Audio Compression Using Singular Value Decomposition

Michele Magrini

July 20, 2024

Contents

1 Introduction

In this project, we explore the use of Singular Value Decomposition (SVD) for audio compression. The process involves converting audio files into spectrograms using the Short-Time Fourier Transform (STFT), applying SVD to these spectrograms, truncating the singular values, and reconstructing the audio using the inverse STFT with the Griffin-Lim algorithm. This report details each step, including the mathematical foundations and practical implementations.

2 Short-Time Fourier Transform (STFT)

The Short-Time Fourier Transform (STFT) is a technique used to analyze the frequency content of signals that vary over time. It divides the signal into overlapping segments, applies the Fourier Transform to each segment, and produces a time-frequency representation of the signal.

2.1 Mathematical Formulation of STFT

Given a signal $x(t)$, the STFT is defined as:

$$\text{STFT}\{x(t)\}(m, \omega) = \sum_{n=-\infty}^{\infty} x[n] \cdot w[n-m] \cdot e^{-j\omega n} \tag{1}$$

Here:

- $x[n]$ is the discrete signal.
- $w[n-m]$ is a window function centered at time m.
- ω is the angular frequency.

The STFT provides a complex-valued matrix where each element represents the amplitude and phase of the signal at a specific time and frequency.

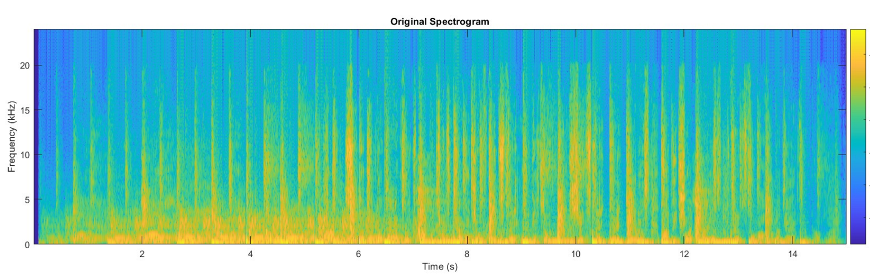

Figure 1: Spectrogram example

3 Singular Value Decomposition (SVD)

Singular Value Decomposition is a powerful mathematical technique used in linear algebra to factorize a matrix into three component matrices. For a given matrix A, the SVD is given by:

$$A = U\Sigma V^T \tag{2}$$

Where:

- U is an orthogonal matrix containing the left singular vectors.
- Σ is a diagonal matrix with singular values.
- V^T is the transpose of an orthogonal matrix containing the right singular vectors.

3.1 SVD for Compression

In the context of image-like data such as spectrograms, SVD is particularly effective for compression. The singular values in Σ represent the energy or information content of the matrix. By truncating the smaller singular values, we can approximate the matrix with lower rank, reducing the amount of data while preserving essential features.

3.2 Error in approximating a matrix with SVD (EYM Theorem)

Theorem (Eckart–Young–Mirsky). *Given a matrix $A \in \mathbb{R}^{m\times n}$, let $U\Sigma V^T$ be its singular value decomposition (SVD). Let $\widetilde{\Sigma}$ be the diagonal matrix obtained from Σ by setting all but the first ℓ largest singular values to zero. Then, the matrix $\widetilde{A} \equiv U\widetilde{\Sigma}V^T$ minimizes the error $\|A - \widetilde{A}\|$ subject to the constraint that the column space of $\widetilde{A}$ has at most dimension ℓ.*

3.3 Why SVD Works Well on Grid-Like Images

- **Energy Compaction:** Most of the energy (information) in an image is concentrated in the largest singular values. Truncating the smaller values results in minimal loss of important information.
- **Dimensionality Reduction:** SVD efficiently reduces the dimensionality of the data, which is ideal for compression.
- **Noise Reduction:** Truncating small singular values can also help in reducing noise and improving the quality of the reconstructed image.

4 Truncating Singular Values

To compress the spectrogram, we retain only the top k singular values, resulting in an approximation of the original spectrogram:

$$A_k = U_k \Sigma_k V_k^T \tag{3}$$

Here:

- U_k consists of the first k columns of U.
- Σ_k is a $k \times k$ diagonal matrix with the top k singular values.
- V_k consists of the first k columns of V.

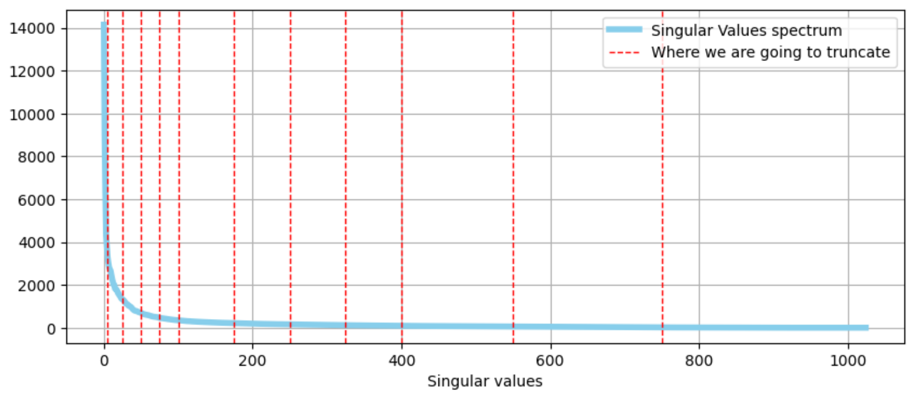

Figure 2: Plot of the singular values spectrum

5 Inverse STFT with Griffin-Lim Algorithm

After truncating the singular values and obtaining the compressed spectrogram, the next step is to reconstruct the audio signal. This involves applying the inverse STFT, which synthesizes the time-domain signal from its spectrogram.

