THE EUROPEAN PHYSICAL JOURNAL C

Principal component analysis of collective flow in relativistic heavy-ion collisions

Ziming Liu[1], **Wenbin Zhao**[1], **Huichao Song**[1,2,3,a]

[1] Department of Physics and State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing 100871, China
[2] Collaborative Innovation Center of Quantum Matter, Beijing 100871, China
[3] 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 v_n' are also similar to the traditional v_n 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:

$$
\begin{aligned}
\frac{\mathrm{d}N}{\mathrm{d}\varphi} &= \frac{1}{2\pi} \sum_{-\infty}^{\infty} \mathbf{V}_n e^{-in\varphi} \\
&= \frac{1}{2\pi} \left(1 + 2 \sum_{n=1}^{\infty} v_n \cos(n(\varphi - \Psi_n)) \right)
\end{aligned}
\tag{1}
$$

where $\mathbf{V}_n = v_n e^{in\Psi_n}$ is the nth order flow-vector, v_n is the n-th order flow harmonics and Ψ_n is the corresponding event

[a] e-mail: Huichaosong@pku.edu.cn

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 [14].

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 η bins. Besides, PCA has also been used to study the non-linear mode coupling between different flow harmonics [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 [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 [13,15–18] 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. Section 2 introduces relativistic hydrodynamics, principal component analysis (PCA) and the corresponding flow analysis. Section 3 shows and discusses the flow results from PCA and compares them with the ones from traditional Fourier expansion. Section 4 summarizes and concludes the paper.

2 Model and method

2.1 VISH2+1 hydrodynamics

In this paper, we implement VISH2+1 [19–22] to generate the final particle distributions for the PCA analysis. VISH2+1 [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 $\pi^{\mu\nu}$ and bulk pressure Π with an equation of state s95-PCE [23,24] 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 [24,25]. These parameters, together with the temperature dependent specific shear viscosity and bulk viscosity, hydrodynamic starting time ($\tau_0 = 0.6$ fm/c) and decou-

pling /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 [24], which also nicely described various flow data at the LHC [26]. 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 iss based on the Cooper–Fryer formula [27]:

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

where $f(x, p)$ is the distribution function of particles, g is the degeneracy factor, and $d^3\sigma_\mu$ is the volume element on the freeze-out hypersurface.

For the following PCA analysis, as well as for the traditional flow analysis in comparison, we run the event-by-event VISH2+1 simulations with 12000 fluctuating initial conditions generated from TRENTO for 2.76 A TeV Pb–Pb collisions at 0–10%, 10–20%, 20–30%, 30–40%, 40–50% and 50–60% centrality bins. The default iss sampling for each VISH2+1 simulation is 1000 events, which corresponds to the main results presented in Sect. 3. In the appendix of this paper, we also investigate the ability of PCA to distinguish signal and noise. We thus implement 25, 100 and 500 iss samplings for each VISH2+1 simulation for such investigation. Note that the default 1000 iss sampling used in this paper has already dramatically suppressed the statistical fluctuations from noises for the final hadron distributions.

With the final particle distributions obtained from hydrodynamic simulations, various flow observables can be calculated based on the traditional flow harmonics defined by the Fourier decomposition in Eq. (1). In Sect. 3, the traditional flow results will be served as the comparison to the PCA results.

2.2 Principal component analysis (PCA)

Principal component analysis (PCA) is a statistical method to analyze complicated data, which aims to transform a set of correlated variables into various independent variables via orthogonal transformations. These obtained main eigenvectors, associated with large or unnegligible singular values, are also called the principal components, which reveal the most representative characteristics of the data. In practice, PCA implements the singular value decomposition (SVD) to a real matrix, which obtains a diagonal matrix with the diagonal elements arranged in a descending order. Therefore, one needs to first construct a related matrix before the following PCA and SVD analysis. Since this paper focuses on investigating the integrated flow with PCA, such final state matrix $\mathbf{M_f}$ is constructed from the angular distribution

of all charged hadrons $dN/d\varphi$ ($|y| < 1.0$) (obtained from Eq. (2)) of $N = 2000$ independent events in each centrality bin, using VISH2+1 simulations with TRENTo initial conditions. In more details, we divide the azimuthal angle $[-\pi, \pi]$ into $m = 50$ bins and count the number of particles in each bin. For the jth bin in event (i), the number of particles is denoted as $dN/d\phi_j^{(i)}$, which is also the i_{th} row and j_{th} column of the matrix $\mathbf{M_f}$.[1]

