# Analysis and compression of six-dimensional gyrokinetic datasets using higher order singular value decomposition

D. R. Hatch[1,2], D. del-Castillo-Negrete[3], P. W. Terry[1]

[1] *University of Wisconsin-Madison, Madison, Wisconsin 53706*

[2] *Max-Planck-Institut für Plasmaphysik, EURATOM Association, 85748 Garching, Germany*

[3] *Oak Ridge National Laboratory, Oak Ridge TN, 37831-8071*

**Abstract**

Higher order singular value decomposition (HOSVD) is explored as a tool for analyzing and compressing gyrokinetic data. An efficient numerical implementation of an HOSVD algorithm is described. HOSVD is used to analyze the full six-dimensional (three spatial, two velocity space, and time dimensions) gyrocenter distribution function from gyrokinetic simulations of ion temperature gradient, electron temperature gradient, and trapped electron mode driven turbulence. The HOSVD eigenvalues for the velocity space coordinates decay very rapidly, indicating that only a few structures in velocity space can capture the most important dynamics. In almost all of the cases studied, HOSVD extracts parallel velocity space structures which are very similar to orthogonal polynomials. HOSVD is also used to compress gyrokinetic datasets, an application in which it is shown to significantly outperform the more commonly used singular value decomposition. It is shown that the effectiveness of the HOSVD compression improves as the dimensionality of the dataset increases.

## I. INTRODUCTION

The purpose of this paper is to explore higher order singular value decomposition [1]-[4] (HOSVD) as a tool for analyzing and compressing gyrokinetic data. HOSVD is a high-dimensional variant of singular value decomposition (SVD) [5]. SVD is a powerful and commonly used matrix decomposition which, among other applications, constitutes the basis of the proper orthogonal decomposition techniques which have been used to analyze turbulent dynamics. These techniques have been widely applied to hydrodynamic turbulence for several decades [6] and more recently to a variety of applications in fusion research including analysis of impurity transport [7], compression of magnetohydrodynamic simulation data [8], filtering of particle noise in particle-in-cell calculations [9], and analysis of the excitation of damped eigenmodes in gyrokinetic simulations [10, 11].

HOSVD has been used in a much more limited fashion than SVD, particularly with regard to analysis of turbulent systems. Nonetheless it has found several applications in a variety of fields (for specific examples see Refs. [4] and [3] and references therein). Recently HOSVD has also been used to compress aerodynamic databases [12]. In the present study, HOSVD is applied for the first time to analysis of gyrokinetic turbulence data.

The gyrokinetic model is the most comprehensive model for describing microturbulence in fusion and astrophysical plasmas [13]. One important application, which is the topic of the present work, is the turbulent transport of heat in fusion devices. The gyrokinetic model describes the time evolution of the distribution of particle gyro-centers in three spatial and two velocity dimensions. The high-dimensional nature of the HOSVD makes it a promising tool for analyzing gyrokinetic data. In contrast with a tensor decomposition, a matrix decomposition, like a matrix, is inherently two dimensional. The matrix decomposition, SVD, can be applied to higher dimensional datasets by matricizing the data (unstacking it to two dimensions), as has been done, e.g., in Refs. [7, 10, 11]. While this is effective and appropriate for many applications, it would be useful to have the ability to treat each dimension independently when analyzing higher-dimensional datasets - i.e., to analyze a tensor using a *tensor* decomposition rather than a matrix decomposition. Here we explore the use of the tensor decomposition HOSVD to analyze and compress the full six-dimensional (three spatial

dimensions, two velocity dimensions and time) gyrokinetic distribution function from nonlinear gyrokinetic simulations produced by the GENE code [14].

Gyrokinetic simulations produce a very large amount of information, but only a small fraction of this is typically analyzed, or even output to a file. An extreme example of this is when only a zero-dimensional set of diffusivities is taken from the six-dimensional dataset. Often, data output is limited to a few moments of the distribution function (consisting of the three spatial dimensions and time). Relatively little has been done to characterize velocity space dynamics (for exceptions see Refs. [10]-[11],[15]-[17]), in part because of the high-dimensional, memory-intensive nature of the task. Thus, a technique that can compress large datasets and facilitate their analysis could be very valuable. In this paper, HOSVD is used to construct tensor decompositions of such data sets in an attempt to characterize the full distribution function and gain insight into the turbulence it describes. Another very practical application is in data compression, i.e., capturing the important features of the turbulence with a reduced amount of memory usage.

One persistent theme in this work is comparison of HOSVD with the more commonly used SVD. Three measures of merit are used in this regard: the degree of computer resources that are necessary to apply the decomposition (or alternatively, the limit on the size of a dataset which can be realistically analyzed); the effectiveness of the decomposition in compressing the data set of interest; and the utility of the decomposition in providing insight and facilitating a physical understanding of the dataset. It is found that SVD is superior with regard to the first criterion (although this is possibly due only to the exhaustive optimization of SVD routines motivated by its widespread use), HOSVD is superior for the second criterion, and with regard to the third criterion, both techniques have certain advantages depending on the application. These measures of merit will be discussed throughout this paper and summarized in the conclusion.

This paper will proceed as follows: in Sec. II, the mathematics of the HOSVD will be outlined. We will discuss in what sense HOSVD is a natural extension of SVD, define notation and naming conventions, and explain the procedure for constructing an HOSVD. In Sec. III, numerical optimization and implementation of an efficient, scalable HOSVD routine is discussed. In Sec. IV

we introduce the data sets that are used. The HOSVD of the full distribution function dataset is used to analyze and understand the underlying structures for each dimension of the turbulence. Sec. V presents a description of a series of truncated HOSVDs which have been constructed from the full gyrokinetic distribution function and numerical performance, compression and truncation are discussed. Comparisons in this regard are also made with SVD. In Sec. VI the major results are summarized.

## II. Mathematics of HOSVD

In this section we will briefly review the mathematical foundations of both SVD and HOSVD. As mentioned above, the HOSVD is a higher-dimensional variant of the SVD. In this section we merely draw some analogies between the two; there exists a much more fundamental and rigorous connection between the two decompositions which is described in Ref. [3].

The SVD decomposition of a matrix $M \in \mathbb{C}^{n_1 \times n_2}$ is

$$M = USV^{\dagger}, \tag{1}$$

where $U \in \mathbb{C}^{n_1 \times min(n_1,n_2)}$ and $V \in \mathbb{C}^{min(n_1,n_2)\times n_2}$ are unitary matrices and $S \in \mathbb{C}^{min(n_1,n_2)\times min(n_1,n_2)}$ is a positive definite diagonal matrix (the superscript † denotes a conjugate transpose). The columns of $U$ are called left singular vectors, the columns of $V$ are called right singular vectors, and the non-zero elements of $S$ are called singular values. The singular values are positive and ordered in terms of magnitude.

The construction of the SVD of a matrix $M$ is straightforward. The left singular vectors are the normalized eigenvectors of the correlation matrix $MM^{\dagger}$, and the right singular vectors are the normalized eigenvectors of the correlation matrix $M^{\dagger}M$. The square roots of the first $min(n_1, n_2)$ eigenvalues (all positive) of either correlation matrix are the singular values.

An alternative notation for the SVD is,

$$M_{ij} = \sum_{l=1}^{min(n_1,n_2)} s_l u_i^{(l)} v_j^{(l)}, \tag{2}$$

where $u_i^{(l)}$ ($v_j^{(l)}$) are the left (right) singular vectors (the subscript labels the elements in each vector, and the superscript labels the vector number), and the $s_l$ are the singular values. This notation

facilitates understanding of important properties of the SVD and will also be used in describing the HOSVD. This notation reveals the SVD to be a superposition of matrices defined by outer products between the left and right singular vectors. Since the singular vectors are normalized, all amplitude information is contained in the singular values, i.e., the matrices in the series are weighted according to the amplitude of their corresponding singular values. One important property of the SVD is 'optimality': a truncated decomposition, $M_{ij}^{(r)} = \sum_{l=1}^{r<min(n_1,n_2)} s_l u_i^{(l)} v_j^{(l)}$, (a decomposition keeping a reduced number $r < min(n_1, n_2)$ of outer products in the series) is guaranteed to reproduce the original matrix better than any other decomposition of the same rank $r$. Formally this is described as

$$\epsilon^{(r)} = \|M - M_{SVD}^{(r)}\|_F^2 \leq \|M - M^{(r)}\|_F^2 \ \forall \ M^{(r)} \in \mathbb{C}^{n_1 \times n_2}, \tag{3}$$

where $M^{(r)}$ denotes an arbitrary rank-$r$ matrix, and $\|A\|_F^2 = \sum_{i,j} A_{ij}^2$ denotes the square of the Frobenius norm (note that we use the squared norm in order to give more meaning to the norm; squared physical quantities are often associated with energy-like quantities). The magnitude (as defined by the square of the Frobenius norm) of a matrix is equal to the sum of its singular values squared. Also, it can be shown that the truncation error is equal to the sum of the squares of the singular values corresponding to all truncated modes: $\epsilon^{(r)} = \sum_{l=r+1}^{min(n_1,n_2)} s_l^2$.

Having briefly reviewed the SVD, let us now transition to tensor decompositions. Consider a three dimensional tensor, $M \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ (in the discussion that follows, three dimensions will be used for simplicity of notation; it is a straightforward exercise to extend these results to higher dimensions). An appealing extension of the SVD decomposition to higher order tensors is to decompose an N-dimensional tensor as a series of outer products between N vectors with each term in the series weighted by a singular value,

$$M_{ijk} = \sum_{l=1}^{min(n_1,n_2,n_3)} s_l u_i^{(l)} v_j^{(l)} w_k^{(l)}. \tag{4}$$

Such a decomposition, in general, does not exist. If one uses a generalized notion of a singular value, replacing a series of singular values with a tensor $S_{m_1 m_2 m_3}$, a feasible decomposition can be

defined. The HOSVD is based on a decomposition of this form:

$$M_{ijk} = \sum_{m_1=1}^{n_1} \sum_{m_2=1}^{n_2} \sum_{m_3=1}^{n_3} S_{m_1 m_2 m_3} u_i^{(m_1)} v_j^{(m_2)} w_k^{(m_3)} \tag{5}$$

In this expression, the data tensor, $M_{ijk}$, is reproduced as a superposition of tensors - outer products between sets of modes defined independently for each coordinate. In this work the modes, $u_i^{m_1}$ ($v_j^{m_2}$ . . .) will be identified with the name of the corresponding coordinate (e.g., radial modes or x-modes). The tensor $S_{m_1 m_2 m_3}$ is called the core tensor.

The modes, $u_i^{m_1}$, in an HOSVD are calculated in a way that is similar to the procedure for calculating singular vectors of an SVD. Correlation matrices can be defined for each coordinate:

$$C_{pq}^{(1)} = \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} M_{pjk}^* M_{qjk}, \tag{6}$$

for coordinate one, and,

$$C_{pq}^{(2)} = \sum_{i=1}^{n_1} \sum_{k=1}^{n_3} M_{ipk}^* M_{iqk}, \tag{7}$$

for coordinate two, and so forth (the asterisk denotes the complex conjugate). The eigenvectors of these correlation matrices are chosen to be the coordinate modes:

$$\sum_{q=1}^{n_1} C_{pq}^{(1)} u_q^{(m_1)} = \sigma_{m_1}^2 u_p^{(m_1)}, \tag{8}$$

$$\sum_{q=1}^{n_2} C_{pq}^{(2)} v_q^{(m_2)} = \sigma_{m_2}^2 v_p^{(m_2)}, \tag{9}$$

and so on. Note that the correlation matrices are Hermitian by construction and so the eigenvectors are orthogonal. For SVD the eigenvalues of the correlation matrices are directly related to the singular values. In contrast, the eigenvalues of the HOSVD correlation matrices, $\sigma_{m_1}^2$, are not directly related to the core tensor $S_{ijk}$ (however, they do provide an error bound as will be described below). Rather, the orthogonality of the coordinate modes can be exploited to solve for the core tensor:

$$S_{m_1 m_2 m_3} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} M_{ijk} u_i^{*(m_1)} v_j^{*(m_2)} w_k^{*(m_3)}. \tag{10}$$

This operator is often called the Tucker operator [4] and will be called by that name in this paper. This defines a method for determining the elements of the core tensor and thus producing a decomposition of the form of Eq. 5.

