Signal Modeling, Statistical Inference and Data Mining in Astrophysics

Spring 2013

Lectures 23

Pulsar Searches (Dusty Madison)

Signal Detection (Shami Chatterjee)

An Introduction to MCMC for Machine Learning and Learning Algorithms (C. Andrieu et al.)

Reading:

- Efficient Monte Carlo Methods, Chapter 29 of Information Theory, Inference, and Learning Algorithms (D. MacKay)
- Monte Carlo Methods, Chapter 29 of Information Theory, Inference, and Learning Algorithms (D. MacKay)
- Selection Applied to Computation (S. Forrest)
- Genetic Algorithms: Principles of Natural
- An Introduction to MCMC for Machine Learning

- (Note the complete book and individual chapters can be viewed at http://www.inference.phy.cam.ac.uk/itila/)

- Genetic Algorithms: Principles of Natural

Spring 2013

Merging in Astrophysics

Signal Modeling, Statistical Inference and Data Mining
Dusty Madison

Discovering Pulsars
Pulsar Search Procedure

1. **Eye**: Candidates by Inspect Promising
2. **Period**: Most Significant according to the Period
3. **Fold Time Series**: S/N Threshold
   - Events above Some Time Series Flag
   - Smooth Dispersed
4. **S/N Threshold**: Clusters above Some Time Series Flag
5. **Single Pulse Searches**: Single Pulse
6. **S/N Threshold**: DM Bright
   - Band and Zero Band Narrow
   - Likely RF
7. **Channels**: Frequency
   - Initial DM Combining
8. **Frequency-Time Values**: Values in Collected Intensity
9. **Frequency-Time Values**: Frequency-Time Values
10. **Periodicity Searches**: Periodicity
11. **Famiili**: Discovery/Camera Report
12. **Famiili**: Discover Candidates
13. **Inspect Promising**: Period
14. **Inspect Promising**: Detect Most Significant
15. **Inspect Promising**: Events above Some Time Series Flag
16. **Inspect Promising**: Smooth Dispersed
Time domain resolution becomes limited by dispersive spreading across lowest frequency channels.

Floating point operations needed from $N^2$ to $N \log N$.

With some care, the so-called "free algorithm" (Taylor 1974) can reduce the number of conceptually and computationally easy to implement.

Lohmer & Krämer 2004

\[ \left( \frac{MH}{f} \right) - \left( \frac{MH}{f} \right)^{-2} \left( \frac{cm-3 pc}{DM} \right) \left( \frac{10^6 ms}{\text{sec}} \right) = 2 \]

Incoherent Dispersion

Original data is a two-dimensional array of intensity measurements at different times and frequencies.

Incoherent Dispersion
Not used when searching for pulsars, but preferred when trying to do high precision timing.

This scheme retains the native time resolution of your instrument.

\[
\frac{2\pi e}{df} = D
\]

\[
\frac{\partial f}{\partial t} = (f + 0f)H
\]

\[
(f + 0f)H(f + 0f)^{\text{lim}}\Lambda = (f + 0f)\Lambda
\]

\[
\int_{-o-f}^{o+f} (f)^{\text{lim}}\Lambda = (t)^{\text{lim}} \quad \int_{-o-f}^{o+f} (f)\Lambda = (t)\Lambda
\]

Interstellar dispersion acts as a transfer function. We can explicitly deconvolve it.

Receiver. This is known as baseline sampling.

Record the complex voltagre (amplitude and phase) of the radio signal as it reaches your.

Coherent Deispersion
normalization are used to whiten the data.

Running median subtraction and local RMS assess the significance of spectral lines.

**Colored Noise**: Undermines one's ability to

\[
\int_{0}^{\pi/2} \frac{1}{\sqrt{1 - (\cos^2 \phi)}} \approx \frac{\pi}{2} \left( 1 + \frac{m/2}{\nu} \right)
\]

Zero padding and Fourier domain interpolation. 

Error function in frequency bins. For workarounds, see into nearby frequency bins. For signal processing, see one of your Fourier bins, the signal power bleeds has a frequency unequal to the central frequency of

Scalloping: If the periodic signal you're looking for 

via special techniques.

from pulsars. With discrete Fourier methods, several problems must be dealt with

Fourier methods are well suited for searching time series for periodic pulsations

**Periodicity Searches**
The fundamental of a periodic signal has a small duty cycle (pulse width divided by period). If a periodic signal has a small duty cycle, the fundamental and many harmonics.

Incoherent Harmonic Summing can be used to add power in harmonics back into the fundamental. This can be done roughly N times where N is the inverse of the duty cycle. It leads to a \( N^{1/2} \) enhancement to the S/N of the fundamental.
searched over.

observation and only 1 parameter must be
approximated as constant during an
Often in practice, the radial acceleration
dimensional parameter space of orbital
For pulsars in binaries, a search over a 5+
parameters is necessary (in principle).
Effects from Earth's orbit and rotation
Radial velocity:
compressed according to a model for the
to fix this, the time-axis is stretched or

problematic since longer observations are used to spot dimmer pulsars.
and the efficacy of Fourier techniques will be deteriorated. This can be especially
during an observation, the rotational period of the pulsar will be Doppler shifted
If the radial velocity between a pulsar and an observatory changes substantially

Time-Domain Resampling
The radial velocity, compressed according to a model for the radial velocity, must be corrected for: 

\[ \tau(t) = \tau_0(t^0 + v_u) \]

To fix this, the time-axis is stretched or compressed over. 

