Application of High-Dimensional Statistics to Astrophysics and its Perspective

Tsutomu T. TAKEUCHI

1. Division of Particle and Astrophysical Science, Nagoya University, Japan
2. The Research Center for Statistical Machine Learning, the Institute of Statistical Mathematics

New Statistical Science, Kanazawa, 23-24 Sep., 2024

Collaborators

Kazuyoshi YATA (矢田 和善), Makoto AOSHIMA(青嶋 誠)
Institute of Mathematics, University of Tsukuba, Japan

Kento EGASHIRA (江頭 健斗), Aki ISHII (石井 晶)
Department of Information Sciences, Tokyo University of Science, Japan

Hiroma OKUBO (大久保 宏真)
School of Science and Engineering, University of Tsukuba, Japan

Suchetha COORAY (クレ スチェータ)
Kavli Institute Particle Astrophysics and Cosmology, Stanford University, USA

Aina May SO (曹 愛奈), Wen SHI (施 文), Ryusei R. KANO (加納龍生), Hai-Xia MA (馬 海霞), Sena A. MATSUI (松井 瀬奈)
Division of Particle and Astrophysical Science, Nagoya University, Japan

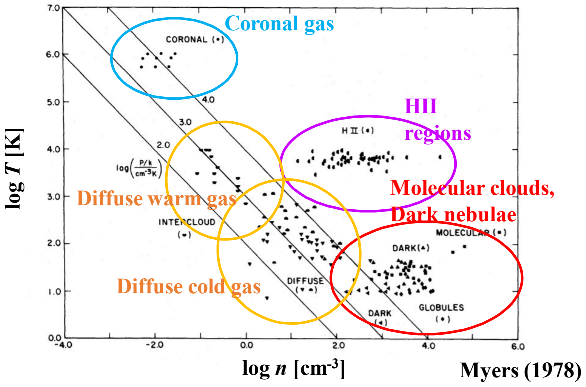
Kohji YOSHIKAWA (吉川 耕司)
Center for Computational Sciences, University of Tsukuba, Japan

Kouichiro NAKANISHI (中西 康一郎)
ALMA Project, National Astronomical Observatory of Japan

Kotaro KOHNO (河野 孝太郎)
Institute of Astronomy, The University of Tokyo, Japan

1. Interstellar Medium (ISM)

1.1 Phase in ISM



Myers (1978)

1.2 ISM phases and star formation

ISM has various phases

1. Plasma (ionized diffuse phase)
2. Neutral gas (mainly neutral hydrogen HI)
3. Molecular gas (mainly molecular hydrogen H_2)

Since gas must become dense enough to form stars, star formation occurs in molecular clouds. Namely,

Atomic gas ⇒ Molecular gas ⇒ Stars

Spatial scales

Spatial scales of galaxies and star formation (SF) are some orders of magnitude different:

Galaxies ~ kpc
Star formation ~ a few pc (for molecular clouds)

Spatial scales

Spatial scales of galaxies and star formation (SF) are some orders of magnitude different:

Galaxies ~ kpc
Star formation ~ a few pc (for molecular clouds)

However, global properties of galaxies and SF activity are mysteriously correlated in various aspects!

Spatial scales

Spatial scales of galaxies and star formation (SF) are some orders of magnitude different:

Galaxies ~ kpc
Star formation ~ a few pc (for molecular clouds)

However, global properties of galaxies and SF activity are mysteriously correlated in various aspects!

⇒ Meso-scale physics to connect the scales of a galaxy and SF should be explored.

2. High-Dimensional Statistical Analysis

2.1 General situation in astrophysics

Classical statistical analysis

Sample size: n
Data dimension: d

The following condition is implicitly assumed

$$n >> d$$

But this is not the case for many cases in scientific researches. Astronomers and astrophysicists have ever simply given up when they face such type of problem.

Star formation in the ISM

Hydrogen is overwhelmingly dominant among others.
⇒ Molecular clouds consist of hydrogen molecules (H_2).

Molecules are not only formed but also dissociated and turn back into atoms by an ultraviolet (UV) radiation.

The layer on which the formation and dissociation of H_2 balance forms the surface boundary of a molecular cloud.

⇒ Since UV is shielded by H_2, the center of a molecular cloud can become cooler and cooler, finally to form a very dense molecular core, where stars form.

2. High-Dimensional Statistical Analysis

2.1 General situation in astrophysics

High-dimensional low-sample size (HDLSS) data analysis

Sample size: n
Data dimension: d

For the HDLSS data, the condition is

$$n << d$$

This condition is often found in e.g., genomic analysis, medical analysis, etc.

In astrophysics, for example, 2-dim spectral map such as integral field spectroscopy has this property.

Kennicutt-Schmidt (K-S) law

Stars form in molecular cores.

⇒ It is natural to suppose a relation between the star formation rate (SFR) and gas density. Schmidt (1959) proposed a relation

$$\mathrm{SFR} \propto \rho^n.$$

i. $n = 1$ Density controls star formation.
ii. $n = 2$ Collision-like process plays a role for star formation

⇒ The power-law index contains substantial information on what triggers the star formation.

It is crucial to reveal spatially resolved SF law in galaxies!

2.2 Unusual behavior of high-dimensional data

For high-dimensional data, classical limit theorems do not work. If we wrongly assume them, we would be lead to a wrong conclusion.

