UNIVERSITY OF WATERLOO | FACULTY OF SCIENCE

Compression techniques

Marco Bonici

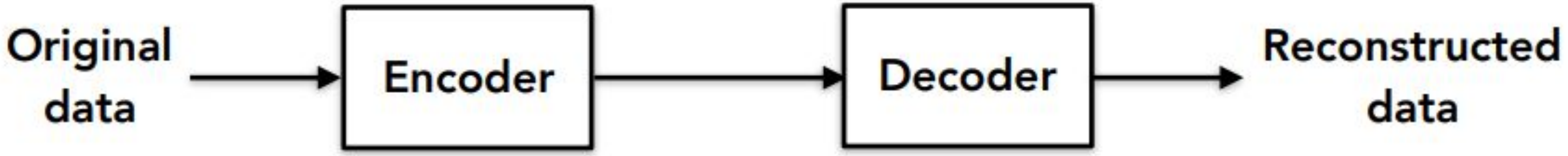
Compression techniques

In Cosmology and Astrophysics, we usually deal with huge datasets

Euclid 2D observables datavector will contain a few thousands elements

This puts huge pressure on constructing the covariance matrices for the analysis (see Pierre's lecture on covariance matrices) as we are gonna need many expensive simulations to compute them

In general, you might have a big dataset and ask yourself: is there a way to reduce dimension of my dataset without losing the information contained in it? (Also useful as de-noiser)



The general idea of compression is to represent the same information you have in your dataset with fewer bits

1. You take an encoder (a compression algorithm), you apply it to your data
2. You decode, i.e. you try to reconstruct your original data given the compressed ones
3. You compare the reconstructed data with the original one: are they close enough (i.e. was the compression loss-less)?

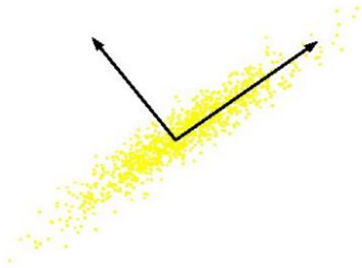
Several approaches exist to perform the compression step

- Principal Component Analysis
- Massively Optimised Parameter Estimation and Data compression
- Singular Value Decomposition
- Normalizing Flows
- Neural Networks

We will cover the first two in our lectures, will giving some hints for the latter in the remainder

What is PCA

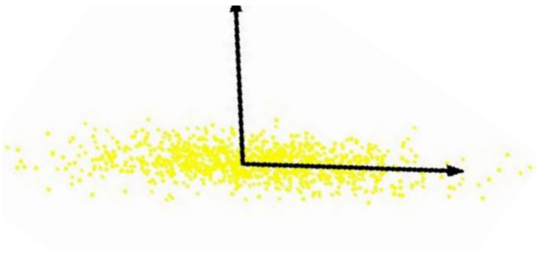
What is PCA: Unsupervised technique for extracting variance structure from high dimensional datasets.



PCA is an orthogonal projection or transformation of the data into a (possibly lower dimensional) subspace so that the variance of the projected data is maximized.

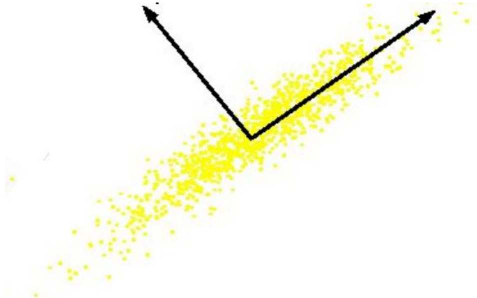
How does it work?

Intrinsically lower dimensional than the dimension of the ambient space.



