Principal Component Analysis
And its application to Relativistic Heavy-Ion Collisions

For Initial Stages 2019

Ziming Liu
Peking University
June 25, 2019
Overview

• Introduction to PCA

• Paper I:
  PCA Analysis of Collective Flow

• Paper II:
  PCA Analysis of flow factorization breaking
Machine learning

- AlphaGo
- Autonomous Driving
- Machine Translation
- Google Assistant
What is PCA?
An intuitive way for PCA

Eigenvectors $z$: correlations between features
Singular values $\sigma$: importance of eigenvectors
Principal component analysis

Application

Dataset: Faces of different people

Eigenfaces

Mean: $\mu$

Top eigenvectors: $u_1 \ldots u_k$
Principal component analysis

---

Application

- Each face is decomposed into superposition of eigenfaces.

\[ \hat{x} = \mu + w_1 u_1 + w_2 u_2 + w_3 u_3 + w_4 u_4 + \ldots \]

- We can drop unimportant high order eigenfaces. So we can use a few coefficients and corresponding eigenfaces to reconstruct original faces. (Image compression)
Principal component analysis --math

**Theorem: SVD (Singular Value Decomposition)**

For a complex (real) matrix $A \in \mathbb{R}^{n \times m}$, there exist unitary (orthogonal) matrices $U_{n \times n}$ and $V_{m \times m}$, along with a sub-diagonal matrix $\Sigma_{n \times m}$ such that

$$A = U \Sigma V$$

Where $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots)$ such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$

$\sigma$: singular values

$v_i$: eigenmodes/ eigenvectors/principal components
Now, the $i$th event can be decomposed as summation of eigenvectors $z_j$:

$$A_{i,:} \approx \sum_{j=1,2,\ldots,k} \sigma_j u_{i,j} v_{:,j}$$

$k$ is the cut we choose to drop out minor modes.
PCA in physics

- Eigenfrequencies in particle motion

- Multi-resolution PCA to discover El Nino.

H. Y. Chen, Raphal Ligeois, John R. de Bruyn, and Andrea Soddu Phys. Rev. E 91, 042308  Published 15 April 2015

PCA in physics

Machine learning helps discover

- Correlations between spin configurations
- Phase transition

\[ \mathcal{H} = J \sum_{\langle ii \rangle} \cos(\theta_i - \theta_j) \]

C Wang, H Zhai - Physical Review B, 96 (2017), 14, 144432
PCA in Heavy-Ion

- subleading modes of factorization breaking

- Best linear descriptor
  \[ \zeta_{n,\text{pred}}^{(a)} = \varepsilon_{n,n} + c_1 \varepsilon_{n,n+2} \]

- Nonlinear response coefficients
  Rajeev S. Bhalerao, Jean-Yves Ollitrault, Subrata Pal, Derek Teaney

- Experimental data
  CMS collaboration, Phys.Rev. C96 (2017) no.6, 064902
Principal Component Analysis (PCA) in Heavy-ion Physics

arXiv: 1903.09833
Ziming Liu, Wenbin Zhao, Huichao Song
Principal Component Analysis (PCA) of Collective Modes in Relativistic Heavy-Ion Collisions

Recent work, paper in preparation
Ziming Liu, Arabinda Behera, Huichao Song, and Jiangyong Jia
Reconsideration on studying sub-leading flow with PCA
Can a machine automatically discover flow?

arXiv: 1903.09833
Ziming Liu, Wenbin Zhao, Huichao Song
Principal Component Analysis (PCA) of Collective Modes in Relativistic Heavy-Ion Collisions
Previous work utilizes Fourier Transformation in the $\phi$ direction:

$$\frac{dN}{dp} = \sum_{n=-\infty}^{+\infty} V_n(p) e^{in\phi} \quad p = (p_t, \eta)$$

PCA decomposes $V_n(p)$ into eigenmodes:

$$V_n(p) = \sum_{\alpha=1}^{k} \xi^{(\alpha)} V_n^{(\alpha)}(p)$$

However, we apply PCA directly to $dN/d\phi$ data without FT:

$$\frac{dN}{d\phi} = \sum_{\alpha=1}^{k} \xi^{(\alpha)} \left(\frac{dN}{d\phi}\right)^{(\alpha)}$$
Pb+Pb collisions at 2.76 A TeV

