Lecture 8
Time-Frequency Representations

Dennis Sun
Stats 253

July 16, 2014

Outline of Lecture

1. Short-Time Fourier Transform
2. Non-negative Matrix Factorization
3. Audio Source Separation
4. Wrapping Up

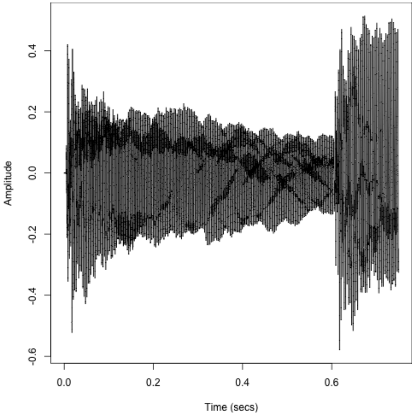
Where are we?

Stationary Processes

- We've seen the **Spectral Representation Theorem**, which says that all stationary processes are essentially just linear combinations of sinusoids.
- So: is a typical music recording stationary?

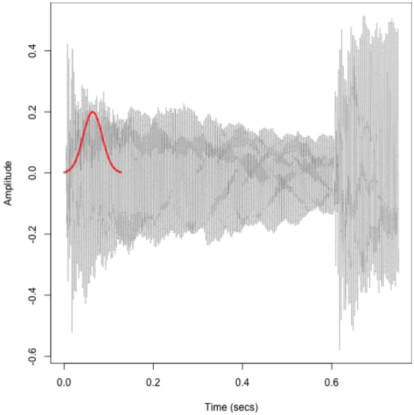
Short-Time Fourier Transform

- **Idea:** Audio is locally stationary.
- Take Fourier transform of only a chunk of a signal at a time.



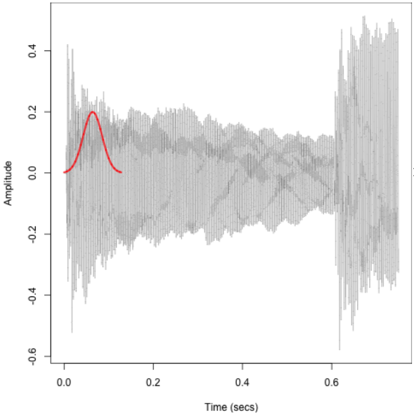
Short-Time Fourier Transform

- **Idea:** Audio is locally stationary.
- Take Fourier transform of only a chunk of a signal at a time.

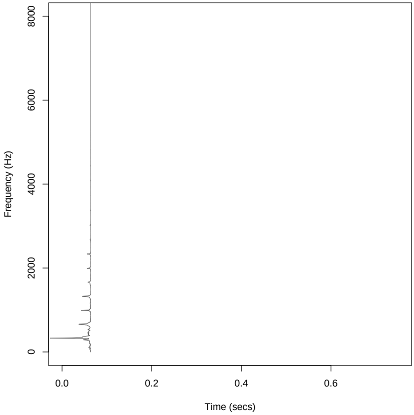


Short-Time Fourier Transform

- **Idea:** Audio is locally stationary.
- Take Fourier transform of only a chunk of a signal at a time.

