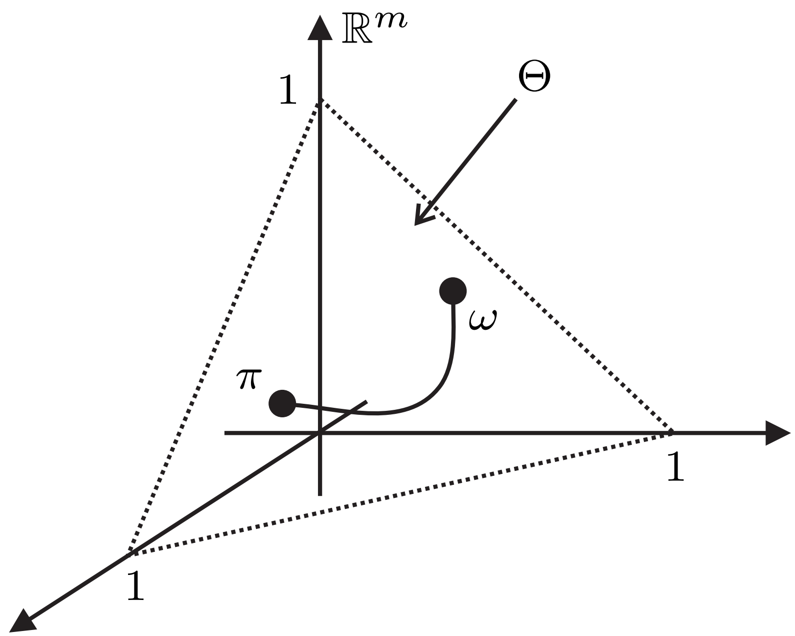


Application: information geometry

- **Example:** $\mathcal{F} = \{p(x;\pi) \sim \text{Multinomial}(n,\pi) \,:\, \pi \in \Theta = \text{simplex}(\mathbb{R}^m)\}$

  $x = (x_1,\ldots,x_m) \in \mathbb{N}^m$, $\sum_{i=1}^m x_i = n$, $\pi = (\pi_1,\ldots,\pi_m)$, $\sum_{i=1}^m \pi_i = 1$

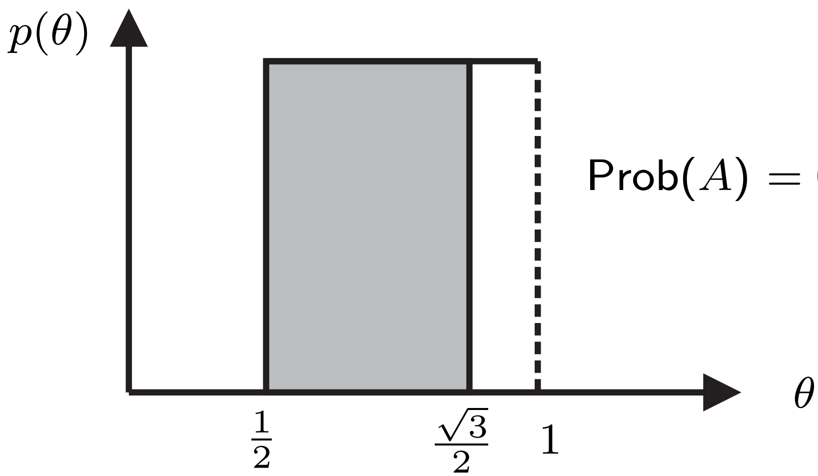
$$p(x;\pi) = \frac{n!}{x_1!\cdots x_m!}\pi_1^{x_1}\cdots\pi_m^{x_m} \qquad d(\pi,\omega) = 2\sqrt{n}\arccos\left(\sum_{i=1}^m \pi_i\omega_i\right)$$



Application: geometrical interpretation of Jeffreys' prior

- **Problem:** given a parametric statistical family $\mathcal{F} = \{p(x;\theta) \,:\, \theta \in \Theta\}$ assign a non-informative prior $p(\theta)$ for the parameter θ

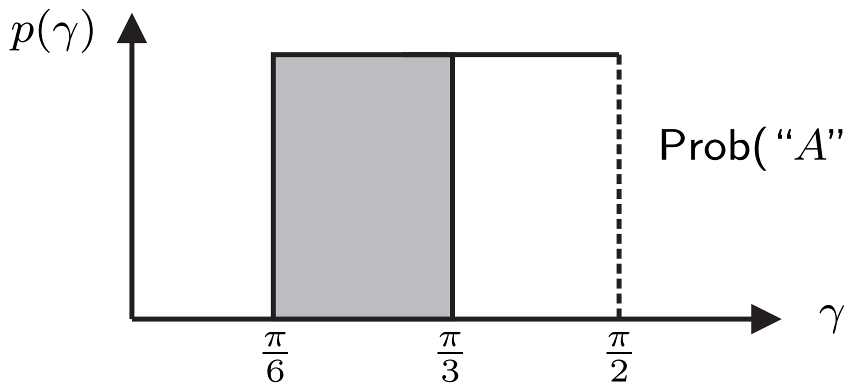
- **Example:** $\mathcal{F} = \left\{p(x;\theta) \sim \mathcal{N}(0,\theta^2) \,:\, \theta \in \Theta = (1/2, 1)\right\}$
- Naive choice (uniform distribution):



- This method does not produce "intrinsic" priors (parameter invariant)

Application: geometrical interpretation of Jeffreys' prior

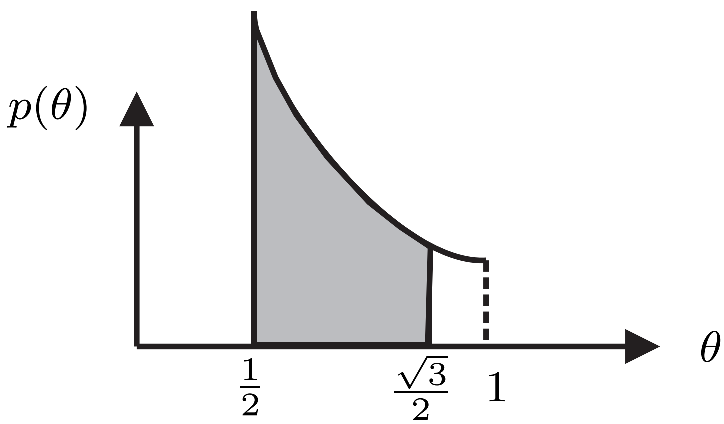
- With $\theta = \sin(\gamma)$: $\mathcal{F} = \left\{ p(x;\gamma) \sim \mathcal{N}(0, \sin^2(\gamma)) \ : \ \gamma \in \Gamma = (\pi/6, \pi/2) \right\}$

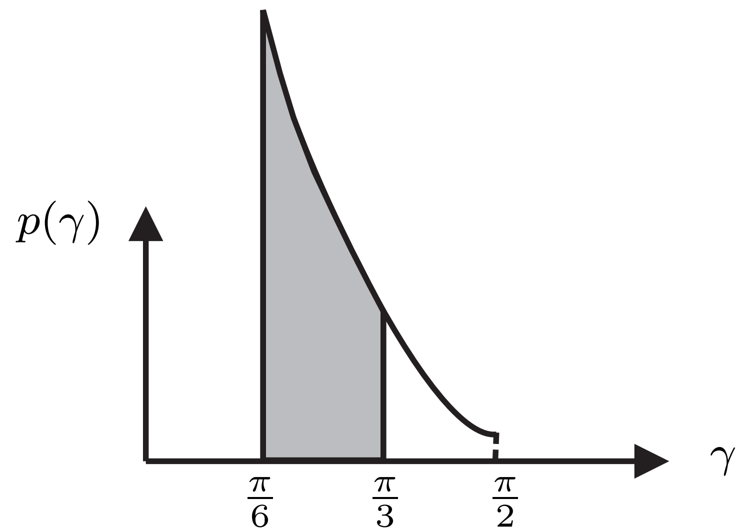


- Jeffreys' prior: $p(\theta) \propto \sqrt{\det(I(\theta))}$ where $I(\theta)$ is the Fisher information matrix

Application: geometrical interpretation of Jeffreys' prior

- **For the current example:** $p(\theta) \propto \dfrac{1}{\theta}$ and $p(\gamma) \propto \mathrm{cotg}(\gamma)$



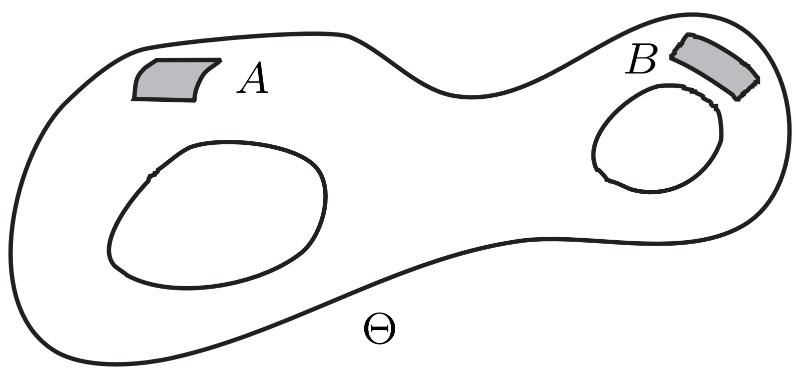


$$\text{Prob}(A) = \text{Prob}(\text{“}A\text{”}) = 0.79$$

Application: geometrical interpretation of Jeffreys' prior

- **Differential geometric interpretation:** Jeffreys' prior is simply the Riemannian volume element induced by the Fisher metric!