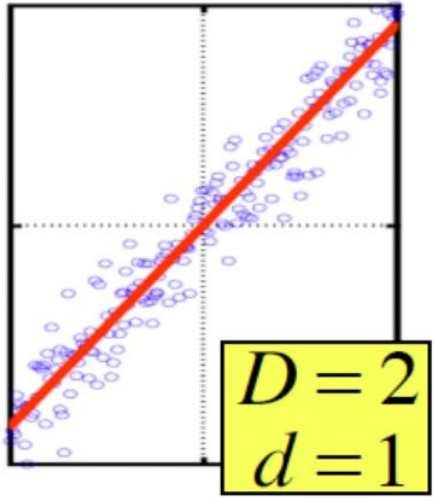
One relevant feature

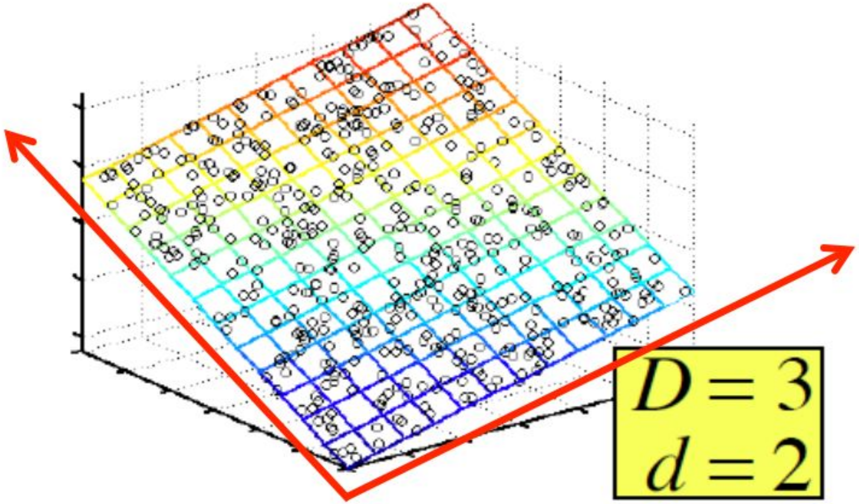
If we rotate data again, only one coordinate is more important.



Two relevant features

Question: Can we transform the features so that we only need to preserve one latent feature?





In case where data lies on or near a low d-dimensional linear subspace, axes of this subspace are an effective representation of the data.

Identifying the axes is known as **Principal Components Analysis**, and can be obtained by using classic matrix computation tools (Eigen or Singular Value Decomposition).

PCA in practice (1)

mean subtracted

Given the **centered** data $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, compute the principal vectors:

$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m}\sum_{i=1}^{m}\{(\mathbf{w}^T\mathbf{x}_i)^2\}$$

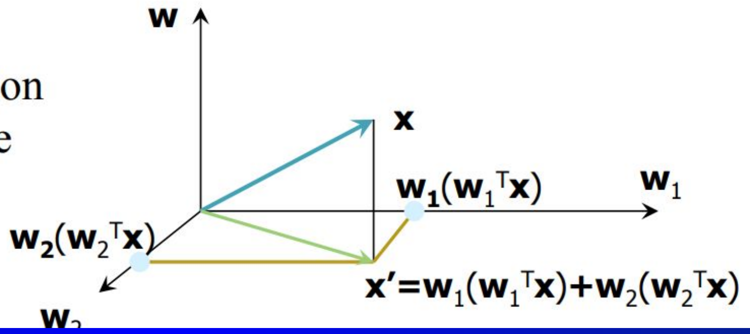
1st PCA vector

We maximize the variance of projection of $\mathbf{x}$

$$\mathbf{w}_k = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m}\sum_{i=1}^{m}\{[\mathbf{w}^T(\mathbf{x}_i - \underbrace{\sum_{j=1}^{k-1}\mathbf{w}_j\mathbf{w}_j^T\mathbf{x}_i}_{\mathbf{x}' \text{ PCA reconstruction}})]^2\}$$

kth PCA vector

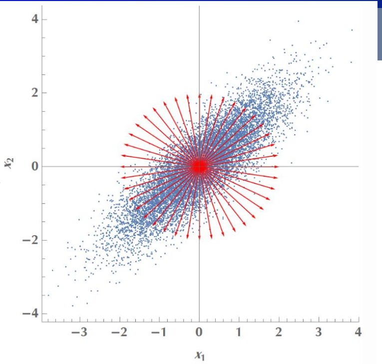
We maximize the variance of the projection in the residual subspace



