MATH-GA 2840 HW#9

Yifei(Fahy) Gao yg1753

1. (Impulse response of a room) We are interested in analyzing the acoustic characteristics of an auditorium during a concert. To this end, we want to study the sound traveling from the singer's location to a seat in the last row.

(a) Under what assumptions would the sound propagation be characterized by the convolution of the emitted sound and an impulse response?

My assumption is that: the propagation of sound from a source can be modified by a linear translation in-variant system:

$$\begin{aligned}\mathcal{F}(x) &= \mathcal{F}\left(\sum_{j=0}^{N-1} x[j]e_j\right)\\ &= \sum_{j=0}^{N-1} x[j]\mathcal{F}\left(e_j\right)\\ &= \sum_{j=0}^{N-1} x[j]\mathcal{F}\left(e_0^{\downarrow j}\right)\\ &= \sum_{j=0}^{N-1} x[j]h_{\mathcal{F}}^{\downarrow j} = y\end{aligned}$$

(b) Assume the assumptions hold, and we measure the following impulse response: Why do you think the response is zero before 25 ms ?

Because the sound is trying to reach the seat in the last row. By the definition of impulse response,

$$h_{\mathcal{F}} := \mathcal{F}\left(e_0\right) = \mathcal{F}\left(\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}\right)$$

Notice that the first component of $y = x * h_{\mathcal{F}} = \sum_{j=0}^{N-1} x[j]h_{\mathcal{F}}^{\downarrow j}$ is

$$x[0]h_{\mathcal{F}}^{\downarrow 0} = x[0]h_{\mathcal{F}}$$

Since $h_{\mathcal{F}}$ has zero entries before 25 ms, the overall $x[0]h_{\mathcal{F}}$ is gonna be zero for entries before 25 ms.

(c) What do the two first spikes in the impulse response correspond to?

They correspond to a direct sound and an early reflection of the original sound.

(d) Suggest a way to measure the impulse response by bursting a balloon. What are the possible limitations of this approach?

We inflat ballons to fullest size until a perceivable thereshold in the inflation is reached. Then we set up an apparatus for the measurement of the impulse response in a seat in the last row. Then we burst the ballons at the same location as the location of the singer, then we take same measurements.

For the limitations:

1. Ballons may be inflation to different level, so the measurements may be different.
2. The speed of sound may be different for the ballons in different size.

(e) Suggest a way to measure the impulse response using a device that produces pure tones (i.e. sinusoidal sounds).

We can find the impulse response to a complex sinusoid with frequency. Complex sinusoid are eigenfunction since from our notes:

$$\mathcal{F}(e^{i2\pi wt}) = h_{\mathcal{F}} e^{i2\pi wt}$$

Then after taking the fourier translation we have:

$$\begin{aligned} \widehat{\mathcal{F}}(s) &= \hat{h}_{\mathcal{F}}(s)\delta(s-w) \\ &= \hat{h}_{\mathcal{F}}(w)\delta(s-w) \end{aligned}$$

Then taking the inverse of FT, we get:

$$F(e^{i2\pi wt}) = \hat{h}_{\mathcal{F}}(w) e^{i2\pi wt}$$

And we know $\hat{h}_{\mathcal{F}}(w)$ is a constant so we have a device that produces pure sinusoidal, so we can just measure the constant scaling factor $\hat{h}_{\mathcal{F}}(w)$ to get the impluse response.

2. (Discrete filter) Let us index the DFT coefficients of the N-dimensional vectors from −(N − 1)/2 to (N − 1)/2 (assuming N is odd). We define the bandlimited signals in this space as those for which the nonzero Fourier coefficients are zero beyond a certain value kc, i.e. x ∈ CN is bandlimited if xˆ[k] = 0 for all |k| > kc. Let y be the vector with the smallest l2 norm such that x ∗ y = x for all bandlimited vectors with cut-off frequency kc (where kc is a fixed integer smaller than (N − 1)/2). Derive an explicit expression for the entries of y, showing that they are real valued.

