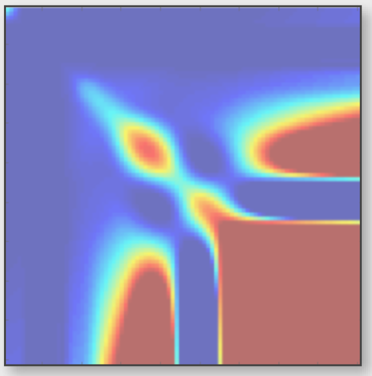
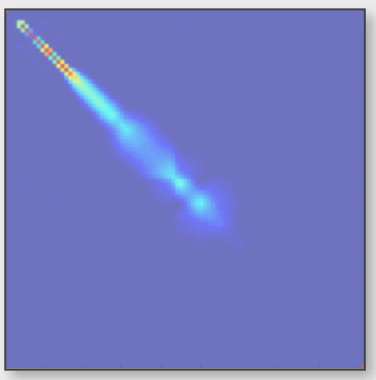




An introduction to geophysical inversion, with comparisons to analytical inversion

Andrew Ganse

Dept. of Earth & Space Sciences (ESS-UW)

Applied Physics Laboratory (APL-UW)

`aganse@apl.washington.edu`

`http://staff.washington.edu/aganse`

UW Math department Inverse Problems seminar

17 Oct 2007

Briefly before I begin…

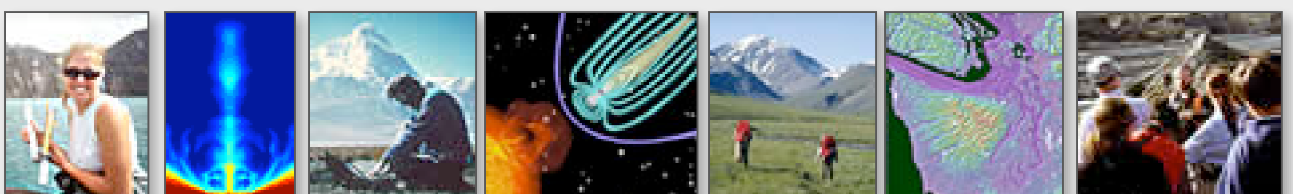
A pretty good beer!

"Jim! I'm a geophysicist, not a mathematician…"