Multiple truncation ranks are necessary to define a truncated HOSVD,

$$M_{ijk}^{(r_1 r_2 r_3)} = \sum_{m_1=1}^{r_1} \sum_{m_2=1}^{r_2} \sum_{m_3=1}^{r_3} S_{m_1 m_2 m_3} u_i^{(m_1)} v_j^{(m_2)} w_k^{(m_3)}, \tag{11}$$

where $r_1(r_2...)$ are the truncation ranks and correspond to the number of 1(2...)-modes retained in the decomposition. In analogy to SVD, the magnitude of the data tensor is equal to the sum of the squares of the elements of the core tensor: $\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} M_{ijk}^2 = \sum_{m_1=1}^{n_1}\sum_{m_2=1}^{n_2}\sum_{m_3=1}^{n_3} S_{m_1 m_2 m_3}^2$. Also, the truncation error is equal to the sum of the squares of the $m_n > r_n$ elements of the core tensor,

$$\epsilon_{HOSVD}^{(r_1 r_2 r_3)} = \| M_{ijk} - M_{ijk}^{(r_1 r_2 r_3)} \|_F^2 = \sum_{m_1=r_1+1}^{n_1} \sum_{m_2=r_2+1}^{n_2} \sum_{m_3=r_3+1}^{n_3} S_{m_1 m_2 m_3}^2. \tag{12}$$

It can be shown that the eigenvalues of the correlation matrices can be used to determine an error bound for a given set of truncation ranks:

$$\epsilon_{\text{Bound}}^{(r_1 r_2 r_3)} = \sum_{m_1=r_1+1}^{n_1} \sigma_{m_1}^2 + \sum_{m_2=r_2+1}^{n_2} \sigma_{m_2}^2 + \sum_{m_3=r_3+1}^{n_3} \sigma_{m_3}^2 \geq \epsilon_{HOSVD}^{(r_1 r_2 r_3)} \tag{13}$$

This is a very useful property of the HOSVD; the most computationally expensive step in calculating the decomposition is the calculation of the core tensor (depending on the truncation ranks). This error bound allows one to determine a sufficient set of truncation ranks for a specified error tolerance after solving the correlation matrix eigenproblems but before calculating the core tensor.

One drawback of the HOSVD is that there is no analogue to the SVD optimality theorem; one must perform the calculation in order to see if it has good properties in terms of compression or extracting important features. However, the error bound in Eq. 13 gives an indication why the HOSVD can be useful for these purposes; if the magnitudes of the eigenvalues of the correlations matrices decrease rapidly then the truncation error is ensured to be correspondingly small.

## III. Numerical Optimization and Implementation

Two challenges are presented in producing a numerical implementation of the HOSVD algorithm. First, due to its high-dimensional nature, the important operations in the HOSVD are tensor operations. Unfortunately, nearly all optimized numerical libraries cater to matrix operations rather than tensor operations. Second, the datasets of interest are very large, so memory

constraints quickly become an issue. The solution to the first problem is to 'matricize' the tensors and tensor operations, and the solution to the second problem is to parallelize the algorithm. These efforts will be described below.

Recall from the previous section that there are three steps in calculating an HOSVD:

1). Calculation of the correlation matrices. For a D-dimensional tensor with $n = n_1 = ...$, this requires $\mathcal{O}(n^{D+1})$ operations.

2). Calculation of the eigenvectors and eigenvalues of the correlation matrices. Since each correlation matrix is only $n \times n$ (compared to the entire $n^D$ tensor) this is a very inexpensive calculation; if the problem as a whole is tractable then this step is trivial.

3). Extraction of the core tensor. For the entire decomposition this scales as $\mathcal{O}(n^{2D})$ (for every value of the core tensor ($n^D$), one must cycle through every element of the data-tensor as seen in Eq. 5). As such, optimization of this step is critical to producing an efficient implementation of the HOSVD algorithm.

The optimization strategy that has been used here is to formally map the data-tensor onto a 2-D matrix and then convert the tensor version of the Tucker operator into a matrix operator. At this point, standard optimized numerical libraries can be exploited to efficiently calculate the core tensor. The important steps are outlined in Refs. [4],[19]. First the data-matrix must be matricized - mapped from a D-dimensional tensor onto a 2-dimensional matrix (see Sec. 3.4 in Ref. [19] for a formal description of this intuitive but notationally awkward procedure). Then the Tucker operator must be converted into a matrix operator. For a 6-D tensor (like those analyzed in this study) the tensor version of the Tucker operation is,

$$S_{m_1 m_2 m_3 m_4 m_5 m_6} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_4}\sum_{m=1}^{n_5}\sum_{n=1}^{n_6} M_{ijklmn} u_i^{*(m_1)} v_j^{*(m_2)} w_k^{*(m_3)} x_l^{*(m_4)} y_m^{*(m_5)} z_n^{*(m_6)}. \quad (14)$$

The matricized version of this operation is,

$$S^{(6\to 2)} = (U \otimes V \otimes W) M^{(6\to 2)} (Z \otimes Y \otimes X)^T. \quad (15)$$

In this equation the notation $S^{(6\to 2)}$ is used to denote the remapping of a 6-D tensor to a 2-D matrix, and $U, V$ (etc.) are the matrices containing the coordinate modes. The symbol $\otimes$ denotes

the Kronecker product,

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n_2}B \\ \vdots & \ddots & \vdots \\ a_{n_11}B & \cdots & a_{n_1n_2}B \end{pmatrix}. \tag{16}$$

In this construction, the first three coordinates of $M_{ijklmn}$ are mapped to the rows of the matrix (so that $U$, $V$, and $W$ are on the left side of the matrix multiplication) and the other three coordinates are mapped to the columns (so that $Z$, $Y$, and $X$ are on the right side of the matrix multiplication). Note that the tensor version of the operation requires $\mathcal{O}(n^{12})$ operations, whereas the matrix version requires only $\mathcal{O}(n^9)$ operations, (two matrix multiplications of $n^3 \times n^3$ matrices). For $D$ dimensions this corresponds to a reduction from $\mathcal{O}(n^{2D})$ operations for the tensor version to $\mathcal{O}(n^{3D/2})$ operations for the matrix version. Thus the matrix version of the operation is not only suitable for optimized numerical linear algebra libraries, but also requires significantly fewer operations. For a small test problem, matricizing the HOSVD operator reduced computation time by a factor of $\sim 300$.

As with the matricized HOSVD, the full SVD of a D-dimensional tensor also requires $\mathcal{O}(n^{3D/2})$ operations. Even so, trials indicate that a serial version of the HOSVD code used in this analysis is significantly slower than the LAPACK [21] SVD routine CGESVD. The HOSVD routine took $\sim 34$ seconds to perform a full HOSVD on a 13 MB, 5-dimensional, test dataset, and the SVD routine took $\sim 12$ seconds. The LAPACK SVD routine has been extensively optimized and, in many cases, tuned specifically for a machine. In light of this fact, it is plausible that if an HOSVD routine were optimized to the same level, it may be competitive with SVD with regard to this performance measure.

In order to deal with large datasets, parallelization is also necessary. The PBLAS [22] matrix multiplication routine, PCGEMM, was used for the matrix multiplications in the matricized Tucker operator. A parallel routine for calculating the correlation matrices was written from scratch using MPI. Parallel scaling tests indicate that the code scales as expected with problem size; for a 6-dimensional test tensor (with equal length of $n$ in each dimension), calculation of the full core tensor scales as $n^9$, and calculation of the correlation matrices scales as $n^7$. The calculation of the

correlation matrices is comparatively slow - depending on truncation ranks, it often takes more time than the calculation of the core tensor. Efforts to optimize this step would be worthwhile if computer resources are limited.

## IV. Analysis of Gyrokinetic Data

### A. Description of gyrokinetic dataset

In this work we analyze gyrokinetic data describing the nonlinear turbulence produced by a variety of plasma microinstabilities thought to be important transport mechanisms in fusion plasmas. The gyrokinetic model describes the time evolution of the distribution of particle gyro-centers in three spatial and two velocity dimensions. In many applications, distribution functions for multiple particle species (i.e., electrons and ions) are evolved self-consistently. In this study, we use toroidal $s - \alpha$ geometry and employ the flux tube assumption [23], resolving an extended domain in the direction following an equilibrium magnetic field line, and a much more limited domain in the directions perpendicular to the magnetic field. The perpendicular coordinates are the radial $x$ coordinate and the mostly-poloidal (sometimes called binormal) $y$ coordinate. In this work, these perpendicular coordinates will often be expressed in terms of the corresponding Fourier representation, $k_x$ and $k_y$. The parallel coordinate (called here the $z$-coordinate) is parameterized by the poloidal angle, and covers one poloidal circuit. A periodic boundary condition cannot, in general, be applied because of the distortion of the simulation domain due to a radially sheared magnetic field. As such, the parallel boundary condition connects the parallel boundaries of different $k_x$ modes at the same $k_y$ [23]. The velocity space is also parameterized in relation to the magnetic field with $v_{||}$ being the velocity parallel to the field, and $\mu$ being the magnetic moment (proportional to the perpendicular velocity squared). In the gyrokinetic description, the third velocity coordinate is removed from the equations by averaging over the angle of gyro-motion. This distribution function evolves in time as described by the coupled gyrokinetic-Maxwell equations (see, e.g., [13]). The GENE code [14] is used for all gyrokinetic simulations in this study. When using the flux tube assumption, the GENE code employs a Fourier representation in the $x$ and $y$ coordinates so

| | $q$ | $\hat{s}$ | $\epsilon = r/R$ | $n_i/n_e$ | $T_i/T_e$ | $R/L_{T_i}$ | $R/L_{T_e}$ | $R/L_n$ | $\beta$ | $\nu$ |
|---|---|---|---|---|---|---|---|---|---|---|
| ITG | 1.4 | 0.8 | 0.18 | 1.0 | 1.0 | 6.9 | n/a | 2.2 | 0.0 | 0.0 |
| ITG-coll | 1.4 | 0.8 | 0.18 | 1.0 | 1.0 | 6.9 | n/a | 2.2 | 0.0 | $5.0 \times 10^{-3}$ |
| ITG-KE | 1.4 | 0.8 | 0.18 | 1.0 | 1.0 | 6.9 | 6.9 | 2.2 | $1.0 \times 10^{-4}$ | 0.0 |
| ETG | 1.4 | 0.35 | 0.18 | 1.0 | 1.0 | n/a | 6.9 | 2.2 | 0.0 | 0.0 |
| TEM | 1.4 | 0.8 | 0.16 | 1.0 | 1.0/3.0 | 0.0 | 6.0 | 3.0 | $1.0 \times 10^{-3}$ | 0.0 |

Table 1: Physical parameters used in the GENE simulations. The parameters are safety factor $q$, magnetic shear $\hat{s}$, inverse aspect ratio $r/R$, ion (electron) density $n_{i(e)}$, ion (electron) temperature $T_{i(e)}$, ion (electron) temperature gradient scale length $R/L_{T_{i(e)}}$, density gradient scale length $R/L_n$, plasma $\beta$, and collision frequency $\nu$. In all cases fourth order hyperdiffusion is applied in the $z$ and $v_{||}$ coordinates.

that the full functional dependence of the distribution function is, $g_{k_x,k_y}(z, v_{||}, \mu, t)$. The numerical representation of this distribution function is a six-dimensional tensor. In a tensor context, the gyrokinetic distribution function will be denoted as $\mathcal{G}$.