From the question implication,

$$x = x * y$$

And by the Corollary on the notes,

$$\mathcal{F}(x) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{h}_{\mathcal{F}}[k]\hat{x}[k]\psi_k.$$

Then:

$$\begin{aligned} x &= \tfrac{1}{N}\sum_{k=0}^{N-1} y[k]\hat{x}[k]\psi_k \\ &= \tfrac{1}{N}\sum_{k=-k_c}^{k} \hat{y}[k]\hat{x}[k]\psi_k \end{aligned}$$

Thus $\hat{y}$ has to be: 1, if $|K| > K_c$, then:

$$y = \mathcal{F}^{-1}(x)\hat{y} = \mathcal{F}^*\hat{y}$$

Then we can find the ℓ_2 norm of y using $\hat{y}$ from

$$\begin{aligned} \|y\|_2 = y^* y &= \frac{1}{N^2}\hat{y}^*\mathcal{F}\mathcal{F}^*\hat{y} \\ &= \frac{1}{N}\hat{y}^*\hat{y} \end{aligned}$$

Therefore, we must minimize the norm of $\hat{y}$ to minimize the norm of y.

Thus we can make a conclusion that:

$$\hat{y}[k] = \begin{cases} 1 & \text{if } |k| < k_c \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$\begin{aligned} y &= \frac{1}{2k_c}\sum_{k=-k_c}^{k_c}\hat{y}[k]\psi_k \\ &= \frac{1}{2k_c}\sum_{k=0}^{k_c}\hat{y}[k]\psi_k + \frac{1}{2k_c}\sum_{k=-1}^{-k_c}\hat{y}[k]\psi_k \\ &= \frac{1}{2k_c}\sum_{k=0}^{k_c}\hat{y}[k]\psi_k + \frac{1}{2k_c}\sum_{k=1}^{k_c}\hat{y}[-k]\psi_{-k} \\ &= \frac{1}{2k_c}\sum_{k=0}^{k_c}(\cos\left(\frac{2\pi k}{N}\right) + i\sin\left(\frac{2\pi k}{N}\right) + \cos\left(-\frac{2\pi k}{N}\right) + i\sin\left(-\frac{2\pi k}{N}\right)) \\ &= \frac{1}{2k_c}\sum_{k=0}^{k_c} 2\cos\left(\frac{2\pi k}{N}\right) \end{aligned}$$

And now we also can apparently say that y is real-valued.

3. (PCA of stationary vector) Let $\tilde{x}$ be a wide-sense stationary vector with real-valued autocovariance vector $a_{\tilde{x}}$, with covariance matrix $\Sigma_{\tilde{x}}$. In the notes we showed that the eigenvectors and eigenvalues of $\Sigma_{\tilde{x}}$ are complex exponentials and the DFT coefficients of 1

$a_{\tilde{x}}$ respectively. Here we will show that we can derive an equivalent real-valued eigendecomposition because the autocovariance vector is real. We will assume that N is an odd number.

(a) Show that the DFT coefficients of $a_{\tilde{x}}$ are real, and satisfy $\hat{a}_{\tilde{x}}[k] = \hat{a}_{\tilde{x}}[N-k]$ for $k = 1, \ldots, \frac{N-1}{2}$

Since the DFT of $a_{\tilde{x}}$ can be:

$$\hat{a}_{\tilde{x}}[k] = \Sigma_{j=0}^{N} a_{\tilde{x}}[j]\psi_k[j]$$

$$= a_{\tilde{x}}[0] + \Sigma_{t=1}^{\frac{N-1}{2}} (a_{\tilde{x}}[t]\psi_k[t] + a_{\tilde{x}}[N-k]\psi_k[N-k]) \tag{1}$$

since we have already knew that $a_{\tilde{x}}[j]$ is real-valued, so let us focus on the second term:

$$\begin{aligned} a_{\tilde{x}}[t]\psi_k[t] + a_{\tilde{x}}[N-k]\psi_k[N-k] \\ &= a_{\tilde{x}}[t](\psi_k[t] + \psi_k[N-k]) \\ &= a_{\tilde{x}}[t]\cos[\left(\frac{2\pi kt}{N}\right) + i\sin\left(\frac{2\pi kt}{N}\right) \\ &+ \cos\left(2\pi k - \frac{2\pi kt}{N}\right) + i\sin\left(2\pi k - \frac{2\pi kt}{N}\right)\Big] \\ &= 2a_{\tilde{x}}[t]\cos(\frac{2\pi kt}{N}) \end{aligned}$$

Therefore, plug the value back to (1), we get:

$$\hat{a}_{\tilde{x}}[k] = a_{\tilde{x}}[0] + \Sigma_{t=1}^{\frac{N-1}{2}} (2a_{\tilde{x}}[t]\cos(\frac{2\pi kt}{N}))$$

So it is real valued now.