Please correct me on any missteps in my mathematical descriptions

A little bit about ESS-UW

Department of Earth and Space Sciences, University of Washington



graphics credits: ESS-UW

A little bit about APL-UW

Applied Physics Laboratory, University of Washington

graphics credits: APL-UW

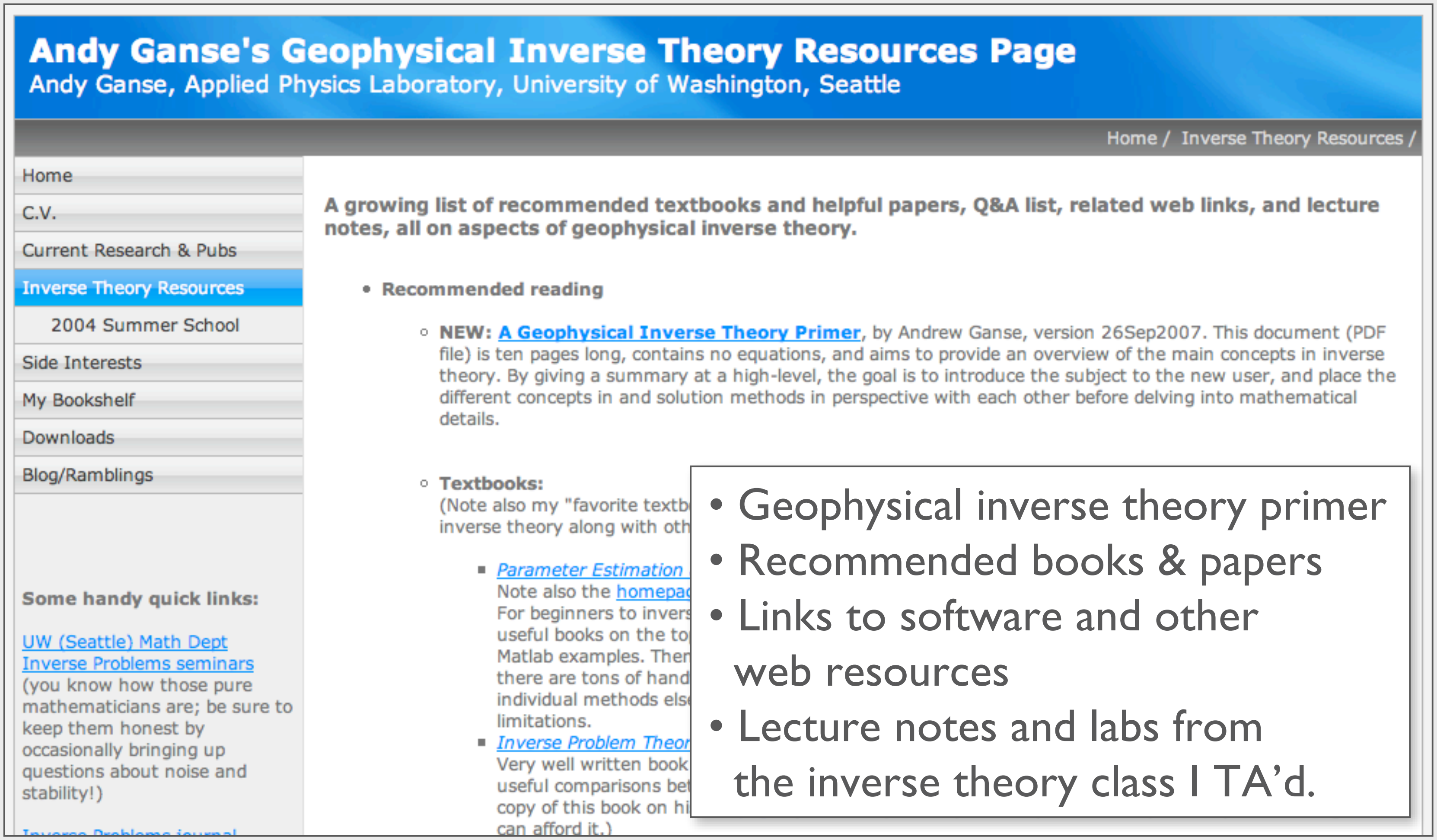
Outline

1. Introducing two approaches to geophysical inversion:
    a. Frequentist: statistical, find the deterministic but unknown model.
    b. Bayesian: probabilistic, the model is a random variable; find PDF.
2. A comparison with analytical inversion in linear problems.
3. Details of Frequentist inversion in linear problems
4. Details of Bayesian inversion in linear problems
5. Weakly nonlinear problems
6. More strongly nonlinear problems
7. Filters and smoothers
8. My own PhD work
9. Summarize

Inverse theory resources on my APL website

http://staff.washington.edu/aganse

(also linked via ESS and APL directory pages)



- Geophysical inverse theory primer
- Recommended books & papers
- Links to software and other web resources
- Lecture notes and labs from the inverse theory class I TA'd.

Introducing two approaches

Frequentist inversion *- find the deterministic but unknown model*

model

$m(x)$

(wavespeeds over depth)