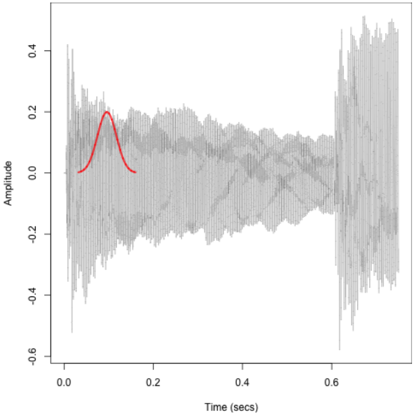


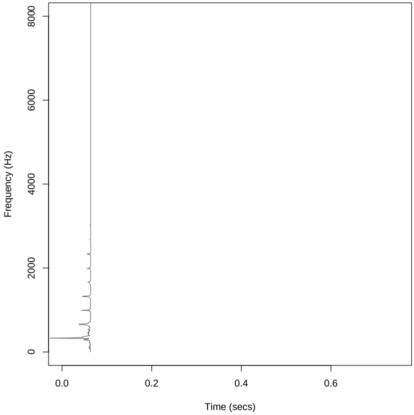
$\Longrightarrow$



Short-Time Fourier Transform

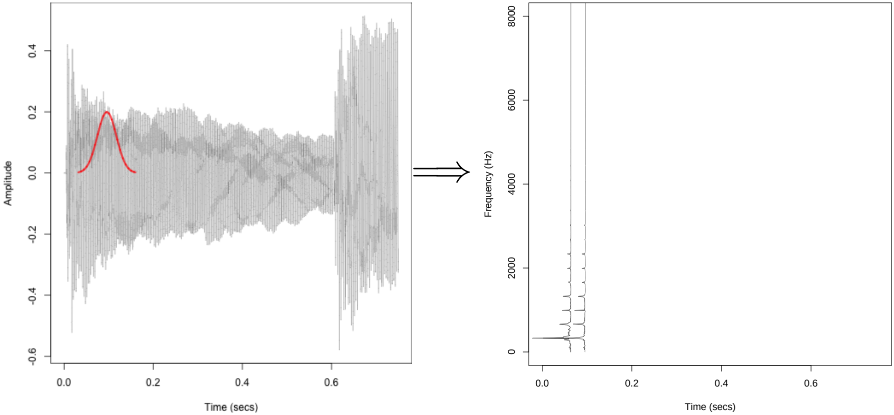
- **Idea:** Audio is locally stationary.
- Take Fourier transform of only a chunk of a signal at a time.





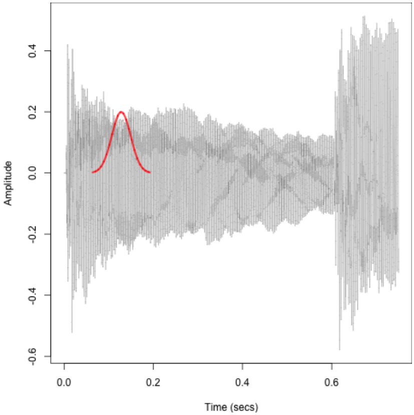
Short-Time Fourier Transform

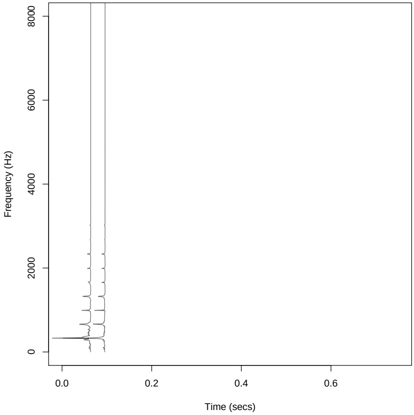
- **Idea:** Audio is locally stationary.
- Take Fourier transform of only a chunk of a signal at a time.



Short-Time Fourier Transform

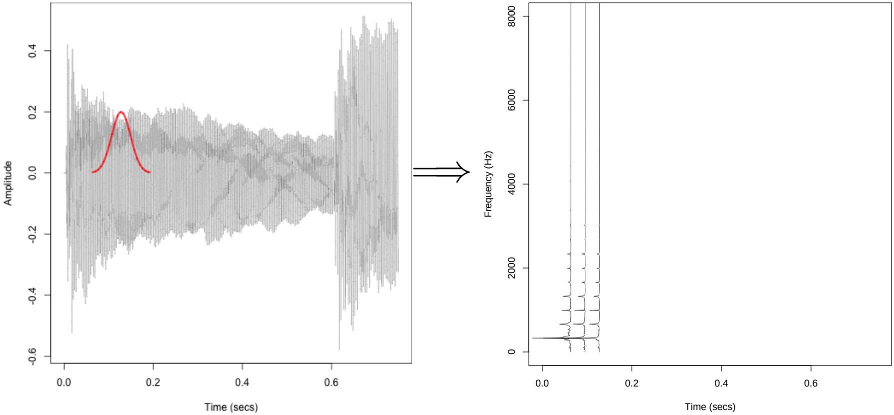
- **Idea:** Audio is locally stationary.
- Take Fourier transform of only a chunk of a signal at a time.