And we can see that:

$$\begin{aligned} \hat{a}_{\tilde{x}}[N-k] - \hat{a}_{\tilde{x}}[k] &= a_{\tilde{x}}[0] + \Sigma_{t=1}^{\frac{N-1}{2}} (2a_{\tilde{x}}[t]\cos(\frac{2\pi(N-k)t}{N})) - a_{\tilde{x}}[0] \\ &+ \Sigma_{t=1}^{\frac{N-1}{2}} (2a_{\tilde{x}}[t]\cos(\frac{2\pi kt}{N})) \\ &= a_{\tilde{x}}[0] + \Sigma_{t=1}^{\frac{N-1}{2}} (2a_{\tilde{x}}[t]\cos(2\pi t - \frac{2\pi kt}{N})) - a_{\tilde{x}}[0] \\ &+ \Sigma_{t=1}^{\frac{N-1}{2}} (2a_{\tilde{x}}[t]\cos(\frac{2\pi kt}{N})) \\ &= 0\blacksquare \end{aligned}$$

(b) Show that

$$\Sigma_{\tilde{x}}[j_2, j_1] = \frac{1}{N}\sum_{k=0}^{N-1} \hat{a}_{\tilde{x}}[k]\exp\left(\frac{i2\pi k(j_2 - j_1)}{N}\right)$$

Based on theroem 3.5:

$$\begin{aligned}\Sigma_{\tilde{x}}\left[j_2, j_1\right] &= \boldsymbol{a}_{\tilde{x}}^{\downarrow j_1}\left[j_2\right] = \left\langle \hat{\boldsymbol{a}}_{\tilde{x}}^{\downarrow j_1}, \psi_{j_2} \right\rangle \\ &= e_{j_2}^{*} \frac{1}{N} \sum_{k=0}^{N-1} \exp\left(-\frac{i2\pi k j_1}{N}\right) \hat{a}_{\tilde{x}}[k] \psi_k \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \exp\left(-\frac{i2\pi k j_1}{N}\right) \hat{a}_{\tilde{x}}[k] \exp\left(\frac{i2\pi k j_2}{N}\right) \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \hat{a}_{\tilde{x}}[k] \exp\left(\frac{i2\pi k\left(j_2 - j_1\right)}{N}\right) \blacksquare\end{aligned}$$

(c) Show that $\Sigma_{\tilde{x}}$ has the following decomposition

$$\Sigma_{\tilde{x}}\left[j_2, j_1\right] = \sum_{k=0}^{N-1} \lambda_k u_k\left[j_1\right] u_k\left[j_2\right]$$

where the eigenvectors correspond to the orthonormal vectors

$$\frac{1}{\sqrt{N}}\begin{bmatrix}1\\1\\\cdots\\1\end{bmatrix}, \quad \sqrt{\frac{2}{N}}\begin{bmatrix}1\\\cos\left(\frac{2\pi k}{N}\right)\\\cdots\\\cos\left(\frac{2\pi k(N-1))}{N}\right)\end{bmatrix}, \quad \sqrt{\frac{2}{N}}\begin{bmatrix}0\\\sin\left(\frac{2\pi k}{N}\right)\\\cdots\\\sin\left(\frac{2\pi k(N-1))}{N}\right)\end{bmatrix}, \quad 1 \le k \le \frac{N-1}{2}$$

and report the value of $\lambda_0, \ldots, \lambda_k$

From the part b, we know: $\Sigma_{\tilde{x}}[j_2, j_1] = \frac{1}{N}\sum_{k=0}^{N-1} \hat{a}_{\tilde{x}}[k] \exp(\frac{i2\pi k(j_2 - j_1)}{N} t)$, then:

$$\Sigma_{\tilde{x}}\left[j_2, j_1\right] \qquad = \frac{1}{N}\left[\hat{a}_{\tilde{x}}[0] + \sum_{k=1}^{\frac{N-1}{2}} \hat{a}_{\tilde{x}}[k] \exp\left(\frac{i2\pi l\left(j_2 - j_1\right)}{N}\right)\right.$$