Then, we apply SVD to the final state matrix $\mathbf{M_f}$ with the size $N \times m$ (Here, $N = 2000$ and $m = 50$), which gives

$$\mathbf{M_f} = \mathbf{X\Sigma Z} = \mathbf{VZ} \tag{3}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are two orthogonal matrices with the size of $N \times N$ and $m \times m$, respectively. $\mathbf{\Sigma}$ is a diagonal matrix with diagonal elements (singular values) arranged in the descending order $\sigma_1 > \sigma_2 > \sigma_3 \cdots > 0$.

With such matrix multiplication, the i_{th} row of matrix $\mathbf{M_f}$, denoted as $dN/d\varphi^{(i)}$, can be expressed by the linear combination of the eigenvectors z_j (the j_{th} row of matrix $\mathbf{Z}$) with $j = 1, 2, \ldots, m$:

$$
\begin{aligned}
dN/d\varphi^{(i)} &= \sum_{j=1}^{m} x_j^{(i)} \sigma_j z_j = \sum_{j=1}^{m} \tilde{v}_j^{(i)} z_j \\
&\approx \sum_{j=1}^{k} \tilde{v}_j^{(i)} z_j \quad (i = 1, \ldots, N)
\end{aligned}
\tag{4}
$$

where (i) $= 1, 2, \ldots, N$, represents the index of the event, z_j are a set of orthogonal vectors such that $z_i^T z_j = \delta_{ij}$, m is the number of angular bins of the inputting events. $\tilde{v}_j^{(i)}$ is the corresponding coefficient of z_j for the i_{th} event.[2] 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 Sect. 3, 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 ($j = 1, \ldots, k$) is similar to but not identical with the Fourier transformation bases $\cos(n\varphi)$ and $\sin(n\varphi)$ ($n = 1, \ldots, 6$) used in the traditional method. Correspondingly, $\tilde{v}_j^{(i)}$ ($j = 1, \ldots, k$) is identified as the real or imaginary part of the flow harmonics for event (i), and the singular values σ_j are associated with the corresponding event averaged flow harmonics at different orders. For more details, please also refer to Sect. 3.

---

[1] In practice, we normalize the event vector in $\mathbf{M_f}$ to get rid of the multiplicity fluctuations.

[2] Note that this paper focuses on investigating whether machine could discover flow from the single particle distributions of hydrodynamics. For more realistic implementation to experimental data, one should perform the PCA analysis for the two-particle correlations with self-correlation and non-flow effects eliminated, which we would like to leave it to future study.

3 Results

In this section, we implement PCA to analyze the single particle distributions $dN/d\varphi$ 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 $dN/d\varphi$ distributions for 10–20% Pb + Pb collisions at $\sqrt{s_{NN}} = 2.76$ A TeV. With these $dN/d\varphi$ distributions, we construct the final state matrix $\mathbf{M_f}$ and then implement SVD to $\mathbf{M_f}$ as described in Sect. 2. Figure 1 shows these obtained first 12 eigenvectors z_j ($j = 1, 2, \ldots, 12$) and the first 20 singular values σ_j ($j = 1, 2, \ldots, 20$) of PCA, arranged by the descending order of magnitude.[3] As introduced in Sect. 2, these eigenvectors contain the most representative information on correlations among final particles. Figure 1 shows that the 1st and 2nd eigenvectors from PCA are similar to the Fourier decomposition bases $\sin(2\varphi)$ and $\cos(2\varphi)$, and the 3rd and 4th components are similar to $\sin(3\varphi)$ and $\cos(3\varphi)$, etc. Meanwhile, Fig. 1b shows that singular values σ_j ($j = 1, 2, \ldots, 12$) 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 Sect. 2, 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 $\tilde{v}_j^{(i)}$, $j = 1 \ldots k$ in Eq. (4). 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 1. For example, $v_2' = \sqrt{\frac{m}{2}}\sqrt{\tilde{v}_1^2 + \tilde{v}_2^2}$ and $v_3' = \sqrt{\frac{m}{2}}\sqrt{\tilde{v}_3^2 + \tilde{v}_4^2}$ ($m = 50$), 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