- **Insight:** A parametric statistical family is an autonomous geometrical object carrying its own "uniform" prior (applies equal mass to sets of equal area)



$$\text{Area}(A) = \text{Area}(B) \Rightarrow \text{Prob}(\theta \in A) = \text{Prob}(\theta \in B)$$

Application: performance bounds

- Classical setup for Cramér-Rao Bound (CRB):
  - $\Omega = \mathbb{R}^n$ is the observation space and $y \in \Omega$ is the observed data point
  - $\mathcal{F} = \{f_\theta \, : \theta \in \Theta\}$ is a given parametric family of positive pdf's
  - $\widehat{\theta} \, : \, \Omega \to \Theta$ is an unbiased estimator of θ, i.e, $\mathsf{E}_\theta \left\{ \widehat{\theta}(Y) \right\} = \theta, \; \forall_{\theta \in \Theta}$
  - Θ denotes an **open** subset of the Euclidean space $\mathbb{R}^p$

- CRB inequality:

$$\mathsf{Cov}_\theta \left(\widehat{\theta} \right) \succeq I(\theta)^{-1}$$

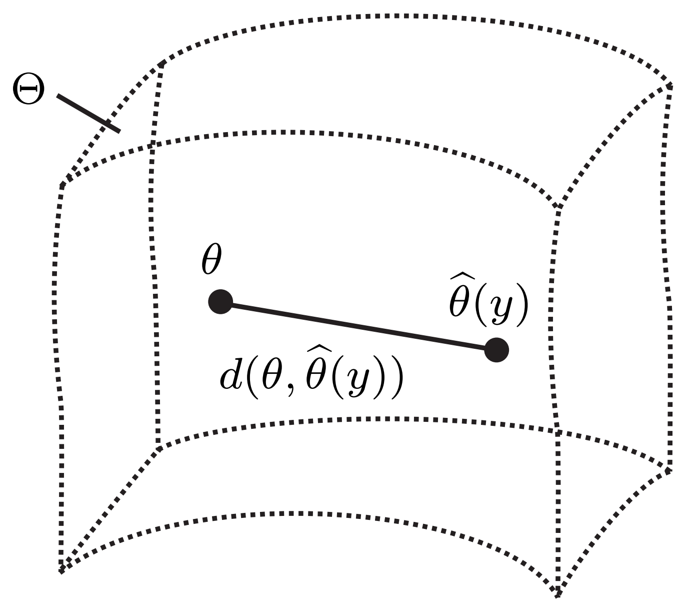
  - $\mathsf{Cov}_\theta \left(\widehat{\theta} \right) = \mathsf{E}_\theta \left\{ \left(\widehat{\theta}(Y) - \theta \right) \left(\widehat{\theta}(Y) - \theta \right)^\top \right\}$ is the covariance matrix of $\widehat{\theta}$
  - $I(\theta) = \mathsf{E}_\theta \left\{ \nabla_\theta \ln f(Y;\theta) \, \nabla_\theta \ln f(Y;\theta)^T \right\}$ is the Fisher Information Matrix (FIM)

Application: performance bounds

- Distance lower bound:

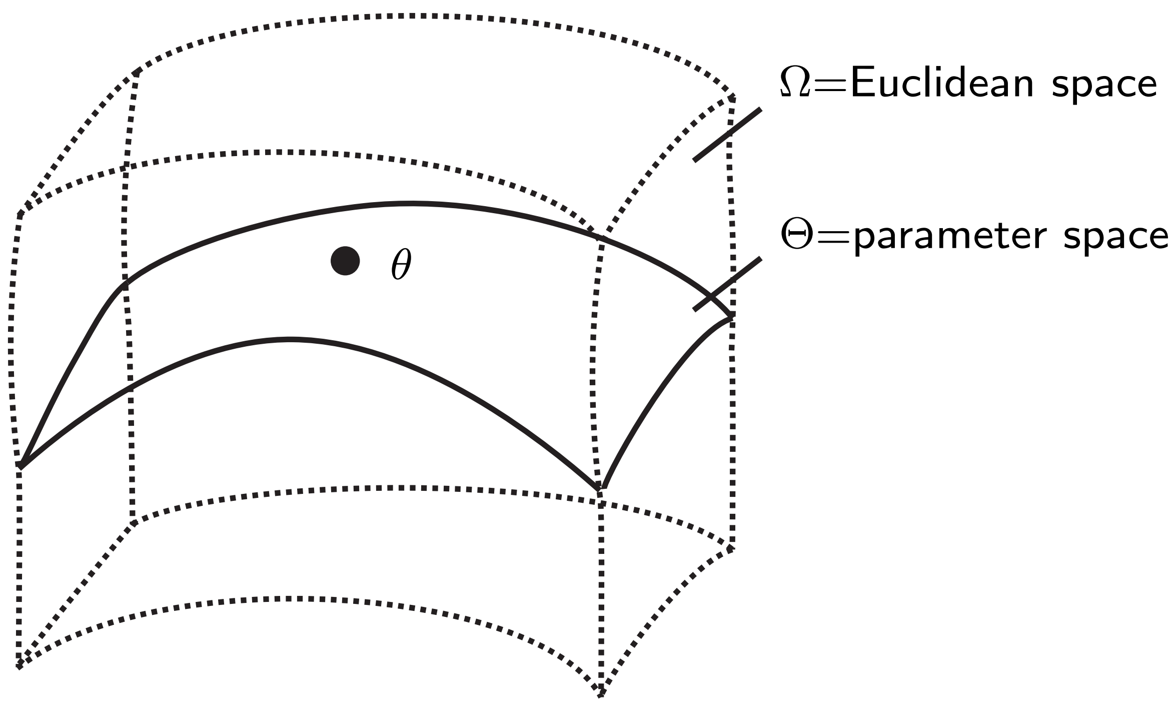
$$\mathsf{Cov}_\theta\left(\widehat{\theta}\right) \succeq I(\theta)^{-1} \quad \Rightarrow \quad \mathsf{var}_\theta\left(\widehat{\theta}\right) \geq \mathsf{tr}\left(I(\theta)^{-1}\right)$$

  - $\mathsf{var}_\theta\left(\widehat{\theta}\right) = \mathsf{E}_\theta\left\{d\left(\theta, \widehat{\theta}(Y)\right)^2\right\}$ is the variance of the estimator $\widehat{\theta}$
  - $d\left(\theta, \widehat{\theta}(y)\right) = \left\|\theta - \widehat{\theta}(y)\right\|$ is the Euclidean distance between θ and $\widehat{\theta}(y)$



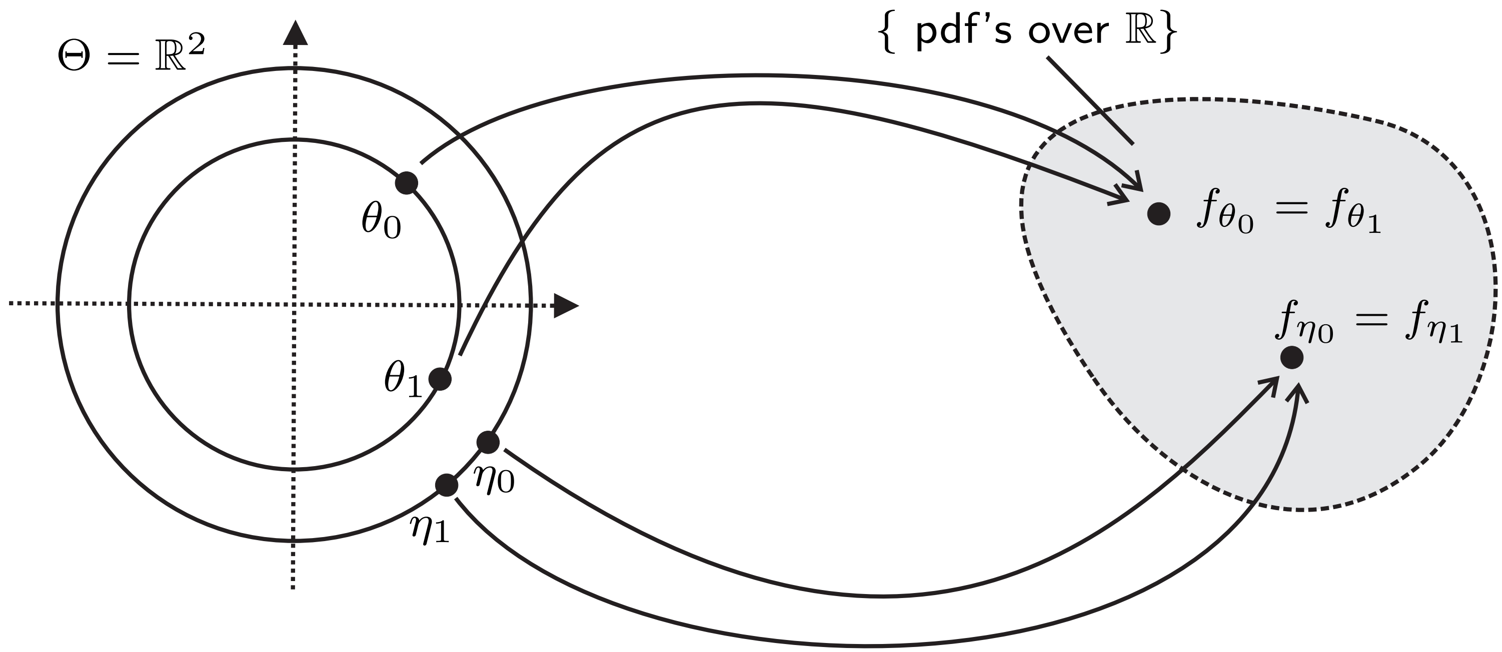
Application: performance bounds

- In practice, we need extensions of the CRB
- **Extension 1**: there are deterministic constraints on the parameter θ
  - Example (θ is an orthogonal matrix): $\Omega = \mathbb{R}^{n \times n}$, $\Theta = \mathsf{O}(n)$
- Parameter space Θ becomes a submanifold of an Euclidean space



