Lecture 2

Linear least squares, matched filter with uncorrelated noise

Mechanics

- Office hours currently set to 1-2:30 on Tuesday. Is that OK?
- Problem sets - would you prefer due on Mondays or on Wednesdays?

Poisson Example

- You’ve discovered a new object! Your theorist friend has a model, and thinks it will flare randomly, with a mean rate of once per month.
- You, an observer, think your friend is wrong. How long would you need to observe to rule out their model?

Expectation/Variance

- expectation of a random variable $<x>$ = average value of many realizations. Mathematically=$\int x p(x) dx$.
- Variance is scatter (squared) about the mean - $Var(x)= <(x-<x>)^2>$. Show this equals $<x^2>-<x>^2$.
- Expectation and variance fully describe a Gaussian random variable.
- Let c be a constant. What is $<cx>$? What is $Var(cx)$?
- Let y be another random variable. What is $<x+y>$? What is $Var(x+y)$?

Covariance

- Let's look at Var(x+y): $\langle (x+y)^2\rangle-\langle x+y\rangle^2=\langle x^2+2xy+y^2\rangle-\langle x\rangle^2-2\langle x\rangle\langle y\rangle-\langle y\rangle^2=\langle x^2\rangle-\langle x\rangle^2+2\langle xy\rangle-2\langle x\rangle\langle y\rangle+\langle y^2\rangle-\langle y\rangle^2$.

- =Var(x)+Var(y)+2Cov(x,y), where covariance defined to be $\langle xy\rangle-\langle x\rangle\langle y\rangle$.

- If x and y are uncorrelated, what is the variance of (x+y)? And of (x-y)?

Linear Least-Squares

- Recall matrix description of $-2\ln(L)=\chi^2$ is $(d-p)^T N^{-1}(d-p)$ for data values d and model prediction p. If p is true model, then $\langle d\rangle=p$.

- Take case where predicted values depend linearly on a set of model parameters: $p=Am$, so $\langle d\rangle=Am$

- $\chi^2=(d-Am)^T N^{-1}(d-Am)$. What are the dimensions of the various things?

- What values of *m* are the “best fit”?

LLS cont'd

- Maximizing likelihood equivalent to minimizing χ^2. Take gradient w.r.t what?

- $\nabla\chi^2=-A^TN^{-1}(d-Am)=0$. $A^TN^{-1}Am=A^TN^{-1}d$.

- How can you solve this? Could you multiply by various combinations of A^{-1T} and N to get $m=A^{-1}d$?

- Let's say we have two data sets d_1 and d_2 with best-fit solutions m_1 and m_2. What are the best-fit parameters if we fit (d_1+d_2)?

Parameter Errors

- Since we are scientists, we need errors on our parameters. If we subtract the true model from the data, we can look at the covariance of what's left, $\langle mm^T \rangle$ (why not $\langle m^Tm \rangle$?)
- Let's show that $m=(A^TN^{-1}A)^{-1}A^TN^{-1}d$ gives $\langle mm^T \rangle=(A^TN^{-1}A)^{-1}$.
- What are the standard deviations of the parameters?

Stability

- Let's fit polynomials. How did that go? Why?
- Let's ignore N for now, and use SVD of A - $A=USV^T$, where U is orthogonal (and rectangular), S is diagonal, and V is orthogonal (and square).
- $ATA = VSU^TUSV^T=VS^2V^T$. $(ATA)^{-1}=VS^{-2}V^T$, so if an entry of S was very small, it becomes very large.
- By writing out an analytic cancellation, we can get rid of one copy of S, making problem better behaved numerically: $VS^2V^Tm=VSU^Td$. V and S are square and invertible (usually!), so leaves us with $SV^Tm=U^Td$. No squaring of S… or, $m=VS^{-1}U^Td$.
- For polys, real solution is to switch bases to e.g. Legendre, Chebyshev… Good idea to check condition number (ratio of largest to smallest entries of S) before trying LLSQs.

