Principal Component Analysis of collective flow in Relativistic Heavy-Ion Collisions

Ziming Liu\textsuperscript{a}, Wenbin Zhao\textsuperscript{a,b}, Huichao Song\textsuperscript{a,b,c}

\textsuperscript{a}Department of Physics and State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing 100871, China
\textsuperscript{b}Collaborative Innovation Center of Quantum Matter, Beijing 100871, China
\textsuperscript{c}Center for High Energy Physics, Peking University, Beijing 100871, China

Abstract

In this paper, we implement Principal Component Analysis (PCA) to study the single particle distributions generated from thousands of VISH2+1 hydrodynamic simulations with an aim to explore if a machine could directly discover flow from the huge amount of data without explicit instructions from human-beings. We found that the obtained PCA eigenvectors are similar to but not identical with the traditional Fourier bases. Correspondingly, the PCA defined flow harmonics \( \Psi \) are also similar to the traditional \( \Psi \) for \( n = 2 \) and 3, but largely deviated from the Fourier ones for \( n \geq 4 \). A further study on the symmetric cumulants and the Pearson coefficients indicates that mode-coupling effects are reduced for these flow harmonics defined by PCA.

1. Introduction

Collective flow is one of the most important observables in relativistic heavy-ion collisions, which provides valuable information on the initial state fluctuations, final state correlations and the QGP properties. In the past decades, various flow observables have been extensively measured in experiments and studied in theory \([1–6]\). In general, these flow observables are defined based on the Fourier decomposition. For example, the integrated flow harmonics are defined as:

\[
\frac{dN}{d\varphi} = \frac{1}{2\pi} \sum_{n=2}^{\infty} V_n e^{-in\varphi} = \frac{1}{2\pi} \left( 1 + 2 \sum_{n=1}^{\infty} V_n e^{-in(\varphi-\Psi_n)} \right)
\]  

(1)

where \( V_n = \tilde{V}_n e^{in\varphi} \) is the \( n \)-th order flow vector, \( V_n \) is the \( n \)-th order flow harmonics and \( \Psi_n \) is the corresponding event plane angle. In general, the first coefficient, \( v_1 \), is called the directed flow, the second coefficient, \( v_2 \), is called the elliptic flow and the third coefficient \( v_3 \), is called the triangular flow. For \( n \geq 3 \), \( v_n \) is also referred as the higher order flow harmonics.

In spite of the success of the flow measurements and the hydrodynamic descriptions, one essential question is why the Fourier expansion is a natural way to analyze the flow data. In this paper, we will address these questions with one of the machine learning techniques, called the Principal Component Analysis (PCA). In more details, we will investigate if a machine could directly discover flow from the huge amount of data of the relativistic fluid systems without explicit instructions from human beings.

PCA is one of the unsupervised algorithms of machine learning \([7]\) based on the Singular Value Decomposition (SVD) that diagonalize a random matrix with two orthogonal matrices. Compared with other deep learning algorithms, the advantage of PCA lies in its simple and elegant mathematical formulation, which is understandable and traceable to human beings, and is able to reveal the main structure of data in a quite transparent way.

Due to its strong power in data mining, PCA has been implemented to various research area of physics \([8–13]\). In molecular dynamics, PCA has been utilized to distinguish break junction trajectories of single molecules \([8]\), which is time efficient and can transfer to a wide range of multivariate data sets. In the field of quantum mechanics, the quantum version of PCA was applied to study quantum coherence among different copies of the system \([9]\), which are exponentially faster than any existing algorithm. In condensed matter physics, PCA has been implemented to study the phase transition in Ising model \([11]\), which found that eigenvectors of PCA can aid in the definition of the order parameter, as well as provide reasonable predictions for the critical temperature without any prior knowledge. Besides, PCA is a widely used tool in engineering for model reduction to make computations more efficient \([13]\).