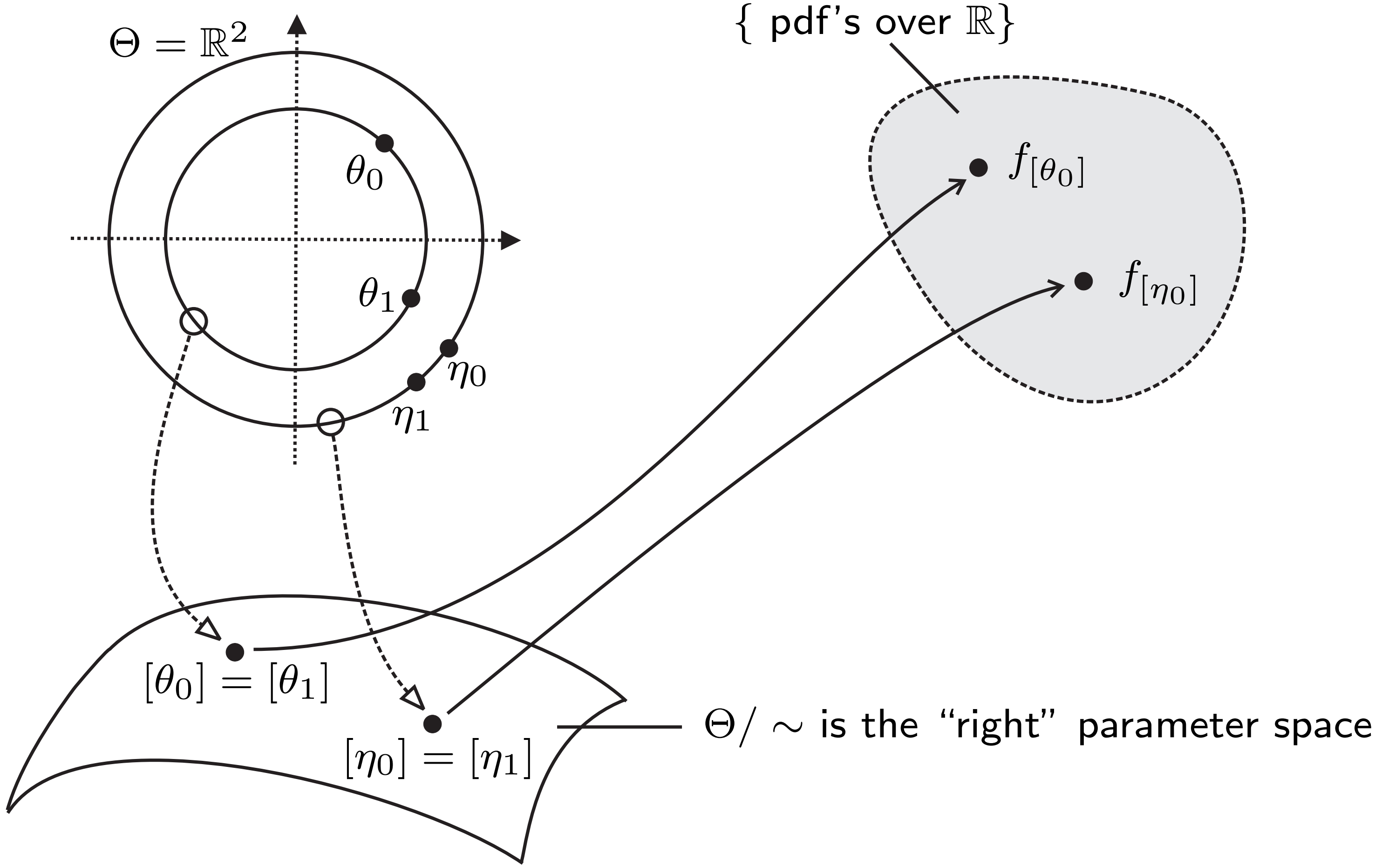
Application: performance bounds

- **Extension 2**: model has intrinsic ambiguities (e.g., over-parameterized)
- **Simple example:** $\Theta = \mathbb{R}^2$
  - Observation model: $y = \|\theta\| + \text{AWGN}$



Application: performance bounds

- Introduce equivalence relation on Θ: $\theta_1 \sim \theta_2 \iff \|\theta_1\| = \|\theta_2\|$

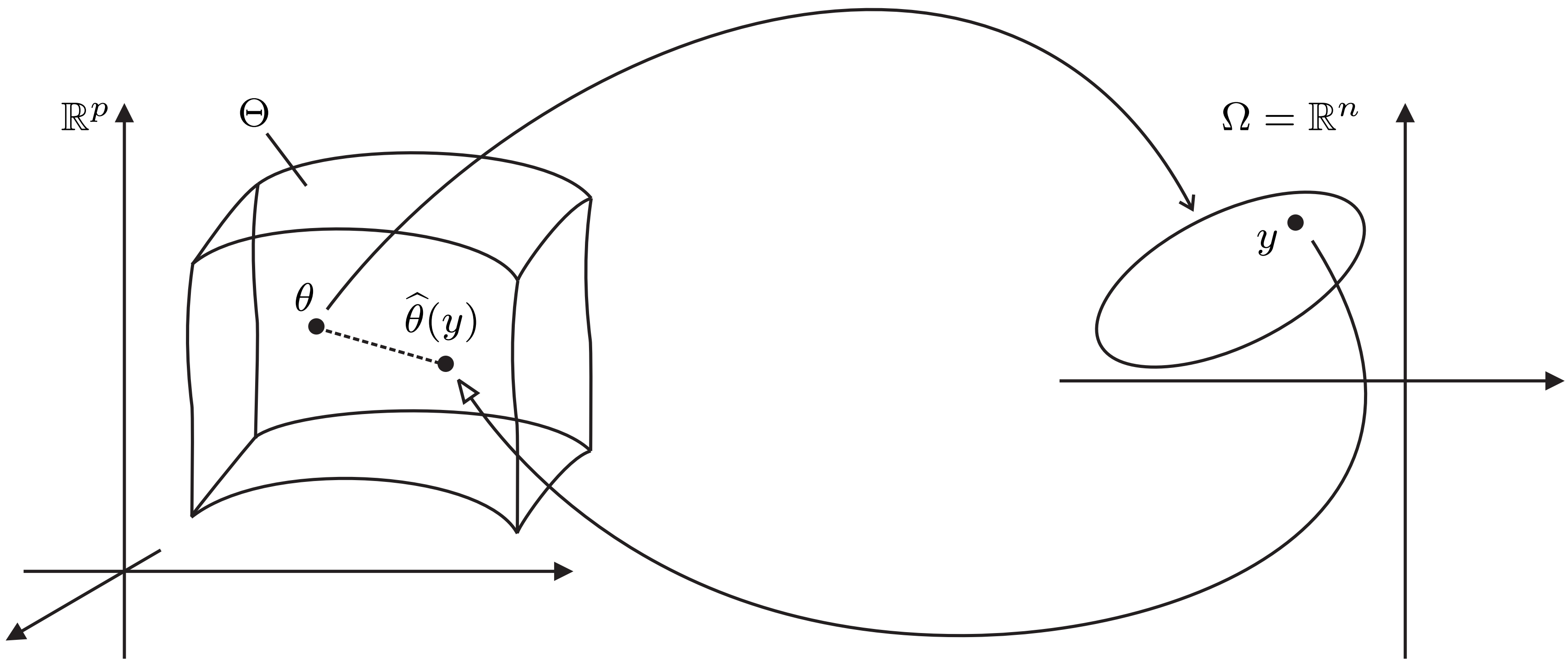


Application: performance bounds

- **Key-idea**: Riemannian manifold theory unifies treatment of
  - Extension 1: Parametric estimation with constraints
  - Extension 2: Parametric estimation over quotient spaces