Trento initial model → Vishnew Hydrodynamics → iss particle sampling

No hadron rescattering or resonance decays to simplify problem settings.
PCA for flow analysis

Data sets: \( \frac{dN}{d\phi} \)

With PCA, each flow distribution is decomposed into superposition of eigenmodes.

\[
\frac{dN}{d\phi} = \mu + x_1 z_1 + x_2 z_2 + x_3 z_3 + \ldots
\]
Singular values

Degeneracy → Rotational symmetry
Eigenvectors/ Principal components

Elliptic flow  Triangular flow  ...... 

Machines can automatically discover flow without any guidance from human beings!
Define flow harmonics with PCA

\[
\frac{dN}{d\phi} = \mu + \sum_{i=1}^{k} x_k z_k
\]

Event average comparisons

<table>
<thead>
<tr>
<th>n</th>
<th>( v'_n \times 10^2 )</th>
<th>( \bar{v}_n \times 10^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.03</td>
<td>6.08</td>
</tr>
<tr>
<td>3</td>
<td>2.57</td>
<td>2.53</td>
</tr>
<tr>
<td>4</td>
<td>1.21</td>
<td>1.25</td>
</tr>
<tr>
<td>5</td>
<td>0.57</td>
<td>0.66</td>
</tr>
<tr>
<td>6</td>
<td>0.26</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Event-by-event comparisons

Define flow harmonics with Fourier (Tradition)

PCA flow harmonics \( \approx \)

Traditional flow harmonics

\( v'_2 \times 1 \quad v'_3 \times 2 \quad v'_4 \times 4 \quad v'_5 \times 10 \quad v'_6 \times 15 \)
Symmetric cumulants

Fourier: \( SC^\nu(m, n) = \langle v_m^2 v_n^2 \rangle - \langle v_m^2 \rangle \langle v_n^2 \rangle \)

PCA: \( SC'^\nu(m, n) = \langle v'_m^2 v'_n^2 \rangle - \langle v'_m^2 \rangle \langle v'_n^2 \rangle \)

Correlation between different harmonics decrease for PCA!
Pearson correlation between initial and final

PCA
Fourier

Off-diagonal correlation
Diagonal correlation

Centrality%  Centrality%  Centrality%  Centrality%
Pearson correlation between initial and final

20%-30% centrality data

Fourier:

PCA:

PCA has a more diagonal pattern!
Conclusion for paper 1

• Without defining Fourier bases, PCA can automatically discover flow.
• We use PCA bases to re-define flow harmonics and find that
  • PCA bases lead to less correlations between different flow harmonics.
  • PCA flow harmonics have a more diagonal pattern with initial eccentricities as compared to traditional one.
The limitation of studying sub-leading flow with PCA

Recent work, paper in preparation
Ziming Liu, Arabinda Behera, Huichao Song, and Jiangyong Jia
Reconsideration on studying sub-leading flow with PCA
Principal component analysis of event-by-event fluctuations

PRL 114, 152301 (2015)
Rajeev S. Bhalerao, Jean-Yves Ollitrault, Subrata Pal, and Derek Teaney

**Single particle distribution**

\[
\frac{dN}{dp} = \sum V_n(p)e^{-i n \phi}
\]

**Two-particle correlation**

\[
\langle V_{n\Delta}(p_a, p_b) \rangle = \langle V_n(p_a)V_n^*(p_b) \rangle
\]

\[
\downarrow \text{diagonalization}
\]

Decompose \( V_n(p_T) \) with PCA modes

\[
V_n(p_T) = \zeta_n^{(1)} V_n^{(1)}(p_T) + \zeta_n^{(2)} V_n^{(2)}(p_T) + \zeta_n^{(3)} V_n^{(3)}(p_T) + \ldots + \zeta_n^{(k)} V_n^{(k)}(p_T)
\]

\[
\downarrow \text{Leading flow}
\]
\[
\downarrow \text{Sub-leading flow}
\]
\[
\downarrow \text{Sub-sub-leading flow}
\]
\[
\downarrow \text{k-th mode}
\]
Principal component analysis of event-by-event fluctuations

PRL 114, 152301 (2015)

Rajeev S. Bhalerao, Jean-Yves Ollitrault, Subrata Pal, and Derek Teaney

Non-zero sub-leading flow?
Causes for Sub-leading flow

Aleksas Mazeliauskas, Derek Teany

\[ V_3(p_T) \]

Radial excitation leads to sub-leading flow!