In relativistic heavy-ion collisions, PCA has been implemented to study the event-by-event flow fluctuations, using the 2-particle correlations with the Fourier expansion \([13, 15, 18]\). Compared with the traditional method, PCA explores all the information contained in the 2-particle correlations, which reveals the substructures in flow fluctuations \([13, 15, 16]\). It was found that the leading components of PCA correspond to the traditional flow harmonics and the sub-leading components evaluate the breakdown of the flow factorization at different \( p_t \) or \( \eta \) bins. Besides, PCA has also been used to study the non-linear mode cou-
pling between different flow harmonics\cite{17}, which helps to discover some hidden mode-mixing patterns. Recently, the CMS Collaboration further implemented PCA to analyze 2-particle correlation in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV and p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV \cite{18}, showing the potential of largely implementing such machine learning technique to realistic data in relativistic heavy ion collisions.

These early PCA investigations on flow \cite{13,15,16,18} are all based on the preprocessed data with the Fourier expansion, which still belong to the category of traditional flow analysis. In this paper, we will directly apply PCA to study the single particle distributions from hydrodynamic simulations without any priori Fourier transformation. We aim to explore if PCA could discover flow with its own bases.

This paper is organized as follows. Sec. II introduces relativistic hydrodynamics, principal component analysis (PCA) and the corresponding flow analysis. Sec. III shows and discusses the flow results from PCA and compares them with the ones from traditional Fourier expansion. Sec. IV summarizes and concludes the paper.

2. Model and method

2.1. VISH2+1 hydrodynamics

In this paper, we implement VISH2+1 \cite{19-22} to generate the final particle distributions for the PCA analysis. VISH2+1 \cite{19-22} is a 2+1-dimensional viscous hydrodynamic code to simulate the expansion of the QGP fireballs, which solves the transport equations for the energy-momentum tenor $T^\mu\nu$ and the second order Israel-Stewart equations for the shear stress tensor $\tau^{\mu\nu}$ and bulk pressure $P$ with an equation of state $\Pi(s, \tau)$ as an input. The initial profiles for VISH2+1 are provided by TRENTo, a parameterized initial condition model that generates event-by-event fluctuating entropy profiles with several tunable parameters \cite{23,24}. These parameters, together with the temperature dependent specific shear viscosity and bulk viscosity, hydrodynamic starting time ($t_0 = 0.6$ fm/c) and decoupling/switching temperature ($T_{sw} = 148$ MeV) have been fixed through fitting all charged and identified particle yields, the mean transverse momenta and the integrated flow harmonics in 2.76 A TeV Pb-Pb collisions using the Bayesian statistics \cite{24}, which also nicely described various flow data at the LHC \cite{25}. In practice, the transition from the hydrodynamic fluid to the emitted hadrons on the freeze-out surface is realized by a Monte-Carlo event generator $\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\pi\p
Z) with \( j = 1, 2, ..., m \):

\[
dN/d\phi^j = \sum_{i=1}^{m} x_i^{(j)} \sigma_j z_i = \sum_{i=1}^{m} \tilde{v}_i^{(j)} z_i
\]

\[\approx k \sum_{j=1}^{k} \tilde{v}_i^{(j)} z_j \]  

(4)

where \( i = 1, 2, ..., N \), represents the index of the event, \( m \) is the number of angular bins of the inputting events. \( v_i^{(j)} \) is the corresponding coefficient of \( z_j \) for the \( i \)th event. In the spirit of PCA, one only focuses on the most important components, so there is a cut at the indices \( k \) in the last approximation of Eq. (4). In Sec. III, we will show that \( k = 12 \) is a proper truncation for the integrated flow analysis, and the shape of the bases or eigenvectors \( z_j \) \( (j = 1, ..., k) \) is similar to but not identical with the Fourier transformation bases \( \cos(n\phi) \) and \( \sin(n\phi) \) \( (n = 1, ..., 6) \) used in the traditional method. Correspondingly, \( v_i^{(j)} \) \( (j = 1, ..., k) \) is identified as the real or imaginary part of the flow harmonics for event \( i \), and the singular values \( \sigma_j \) are associated with the corresponding event averaged flow harmonics at different orders. For more details, please also refer to Sec. III.