Application: performance bounds

- **Classical Euclidean setup:**

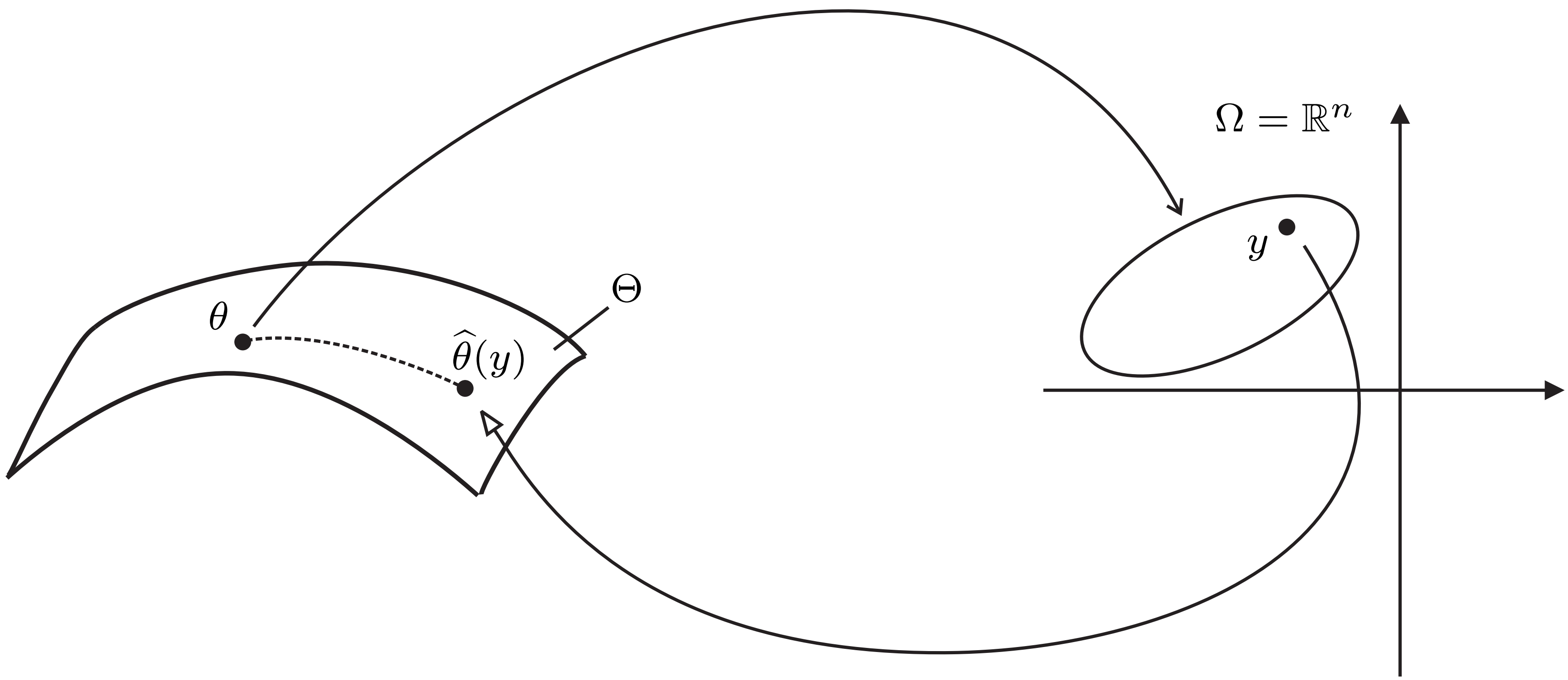


- **Cramér-Rao Bound (CRB):**

$$\mathsf{var}_\theta\left(\widehat{\theta}\right) = \mathsf{E}_\theta\left\{d\left(\theta, \widehat{\theta}(Y)\right)^2\right\} \geq \mathsf{tr}\left(I(\theta)^{-1}\right)$$

Application: performance bounds

- **Riemannian setup:**



- **Intrinsic Variance Lower Bound (IVLB):**

$$\mathsf{var}_\theta\left(\widehat{\theta}\right) = \mathsf{E}_\theta\left\{d\left(\theta, \widehat{\theta}(Y)\right)^2\right\} \geq \text{IVLB}$$

Application: performance bounds

- **Theorem (IVLB).** Suppose:
  - The sectional curvature of Θ is upper bounded by $C \geq 0$
  - $+$ some technical conditions

Then,

$$\operatorname{var}_\theta\left(\widehat{\theta}\right) \geq \begin{cases} \lambda_\theta & , \quad \text{if } C = 0 \\ \\ \dfrac{\lambda_\theta C + 1 - \sqrt{2\lambda_\theta C + 1}}{C^2 \lambda_\theta / 2} & , \quad \text{if } C > 0 \end{cases}$$

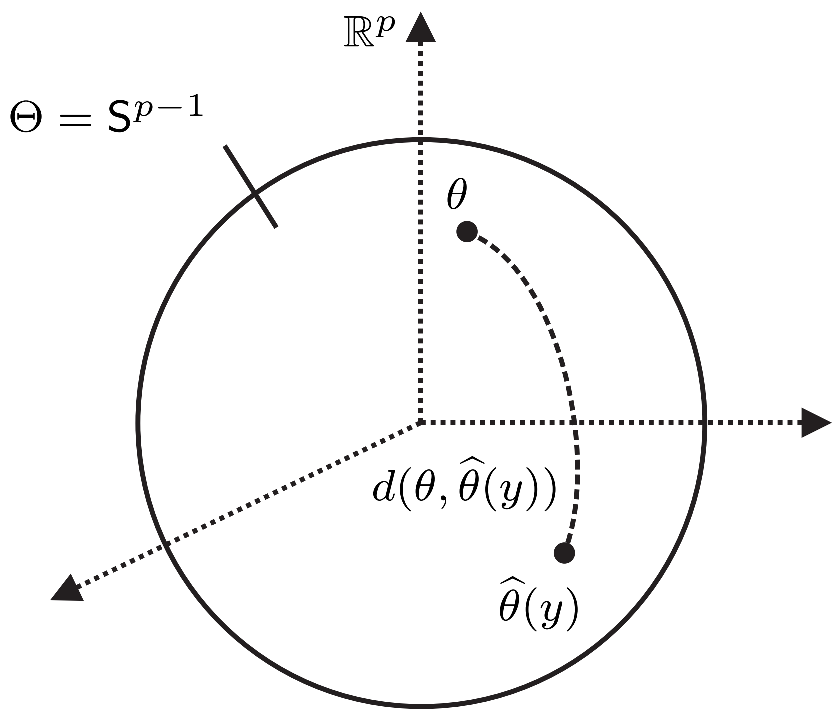
where:

- $\lambda_\theta = \operatorname{tr}(I_\theta^{-1}) \qquad (I_\theta = \text{ Fisher tensor })$

- When $C = 0$, IVLB$\equiv$CRB

Example: inference on S^{p-1}

- $\mathsf{S}^{p-1} = \{x \in \mathbb{R}^p \,:\, \|x\| = 1\}$ is the unit-sphere in $\mathbb{R}^p$



- **Geometry of** Θ: $d(\theta, \widehat{\theta}(y)) = \mathrm{acos}(\theta^T \widehat{\theta}(y))$ and $C = 1$

Example: inference on S^{p-1}

- **Observation**: $y = \theta + w \in \mathbb{R}^p \;\; (p = 10)$
  - $\theta \in \Theta = \mathsf{S}^{p-1}$
  - $w \sim \mathcal{N}(0, \sigma^2 I_p)$

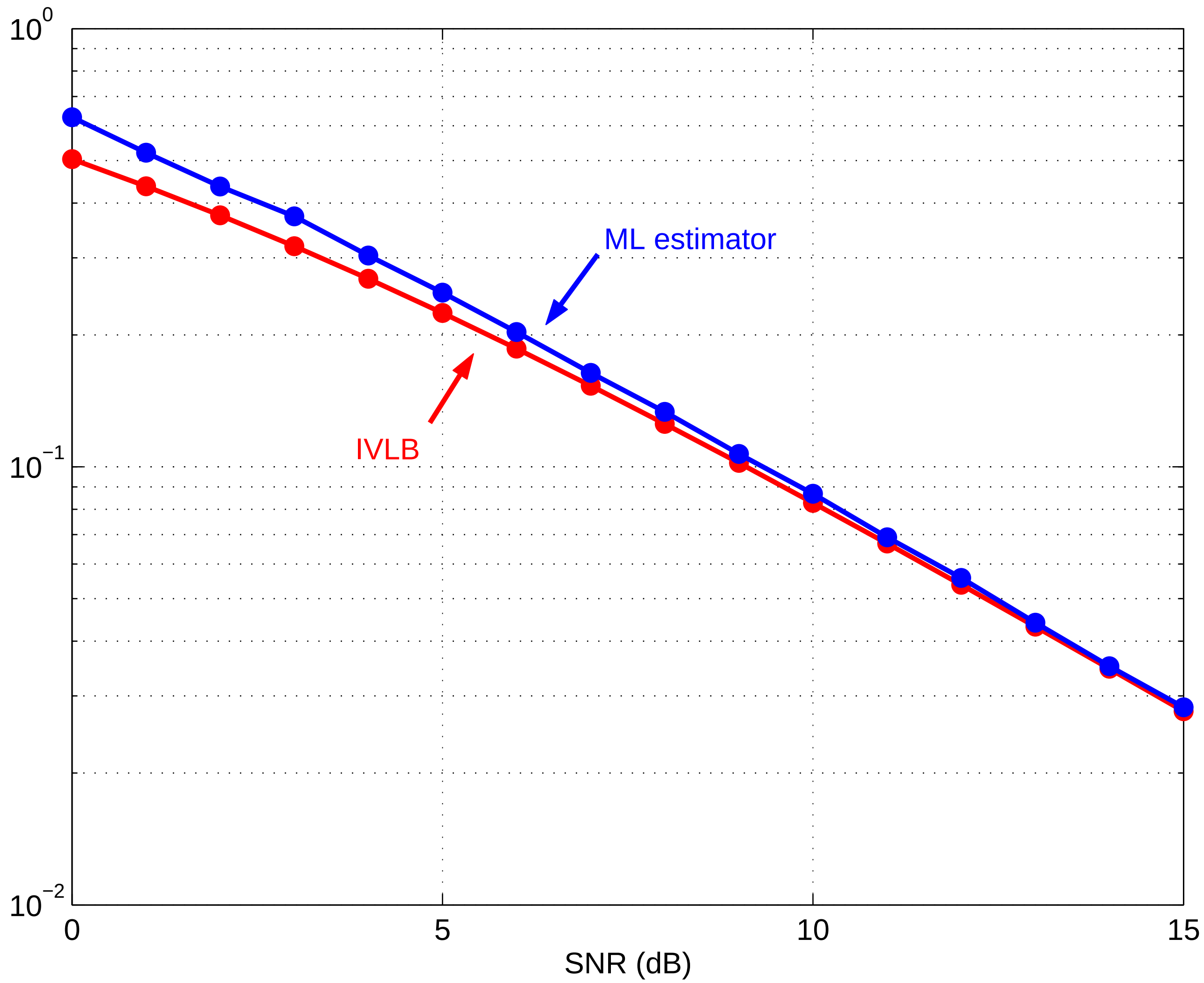
- Maximum-likelihood estimator:

$$\widehat{\theta}(y) = \frac{y}{\|y\|}$$

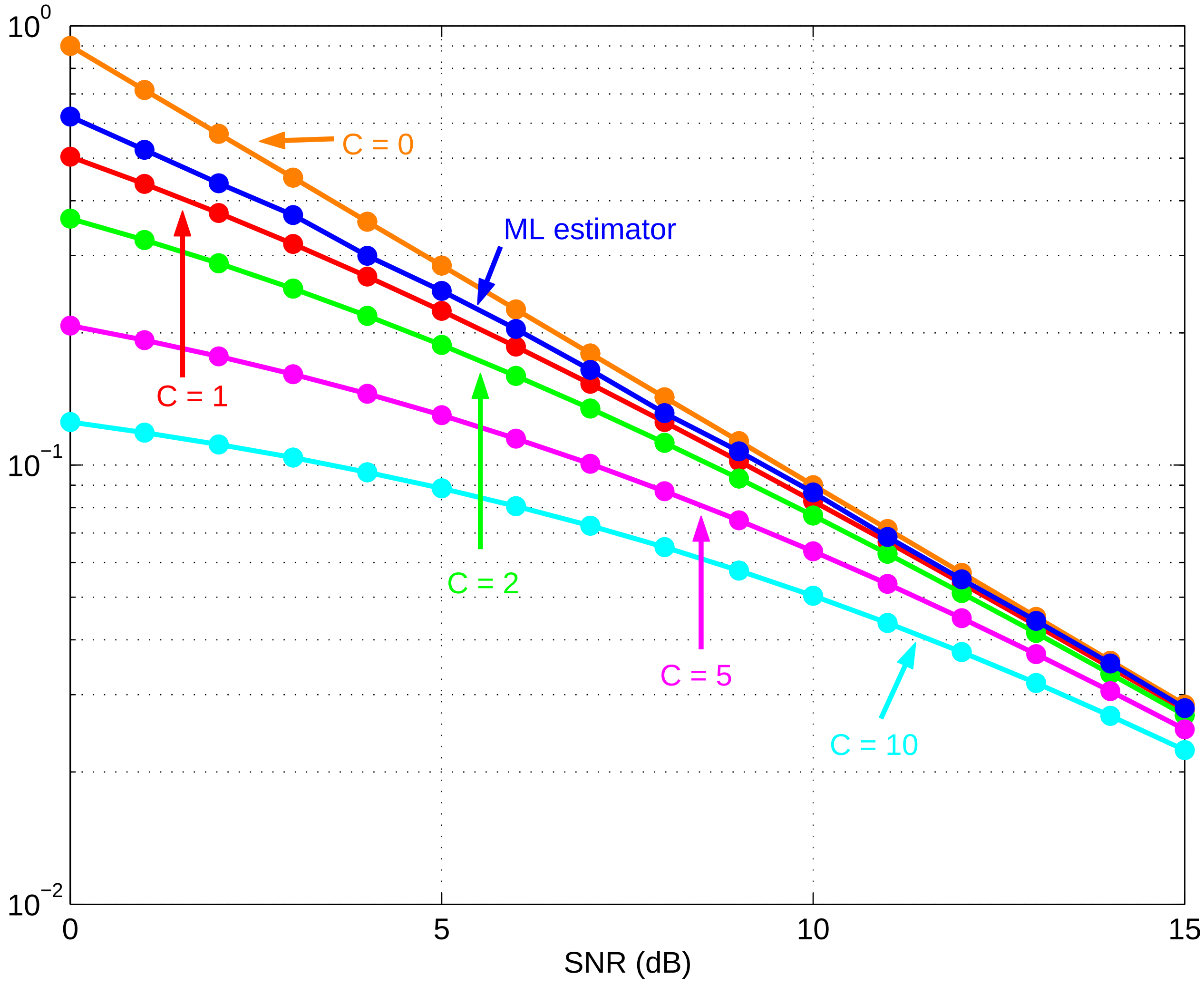
- Signal-to-noise ratio:

$$\mathsf{SNR} = \frac{\mathsf{E}\left\{\|\theta\|^2\right\}}{\mathsf{E}\left\{\|w\|^2\right\}} = \frac{1}{p\,\sigma^2}$$

Example: inference on $\mathbb{S}^{p-1}$



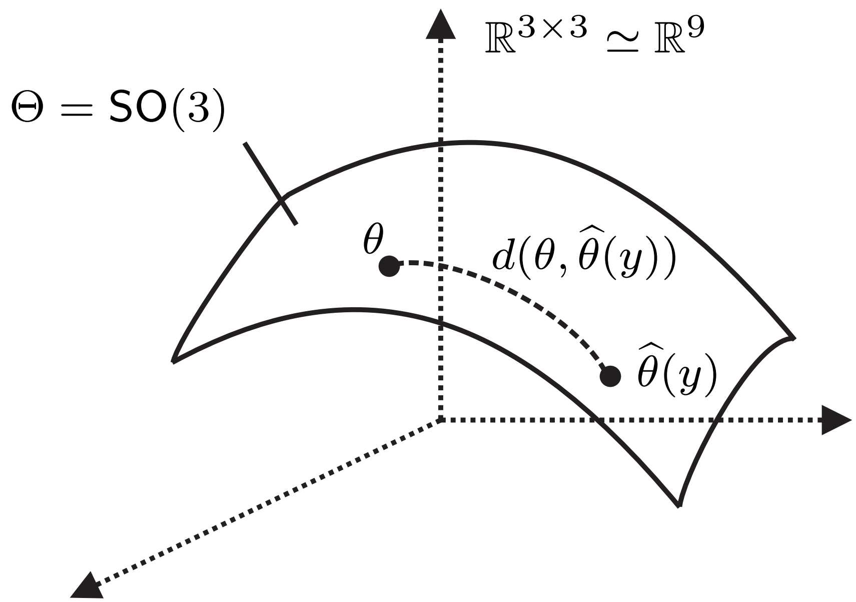
Example: inference on $\mathbb{S}^{p-1}$



Example: inference on SO(3)

- SO(3) is the special orthogonal group:

$$\mathsf{SO}(3) = \left\{ Q \in \mathbb{R}^{3\times 3} \, : \, Q^\top Q = I_3, \det(Q) = 1 \right\}$$



- **Geometry of** Θ: $d(\theta, \widehat{\theta}(y)) = \sqrt{2}\,\mathsf{acos}(0.5[\mathsf{tr}(\theta^\top \widehat{\theta}(y)) - 1])$ and $C = 1/8$

Example: inference on SO(3)

- **Observation**: $Y = \theta X + W \in \mathbb{R}^{3\times k}$ $(k = 10)$
  - $\theta \in \Theta = \mathsf{SO}(3)$: unknown rotation matrix [Procrustean analysis]
  - $X = [\, x_1\, x_2\, \cdots\, x_k\,]$: constellation of known k landmarks in $\mathbb{R}^3$ $(XX^\top = I_3)$
  - $W = [\, w_1\, w_2\, \cdots\, w_k\,]$, $w_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_3)$: additive observation noise

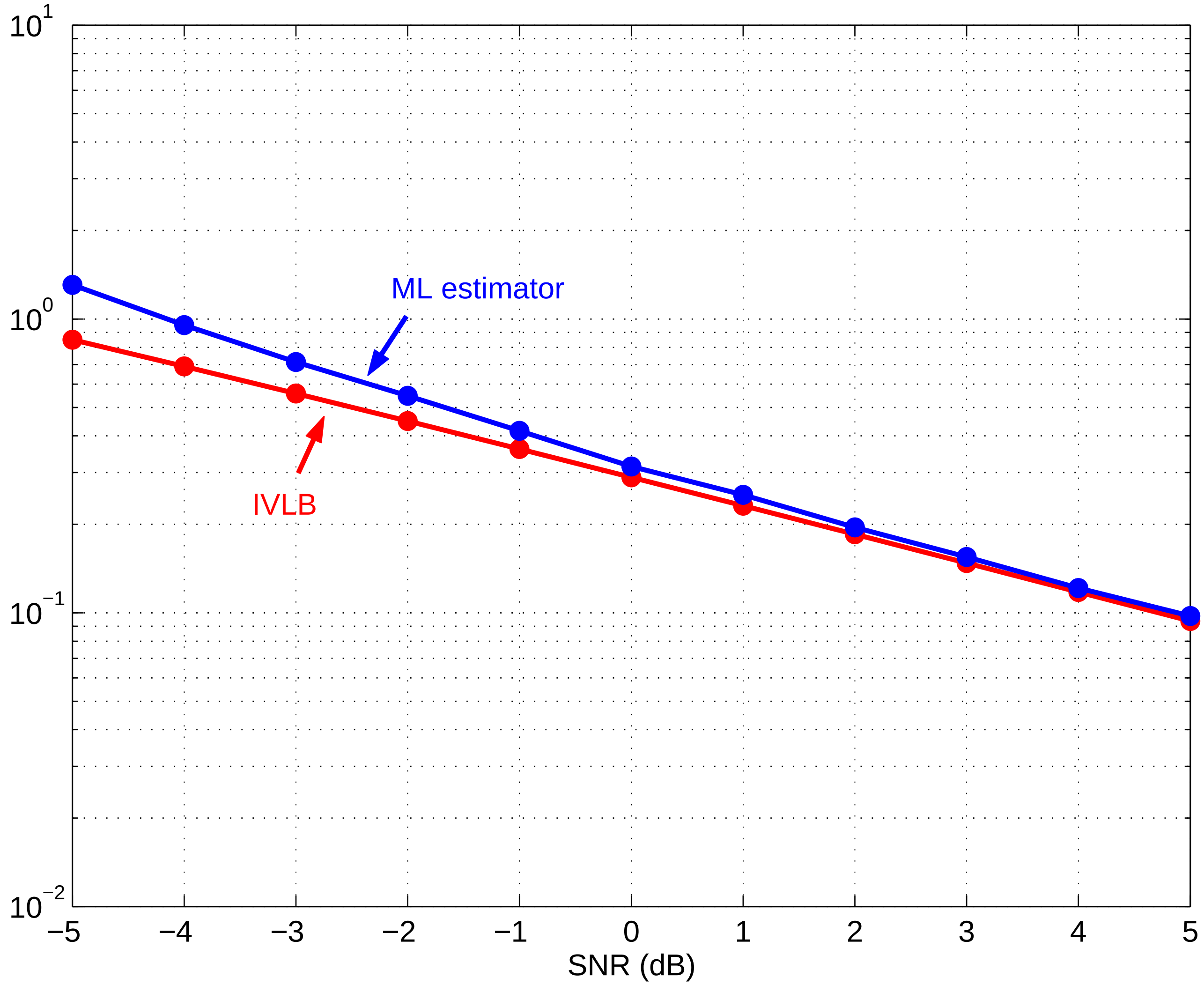
- Maximum-likelihood estimator:

$$\widehat{\theta}(Y) = \cdots \text{ (closed} - \text{form)}$$

- Signal-to-noise ratio:

$$\mathsf{SNR} = \frac{\mathsf{E}\left\{\|\theta X\|^2\right\}}{\mathsf{E}\left\{\|W\|^2\right\}} = \frac{1}{k\,\sigma^2}$$