Simplest example: for the sample mean

$$\bar{\vec{x}} = \frac{1}{n}\sum_{i=1}^{n} \vec{x}_i$$

1. as $d/n \to 0$

$$\| \bar{\vec{x}} - \vec{\mu} \| \xrightarrow{P} \vec{0}$$

2. as $d/n \to \infty$

$$\| \bar{\vec{x}} - \vec{\mu} \| \xrightarrow{P} \infty$$

This striking property is referred to as the strong inconsistency.

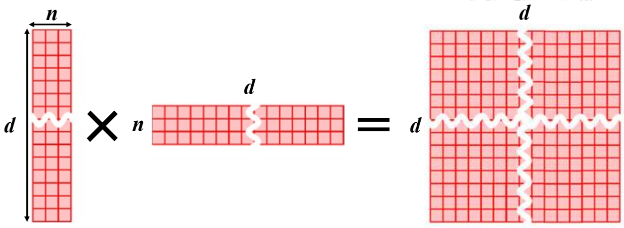
2.2 Geometric Representation

Dual representation of sample covariance matrix

When we draw a set of n samples from the parent population ($d > n$), $\vec{x}_1, \ldots, \vec{x}_n$.

The sample covariance matrix ($d \times d$) is $\tilde{S} = \frac{1}{n}\tilde{X}\tilde{X}^\top$,

$\tilde{X} \equiv (x_1, x_2, \ldots, x_n)$



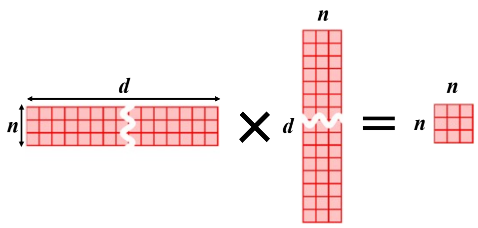
Note that this is a tremendously huge matrix!

2.2 Geometric Representation

Dual representation of sample covariance matrix

When we draw a set of n samples from the parent population ($d > n$), $\vec{x}_1, \ldots, \vec{x}_n$.

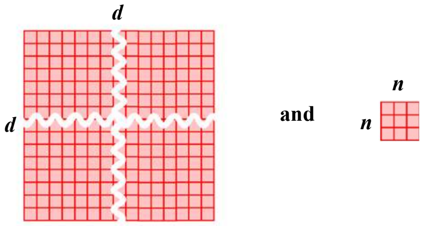
Consider a dual sample covariance matrix ($n \times n$), $\tilde{S}_\mathrm{D} = \frac{1}{n}\tilde{X}^\top\tilde{X}$



This can be handled much more easily!

Eigenvalues of the dual covariance matrix

When we draw a set of n samples from the parent population ($d > n$), $\vec{x}_1, \ldots, \vec{x}_n$.

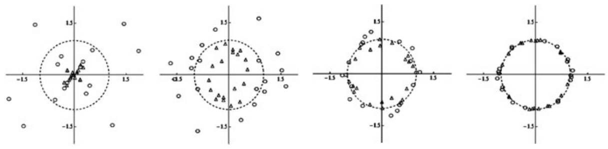


share the first n eigenvalues, i.e., the same important statistical information!

Unusual behavior of high-dimensional data: details

We can visualize the behavior of high-dimensional data vectors with dual representation. We omit all the mathematical details and jump onto the result.

1. The population has a similar property with Gaussian
   ⇒ The data converge on a sphere!!



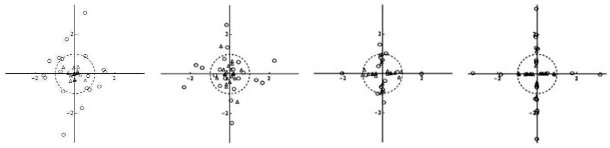
$d = 2$ $d = 20$ $d = 200$ $d = 2000$

Yata & Aoshima (2012)

Unusual behavior of high-dimensional data

We can visualize the behavior of high-dimensional data vectors with dual representation. We omit all the mathematical details and jump onto the result.

2. The population has a similar property with non-Gaussian
   ⇒ The data converge on the axes!!

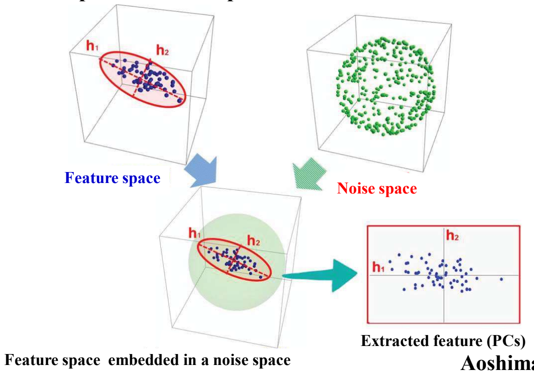


$d = 2$ $d = 20$ $d = 200$ $d = 2000$

Yata & Aoshima (2012)

High-dimensional PCA

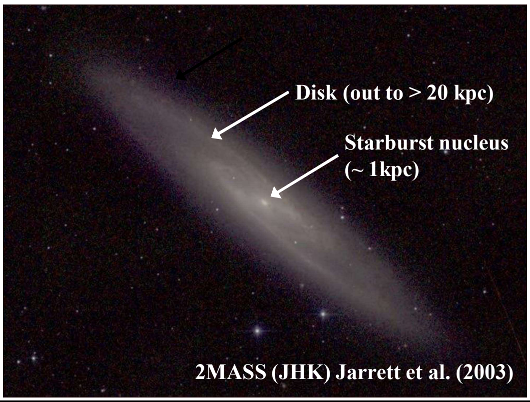
A specially designed PCA, the high-dimensional PCA, can sweep out the noise sphere and extract features of the data.