3. Results

In this section, we implement PCA to analyze the single particle distributions \( dN/d\phi \) from hydrodynamics simulations in Pb+Pb collisions at \( \sqrt{s_{NN}} = 2.76 \) A TeV. Firstly, we focus on the singular values, eigenvectors as well as the associated coefficients of PCA and explore if such unsupervised learning could discover flow with its own bases.

In practice, we run 2000 event-by-event VISH2+1 hydrodynamic simulations with TRENTo initial conditions to generate the \( dN/d\phi \) distributions for 10%-20% Pb+Pb collisions at \( \sqrt{s_{NN}} = 2.76 \) A TeV. With these \( dN/d\phi \) distributions, we construct the final state matrix \( M_f \) and then implement SVD to \( M_f \) as described in Sec. II. Fig. 1(a) shows these obtained first 12 eigenvectors \( z_j \) \( (j = 1, 2, ..., 12) \) and the first 20 singular values \( \sigma_j \) \( (j = 1, 2, ..., 20) \) of PCA, arranged by the descending order of magnitude.

As introduced in Sec. II, these eigenvectors contain the most representative information on correlations among final particles. Fig. 1 shows that the 1st and 2nd eigenvectors from PCA are similar to the Fourier decomposition bases \( \cos(2\phi) \) and \( \sin(2\phi) \), and the 3rd and 4th components are similar to \( \cos(3\phi) \) and \( \cos(3\phi) \), etc. Meanwhile, Fig. 1(b) shows that singular values \( \sigma_j \) \( (j = 1, 2, ..., 12) \) are arranged in pairs. These results indicate that each pair of the singular values may associate with the real and imaginary parts of the event averaged flow vectors at different orders. Therefore, we define the event averaged flow harmonics of PCA with these paired singular values, as outlined in the the second column of Table 1. The values of these PCA flow at different order are compared with the traditional flow harmonics from the Fourier expansion in Table 1 which are close, but not exactly the same values for \( n \leq 6 \).

As explained in Sec. II, one could also read the event-by-event flow harmonics from the results of PCA. In more details, such PCA flow harmonics for event \( i \) is associated with these coefficients \( \tilde{v}_i^{(j)}, j = 1...k \) in Eq. (4).

\( ^2 \)Each eigenvector is automatically normalized with \( ||z_j||^2 = \sum_{n=1}^{m} (z_i(n))^2 = 1 \) (\( m = 50 \)), due to the orthogonality of the eigenvector matrix \( Z \).

\[
\begin{array}{cccc}
 n & \nu'_i (PCA) & \nu'_i \times 10^2 & \nu_i \times 10^2 \\
 2 & 2\sqrt{\pi} \sigma_1^2 + \sigma_2^2 & 6.03 & 6.08 \\
 3 & 2\sqrt{\pi} \sigma_3^2 + \sigma_4^2 & 2.57 & 2.53 \\
 4 & 2\sqrt{\pi} \sigma_5^2 + \sigma_6^2 & 1.21 & 1.25 \\
 5 & 2\sqrt{\pi} \sigma_7^2 + \sigma_8^2 & 0.57 & 0.66 \\
 6 & 2\sqrt{\pi} \sigma_9^2 + \sigma_{10}^2 & 0.26 & 0.37 \\
\end{array}
\]
Therefore, we define the event-by-event flow harmonics $v'_n$ with magnitudes projected onto PCA bases, similar to the event averaged ones defined in Table I. For example, $v'_2 = \sqrt{\sum} \sqrt{v_1^2 + v_2^2}$ and $v'_4 = \sqrt{\sum} \sqrt{v_3^2 + v_4^2}$ ($m = 50$), etc. Fig. 2 compares $v'_n$ from PCA and $v_n$ from the traditional Fourier expansion at different orders. For the event-by-event elliptic flow $v_2$ and $v'_2$ and triangular flow $v_3$ and $v'_3$, the definitions from PCA and that from Fourier expansion are highly agree with each other, which mostly fall on the diagonal lines. For these higher order flow harmonics with $n \geq 4$, these PCA results are largely deviated from the traditional Fourier ones. We also noticed that the first two PCA eigenvector $z_1$ and $z_2$ for $v'_2$ are similar to but not identical with the Fourier bases $\sin(2\varphi)$ and $\cos(2\varphi)$ with $n = 2$, which contain the contributions from $\sin(4\varphi)$ and $\cos(4\varphi)$. Similarly, the PCA eigenvectors $z_3$ and $z_4$ also contain the contributions from other Fourier bases. Such mode mixing in the PCA eigenvectors leads to the large deviations between $v_4$ and $v'_4$, as well as between $v_5$ and $v'_5$, etc.