Example: inference on $\mathrm{SO}(3)$



Example: inference on Grassmann $\mathrm{G}(4,2)$

- Array snapshot: $y[t] = Us[t] + w[t] \in \mathbb{R}^4$
  - $U \in \mathbb{R}^{4\times 2}$: **unknown** orthonormal frame ($U^\top U = I_2$)
  - $s(t) \in \mathbb{R}^2$: vector of i.i.d., zero-mean, unit-power, Gaussian sources
  - $w(t) \in \mathbb{R}^4$: zero-mean, white spatio-temporal Gaussian noise with power σ^2
  - **Observation:** $y = \mathrm{vec}([\, y(1)\, y(2)\, \cdots\, y(T)\,]) \in \mathbb{R}^{4T}$

- **Parameter space:** $\Theta = \left\{U \in \mathbb{R}^{4\times 2} \,:\, U^T U = I_2\right\}$ [Stiefel manifold]

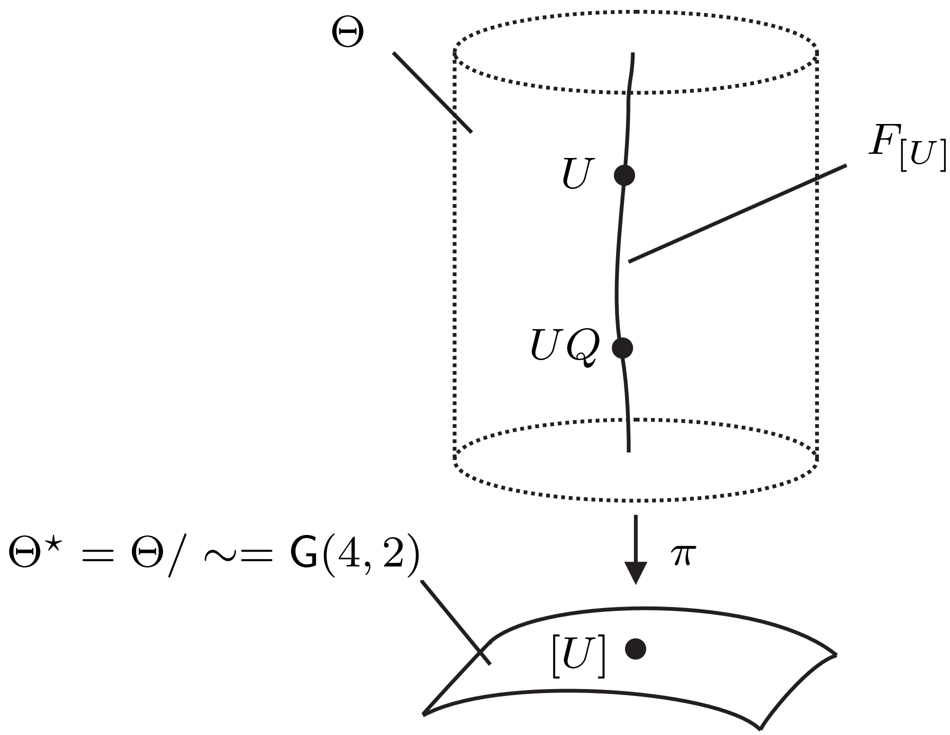
Example: inference on Grassmann $\mathrm{G}(4,2)$

- **Ambiguous parameterization:** y is distributed as $\mathcal{N}(0, C(U))$ where

$$C(U) = I_T \otimes (UU^\top + \sigma^2 I_4)$$

$C(U) = C(UQ)$ for $QQ^\top = I_2 \Rightarrow$ only the 2D-subspace spanned by U is identifiable

- **New parameter space:** $\Theta^\star = \Theta/\sim$ where $U \sim V$ iff $U = VQ$ with $QQ^\top = I_2$



Example: inference on Grassmann $\mathrm{G}(4,2)$

- $\Theta^\star$ can be given the structure of a Riemannian manifold

- **Geodesic distance on** $\Theta^\star$:

$$d([U],[V]) = \sqrt{2}\sqrt{(\mathsf{acos}(\sigma_1))^2 + (\mathsf{acos}(\sigma_2))^2}$$

  where σ_1, σ_2 are the singular values of $U^\top V$

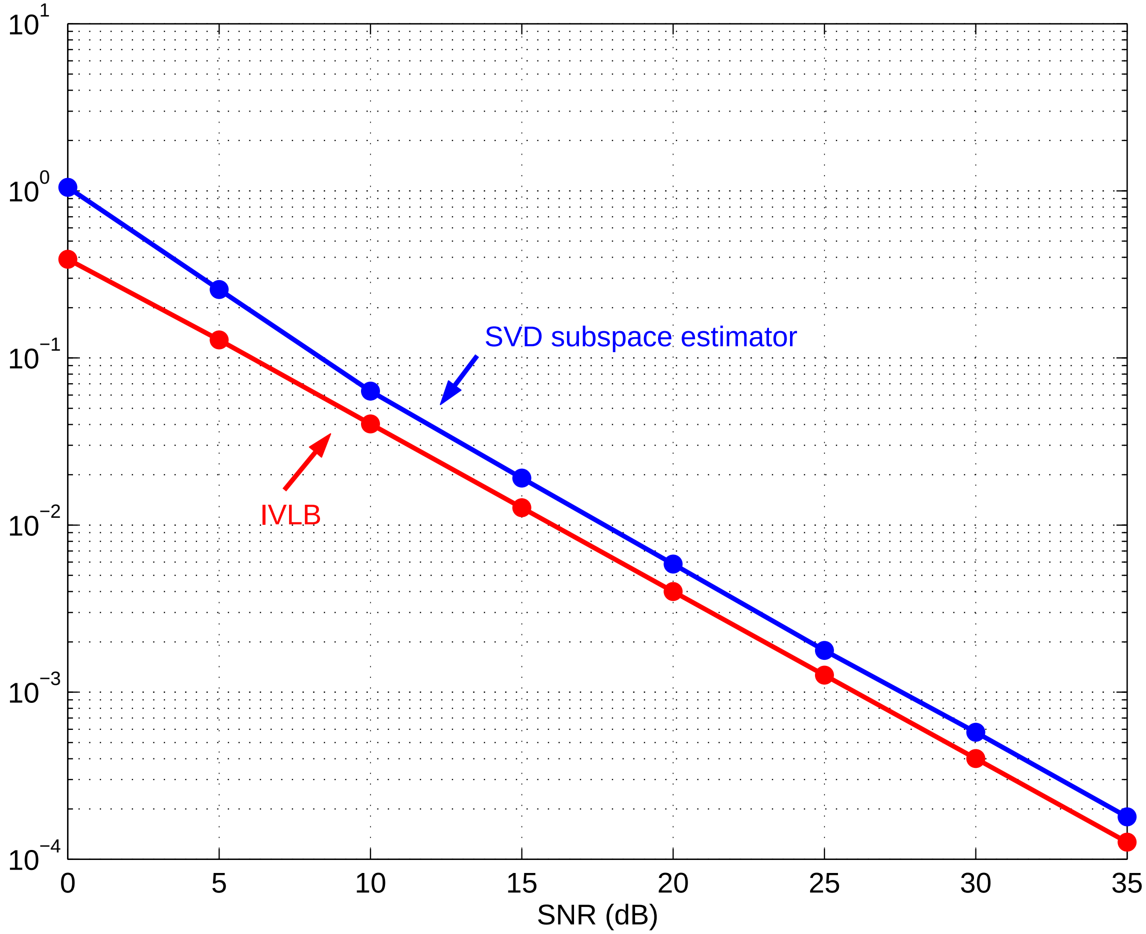
- Bound on sectional curvature:

$$C = 1$$

- $[\widehat{U}]$ is the dominant 2D-subspace from the SVD of $\widehat{R_y} = \frac{1}{T}\sum_{t=1}^{T} y(t)y(t)^\top$