5.1 Inverse STFT

The inverse STFT is given by:

$$x[n] = \frac{1}{W(n)} \sum_{m=-\infty}^{\infty} X(m, \omega) \cdot w[n-m] \cdot e^{j\omega n} \tag{4}$$

Where $X(m, \omega)$ is the STFT of the signal, $w[n-m]$ is the window function, and $W(n)$ is a normalization factor to correct for the windowing.

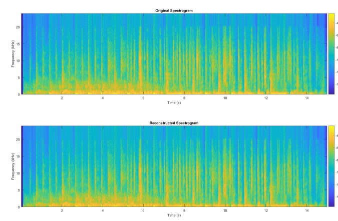

Figure 3: Comparing the compressed spectrogram

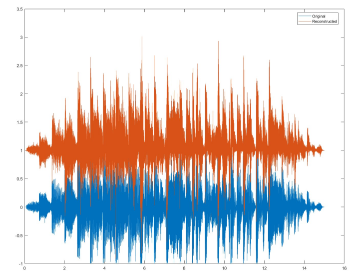

Figure 4: Comparing the compressed waveform

5.2 Griffin-Lim Algorithm

The Griffin-Lim algorithm is an iterative method used to improve the quality of the reconstructed audio signal from its spectrogram. It addresses the phase reconstruction problem, ensuring that the reconstructed signal's STFT closely matches the original spectrogram.

Algorithm 1 Griffin-Lim Algorithm

```
Initialize the phase of the spectrogram randomly or with an estimate.
while not converged do
    Compute the inverse STFT to get a time-domain signal.
    Recompute the STFT of the time-domain signal.
    Replace the magnitude of the STFT with the original spectrogram's
    magnitude while keeping the phase.
end while
```

The Griffin-Lim algorithm enhances the reconstruction process by iteratively refining the phase information, resulting in a more accurate and perceptually better audio signal.

6 Evaluation of Singular Value Truncation

The evaluation of the singular value components is crucial for understanding the trade-off between compression and audio quality. In the plot shown in Figure 5, we analyze the size of the singular value components compared to the original file size.

6.1 Analysis

From the Figure 5, it is evident that the original audio file size is approximately 28 MB. As singular values are added, the size of the retained components increases linearly.

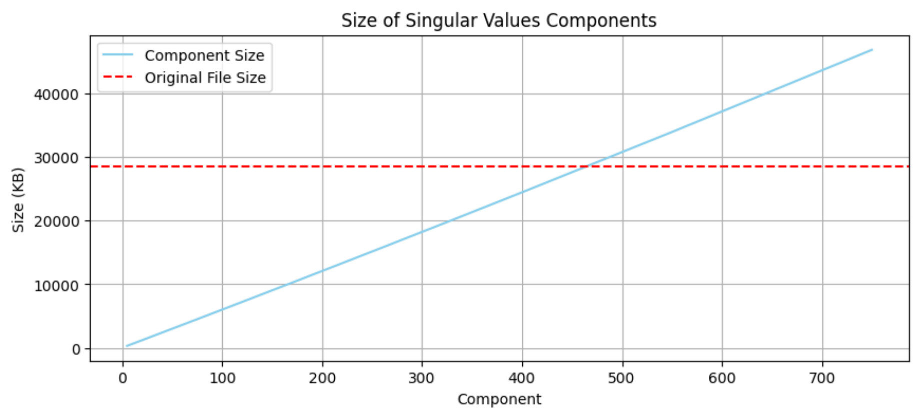

Figure 5: Size of compressed files truncated at the k-th component compared to the Original File Size

The experient also shows that retaining around one-quarter of the total number of singular values results in a substantial reduction in file size. Despite this reduction, audio quality remains nearly perfect when this fraction of singular values is retained, demonstrating the effectiveness of SVD in compressing audio data without significant loss in quality.

6.2 Optimal Compression Point

The evaluation suggests that retaining about one-quarter of the total number of singular values strikes a favorable balance between compression and audio quality. Beyond this point, the incremental increase in component size does not correspond to a proportionate improvement in audio quality. Therefore, for practical applications, retaining this fraction of singular values is recommended for achieving high-quality audio reconstruction with significant data size reduction.

6.3 General Considerations

While the compression efficiency diminishes beyond this point, it's important to consider specific use cases. For scenarios where maximum audio fidelity is required, retaining more singular values may be justified despite the larger file size. Conversely, for applications prioritizing data size over audio quality, fewer singular values may be retained.

The results demonstrate that SVD is a powerful tool for audio compression, offering a significant reduction in file size while maintaining excellent audio quality. This balance is essential for applications such as streaming, storage, and transmission of audio data.