ME 5013
Big Data Analysis Final Project Report
Data reduction and spectral analysis of turbulent field images
Eugene Hoffman

Introduction

The advent of high-speed cameras and non-intrusive diagnostic techniques for the evaluation of turbulent fields has led to the creation of larger and larger data sets and has brought about the need to develop the capability of handling large amounts of data typically in the form of images and analyzing said images to develop a better understanding of turbulent flow properties. This project sets a foundation for handling large image data sets and begins the analysis of the data set. The primary programming tool used in this evaluation is python and the code was written in Jupyter notebook.

The goals and hypothesis of this project are as follows:

- Frequency filtration of data can reduce memory usage of code the code by an order of magnitude. This has been proven and results shown in the presentation.
- Jets of varying pressure (ie varying Reynolds number) share some fundamental spectral content and hence can be correlated.
- The spectral content in a turbulent Jet varies by pressure with an increase in high frequency spectral content with increased pressure.

Experimental Method

Data for this analysis was collected in the UTSA hypersonic lab utilizing a z-type schlieren setup capturing flow emanating from an air duster at two different pressures and recorded using a Photron Fastcam SAZ 2100 high speed camera. The PNG file format was finally settled upon as the format of choice so as to reduce the amount of memory needed to store the image data files. The data for the project is currently hosted in a google drive and the link to these can be found as follows:
([https://drive.google.com/drive/folders/1nu2r1wRfx2k63iDBfLGYdC_K0D3ZFQkS?usp=sharing)](https://drive.google.com/drive/folders/1nu2r1wRfx2k63iDBfLGYdC_K0D3ZFQkS?usp=sharing)
A detailed breakdown of data set and sample images from this data set can be found in **Figure 1** and **Table 1**.

Table 1. Detailed Information about the data set.

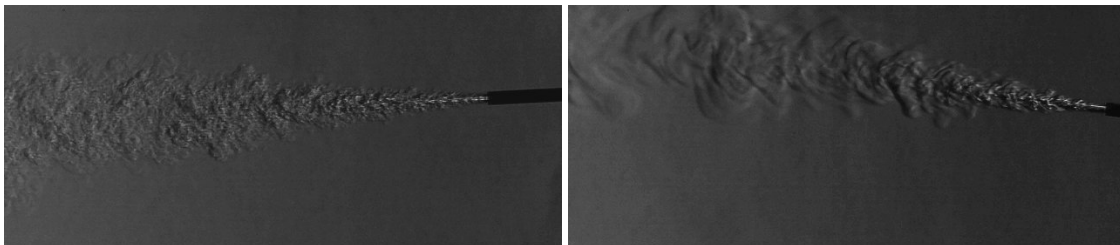| Image Set	Pixel size (rows x columns)	Image Count	Data Set Size
Test 1 background	680 x 280	200	14 MB
Test 1 images	680 x 280	100,000	7.62 GB
Test 2 background	680 x 280	2,000	343 MB
Test 2 (a & b) images	680 x 280	150,000	~11 GB

Figure 1. Sample Images from Data set

Results

a. Code

As stated above, this analysis was performed in python using the Jupyter notebook live text editor. To begin, there were several python libraries employed to perform tasks needed. A list of those used for this code can be seen below alongside their intended purposes.

i.   Numpy – This package was used for most of the numerical computations as well as array manipulations.

ii.   Pandas – This package was used for correlations and simple matrix math

iii.   Matplotlib.pyplot – This was used for plotting various pieces of data.

iv.   PIL – This was used for reading images and some image manipulations.

v.   Time – This was used to time the code to ensure performance.

vi.   Scipy-sparce – This was used to create sparse arrays for data handling purposed.

<u>Image Quality Analysis</u>

The code started off by calling in the required packages and creating libraries and dictionaries for frequently called upon operations. Once these tasks had been completed, a quality check was performed on the images to ensure the fidelity of the data. For image datasets acquired with high speed cameras, one of the major drawbacks to data fidelity is high frequency noise which can be visualized on the image as a host of random speckles dispersed all over the image. This distortion in the image can come about form image saving, transferring or imperfections in the viewing element of the high-speed camera. A hypothetical speckle image shown in **Figure 2** (b) alongside the original image (a) and representation of the norm magnitudes (d) and (c), respectively, are presented.