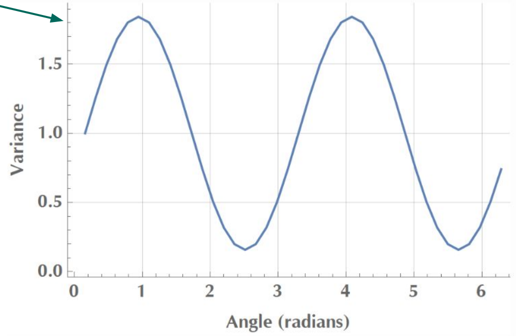
Slide from Barnabas Poczos

PCA in practice (1): an easy example

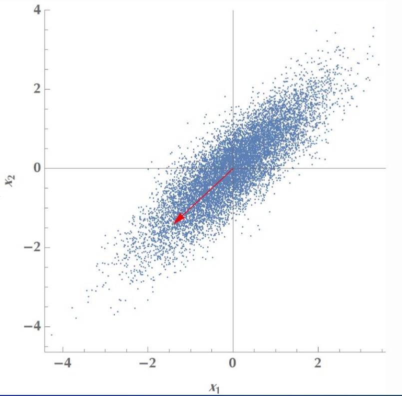
Data distributed in a 2D space



If we plot the variance as a function of the angle, we get



The maximum is reached when the vector is aligned with the elongated axis



PCA in practice (2)

- Given data $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, compute covariance matrix $\boldsymbol{\Sigma}$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T$$

where

$$\overline{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i$$

- **PCA** basis vectors = the eigenvectors of $\boldsymbol{\Sigma}$

We get the eigvectors using an eigendecomposition.
Power iteration (Von Mises iteration is a standard algorithm for this)

- Larger eigenvalue $\Rightarrow$ more important eigenvectors

PCA in practice (2)

Mean subtraction!

- Given data $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, compute covariance matrix $\mathbf{\Sigma}$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T \quad \text{where} \quad \overline{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i$$

- **PCA** basis vectors = the eigenvectors of $\mathbf{\Sigma}$

We get the eigvectors using an eigendecomposition.
Power iteration (Von Mises iteration is a standard algorithm for this)

- Larger eigenvalue $\Rightarrow$ more important eigenvectors

PCA & Lagrange Multipliers

We want to maximise the variance of the projection

subject to the constraint $\|\mathbf{v}\| = 1.$

This is an application for Lagrange Multipliers!

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda(g(\mathbf{x}) - c)$$

quantity to minimize

constraint

$$\left\{ \frac{\partial \mathcal{L}(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = 0, \frac{\partial \mathcal{L}(\mathbf{x}, \lambda)}{\partial \lambda} = 0 \right\}$$

$$\begin{aligned} V &= \frac{1}{n} \sum_{i=1}^{n} y_i^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^\top \mathbf{v} \right)^2 \\ &= \frac{1}{n} \sum_{i=1} \mathbf{x}_i^\top \mathbf{v} \cdot \mathbf{x}_i^\top \mathbf{v} = \frac{1}{n} \sum_{i=1} \mathbf{v}^\top \mathbf{x}_i \cdot \mathbf{x}_i^\top \mathbf{v} \\ &= \mathbf{v}^\top \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top \right)}_{\text{Covariance matrix}} \mathbf{v} = \mathbf{v}^\top C \mathbf{v} \end{aligned}$$

PCA & Lagrange Multipliers

Our Lagrangian is

$$\mathcal{L}(\mathbf{v}, \lambda) = \mathbf{v}^\top C\mathbf{v} - \lambda(\mathbf{v}^\top \mathbf{v} - 1)$$

Our system of equations is

$$\frac{\partial \mathcal{L}}{\partial \mathbf{v}} = 2\mathbf{v}^\top \mathbf{C} - 2\lambda\mathbf{v}^\top = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{v}^\top \mathbf{v} - 1 = 0$$

Which leads us to the eigenvector equation

$$C\mathbf{v} = \lambda\mathbf{v}$$

P.S. we used $\frac{\partial}{\partial \mathbf{v}}(\mathbf{v}^\top \mathbf{C}\mathbf{v}) = \mathbf{v}^\top \left(\mathbf{C} + \mathbf{C}^\top\right) = 2\mathbf{v}^\top \mathbf{C}$

Eigenvalues & Eigenvectors

- For symmetric matrices, eigenvectors for distinct eigenvalues are **orthogonal**

$$Sv_{\{1,2\}} = \lambda_{\{1,2\}} v_{\{1,2\}}, \text{ and } \lambda_1 \neq \lambda_2 \Rightarrow v_1 \bullet v_2 = 0$$

- All eigenvalues of a real symmetric matrix are **real**.

$$\text{if } |S - \lambda I| = 0 \text{ and } \mathrm{S} = \mathrm{S}^{\mathrm{T}} \Rightarrow \lambda \in \Re$$