---

[3] Each eigenvector is automatically normalized with $||z_j||_2^2 = \sum_{i=1}^{m} (z_j)_i^2 = 1$ ($m = 50$), due to the orthogonality of the eigenvector matrix $\mathbf{Z}$.
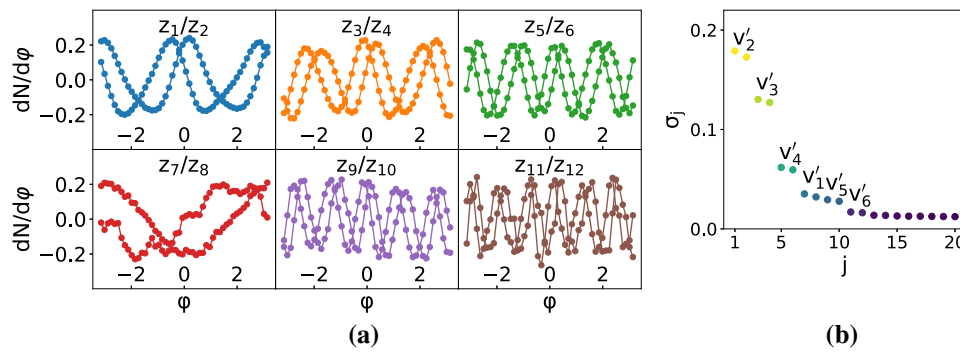
Fig. 1 **a** The first 12 eigenvectors z_j ($j = 1, 2, \ldots, 12$) and **b** the first 20 singular values σ_j ($j = 1, 2, \ldots, 20$), after applying PCA to the final state matrix $\mathbf{M_f}$. The matrix $\mathbf{M_f}$ is constructed from 2000 $dN/d\varphi$ distributions, generated from the event-by-event VISH2+1 simulations with TRENTo initial conditions for 10–20% Pb + Pb collisions at $\sqrt{s_{NN}} = 2.76$ A TeV

Table 1 Event averaged flow harmonics v'_n from PCA and v_n from the Fourier expansion, for VISH2+1 simulated Pb+Pb collisions at 10–20% centrality

n	$\overline{v'_n}$ (PCA)	$\overline{v'_n} \times 10^2$	$\overline{v_n} \times 10^2$
2	$\sqrt{\frac{m}{2}}\sqrt{\sigma_1^2 + \sigma_2^2}$	6.03	6.08
3	$\sqrt{\frac{m}{2}}\sqrt{\sigma_3^2 + \sigma_4^2}$	2.57	2.53
4	$\sqrt{\frac{m}{2}}\sqrt{\sigma_5^2 + \sigma_6^2}$	1.21	1.25
5	$\sqrt{\frac{m}{2}}\sqrt{\sigma_9^2 + \sigma_{10}^2}$	0.57	0.66
6	$\sqrt{\frac{m}{2}}\sqrt{\sigma_{11}^2 + \sigma_{12}^2}$	0.26	0.37

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 and v'_n, we calculate the symmetric cumulants as once defined for traditional flow harmonics [28–30]:

$$SC^{v'}(m, n) = \left\langle v'^2_m v'^2_n \right\rangle - \left\langle v'^2_m \right\rangle \left\langle v'^2_n \right\rangle. \tag{5}$$

Correspondingly, the traditional symmetric cumulants SC^v (m, n) just replace v'_m and v'_n with v_m and v_n from the Fourier expansion.

