The Mathematics Behind Principal Component Analysis

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Principal component analysis

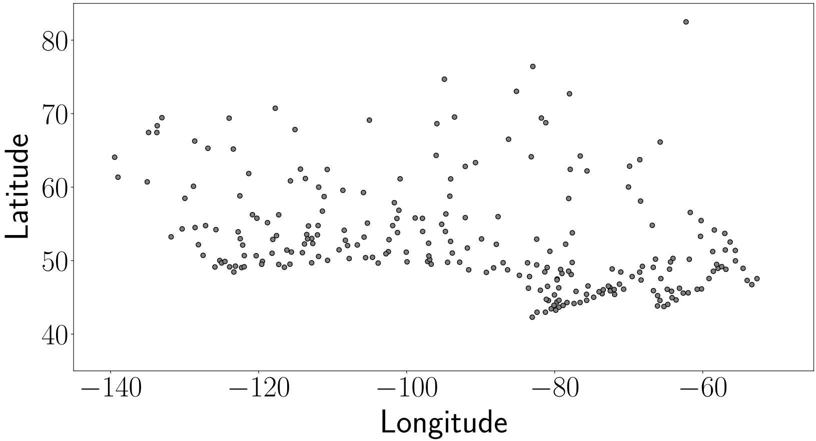
Dataset $X = \{x_1, x_2, \ldots, x_n\}$ with d features

1. Compute sample covariance matrix Σ_X
2. Eigendecomposition of Σ_X yields principal directions $u_1, \ldots, u_d$
3. Center the data and compute principal components

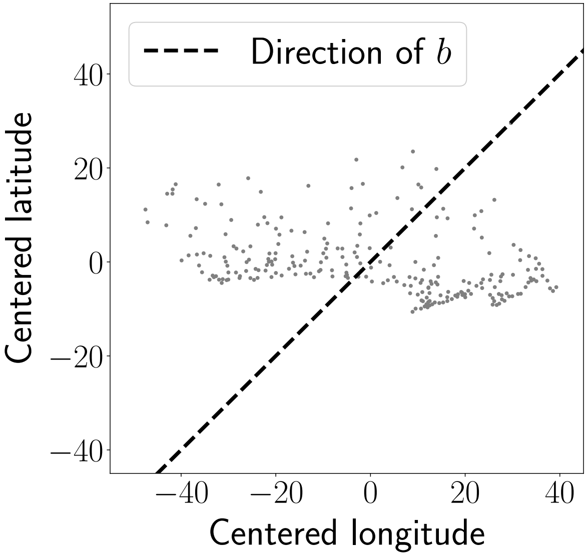
$$w_j[i] := u_j^T \operatorname{ct}(x_i), \quad 1 \leq i \leq n, \ 1 \leq j \leq d$$

where $\operatorname{ct}(x_i) := x_i - m(X)$

Cities in Canada



Variance in a certain direction?