Observations are used to spot dimmer pulsars. The efficacy of Fourier techniques will be deteriorated. This can be especially problematic since longer observations are used to spot dimmer pulsars. During an observation, the rotational period of the pulsar will be Doppler shifted. If the radial velocity between a pulsar and an observatory changes substantially, the parameters must be corrected (in principle).
Both are astrophysically interesting objects and we want techniques to find them out. Both are astrophysically interesting objects and we want techniques to find them out. They are approaching the so-called “death line” and may literally be stripping fields. They are approaching the so-called “death line” and may literally be stripping normal pulsation. For others, pulse emission stochastically turns on and off. These normal pulsations. For others, pulse emission stochastically turns on and off. These may be being created by altogether different emission mechanisms than others.

Not all pulsars are detectable through their periodic pulse emission. For some pulsars, only occasionally, very bright pulses are detectable (100-1000 times brighter than other
Single Pulse Searches

Cordes et al. (2006)

Generated.

Thresholding and a list of candidate events is
These smoothed time series are subjected to S/N
incoherently while any signal adds directly.
neighboring samples). The noise adds
degraded through smoothing (adding
the time resolution is often intentionally
As narrower pulses are more readily detectable,

does as \( W^{-1/2} \).

For fixed pulse area (S) (S = const.), the S/N

\[
\frac{W}{\sqrt{2 \pi}} \sqrt{\frac{S_{\text{peak}}}{S}} = N/S
\]

resolution
white noise, if \( W \) is equal to the time sampling
amplitude \( S_{\text{peak}} \) in the presence of Gaussian
for rectangular pulses of width \( W \) and
16 Apr 2013

Source Detection

Source Identification

• Identify a source in the presence of noise.
• Localize it as best as possible given instrumental resolution.

Resolution \sim \frac{Beam 
Size}{Signal-to-Noise 
Ratio}
Source Detection Algorithms

16 Apr 2013

• Identify “real” sources in the presence of noise in an image.
• Many different approaches to the problem, optimized for different image domains and properties of interest.
• Examples: SFind, IMSAD, SExtractor, Aegean, BlobCat, etc. etc. etc.

• Different trade-offs between speed and completeness.
• False positives and completeness.

• Do we care about extended sources or want compact components?

• Various Photon events (Poisson statistics) and sparse detector arrays vs. Fourier transformed visibilities + CLEANing

• Blurring from diffraction and seeing

• e.g., Optical, Radio, X-ray images?

• Identifying “real” sources in the presence of noise in an image.

Source Detection Algorithms
<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Source Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>16 Apr 2013</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Basic Image Transformations**

<table>
<thead>
<tr>
<th>SD</th>
<th>Radio</th>
<th>X-ray</th>
<th>X-ray and infrared</th>
<th>Multi-band</th>
<th>Hi-band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SD</td>
<td>Radio</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2SD</td>
<td>Audio</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3SD</td>
<td>Audio</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4SD</td>
<td>Audio</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5SD</td>
<td>Audio</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>6SD</td>
<td>Audio</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Multi-scale applications**

<table>
<thead>
<tr>
<th>SD</th>
<th>Audio</th>
<th>Radio</th>
<th>X-ray</th>
<th>X-ray and infrared</th>
<th>Multi-band</th>
<th>Hi-band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SD</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2SD</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>3SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>4SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>5SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>6SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**WE**

<table>
<thead>
<tr>
<th>SD</th>
<th>Audio</th>
<th>Radio</th>
<th>X-ray</th>
<th>X-ray and infrared</th>
<th>Multi-band</th>
<th>Hi-band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SD</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2SD</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>3SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>4SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>5SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>6SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Image applications**

<table>
<thead>
<tr>
<th>SD</th>
<th>Audio</th>
<th>Radio</th>
<th>X-ray</th>
<th>X-ray and infrared</th>
<th>Multi-band</th>
<th>Hi-band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SD</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2SD</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>3SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>4SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>5SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>6SD</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Source Identification: Flooding Fill

- **Background pixels**
- **Pixels in the island**
- **Pixels being considered**
- **Unprocessed pixels**

**Diagram Description**:
- **A**: Initial state with unprocessed pixels.
- **B**: Processed pixels in the island.
- **C**: Further progression of processing.
- **D**: Complete processing of the island.
- **E**: Additional states showing the process.
- **F**: Final state with all pixels processed.
Source detection with Aegean

Source detection
16 Apr 2013

Aegean: Part of Paul Hancock’s thesis work with Murphy, Gaensler, et al.


A curvature map is created from the input noise. The curvature noise (\sigma_{\text{curv}} (\sigma_{\text{map}})) is calculated. The image is set by the user.
Source detection with Aegean

Source detection

16 Apr 2013

Source Detection with Aegean

16 Apr 2013

Aegean: Part of Paul Hancock’s thesis work with Murphy, Gaensler, et al.


Each of the Gaussian components

...converge to an acceptable solution. Red ellipses show the constraints that ensure the fits will converge jointly with appropriate...
Source detection with Aegean: Performance Comparisons
Sourcefinders compared on real VLA observations (Mooley et al. 2013, arXiv:1303.6282)

- Bottom: a wide field area
- Mid: a source with sidelobes
- Top: a blended source

Comparison of source finder performance vs DR2 catalog for

Source Finding: Real World Performance
For sources in each S/N ratio bin (top), Mooley et al. compare completeness (middle) and reliability (bottom) of source finders. Highly blended sources cause dip at SNR $\sim 70$. 
Bottom line for Source Detection

- Source finders are typically optimized for different domains.
- Trade-off between completeness and reliability (false positives).
- All source finders are pretty good. No source finder is perfect. If you’re using one, understand its strengths and weaknesses first.
- There’s room to do better!