Aoshima (2012)

2.3 Actual data: ALMA data cube of NGC253

NGC 253: prototypal starburst



2.4 Structure of the Data

Data: Ando et al. (2017)

~ spatial dimension 231 × spectral dimension 2248

⇒ A case with n = 231 and d = 2248 ($n << d$)

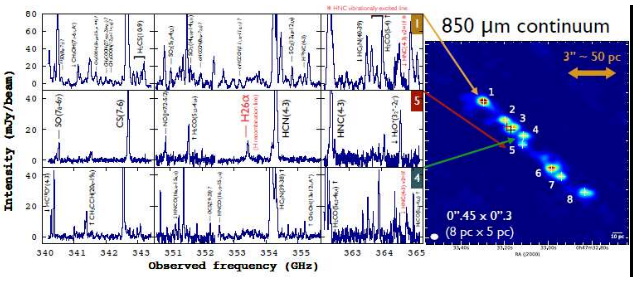
Problems from astrophysical side

- Too much information on spectra.
- Too large variety of spectral lines compared to n.

We apply the high-dimensional statistical analysis to the ALMA spectral mapping data of NGC253.

Rich in molecular lines

ALMA resolved diverse star-forming activities at ~ 10 pc scale.



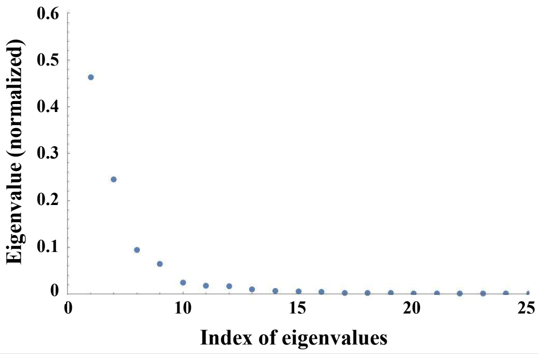
ALMA Band7 spectra

Ando et al. (2017)

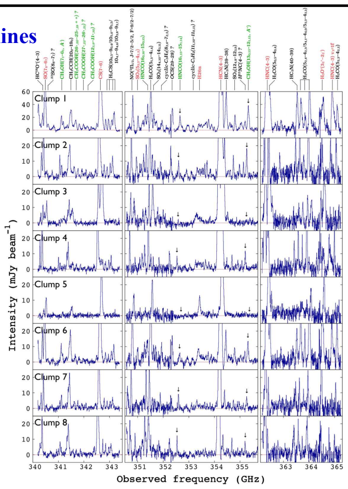
3. Analysis of Starburst Region in NGC253

3.1 Analysis of Raw Data

Eigenvalues of the PCA (contribution)



Rich in molecular lines

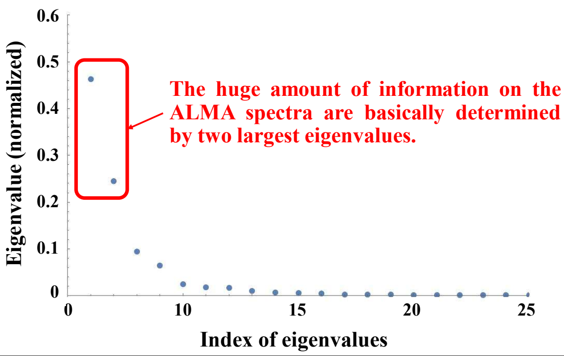


Ando et al. (2017)

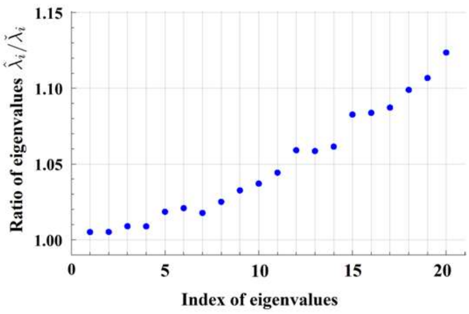
3. Analysis of Starburst Region in NGC253

3.1 Analysis of Raw Data

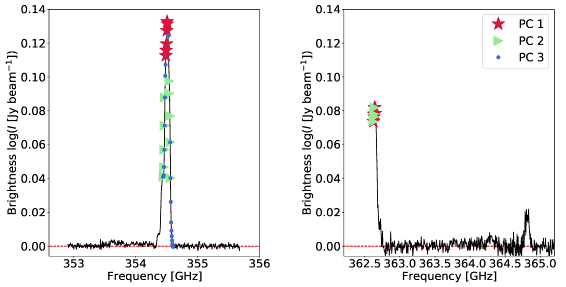
Eigenvalues of the PCA (contribution)



Ratio of eigenvalues obtained by traditional and high-dimensional PCAs (raw data)