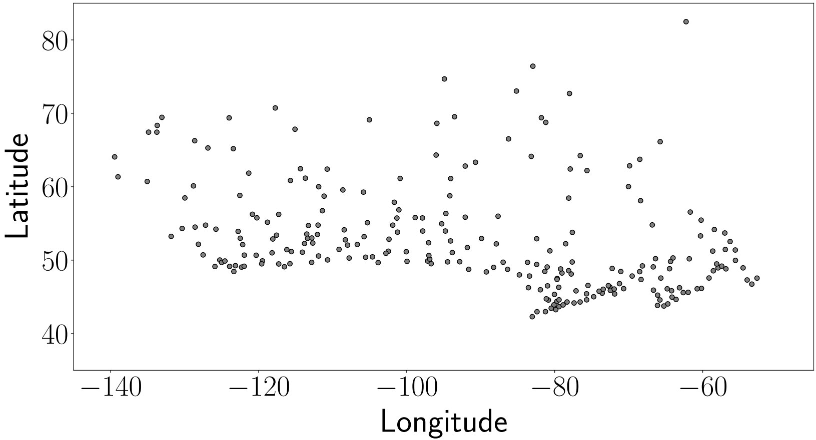
Sample variance of linear combination

Dataset: $X = \{x_1, \ldots, x_n\}$

$$X_a := \left\{ a^T x_1, \ldots, a^T x_n \right\}$$

$$v\left(X_a\right) = a^T \Sigma_X a = q(a)$$

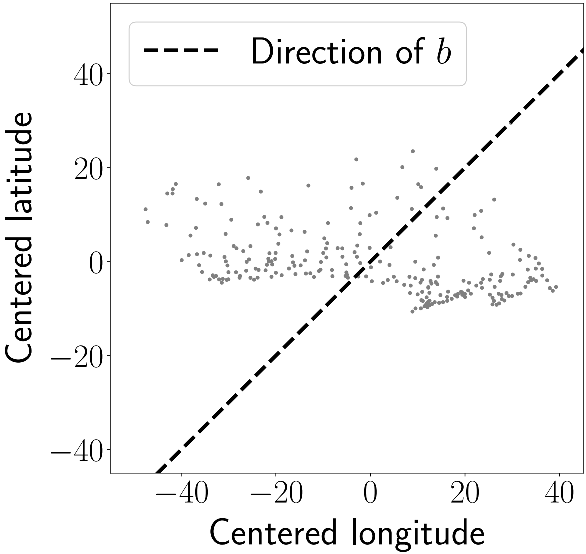
Cities in Canada



Sample covariance matrix:

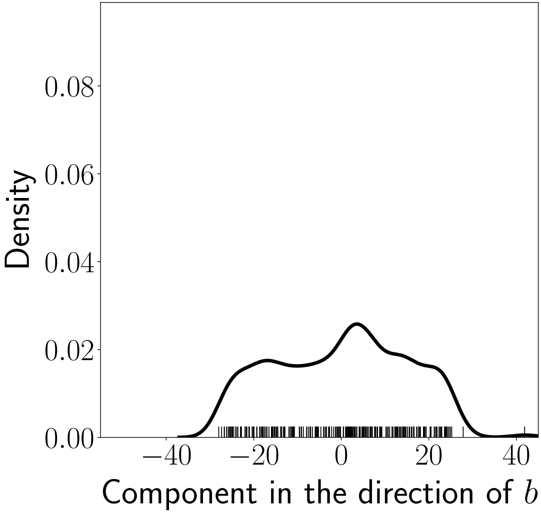
$$\Sigma_X = \begin{bmatrix} 524.9 & -59.8 \\ -59.8 & 53.7 \end{bmatrix}$$

Variance in a certain direction?