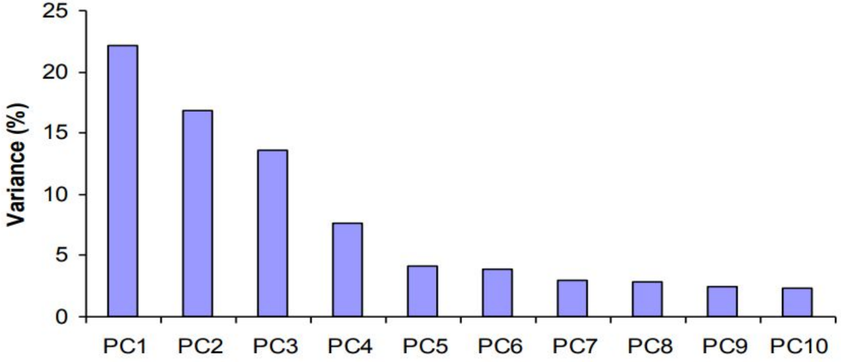
- All eigenvalues of a positive semidefinite matrix are **non-negative**

$$\forall w \in \Re^n, w^T S w \geq 0, \text{ then if } \mathrm{Sv} = \lambda \mathrm{v} \Rightarrow \lambda \geq 0$$

How many components?

- For n original dimensions, sample covariance matrix is nxn, and has up to n eigenvectors. So n PCs.
- Where does dimensionality reduction come from?

Can *ignore* the components of lesser significance.



You do lose some information, but if the eigenvalues are small, you don't lose much

- n dimensions in original data
- calculate n eigenvectors and eigenvalues
- choose only the first p eigenvectors, based on their eigenvalues
- final data set has only p dimensions

PCA interpretation

Maximum Variance Direction: 1st PC a vector v such that projection on to this vector capture maximum variance in the data (out of all possible one dimensional projections)

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}^T\mathbf{x}_i)^2 = \mathbf{v}^T\mathbf{X}\mathbf{X}^T\mathbf{v}$$

Minimum Reconstruction Error: 1st PC a vector v such that projection on to this vector yields minimum MSE reconstruction

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i - (\mathbf{v}^T\mathbf{x}_i)\mathbf{v}\|^2$$

PCA interpretation(s)

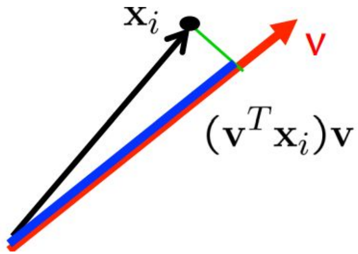
Maximum Variance Direction: 1st PC a vector v such that projection on to this vector capture maximum variance in the data (out of all possible one dimensional projections)

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}^T\mathbf{x}_i)^2 = \mathbf{v}^T\mathbf{X}\mathbf{X}^T\mathbf{v}$$

blue² + green² = black²

black² is fixed (it's just the data)

So, maximizing blue² is equivalent to minimizing green²

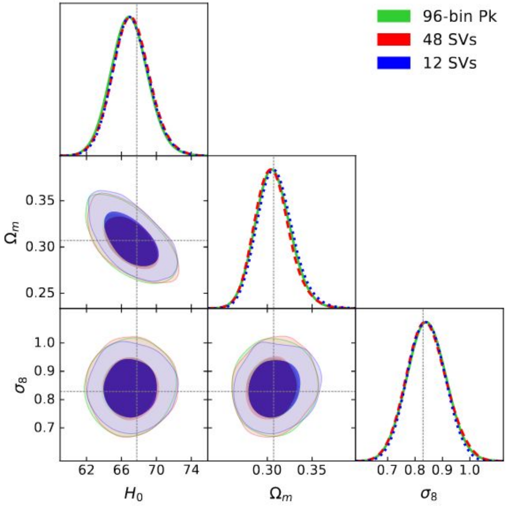


Minimum Reconstruction Error: 1st PC a vector v such that projection on to this vector yields minimum MSE reconstruction

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i - (\mathbf{v}^T\mathbf{x}_i)\mathbf{v}\|^2$$