$$\left. + \hat{a}_{\tilde{x}}[N-k] \exp\left(\frac{i2\pi(N-k)\left(j_2 - j_1\right)}{N}\right)\right]$$

$$= \frac{1}{N}\left[\hat{a}_{\tilde{x}}[0] + 2\sum_{k=1}^{\frac{N-1}{2}} \hat{a}_{\tilde{x}}[k] \cos\left(\frac{2\pi k\left(j_2 - j_1\right)}{N}\right)\right]$$

$$= \frac{1}{N}\left[\hat{a}_{\tilde{x}}[0] + 2\sum_{k=1}^{\frac{N-1}{2}} \hat{a}_{\tilde{x}}[k]\left[\cos\left(\frac{2\pi k j_2}{N}\right)\cos\left(\frac{2\pi k j_1}{N}\right)\right.\right.$$

$$\left.\left. + \sin\left(\frac{2\pi k j_2}{N}\right)\sin\left(\frac{2\pi k j_1}{N}\right)\right]\right]$$

$$= \frac{1}{N}\left[\hat{a}_{\tilde{x}}[0]\right.$$

$$+ 2\sum_{k=1}^{\frac{N-1}{2}} \hat{a}_{\tilde{x}}[N-k] \cos\left(\frac{2\pi(N-k)j_2}{N}\right)\cos\left(\frac{2\pi(N-k)j_1}{N}\right)$$

$$\left. + \hat{a}_{\tilde{x}}[k] \sin\left(\frac{2\pi k j_2}{N}\right)\sin\left(\frac{2\pi k j_1}{N}\right)\right]$$

$$= \sum_{k=0}^{N-1} \lambda_k u_k\left[j_1\right] u_k\left[j_2\right] \blacksquare$$

then when $|k| < k_c$, which means k =0:

$$u_k = \frac{1}{\sqrt{N}}\begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix}$$

When $1 \leq K \leq \frac{N-1}{2}$,

$$u_k = \quad \sqrt{\frac{2}{N}}\begin{bmatrix} 1 \\ \cos\left(\frac{2\pi k}{N}\right) \\ \cdots \\ \cos\left(\frac{2\pi k(N-1))}{N}\right) \end{bmatrix}$$

When $\frac{N-1}{2} \leq K \leq N-1$,

$$u_k = \sqrt{\frac{2}{N}} \begin{bmatrix} 0 \\ \sin\left(\frac{2\pi k}{N}\right) \\ \cdots \\ \sin\left(\frac{2\pi k(N-1))}{N}\right) \end{bmatrix}$$

4. (Household electricity usage) Given the rise of smart electricity meters and the wide adoption of electricity generation technology like solar panels, there is a wealth of electricity usage data available. Here we will explore a household electricity usage dataset which represents a time series of power-related variable.

(a) Load the data file household_power_consumption_days.csv in Python and plot it. Compute and plot the autocorrelation of these data using Definition 4.1 in the note for the first 3 weeks and first 3 years. Provide explanations for each of the plots.

In [12]:
```
import statsmodels.api as sm
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from pandas.plotting import register_matplotlib_converters
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
register_matplotlib_converters()

df = pd.read_csv('household_power_consumption_days.csv')
```

In [13]:
```
df.head()
```

Out[13]:

	datetime	Global_active_power(kW)
0	2006-12-16	1209.176
1	2006-12-17	3390.460
2	2006-12-18	2203.826
3	2006-12-19	1666.194
4	2006-12-20	2225.748