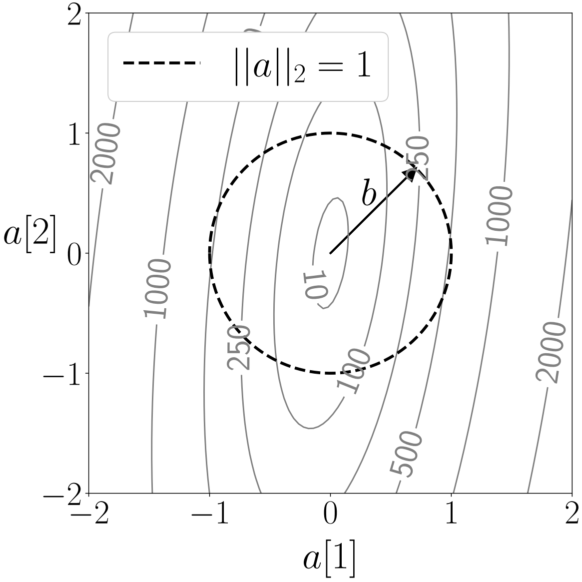


Variance in a certain direction

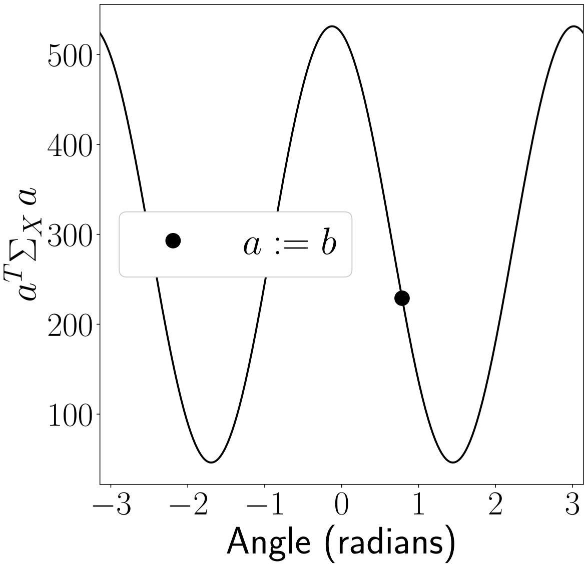
$$v(X_b) = q(b) = b^T \Sigma_X b = 229$$



Quadratic form $q(a) := a^T \Sigma_X a = v(X_a)$



$q(a)$ for $||a||_2 = 1$: Maximum?



Is there a maximum? Yes!

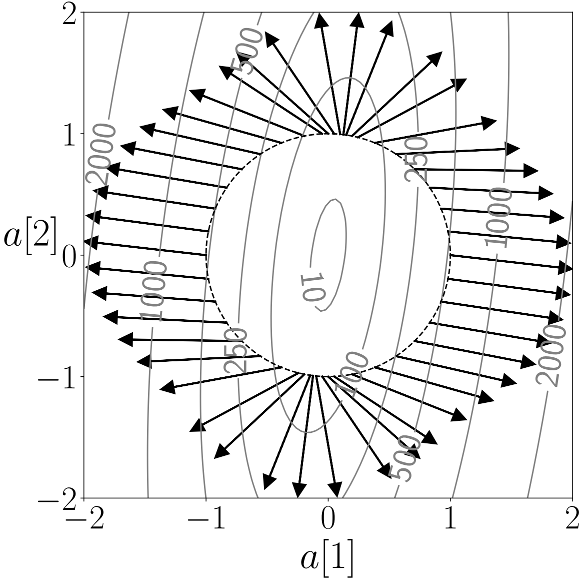
- The function is continuous (second-order polynomial)
- Unit sphere is closed and bounded (contains all limit points)
- Image of unit sphere is also closed and bounded
- Image cannot grow towards limit it does not contain

Maximum

There exists $u_1 \in \mathbb{R}^d$ such that

$$u_1 = \arg\max_{||a||_2=1} q(a)$$

Gradient $\nabla q(a) = 2\Sigma_X a$



Directional derivative

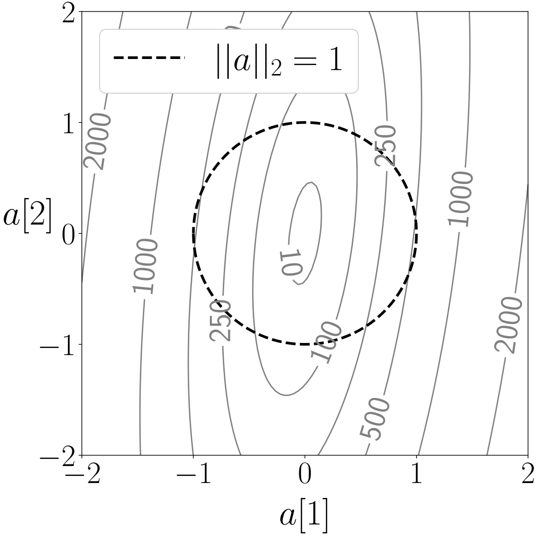
$$
\begin{aligned}
q_h'(b) &:= \lim_{\epsilon \to 0} \frac{q(b + \epsilon h) - q(b)}{\epsilon} \\
&= \nabla q(b)^T h
\end{aligned}
$$

If $q_h'(b) > 0$, then $q(b + \epsilon h) > q(b)$ for small enough $\epsilon > 0$

At the maximum u_1 can we have $\nabla q(u_1)^T h > 0$ if $u_1 + \epsilon h$ is in constraint set? No!

