1: Elements of Statistical Signal Processing

ECE 830, Spring 2017

What do we have here?

The first step in many scientific and engineering problems is often signal analysis. Given measurements or observations of some physical process, we ask the simple question **"what do we have here?"** For instance,

- Is there any information in my measurements, or are they just noise?
- Is my signal in category A or B?
- What is the signal underlying my noisy measurements?

Answering this question can be particularly challenging when

- measurements are corrupted by noise or errors
- the physical process is "transient" or its behavior changes over time.

Fourier analysis

In some contexts, these challenges can be addressed via **Fourier analysis**, one of the major achievements in physics and mathematics. It is central to signal theory and processing for several reasons.
Recall the Fourier series:

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{-j2\pi f_k t}.$$

This is used for

- analysis of physical waves (acoustics, vibrations, geophysics, optics)
- analysis of periodic processes (economics, biology, astronomy)

Fourier analysis and filtering

Recall the Fourier transform

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$$

and the convolution integral

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau) x(t-\tau) d\tau \\ &= \int_{-\infty}^{\infty} H(f) X(f) e^{j2\pi ft} df \end{aligned}$$

which describes, for example, the result of sending a signal x through a filter h. **Two key facts:**

- Convolution in time $\Longleftrightarrow$ multiplication in frequency
- A stationary, zero-mean, Gaussian random process can be represented as a white noise process passed through a linear, time-invariant filter

Limitations of Fourier analysis

The inverse Fourier transform

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df$$

reveals that any value $x(t)$ of a signal at one time instant can be regarded as an infinite superposition of complex exponentials – everlasting and completely non-local waves.

Even though this mathematical representation can aid us in the discovery of signal structure in certain cases (e.g. periodic behavior) it can also distort the physical reality.

In particular...

1. Many signals, especially those which are transient in nature, are not well represented in terms of sinusoidal waves.
    - e.g. Images contain edges which are not efficiently represented with sinusoids.
    - e.g. Suppose that the signal $x(t)$ is exactly zero outside a certain time interval (e.g. by switching a machine on and off).

   Although this signal can still be studied by Fourier techniques, the frequency domain representation has a very artificial behavior. The time signal's zero values are achieved by an infinite superposition of virtual waves that interfere in such a way that they cancel each other out.

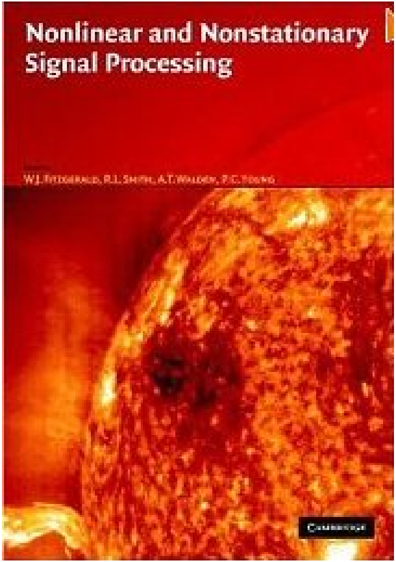
In particular...

2. often, it is the non-stationary or transient components of a signal that carry the important information.

e.g. Imagine that you are on a beach, watching the waves roll in to shore. Your peaceful state is broken as dolphins begin to evacuate earth to make way for an intergalactic superhighway.

In particular...

3. As we have seen, stationary Gaussian processes are intimately linked with Fourier analysis. However, many signals in which we are interested (e.g. speech, images) are not well modeled as stationary and Gaussian



In this course, we will move beyond Fourier analysis and focus on

Statistical Digital Signal Processing

Statistical	based on probabilistic models for signals and noise
Digital	discrete-time, sampled, quantized
Signal	waveform, sequence of measurements or observations
Processing	analyze, modify, synthesize

Examples of digital signals

- sampled speech waveform
- pixelized image
- Dow-Jones index
- stream of internet packets
- vector of medical predictors

A major difficulty

In many DSP applications, we don't have complete or perfect knowledge of the signals we wish to precess. We are faced with many unknowns and uncertainties.

Examples:

- Unknown signal parameters (delay of radar return, pitch of speech signal)
- Environmental noise (multipath signals in wireless communications, ambient electromagnetic waves, radar jamming)
- Sensor noise (grainy images, old phonograph recordings)
- Variability inherent in nature (stock market, internet)

How can we process signals in the face of such uncertainty? Can we model the uncertainty and incorporate this model into the processing?

Statistical Signal Processing is the study of answers to these questions.

