Interpreting Deep Neural Networks Trained for Estimating Sinusoid Frequencies

Bofei Zhang
Center for Data Science
New York University
bz1030@nyu.edu

Yuanxi Sun
Center for Data Science
New York University
ys1879@nyu.edu

Abstract

In this project, we aim at understanding the behavior of a deep neural network trained for the line-spectra super-resolution task. We implement a bias-free version of DeepFreq to enable the analysis by linear-algebraic tools. First, we show that the bias-free network has a comparable performance compared to the original version of DeepFreq. Second, we analyze the Jacobian of this neural network to understand the mechanism of how this network detects the target frequency and denoises signal. Our analysis suggests neural network relies on Periodogram-like transformation to get frequency representation. Moreover, the network projects noisy signal into a low-dimensional subspace that preserves the signal structure but denoising by average around the target area[1].

1 Introduction

In this project, we focus on the interpretation of a deep neural network that is trained for estimating sinusoidal frequencies. The spectral super-resolution problem is one of the fundamental problems in signal processing. It has wide applications in areas like geographical imaging, spectroscopy, imaging radar, and underwater acoustics. The goal of this problem is to estimate the frequency of each component from a multi-sinusoidal signal with some noise. Therefore, after acquiring a finite number of samples from a noisy multi-sinusoidal signal, we need to reconstruct a frequency representation out of the samples, and make sure that the frequencies we construct matches closely with the original true frequencies.

While the traditional methods for spectral super-resolution problem like Periodogram and MUSIC [4] focus on applying linear-algebraic tools and analysis to do the reconstruction, in recent years, [2] starts to use deep learning-based models to solve the same problem. In this paper, the author introduced a Bias-Free Convolution Neural Network that is designed to solve the line spectral super-resolution problem and has achieved a significant improvement.

As most of the neural network-based models typically have non-linearity, which makes the traditional linear algebra-based analysis tools invalid. Therefore, the author of [2] intentionally removes the bias terms of the model, to make the model locally linear. In this project, we focus on learning the properties of this network by analyzing its Jacobian matrix. Our analysis shows that, each row of the Jacobian matrix has a clear relationship with the frequency that it is aimed to represent, which is similar to a Periodogram transformation. Moreover, by applying principal component analysis on each row of the Jacobian, we also find a relationship between the sparsity of the singular values and the noise level. We have shown that, the model performs a projection onto a lower-dimensional space that captures the frequency property, but denoise by averaging in local areas.

[1] Code and all analysis available here https://github.com/bofei5675/DeepFreq

2 Related Work

In signal processing, Periodogram is a traditional method to estimate the spectral representation of a signal. Basically, the Periodogram will generate a superposition of kernels centered at each true position. The Periodogram works similar to Fast Fourier Transform. However, the Periodogram cannot recover the true signal, even when there is no noise in the data. Another method to reconstruct the signal is the **MU**ltiple **SI**gnal **C**lassification(MUSIC)[4] model. The MUSIC model approximates the transformation between the incident waveform to the measured waveform as a linear function, and thus uses the eigendecomposition of the transformation to study the number of incident signals and intensity of signal noise.

Recently, Convolutional Neural Network (CNN) was very successful in the signal processing field. Izacard et al. [1] proposed novel and deep learning-based method called PSNet to address the line-spectra super-resolution task, which shows promising progress compared to the traditional methods such as Periodogram and MUSIC. Second, Izacard et al. [2] proposed another CNN architectures called DeepFreq that consists of two modules, which can first learn a frequency representation of the noisy signal, and then estimate the number of frequencies by searching for the peaks from the representation. This work established the state-of-the-art for the spectra super-resolution task.