Wait a minute, *can* $u_1 + \epsilon h$ *be in the constraint set?*

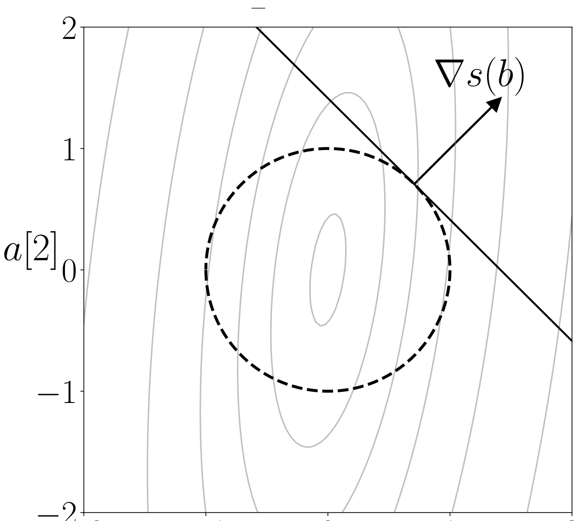
No



Tangent hyperplane

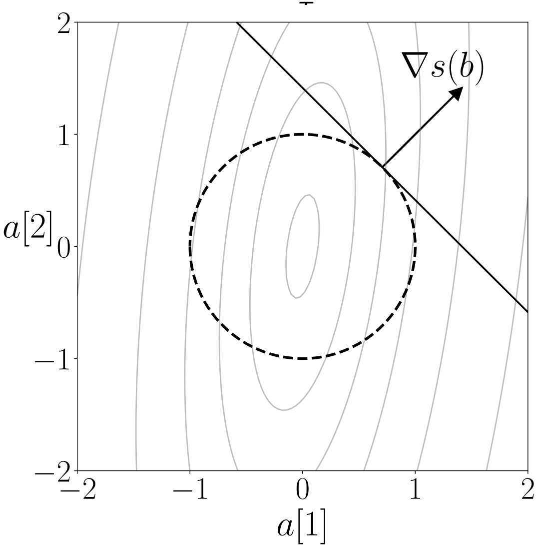
Unit sphere is level surface of $s(a) := a^T a$

y is in the tangent plane at b if



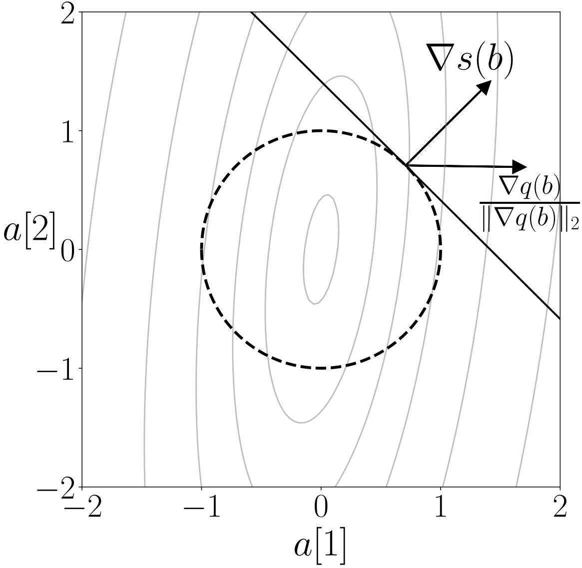
Tangent hyperplane

If $y - b$ is very small



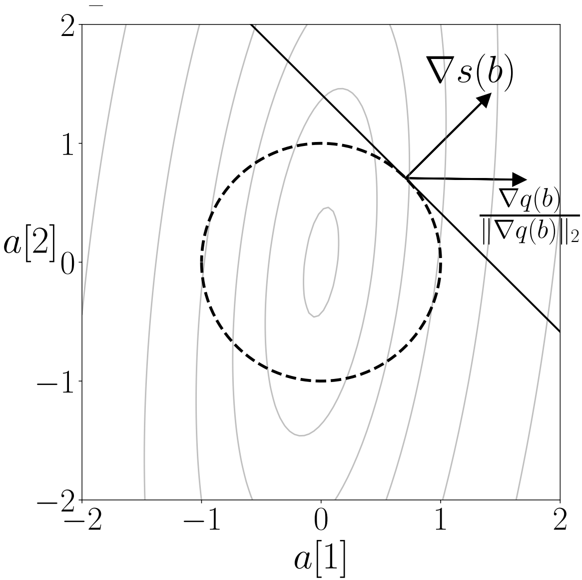
If y is in tangent plane it is almost in the same level set as b

Can this point be a maximum?