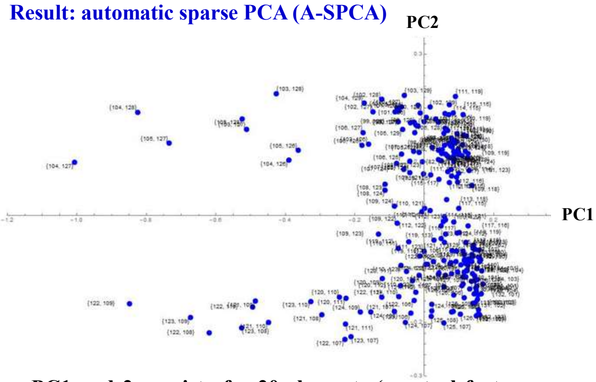


Responsible spectral features for PC1, PC2 and PC3



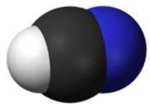
Now PC1 more clearly represents the total intensity, and PC2 and 3 represent smaller-scale velocity structures.
The responsible features are extracted by the A-SPCA (Yata & Aoshima 2024).

Result: automatic sparse PCA (A-SPCA)



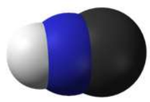
PC1 and 2 consist of ~ 20 elements (spectral features on the resolution units). The key features may be reduced only to a few to several lines!

Spectral features corresponding to PC1 and PC2



H—C≡N

https://en.wikipedia.org/wiki/Hydrogen_cyanide

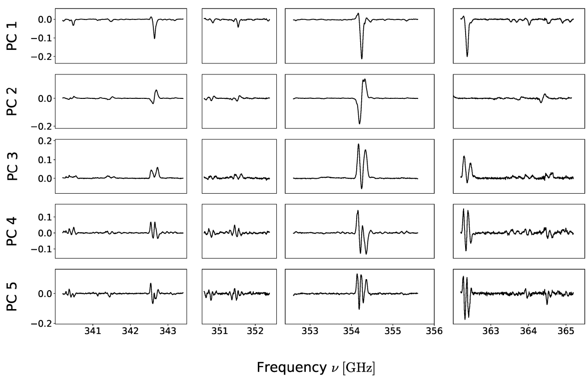


$H—\overset{+}{N}≡C^{-}$

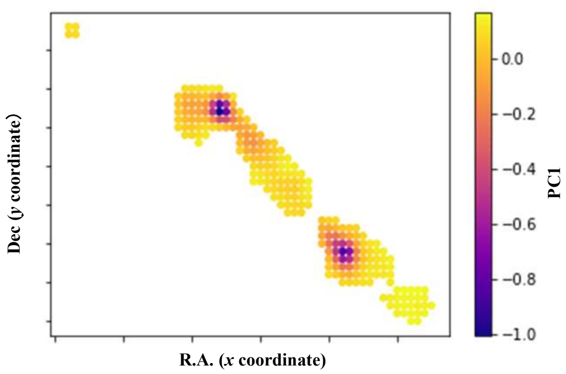
https://en.wikipedia.org/wiki/Hydrogen_isocyanide

HCN (hydrogen cyanide, as known as the hydrocyanic acid) and HNC (hydrogen isocyanide) are linear molecules, which have a quantum mechanical transition corresponding to the rotation states.

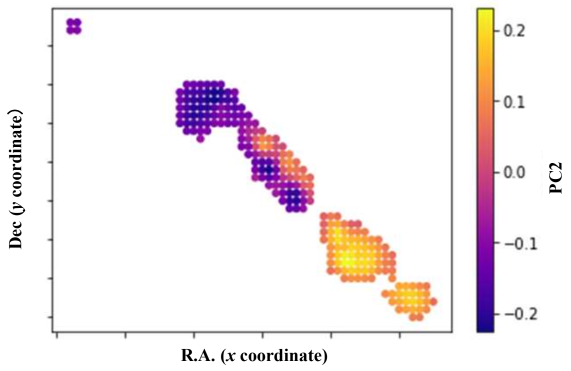
Eigenspectra for PC1-5



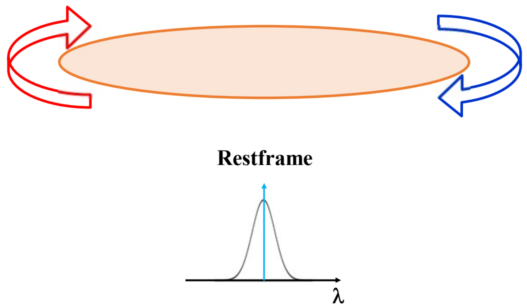
Spatial map of PC1



Spatial map of PC1 and PC2

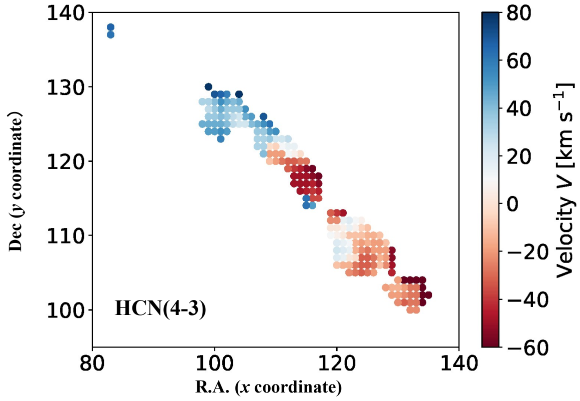


Systemic rotation and Doppler shift



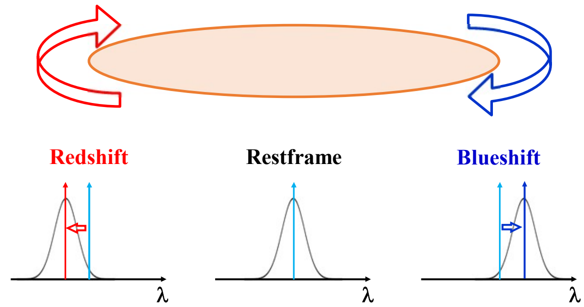
If the system is rotating as a whole, the observed wavelength is affected by the Doppler shift.