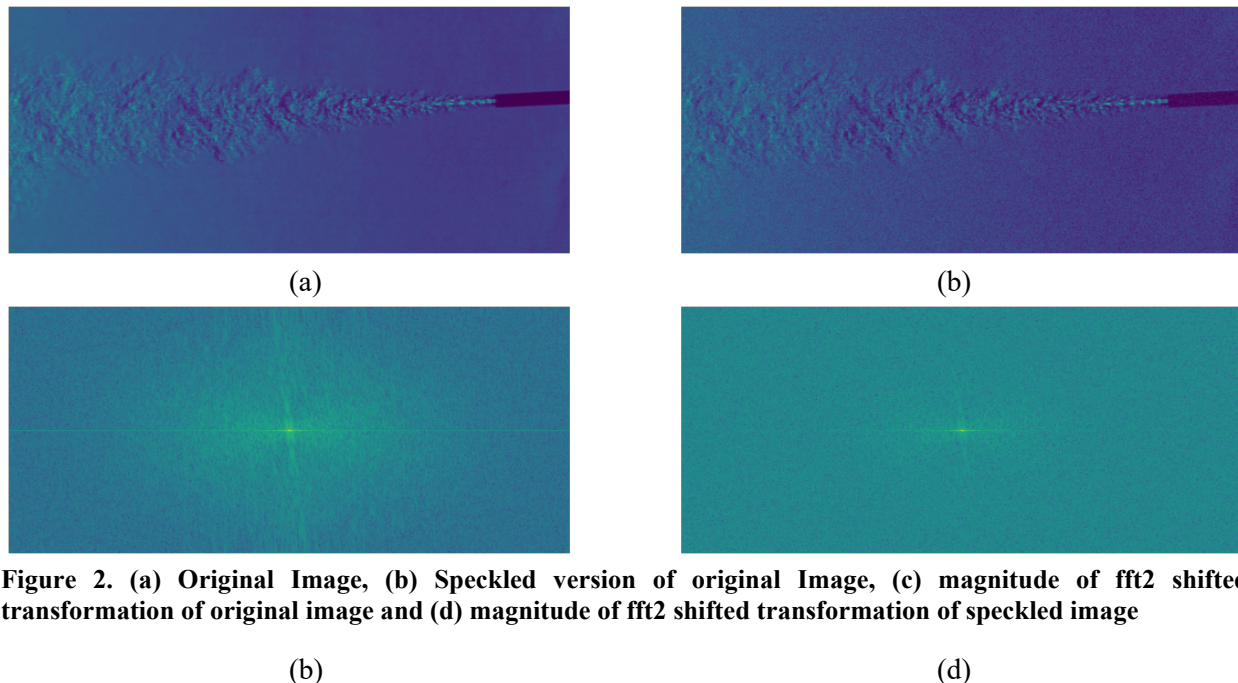(a)   (b)

Figure 2. (a) Original Image, (b) Speckled version of original Image, (c) magnitude of fft2 shifted transformation of original image and (d) magnitude of fft2 shifted transformation of speckled image

(b)   (d)

To check for this, a 2D Fast Fourier transform (fft) was applied to each image and the norm of the image 2D fft calculated and compared to that of an acceptable image. An acceptable standard deviation from the norm was determined and images outside these bounds would be taken out. All images in the data set passed this initial quality check.

<u>Data reduction.</u>

Once the images had passed the quality check, a precisely executed data reduction algorithm written by the author was performed to shrink the data down to manageable sizes. In order to optimize the computer memory usage, the images were imported and analyzed in 5 equally distributed batches (ie 20, 000 images analyzed at a time in a 100,000 image dataset.) The steps involved in this data reduction are listed below

1. Import total images in the batch, reshape, and create image matrix. Each column representing an image in time and each row spanning the set time duration.
2. Calculate fluctuating properties by subtracting mean image from instantaneous image to create a new image matrix set of fluctuating properties.
3. Perform a row-wise fft to convert into frequency domain.
4. Filter data set based on selected frequencies.
5. Convert frequency filtered data set to sparse Matrix for storage.