However, it is very difficult to analyze and interpret CNN due to the deep architecture and non-linearity. To interpret the deep neural network, Mohan et al. [3] the bias-free neural network with Recitified Linear Unit (ReLU) is locally linear. This will enable analysis by linear-algebraic tools by computing the Jacobian matrix of the network. The bias-free network can achieve better generalization on the image denoising task. The analysis of Jacobian indicated that the model implicitly performs a projection onto an adaptively-selected low-dimensional subspace capturing features of natural images[3].

3 Methods

3.1 Bias-Free CNN

To enable linear algebraic analysis, we utilized ReLU non-linearity and remove bias of all linear, convolution, and batch normalization layer in DeepFreq[2]. It's very straightforward about how to remove bias from convolution and linear layer. For batch normalization layer, we change the update equation of 1D batch normalization into

$$\hat{x}_i = \frac{x_i}{\sqrt{\sigma_b^2 + \epsilon}} \qquad\qquad y_i = \beta \hat{x}_i \tag{1}$$

where x_i, y_i is the input and output of batch normalization layer respectively. We disable the mini-batch mean and only keep the mini-batch variance σ_b^2. We only keep the trainable parameter β and remove the bias term. The resulting model is bias free, and we can verify it by checking $f(\alpha x) = \alpha f(x)$ and $f(0) = 0$. We compared the performance in False Negative Rate (FNR) with DeepFreq with bias to see if there is any gain in generalization capability with removal of bias.

3.2 Jacobian Computation

Consider a input signal with 50 data points, $x_i \in C^{50}$, we add an independent Gaussian noise $\tilde{z}_i \in C^{50}$ by $\tilde{x}_i = x_i + \tilde{z}_i$. We follow the exactly same sampling schema that Izacard et al. [2] suggested. We obtain the frequency representation via the bias-free network $y_i = f_{BF}(\tilde{x}_i)$, where $y_i \in R^{1000}$. Then, the Jacobian and final output are obtained by

$$A[j,k] = \frac{\partial y_i[k]}{\partial x_i[j]} \tag{2}$$

$$y_i = A^* \tilde{x}_i \tag{3}$$

Where $A \in C^{50\times 1000}$ and A^* is the conjugate transpose of A. We implemented equation 2 by computing backpropagation of an element of output with respect to an element of input. Based on formulation of equation 3, the magnitude of frequency representation at frequency k can be computed by the inner product of k-th row of A^* and input signal $\tilde{x}_i$. Therefore, we study the row of Jacobian to explore how neural network detect frequency components.

3.3 Jacobian Analysis

We follow the idea of effective dimensionality introduced by Mohan et al. [3] to study projection into signal subspace of a noisy signal. Specifically, for a realization of Jacobian matrix, we can do a Singular Value Decomposition (SVD) such that $A^* = USV^T$. The effective dimensionality is given by sum of squared eigenvalues $d = \sum_i^N s_i^2$.

Next, we compute multiple realizations of Jacobian matrices for a fixed Signal-to-Noise Ratio (SNR), then we perform Principle Component Analysis (PCA) on a row of Jacobian matrices. Let $A_i^*[j]$ denotes j-th row of a Jacobian matrix realization, we obtains unbiased estimation of covariance matrices $\Sigma_{A^*} = \frac{1}{N-1}\sum_{i=1}^N A_i^*[j]A_i^*[j]^T$. This equation only holds when A^* has zero-mean for each row so that we verified this assumption numerically. PCA reveals principle components of a certain row in the time domain by decomposing Σ_{A^*} into $U\Lambda U^T$, where Λ contains eigenvalues λ_i as diagonal entries. We convert the principle component $u_i \in U$ via Fast Fourier Transform (FFT) to analyze the frequency components. Inversely, we can approximate the row by $A^*[j] = \sum_i \sqrt{\lambda_i} u_i$. We refer to each row of Jacobian as a filter. We aim to draw a connection between these filters with classical signal processing method Periodogram [5] shown by

$$\hat{\phi}(w) = \frac{1}{N} | \sum_{t=0}^{N-1} y(t) e^{-iwt} |^2 \tag{4}$$

Where $\hat{\phi}(w)$ is magnitude of frequency representation at frequency w. Stoica et al. [5] introduced a Filter Bank Approach (FBA) to interpret Periodogram. This concept also offers a novel way to understand the behavior of a neural network. In our proposal, we try to draw the connection between MUSIC and the Jacobian, but later we found Periodogram shows similar behavior pattern.