Velocity field of the systemic rotation



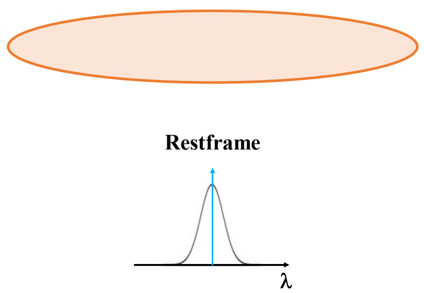
⇒ Doppler shift correction to remove the systemic rotation.

Systemic rotation and Doppler shift



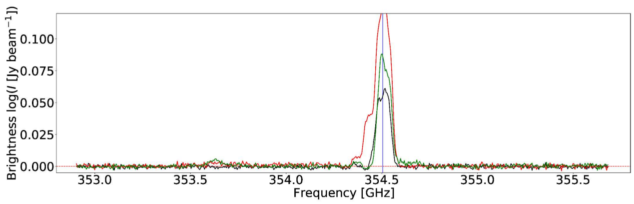
If the system is rotating as a whole, the observed wavelength is affected by the Doppler shift. PC2 beautifully describes the Doppler shift!

Systemic rotation and Doppler shift



3.3 Main analysis

Doppler shift correction

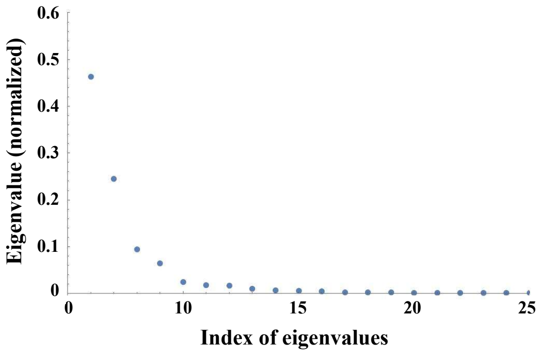


Takeuchi et al. (2021)

We estimated the peculiar velocity field (mainly due to the systemic rotation of the central region of NGC253) by averaging the results from HCN(4-3), HNC(4-3) and CS(7-6) lines, and corrected the Doppler shift.

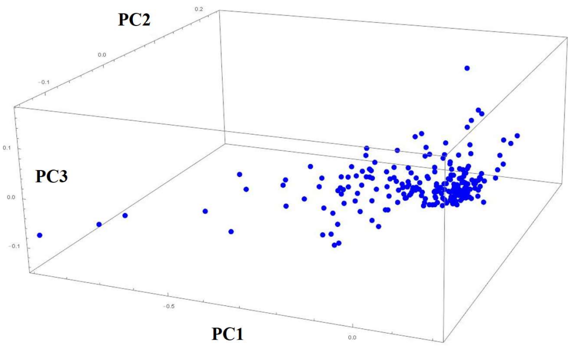
Due to this correction, the final data dimension is $d = 1971$.

Eigenvalues of the NGC253 before Doppler correction



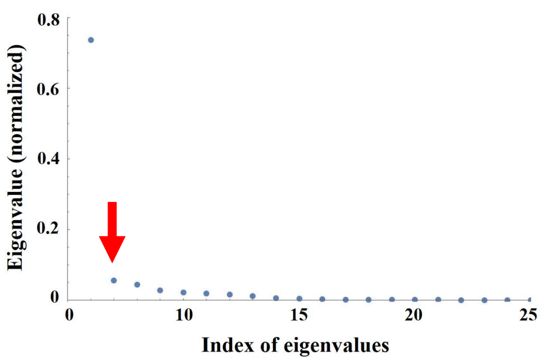
Takeuchi et al. (2023)

PC1, PC2, and PC3 from sparse PCA



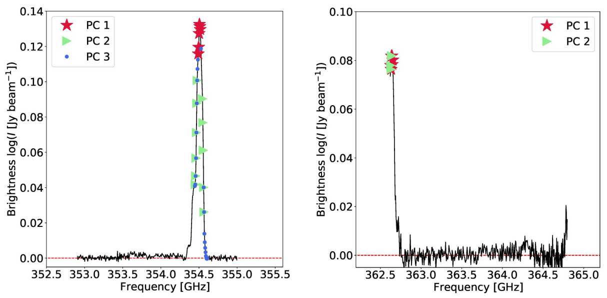
Takeuchi et al. (2023)

Eigenvalues of the NGC253 after Doppler correction



Takeuchi et al. (2023)

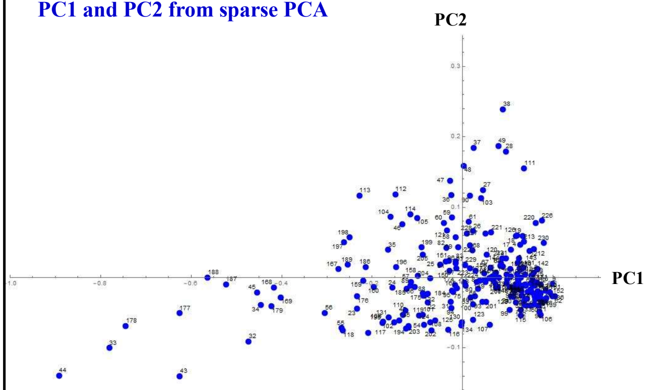
Responsible spectral features for PC1, PC2 and PC3



Takeuchi et al. (2023)

Now PC1 more clearly represents the total intensity, and PC2 and 3 represent smaller-scale velocity structures.