Worked Example

- What is the best-fit mean and error for a set of uncorrelated gaussian variables with same mean but individual errors?

- $A=?$ Show that $A^T N^{-1} A = \sum(\sigma_i^{-2})$, $A^T N^{-1} d = \sum d_i/\sigma_i^2$.

- Define weights $w_i=\sigma_i^{-2}$. Then $m=\sum w_i d_i/\sum w_i$. Variance of our estimator is $1/\sum w_i$.

Worked Example 2

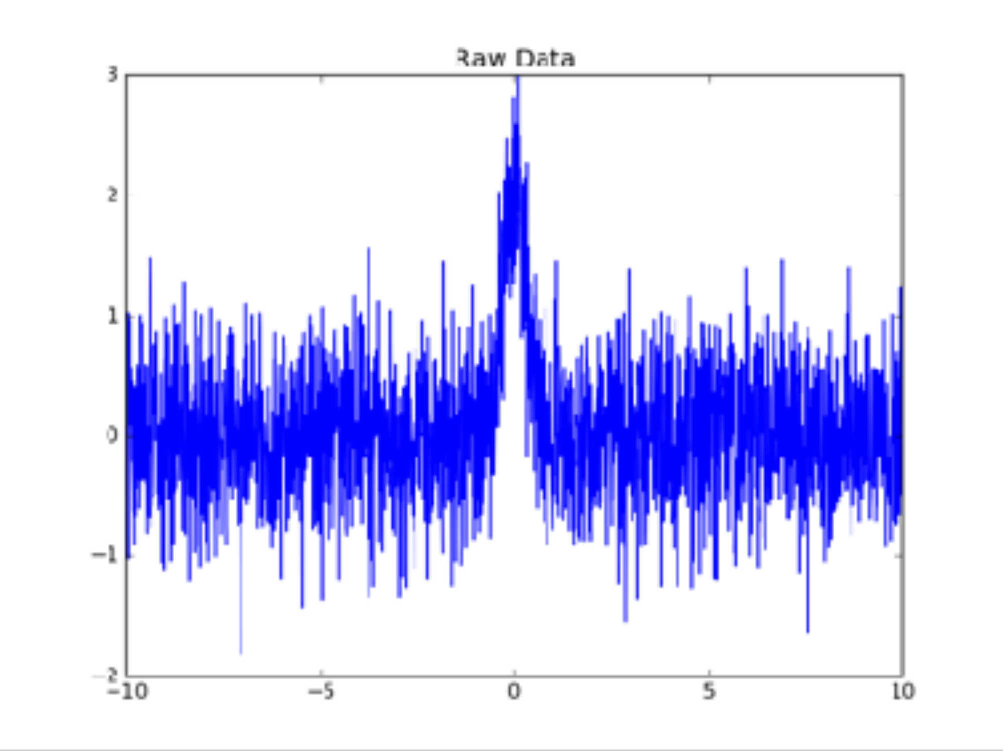
- Let's fit a 1-parameter template to data, but possibly want to shift it (e.g. fitting a for a source amplitude at various positions).
- If A is n by 1, then $A^T N^{-1} A$ is a scalar.
- now we have $m = A^T N^{-1} d / (A^T N^{-1} A)$.
- If N is "constant" (i.e. $N_{ij} = f(i-j)$) and we shift the template $A_i \to A_{i+\delta}$, how does the denominator depend on δ?
- Up to an overall constant, we can make $N^{-1}d$ and dot it against the various shifted A's.

Example 2 ctd.