Modeling uncertainty

There are many ways to model these sorts of uncertainties. In this course we will model them probabilistically. Let $p(x|\theta)$ denote a probability distribution parameterized by θ. The parameter θ could represent characteristics of errors or noise in the measurement process or govern inherent variability in the signal itself. For example, if x is a scalar measurement then we could have

$$p(x|\theta) = \frac{1}{\sqrt{2\pi}}\mathsf{exp}(-(x-\theta)^2/2),$$

a model which says that typically x is close to the value of θ and rarely is very different.

Why Probabilistic Models?

The observations or measurements we make are seldom perfect; often they are impure and contaminated by effects unknown to us. We call these effects noise. Our models are seldom perfect. Even the best choice of θ may not perfectly predict new observations. We call these modeling errors bias.

How do we model noise and bias, these uncertain errors? We need a calculus for uncertainty, and among many that have been proposed and used, the probabilistic framework appears to be the most successful, and in many situations it is physically plausible as well.

Uses of probabilistic models

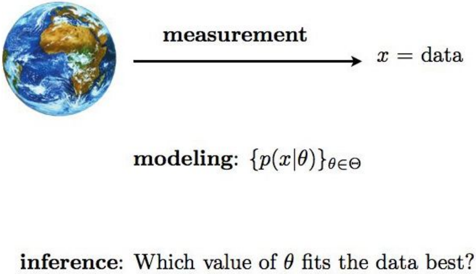
- sensor noise modeled as an additive Gaussian random variable
- uncertainty in the phase of a sinusoidal signal modeled as a uniform random variable on $[0, 2\pi)$
- uncertainty in the number of photons striking a CCD per unit time modeled as a Poisson random variable.

Components of Statistical Signal Processing: Modeling, Measurement, and Inference

Step 1: Postulate a probability model (or collection of models) that can be expected to reasonably capture the uncertainties in the data

Step 2: Collect data.

Step 3: Formulate statistics that allow us to interpret or understand our probability models.



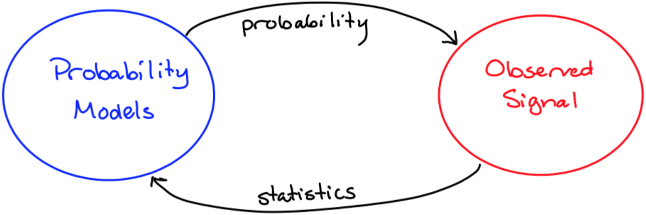
modeling: $\{p(x|\theta)\}_{\theta\in\Theta}$

inference: Which value of θ fits the data best?

Probability and statistics

Probability laws describe the uncertainty in the signals we might observe.

Statistics describe the salient features of the signals we do observe, and allow us to draw conclusions (inferences) about which probability model actually reflects the true state of nature.



Statistics

A statistic is a function of observed data, and may be scalar or vector valued.

Example: Statistics

Supposed we observe n scalar values $x_1, \ldots, x_n$. The following are statistics:

- sample mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- the data itself $[x_1, \ldots, x_n]^T$
- an order statistic $\min\{x_1, \ldots, x_n\}$
- an arbitrary function $[x_1^2 - x_2 \sin(x_3), e^{-x_1 x_3}]^T$

A statistic cannot depend on unknown quantities.

Four main problems

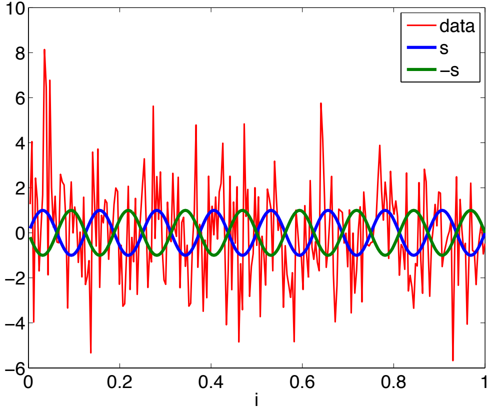
There are four fundamental inference problems in statistical signal processing that will be the focus of this course.

Example: **Detection**

Suppose that θ takes one of two possible values, so that either $p(x|\theta_1)$ or $p(x|\theta_2)$ fit the data x the best. Then we need to "decide" whether $p(x|\theta_1)$ is a better model than $p(x|\theta_2)$. More generally, θ may be one of a finite number of values $\{\theta_1, \ldots, \theta_M\}$ and we must decide among the M models.

A Detection Example

Consider a binary communication system. Let $s = [s_1, \ldots, s_n]$ denote a digitized waveform. A transmitter communicates a bit of information by sending s or $-s$ (for 1 or 0, respectively). The receiver measures a noisy version of the transmitted signal.



A detection example (cont.)