A variety of simulations were performed with the GENE code, representing turbulence driven by different instabilities: ion temperature gradient (ITG) driven turbulence using the cyclone base case parameters [24], electron temperature gradient (ETG) driven turbulence [14, 25], and trapped electron mode (TEM) [26, 27] turbulence using parameters similar to those in Refs. [26, 28]. The physical parameters for these simulations are shown in table 1, and the numerical parameters for both the original simulations and the HOSVD analyses are shown in table 2. For the ITG case several variations have been examined – 1) the standard case using the adiabatic electron approximation (wherein only the ion distribution function is evolved) and no collision operator, 2) the identical case using a linearized Landau-Boltzmann collision operator, and 3) a case with kinetic electrons described by a self-consistently-evolved electron distribution function. The ETG simulation uses the adiabatic ion approximation and thus evolves only the electron distribution function. The TEM simulation necessarily evolves both an ion and an electron distribution function. In all cases, fourth order hyperdiffusion is applied in the $z$ and $v_{||}$ coordinates. This hyperdiffusion eliminates small scale numerical effects [29] and also acts as an energy sink in the absence of a collision operator. We focus largely on the standard ITG case and use the other simulations to gain a sense of the type and extent of the variation observed for a broader class of turbulence scenarios.

| | $L_x$ | $L_y$ | $N_{k_x}$ | $N_{k_y}$ | $N_z$ | $N_{v_{\|\|}}$ | $N_\mu$ | Species | $N_{t_{HOSVD}}$ | $\Delta t_{HOSVD}$ | Size(GB) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ITG | 125.6 | 125.6 | 64 | 16 | 16 | 32 | 8 | i | 240 | 100 | 8.1 |
| ITG-coll | 125.6 | 125.6 | 64 | 16 | 16 | 32 | 8 | i | 230 | 100 | 7.7 |
| ITG-KE | 125.6 | 104.7 | 128 | 24 | 16 | 32 | 8 | i,e | 220 | 400 | 22.5 |
| ETG | 114.3 | 125.6 | 64 | 16 | 16 | 32 | 8 | e | 280 | 200 | 11.7 |
| TEM | 94.2 | 78.5 | 128 | 24 | 16 | 40 | 8 | i,e | 250 | 400 | 31.5 |

Table 2: Numerical parameters used in the GENE simulations and the HOSVD analysis. The parameters are box size in the $x/y$ directions, $L_{x/y}$, number of $k_{x/y}$ modes, $N_{k_{x/y}}$, number of grid points in the $v_{||}/z/\mu$ coordinate, $N_{z/v_{||}/\mu}$, particle species (electron or ion), number of time steps used in the HOSVD analysis, $N_{t_{HOSVD}}$, factor determining frequency of time step data output (e.g. the distribution function was output at every 100 time steps for the ITG simulation), $\Delta t_{HOSVD}$, and size of the data set in gigabytes.

### B. Mode structures and eigenvalue spectra

This subsection presents information from coordinate modes and eigenvalues in order to illustrate the utility of the HOSVD in providing insight into the full gyrokinetic distribution function. We focus here on the standard ITG simulation and discuss the other types of turbulence in the next subsection. In this section HOSVD is applied to a dataset in which the perpendicular coordinates are represented in direct space ($x$,$y$) rather than wavenumber space ($k_x, k_y$). This is done in order to extract modes in the perpendicular directions that are more easily interpreted. This direct space data set is only used to look at the mode structures; all other results (e.g., compression rates, etc.) use the data in its original Fourier representation. The eigenvalue spectra for both cases are very similar, indicating that the form (direct space or Fourier space) of the data set has little impact on such results.

The following paragraphs will discuss mode structures for each of the six dimensions. An indication of the importance of each mode structure is given by its corresponding eigenvalue (recall Eqs. 8,9), since these eigenvalues of the correlation matrices define an error bound for a truncated HOSVD (as described in Eq. 13). The spectra of eigenvalues for all six coordinates are shown in Fig. 1. The eigenvalues for the velocity space coordinates fall off very steeply. One can then infer

that only a few structures in velocity space are important in the dynamics, and that the higher order structures can be dropped in a truncation without losing important velocity space information. In contrast, the eigenvalue spectra for the other coordinates drop of very gradually after a sharp initial decrease. This indicates broadband fluctuations in these coordinates. In spite of the slow decay of the spatial eigenvalues, there is typically a difference of one or two orders of magnitude between the first and last eigenvalues. For most purposes, gyrokinetic simulations are constructed to ensure convergence in transport levels (associated with low order moments of the distribution function). In contrast, the 'energy' quantity associated with the HOSVD analysis is a numerical sum over the six-dimensional distribution function which is not relatable to the transport quantities in a simple way. Further decay of the eigenvalue spectrum would be ensured by increasing the resolution of the corresponding gyrokinetic simulation.

For ITG turbulence with adiabatic electrons, zonal flows [30] (characterized, in part, as $k_y = 0$ fluctuations) are known to attain very high intensity. Thus it is unsurprising that the $n = 1$ binormal ($y$) mode structure represents this feature of the turbulence - it is a near constant function of $y$ as seen in Fig. 2 (note that this represents not only zonal flows but also any other $k_y = 0.0$ fluctuation). A series of other $y$-modes are also shown in Fig. 2, where it is seen that the modes exhibit largely harmonic structures. Fig. 3 shows the Fourier spectra for the entire set of $y$-modes. It is instructive to compare the mode structures in Fig. 2 with their Fourier spectra in Fig. 3. For instance, two scales are plainly visible for the $n = 10$ $y$-mode as seen in Fig. 2. Fig. 3 verifies this by showing strong Fourier components at $k_y\rho_i = 0.25$, $k_y\rho_i = 0.05 - 0.1$. Modes two through five have strong $k_y\rho_i = 0.2$, and $k_y\rho_i = 0.15$ components. Both of these examples suggest that there is some nonlinear phenomenon that couples certain scales. It is interesting to note that the growth rate peaks at $k_y\rho_i = 0.3$, but the nonlinear spectrum peaks at $k_y\rho_i = 0.2$. This observation of coupled scales could lend insight into this nonlinear downshift in the spectrum. For $n > 10$ the Fourier spectra are strongly peaked about single wavenumbers; each increasing increment in mode number $n$ is associated with an increase of one $\Delta k_y\rho_s = 0.05$. This indicates that a Fourier representation is well-suited for those scales.

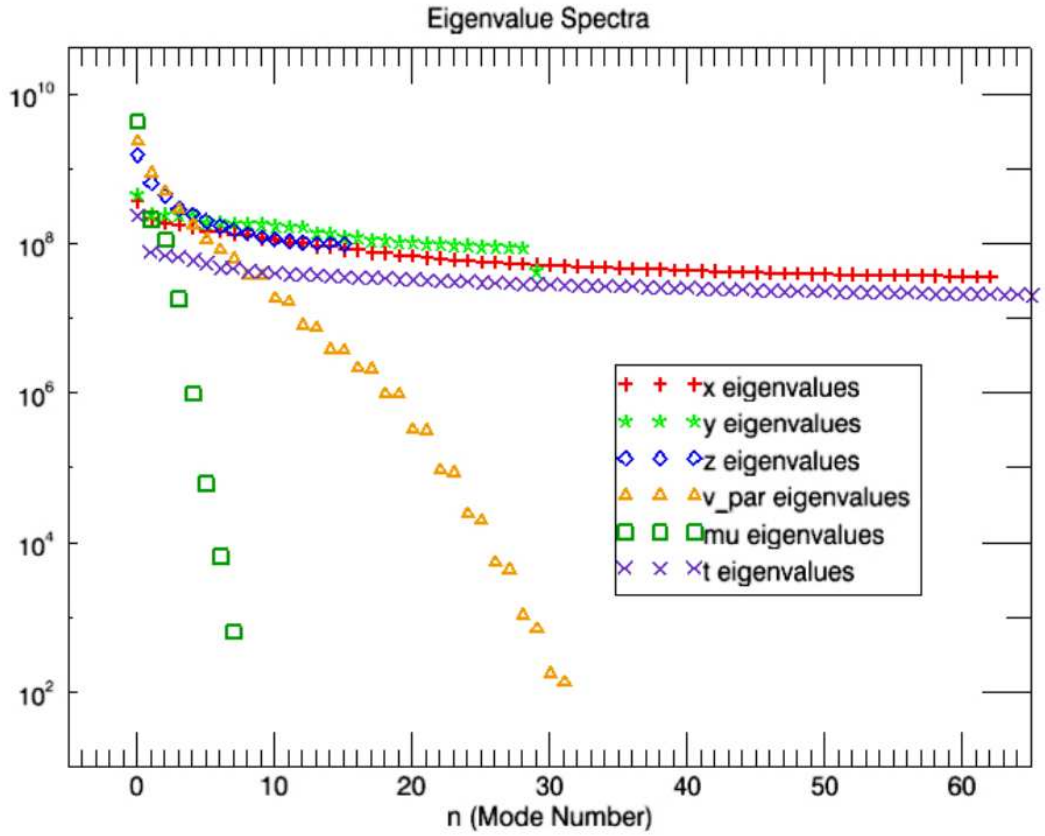


Figure 1: Spectra of eigenvalues of the coordinate correlation matrices. The steep drop-off of the eigenvalues corresponding to the velocity space coordinates indicates that the important velocity structures can be captured with only a few mode structures.

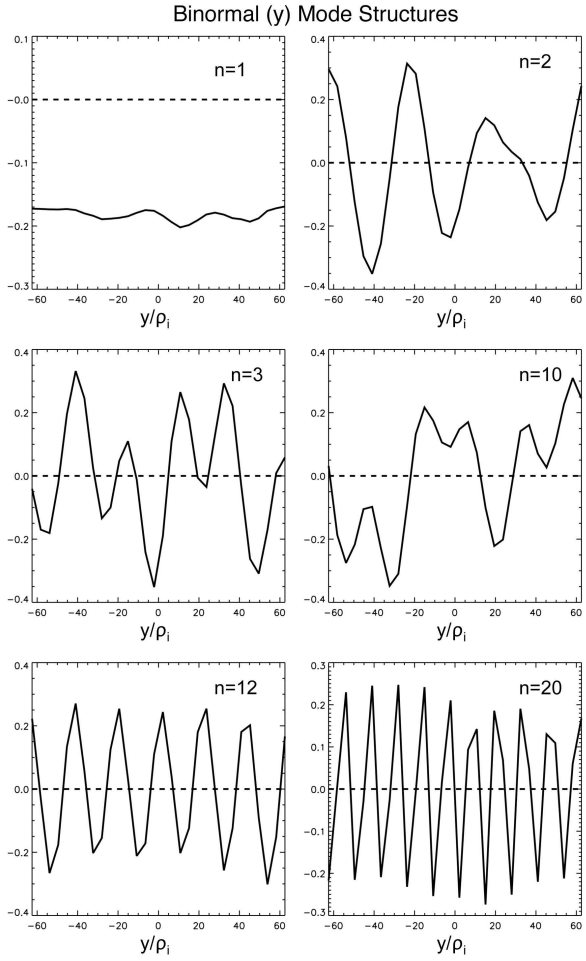


Figure 2: Selected HOSVD mode structures for the $y$ coordinate.