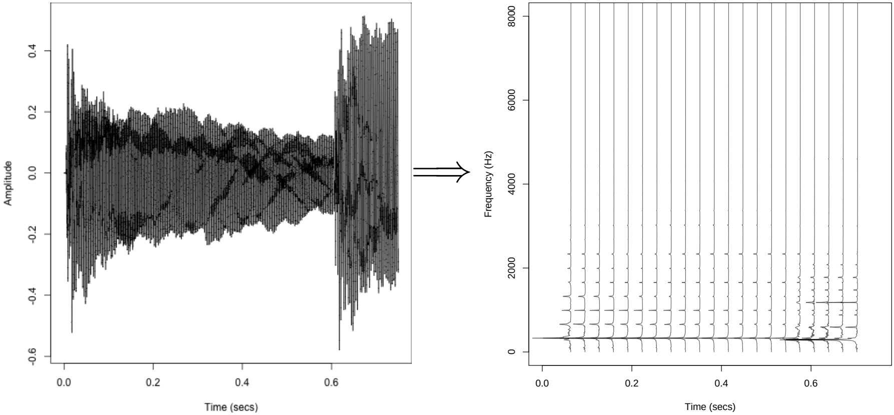
Short-Time Fourier Transform

- **Idea:** Audio is locally stationary.
- Take Fourier transform of only a chunk of a signal at a time.



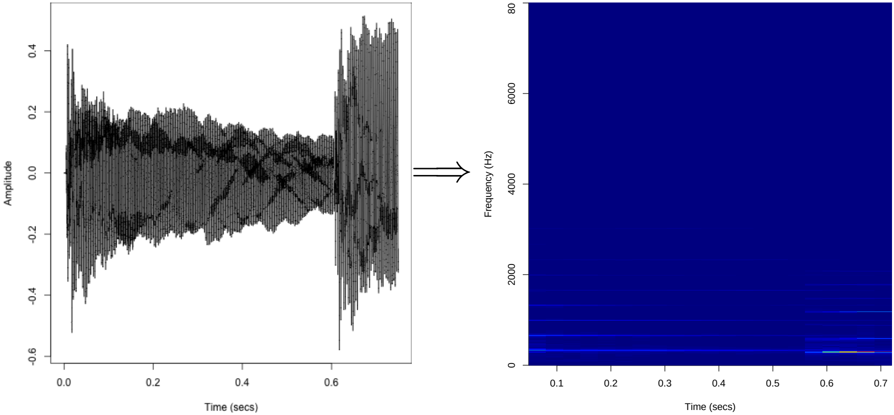
Short-Time Fourier Transform

- **Idea:** Audio is locally stationary.
- Take Fourier transform of only a chunk of a signal at a time.



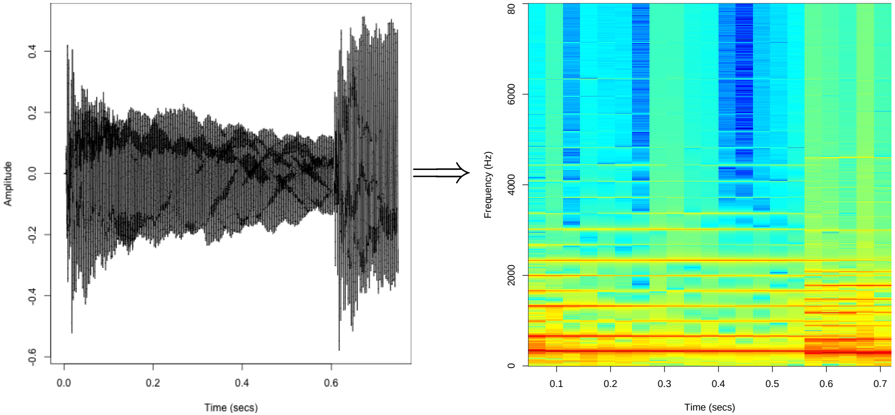
Short-Time Fourier Transform

- **Idea:** Audio is locally stationary.
- Take Fourier transform of only a chunk of a signal at a time.