Sub-leading flow is theoretically important!!!
Experimental results

Phys. Rev. C 96, 064902
The CMS Collaboration

\[ v_3 \{|\Delta \eta| > 2.0\}, \text{CMS} \]

\[ v_3 \{|\Delta \eta| > 0.8\}, \text{ALICE} \]
Think more about it.

Model setup:
• AMPT model
• 1M UCC events
• subevent method

(paper in preparation)
Ziming Liu, Arabinda Behera, Huichao Song, and Jiangyong Jia
Reproduce CMS results

Our model can properly reproduce the results @CMS.

Same $p_T$ cut and proper $\eta$ gap $|\Delta\eta| < 0.8$. 
Problem 1: How to choose $p_T$ bins?

Orthogonal on this large interval

Not orthogonal on this small interval

PCA will try to mix these two modes!

Drop one $p_T$ bin each time, and redo the PCA

Will the modes still be stable?
Problem 1: How to choose $p_T$ bins?

Different choice of $p_T$ bins can introduce systematic errors!
PCA modes are sensitive to our choice of $p_T$ bins!
Problem 2: Normalization

Traditional PCA
\[ \langle V_{n\Delta}(p_a, p_b) \rangle = \langle V_n(p_a)V_n^*(p_b) \rangle \]

PCA

\[ V_n(p_T) = \zeta_n^{(1)} V_n^{(1)}(p_T) + \zeta_n^{(2)} V_n^{(2)}(p_T) + \zeta_n^{(3)} V_n^{(3)}(p_T) + \cdots + \zeta_n^{(k)} V_n^{(k)}(p_T) \]

Normalize with \( \langle M(p_T) \rangle \)

\[ v_n^{(\alpha)}(p_T) = \frac{V_n^{(\alpha)}(p_T)}{\langle M(p_T) \rangle} \]

New PCA
\[ \langle V_{n\Delta}(p_a, p_b) \rangle = \langle V_n(p_a)V_n^*(p_b) \rangle \]

Normalize with \( \langle M(p_T) \rangle \)

\[ v_{n\Delta}(p_a, p_b) = \frac{\langle V_{n\Delta}(p_a, p_b) \rangle}{\langle M(p_a) \rangle \langle M(p_b) \rangle} \]

\[ \tilde{v}_n(p_T) = \frac{V_n(p_T)}{\langle M(p_T) \rangle} \]

PCA

\[ \tilde{v}_n(p_T) = \zeta_n^{(1)} \tilde{v}_n^{(1)}(p_T) + \zeta_n^{(2)} \tilde{v}_n^{(2)}(p_T) + \zeta_n^{(3)} \tilde{v}_n^{(3)}(p_T) + \cdots + \zeta_n^{(k)} \tilde{v}_n^{(k)}(p_T) \]

\[ \tilde{v}_n^{(\alpha)}(p_T) \]

Different?
Problem 2: Normalization

Different!

\[ v_n^{(\alpha)}(p_T) \] or \[ \tilde{v}_n^{(\alpha)}(p_T) \]?

Leading flow

Sub-sub-leading flow

Elliptic

Sub-leading flow

Triangular

Sub-sub-leading: real or fake?

But which scheme can reveal more physics about initial profiles?
More analysis should be done to fully unravel the mystery of PCA!
Conclusion for paper 2

• The choice of $p_T$ bins introduces **systematic errors**, but we have no guidance from physics about how to choose them.