In [14]:
```
df["datetime"]= pd.to_datetime(df.datetime)
```

In [17]:
```
df.set_index("datetime",inplace=True)
```

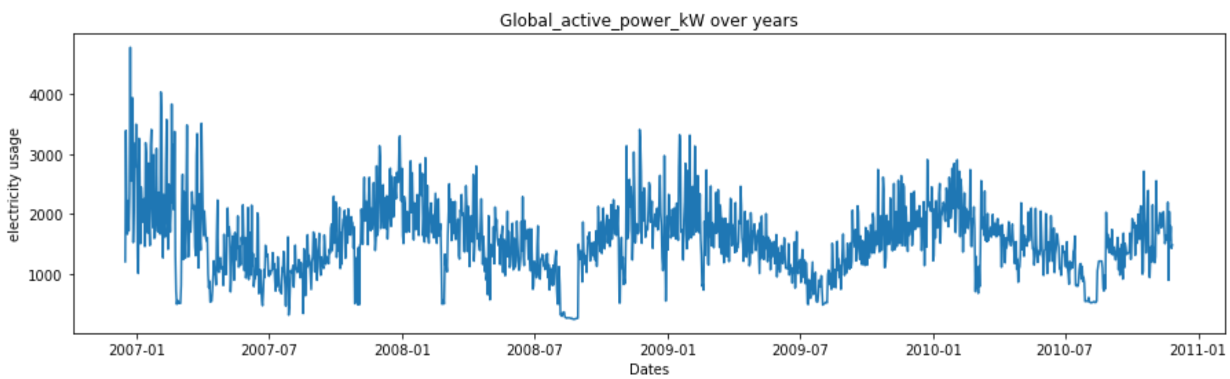
In [18]:
```
df.head()
```

Out[18]:

	Global_active_power(kW)
datetime	
2006-12-16	1209.176
2006-12-17	3390.460
2006-12-18	2203.826
2006-12-19	1666.194
2006-12-20	2225.748

In [23]:
```
plt.figure(figsize=(15,4))
plt.plot(df["Global_active_power(kW)"])
plt.xlabel('Dates')
plt.ylabel('electricity usage')
plt.title('Global_active_power_kW over years')
```

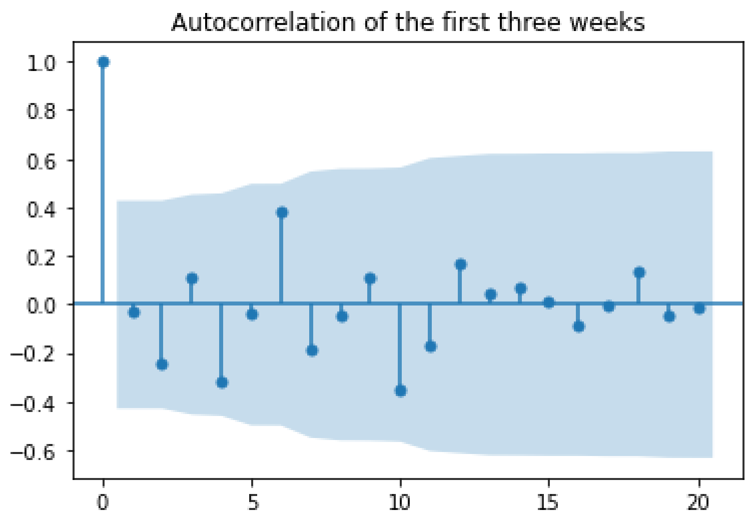
Out[23]:
```
Text(0.5, 1.0, 'Global_active_power_kW over years')
```



In [33]:
```
power_vector_week=df.iloc[0:21, 0].values
autocorr_week=sm.tsa.acf(power_vector_week)
acf_plot = plot_acf(power_vector_week,lags=len(power_vector_week)-1)
plt.title('Autocorrelation of the first three weeks')
```

```
/Users/herculesgao/opt/anaconda3/lib/python3.8/site-packages/statsmodels/
tsa/stattools.py:652: FutureWarning: The default number of lags is changi
ng from 40 tomin(int(10 * np.log10(nobs)), nobs - 1) after 0.12is release
d. Set the number of lags to an integer to  silence this warning.
  warnings.warn(
/Users/herculesgao/opt/anaconda3/lib/python3.8/site-packages/statsmodels/
tsa/stattools.py:662: FutureWarning: fft=True will become the default aft
er the release of the 0.12 release of statsmodels. To suppress this warni
ng, explicitly set fft=False.
  warnings.warn(
```

Out[33]:
```
Text(0.5, 1.0, 'Autocorrelation of the first three weeks')
```



In [30]:
```
df[df.index<"2009-12-17"]
```

Out[30]:

	Global_active_power(kW)
datetime	
2006-12-16	1209.176
2006-12-17	3390.460
2006-12-18	2203.826
2006-12-19	1666.194
2006-12-20	2225.748
...	...
2009-12-12	1816.380
2009-12-13	2138.814
2009-12-14	1794.862
2009-12-15	1940.116
2009-12-16	1394.874