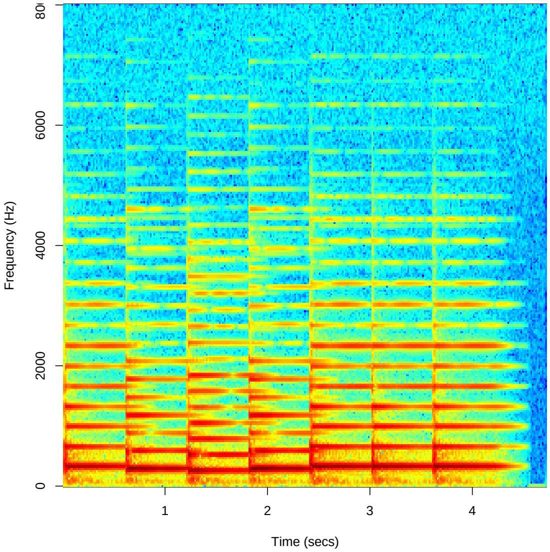


Short-Time Fourier Transform

Plot log-magnitudes (deciBels) to make contrast clearer.



STFT of the Audio File



play stop

Summary

- The **Short-Time Fourier Transform** (STFT) takes in an input signal, and computes local DFTs to obtain a matrix.
- A plot of the magnitudes of this complex-valued matrix is called a **spectrogram**. (Log-magnitudes are often plotted instead of magnitudes.)
- There is also an inverse STFT (ISTFT), which takes in a complex matrix, computes the IDFT of each column, and adds the signal piece by piece to recover the time-domain signal.

Where are we?

Matrix Factorization

In many applications, we wish to decompose a matrix X as a product of two matrices:

$$X = AB.$$

The most important example is **principal components analysis** (PCA).

$$\begin{bmatrix} — & \boldsymbol{x}_1 & — \\ & \vdots & \\ — & \boldsymbol{x}_n & — \end{bmatrix} \approx \begin{bmatrix} d_1 u_{11} & d_2 u_{12} \\ \vdots & \vdots \\ d_1 u_{n1} & d_2 u_{n2} \end{bmatrix} \begin{bmatrix} — & \boldsymbol{v}_1^T & — \\ — & \boldsymbol{v}_2^T & — \end{bmatrix}$$

Non-Negative Matrix Factorization

$$\underbrace{\begin{bmatrix} \quad V \quad \end{bmatrix}}_{\text{Data}} \approx \underbrace{\begin{bmatrix} W \end{bmatrix}}_{\text{Basis Vectors}} \underbrace{\begin{bmatrix} \quad H \quad \end{bmatrix}}_{\text{Weights}}$$

- A matrix factorization where everything is non-negative
- $V \in \mathbb{R}_+^{F \times T}$ - original non-negative data
- $W \in \mathbb{R}_+^{F \times K}$ - matrix of basis vectors, dictionary elements
- $H \in \mathbb{R}_+^{K \times T}$ - matrix of activations, weights, or gains
- Typically, $K \ll F, T$
  - A compressed representation of the data
  - A low-rank approximation to V