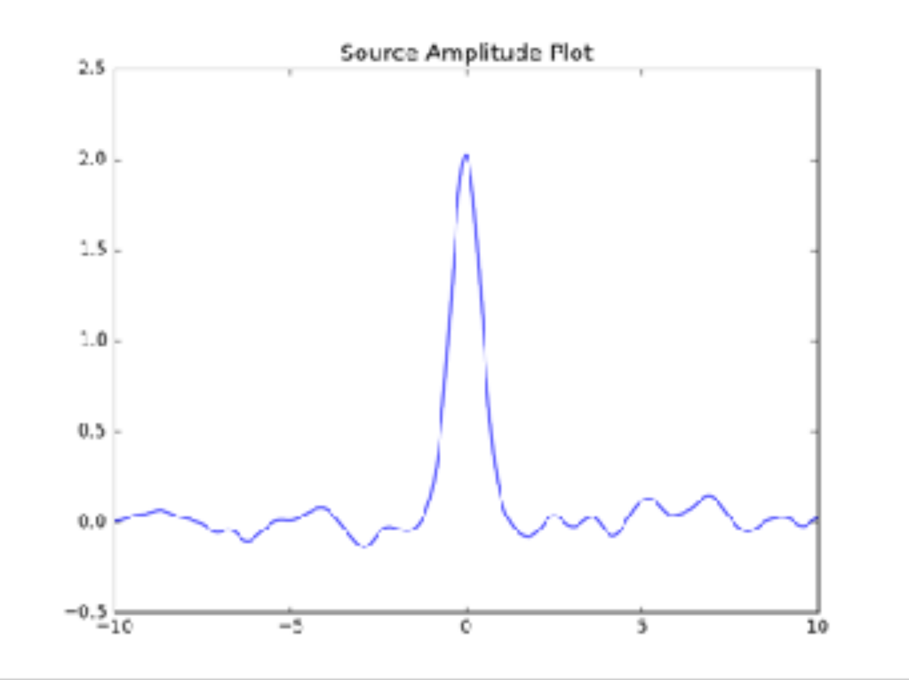
- What if we wanted to plot amplitude in terms of SNR?
- recall $Var(m)=(A^TN^{-1}A)^{-1}$, so $\sigma(m)=(A^TN^{-1}A)^{-1/2}$.
- So, $SNR=signal/noise=A^TN^{-1}d/A^TN^{-1}A/(A^TN^{-1}A)^{-1/2}=A^TN^{-1}d/\sqrt{(A^TN^{-1}A)}$.

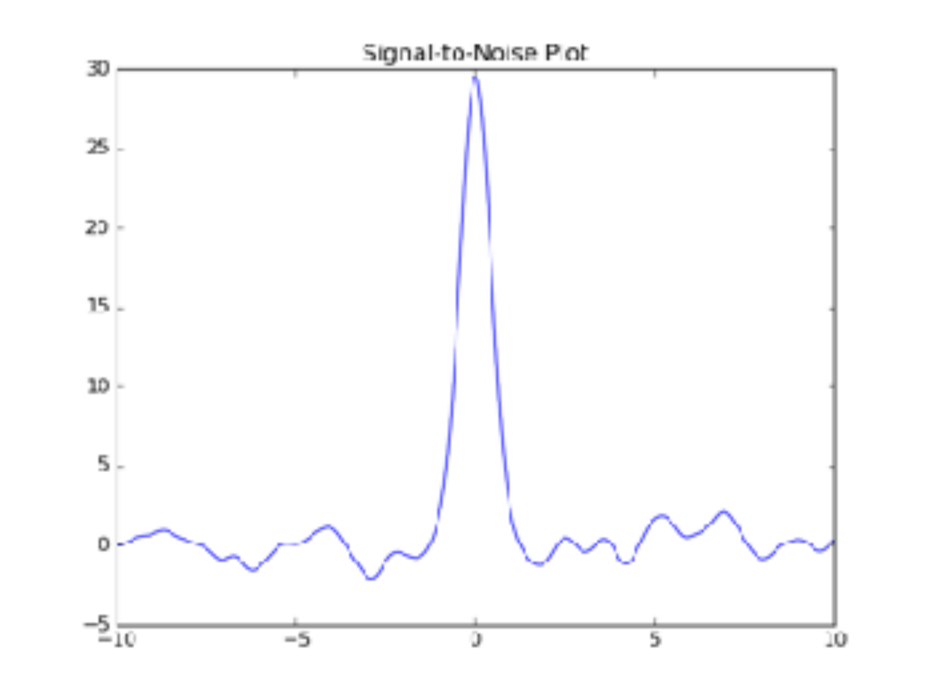
Code Example

```
import numpy
from matplotlib import pyplot as plt

dx=0.01
noise=0.5
Ninv=1.0/noise**2
x=numpy.arange(-10,10,dx)
n=len(x)


x0=0
amp_true=2.0

sig=0.3
template=numpy.exp(-0.5*(x-x0)**2/sig**2)

dat=template*amp_true+numpy.random.randn(n)*noise

snr=numpy.zeros(n)
amp=numpy.zeros(n)
dat_filt=Ninv*dat
denom=(numpy.dot(template,Ninv*template))
rt_denom=numpy.sqrt(denom)
for i in range(n):
    template=numpy.exp(-0.5*(x-x[i])**2/sig**2)
    rhs=numpy.dot(template,dat_filt)
    snr[i]=rhs/rt_denom
    amp[i]=rhs/denom
```




```
plt.clf();
plt.plot(x,snr);
plt.title('Signal-to-Noise Plot')
plt.savefig('snr_plot.png')

plt.clf();
plt.plot(x,amp)
plt.title('Source Amplitude Plot')
plt.savefig('amp_plot.png')

plt.clf();
plt.plot(x,dat)
plt.title('Raw Data')
plt.savefig('dat_raw.png')
```