To evaluate the correlations between different PCA flow harmonics $v'_{m,n}$ and $v'_{m,n}$, we calculate the symmetric cumulants as once defined for traditional flow harmonics [28–30]:

$$SC^{\nu}(m,n) = \langle v_m^2 v_n^2 \rangle - \langle v_m^2 \rangle \langle v_n^2 \rangle.$$  \hspace{1cm} (5)

Correspondingly, the traditional symmetric cumulants $SC^{\nu}(m,n)$ just replace $v'_{m,n}$ and $v'_{m,n}$ with $v_{m,n}$ and $v_{m,n}$ from the Fourier expansion.

FIG. 3 compares the symmetric cumulants $SC^{\nu}(m,n)$ from PCA and $SC^{\nu}(m,n)$ from Fourier expansion, for the event-by-event VISH2+1 simulations in 2.76 A TeV Pb+Pb collisions at various centrality bins. One finds that, except for $SC^{\nu}(2,3)$, almost all PCA symmetric cumulants $SC^{\nu}(m,n)$ reduce significantly compared to the traditional ones. Although $v'_{4}$ from PCA largely deviated from the traditional $v_{4}$ from the Fourier expansion, the obtained $SC^{\nu}(2,4)$ shows a significant suppression, which contradicts to the long believed idea that the nonlinear hydrodynamics evolution strongly couples $v_2$ to $v_4$, leading to an obvious positive correlations between $v_2$ and $v_4$ obtained from Fourier expansion. Similarly, the non-linear mode coupling between $v'_2$ and $v'_3$, $v'_3$ and $v'_5$ and $v'_5$ for these PCA defined flow harmonics also decrease, which results in the reduced symmetric cumulants $SC^{\nu}(2,5)$, $SC^{\nu}(3,5)$ and $SC^{\nu}(3,4)$ correspondingly.

To evaluate the correlations between the initial and final state fluctuations, we use the Pearson coefficients $r(v'_m, \tilde{r}_m)$
and $r(v_n, \epsilon_m)$ to characterize the linearity between the PCA flow harmonics $v'_m$ and the initial eccentricities $\epsilon_m$, as defined as the following:

$$r(v'_m, \epsilon_m) = \frac{(v'_m \epsilon_m) - (v'_m) \langle \epsilon_m \rangle}{\sqrt{(v'_m)^2 - (v'_m)^2}(\epsilon_m - \langle \epsilon_m \rangle)^2}$$  \hspace{1cm} (6)

Here, $\epsilon_m$ is the traditional eccentricities defined by Eq. (A.1). In Appendix A, we will demonstrate that, with a properly chosen smoothing procedure, the event-by-event eccentricities $\epsilon'_m$ from PCA is highly similar to $\epsilon_m$ from the traditional method. We thus use $\epsilon_m$ in the Pearson coefficient definition $r(v'_m, \epsilon_m)$ for PCA. Meanwhile, we can also calculate the Pearson coefficient $r(v'_m, \epsilon_m)$ for the traditional flow with Fourier expansion, which just replaces the flow harmonics $v'_m$ in Eq. (6) by $v_n$. According to the definition, the Pearson coefficient falls in the range [-1, 1], with $r > 0$ implying a positive correlation, and $r < 0$ implying a negative correlation.