Figure 3 compares the symmetric cumulants $SC^{v'}(m, n)$ from PCA and $SC^{v'}(m, n)$ 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 $SC^v(2, 3)$, almost all PCA symmetric cumulants $SC^{v'}(m, n)$ 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 $SC^{v'}(2, 4)$ shows a significant suppression, which contradicts to the long believed idea that the nonlinear hydrodynamics evolution strongly couples v_2^2 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'_5, v'_3 and v'_5 and v'_3 and v'_4 for these PCA defined flow harmonics also decrease, which results in the reduced symmetric cumulants $SC^{v'}(2, 5)$, $SC^{v'}(3, 5)$ and $SC^{v'}(3, 4)$ correspondingly.

To evaluate the correlations between the initial and final state fluctuations, we use the Pearson coefficients $r(v'_n, \varepsilon_m)$ and $r(v_n, \varepsilon_m)$ to characterize the linearity between the PCA flow harmonics v'_n and the initial eccentricities ε_m, as defined as the following:
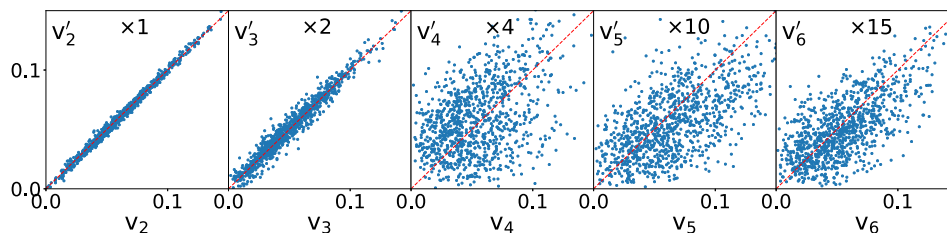
Fig. 2 A comparison between the event-by-event flow harmonics v'_n from PCA and v_n from the Fourier expansion, for VISH2+1 simulated Pb+Pb collisions at 10–20% centrality
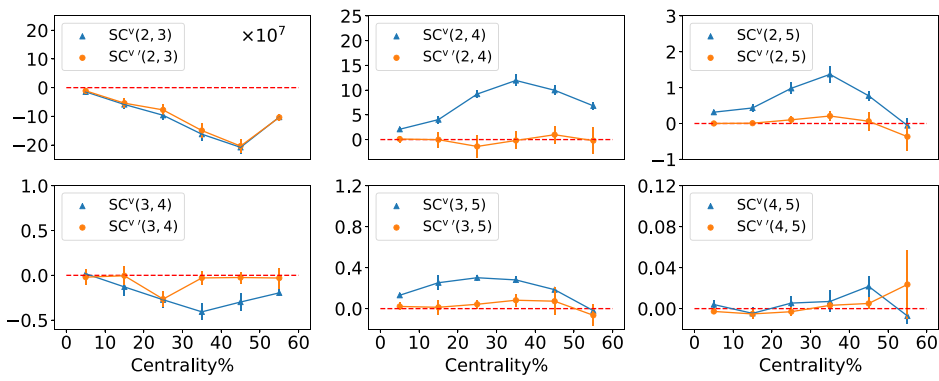
Fig. 3 Symmetric Cumulants $SC^{v\prime}(m, n)$ from PCA and $SC^{v\prime}(m, n)$ from the Fourier expansion, for VISH2+1 simulated Pb+Pb collisions at various centralities
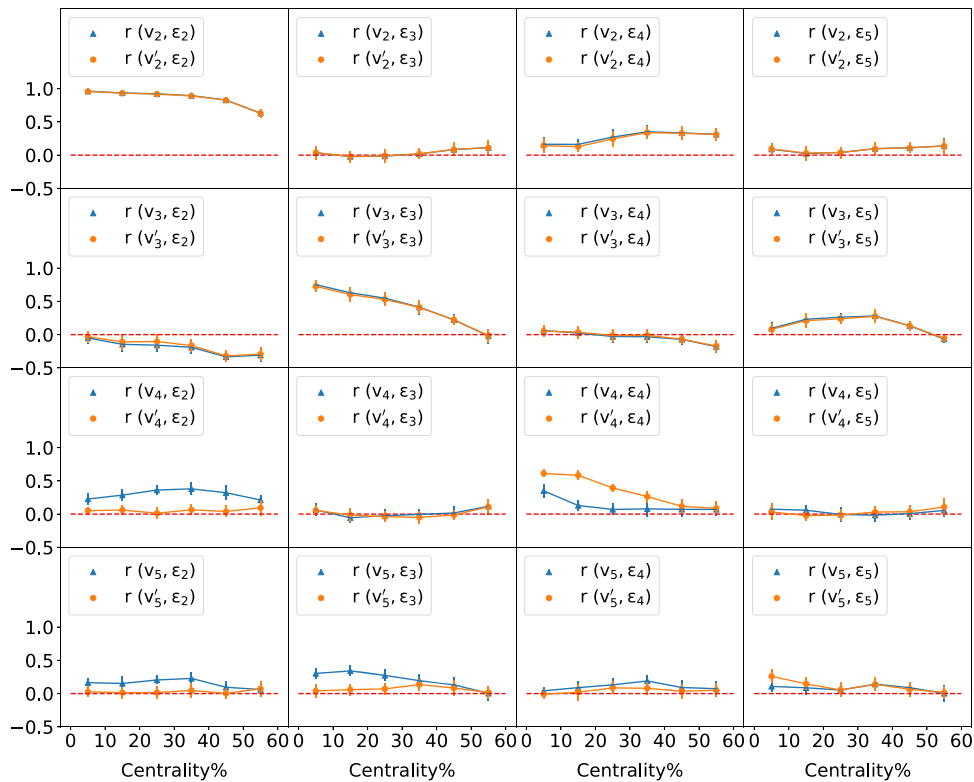
Fig. 4 The Pearson coefficient $r(v'_n, \varepsilon'_m)$ from PCA and $r(v_n, \varepsilon_m)$ from Fourier expansion, for VISH2+1 simulated Pb + Pb collisions at various centralities