But wait!

- We took $\sum d(t)a(t-\tau)$. But, this is just the correlation of d with a. We can do this quickly using Fourier transforms.

- Alternatively, let $a^{\lozenge}=a(-t)$. Then this is $\sum d(t)a^{\lozenge}(\tau-t)=d\otimes a^{\lozenge}$. By convolution theorem, this is $IFT(FT(d)*FT(a^{\lozenge}))$.

- However, $FT=\sum f(x)\exp(-2\pi ikx/N)$. $FT(f(-x))=\sum f(-x)\exp(-2\pi ikx/N)=\sum f(x)\exp(2\pi ikx/N)=F^{*}(k)$. So, our output is just $IFT(FT(d)FT^{*}(a))$.

- NB - if a is symmetric, when A(k) is real (why?) and we can skip the conjugates.

How about changing noise?

- Often we're searching for things where noise is varying - say, searching for sources in a map that is deeper in the middle.
- Top: $A^T N^{-1} d$ is easy - correlate A with $N^{-1}d$.
- Bottom: $A^T N^{-1} A$ we can correlate N^{-1} with A^2.
- Amplitude stays ratio of the two (bottom is now a vector instead of a scalar), and SNR is top/$\sqrt{}$(bottom)

Matched Filter

- We are beginning on a very widespread class of techniques called *matched filters*.
- In various forms, they show up in GW analysis, photometry in optical images, finding galaxy clusters in CMB maps, finding radar return echos…
- Extendable to non-independent noise, multiple (possibly correlated) simultaneous/multi-frequency datasets, many others.
- All arises from writing down χ^2 and minimizing. Make a habit of this.

Linear Algebraing up χ^2 (from last slides)

- Usual expression is $\sum(x_i-\mu_i)^2/2\sigma_i^2$
- Let N be diagonal matrix with $N_{ii}=\sigma_i^2$.
- Element-wise, $(x-\mu)^T N^{-1}(x-\mu)$ is identically χ^2.
- I can put orthogonal matrices ($V^T=V^{-1}$) in without changing anything: $(x-\mu)^T V^T V N^{-1} V^T V(x-\mu)$.
- In new, rotated coordinates: $x \rightarrow Vx$, $\mu \rightarrow V\mu$, $N \rightarrow VNV^T$, χ^2 remains unchanged. Show that expectation of (rotated) x_i noise times x_j noise = (rotated) N_{ij}?

Stationary Noise

- With this plus posted note, we can work out N^{-1} for correlated but stationary noise.
- N^{-1} operator becomes Fourier divide by noise power spectrum transform.
- Numerator becomes $IFT(FT(a)^{*}FT(d)/N^{FT})$