Instructor Lab Notes: Data Science in Radio Astronomy I

Overview and Objectives

This lab introduces students to foundational concepts in data science in radio astronomy, focusing on analyzing spectral data of the 21cm hydrogen line in the Milky Way. Through hands-on coding activities in Python with a Jupyter notebook, students will work will both simulated and real data from the Allen Telescope Array (ATA) to explore the signal processing and data science techniques needed to study spectral radio data. The lab is divided into two main parts:

- **Part 1:** Students will analyze a simulated noisy narrowband signal, experimenting with fast Fourier transforms (FFTs), spectral resolution, and noise reduction techniques to understand their effect on signal detection.
- **Part 2:** Students will analyze real hydrogen line data from the ATA, performing Gaussian curve fitting, Doppler velocity conversion, and use these to interpret the physical properties of the gas, such as velocity and temperature.

Objectives: By the end of this lab, students should be able to:

1. Understand the importance of spectral resolution in identifying narrowband signals, and its importance for SETI research.
2. Apply Welch's method to reduce noise in spectral data, and evaluate its effectiveness.
3. Generate and interpret power spectral density (PSD) plots for both simulated and real data.
4. Use curve fitting techniques to analyze hydrogen line data and extract it's physical properties, including amplitude, center frequency, and line width.
5. Perform a frequency to Doppler velocity conversion and calculate the radial velocity of the hydrogen gas in the Milky Way.
6. Reflect on the broader application of these techniques to SETI, studies of galactic structure, and radio astronomy.

Pre-Lab Preparation

Materials Needed:

- Students will use the lab manual in tandem with the provided Jupyter notebook `DataScience1Lab.ipynb` for the activities in this lab. An instructor copy of the Jupyter notebook DataScience1Lab-InstructorCopy.ipynb is also included, containing the completed and functional version of the code for reference. These can both be found in the **Teaching Resources** section of the website for this module. Ensure students have the correct software and packages installed prior to starting the lab, including:
  - Jupyter: Installing Jupyter

- NumPy: Installing NumPy
- Matplotlib: Installing Matplotlib
- SciPy: Installing SciPy

- The data for part 2 of the lab includes 20 integrated scans of the hydrogen line along the galactic plane from the ATA; the file name indicates the galactic coordinates of the pointing (l,b). Assign these pointings to each student/group, ensuring coverage across the galactic plane.

Instructor Setup:

- Ensure all required software is installed and tested.
- Verify that data files for Part 2 are correctly formatted and accessible.
- Familiarize yourself with the Welch method and Gaussian fitting processes.

Key Teaching Points

Part 1:

- Highlight the role of spectral resolution in SETI.
- Explain the trade-offs between resolution, computation, and noise reduction.
- Introduce Welch's method and its practical applications for global noise reduction.

Part 2:

- Demonstrate the process of Gaussian fitting.
- Link fitted parameters (amplitude, center, width) to physical properties of hydrogen gas.
- Emphasize the Doppler effect and its role in calculating gas velocities.

Common Student Challenges

- **FFT and PSD:** Students may struggle to interpret PSD plots. Encourage discussions about the meaning of peaks and noise floors.
- **Gaussian Fitting:** Setting initial guesses can be confusing; provide tips for estimating amplitude, center, and width.
- **Code Debugging:** Ensure students understand error messages, particularly with curve fitting functions.
- **Frequency-Velocity Conversion:** Clarify the Doppler formula and its implementation in code.

Instructor Tips

- **Encourage Exploration:** Allow students to experiment with FFT sizes and Welch segment lengths. Suggest varying Gaussian parameters to observe changes.

- **Guide with Questions:** Ask why narrowband signals are more detectable with higher resolution. Discuss astrophysical implications of Gaussian models.
- **Provide Checkpoints:**
  - After Part 1, ensure students can identify narrowband signals at different spectral resolutions.
  - Before Part 2, verify that students can mask the region of interest and set initial guesses.

Workflow and Checkpoints

1. **Part 1: Exploring Spectral Resolution**
   - **Checkpoint 1:** Students successfully generate and observe PSD plots at low and high spectral resolutions.
   - **Checkpoint 2:** Students apply Welch's method and understand its effect on noise reduction.
2. **Part 2: Hydrogen Line Analysis**
   - **Checkpoint 3:** Students define the region of interest and set initial guesses for Gaussian fitting.
   - **Checkpoint 4:** Students calculate velocities and temperatures, interpreting results in a physical context.

Example Discussion Questions

Part 1:

- How does increasing FFT size impact the clarity of narrowband signals?
- Why is Welch's method better for global noise reduction compared to filtering?

Part 2:

- What does the width of a Gaussian line reveal about the gas's temperature or turbulence?
- How do your velocity results compare with the expected values for galactic rotation?

Relevance to SETI and Astronomy

Discuss why narrowband signals are key in SETI, linking this to the lab's focus on spectral resolution and noise reduction. Connect Gaussian fitting of the 21cm line to real-world applications in mapping galactic structure and detecting extraterrestrial signals.

Post-Lab Reflection

Encourage students to reflect on how the techniques learned could be applied to other problems in radio astronomy. Discuss advancements in data science tools that could enhance future SETI research.