Lecture Notes: Applications of Linear Regression
Noise and Fourier Analysis

33A Linear Algebra, Puck Rombach

Last updated: May 10, 2016

Linear Regression and Noise

The following is not discussed in Bretscher, but I think it is very important to think about if you will ever want to apply linear regression to a problem in the future. This is a heavily simplified exposition of a far more informative blog post by Jacob Steinhardt (jsteinhardt.wordpress.com/2010/08/22/least-squares-and-fourier-analysis/).

When we apply linear regression, we assume that we have an underlying system that behaves according to a law with unknown parameters, and that the only reason that our system is inconsistent is noise. This noise may be on the input vector, the output vector, or both. This means that in our inconsistent system

\[ A\vec{x} = \vec{b}, \]

we may have a noisy matrix

\[ A = A_t + A_N, \]

where \( A_t \) is the true input matrix and \( A_N \) is the added noise, and

\[ \vec{b} = \vec{b}_t + \vec{b}_N, \]

where \( \vec{b}_t \) is the true output vector and \( \vec{b}_N \) is the added noise. Then, we can rewrite our normal equation as

\[ \vec{x}^* = (A_t^T A_t)^{-1} A_t^T (\vec{b}_t + \vec{b}_N - A_N \vec{x}^*). \]

We assume that our true consistent system is \( \vec{x}^* = A_t \vec{b}_t \). Therefore, on average, we would need the noise term \((A_t^T A_t)^{-1} A_t^T (\vec{b}_N - A_N \vec{x}^*)\) to be 0 (such that indeed \( \vec{x}^* = (A_t^T A_t)^{-1} A_t^T \vec{b}_t \)).

First, consider \( \vec{b}_N \). The noise is unbiased (if the noise is biased, it is probably not noise in the first place). This means that \( \sum [\vec{b}_N]_i = 0 \), i.e. the average value of \( \vec{b}_N \) is 0. The noise on the output should also be uncorrelated to the noise on the input, because, if there is correlated noise, it cannot be distinguished from any real relation between input and output. When two variables \( \vec{v} \) and \( \vec{w} \) are uncorrelated, then we expect the dot product of the deviation vectors to be 0, i.e. \( \mathbb{E}[(\vec{v} - \text{mean}(\vec{v})) \cdot (\vec{w} - \text{mean}(\vec{w}))] = 0 \). You should be able to check that it is enough to normalize only one of them: i.e. \( \mathbb{E}[(\vec{v} - \text{mean}(\vec{v})) \cdot \vec{w}] = 0 \).
Now, the entries of the vector $A^T \vec{b}_N$ are the dot products of the rows of $A^T$ and the vector $\vec{b}_N$. By our assumptions, we have $\mathbb{E}(A^T \vec{b}_N) = \vec{0}$. This means that if we have unbiased noise on the output, this noise does not bias the solution $\vec{x}^\circ$.

However, there is clearly a problem when we consider $A^T A_N \vec{x}^\circ$. Neither the rows of $A^T$ or the vector $A_N \vec{x}^\circ$ will often be heavily correlated. This means that unbiased noise on the input may cause a systematic bias on the output. (This bias would not go away by, for example, taking a larger sample size.) That is terrible.

**Application: Fourier Analysis**

Suppose that we measure a signal at $N$ times with equally spaced intervals $0, \Delta t, 2\Delta t, \ldots, (n-1)\Delta t$. To simplify, let’s set $\Delta t = 1$. This is a sequence $y_1, y_2, \ldots, y_N$, where the $y_j$s are complex scalars.

We can think of this as a scalar function $f : \{0, 1, 2, \ldots, (n-1)\} \to \mathbb{C}$. The **discrete Fourier transform** is a sequence $c_0, \ldots, c_{n-1}$ such that our function can be written as a superposition of $n$ wave functions (with frequencies $\theta_0 = 0, \theta_1 = \frac{2\pi}{n}, \theta_2 = \frac{4\pi}{n}, \ldots, \theta_{n-1} = \frac{2(n-1)\pi}{n}$):

$$f(x) = \sum_{k=0}^{n-1} c_k e^{\theta_k ix}.$$ 

It may be surprising that this is possible, but we can show that it is using linear algebra!

We can think of the signal or the function $f$ as a vector $\vec{f} \in \mathbb{C}^n$. Now, consider the vectors $\vec{f}_0, \vec{f}_1, \ldots, \vec{f}_{n-1}$ given by

$$\vec{f}_k = \begin{pmatrix} 1 \\ e^{\theta_0 i} \\ \vdots \\ e^{\theta_0 i(n-1)} \end{pmatrix}.$$ 

For any $j \neq k$,

$$\langle \vec{f}_j, \vec{f}_k \rangle = 1 + e^{\frac{2\pi i(j-k)}{n}} + \ldots + e^{\frac{2\pi i(j-k)(n-1)}{n}} = 0. \quad \text{(Why?)}$$

When $j = k$,

$$\langle \vec{f}_j, \vec{f}_k \rangle = 1 + \ldots + 1 = n.$$ 

So, $\{\vec{f}_0, \vec{f}_1, \ldots, \vec{f}_{n-1}\}$ is an orthogonal (not quite orthonormal) basis for $\mathbb{C}^n$. This proves that the discrete Fourier transform exists and is unique!