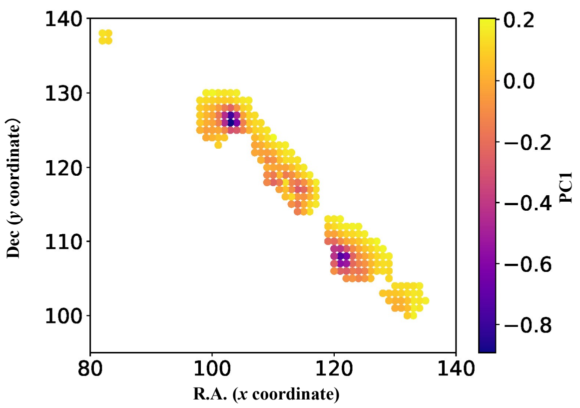
PC1 and PC2 from sparse PCA



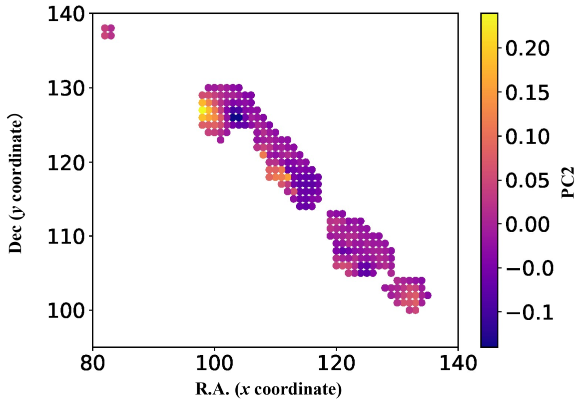
Takeuchi et al. (2023)

Butterfly-like pattern completely disappeared.

Spatial map of PC1 after Doppler correction



Spatial map of PC2 after Doppler correction



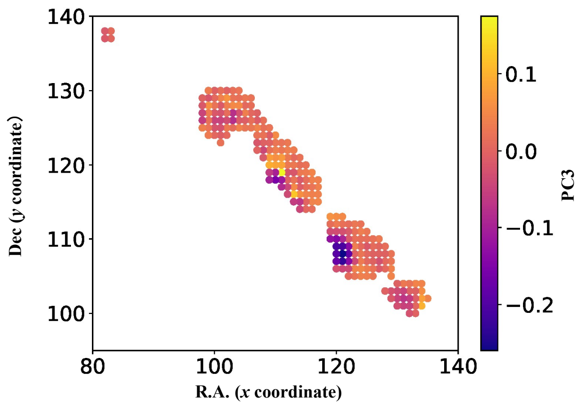
What do we see from the Doppler-corrected map?

NGC253

- Pure starburst: SFR in the central molecular zone is 2 $M_\odot$ yr^{-1} (Rieke et al. 1980; Keto et al. 1999)
- Intense outflow (Matsubayashi et al. 2009; Bolatto et al. 2013)

Indeed the outflow phenomenon is mainly delineated by PC3.

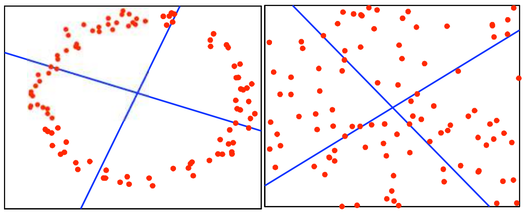
Spatial map of PC3 after Doppler correction



4 Kernel Principal Component Analysis (KPCA)

4.1 Making the PCA nonlinear

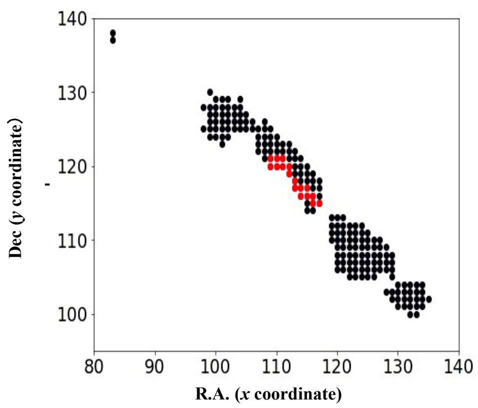
Difficult examples



PCA will make no difference between these examples because the structure on the left is not linear.

Are there ways to find nonlinear, low-dimensional manifolds?

Anomaly regions in the velocity field



Kernel trick: how to make PCA nonlinear

Suppose that instead of using the points $\mathbf{x}_i$ as is, we wanted to go to some different feature space $\phi(\mathbf{x}_i) \in \mathbb{R}^N$.

For example, using polar coordinates, instead of cartesian coordinates, would help us deal with a circle.