4 Results

4.1 Generalization of Bias-Free DeepFreq

In Figure 1 , we show the FNR of different models trained in a different range of noise level (SNR). We found for high noise level like SNR 1 or 10, the model with or without bias generalize well on the test data with noise level they never see. The bias-free version of DeepFreq has very marginal improvement based on left and middle plots in Figure [5]. However, the experiment results show DeepFreq without bias does not perform as well as the original implementation when trained on the entire range of SNR (1-50).

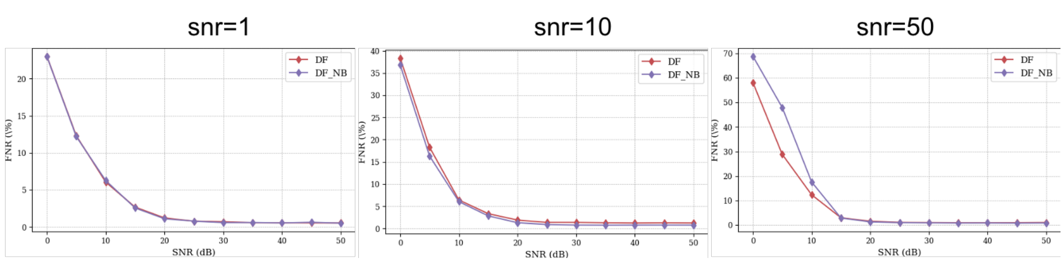

Figure 1: FNR of DeepFreq with or without bias trained in different ranges of SNR. The number of the title indicates training data includes a noisy signal from 1 SNR to the number shown by title. "NB" in the legend indicates "No Bias". Title for each subplot indicates the smallest noise level in training data.

4.2 Jacobian and Effective Dimensionality

Figure 2 shows one example of model outputs (Top) and computed Jacobian (Middle). The bottom figure is obtained by $y_i - A^*x$, which verifies Equation 3. In practice, y_i computed Jacobian may have an imaginary part very close to 0 and we chose to discard the imaginary part. Based on the magnitude of elements in the Jacobian matrix, we found if the network believes there is a frequency component, then the corresponding row contains a vector with the large norm. We experimented with a couple of realization and found that each row of Jacobian can be viewed as a sinusoidal wave with certain frequency components. Numerically, we also verified each row has a mean that is very closed to 0.

Next, we followed the method introduced by Izacard et al.[1] to compute effective dimensionality. We compute 10 realizations of clean signal and noisy signal with SNR 1, and perform SVD on obtained Jacobian matrices. The left plot of Figure 3 indicates that most of the singular values are very small for a clean signal. For noisy signals, the middle plot of Figure 3 suggests a signal subspace with high dimensions. We also observed increasing effective dimensionality based on the right plot of Figure 3, which is contradictory to what Izacard et al.[1] found in the image denoising task. This finding suggests the mechanism for signal denoising might be different from the image denoising.