NMF With Spectrogram Data

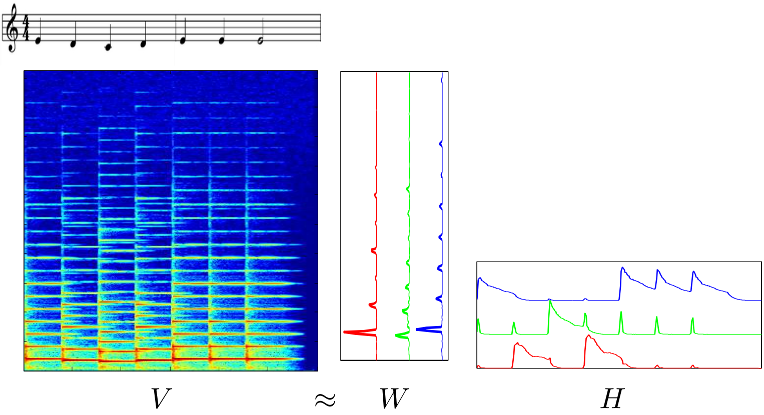
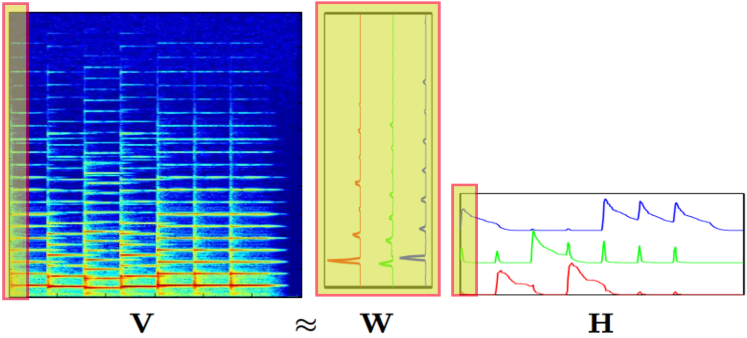


Figure : NMF of Mary Had a Little Lamb with $K = 3$ play stop

Factorization Interpretation I

Each column of V is a weighted sum of the columns of W.

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_T \\ | & | & & | \end{bmatrix} \approx \begin{bmatrix} \sum_{j=1}^{K} H_{j1}\boldsymbol{w}_j & \sum_{j=1}^{K} H_{j2}\boldsymbol{w}_j & \dots & \sum_{j=1}^{K} H_{jT}\boldsymbol{w}_j \end{bmatrix}$$

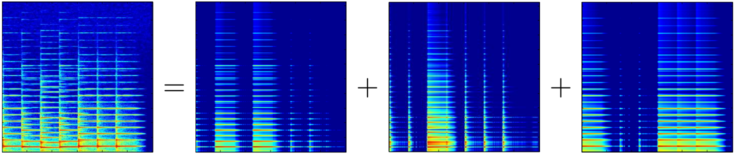


Factorization Interpretation II

V is a sum of rank-1 matrices.

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_T \\ | & | & & | \end{bmatrix} \approx \begin{bmatrix} | & | & & | \\ \boldsymbol{w}_1 & \boldsymbol{w}_2 & \ldots & \boldsymbol{w}_K \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & \boldsymbol{h}_1^T & — \\ — & \boldsymbol{h}_2^T & — \\ & \vdots & \\ — & \boldsymbol{h}_K^T & — \end{bmatrix}$$

$$V \approx \boldsymbol{w}_1\boldsymbol{h}_1^T + \boldsymbol{w}_2\boldsymbol{h}_2^T + \ldots + \boldsymbol{w}_K\boldsymbol{h}_K^T$$



The NMF Cost Function

- So far we've been vague by saying $V \approx WH$.
- Formally, NMF solves the optimization problem

$$\underset{W,H}{\text{minimize}}\ D(V|WH) \ \text{ subject to } W, H \geq 0$$

  for some choice of "distance" measure D.
- Choices of D:
  - Euclidean distance: $D(V|\hat{V}) = \sum_{i,j} (V_{ij} - \hat{V}_{ij})^2$
  - Kullback-Leibler (KL) divergence:
    $$D(V|\hat{V}) = \sum_{i,j} V_{ij} \log \frac{V_{ij}}{\hat{V}_{ij}} - V_{ij} + \hat{V}_{ij}$$
- KL divergence is more appropriate for audio.