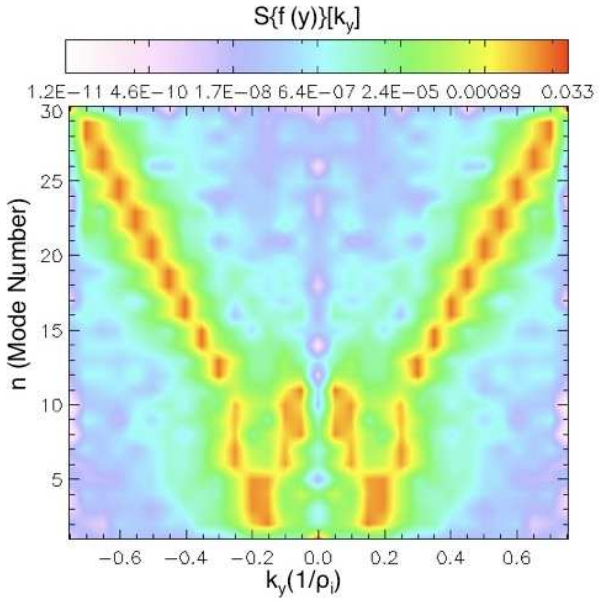


Figure 3: Fourier spectra of the HOSVD $y$-modes. At low $k_y$ some modes represent multiple scales. At higher $k_y$ the modes largely correspond to single Fourier modes (note that $\delta k_y \rho_s = 0.05$).

The radial mode structures exhibit certain features in common with the binormal mode structures; the low $n$ modes exhibit some coupling between different scales while the higher $n$ modes are strongly peaked about single wavenumbers. A selection of radial mode structures is shown in Fig. 4 and the Fourier spectra are shown in Fig. 5.

Because the mode structures in the perpendicular coordinates are largely harmonic for higher mode numbers (as shown in Figs 2-5), the corresponding eigenvalue spectra (shown in Fig. 1) would be very similar to Fourier $k_x$ or $k_y$ spectra for the squared Frobenius norm. Such spectra would in turn find some degree of correspondence with energy spectra, although the latter would entail integrals over phase space and additional terms defined by the gyrokinetic free energy quantity [15].

A series of parallel mode structures is shown in Fig. 6. Again it is seen that the mode structures develop finer scale structure as $n$ increases. Also, striking regularity is exhibited in their structure - they alternate between even and odd parity and add one additional peak with each increasing increment of $n$ - a feature that is similar to many types of orthogonal polynomials. In spite of the lack of periodicity (for even numbers) in these modes, Fourier spectra (constructed by taking a fast Fourier transform) are plotted in Fig. 7 in order to demonstrate their regularly

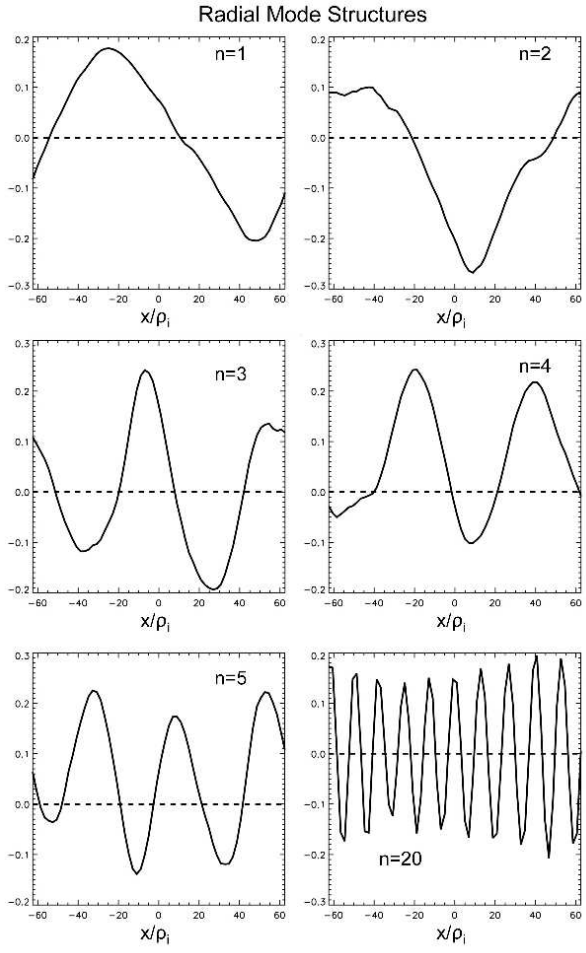


Figure 4: Selected HOSVD mode structures for the radial ($x$) coordinate.

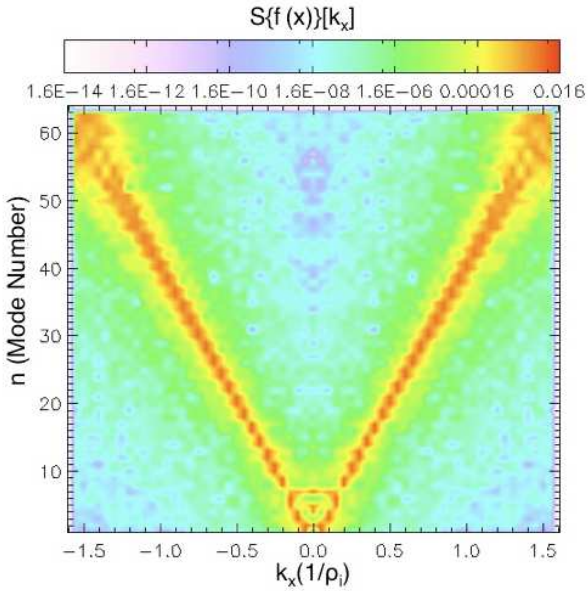


Figure 5: Fourier spectra of the HOSVD $x$-modes. At low $k_x$ some modes represent multiple scales. At higher $k_x$ the modes largely correspond to single Fourier modes (note that $\delta k_x \rho_s = 0.05$).

decreasing scales.

The $v_{||}$-modes also have a strong similarity to orthogonal polynomials; in this case they appear to be very similar to Hermite polynomials (including the Gaussian weight function) as seen in Fig. 8. This is verified for $n < 5$ in Fig. 9 where the Fourier spectra (A) and Hermite spectra (B) are plotted. It is seen that the Hermite spectra are sharply peaked at single values for $n < 5$. In certain applications Hermite polynomials have been widely used in plasma kinetic theory and simulation [17],[31]-[36]. The present analysis indicates that this representation is quite efficient for capturing the most important $v_{||}$ structures. It is also interesting to note that the fine scale structure exhibited in the $v_{||}$-modes is associated with collisional dissipation (or artifical dissipation in the case of $v_{||}$ hyperdiffusivities), since a collision operator involves second derivatives in velocity space. This is related to the results described in Ref. [10] where it is shown that both energy drive *and* dissipation peak at the same scales (low $k_x, k_y$) due to the excitation of damped eigenmodes which exhibit fine scale structure in $v_{||}$.

The $\mu$-modes are shown in Fig. 10. These mode structures exhibit peaks at successively higher values, a property shared with certain special functions.

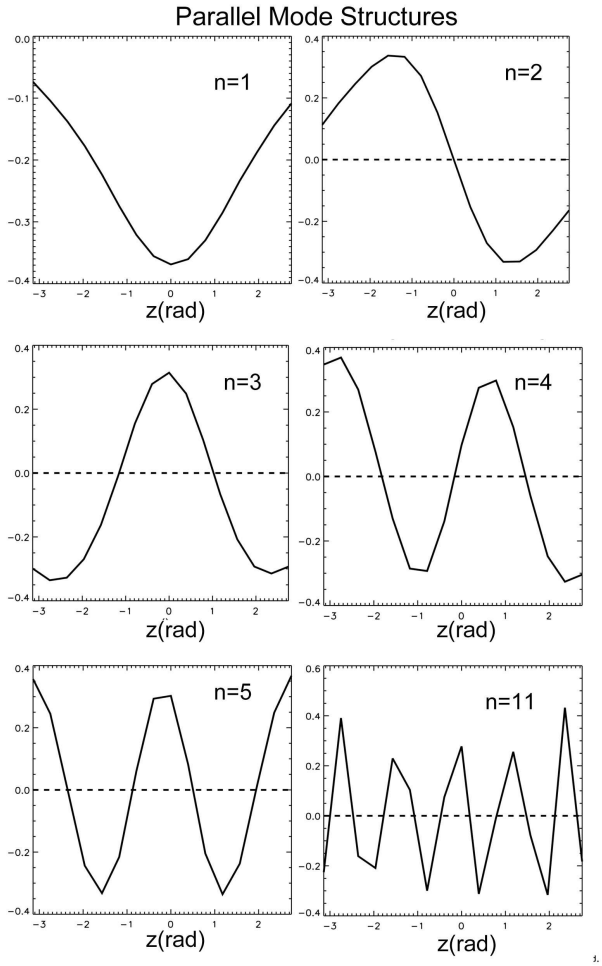


Figure 6: Selected HOSVD mode structures for the parallel ($z$) coordinate.

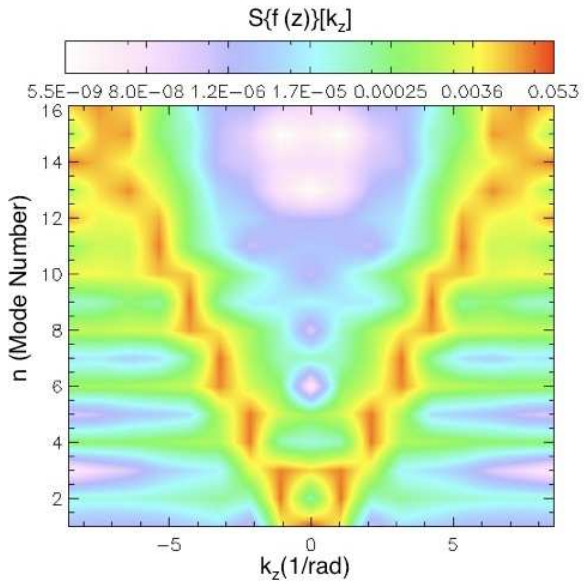


Figure 7: Fourier spectra of the HOSVD $z$-modes.

The $n = 1$ $t$-mode corresponds to the zonal flows ($n = 1$ $y$-mode), which are very low frequency fluctuations. This structure, along with several other $t$-modes, is shown in Fig. 11. The corresponding frequency spectra are shown in Fig. 12 where it is seen that the modes exhibit broadband spectra.