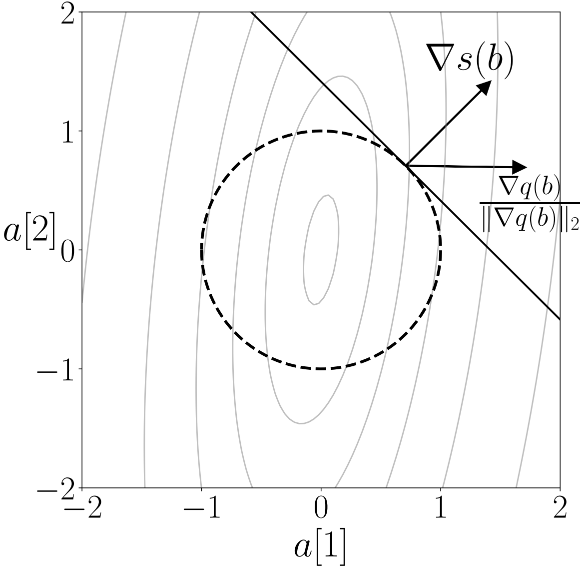
Can this point be a maximum?

For h such that $b + \epsilon h$ is in tangent plane

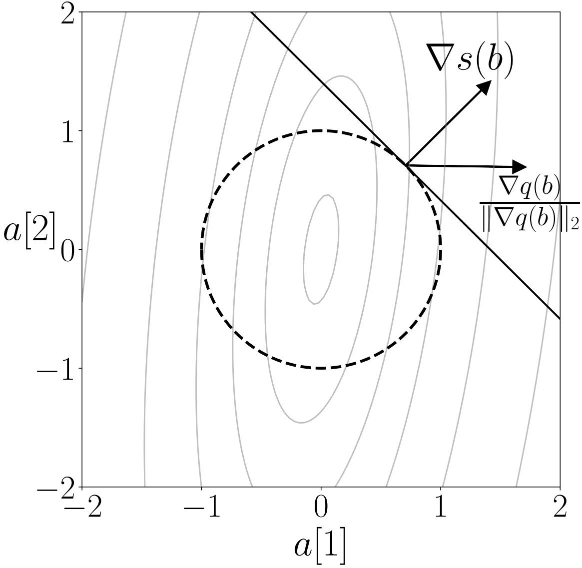


Can this point be a maximum? No!

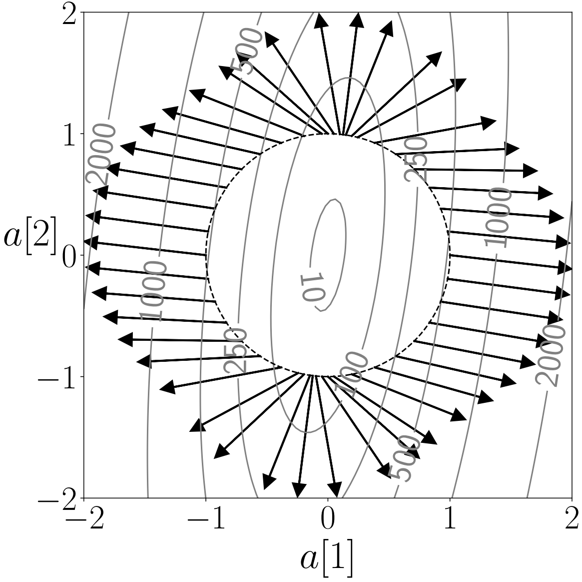
$b + \epsilon h$ is in the tangent plane, there is y on unit sphere, such that



What do we need?