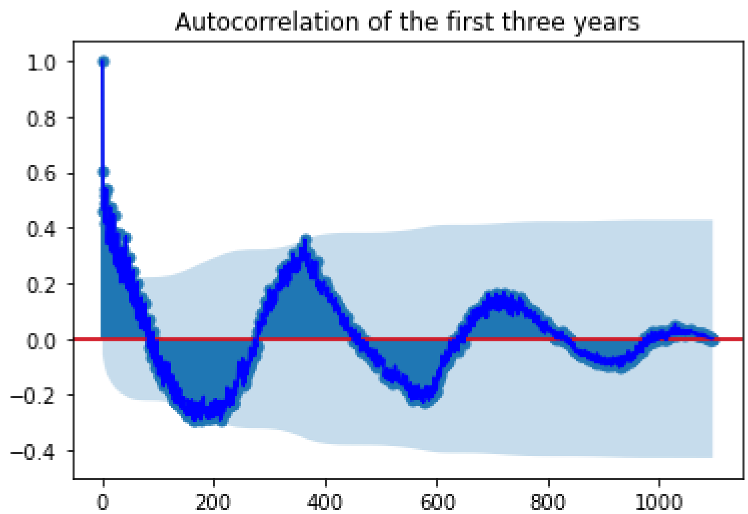
1097 rows × 1 columns

In [48]:
```
plt.figure(figsize=(15,4))
power_vector_year=df.iloc[0:1097, 0].values
autocorr_year=sm.tsa.acf(power_vector_year,nlags=len(power_vector_year))
acf_plot = plot_acf(power_vector_year,lags=len(power_vector_year)-1)
plt.axhline(y=0,color='r',linestyle='-')
plt.title('Autocorrelation of the first three years')
plt.plot(autocorr_year,'b-')
```

```
/Users/herculesgao/opt/anaconda3/lib/python3.8/site-packages/statsmodels/
tsa/stattools.py:662: FutureWarning: fft=True will become the default aft
er the release of the 0.12 release of statsmodels. To suppress this warni
ng, explicitly set fft=False.
  warnings.warn(
```

Out[48]:
```
[<matplotlib.lines.Line2D at 0x7fd590e946a0>]
```

```
<Figure size 1080x288 with 0 Axes>
```

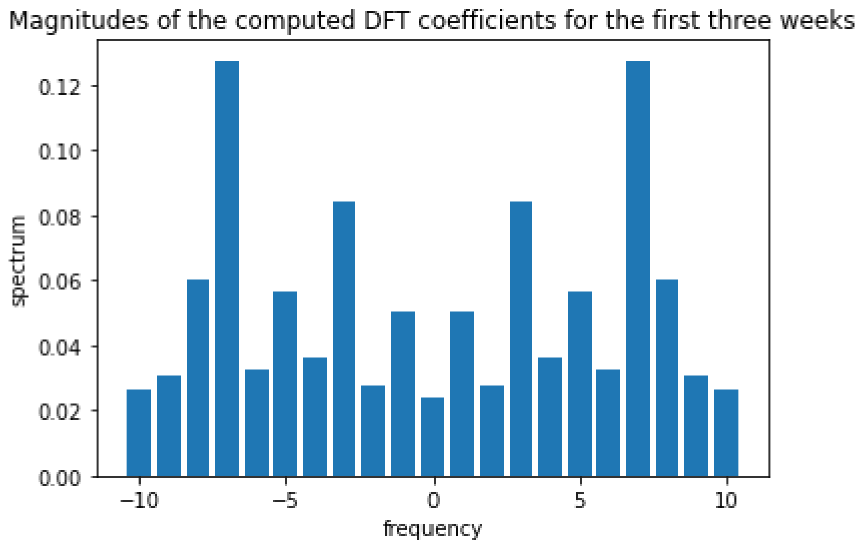


we can tell tha autocorrelation turnes to be lower if the timespan is longer (gap of time).

(b) Apply DFT to the autocorrelation and give a plot of the magnitudes of the computed DFT coefficients. What do you notice from the plot? (Hint: You might want to zoom in to see spikes around zero frequency clearly.)