$$r(v'_n, \varepsilon_m) = \frac{\langle v'_n \varepsilon_m \rangle - \langle v'_n \rangle \langle \varepsilon_m \rangle}{\sqrt{(v'_n - \langle v'_n \rangle)^2 (\varepsilon_m - \langle \varepsilon_m \rangle)^2}} \qquad (6)$$

Here, ε_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 ε'_m from PCA is highly similar to ε_m from the traditional method. We thus use ε_m in the Pearson coefficient definition $r(v'_n, \varepsilon_m)$ for PCA. Meanwhile, we can also calculate the Pearson coefficient $r(v_n, \varepsilon_m)$ for the traditional flow with Fourier expansion, which just replaces the flow harmonics

v'_n 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.

Figure 4 plots the Pearson coefficients $r(v'_n, \varepsilon_m)$ from PCA and $r(v_n, \varepsilon_m)$ 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. 3, 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 $r(v'_2, \varepsilon_m)$ and $r(v'_3, \varepsilon_m)$ are almost overlap with the Fourier ones $r(v_2, \varepsilon_m)$ and $r(v_3, \varepsilon_m)$ as shown by these upper panels in the first two rows. Meanwhile, these diagonal Pearson coefficients $r(v'_2, \varepsilon_2)$ or $r(v_2, \varepsilon_2)$ and $r(v'_3, \varepsilon_3)$ or $r(v_3, \varepsilon_3)$ 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 ε_2 and ε_3 with the approximate linear relationship $v_2 \sim \varepsilon_2$ ($v'_2 \sim \varepsilon_2$) and $v_3 \sim \varepsilon_3$ ($v'_3 \sim \varepsilon_3$) [31,32].

Although v'_4 from PCA is largely deviated from the traditional v_4 in Fig. 3, such PCA definition largely enhances correlations between ε_4, and also largely reduces the correlations between ε_2. For example, at 20-30% centrality, the Pearson coefficients $r(v_4, \varepsilon_4)$ is only 70% of the $r(v'_4, \varepsilon_4)$, while $r(v_4, \varepsilon_2)$ is 200% larger than $r(v'_4, \varepsilon_2)$. Traditionally, it is generally believed that v_4 is largely influenced by ε_2^2 through the non-linear evolution of hydrodynamics and the Cooper–Frye freeze-out procedure. 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'_5 and ε_2, v'_5 and ε_3, 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 $SC^{v'}(2, 3)$, almost all other symmetric cumulants $SC^{v'}(m, n)$ from PCA decrease significantly compared to the traditional $SC^v(m, n)$. Meanwhile, some certain Pearson coefficients $r(v'_n, \varepsilon_m)$ that evaluate the linearity between the PCA flow harmonics and the initial eccentricities are obviously enhanced (especially for $n \geq 4$), 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 emphasis that these