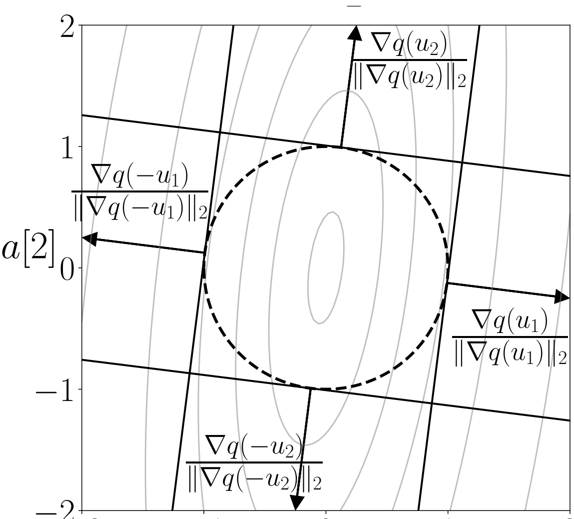
Where is the maximum?



Where's the maximum u_1?

$\nabla q(u_1)$ and $\nabla s(u_1)$ must be collinear

For $u_1 + \epsilon h$ in tangent plane



Eigenvectors

At maxima/minima $\nabla q(u) = \lambda \nabla s(u)$ for some λ

$$\begin{aligned}\nabla q(u) &= \nabla(u^T \Sigma_X u) \\ &= 2\Sigma_X u\end{aligned}$$

$$\begin{aligned}\nabla s(u) &= \nabla(u^T u) \\ &= 2u\end{aligned}$$

$$\Sigma_X u = \lambda u$$

so u is an eigenvector!

Spectral theorem

Same argument can be applied to minima

And to maxima in directions orthogonal to u_1

Spectral theorem

If $M \in \mathbb{R}^{d \times d}$ is symmetric, then it has an eigendecomposition

$$M = \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix}^T,$$

Eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ are real

Eigenvectors $u_1, u_2, \ldots, u_n$ are real and orthogonal

Spectral theorem

$$u_1 = \arg \max_{||a||_2=1} a^T M a$$

$$\lambda_1 = \max_{||a||_2=1} a^T M a$$

$$u_k = \arg \max_{||a||_2=1, a \perp u_1, \ldots, u_{k-1}} a^T M a, \quad 2 \leq k \leq d-1$$

$$\lambda_k = \max_{||a||_2=1, a \perp u_1, \ldots, u_{k-1}} a^T M a, \quad 2 \leq k \leq d-1$$

$$u_d = \arg \min_{||a||_2=1} a^T M a$$

$$\lambda_d = \min_{||a||_2=1} a^T M a$$

Principal directions

Let $u_1, \ldots, u_d$ be the eigenvectors and $\lambda_1 > \ldots > \lambda_d$ the eigenvalues of Σ_X

$$\lambda_1 = \max_{||a||_2=1} a^T \Sigma_X a = \max_{||a||_2=1} v(X_a)$$

$$u_1 = \arg \max_{||a||_2=1} v(X_a)$$

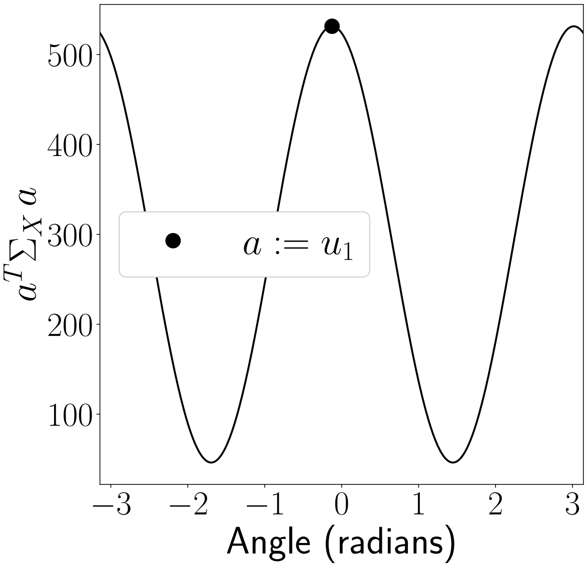
$$\lambda_k = \max_{||a||_2=1, a \perp u_1, \ldots, u_{k-1}} v(X_a), \quad 2 \leq k \leq d$$

$$u_k = \arg \max_{||a||_2=1, a \perp u_1, \ldots, u_{k-1}} v(X_a), \quad 2 \leq k \leq d$$