Suppose that we are interested in a set of $p < n$ frequencies, in other words, we would like to approximate our signal using fewer frequencies than necessary for a perfect reconstruction. Any function of the form

$$y = c_1 e^{\theta_1 x} + c_2 e^{\theta_2 x} + \ldots + c_p e^{\theta_p x},$$

lends itself perfectly for linear regression.

If the set of frequencies we are interested in is a subset of the set given by the Fourier transform, then a linear regression approximation of our function corresponds to a projection onto a subspace.
spanned by a subset of an orthogonal basis of $\mathbb{C}^n$. You should be able to see that we obtain identical coordinates for the relevant frequencies as they would in the full Fourier transform.

Furthermore, if the observed signal is real, then the approximated signal is real, as long as positive and “negative” frequencies cancel each other out. (This is a bit beyond the scope of this material, but I will show examples of this in the lecture.)

Suppose that we have the following signal of 10 equally spaced points in the interval $[0, 10]$.

<table>
<thead>
<tr>
<th>Time</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>-5</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>-6</td>
</tr>
<tr>
<td>6</td>
<td>-10</td>
</tr>
<tr>
<td>7</td>
<td>-3</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
</tr>
</tbody>
</table>

To find the full discrete Fourier transform, we simply need to write our vector $\vec{y} = (9, -5, \ldots, 7)^T$ in terms of the complex basis $\vec{f}_0, \vec{f}_1, \ldots, \vec{f}_{n-1}$. We can do this using the inner product method that we are familiar with for projections. (This basis is not quite orthonormal, but that is easily corrected with a factor $\frac{1}{n}$.) We get

$$c_k = \frac{1}{n} \langle \vec{y}, \vec{f}_k \rangle = \frac{1}{n} \sum_{t=0}^{9} y_t e^{-\frac{2\pi i kt}{10}}, \quad 0 \leq k \leq n - 1.$$  

Now we have 10 complex scalars $c_k$, and this gives us our expression

$$\vec{y} = (c_0\vec{f}_0 + \ldots + c_{n-1}\vec{f}_{n-1}),$$

and continuous function

$$y(t) = \sum_{k=0}^{9} c_k e^{\frac{2\pi i kt}{10}}.$$  

This is indeed a real function that gives an exact solution for all for our observations.
This solution has equal amplitudes for the frequencies 1, 9, the frequencies 2, 8, the frequencies 3, 7, etc. we can combine these and only use the lower frequencies. In other words, we remove the signal of frequency 9, and double the signal of frequency 1. And do the same for the other pairs. This looks as follows:

The reason this works is that we can write

\[ y(t) = \sum_{k=0}^{n-1} c_k e^{i\theta_k t} = \sum_{k=0}^{n-1} (\text{Re}(c_k) + i \text{Im}(c_k) \cdot i) (\cos(\theta_k t) + i \sin(\theta_k t)) \]

\[ = \sum_{k=0}^{n-1} \text{Re}(c_k) (\cos(\theta_k t) + i \sin(\theta_k t)) - \sum_{k=0}^{n-1} \text{Im}(c_k) (\sin(\theta_k t) - i \cos(\theta_k t)) \]

\[ = \begin{cases} 
  c_0 + c_{n/2} \cos(\theta_{n/2} t) + 2 \sum_{k=1}^{n/2-1} (\text{Re}(c_k) \cos(\theta_k t) + \text{Im}(c_k) \sin(\theta_k t)), & n \text{ even}, \\
  c_0 + 2 \sum_{k=1}^{\lfloor n/2 \rfloor} (\text{Re}(c_k) \cos(\theta_k t) + \text{Im}(c_k) \sin(\theta_k t)), & n \text{ odd}. 
\end{cases} \]

**Exercise 6.1.** By hand, find the complex discrete Fourier transform of the function (0, 2), (1, 4), and write this as a real function

\[ f(t) = a_1 + a_2 \cos(\pi t) + a_3 \sin(\pi t). \]
Then, solve the same problem using just a linear system over $\mathbb{R}$. What is an important difference?

**Exercise 6.2.** By hand, find the complex discrete Fourier transform of the function $(0, 2), (1, -1), (2, -4)$, and write this as a real function

$$f(t) = a_1 + a_2 \cos(\frac{2}{3} \pi t) + a_3 \sin(\frac{2}{3} \pi t).$$

Then, solve the same problem using just a linear system over $\mathbb{R}$.