A sparse array stores data in three separate matrices with these consisting of the data values, and location information necessary to reconstruct the original matrix. As such there is a minimum image count necessary to make this option viable. A constant known as the sparsity of a matrix can be calculated to ensure this parameter is being applied properly. This data reduction technique performed very satisfactorily and was able to shrink data usage by up to about 75% at the highest image count (100, 000 images). For a low fidelity capturing of data utilizing int8 data types with the amplitudes of the first 500 selected frequencies, this method was able to shrink the data matrix down to 4.5 Gb of Ram as compared to 17.9 Gb of Ram for the entire data set at int8 integers. A higher fidelity sparce matrix made from float 32 data types captured the first 50 frequencies in a sparse matrix of size 1.4 Gb, as compared to a 8 bit integer matrix of 8.96 Gb. The latter higher fidelity sparse matrix was selected as the option of choice due to the great representation and minimal data loss when converted back to the image domain. Images captured from various points along the data reduction phase are presented in **Figure 3** for viewer considerations. A plot of data after being compressed with the sparce arrays is presented as well in **Figure 4** . Results presented above speak to the first goal/ hypothesis of this project and meet it satisfactorily.
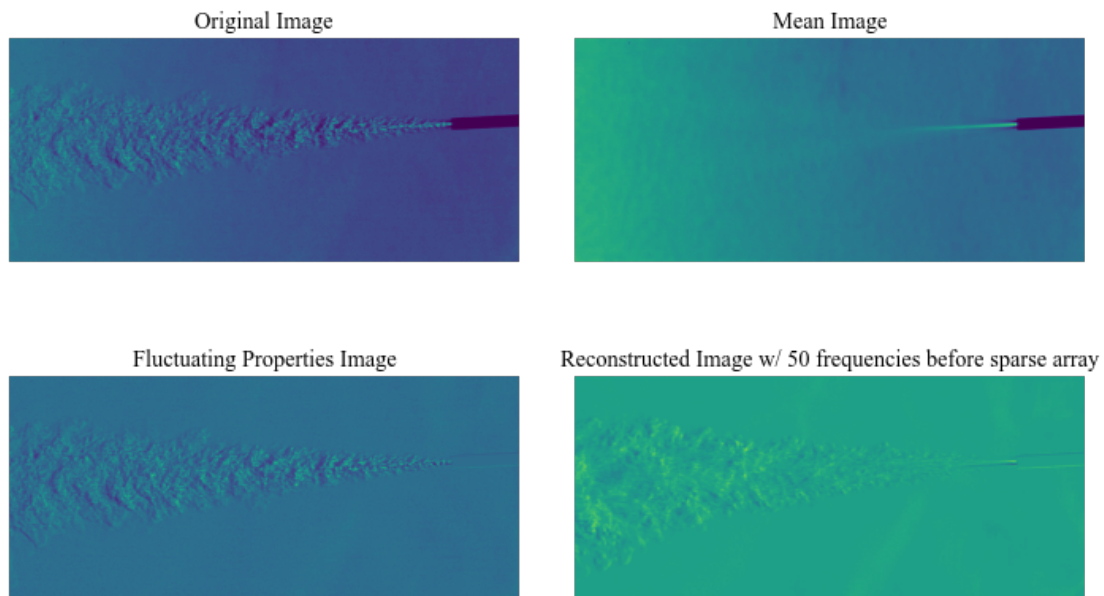

Figure 3. Image reconstruction form various points in the data reduction phase.