eigenvectors from PCA are modeled to be orthogonal and uncorrelated to each other. As a result, most of the symmetric cumulants $SC^{v'}(m, n)$ 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 ε_n, especially for $n = 4$. These results seem contradictory to the long believed idea that hydrodynamics evolution are highly non-linear, which leads to strong mode-coupling between different flow harmonics. Our PCA investigation has shown that such mode coupling effects could be reduced with new-defined bases for the flow analysis. With such finding, the non-linearity of the hot QGP systems created in heavy ion collisions should be re-evaluated, which we would like to further explore it with such PCA method in the near future.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: Source data are too large, so we haven't uploaded the data. If other researchers need the data of our calculations, please send us an email, we would like to share the data to them by other way.]

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 ε_n at different order, which are defined as [31]:

$$\varepsilon_n e^{in\Phi_n} = -\frac{\int r\, dr\, d\varphi\, r^n\, e^{in\varphi}\, s(r, \varphi)}{\int r\, dr\, d\varphi\, r^n\, s(r, \varphi)}, \tag{A.1}$$

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

For the PCA analysis, we first construct the initial state matrix $\mathbf{M_i}$, using the azimuthal angle distribution of the initial entropy $dS/d\varphi$ which is defined by
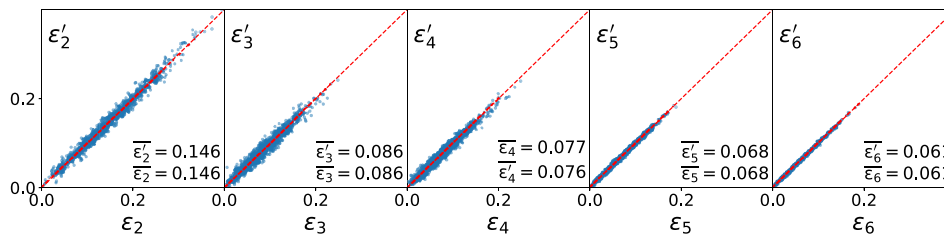
Fig. 5 A comparison between the event-by-event eccentricites ε'_n from PCA and ε_n from the traditional definition (A1), for TRENTo initial conditions at 10-20% centrality. In the corresponding panels, we also write the values of event averaged eccentricites $\overline{\varepsilon'_n}$ from PCA and $\overline{\varepsilon_n}$ from the traditional definition

$$\frac{dS}{d\varphi} = \int r^2 dr s(r, \varphi) \qquad (A.2)$$

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 $dS/d\varphi$, which is written as:

$$\left(\frac{dS}{d\varphi}\right)_{smooth} = \int_{-\pi}^{\pi} K(\varphi', \varphi) \frac{dS}{d\varphi'} d\varphi' \qquad (A.3)$$

Here, $K(\varphi', \varphi)$ is the convolution kernel, which is taken a gaussian form $K(\varphi', \varphi) = \frac{1}{\sqrt{2\pi} a} e^{-\frac{(\varphi' - \varphi)^2}{2a^2}}$. 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 $\mathbf{M_i}$ with 2000 event-by-event $(\frac{dS}{d\varphi})_{smooth}$ distributions from TRENTo for each selected centralities. As the case for the flow analysis in Sects. 2.2B and 3, the implementation of SVD and PCA to the initial state matrix, $\mathbf{M_i} = \hat{\mathbf{Y}}\hat{\mathbf{\Sigma}}\hat{\mathbf{Z}} = \hat{\mathbf{E}}\hat{\mathbf{Z}}$, gives the singular value $\hat{\sigma}_j$, eigenvectors $\hat{z}_j$ and the corresponding eccentricity coefficients $\hat{\varepsilon}_j^{(i)}$, $(j = 1, \ldots \hat{k})$ such that

$$dS/d\varphi^{(i)} = \sum_{j=1}^{m} y_j^{(i)} \hat{\sigma}_j \hat{z}_j = \sum_{j=1}^{m} \hat{\varepsilon}_j^{(i)} \hat{z}_j$$

$$\approx \sum_{j=1}^{k} \hat{\varepsilon}_j^{(i)} \hat{z}_j \qquad (i) = 1, \ldots, N \qquad (A.4)$$