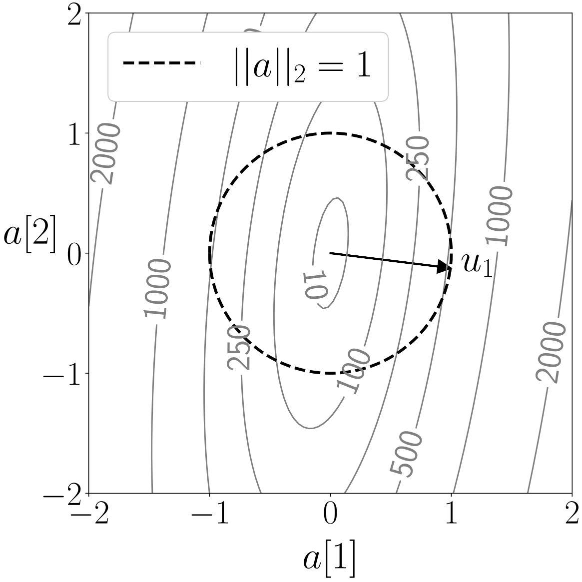
$$\lambda_d = \min_{||a||_2=1} v(X_a)$$

$$u_d = \arg \min_{||a||_2=1} v(X_a)$$

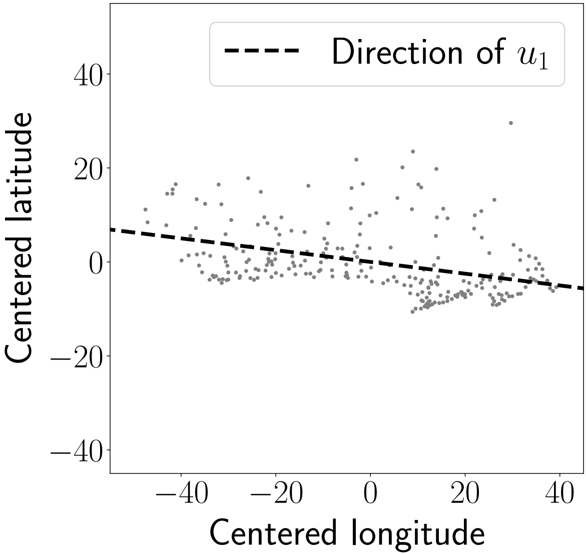
First principal direction



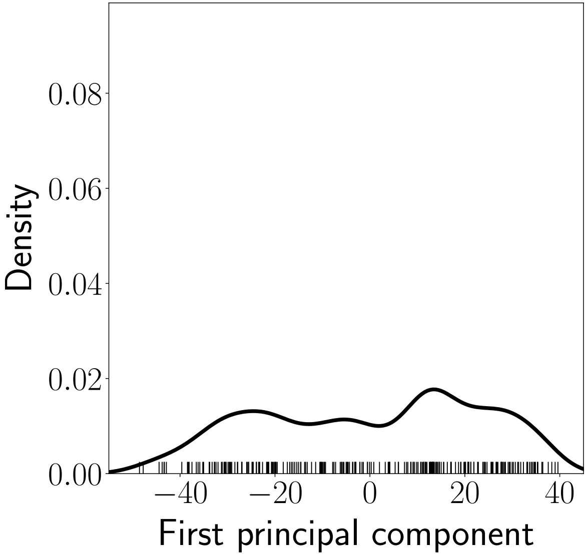
Quadratic form $a^T \Sigma_X a = v(X_a)$



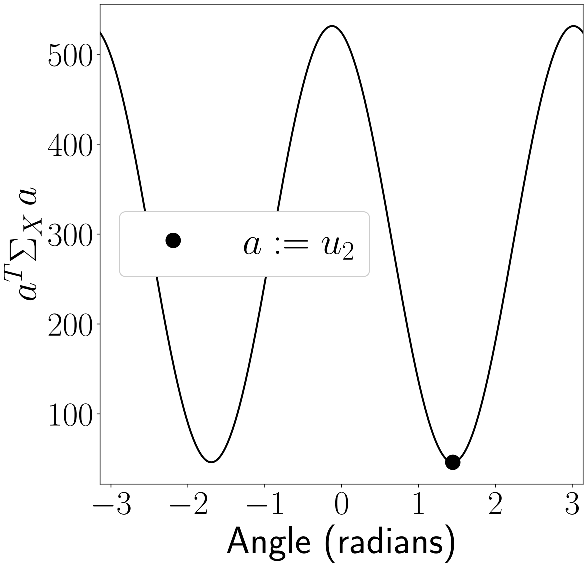
Data


Direction of u_1
40
20
0
−20
−40
Centered latitude
−40
−20
0
20
40
Centered longitude