Reconstructed Image w/ 50 frequencies after sparse array

Figure 4. Image plot after extraction from sparce arrays.

b. Frequency analysis of datasets

Frequency analysis of points in flow

Once fully loaded, a frequency analysis of various points of both turbulent flows was performed and results from each presented in **Figure 5** and **Figure 6**. In both scenarios, the last point (Point 5) was chosen to serve as a reference and as expected there was no distinguishable frequency in that section. An "atmospheric" (Point 4) reading was selected as well plotted as well for comparison to the other fields. In this atmospheric probe, the frequency plot showed a low energy broadband signal which was consistent between the two plots. The high velocity turbulent flow (Test 1 in **Figure 6**) maintained higher amplitudes of energy at a further distance from the jet nozzle as compared to the low velocity jet (Test 2 in **Figure 5**). As expected, the high frequency content is the first to minimize in the frequency spectrum as seen in Test 2, Point 3 in **Figure 5** and Test 1, Point 1 in **Figure 6** when dissipation begins.

A cross correlation of the selected points was preformed to see whether any of the points shared any similar patterns. The result showing the correlation coefficients scaled in a range from -1 to 1 can be seen in the **Table 1**. From this table it can be witnessed that there is no strong correlation between any of the points. The strongest correlation is between Test 2, Point 3 in **Figure 5** and Test 1, Point 1 in **Figure 6** which describes similar levels of flow dispersion and the least correlated being between both background points.
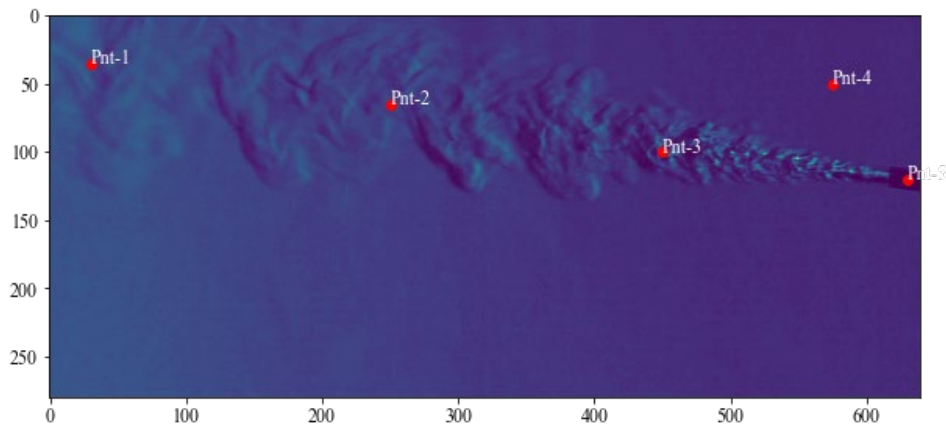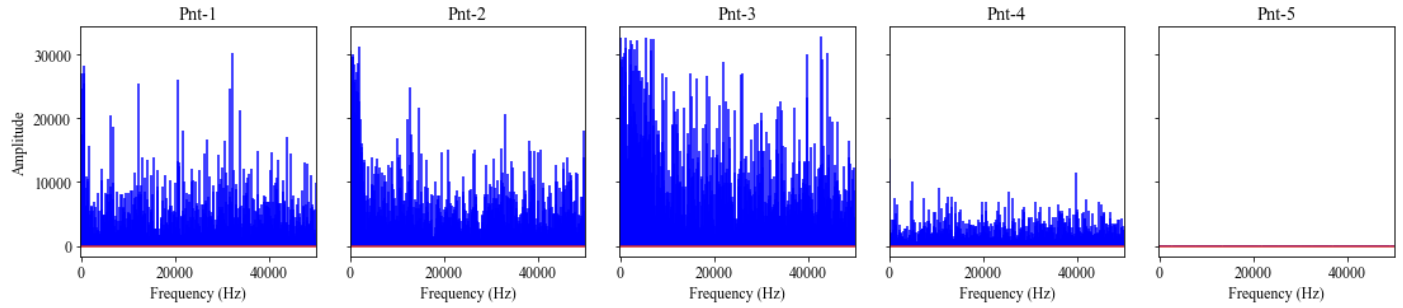
Figure 5. Frequency plot at various points in the low-velocity turbulent flow (Test 2).
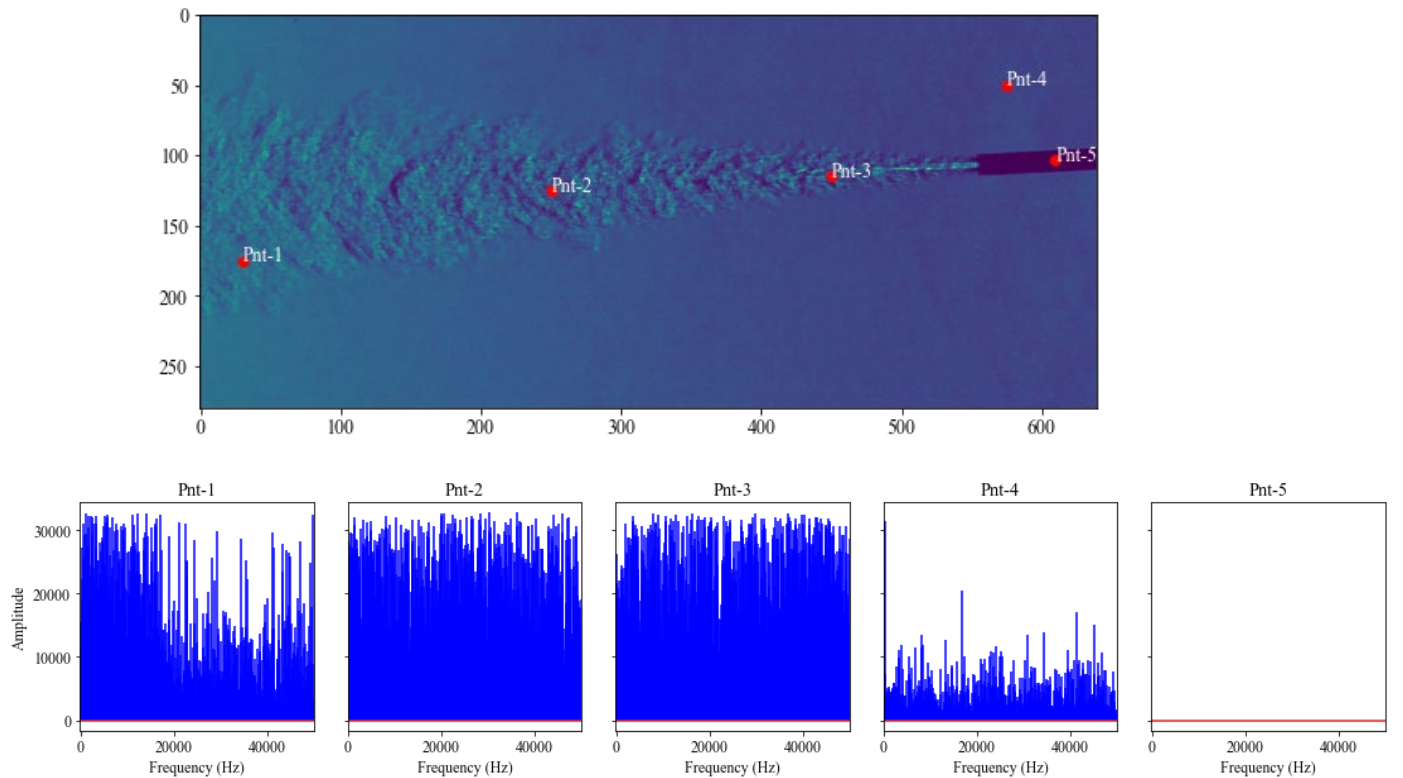