Fig. 4 plots the Pearson coefficients $r(v'_m, \epsilon_m)$ from PCA and $r(v_n, \epsilon_m)$ from the Fourier expansion, for VISH2+1 simulated Pb+Pb collisions at various centralities. With these Pearson coefficients, we focus on evaluating if the PCA defined flow harmonics reduce or increase the correlations with the corresponding initial eccentricities. As shown in Fig. 4, the event-by-event flow harmonics $v'_2$ or $v'_3$ from PCA are approximately equal to the Fourier ones $v_2$ or $v_3$. As a result, these Pearson coefficients involved with these two flow harmonics $r(v'_2, \epsilon_m)$ and $r(v'_3, \epsilon_m)$ are almost overlap with the Fourier ones $r(v_2, \epsilon_m)$ and $r(v_3, \epsilon_m)$ as shown by these upper panels in the first two rows. Meanwhile, these diagonal Pearson coefficients $r(v'_2, \epsilon_2)$ or $r(v'_3, \epsilon_3)$ or $r(v'_4, \epsilon_4)$ are much larger than other ones, which confirms the early conclusion that the elliptic flow and triangular flow are mainly influenced by the initial eccentricity $\epsilon_2$ and $\epsilon_3$, with the approximate linear relationship $v_2 \approx \epsilon_2$ ($v'_2 \approx \epsilon_2$) and $v_3 \approx \epsilon_3$ ($v'_3 \approx \epsilon_3$) \cite{31, 32}.

Although $v'_4$ from PCA is largely deviated from the traditional $v_4$ in Fig. 4, such PCA definition largely enhances correlations between $\epsilon_4$, and also largely reduces the correlations between $\epsilon_2$. For example, at 20-30% centrality, the Pearson coefficients $r(v_4, \epsilon_4)$ is only 70% of the $r(v'_4, \epsilon_4)$, while $r(v_4, \epsilon_2)$ is 200% larger than $r(v'_4, \epsilon_2)$. Traditionally, it is generally believed that $v_4$ is largely influenced by $\epsilon_2$, through the non-linear evolution of hydrodynamics. Our
PCA analysis showed that such mode mixing could be deduced through a redefined PCA bases. Meanwhile, such PCA defined bases also significantly reduce the mode mixing for other higher order flow harmonics such as between $v_2$, $v_3$, and $v_4$, etc.

4. Conclusions

In this paper, we implemented Principal Components Analysis (PCA) to study the single particle distributions of thousands of events generated from VISH2+1 hydrodynamic simulations. Compared with the early PCA investigations on flow that imposed the Fourier transformation in the input data [13–15,18], we focused on analyzing the raw data of hydrodynamics and exploring if a machine could directly discover flow from the huge amount of data without explicit instructions from human-beings. We found that the PCA eigenvectors are similar to but not identical with the traditional Fourier basis. Correspondingly, the obtained flow harmonics $v'_n$ from PCA are also similar to the traditional $v_n$ for $n = 2$ and 3, but largely deviate from the Fourier ones for $n \geq 4$. With these PCA flow harmonics, we found that, except for $SC^{(2,3)}$, almost all other symmetric cumulants $SC^{(m,n)}$ from PCA decrease significantly compared with the traditional $SC^{(m,n)}$. Meanwhile, some certain Pearson coefficients $r(v'_i, \epsilon_n)$ that evaluate the linearity between the PCA flow harmonics and the initial eccentricities are obviously enhanced (especially for $n = 4$), together with an corresponding reduction of the off-diagonal elements.