We find that the PCA eigenvectors $\hat{z}_j$ of the initial states are highly similar to traditional Fourier bases $\cos(2\varphi)$, $\sin(2\varphi)$,

$\cos(3\varphi)$, $\sin(3\varphi)$, etc. Meanwhile, we could associate the singular value $\hat{\sigma}_j$ to the event averaged initial eccentricities of PCA, $\bar{\varepsilon}'_n$, at different orders and connect the coefficients $\hat{\varepsilon}_j^{(i)}$, $(j = 1, \ldots \hat{k})$ to the real or imaginary part of the PCA event-by-event initial eccentricities $\varepsilon'_n (n = 1, \ldots \hat{k}/2)$ as the case for flow.[4]

Figure 5 compares the event-by-event eccentricites ε'_n from PCA and ε_n from the traditional definition (A.1) for the TRENTo initial conditions at 10–20% centrality. It shows, with a properly chosen smoothing procedure of the initial conditions, ε'_n and ε_n agree with each other well till $n = 6$. Meanwhile, the event averaged eccentricites $\overline{\varepsilon'_n}$ from PCA and $\overline{\varepsilon_n}$ from (A.1) 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 ε_m to define the Pearson coefficient $r(v'_n, \varepsilon_m)$, and $r(v_n, \varepsilon_m)$ for both PCA and traditional flow in Sect. 3.

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 iss 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 unfold-

---

[4] Here, $\bar{\varepsilon}'_2 = \sqrt{\frac{m}{2}}\sqrt{\hat{\sigma}_3^2 + \hat{\sigma}_4^2}$, $\bar{\varepsilon}'_3 = \sqrt{\frac{m}{2}}\sqrt{\hat{\sigma}_5^2 + \hat{\sigma}_6^2}$, $\bar{\varepsilon}'_4 = \sqrt{\frac{m}{2}}\sqrt{\hat{\sigma}_7^2 + \hat{\sigma}_8^2}$, etc. with $m = 50$ the number of bins. For event-by-event definition of $\varepsilon'_n (n = 1, \ldots \hat{k}/2)$, we could simply replace $\hat{\sigma}_j$ with $\hat{\varepsilon}_j$ $(j = 1, \ldots \hat{k})$, correspondingly.
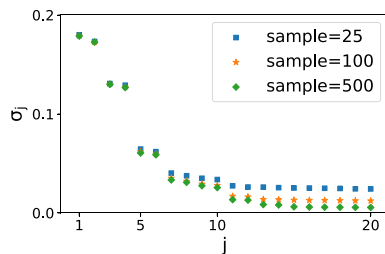
Fig. 6 The first 20 singular values of PCA σ_j ($j = 1, 2, \ldots, 20$) for the final state matrixes $\mathbf{M_f}$ with different weighted signal and noise. Such matrix $\mathbf{M_f}$ are constructed with the $dN/d\varphi$ distributions from 2000 event-by-event `VISH2+1` simulations with 25, 100 and 500 `iss` samplings