Now, suppose that we would like to approximate this signal using $< n$ frequencies. We have now seen that this amounts to finding a solution within a subspace of $\mathbb{C}^n$ that has $p < n$ dimensions. With complex vectors, linear regression looks almost the same as with real vectors, except that we have to add some conjugates (just like with the inner product). Instead of using the transpose of a matrix $A$, we use the **Hermitian transpose**:

$$A^H = (\overline{A})^T = (A^T).$$

So, suppose that we would like to only use the frequencies 0, 1.5, then we use the normal equation

$$\hat{c} = (A^H A)^{-1} A^H \hat{y},$$

where we add negative frequencies so that they can cancel out to produce the real signal:

$$A = \begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \downarrow & f_0 & f_{1.5} & f_{-1.5} \end{pmatrix}.$$

Alternatively (equivalently), we can use linear regression to fit our points to the expression

$$f(t) = a_1 + a_2 \cos(\frac{3}{2} \pi t) + a_3 \sin(\frac{3}{2} \pi t).$$

This looks like this.

<table>
<thead>
<tr>
<th>Time</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>-5</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>-6</td>
</tr>
<tr>
<td>6</td>
<td>-10</td>
</tr>
<tr>
<td>7</td>
<td>-3</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
</tr>
</tbody>
</table>
Solutions to Exercises

Solution 6.1. We wish to decompose the vector \( \vec{y} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = c_0 \vec{f}_0 + c_1 \vec{f}_1 \), where \( \vec{f}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \) and \( \vec{f}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \). Using any method you like, we find \( c_0 = 3 \) and \( c_1 = -1 \). This gives us

\[
 f(t) = 3 - e^{\pi i t} = 3 - (\cos(\pi t) + i \sin(\pi t)),
\]

and, since \( \sin(\pi t) = 0 \) when \( t \) is an integer, we can safely call this function

\[
 f(t) = 3 - \cos(\pi t).
\]

Indeed, \( f(0) = 2, f(1) = 4 \). Next, let’s solve this using real linear equations. We get system

\[
 \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.
\]

This gives us infinitely many solutions \( a_1 = 3, a_2 = -1 \) and \( a_3 \) is free. What this means is that we can add a \( \sin(\pi t) \) as much as we want, because this does not affect the values of \( f(t) \) at integer values. We get infinitely many functions:

\[
 f_s(t) = 3 - \cos(\pi t) + s \sin(\pi t),
\]

with, \( f_s(0) = 2, f_s(1) = 4 \), for all \( s \in \mathbb{R} \).

Solution 6.2. We wish to decompose the vector \( \vec{y} = \begin{pmatrix} 2 \\ -1 \\ -4 \end{pmatrix} = c_0 \vec{f}_0 + c_1 \vec{f}_1 + c_2 \vec{f}_2 \), where \( \vec{f}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \), \( \vec{f}_1 = \begin{pmatrix} e^{\frac{2\pi}{3} i} \\ e^{\frac{2\pi}{3} i} \end{pmatrix} \), and \( \vec{f}_2 = \begin{pmatrix} e^{\frac{4\pi}{3} i} \\ e^{\frac{4\pi}{3} i} \end{pmatrix} \). Using dot products or any other method you like, we find

\[
 c_0 = \frac{3}{\sqrt{3}}
\]

\[
 c_1 = \frac{1}{\sqrt{3}} (2 - e^{-\frac{2\pi}{3} i} - 4e^{-\frac{4\pi}{3} i}) = \frac{1}{\sqrt{3}} \left( \frac{9}{2} - i \frac{3\sqrt{3}}{2} \right)
\]

\[
 c_2 = \frac{1}{\sqrt{3}} (2 - e^{-\frac{2\pi}{3} i} - 4e^{-\frac{4\pi}{3} i}) = \frac{1}{\sqrt{3}} \left( \frac{9}{2} + i \frac{3\sqrt{3}}{2} \right).
\]

This gives us

\[
 f(t) = \frac{1}{\sqrt{3}} \sum_{k=0}^{2} \text{Re}(c_k) \left( \cos \left( \frac{2k}{3} \pi t \right) + i \sin \left( \frac{2k}{3} \pi t \right) \right) - \frac{1}{\sqrt{3}} \sum_{k=0}^{2} \text{Im}(c_k) \left( \sin \left( \frac{2k}{3} \pi t \right) - i \cos \left( \frac{2k}{3} \pi t \right) \right)
\]

\[
 = \frac{1}{3} \left( -3 + 9 \cos \left( \frac{2\pi}{3} t \right) + 3 \sqrt{3} \sin \left( \frac{2\pi}{3} t \right) \right).
\]

Indeed, \( f(0) = 2, f(1) = -1, f(2) = -4 \). If we solve this using real linear equations. We get system

\[
 \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1/2 & \sqrt{3}/2 \\ 1 & -1/2 & -\sqrt{3}/2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix}.
\]

This gives us solutions \( a_1 = -1, a_2 = 3 \) and \( a_3 = \sqrt{3} \).