These results indicate that PCA has the ability to discover flow with its own basis, which also reduce the related mode coupling effects, when compared with traditional flow analysis based on the Fourier expansion. We emphasize that these eigenvectors from PCA are modeled to be orthogonal and uncorrelated to each other. As a result, most of the symmetric cumulants $SC^{(m,n)}$ from PCA that evaluate the correlations between different flow harmonics are naturally reduced compared with the traditional ones. Besides, the PCA flow harmonics $v'_n$ presents an enhanced linear relationship to the corresponding eccentricities $\epsilon_n$, especially for $n = 4$. These results seem contradictory to the long believed idea that hydrodynamics evolution tends to smear out inhomogeneity of the evolving systems. In order to connect and compare these PCA singular values from the initial and final states, we implement a smoothing procedure for the initial profiles before the PCA analysis.

Appendix A. PCA for initial profiles with smoothing procedure

In this appendix, we focus on analyzing the initial state fluctuations using the PCA method. Traditionally, the initial state fluctuations are evaluated by the eccentricity coefficients $\epsilon_n$ at different order, which are defined as [31]:

$$e_n e^{i \phi_n} = -\frac{\int r dr d\phi r^n e^{i \alpha_n} s(r, \phi)}{\int r dr d\phi r^n s(r, \phi)},$$

(A.1)$

where $\Phi_\phi$ is the participant plane angle, $s(r, \phi)$ is the initial entropy density and $\phi$ is the azimuthal angle in the transverse plane [31].

For the PCA analysis, we first construct the initial state matrix $M_i$, using the azimuthal angle distribution of the initial entropy $dS/d\phi$ which is defined by

$$\frac{dS}{d\phi} = \int r^2 dr s(r, \phi)$$

(A.2)$

obtained from 2000 event-by-event TRENTo initial conditions. A direct PCA analysis shows that more than 100 eigenvectors are needed to capture the rich structures of the initial state fluctuations. In contrast, 12 PCA eigenvectors are enough to describe the final state ones since the hydrodynamic evolution tends to smear out inhomogeneity of the evolving systems. In order to connect and compare these PCA singular values from the initial and final states, we implement a smoothing procedure for the initial profiles before the PCA analysis.

In more details, we apply a circular convolution with to the initial density profile $dS/d\phi$, which is written as:

$$\frac{dS}{d\phi} \text{smooth} = \int_{-\pi}^{\pi} K(\phi', \phi) \frac{dS}{d\phi'} d\phi'$$

(A.3)$

Here, $K(\phi', \phi)$ is the convolution kernel, which is taken a gaussian form $K(\phi', \phi) = \frac{1}{\sqrt{2\pi} a} e^{-\frac{(\phi'-\phi)^2}{2a^2}}$. Here, we fine tune the radius $a$ to ensure the the same decaying rate for the PCA singular values from the initial profiles and final profiles. The obtained optimized value for $a$ is 0.251 rad.

With such smoothing procedure, we reconstruct the initial state matrix $M_i$ with 2000 event-by-event $(\frac{dS}{d\phi})_{\text{smooth}}$ distributions from TRENTo for each selected centralities. As the case for the flow analysis in Sec. II B and Sec.III, the implementation of SVD and PCA to the initial state matrix, $M_i = \hat{Y} \hat{Z} = \hat{E} \hat{Z}$, gives the singular value $\hat{\sigma}_j$, eigenvectors $\hat{z}_j$ and the corresponding eccentricity coeffi-
We find that the PCA eigenvectors $\tilde{v}_i$ of the initial states are highly similar to traditional Fourier bases $\cos(2\phi)$, $\sin(2\phi)$, $\cos(3\phi)$, $\sin(3\phi)$, etc. Meanwhile, we could associate the singular value $\tilde{\sigma}_j$ to the event averaged initial eccentricities of PCA, $\bar{\epsilon}'_n$, at different orders and connect the coefficients $\tilde{v}_i$ (j = 1, ...k) to the real or imaginary part of the PCA event-by-event initial eccentricities $\tilde{\epsilon}'_n(n = 1, \ldots \hat{k}/2)$ as the case for flow.