We model our observations as

$$x_i = \theta s_i + \epsilon_i \ , \quad i = 1, \ldots, n$$

The parameter θ is either $+1$ or -1, depending on which bit the transmitter is sending. The $\{\epsilon_i\}$ represent errors incurred during the transmission process. So we have two models, or *hypotheses*, for the data:

$$\begin{aligned} H_0 : \ x_i &= +s_i + \epsilon_i \ , \quad i = 1, \ldots, n \\ H_1 : \ x_i &= -s_i + \epsilon_i \ , \quad i = 1, \ldots, n \end{aligned}$$

How well does $\{s_i\}$ match $\{x_i\}$? How well does $\{-s_i\}$ match $\{x_i\}$?

This comparison can be made by computing a function of the data or a statistic. A natural statistic in this problem is the correlation statistic:

$$\begin{aligned} t &= \sum_{i=1}^{n} s_i x_i \\ &= \theta \sum_{i=1}^{n} s_i^2 + \sum_{i=1}^{n} s_i \epsilon_i \end{aligned}$$

If the errors are noise-like and don't resemble the signal $\{s_i\}$, then $\sum_{i=1}^{n} s_i \epsilon_i \approx 0$. So a reasonable way to decide which value of the bit was sent is to decide that 0 was sent if $t < 0$ and that 1 was sent if $t > 0$. To quantify the performance of this test we need a mathematical model for the errors $\{\epsilon_i\}$.

Four main problems

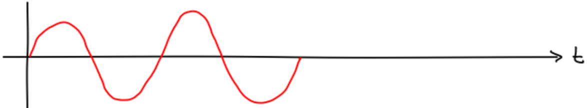
Example: **Parameter Estimation**

Suppose that θ belongs to an infinite set. Then we must decide or choose among an infinite number of models. In this sense, estimation may be viewed as an extension of detection to infinite model classes. This extension presents many new challenges and issues and so it is given its own name.

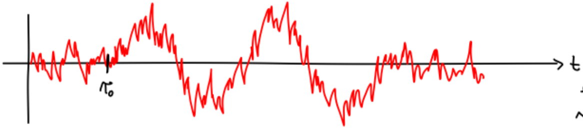
A Parameter Estimation Example

Radar example:

Transmit pulse:



Received waveform:



$\tau_0 = 2d/c$,

c = speed of propagation

estimate $\tau_0 \to \hat{\tau}_0$

$$\hat{d} = \frac{c\hat{\tau}_0}{2}$$

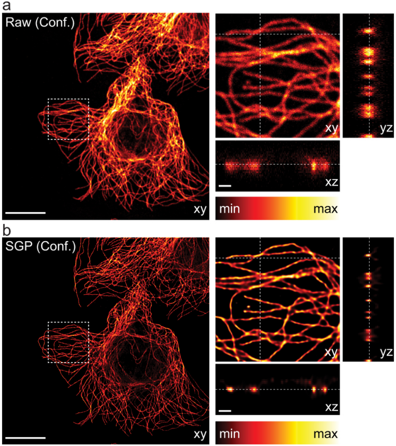
Four main problems

Example: **Signal Estimation/Prediction**

In many problems we wish to predict the value of a signal x given an observation of another related signal y. We can model the relationship between x and y using a *joint* probability distribution, $p(x, y)$. The *conditional* distribution of x given y, denoted by $p(x|y)$, can be derived from the joint distribution and the prediction problem can then be viewed as determining a value of x that is highly probable given y.

A Signal Estimation Example

Imagine that you are collaborating with biologists who are interested in imaging biological systems using a new type of microscopy. The imaging system doesn't produce perfect images: the data collected is distorted and noisy. As a signal processing expert, you are asked to develop an image processing algorithm to "restore" the image.



http://www.nature.com/srep/2013/130828/srep02523/full/srep02523.html

A Signal Estimation Example (cont.)

Let us assume that the distortion is a linear operation. Then we can model the collected data by the following equation.

$$y \quad = \quad Hx + w$$

where

- x is the ideal image we wish to recover (represented as a vector, each element of which is a pixel),
- H is a known model of the distortion (represented as a matrix), and
- w is a vector of noise.

A Signal Estimation Example (cont.)

It is tempting to ignore w and simply attempt to solve the system of equations $y = Hx$ for x. There are two problems with this approach.

1. The system of equations may not admit a unique solution, depending on the physics of the imaging system. If a unique solution exists, the problem is said to be well-posed, otherwise it is called ill-posed.
2. Even if the system of equations is invertible, it may be ill-conditioned, which means that small perturbations due to noise and numerical methods can lead to large errors in the restoration of x.