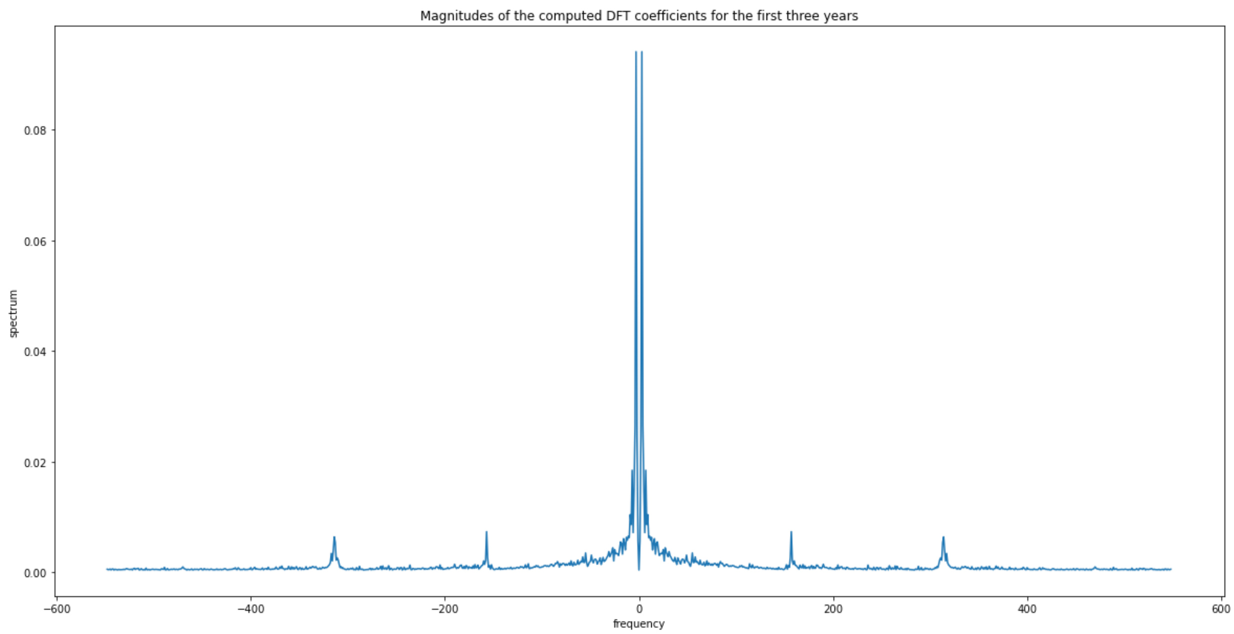
In [36]:
```
spectrum_week = np.fft.fft(autocorr_week)
spectrum_week =np.fft.fftshift(spectrum_week)/len(autocorr_week)
freq =np.fft.fftfreq(len(autocorr_week),1/len(autocorr_week))
freq = np.fft.fftshift(freq)
plt.xlabel('frequency')
plt.ylabel('spectrum')
plt.title('Magnitudes of the computed DFT coefficients for the first three
plt.bar(freq,np.abs(spectrum_week))
```

Out[36]: <BarContainer object of 21 artists>



In [40]:
```
plt.figure(figsize=(20,10))
spectrum_yr = np.fft.fft(autocorr_year)
spectrum_yr =np.fft.fftshift(spectrum_yr)/len(autocorr_year)
freq_yr =np.fft.fftfreq(len(autocorr_year),1/len(autocorr_year))
freq_yr = np.fft.fftshift(freq_yr)
plt.xlabel('frequency')
plt.ylabel('spectrum')
plt.title('Magnitudes of the computed DFT coefficients for the first three
plt.plot(freq_yr,np.abs(spectrum_yr))
```

Out[40]: [<matplotlib.lines.Line2D at 0x7fd5afa59e50>]



For the first three weeks and first three years graph, the magnitudes reach the lowest point on 0 frequency. However, for three years graph, there are some small spikes spread from the center around -200, 200 and -300 and 300.

(c) Based on what you have observed, what are the advantages and disadvantages of modeling this time series data as stationary?

Advantage: As we can longer time series (timespan) the autocorrelation will be lower and the data will be consistent adn nnot change over time. It can help to clean the data to find the true signal.

Besides, it can then be the basis for us to forecast the data.

Disadvantage: It is hard for us to obtaining appropriate measures to accurately identifying the correct model to respesent the data. And it can lead to unrealistic forecasts of values far beyond the temporal horizon of study.