Signal and Information Processing

Alejandro Ribeiro
Dept. of Electrical and Systems Engineering
University of Pennsylvania
aribeiro@seas.upenn.edu
https://alelab.seas.upenn.edu/

January 20, 2021

Presentations

Data, Signals, and Information

Me

- Alejandro Ribeiro
- Professor, Department of Electrical and Systems Engineering
- Warren Center on Data and Network Sciences in Walnut 3401, 4th floor
- Email: aribeiro@seas.upenn.edu
- Webpage: `https://alelab.seas.upenn.edu/`

- My research is on signal and information processing. Best possible class for me to teach

- We meet online $\Rightarrow$ Mondays, Wednesdays, Fridays 10 am to 11 am
- I would like to see you in class. A minimum level of social connection is important!
- There is ample empirical evidence on the large correlation of grades and class attendance

How to Stay on Top

- Course's website $\Rightarrow$ https://ese224.seas.upenn.edu/
- Discussion forum in Piazza $\Rightarrow$ https://piazza.com/class/kjuecpey9hb4hr

Teaching Assistants

- Luana Ruiz
- Vinicius Lima
- Zhan Gao
- Zhiyang Wang
- Juan Cerviño

Midterm, final, and grading

- First midterm will be in class, during the lecture on Friday, March 6
  - $\Rightarrow$ This is the Friday right before the Spring break. Plan accordingly!
- Second midterm on the Wednesday, April 29 lecture $\Rightarrow$ Worth 26 points. Not cumulative.
  - $\Rightarrow$ This is the last day of classes. Plan accordingly!
- You will be handing in 13 lab reports which we will grade from 0 to 4
  - $\Rightarrow$ No show (0). Poor (1). Fair (2). Good (3). Excellent (4)
- Maximum possible points: $104 = 26 + 26 + 13 \times 4$
- You pass with 60 points and 13 points in midterm and final (each)
- C requires at least 70 points. B at least 80. A at least 90
- Don't worry. We will work hard. We will learn. We will get A's.

Presentations

Data, Signals, and Information

Signals, Data, and Information

- Data and signals are collections of values (numbers) we have acquired

  ⇒ A conversation recording, an image, car fuel efficiency
- People tend to say "signal" if they are electrical engineers. and to say "data" if they are not

- In any event, this disagreement is irrelevant ⇒ Signals and data are functions

Signals, Data, and Information

- Information is the answer to a question of interest
  - ⇒ For example, these two are pictures of the Eiffel tower
  - ⇒ But also, “the tower is illuminated” (left) or “the tower is under construction (right)

- The goal of signal processing is to extract information from data

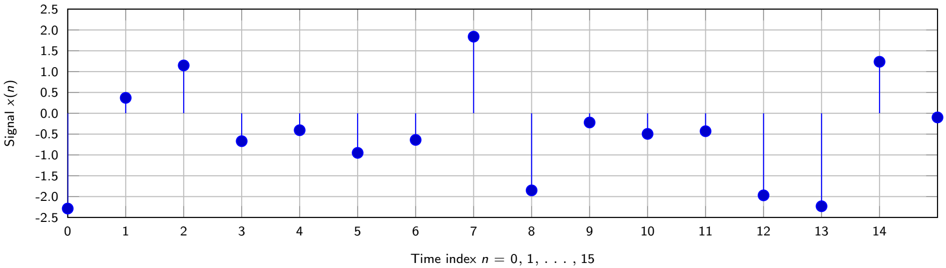
The purist and the pragmatist

- The pragmatist: Data is different from information
- The purist: Information is not created by processing $\Rightarrow$ Data IS information
- I am a pragmatist, but the purist is right
- Dodo principle: Everybody has won and all must have prizes

  $\Rightarrow$ Process signals and information to uncover patterns of interest

The number of atoms in the universe

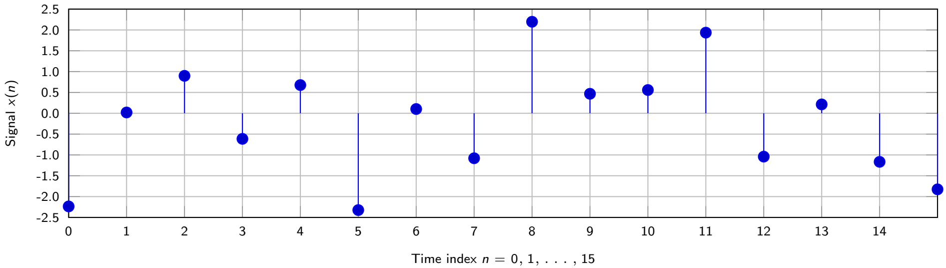
- What is so difficult about signal processing?
  $\Rightarrow$ Dimensionality $\Rightarrow$ The number of possible signals is too large