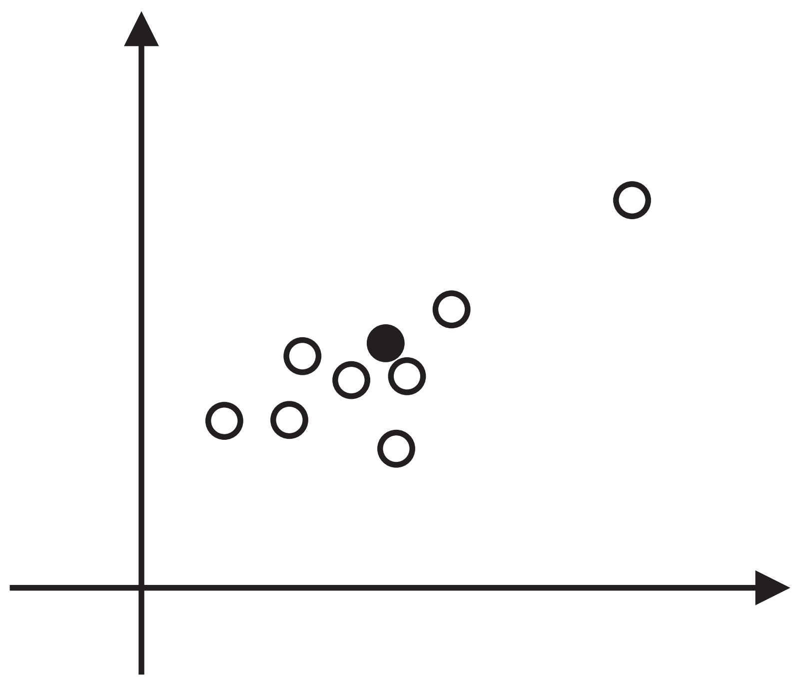
Example: inference on Grassmann $\mathrm{G}(4, 2)$

- **Example:** $T = 10$ data samples



Application: statistics on manifolds

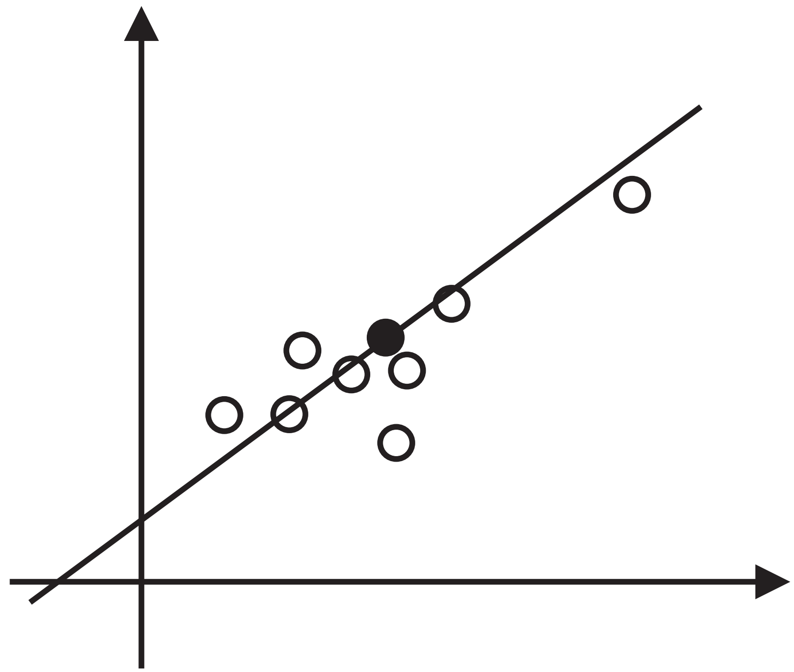
- Basic data compression: clustering



- Simple expression for mean-value: $\overline{x} = \frac{1}{K}\sum_{k=1}^{K} x_k$

Application: statistics on manifolds

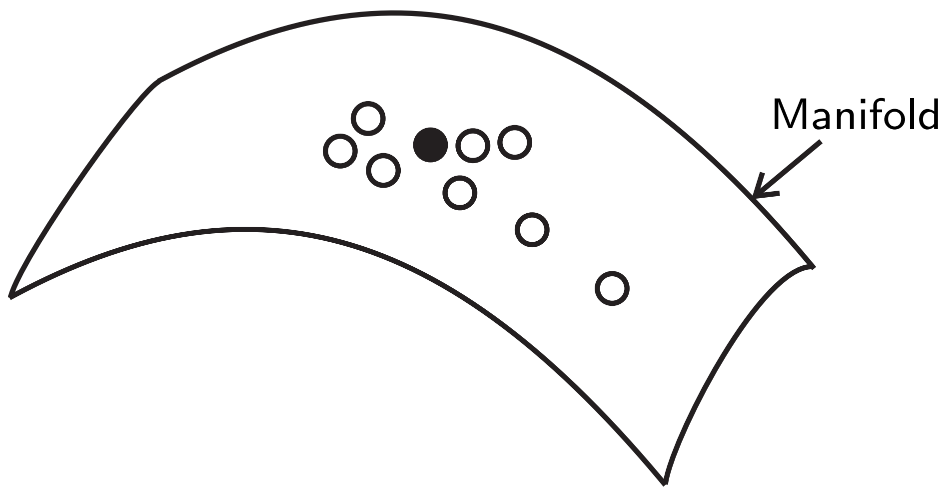
- Basic data compression: principal component analysis (PCA)



- Simple formulas for PCA (eigendecomposition)

Application: statistics on manifolds

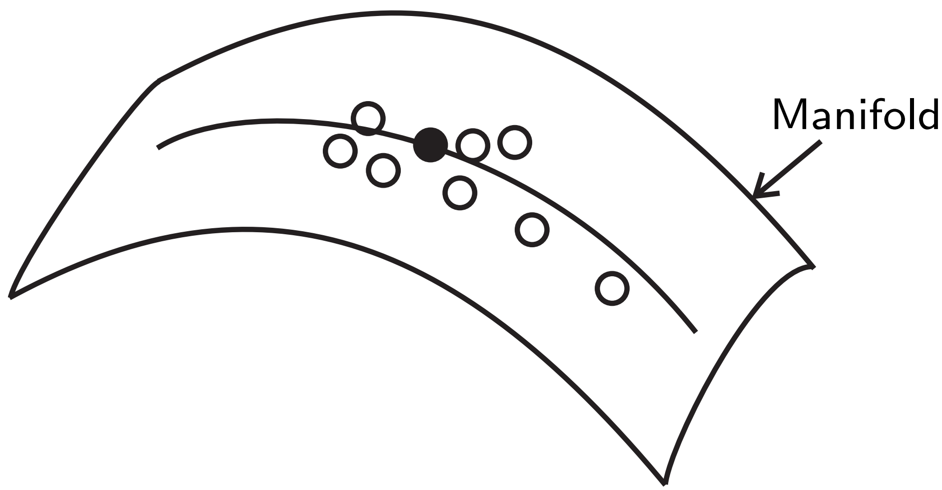
- Generalizations:
  - What is the mean rotation matrix in $\{Q_1, Q_2, \ldots, Q_K\} \subset \mathsf{O}(n)$?
  - What is the mean subspace in $\{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_K\} \subset \mathsf{G}(n, k)$?



- No closed-formulas anymore !

Application: statistics on manifolds

- Generalizations:
  - What is the principal curve through $\{Q_1, Q_2, \ldots, Q_K\} \subset \mathsf{O}(n)$?
  - What is the principal curve through $\{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_K\} \subset \mathsf{G}(n, k)$?



- No closed-formulas anymore !

Application: statistics on manifolds

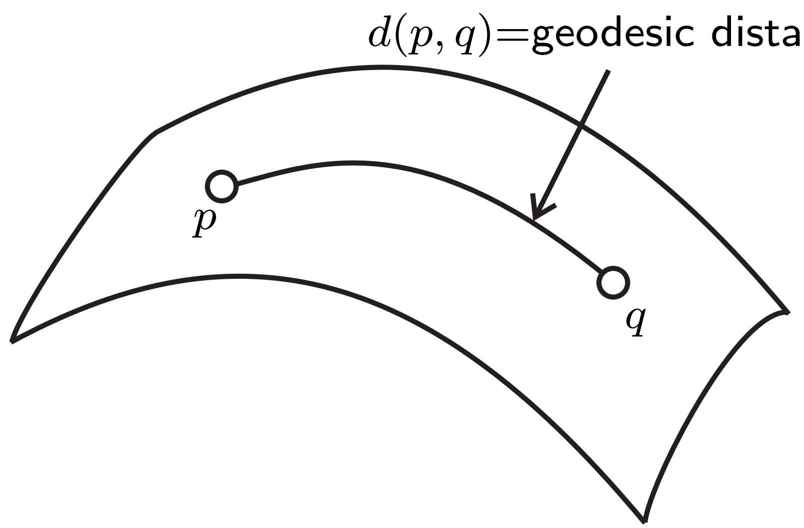
- Applications:
  - Data compression on manifolds (clustering, etc)
  - Study of plate tectonics
  - Sequence-dependent continuum modeling of DNA
  - Encoding of principal diffusion directions in Diffusion Tensor Imaging
  - Analysis of shape in medical imaging
  - . . . many more

Application: statistics on manifolds

- Concepts must be re-formulated:

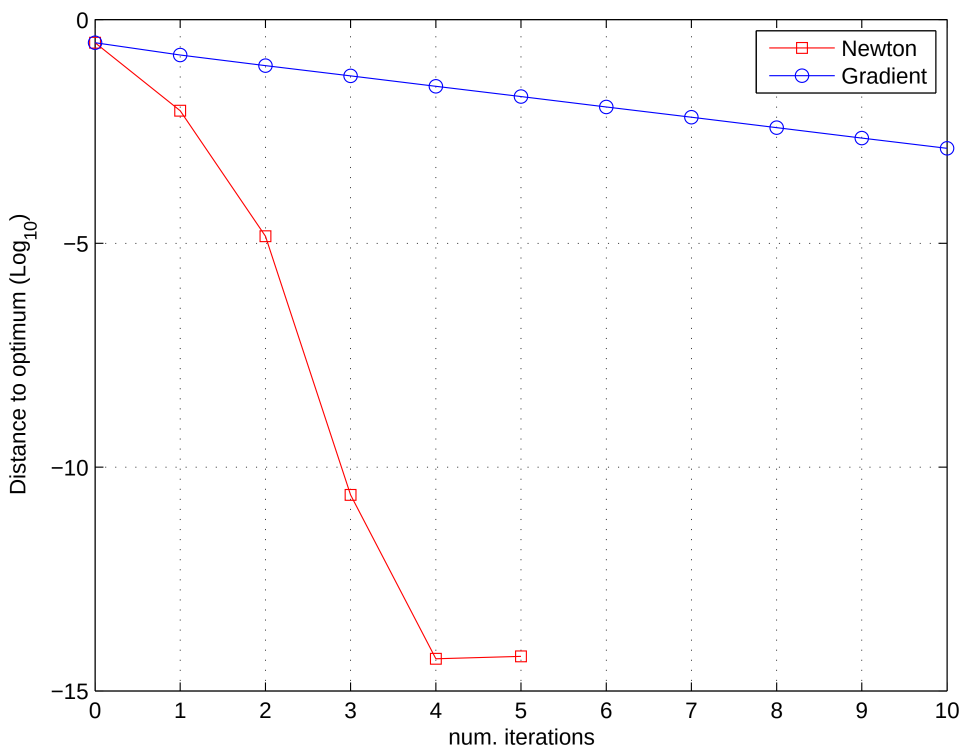
$$\overline{x} = \frac{1}{K}\sum_{k=1}^{K} x_k \quad \rightarrow \quad \overline{x} = \arg\min_{x\in\mathbb{R}^n}\sum_{k=1}^{K}\|x - x_k\|^2 \quad \rightarrow \quad \overline{x} = \arg\min_{x\in\mathbb{R}^n}\sum_{k=1}^{K} d(x_k, x)^2$$

- Center-of-mass on a Riemannian manifold: $\overline{x} \in \arg\min_{x\in\mathbb{R}^n}\sum_{k=1}^{K} d(x_k, x)^2$



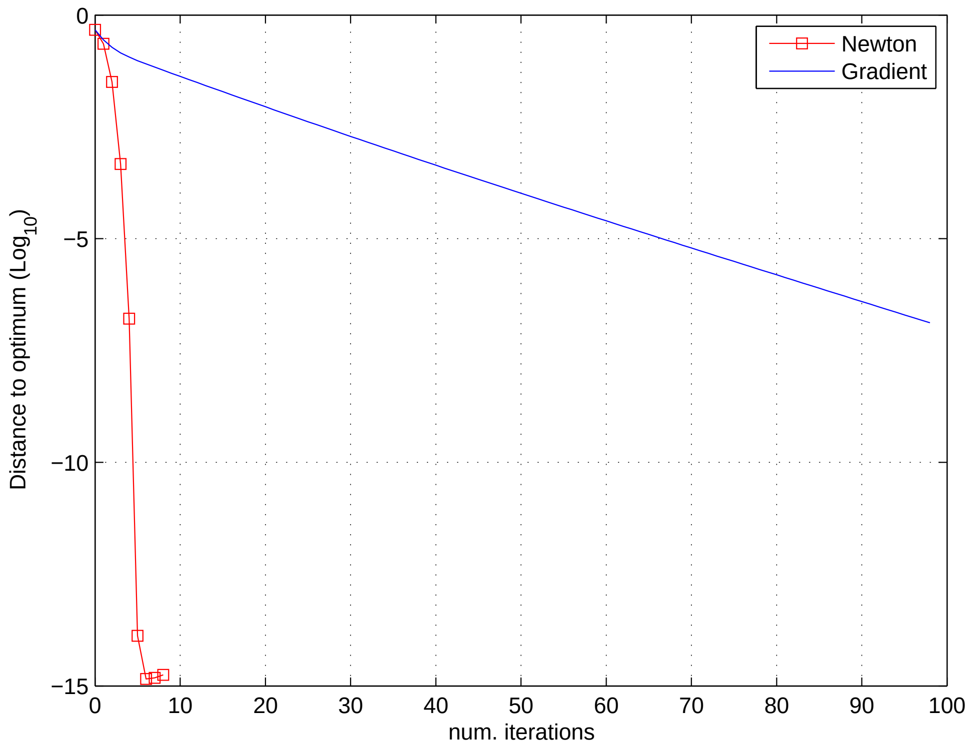
Application: statistics on manifolds

- Example: 5 points in Grassmann $\mathrm{G}(6, 3)$



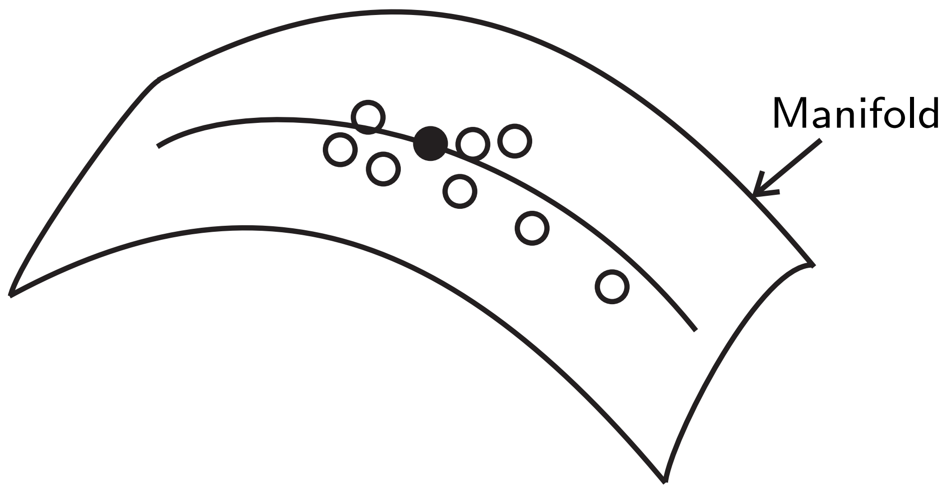
Application: statistics on manifolds

- By-product: MAP estimation on $SE(3)$



Application: statistics on manifolds

- Results for geodesic PCA ...



- ...coming soon !

Course's Table of Contents

- **Three main topics:**
  - Topological manifolds
  - Differentiable manifolds
  - Riemannian manifolds

- **Three layers of structure:**

Riemannian structure	Length of curves ; Geodesics ; Distance ; Connections ; etc
Differentiable structure	Tangent vectors; Smooth maps; Tensors; Integration ; etc
Topological structure	Boundary of sets; Convergent sequences; Continuous maps ; etc
Plain set	

Course's Table of Contents

- **Topological manifolds:** "Introduction to Topological Manifolds", J. Lee, Springer-Verlag
  - Ch.2: Topological spaces
  - Ch.3: New spaces from old
  - Ch.4: Connectedness and compacteness

- **Smooth manifolds:** "Introduction to Smooth Manifolds", J. Lee, Springer-Verlag
  - Ch.2: Smooth maps
  - Ch.3: The tangent bundle
  - Ch.5: Submanifolds
  - Ch.7: Lie group actions
  - Ch.8: Tensors
  - Ch.9: Differental forms
  - Ch.10: Integration on manifolds

Course's Table of Contents

- **Riemannian manifolds:** "Riemannian Manifolds", J. Lee, Springer-Verlag
  - Ch.3: Definitions and examples of Riemannian metrics
  - Ch.4: Connections
  - Ch.5: Riemannian geodesics

Bibliography for the Course

- **Topological manifolds**
  - “Introduction to Topological Manifolds”, J. Lee, Springer-Verlag, 2000
  - “Introduction to Topology and Modern Analysis”, G. Simmons, 1963

- **Smooth manifolds**
  - “Introduction to Smooth Manifolds”, J. Lee, Springer-Verlag, 2002
  - “ An Introduction to Differentiable Manifolds and Riemannian Geometry”, 2nd ed., W.Boothby, Academic Press, 1986
  - “Manifolds, Tensor Analysis and Applications”, R. Abraham *et al.*, Springer-Verlag, 1988
  - “A Comprehensive Introduction to Differential Geometry”, vol.I, M. Spivak, Publish or Perish, 1979
  - “Lectures on Differential Geometry”, S. Chern, W. Chern and K. Lam, World Scientific, 1999

- **Riemannian manifolds**
  - “Riemannian Manifolds”, J. Lee, Springer-Verlag
  - “Riemannian Geometry”, M. Carmo, Birkhauser, 1992

Bibliography

- **Other references (introductory):**
  - "Differential Forms with Applications to the Physical Sciences", H. Flanders, Dover, 1963
  - "Differential Forms with Applications", M. Carmo, Springer-Verlag, 1994

- **Other references (advanced):**
  - "Riemannian Geometry", S. Gallot, D. Hulin and J. Lafontaine, Springer-Verlag, 1987
  - "A Comprehensive Introduction to DG", vol.II-V, M. Spivak, Publish or Perish, 1979
  - "Riemannian Geometry: A Modern Introduction", I. Chavel, Cambridge Press, 1993
  - "Riemannian Geometry and Geometric Analysis", J. Jost, Springer-Verlag, 1998
  - "Foundations of Differential Geometry", vol. I-II, S. Kobayashi and K. Nomizu, Wiley 1969
  - "DG, Lie Groups and Symmetric Spaces", S. Helgason, Academic Press, 1978

- Many others. . .

Grading

- **Grade** $=$ Homework (50%) $+$ Project (50%)

- **Homeworks:** 3 sets

- **Project (individual)**: 1 of 2 choices
  - I assign a paper
  - the student proposes a topic

In either case: the output is a public presentation of the project