ing procedure to suppress effects from the finite multiplicites 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 `iss` samplings for each `VISH2+1` simulation to generate the $dN/d\varphi$ distributions of final particles and the related final state matrixes $\mathbf{M_f}$ with different weighted signal and noise. Then, we implement PCA to analyze these matrixes. As shown in Fig. 6, the distribution of the PCA singular values is changed with the number of `iss` samplings. For these systems with large statistical fluctuations, for example with 25 `iss` samplings, the singular values σ_j at large j tend to have a long and high tail. For these systems with reduced statistical fluctuations with more `iss` 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 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 $\frac{1}{\sqrt{25}} : \frac{1}{\sqrt{100}} : \frac{1}{\sqrt{500}}$, 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 `iss` 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 *Quark–gluon plasma*, ed. by R.C. Hwa, X.N. Wang 4th edn. (2010), pp. 207–266
2. P. Romatschke, Int. J. Mod. Phys. E **19**, 1 (2010)
3. P. Huovinen, Int. J. Mod. Phys. E **22**, 1330029 (2013)
4. U. Heinz, R. Snellings, Ann. Rev. Nucl. Part. Sci. **63**, 123 (2013)
5. C. Gale, S. Jeon, B. Schenke, Int. J. Mod. Phys. A **28**, 1340011 (2013)
6. H. Song, Y. Zhou, K. Gajdosova, Nucl. Sci. Technol. **28**(7), 99 (2017)
7. J. Shlens, CoRR. arXiv:1404.1100 (2014)
8. J.M. Hamill, X.T. Zhao, G. Mészáros, M.R. Bryce, M. Arenz, Phys. Rev. Lett. **120**, 016601 (2018)
9. S. Lloyd, M. Mohseni, P. Rebentrost, Nat. Phys. **10**(9), 631 (2014)
10. S.M. Dickerson, J.M. Hogan, A. Sugarbaker, D.M.S. Johnson, M.A. Kasevich, Phys. Rev. Lett. **111**, 083001 (2013)
11. C. Wang, H. Zhai, Phys. Rev. B **96**, 144432 (2017)
12. N.C. Costa, W. Hu, Z.J. Bai, R.T. Scalettar, R.R.P. Singh, Phys. Rev. B **96**, 195138 (2017)
13. R.S. Bhalerao, J.Y. Ollitrault, S. Pal, D. Teaney, Phys. Rev. Lett. **114**(15), 152301 (2015)
14. S.Z. Rizvi, J. Mohammadpour, R. Tóth, N. Meskin, IEEE Trans. Control Syst. Technol. **24**(5), 1883 (2016)
15. A. Mazeliauskas, D. Teaney, Phys. Rev. C **91**(4), 044902 (2015)
16. A. Mazeliauskas, D. Teaney, Phys. Rev. C **93**(2), 024913 (2016)
17. P. Bozek, Phys. Rev. C **97**(3), 034905 (2018)
18. A.M. Sirunyan et al., Phys. Rev. C **96**(6), 064902 (2017)
19. H. Song, U.W. Heinz, Phys. Lett. B **658**, 279 (2008)
20. H. Song, U.W. Heinz, Phys. Rev. C **77**, 064901 (2008)
21. H. Song, Ph.D. thesis, Ohio State U (2009)
22. C. Shen, Z. Qiu, H. Song, J. Bernhard, S. Bass, U. Heinz, Comput. Phys. Commun. **199**, 61 (2016)
23. A. Bazavov et al., Phys. Rev. D **90**, 094503 (2014)
24. J.E. Bernhard, J.S. Moreland, S.A. Bass, J. Liu, U. Heinz, Phys. Rev. C **94**(2), 024907 (2016)
25. J.S. Moreland, J.E. Bernhard, S.A. Bass, Phys. Rev. C **92**(1), 011901 (2015)
26. W. Zhao, H. Xu, H. Song, Eur. Phys. J. C **77**(9), 645 (2017)
27. H. Song, S.A. Bass, U. Heinz, Phys. Rev. C **83**, 024912 (2011)
28. J. Adam et al., Phys. Rev. Lett. **117**, 182301 (2016)
29. R.S. Bhalerao, J.Y. Ollitrault, S. Pal, Phys. Lett. B **742**, 94 (2015)
30. X. Zhu, Y. Zhou, H. Xu, H. Song, Phys. Rev. C **95**(4), 044902 (2017)
31. Z. Qiu, U.W. Heinz, Phys. Rev. C **84**, 024911 (2011)
32. D. Teaney, L. Yan, Phys. Rev. C **86**, 044908 (2012)
33. G. Aad et al., JHEP **11**, 183 (2013)