Algorithms for NMF

- **General Strategy:** Fix W and update H. Fix H and update W. Iterate until convergence.
- The two problems are symmetric. So let's look at fixing W and updating H (for KL divergence).

$$D(V|WH) = -\sum_{i,j} V_{ij} \log \sum_k W_{ik} H_{kj} + \sum_{i,j} \sum_k W_{ik} H_{kj} + \text{const.}$$

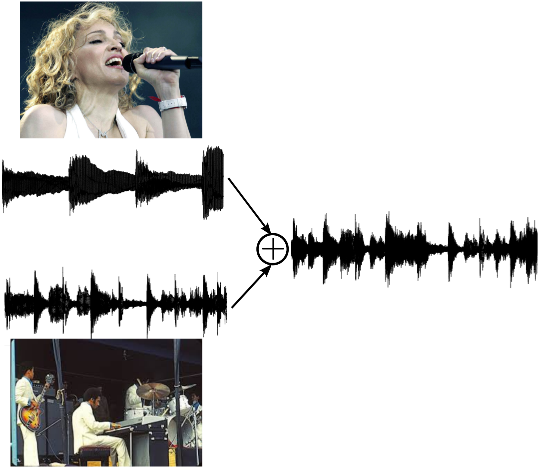
- This cannot be minimized in closed form for H. (Try it!)
- Suppose $\tilde{H}_{kj}$ is our current guess and define $\pi_{ijk} = \frac{W_{ik} H_{kj}}{\sum_k W_{ik} H_{kj}}$. Then we can write

$$D(V|WH) = -\sum_{i,j} V_{ij} \log \sum_k \pi_{ijk} \frac{W_{ik} H_{kj}}{\pi_{ijk}} + \sum_{i,j} \sum_k W_{ik} H_{kj} + \text{const.}$$

  and use Jensen's inequality on the $\log$.
- This "majorizing" function can easily be minimized!

Where are we?

The Problem of Source Separation



play stop

How do we recover the individual sources from just the mixture?

Source Separation Pipeline

$$V \approx WH$$

Ideal Pipeline:

1. Find a segment of music where only the backing band is playing. Take the magnitude STFT and run NMF. Keep W_{bg}.
2. Find a segment of music where only vocals are present. Take the magnitude STFT and run NMF. Keep W_{vocal}.
3. Now take the STFT of the mixture. Run NMF on the magnitudes, where W_{bg} and W_{vocal} are fixed, and estimate H_{bg} and H_{vocal}.

$$V \approx \begin{bmatrix} W_{bg} & W_{vocal} \end{bmatrix} \begin{bmatrix} H_{bg} \\ H_{vocal} \end{bmatrix}$$

4. The backing band magnitudes can be recovered as $W_{bg}H_{bg}$. Use the phases from the STFT, and take the ISTFT to recover the time-domain signal.

Source Separation Pipeline

$$V \approx WH$$

In practice:

1. Find a segment of music where only the backing band is playing. Take the magnitude STFT and run NMF. Keep W_{bg}.
2. ~~Find a segment of music where only vocals are present. Take the magnitude STFT and run NMF. Keep W_{vocal}.~~
3. Now take the STFT of the mixture. Run NMF on the magnitudes, where W_{bg} ~~and W_{vocal}~~ are fixed, and estimate W_{vocal}, H_{bg} and H_{vocal}.

$$V \approx \begin{bmatrix} W_{bg} & W_{vocal} \end{bmatrix} \begin{bmatrix} H_{bg} \\ H_{vocal} \end{bmatrix}$$

4. The backing band magnitudes can be recovered as $W_{bg}H_{bg}$. Use the phases from the STFT, and take the ISTFT to recover the time-domain signal.

A Demo

Let's try this in R!

Where are we?

Projects

- Project proposals due Friday. (Please fill in survey first.)
- I will be available after class. I will also be available tomorrow morning. Please e-mail me to schedule a time.

Homework 3

- Geostatistics packages in R: gstat and geoR
- Prediction competition