Note that all the results presented so far in this section have only used the coordinate modes and eigenvalues. This can be done without extracting any portion of the core tensor. As described in Sec. II, the core tensor assigns a weight to each tensor (outer product between different combinations of coordinate modes) in the decomposition. The elements of the core tensor can be ordered by magnitude. This is plotted in Fig. 13 for the 5000 largest elements of the core tensor. The largest element of the core tensor is $S_{111211}$. This defines a weight for the outer product between the $n = 2$ $v_{||}$-mode and the $n = 1$ modes for all the other coordinates. These mode structures are plotted together in Fig. 14 for easy reference. Since the $n = 1$ $y$-mode is part of this outer product, this reproduces some aspect of the zonal flow. The next four largest combinations also include the first $y$-mode. The largest combination which doesn't include the $n = 1$ $y$-mode is weighted by $S_{221113}$ - i.e., $n = 2$ for $x$ and $y$-modes, $n = 1$ for $z$, $v_{||}$, and $\mu$-modes, and $n = 3$ for the time coordinate. These are plotted together in Fig. 15 for easy reference.

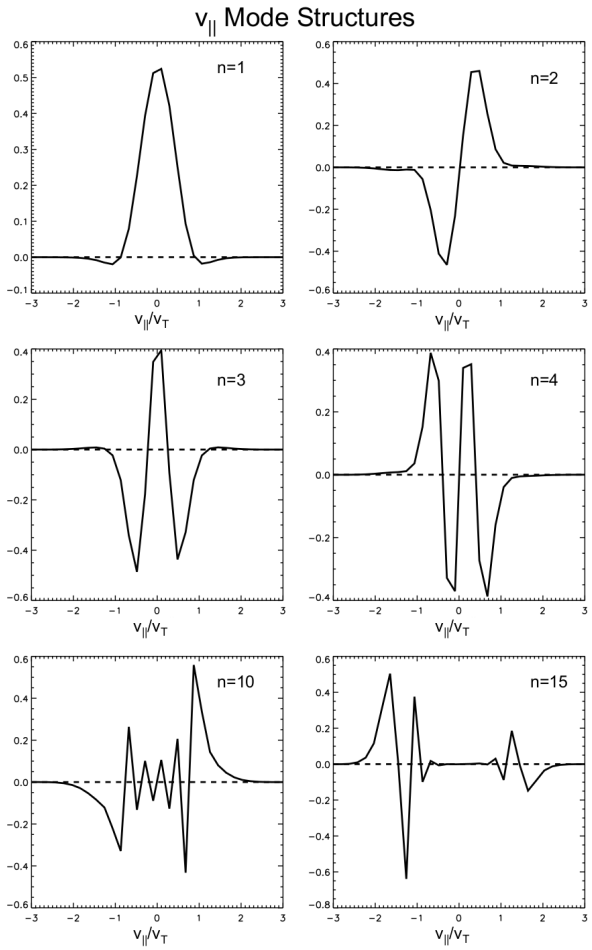


Figure 8: Selected HOSVD mode structures for the $v_{||}$ coordinate. The fine scale structures that develop as $n$ increases are associated with the collisional dissipation.

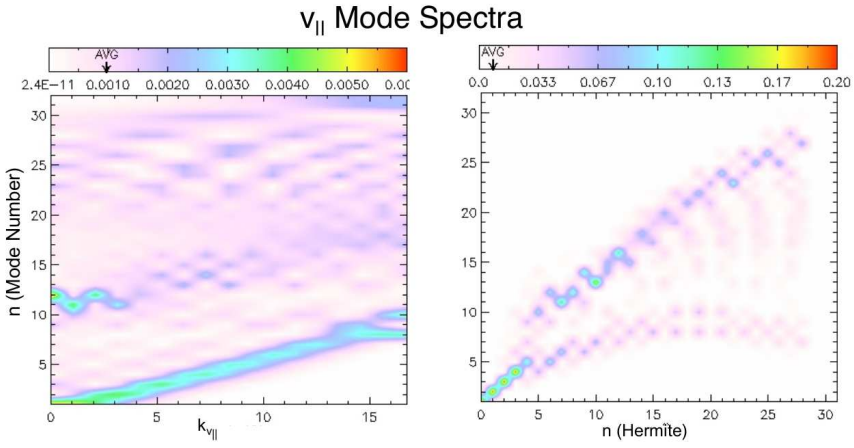


Figure 9: Fourier (A) and Hermite (B) spectra of the HOSVD $v_{||}$-modes.

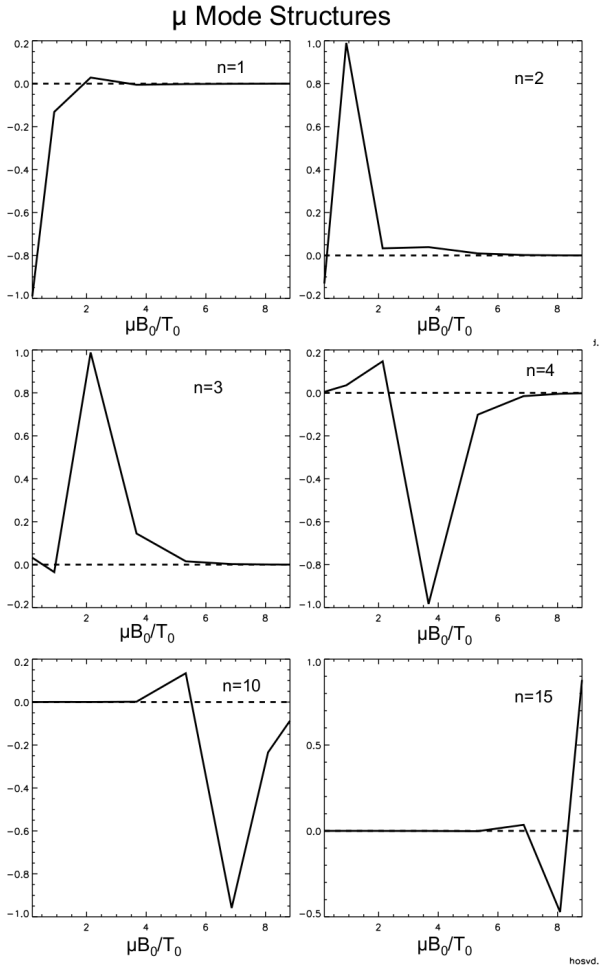


Figure 10: Selected HOSVD mode structures for the $\mu$ coordinate.

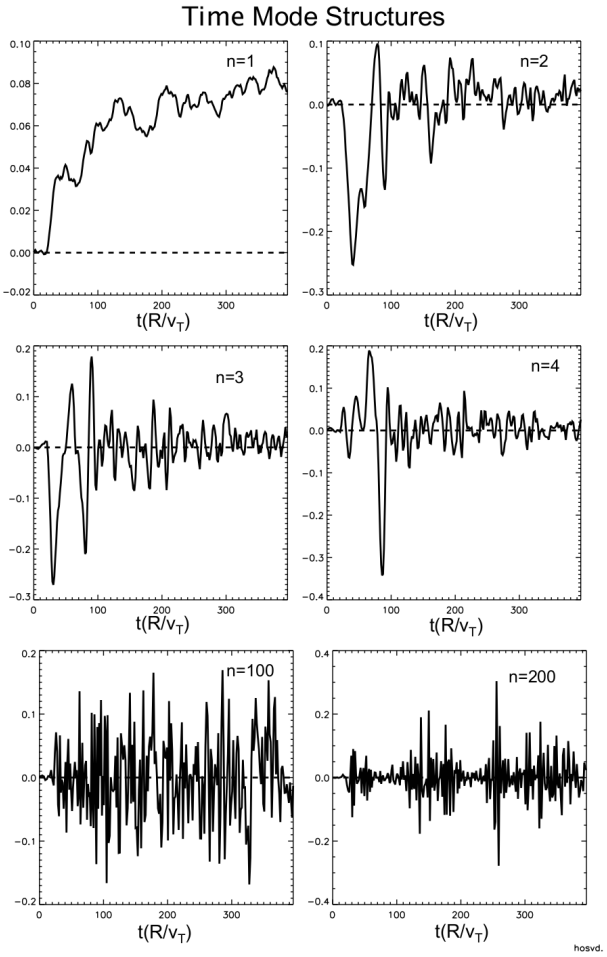


Figure 11: Selected HOSVD mode structures for the time coordinate.

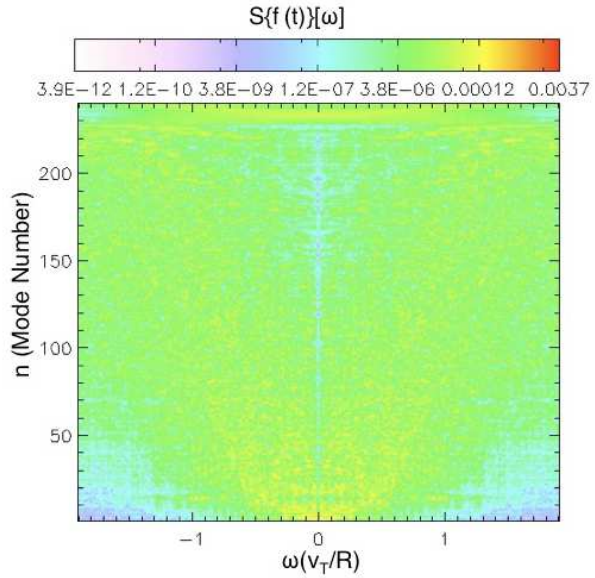


Figure 12: Frequency spectra of the time modes.

### C. Discussion of utility of HOSVD analysis

The figures discussed in the previous subsection demonstrate some of the useful features of HOSVD. Note that various coordinates are characterized by widely different types of structures. The perpendicular directions $x$ and $y$ exhibit largely harmonic structures, the parallel $z$ and $v_{||}$-modes are more akin to orthogonal polynomials, and the time structures fluctuate randomly. The HOSVD is able to extract these widely different behaviors in a single analysis. The HOSVD takes very little data to represent the mode structures and a large amount of data to explain how they combine to reproduce the tensor. While this is not always advantageous, it is easy and appealing to be able to visualize the structures that make up the entire data set at a glance. This would be useful, for instance, if one were exploring different possible coordinate bases in the process of designing or refining a code.

While Hermite representations have been used in a variety of applications in plasma kinetic theory [31]-[36], their use has not been widespread in gyrokinetic treatments. The current analysis would indicate that representing $v_{||}$ with Hermite polynomials could offer significant advantages. Since the HOSVD $v_{||}$-modes are similar to Hermite polynomials and the corresponding $v_{||}$ eigenvalue spectrum drops off steeply, a simulation could plausibly capture with a few Hermite polynomials what would otherwise require tens of gridpoints using a finite difference scheme. Furthermore, one could envision implementing large-eddy techniques to filter higher order polynomials, as has been demonstrated recently in the perpendicular wavenumbers in gyrokinetic simulations [37]. Such a scenario would utilize two favorable properties of a polynomial representation - the efficient representation of velocity space structures (as described in this work) along with the property that spectral representations are more amenable to large-eddy techniques.