In the higher-dimensional space, we can then do PCA.

The result will be nonlinear in the original data space.

4.2 PCA in feature space: kernel PCA

Kernel PCA

For the moment, we suppose that the mean of the data in feature space is 0 (centered). In this case, the covariance matrix is

$$\mathbf{C} = \frac{1}{m}\sum_{i=1}^{m}\phi(\mathbf{x}_i)\phi(\mathbf{x}_i)^T$$

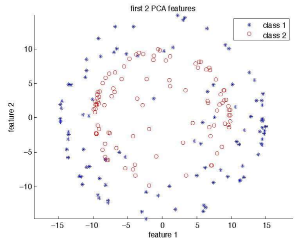
and the eigenvectors are

$$\mathbf{C}\mathbf{v}_j = \lambda_j \mathbf{v}_j, j = 1, \ldots N$$

We want to avoid explicitly going to feature space - instead we want to work with kernels:

$$K(\mathbf{x}_i, \mathbf{x}_k) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_k)$$

Two concentric spheres



Wang (2012)

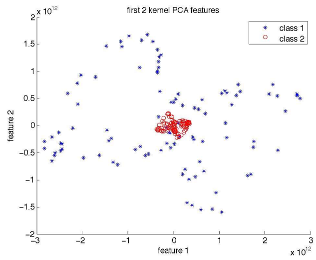
Classical PCA

Classical PCA cannot separate the points from the two spheres.

Summary of kernel PCA

1. Pick a kernel.
2. Construct a normalized kernel matrix $\tilde{\mathrm{K}}$ of the data (this will be of dimension $m \times m$).
3. Find the eigenvalues and eigenvectors of this matrix λ_j, $\mathbf{a}_j$.
4. For any data point (new or old), we can represent it as the following set
$$y_j = \sum_{i=1}^{m} a_{ji}K(\mathbf{x}, \mathbf{x}_i), j = 1, \ldots m$$
5. We can limit the number of components to $k < m$ for a more compact representation (by picking the **a**'s corresponding to the highest eigenvalues)

Two concentric spheres



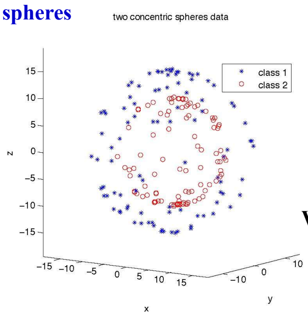
Wang (2012)

Kernel PCA with a polynomial kernel (d = 5)

Points from one sphere are much closer together, the others are scattered. The projected data is not linearly separable.

4.3 Examples

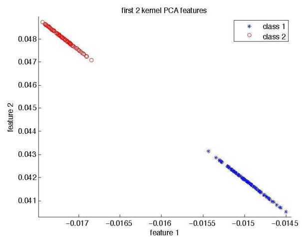
Two concentric spheres



Wang (2012)

Data points are color-coded for visual clarity, but the actual data are unlabeled. We want to project the data distribution from 3D to 2D.

Two concentric spheres



Wang (2012)

Kernel PCA with a Gaussian kernel (σ = 20)

Points from the two spheres are really well separated. We should note that the choice of parameter for the kernel matters!

Validation can be used to determine good kernel parameter values.

4.4 Problem of kernel PCA

Feature extraction

Extraction of responsible spectral features by A-SPCA is not possible for the case of kernel PCA.

Eigenspectra cannot be determined for the kernel PCA.

⇒ We should use both classical and kernel PCA for the physical application.

More careful consideration is needed.

5. Summary

1. **Spectroscopic mapping and similar methods are fundamentally important to reveal the ISM physics, but the data are high-dimensional low sample size.**

2. **We applied the high-dimensional PCA on the NGC253 spectral map. ALMA mapping data are typically HDLSS in general, and in this case n = 231 and d = 2228.**

3. **The controlling feature was HCN(4-3) rotational lines. PC1 describes the total intensity of the lines, and PC2 represents the Doppler shift caused by the systemic rotation.**

Appendix

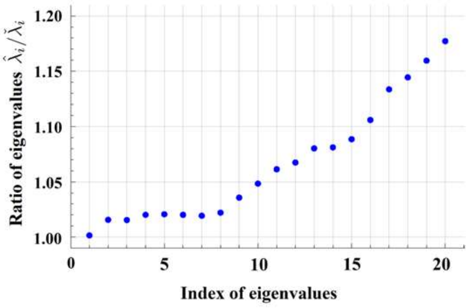
5. Summary

4. **After correcting the Doppler shift due to the systemic rotation, we could obtain information on the smaller-scale velocity field described by PC2 (new) and PC3. These may be caused by outflow phenomena of starburst regions.**

5. **Kernel PCA is a powerful tool to characterize nonlinear relations in the data. However, since we cannot determine the eigenspectral, A-SPCA cannot be applied and then we cannot extract the responsible features. Further consideration is needed.**

If you are interested in details, see Takeuchi et al. 2024, ApJS, 271, 44.