Figure 6. Frequency plot at various points in the low- velocity turbulent flow (Test 1).

Table 2. Cross correlation coefficients comparing signals from the two tests.

	Test_1-Pt1	Test_1-Pt2	Test_1-Pt3	Test_1-Pt4
Test_2-Pt1	0.034854	0.022765	0.019654	0.021102
Test_2-Pt2	0.128545	0.039783	-0.027614	0.111134
Test_2-Pt3	0.194619	0.023648	0.013480	0.080446
Test_2-Pt4	-0.038751	-0.020953	-0.039809	0.000972

Frequency Mode representations

      Unlike the previous analysis which delved into plotting all the frequencies for a single point to get an understanding of the spectrum, this section takes a different perspective on the data. The analysis in this portion involved shrinking the Fourier coefficient matrix of the data set down to a single frequency, calculating the inverse FFT, and plotting a representative image from all the pixels at the particular

frequency. A representation is shown in **Figure 7**. The chosen frequencies being analyzed include 50Hz, 500 Hz, and 2500 Hz. Across both test data, the larger flow structures were witnessed to occur at the lower frequencies. These larger structures occurred further downstream from the nozzle outlet. As the frequencies increased, the structures in the flow shrunk in physical size and were located closer to the nozzle outlet. A conclusion that could be potentially drawn from this is that flow emanating form the jet has the highest frequency and begins to slow down as it disperses and losses energy further away from the nozzle outlet.
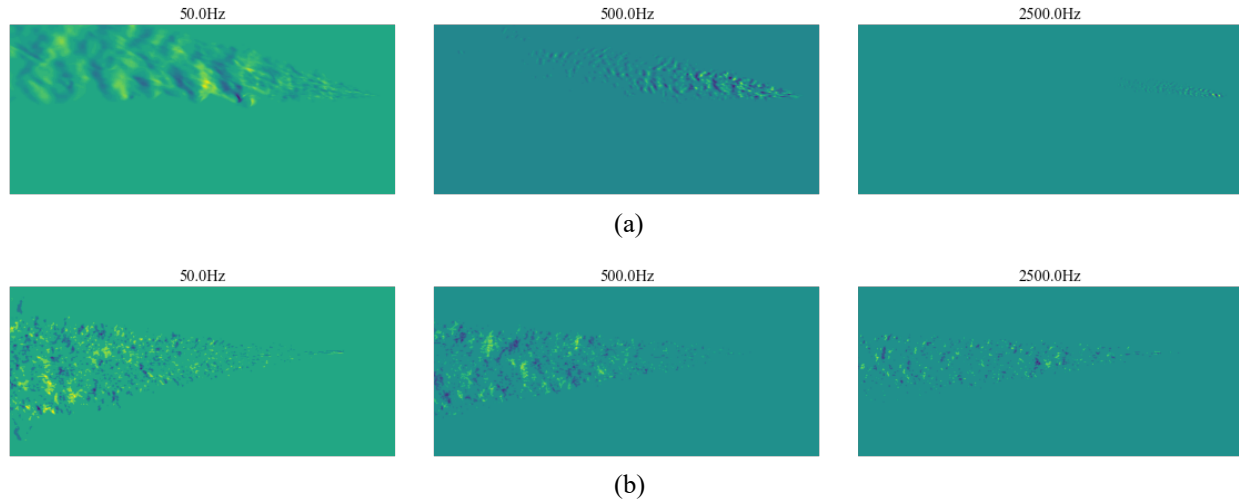

(a)

(b)

Figure 7 Frequency modes of the two tests (a) Test 2 modes (b) Test 1 modes

Conclusion & Future Work

A novel python code for performing data reduction has be created and implemented successfully with turbulent jet schlieren image data. Image data information was able to be reduced by at most 75% of the original image matrix data size requirement. This technique shows promise and can be improved and more refined and possibly applied in a host of other areas. Some preliminary frequency analysis of the turbulent data was performed, and trends discussed based on observations. Both the high velocity and low velocity turbulent jets shared similar types of distribution in frequency at various points. A correlation of the various points being analyzed yielded inconclusive results and would need further studies to be understood. Finally, frequency modal analysis was performed to gain insight into the physical structures associated with the certain frequencies. It was witnessed that the lower frequencies were associated with larger structures furthers away from the jet nozzle while the higher frequencies captures the faster moving flow directly out of the nozzle with smaller physical structures.

There are several areas of improvement to the current study and work to yield more precise data reduction and more in-depth frequency analysis. These include the use of the wavelet transform instead of the FFT transform to convert data more efficiently to the frequency spectrum for data reduction and analysis. A more in-depth look at the correlations between both data sets will help distinguish patterns between the two data sets. Finally, correlations between images at similar frequencies can be used to obtain distance traveled and from there low-level velocimetry readings.