• Technically, the **normalization** procedure before/after PCA also lead to different results. Which is the real physics? Need more discussion.
Summary & Outlook
Principal Component Analysis:
• Unsupervised learning (dimensionality reduction)
• Good at discovering hidden correlations in data

PCA for flow discovery
• Integrate $p_T$, so stable
• discover flow automatically
• Reduce mode coupling

PCA for sub-leading flow
• $p_T$ differential, sensitive to $p_T$ bins
• Ambiguity in normalization

All in all, PCA is a transparent yet powerful machine learning tool to extract main information/hidden correlations in data!
But we should also be more careful about its results.

Outlook
Can PCA detect modes or structures from the massive data that is not realized or easily defined by human being?
More advanced PCAs …..

- **High-order PCA**

  Suitable for high-order data, not constrained to 2-part correlation matrix

- **Kernel PCA**

  Able to capture non-linearity. Hope can help study non-linearity in hydro.

- **Robust PCA**

  Hope: Eliminate non-flow in the single particle distribution level, without the trouble to construct 2-part data.

Observed = Collective Flow + non-flow
Thank you for your attention!
VI. Back up
Model Details

- 2.76 A TeV Pb+Pb
- Viscous Hydro: VISH2+1
- EOS: s95-PCE
- Initial condition: TRENTo
- Hydrodynamic starting time $\tau_0 = 0.6 \text{fm/c}$
- Decoupling temperature $T_{sw} = \frac{148 \text{MeV}}{c}$
- $0.3 < p_T < 3 \text{GeV}$, Pions only
- 0%-10%,......,50%-60% totally 6 centrality bins, 2000 events for each bin
PCA Implementation

• Python – sklearn
• From sklearn.decomposition import PCA
• Mode cut $k = 12$
Copper-Fryer:

\[
\frac{dN}{dy p_T dp_T d\varphi} = \int_{\Sigma} \frac{g}{(2\pi)^3} p^\mu d^3 \sigma_\mu f(x, p)
\]

\[\begin{align*}
\text{sample=25} \\
\text{sample=100} \\
\text{sample=500}
\end{align*}\]
Bases Mixing (2\textsuperscript{nd} and 4\textsuperscript{th})

- Left: Fourier, $v_4 \sim \varepsilon_2$
- Right: PCA, $v'_4 \sim \varepsilon_2$
Bases Mixing (2\textsuperscript{nd} and 4\textsuperscript{th})

- Left: Fourier, $\nu_4 \sim \varepsilon_4$
- Right: PCA, $\nu'_4 \sim \varepsilon_4$
PCA for Initial Profiles

Smoothing Procedure

\[
\left( \frac{dS}{d\varphi} \right)_{\text{smooth}} = \int K(\varphi', \varphi) \frac{dS}{d\varphi'} d\varphi'
\]

\[
K(\varphi', \varphi) = \frac{1}{\sqrt{2\pi a}} e^{-\frac{(\varphi'-\varphi)^2}{2a^2}}, \ a = 0.251 \text{ rad}
\]

For initial states, PCA=Fourier
Angle shift

Event plane Angle shift (alignment)

\[ \sigma_j \]

- **No shift (Reaction Plane)**
- **Angle Shift (Align)**
- **Random Shift**

![Graph showing angle shifts with data points for different conditions.](image-url)
Angle Shift

• The PCA is implemented in the reaction plane, so that eigenvectors mix $2^{nd}$ and $4^{th}$ flow harmonics.

• If randomly shifting every event (as in experiments), the bases will be exact Fourier bases due to rotational symmetry.
Good about Neural Networks

Universal Approximation Theorem:
A neural network with one layer but infinitely many neurons can approximate any continuous function.

Width v.s. Deep Theorem:
An exponentially number of neurons in one layer = linear number of layers

That is why we need DEEP learning!
How to train a model?
Train/Evaluate a model

- The evaluation is to evaluate the difference between the network’s outputs and learning targets. — loss function

For supervised learning, since there has been a target $y(x)$, loss function can be defined as

- $\ell(\theta) = \frac{1}{2n} \sum_x [y(x) - \hat{y}(x)]^2$
- $\ell(\theta) = -\frac{1}{n} \sum_x [y(x) \ln \hat{y}(x) - (1 - y(x)) \ln(1 - \hat{y}(x))]$

Train a model $\Leftrightarrow$ Minimize the loss function

Stochastic Gradient Descent

$$\theta' = \theta - \epsilon \frac{\partial \ell(\theta)}{\partial \theta}$$

What is Machine Learning?