Fig. A.5 compares the event-by-event eccentricites $\tilde{\epsilon}'_n$ from PCA and $\epsilon_n$ from the traditional definition (A1) for the TRENTo initial conditions at 10-20% centrality. It shows, with a properly chosen smoothing procedure of the initial conditions, $\tilde{\epsilon}'_n$ and $\epsilon_n$ agree with each other till $n = 6$. Meanwhile, the event averaged eccentricites $\bar{\epsilon}'_n$ from PCA and $\bar{\epsilon}_n$ from (A1) also fit each other very well, which is much better than ones for flow shown in Table[1] and Fig.[2]. Therefore, for the investigation of initial state and final state correlations, we only use the traditional $\epsilon_n$ to define the Pearson coefficient $r(\tilde{v}_n, \epsilon_m)$, and $r(\tilde{v}_n, \epsilon_m)$ for both PCA and traditional flow in Sec.III.

Appendix B. Signal and noise distinguishment from PCA

In the event-by-event VISH2+1 simulations, both initial state fluctuations and statistical fluctuations from the 1% particle sampling during the Cooper-Fryer freeze-out influence the emissions and distributions of final particles. It is generally believed that the hydrodynamic evolution translate the initial state fluctuations into final state correlations, which directly relate to flow signals. Meanwhile, the statistical fluctuations during Cooper-Fryer freeze-out with a finite number of particle emission introduce statistical noise for the flow definition in each event. As a result, flow harmonics from traditional Fourier expansion are generally analyzed with an event average of millions of events. For the event-by-event flow analysis, one implements the standard Bayesian unfolding procedure to suppress effects from the finite multiplicities and non-flow.[33]

In this appendix, we further explore the ability of PCA to distinguish the signal and noise. With such purpose, we implement 25, 100 and 500 % samplings for each VISH2+1 simulation to generate the $dN/d\phi$ distributions of final particles and the related final state matrices $M_j$ with different weighted signal and noise. Then, we implement PCA to analyze these matrices. As shown in Fig.[B.6], the distribution of the PCA singular values is changed with the number of % samplings. For these systems with large statistical fluctuations, for example with 25 % samplings, the singular values $\sigma_j$ at large $j$ tend to have a long and high tail. For these systems with reduced statistical fluctuations with more % samplings, the height of the tail is largely decreased. Meanwhile, we noticed that these eigenvectors with an index $j$ smaller than a certain “magic number”(12 in this case) is signal-like which has a basis similar to the Fourier one, while these eigenvectors with larger $j$ behave so randomly and chaotically, that we as-

![Figure A.5: A comparison between the event-by-event eccentricites $\tilde{\epsilon}'_n$ from PCA and $\epsilon_n$ from the traditional definition (A1), for TRENTo initial conditions at 10-20% centrality. In the corresponding panels, we also write the values of event averaged eccentricites $\bar{\epsilon}'_n$ from PCA and $\bar{\epsilon}_n$ from the traditional definition.

![Figure B.6: The first 20 singular values of PCA $\sigma_j$ (j = 1, 2, ..., 20) for the final state matrices $M_j$ with different weighted signal and noise. Such matrix $M_j$ are constructed with the $dN/d\phi$ distributions from 2000 event-by-event VISH2+1 simulations with 25, 100 and 500 % samplings.](image-url)
associate these eigenvectors with the noise patterns of the systems. Besides, we check the height of these PCA tails and found the ratios among these heights for different iss samplings approximately satisfy \( \sqrt{\frac{1}{25}} : \sqrt{\frac{1}{100}} : \sqrt{\frac{1}{500}} \), such relation is known as the Law of Large Numbers for statistical noise. With more number of samplings, the height of the tail would further decrease. In the main part of this paper, we thus set the 1σ samplings for each VISH2+1 simulation to 1000, which largely suppresses the noise effects from the statistical fluctuations and makes PCA analysis focus on studying flow signal itself.