### D. Comments on HOSVD analysis of additional turbulence scenarios

In this subsection we compare and contrast the additional turbulent datasets described in tables 2 and 3 with the ITG dataset described in the preceding subsections. The TEM and ITG-KE simulations use a kinetic treatment of both the electrons and ions and an HOSVD is constructed for each of these distribution functions independently. One distinguishing feature of the ITG

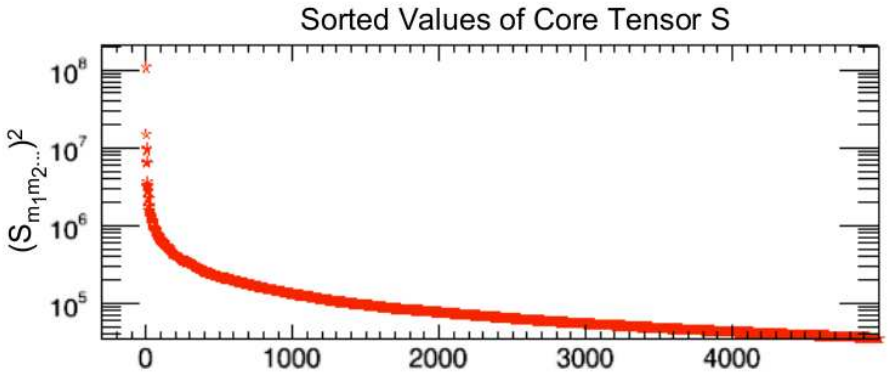


Figure 13: Spectrum of the largest 5000 squared elements of the core tensor

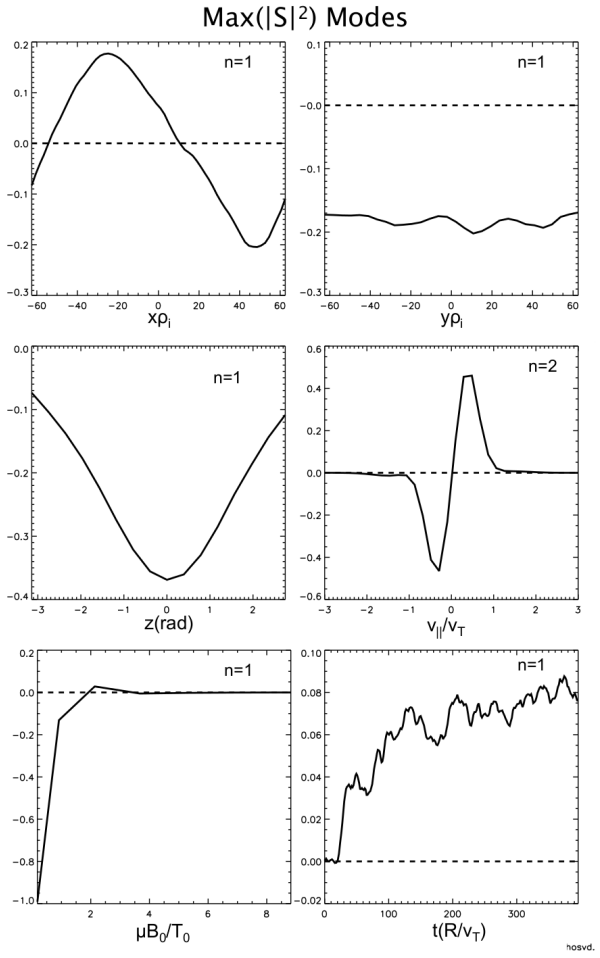


Figure 14: Plots of the coordinate mode structures associated with the largest value of the core tensor.

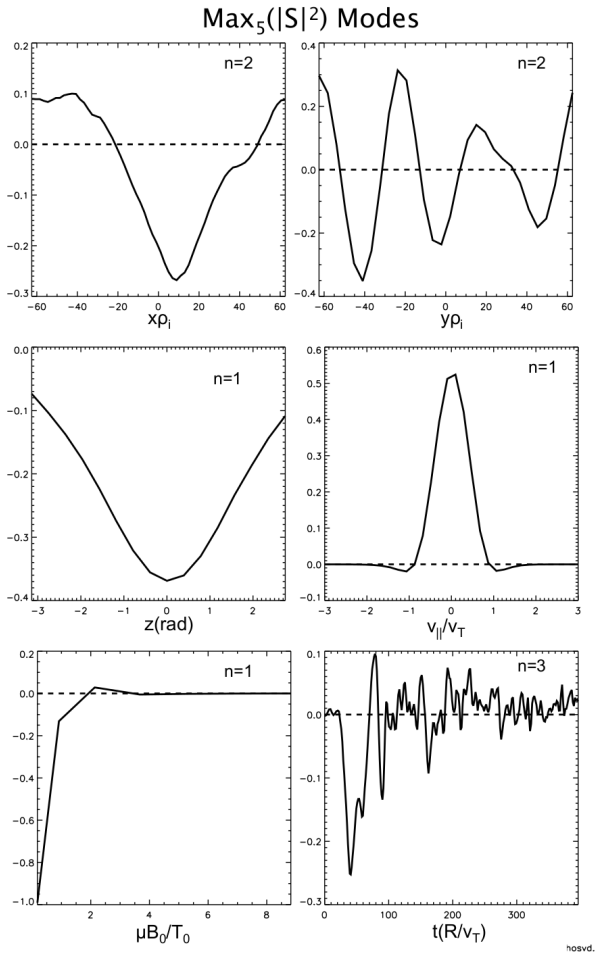


Figure 15: Plots of the coordinate mode structures corresponding to the largest value of the core tensor (the fifth largest value - denoted by $\mathrm{Max}_5$ in the title) not associated with zonal modes ($n = 1$ $y$-mode).

eigenvalue spectra (shown in Fig. 1) is the sharp decay of the velocity space eigenvalues. This feature is reproduced in all cases examined, indicating that this behavior persists in a wide range of parameter regimes. In almost all cases the $\mu$-modes are quite similar to those shown in Fig. 10, the exception being the collisional case wherein more structure is observed in the higher mode numbers. The $v_{||}$-modes produced from the ETG distribution function and the ion distribution functions for the TEM and ITG-KE case are also very similar to the Hermite-like modes shown in Fig. 8 for the ITG case. In contrast, the electron distribution functions for the TEM and ITG-KE cases produce more distinctive modes that cannot be easily related to the Hermite spectrum. Nonetheless, even in these cases, there is a steep decrease in the $v_{||}$ eigenvalue spectrum.

With regard to the other coordinate modes, some features appear in all cases observed, while others are distinctive to a certain type of turbulence. The $z$-modes are very similar in all cases. The $t$-modes for the ETG and TEM cases are very different from those in the ITG cases; they are more harmonic in nature, indicating that the wave characteristics of the turbulence may be largely retained in the nonlinear state. In addition the modes in the perpendicular directions are distinctive in each case. For the ETG case the modes peak strongly about individual Fourier modes for low $n$ (also low $k_x$ and $k_y$), but spread over many Fourier modes as $n$ increases. Additionally, the $n = 1$ $k_x$-mode is peaked almost exclusively at $k_x = 0$ as one would expect when streamers [14, 25] are dominant. In the TEM case the $k_x$ and $k_y$-modes peak at single Fourier numbers nearly throughout the entire mode spectrum. These insights likely reflect the different nonlinear saturation mechanisms in each type of turbulence.

## V. COMPRESSION OF GYROKINETIC DATA

### A. Compression and truncation errors for gyrokinetic dataset

In this section performance results and compression rates are presented for a series of truncated HOSVDs with successively smaller error tolerances. Since a truncated HOSVD is characterized by multiple truncation ranks, there is a high degree of flexibility in selecting these ranks. The scheme used in this analysis takes an error tolerance as an input. After the coordinate-eigenvalues are calculated, a routine cycles through all possible truncation ranks that satisfy this error toler-

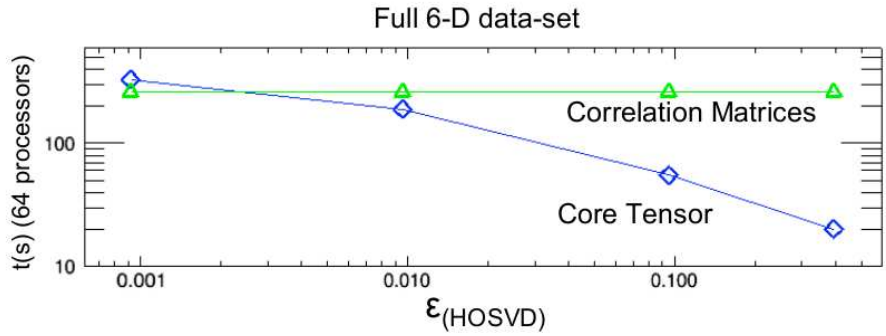


Figure 16: Plot of the computation time necessary to solve for the truncated core tensor (blue diamonds) plotted against truncation error. The computation time for the construction of the correlation matrices (independent of truncation) is also shown for reference (green triangles).

ance (as calculated using the error bound defined in Eq. 13), and selects the truncation ranks that correspond to the smallest possible inverse compression ratio, $\delta_{HOSVD}$ (defined below in Eq. 17). The cpu time necessary to calculate the truncated core tensor is plotted against the truncation error in Fig. 16 along with the time to calculate the correlation matrices (independent of truncation ranks). The matrix multiplication (Eq. 15), which is the most expensive step in calculating the core tensor, uses the highly optimized PCGEMM routine. As such, it is faster than the calculation of the correlation matrices for larger error tolerances (e.g., error above $\sim 0.002$ in Fig. 16) even though it entails more operations. Approximately 10.5 cpu hours (589 seconds on 64 processors) were necessary in order to calculate a truncated HOSVD within the 0.1% error tolerance. This use of computer resources is very manageable and much bigger problems would also be tractable. On the other hand it is expected that the data from many large gyrokinetic simulations (which include things electromagnetic effects, electron gyro-scale instabilities, etc.) could not be analyzed with an HOSVD with a reasonable allocation of computer resources. In cases like this, analysis and compression of individual time steps in the simulation could be useful and manageable.

The error for this series of truncated HOSVDs is plotted against the inverse compression ratio in Fig. 17. The inverse compression ratio is,

$$\delta_{HOSVD}^{(r_{k_x} r_{k_y} r_z r_{v_{||}} r_\mu r_t)} = \frac{r_{k_x} n_{k_x} + r_{k_y} n_{k_y} + r_z n_z + r_{v_{||}} n_{v_{||}} + r_\mu n_\mu + r_t n_t + r_{k_x} r_{k_y} r_z r_{v_{||}} r_\mu r_t}{n_{k_x} n_{k_y} n_z n_{v_{||}} n_\mu n_t}, \qquad (17)$$

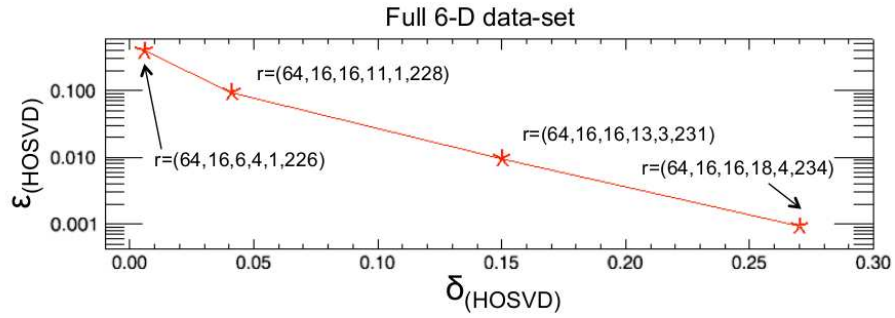


Figure 17: Plot of the truncation error versus inverse compression ratio for a series of truncated HOSVDs. The truncation ranks are noted for the $(k_x, k_y, z, v_{||}, \mu, t)$ coordinates respectively (out of a total of (64,16,16,32,8,240)).

and the (normalized) truncation error is,

$$\epsilon^{(r_{k_x} r_{k_y} r_z r_{v_{||}} r_\mu r_t)} = \frac{\|\mathcal{G} - \mathcal{G}^{r_{k_x} r_{k_y} r_z r_{v_{||}} r_\mu r_t}\|_F^2}{\|\mathcal{G}\|_F^2}, \tag{18}$$

where $\|\cdot\|_F^2$ denotes the square of the Frobenius norm. Fig. 17 shows that if one defines an error tolerance of, e.g., 10% (1%), then the dataset can be compressed to $\sim 4\%$ ($\sim 15\%$) of its original size. Decompositions like the SVD perform particularly well at extracting complex coherent structures that are not well described by the standard toolkit of special functions [20]. In light of this, it is noteworthy that these compression rates can be achieved even for a dataset like the one considered here - a data set seemingly without dominant coherent structures. The truncation ranks are also noted in Fig. 17 for the $k_x, k_y, z, v_{||}, \mu, t$ coordinates, respectively (for reference, the full rank values are $(64, 16, 16, 32, 8, 240)$). Nearly all of the truncation is in the velocity space coordinates. This is consistent with the steep drop-off in the spectra of $v_{||}$ and $\mu$ eigenvalues shown in Fig. 1. This is another indication that a basis of some sort of orthogonal polynomials could be used to advantage in a code; for example, it may take a few dozen grid points for a finite difference method to resolve the same features that can be captured by only a few orthogonal polynomials.