Ratio of eigenvalues obtained by traditional and high-dimensional PCAs (Doppler-corrected)



Kernel PCA

Rewrite the PCA equation as

$$\frac{1}{m}\sum_{i=1}^{m}\phi(\mathbf{x}_i)\phi(\mathbf{x}_i)^T\mathbf{v}_j = \lambda_j\mathbf{v}_j, j = 1,\dots N$$

So the eigenvectors can be written as a linear combination for features

$$\mathbf{v}_j = \sum_{i=1}^{m} a_{ji}\phi(\mathbf{x}_i)$$

Finding the eigenvectors is equivalent to finding the coefficients $a_{ji}, j = 1,\dots N, i = 1,\dots m.$

Kernel PCA

We have a normalization condition for the $\mathbf{a}_j$ vectors as

$$\mathbf{v}_j^T\mathbf{v}_j = 1 \Rightarrow \sum_{k=1}^{m}\sum_{l=1}^{m} a_{jl}a_{jk}\phi(\mathbf{x}_l)^T\phi(\mathbf{x}_k) = 1 \Rightarrow \mathbf{a}_j^T\mathbf{K}\mathbf{a}_j = 1$$

Plugging this into $\mathbf{K}\mathbf{a}_j = m\lambda_j\mathbf{a}_j$, we get

$$\lambda_j m\mathbf{a}_j^T\mathbf{a}_j = 1, \forall j$$

For a new point x, its projection onto the principal components is

$$\phi(\mathbf{x})^T\mathbf{v}_j = \sum_{i=1}^{m} a_{ji}\phi(\mathbf{x})^T\phi(\mathbf{x}_i) = \sum_{i=1}^{m} a_{ji}K(\mathbf{x},\mathbf{x}_i)$$

Kernel PCA

By substituting this back into the equation, we get

$$\frac{1}{m}\sum_{i=1}^{m}\phi(\mathbf{x}_i)\phi(\mathbf{x}_i)^T\left(\sum_{l=1}^{m} a_{jl}\phi(\mathbf{x}_l)\right) = \lambda_j\sum_{l=1}^{m} a_{jl}\phi(\mathbf{x}_l)$$

We can rewrite this as

$$\frac{1}{m}\sum_{i=1}^{m}\phi(\mathbf{x}_i)\left(\sum_{l=1}^{m} a_{jl}K(\mathbf{x}_i,\mathbf{x}_l)\right) = \lambda_j\sum_{l=1}^{m} a_{jl}\phi(\mathbf{x}_l)$$

A small trick: multiply this by $\phi(\mathbf{x}_k)^T$ to the left, we obtain

$$\frac{1}{m}\sum_{i=1}^{m}\phi(\mathbf{x}_k)^T\phi(\mathbf{x}_i)\left(\sum_{l=1}^{m} a_{jl}K(\mathbf{x}_i,\mathbf{x}_l)\right) = \lambda_j\sum_{l=1}^{m} a_{jl}\phi(\mathbf{x}_k)^T\phi(\mathbf{x}_l)$$

Normalizing the feature space

In general, the features may not have a zero mean. Then, we work with

$$\tilde{\phi}(\mathbf{x}_i) = \phi(\mathbf{x}_i) - \frac{1}{m}\sum_{k=1}^{m}\phi(\mathbf{x}_k)$$

The corresponding kernel matrix entries are given by:

$$\tilde{K}(\mathbf{x}_k,\mathbf{x}_l) = \tilde{\phi}(\mathbf{x}_l)^T\tilde{\phi}(\mathbf{x}_j)$$

After some algebra, we get

$$\tilde{\mathbf{K}} = \mathbf{K} - 2\mathbf{1}_{1/m}\mathbf{K} + \mathbf{1}_{1/m}\mathbf{K}\mathbf{1}_{1/m}$$

where $\mathbf{1}_{1/m}$ is the matrix with all elements equal to $1/m$. This operation is referred to as the double centering.

Kernel PCA

We plug in the kernel again

$$\frac{1}{m}\sum_{i=1}^{m}K(\mathbf{x}_k,\mathbf{x}_i)\left(\sum_{l=1}^{m} a_{jl}K(\mathbf{x}_i,\mathbf{x}_l)\right) = \lambda_j\sum_{l=1}^{m} a_{jl}K(\mathbf{x}_k,\mathbf{x}_l), \forall j,k$$

By rearranging, we get

$$\mathbf{K}^2\mathbf{a}_j = m\lambda_j\mathbf{K}\mathbf{a}_j$$

We can remove a factor of K from both sides of the matrix (this will only affect eigenvectors with eigenvalues 0, which will not be principle components)

$$\mathbf{K}\mathbf{a}_j = m\lambda_j\mathbf{a}_j$$

Representation obtained by kernel PCA

Each y_j is the coordinate of $\phi(\mathbf{x})$ along one of the feature space axes $\mathbf{v}_j$.
Recall that $\mathbf{v}_j = \sum_{i=1}^{m} a_{ji}\phi(\mathbf{x}_i)$
Since $\mathbf{v}_j$ are orthogonal, the projection of $\phi(\mathbf{x})$ onto the space spanned by them is

$$\Pi\phi(\mathbf{x}) = \sum_{j=1}^{m} y_j\mathbf{v}_j = \sum_{j=1}^{m} y_j\sum_{i=1}^{m} a_{ji}\phi(\mathbf{x}_i)$$

(again, sums go to k if $k < m$).
The reconstruction error in feature space can be evaluated by

$$\|\phi(\mathbf{x}) - \Pi\phi(\mathbf{x})\|^2$$

This can be rewritten by expanding the norm; we obtain dot products which can all be replaced by kernels.
Note that the error will be 0 on the training data if enough $\mathbf{v}_j$ are retained.