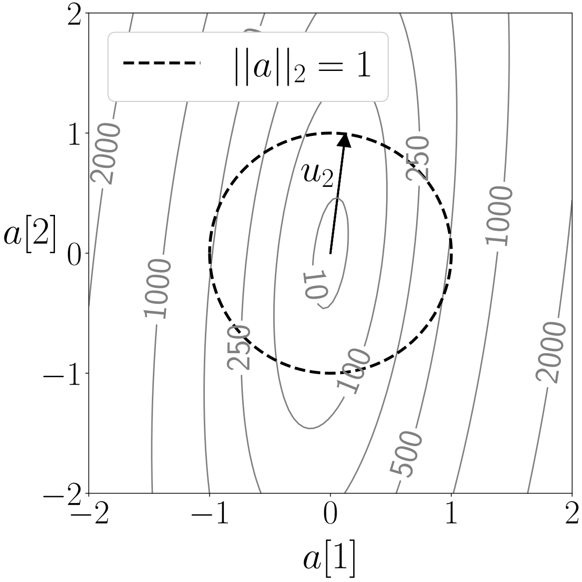
First principal component



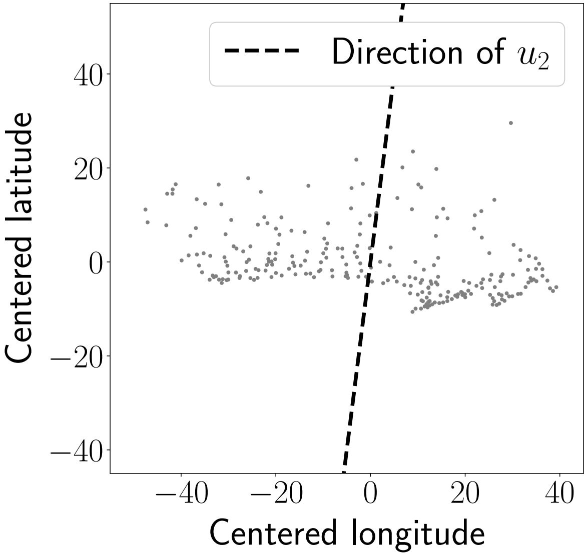
Second principal direction



Quadratic form $a^T \Sigma_X a = v(X_a)$

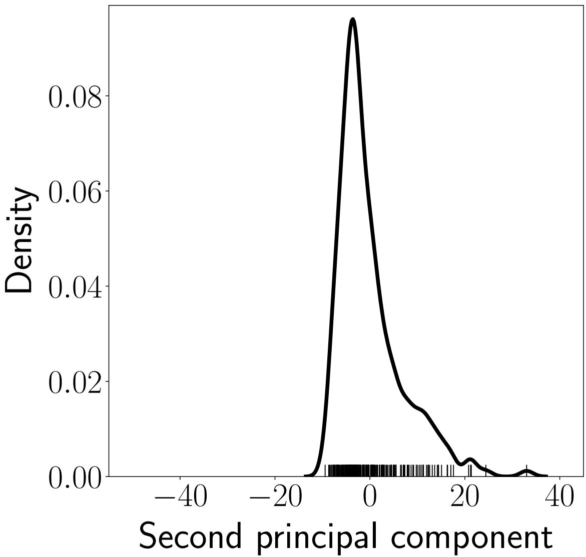


Data



Direction of u_2
40
20
0
−20
−40
Centered latitude
−40
−20
0
20
40
Centered longitude

Second principal component



What have we learned

How to prove the spectral theorem

Why eigendecomposition of the covariance matrix yields directions of maximum/minimum variance