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions.

—wikipedia

Linear Regression is also Machine Learning!
How PCA bases mix Fourier bases?
Bases Mixing (2\textsuperscript{nd} and 4\textsuperscript{th})

\[ z: \text{PCA eigenmodes} \]

\[
\begin{pmatrix}
  z_1 \\
  z_2 \\
  z_5 \\
  z_6 \\
\end{pmatrix}
 =
\begin{pmatrix}
  a_{11} & a_{12} & a_{13} & a_{14} \\
  a_{21} & a_{22} & a_{23} & a_{24} \\
  a_{31} & a_{32} & a_{33} & a_{34} \\
  a_{41} & a_{42} & a_{43} & a_{44} \\
\end{pmatrix}
\begin{pmatrix}
  \cos(2\phi) \\
  \sin(2\phi) \\
  \cos(4\phi) \\
  \sin(4\phi) \\
\end{pmatrix}
\]

If the eigenmodes of PCA is the same as fourier bases, the mixing matrix \( A \) should be identity. But actually, the matrix is not diagonal. Take data from centrality 30\% – 40\% for example

\[
A =
\begin{pmatrix}
  0.956 & 0.295 & 0.213 & 0.053 \\
  -0.295 & 0.956 & -0.057 & 0.215 \\
  -0.217 & -0.041 & 0.960 & 0.209 \\
  0.035 & -0.215 & -0.219 & 0.951 \\
\end{pmatrix}
\]
Bases Mixing (2\textsuperscript{nd} and 4\textsuperscript{th})

It is interesting to find that the mixing matrix $A$ follows the form below for all centrality classes. The parameters do not hold, but the form does.

$$A = \begin{pmatrix}
\cos(\theta_2) & \sin(\theta_2) & a\cos(\theta_2) & a\sin(\theta_2) \\
-s\sin(\theta_2) & \cos(\theta_2) & -a\sin(\theta_2) & a\cos(\theta_2) \\
-a\cos(\theta_4) & -a\sin(\theta_4) & \cos(\theta_4) & \sin(\theta_4) \\
as\sin(\theta_4) & -a\cos(\theta_4) & -\sin(\theta_4) & \cos(\theta_4)
\end{pmatrix}$$

To make notations easier, we denote

$$U_1 = \begin{pmatrix}
\cos(\theta_2) & \sin(\theta_2) \\
-s\sin(\theta_2) & \cos(\theta_2)
\end{pmatrix}, \quad U_2 = \begin{pmatrix}
\cos(\theta_4) & \sin(\theta_4) \\
-s\sin(\theta_4) & \cos(\theta_4)
\end{pmatrix}$$

Note that $U_1$ and $U_2$ are just rotation matrix in 2-d Cartesian coordinate.
Bases Mixing (2\textsuperscript{nd} and 4\textsuperscript{th})

\[ A = \begin{pmatrix} U_1 & aU_1 \\ -aU_2 & U_2 \end{pmatrix} \]

It is interesting to note that \( A \) can be decomposed into multiplication of simpler matrices.

\[ A = \begin{pmatrix} U_1 & 0 \\ 0 & l_2 \end{pmatrix} \begin{pmatrix} l_1 & 0 \\ 0 & U_2 \end{pmatrix} \begin{pmatrix} l_2 & a \\ -a & l_2 \end{pmatrix} \]

A new observable defined by PCA to characterize v2/v4 correlation

If we see matrix as an operation, then the operation was decomposed into three steps:

- First, PCA mixed 2nd harmonic flow and 4th harmonic flow, by adjusting \( a \).
- Second, PCA mixed within 4\textsuperscript{th} order plane by adjusting \( \theta_4 \).
- Third, PCA mixed within 2\textsuperscript{nd} order plane by adjusting \( \theta_2 \).