- This is one signal with 16 components

The number of atoms in the universe

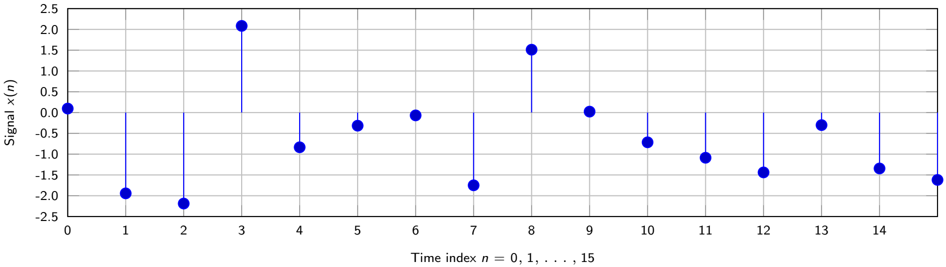
- What is so difficult about signal processing?
  $\Rightarrow$ Dimensionality $\Rightarrow$ The number of possible signals is too large



- This is a second signal with 16 components

The number of atoms in the universe

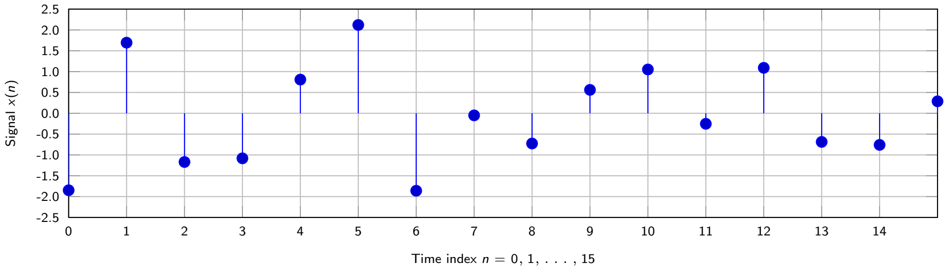
- What is so difficult about signal processing?
  $\Rightarrow$ Dimensionality $\Rightarrow$ The number of possible signals is too large



- This is a third signal with 16 components

The number of atoms in the universe

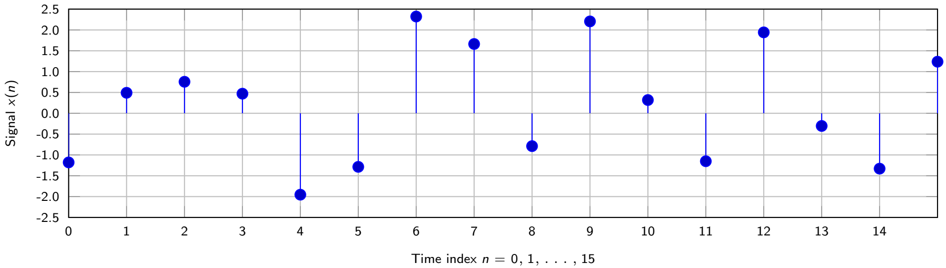
- What is so difficult about signal processing?
  - $\Rightarrow$ Dimensionality $\Rightarrow$ The number of possible signals is too large



- This is a fourth signal with 16 components

The number of atoms in the universe

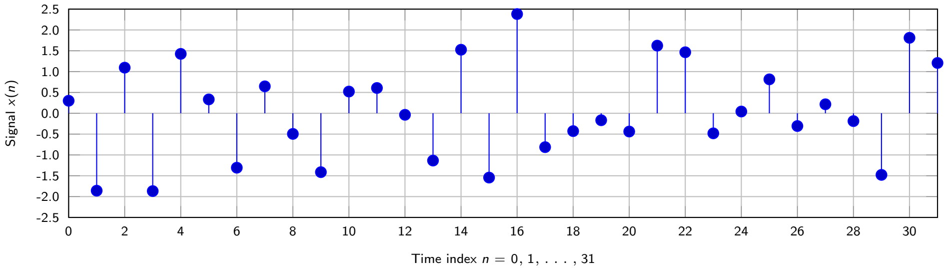
- What is so difficult about signal processing?
  $\Rightarrow$ Dimensionality $\Rightarrow$ The number of possible signals is too large