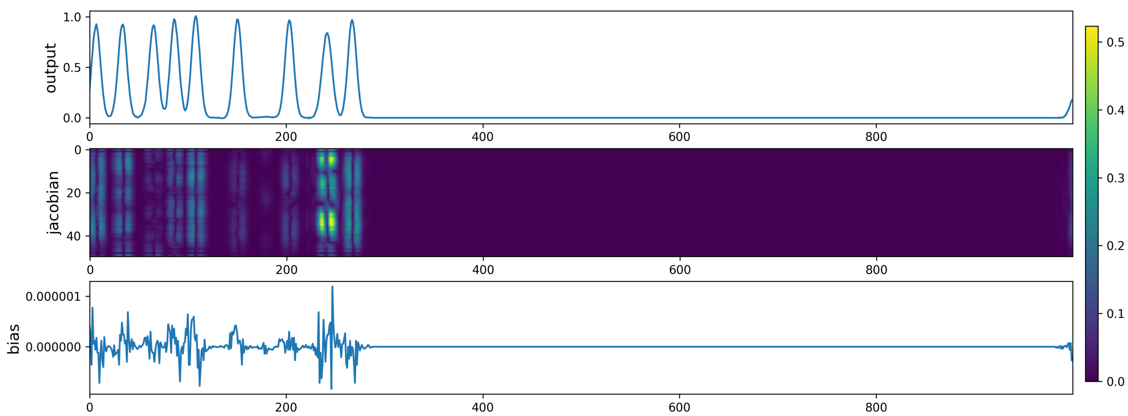

Figure 2: Top: Learned frequency representation from a noisy signal. Middle: Magnitude of elements in the Jacobian matrix. Bottom: Bias computed by subtracting model outputs from $A^* \tilde{x}_i$

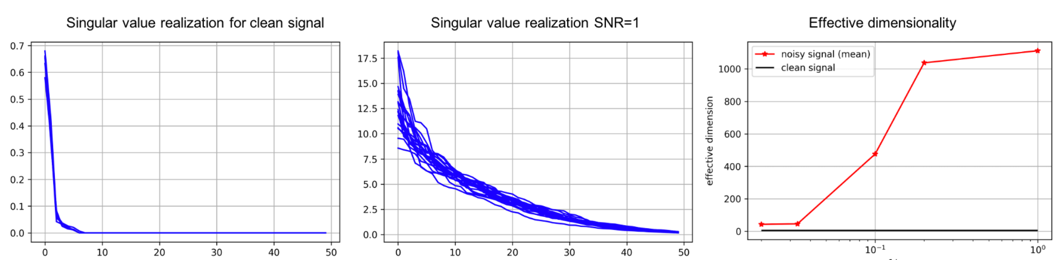

Figure 3: Left: Singular value vector from 10 realizations of Jacobian from a clean signal; Middle: Singular value vector from 10 realizations from a noisy signal with SNR=1; Right: Effective dimensionality vs. noise level.

4.3 DeepFreq and Periodogram

Based on the method we described in section 3.3, we perform PCA on the row of 20 Jacobian realizations. Figure 4 indicates that underlying space has a very low dimension such as only first couple eigenvalues are non-negligible. Additionally, FFT of principle components in Figure 4 indicates the first couple of eigenvectors usually are a sinusoidal wave with negative frequency to the corresponding frequency representation outputs. For the case shown in the bottom row of Figure 4, it's possible to see the filter for some frequencies to have only one non-negligible eigenvalue. To get the frequency representation at w, we can estimate Jacobian row $A^*[w] \approx \sqrt{\lambda_1} u_1$, where λ_1, u_1 is the largest eigenvalue and corresponding eigenvector. Based on the FFT results shown by Figure 4, u_1 is a signal with negative frequency $-w$. Then, we can let $u_1 = exp(-iwt)$. For a magnitude of frequency representation y_i at frequency w, the inner product between Jacobian's row and an input signal resembles the notation of Peroidogram shown in Equation 4

$$y_i[w] \approx \lambda_1 \left| \sum_{t=1}^{N} \tilde{x}_i(t) exp(-iwt) \right|^2 \tag{5}$$

Equation 5 indicates a Periodogram-like transformation accounts for the most part of an output frequency representation. Note in Figure 4, the frequency components that are not label have a small eigenvalue compared to label (0.4 Hz).