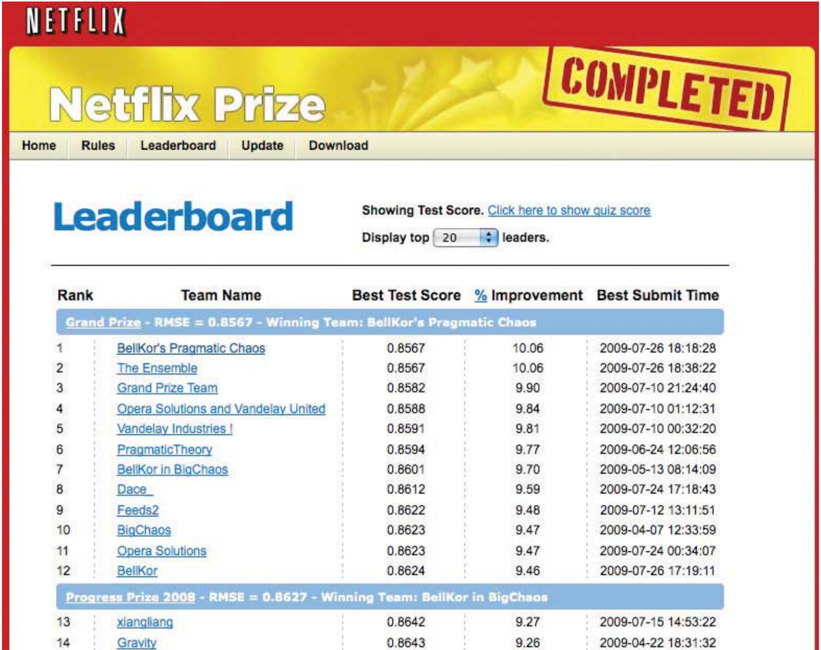
Four main problems

Example: Learning

Sometimes we don't know a good model the relationship between x and y, but we do have a number of "training examples", say $\{x_i, y_i\}_{i=1}^{n}$, that give us some indication of the relationship. The goal of learning is to design a good prediction rule for y given x using these examples, instead of $p(y|x)$.

The Netflix problem

The Netflix prize



Leaderboard

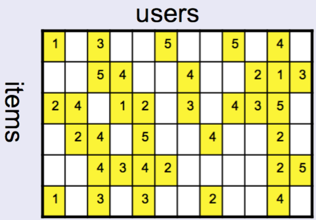
Showing Test Score. Click here to show quiz score

Display top 20 leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11
Progress Prize 2008 - RMSE = 0.8627 - Winning Team: BellKor in BigChaos				
13	xiangliang	0.8642	9.27	2009-07-15 14:53:22
14	Gravity	0.8643	9.26	2009-04-22 18:31:32

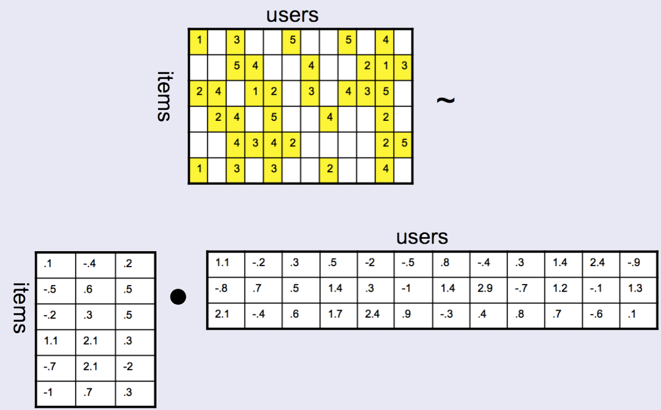
Example: Predicting Netflix ratings

Here x contains the measured movie ratings and y are the unknown movie ratings we wish to predict.



1		3			5			5		4	
		5	4			4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

Example: Predicting Netflix ratings (cont.)



users

items

1		3			5			5		4	
		5	4			4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

~

items

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

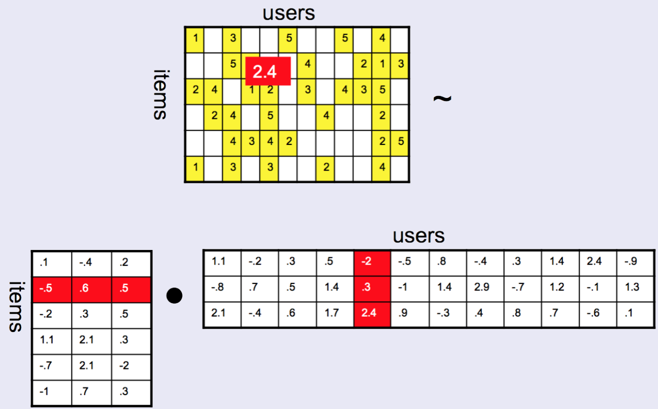
•

users

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

One probabilistic model says the underlying matrix of “true” ratings can be factored into the product of two smaller matrices.

Example: Predicting Netflix ratings (cont.)



1		3			5			5		4	
		5		2.4		4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

One probabilistic model says the underlying matrix of "true" ratings can be factored into the product of two smaller matrices.