References

[1] D. A. Teaney, in: R. C. Hwa, X.-N. Wang (Eds.), Quark-gluon plasma 4, 2010, pp. 207–266.
[2] P. Romatschke, Int. J. Mod. Phys. E19 (2010) 1–53.
[3] P. Huovinen, Int. J. Mod. Phys. E22 (2013) 1330029.
[4] U. Heinz, R. Snellings, Ann. Rev. Nucl. Part. Sci. 63 (2013) 123–151.
[5] C. Gale, S. Jeon, B. Schenke, Int. J. Mod. Phys. A28 (2013) 1340011.
[6] H. Song, Y. Zhou, K. Gajdosova, Nucl. Sci. Tech. 28 (7) (2017) 99.
[7] J. Shlens, CoRR abs/1404.1100.
[8] J. M. Hamill, X. T. Zhao, G. Mészáros, M. R. Bryce, M. Arenz, Phys. Rev. Lett. 120 (2018) 016601.
[9] S. Lloyd, M. Mohseni, P. Rebentrost, Nature Physics 10 (9) (2014) 631.
[10] S. M. Dickerson, J. M. Hogan, A. Sugarbaker, D. M. S. Johnson, M. A. Kasevich, Phys. Rev. Lett. 111 (2013) 083001.
[11] C. Wang, H. Zhai, Phys. Rev. B 96 (2017) 144332.
[12] N. C. Costa, W. Hu, Z. J. Bai, R. T. Scalettar, R. R. P. Singh, Phys. Rev. B 96 (2017) 195138.
[13] R. S. Bhalerao, J.-Y. Ollitrault, S. Pal, D. Teaney, Phys. Rev. Lett. 114 (15) (2015) 152301.
[14] S. Z. Rizvi, J. Mohammadpour, R. Th, N. Meskin, IEEE Transactions on Control Systems Technology 24 (5) (2016) 1883–1891.
[15] A. Mazeliauskas, D. Teaney, Phys. Rev. C91 (4) (2015) 044902.
[16] A. Mazeliauskas, D. Teaney, Phys. Rev. C93 (2) (2016) 024913.
[17] P. Bozek, Phys. Rev. C97 (3) (2018) 034905.
[18] A. M. Sirunyan, et al., Phys. Rev. C96 (6) (2017) 064902.
[19] H. Song, U. W. Heinz, Phys. Lett. B658 (2008) 279–283.
[20] H. Song, U. W. Heinz, Phys. Rev. C77 (2008) 064901.
[21] H. Song, Ph.D. thesis, Ohio State U. (2009).
[22] C. Shen, Z. Qu, H. Song, J. Bernhard, S. Bass, U. Heinz, Comput. Phys. Commun. 199 (2016) 61–85.
[23] A. Bazavov, et al., Phys. Rev. D90 (2014) 094503.
[24] J. E. Bernhard, J. S. Moreland, S. A. Bass, J. Liu, U. Heinz, Phys. Rev. C94 (2) (2016) 024907.
[25] J. S. Moreland, J. E. Bernhard, S. A. Bass, Phys. Rev. C92 (1) (2015) 011901.
[26] W. Zhao, H.-j. Xu, H. Song, Eur. Phys. J. C77 (9) (2017) 645.
[27] H. Song, S. A. Bass, U. Heinz, Phys. Rev. C83 (2011) 024912.
[28] J. Adam, et al., Phys. Rev. Lett. 117 (2016) 182301.
[29] R. S. Bhalerao, J.-Y. Ollitrault, S. Pal, Phys. Lett. B742 (2015) 94–98.
[30] X. Zhu, Y. Zhou, H. Xu, H. Song, Phys. Rev. C95 (4) (2017) 044902.
[31] Z. Qu, U. W. Heinz, Phys. Rev. C84 (2011) 024911.
[32] D. Teaney, L. Yan, Phys. Rev. C86 (2012) 044908.
[33] G. Aad, et al., JHEP 11 (2013) 183.