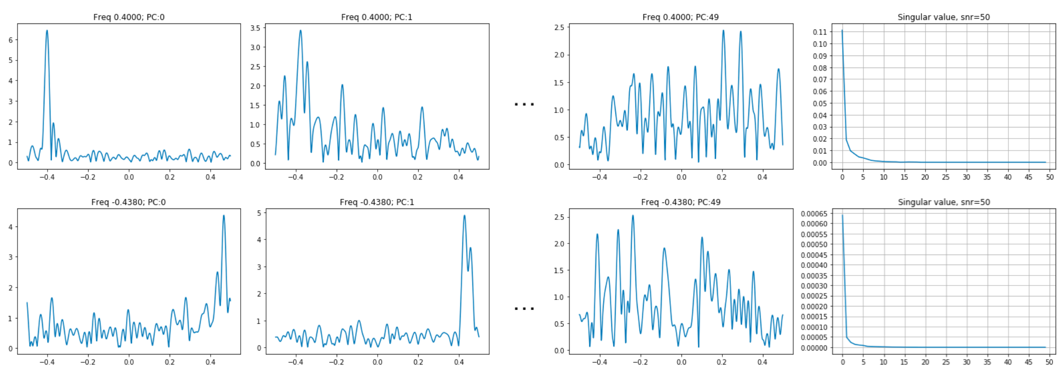

Figure 4: FFT on 50 principle components and singular values obtained by PCA for frequency 0.4 and -0.4380. We simulate this signal with 0.3 and 0.4 Hz as the ground truth frequencies. This signal has SNR 1.

Previous study [2] has shown DeepFreq has much better performance than Periodogram, therefore the network is not only doing a Periodogram. Based on Figure 4, the later principle components will have broader peaks around target frequency and multiple frequency components. Projecting into these basis vectors will average noise signal around target frequency, which leads to the denoising effect. Lamma 5.2.6 of Stoica et al. [5] indicates Periodogram is an FBA with a certain bandpass filter. Since DeepFreq suggests a similar behavior to Periodogram, we hypothesize that the neural network addresses the signal denoising task by learned a series of complicated bandpass filters. During inference, the filters firstly project into a subspace that preserves the signal structure by first principle components. Then, the projection into the rest of eigenvectors with non-negligible eigenvalues can be seen as a weighted average around the target frequency area. This weighted average removed noise, which is similar to the conclusion about image denoising task[3]. The question of how to formulate this filter remains open.

5 Discussion

In this project, we utilized bias-free DeepFreq to address the line-spectral super-resolution task. The model yields comparable performance compared to the previous results. It also has a marginal improvement in terms of generalization on the test data with a noise level that does not appear during training. Second, the bias-free network enables linear algebraic-based analysis. We found the rows of Jacobian are basically sinusoidal waves just like inputs. Analysis of SVD on Jacobian indicates contradictory results regarding effective dimensionality introduced in the image denoising field[3]. Finally, PCA reveals that the behavior of the neural network can be approximate by a Periodogram-like transformation. This result draws a connection between deep learning-based methods with classical signal processing methods.

However, the exact formulation for the filter learned by neural network remains mysterious. For the classical signal processing method Periodogram, the bandpass filter can be enhanced to gain more statistics accuracy and resolution. For examples, Capon method can learns a data-dependent bandpass filter [5]. It would be interesting to dig down the concept of the Capon method to see if we can draw a connection with DeepFreq.

References

[1] Gautier Izacard, Brett Bernstein, and Carlos Fernandez-Granda. A learning-based framework for line-spectra super-resolution.

[2] Gautier Izacard, Sreyas Mohan, and Carlos Fernandez-Granda. Data-driven estimation of sinusoid frequencies.

[3] Sreyas Mohan, Zahra Kadkhodaie, Eero P. Simoncelli, and Carlos Fernandez-Granda. Robust and interpretable blind image denoising via bias-free convolutional neural networks.

[4] R. Schmidt. Multiple emitter location and signal parameter estimation. 34(3):276–280.

[5] Petre Stoica, Randolph L Moses, et al. Spectral analysis of signals. 2005.