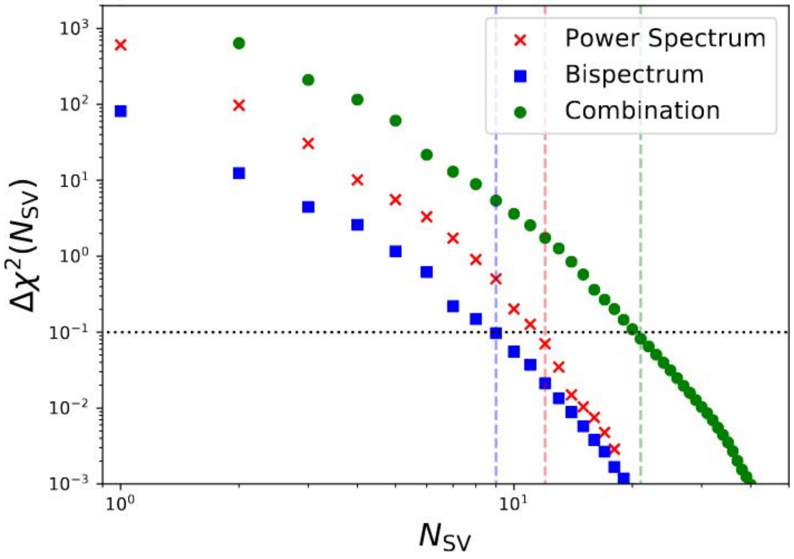
Applications in Cosmology 1: Covariances

In order to compute the covariance matrix from a set of simulations, we need roughly 10 simulations for each element in the data vector: this can be expensive! With PCA/SVD, you can reduce the dimensionality of your problem (and hence the number of required simulations!



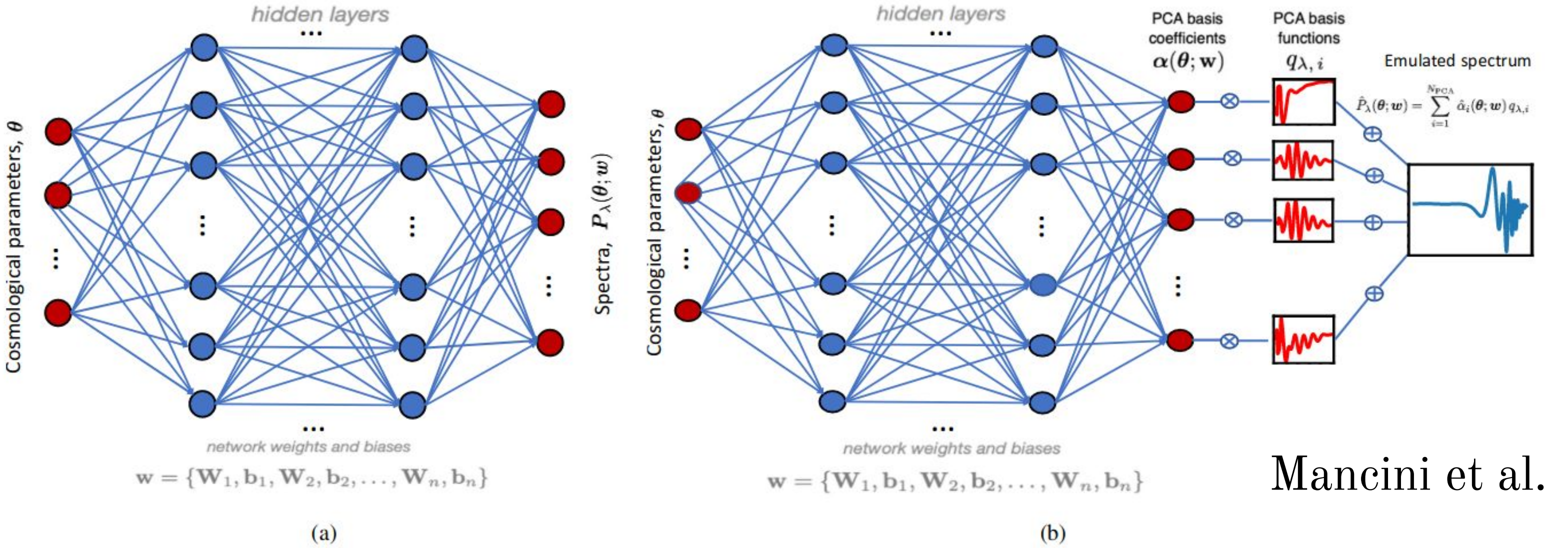
Loss of information

Philcox et al. 2021



Applications in Cosmology 2: Emulators

Emulators are methods to interpolate the prediction of Boltzmann solvers over a cosmological parameters grid. PCA can help to reduce the dimensionality of the datavector you have to emulate



Mancini et al. 2021