- There are 10^{16} signals that we can generate with 16 components

The number of atoms in the universe

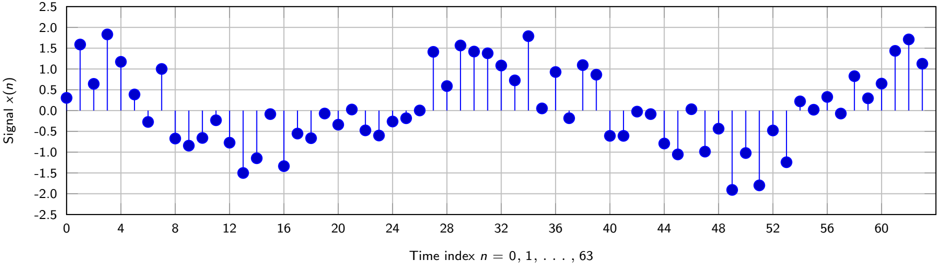
- What is so difficult about signal processing?
  - $\Rightarrow$ Dimensionality $\Rightarrow$ The number of possible signals is too large



- There are 10^{32} signals that we can generate with 32 components

The number of atoms in the universe

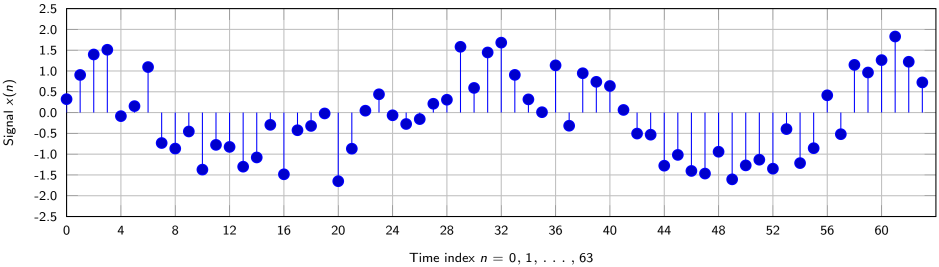
- What is so difficult about signal processing?
  $\Rightarrow$ Dimensionality $\Rightarrow$ The number of possible signals is too large



- There are 10^{64} signals that we can generate with 64 components

The number of atoms in the universe

- What is so difficult about signal processing?
  $\Rightarrow$ Dimensionality $\Rightarrow$ The number of possible signals is too large



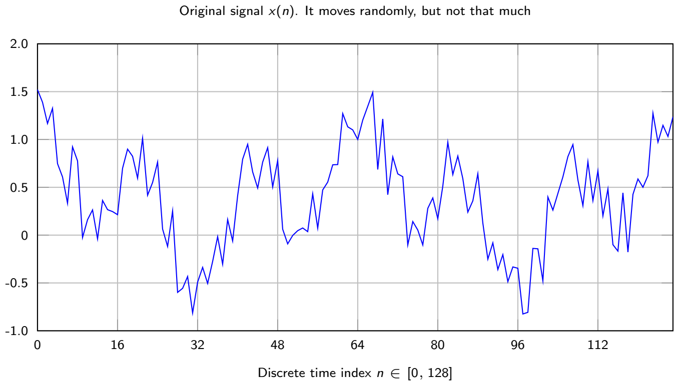
- Let's stop. There are 10^{78} to 10^{82} atoms in the universe

The stars in the sky

- Impossible to explore any large dimensional space with any meaningful degree of accuracy.
- Data points are like the stars in the sky $\Rightarrow$ A lot. But far between.
- And yet, our eyes can identify patterns with so much ease...

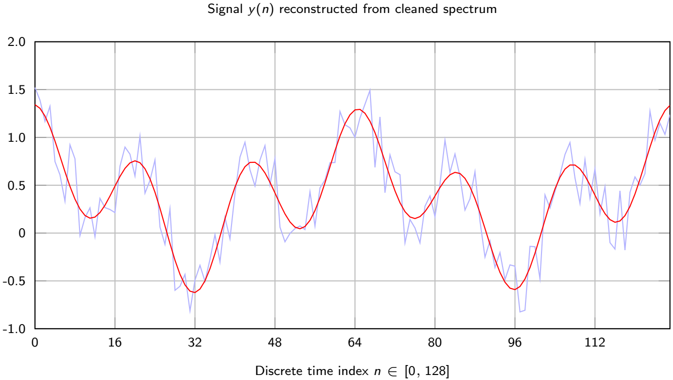
What's the pattern?