### B. Compression and truncation errors: comparison with SVD

One purpose for which HOSVD is particularly valuable is in data compression. The mathematical reasons why this is often expected to be the case will be discussed, and then computational results in this regard will be presented. For the purposes of this discussion, consider a D-dimensional

tensor, $\mathcal{A}^{(D)}$, with an equal number of elements, $n$, in each coordinate, and HOSVD truncation ranks, $r$ (also equal in each coordinate). The number of data elements it takes to store this truncated decomposition is $Drn + r^D$ ($rn$ for each set of modes and $r^D$ for the core tensor), and the size of the original dataset is $n^D$. Thus the inverse compression ratio for this truncation is,

$$\delta^{(r)}_{HOSVD} = \frac{Drn + r^D}{n^D} \approx \frac{r^D}{n^D}. \tag{19}$$

In order to perform an SVD on the same data set, the D-dimensional data tensor must be matricized. Assume that D is an even number and the tensor is matricized in such a way that the dimensions are split evenly between the columns and rows of the matrix: $\mathcal{A}^{(D)} \rightarrow A^{(n^{D/2} \times n^{D/2})}$. Now define a truncation rank, $r_{SVD} = r^D$, for the SVD so that the truncated SVD keeps the same number of terms (D-dimensional tensors) in the series as the truncated HOSVD. This truncated SVD requires $r_{SVD}(1 + 2n^{D/2}) = r^D(1 + 2n^{D/2})$ data elements. The compression ratio for this SVD truncation is

$$\delta^{(r_{SVD}=r^D)}_{SVD} = \frac{r^D(1 + 2n^{D/2})}{n^D} \approx \frac{2r^D}{n^{D/2}}. \tag{20}$$

Note that the majority of the SVD data elements are in the singular vectors and very few data elements (singular values) are necessary to define how the vectors combine in the decomposition. In contrast, the reverse is true for HOSVD; the core tensor (which defines how the modes combine) requires the vast majority of the storage and the modes themselves require relatively few data elements.

A comparison of Eqs. 19 and 20 indicates that the HOSVD truncation requires only a small fraction,

$$\frac{\delta^{(r)}_{HOSVD}}{\delta^{(r_{SVD}=r^D)}_{SVD}} = \frac{1}{2} n^{-D/2}, \tag{21}$$

of the memory required to store the SVD truncation, a feature that *improves* as the size $n$ and the dimensionality, $D$, of the dataset increase. However, there is a large caveat in the fact that when the truncation ranks are equated in this manner ($r_{SVD} = r^D$), the SVD truncation error is guaranteed to not exceed the HOSVD truncation error (see Eq. 3). The HOSVD stores a given number of tensors with far fewer data elements than the SVD, but the SVD produces a smaller truncation error than the HOSVD for the same number of tensors. Which decomposition technique

wins the compression comparison depends on a balance between these two properties. For the data set considered here, it is found that the storage advantages of HOSVD trump the optimality of SVD so that HOSVD achieves a much smaller truncation error for a given compression ratio.

In order to compare compression results for the two decompositions, one can plot truncation error against inverse compression ratio. Equations 17 and 18 define the inverse compression ratio and the truncation error for a truncated HOSVD of the full 6-D dataset. The following relations define the corresponding expressions for a truncated SVD (assuming the coordinates are split evenly in the matricization procedure).

$$\delta_{SVD}^{(r)} = \frac{r(1 + n_{k_x} n_{k_y} n_z + n_{v_{||}} n_\mu n_t)}{n_{k_x} n_{k_y} n_z n_{v_{||}} n_\mu n_t}. \tag{22}$$

$$\epsilon^{(r)} = \frac{\|\mathcal{G} - \mathcal{G}^{(r)}\|_F^2}{\|\mathcal{G}\|_F^2}. \tag{23}$$

Compression rates are tested for a series of datasets with increasing dimensionality - a 3-D data set $[g|_{(k_x\rho_i=0.0,k_y\rho_i=0.2)}(z=0,v_{||},\mu,t)]$ consisting of the gyrokinetic distribution function at the outboard midplane $z=0$ for $k_x\rho_i = 0.0$, and $k_y\rho_i = 0.2$, a 4-D data set which retains the $z$ coordinate $[g|_{(k_x\rho_i=0.0,k_y\rho_i=0.2)}(z,v_{||},\mu,t)]$, and a 5-D data set which additionally retains the $k_y$ dependence $[g|_{(k_x\rho_i=0.0),k_y}(z,v_{||},\mu,t)]$. The expressions for inverse compression ratio and truncation error for these reduced-dimensional datasets can be recovered from the expressions in Eqs. 17, 18, 22, and 23 by dropping the terms corresponding to the reduced dimensions.

In Fig. 18, compression rates are plotted for the 3-D dataset for SVD (blue asterisks) and HOSVD (red diamonds). Even for 3 dimensions, HOSVD produces superior compression rates. Compression rates for the 4 and 5-D data sets are plotted in Figs. 19 and 20. For the four and five dimensional SVDs, two different schemes have been used to distribute the tensor into a matrix. First, all the 'spatial' coordinates are distributed in the rows of the matrix, and only the time coordinate varies with the columns (blue asterisks). This is a common technique which produces separate spatial and time basis vectors (sometimes topos and chronos) as done in Refs. [7],[10],[11]. The second method is to distribute the coordinates more evenly between the columns and the rows

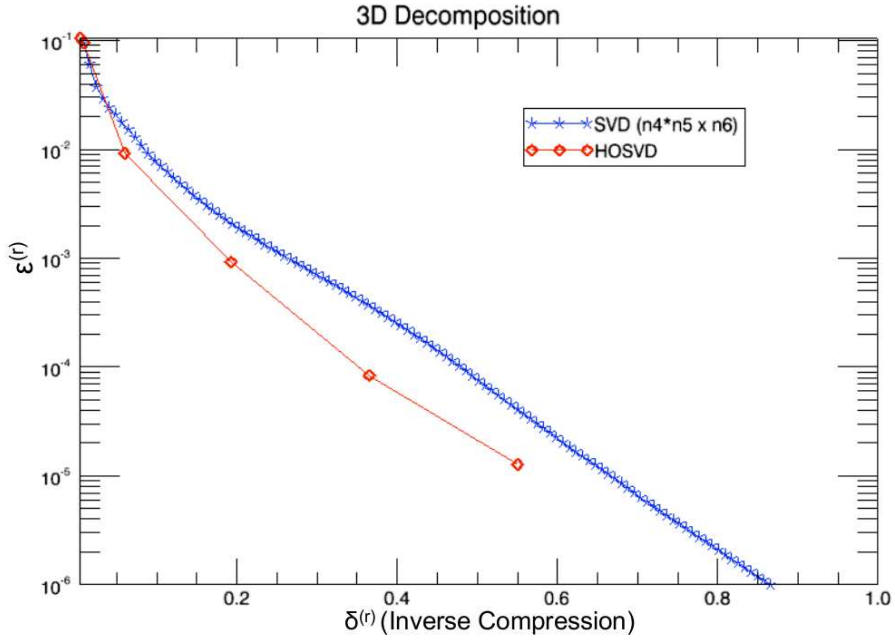


Figure 18: Comparison of truncation error versus inverse compression ratio for HOSVD (red diamonds) and SVD (blue asterisks) for a three-dimensional dataset. HOSVD achieves a smaller truncation error for a given inverse compression ratio.

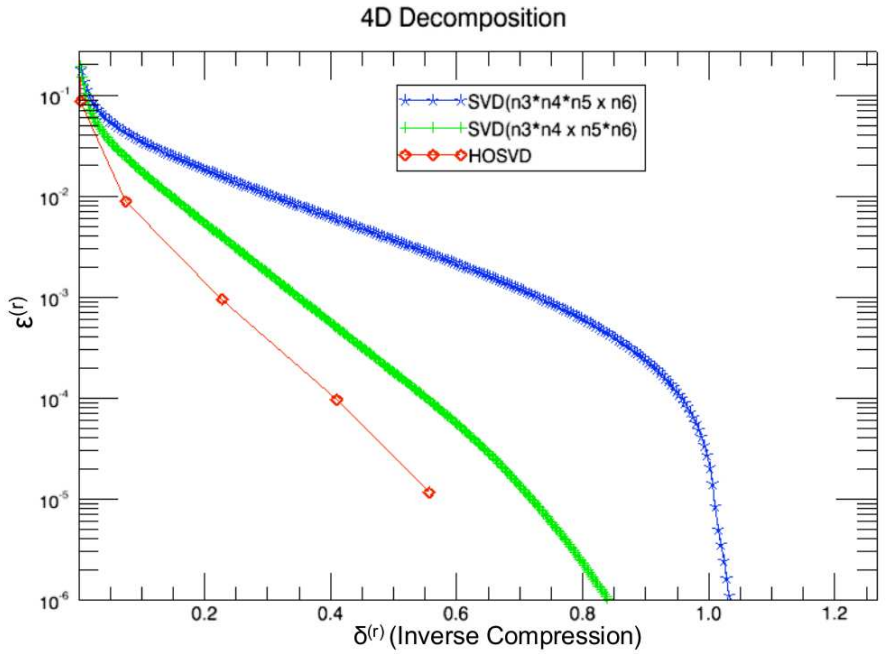


Figure 19: Comparison of truncation error versus inverse compression ratio for HOSVD (red diamonds) and two applications of SVD for a four-dimensional dataset. The blue asterisks denote SVD applied to a matrix where the time coordinate is the only coordinate varying along the columns, and the green plus signs denote SVD applied to a matrix where the coordinates are more evenly distributed. HOSVD achieves a smaller truncation error for a given inverse compression ratio.

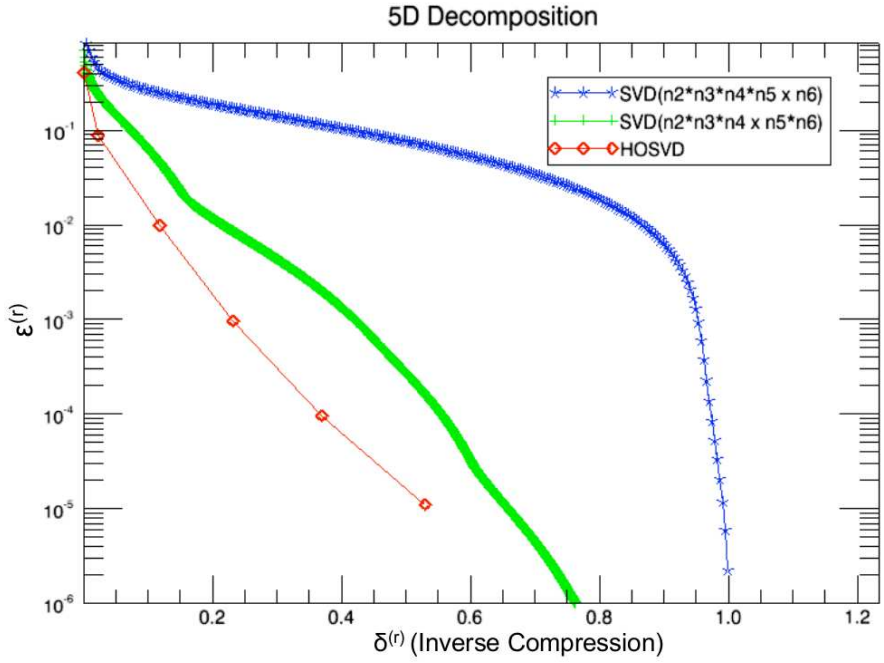


Figure 20: Comparison of truncation error versus inverse compression ratio for HOSVD (red diamonds) and two applications of SVD for a five-dimensional dataset. The blue asterisks denote SVD applied to a matrix where the time coordinate is the only coordinate varying along the columns, and the green plus signs denote SVD applied to a matrix where the coordinates are more evenly distributed. HOSVD achieves a smaller truncation error for a given inverse compression ratio. This distinction is amplified as the dimensionality increases as can be seen by comparing this figure with Figs 18 and 19.

(green plus signs). This makes interpretation of the basis vectors more difficult, but is advantageous for compression as seen in Figs. 19 and 20. In either case, HOSVD (red diamonds) significantly out-performs SVD. To illustrate this consider a situation where a truncation is held to an error tolerance of 0.01. As illustrated in Fig. 20 for the 5-D data set, a truncated HOSVD could compress the data set to $\sim 12\%$ of its original size whereas the truncated SVDs could only compress it to $\sim 21\%$ (or $\sim 87\%$ for the topos-chronos method). Analysis of this series of increasingly high dimensional datasets also demonstrates that the advantages of HOSVD with regard to compression are augmented as the dimensionality increases.

An alternative comparison of SVD and HOSVD is to compare the spectra of SVD singular values and their HOSVD analogues - the elements of the core tensor. By the theorem shown in Eq. 3, SVD must perform better in this comparison. Recall that the sum of the squares of the SVD singular values (elements of the HOSVD core tensor) is equal to the 'energy' in the entire dataset, and the square of each singular value (element of the core tensor) denotes the amount of 'energy' in each tensor in the series. There are a total of 1920 singular values and $\sim 15$ million elements in the core tensor. It is observed that the first ten or so elements of the core tensor are smaller but comparable to the SVD singular values. For numbers beyond a few dozen, the HOSVD is quite inefficient when measuring by this standard. This drawback is compensated by the favorable scaling of the compression ratio as described in Eq. 21.

## VI. SUMMARY AND CONCLUSIONS

This paper has explored the use of the tensor decomposition, HOSVD, as a technique for analyzing and compressing gyrokinetc data produced with the GENE code. As a tensor decomposition, HOSVD has the capability of treating each coordinate in a high-dimensional data set independently for an arbitrarily high number of dimensions. Thus it is well-suited for analyzing the high dimensional gyrokinetic distribution function. We have provided a mathematical description of HOSVD and drawn some analogies with the matrix decomposition, SVD. A description of the numerical implementation and optimization of an HOSVD routine was described; the major features of this effort were 1) parallelization of the routine, and 2) manipulation of the algorithm

into a form compatible with efficient linear algebra libraries.

This routine was used to analyze the entire six-dimensional (three spatial, two velocity, and time dimensions) distribution function from gyrokinetic simulations of ITG, ETG and TEM driven turbulence. In all three cases, the spectra of eigenvalues associated with the velocity space coordinates fall off very sharply, indicating that the important features of the turbulence can be captured by a small number of velocity space structures. In many cases (ITG, ETG, and the ion distribution function for the TEM and ITG with kinetic electrons cases), the $v_{||}$-mode structures are very similar to Hermite polynomials. This may suggest ways in which one could optimize velocity space discretization in gyrokinetic codes and facilitate reduced gyrokinetic modelling. Additionally this analysis provides insight into the phenomenon of damped eigenmode excitation, as the higher order phase space coordinate modes are associated with the damped modes which provide an energy sink in gyrokinetic systems.

It was also found that HOSVD is powerful tool for data-compression; for example, 90% (99%) of the data can be captured with a truncated HOSVD which uses only 4% (15%) as much memory as the total data set. In this role, HOSVD significantly outperforms SVD. With regard to compression, HOSVD improves as the dimensionality of the dataset increases. This observation suggests that HOSVD could be applied to even higher dimensional datasets expanded by including scans in additional parameters (like, e.g., gradient scale lengths, plasma $\beta$, etc.).

A persistent theme in this paper is the comparison of HOSVD with the more commonly used SVD. It is concluded that each technique has distinctive advantages. SVD is more efficient computationally while HOSVD is more effective at data compression. We have also considered the utility of these decomposition as analysis tools aimed at providing insight into the high-dimensional gyrokinetic system. In this regard HOSVD is appealing because of the independent treatment of each coordinate. Such a treatment allows one to observe independently the important structures in each dimension and thus gain intuition into a large and complex dataset.

## ACKNOWLEDGMENTS

This work was supported by Department of Energy Grant DE-FG02-89ER53291. DdCN

was sponsored by the Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S.Department of Energy under contract DE-AC05-00OR22725.

## References

[1] L. R. Tucker, Psychometrika 31 (3): 279-311 (1966).

[2] P. M. Kroonenberg "Three-mode principal component analysis: Theory and applications" DSWO Press, Leiden (1983).

[3] L. De Lathauwer, B. De Moor, and J. Vandewalle, SIAM J. Matrix Anal. Appl. **21**, 4, 1253-1278 (2000).

[4] T.G. Kolda, B.W. Bader, SIAM Review, **51**, 3, 455-500 (2009).

[5] G. H. Golub, C. F. Van Loan, Matrix Computations, (The Johns Hopkins University Press, Baltimore and London, 1989) Ed. 2, p. 71.

[6] G. Berkooz, P. Holmes, and J.L. Lumley, Annu. Rev. Fluid Mech. **25**, 539-575 (1993).

[7] S. Futatani, S. Benkadda, and D. del-Castillo-Negrete, Phys. Plasmas **16**, 042506 (2009).

[8] D. del-Castillo-Negrete, S. P. Hirshman, D. A. Spong, and E. F. D'Azevedo, Journal of Computational Physics, **222**, 265-286 (2007).

[9] D. del-Castillo-Negrete, D. A. Spong, and S. P. Hirshman, Phys. Plasmas **15**, 092308 (2008).

[10] D. R. Hatch, P. W. Terry, F. Jenko, F. Merz, W. M. Nevins, Phys. Rev. Lett. **106**, 115003 (2011).

[11] D. R. Hatch, P. W. Terry, F. Jenko, F. Merz, M. J. Pueschel, W. Nevins, E. Wang, Phys. Plasmas **18**, 055706 (2011).

[12] L.S. Lorente , J.M. Vega, A. Velazquez, Aerospace Science and Technology **14** 168-177 (2010).

[13] A. J. Brizard, and T. S. Hahm, Rev. Mod. Phys. **79**, 421 (2007).

[14] F. Jenko, W. Dorland, M. Kotschenreuther, and B. N. Rogers , Phys. Plasmas, **7**, 1904 (2000).

[15] T.-H. Watanabe, H. Sugama, Nucl. Fusion **46** , 24 (2006).

[16] T. Tatsuno, W. Dorland, A.A. Schekochihin, G.G. Plunk, M. Barnes, S.C. Cowley, and G.G. Howes, Phys. Rev. Lett. **103**, 015003 (2009).

[17] T.-H. Watanabe, and H. Sugama, Phys. Plasmas, **11**, 1476 (2004).

[18] L.R. Tucker, Psychometrika, **31**, 279311 (1966).

[19] T.G. Kolda, "Multilinear operators for higher-order decompositions", Sandia Reports, SAND2006-2081, (2006).

[20] S. Futatani, W.J.T. Bos, D. del-Castillo-Negrete, K. Schneider, S. Benkadda, M. Farge: "Coherent Vorticity Extraction in Resistive Drift-wave Turbulence: Comparison of Orthogonal Wavelets versus Proper Orthogonal Decomposition." Comptes Rendus Physique, 12, 2, 123-131 (2011).

[21] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen, LAPACK User's Guide, (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999) Ed. 3.

[22] L.S. Blackford, J. Choi, A. Cleary, E. D'azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R.C. Whaley, ScaLAPACK Users' Guide, (Society for Industrial and Applied Mathematics, Phlidelphia, PA, 1997), ISBN= 0 - 89871 - 397 - 8 (paperback).

[23] M. A. Beer, S. C. Cowley, and G. W. Hammett, Phys. Plasmas **2**, 2687 (1995).

[24] A. M. Dimits, A. M., G. Bateman, M. A. Beer, B. I. Cohen, W. Dorland, G. W. Hammett, C. Kim, J. E. Kinsey, M. Kotschenreuther, A. H. Kritz, *et al.*, Phys. Plasmas **7**, 969 (2000).

[25] W.M. Nevins, J. Candy, S. Cowley, T. Dannert, A. Dimits, W. Dorland, C. Estrada-Mila, G.W. Hammett, F. Jenko, and M.J. Pueschel, Phys. Plasmas **13**, 122306 (2006).

[26] T. Dannert and F. Jenko, Phys. Plasmas **12**, 072309 (2005).

[27] D. R. Ernst, J. Lang, W. M. Nevins, M. Hoffman, Y. Chen, W. Dorland, and S. Parker, Phys. Plasmas **16**, 055906 (2009).

[28] F. Merz and F. Jenko, Phys. Rev. Lett. **100**, 035005 (2008).

[29] M. J. Pueschel, T. Dannert, F. Jenko, Comput. Phys. Commun. **181**, 1428, (2010).

[30] P.H. Diamond, S.-I. Itoh, K. Itoh and T.S. Hahm, Plasma Phys. Controlled Fusion **47**, R35 (2005).

[31] T. P. Armstrong, Phys. Fluids 10, 1269 (1967).

[32] G. Joyce, G. Knorr, and H. K. Meier, J. Comput. Phys. **8**, 53 (1971).

[33] G. W. Hammett, M. A. Beer, W. Dorland, S. C. Cowley, and S. A. Smith, Plasma Phys. Controlled Fusion **35**, 973 (1993).

[34] J. W. Schumer, and J. P. Holloway, J. Comput. Phys. **144**, 626 (1998).

[35] C. S. Ng, A. Bhattacharjee, and F. Skiff, Phys. Rev. Lett. 83, 1974 (1999).

[36] B. Scott, Phys. Plasmas, **17**, 102306 (2010).

[37] P. Morel, A. Bañón Navarro, M. Albrect-Marc, D. Carati, F. Merz, T. Görler, and F. Jenko, Phys. Plasmas **18**, 072301 (2011).