- When you look at this signal ALL of you can tell me what it is



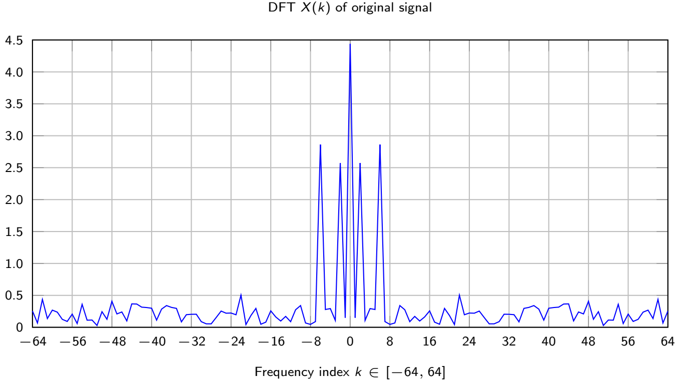
What's the pattern?

- When you look at this signal ALL of you can tell me what it is



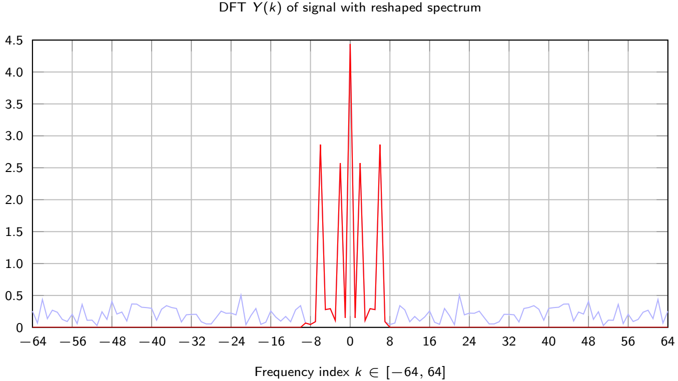
- The slow oscillation is a pattern. The fast oscillation is noise

- We represent the signal in the frequency domain.
  - $\Rightarrow$ We decompose into different modes of variability (slow and fast)



- The pattern (spikes) is clearly separated from the noise (the floor)

- We represent the signal in the frequency domain.
  - $\Rightarrow$ We decompose into different modes of variability (slow and fast)



- Keep the pattern. Remove the noise

The Little Prince

- The Fourier transform is an alternative representation of the signal
- It is equivalent $\Rightarrow$ Can go back and forth. Information is preserved
- But patterns are easier to explain $\Rightarrow$ Information is uncovered
- Fourier representation is called frequency (variability) representation
- Literally a new sense to view things that are otherwise invisible

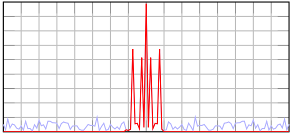
"On ne voit bien qu'avec le coeur.
L'essentiel est invisible pour les yeux."

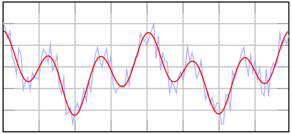
The Little Prince

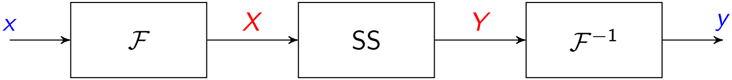
- Coeur being, of course, the french word for frequency

Uncovering the pattern

- This does work. This is how I generated the red (clean) signal







- Don't have to undo transformation $\Rightarrow$ Except for the benefit of our eyes

Good technology is indistinguishable from magic

- Indeed. If you are not in awe, you don't have a heart.
  - $\Rightarrow$ Change majors. We don't want you here.
  - $\Rightarrow$ Just kidding! Stay here, and let me convince you!
- But magic is not real magic $\Rightarrow$ It's a trick the magician can explain
- The Fourier transform reduces the dimensionality of the problem
  - $\Rightarrow$ There are only a few Fourier coefficients that are meaningful
  - $\Rightarrow$ These are not stars in the sky anymore
  - $\Rightarrow$ The Fourier transform concentrates information

- Frequency representations is how we uncover patterns in time signals
  $\Rightarrow$ Noise removal. Compression. Speech recognition.
- Images $\Rightarrow$ Two (multi) dimensional Fourier transforms
- Random Signals $\Rightarrow$ Principal Component Analysis
- Signals with arbitrary strucutres $\Rightarrow$ Graph Signal Processing

Beyond ESE 224

- We talk representations and their role in concentrating information
  - ⇒ But we don't dwell too much on how to exploit representations
- ESE 305 Foundations of Data Science
- ESE 545 Data Mining: Learning from Massive Datasets
- CIS 520 Machine Learning
- Deep Learning ⇒ Convolutional Neural Networks (CNNs)
  - ⇒ Fourier transforms and Convolutional filters are equivalent
  - ⇒ CNNs compose convolutional filters with local nonlinearities
- Graph Neural Networks ⇒ `https://gnn.seas.upenn.edu`