noise

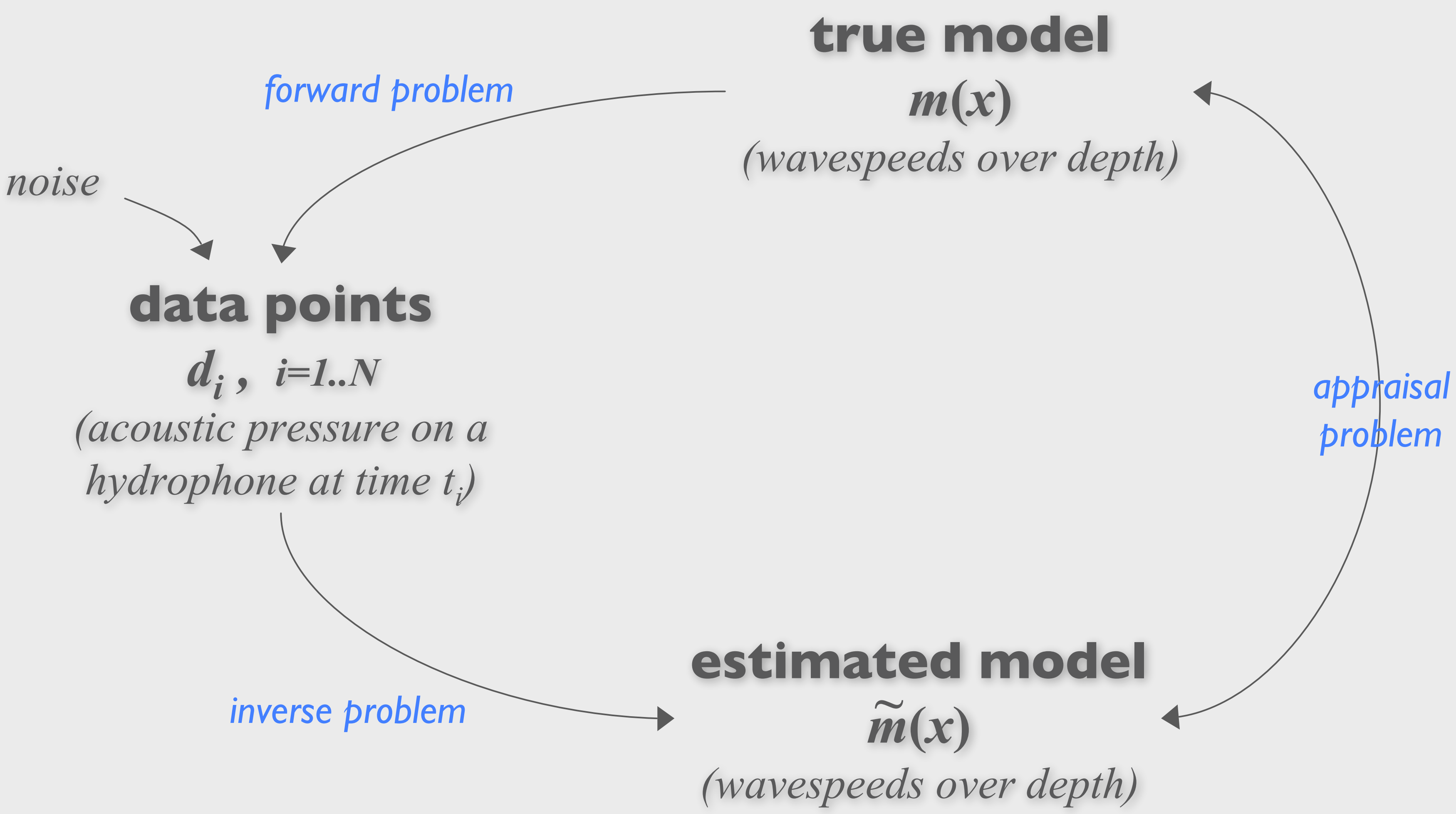
data points

d_i , $i=1..N$

(acoustic pressure on a hydrophone at time t_i)

Introducing two approaches

Frequentist inversion - *find the deterministic but unknown model*



Introducing two approaches

Bayesian inversion *- parameterized model has parameters that are random variables, find their joint PDF*

noise → **data points**

d_i , $i=1..N$

↓

data likelihood

$p(d|m)$

(includes forward problem)

prior model PDF $p(m)$

(perhaps from supplemental direct measurements of model)

Bayes' Rule

↓

posterior model PDF

$p(m|d)$

Geophysical vs. analytical inversion

Given a set of discrete data d_i we want to solve for a continuum model $m(x)$

Linear special cases

- Integral equation (IFK):

$$d(s) = \int g(s,x) m(x) dx$$

- Inverse problem :

$$d_i = \int g_i(x)\, m(x) dx + \epsilon_i$$

- Parameterize m(x) :

$$m(x) = \sum_j m_j\, b_j(x)$$

- Parameter estimation problem :

$$d_i = G_{ij}\, m_j + \epsilon_i$$

General nonlinear problems

- Inverse problem :

$$d_i = F_i[m(x)] + \epsilon_i$$

- Still parameterize the same way :

$$m(x) = \sum_j m_j\, b_j(x)$$

- Parameter estimation problem :

$$d_i = F_i[m_1, m_2, ..., m_n] + \epsilon_i$$

Loosely based on Menke, 1989

Some analytical inversion attempts for geophysical problems

Mainly developed for use on Schroedinger equation, with transforms for wave equation. So doesn't handle elasticity. Also too unstable for use in geophysical problems, where we have lots of noise and terrible geometric coverage.

- Exact inversion
- Downward-continuation
- Layer-stripping
- Gelfand-Levitan method

But noting work by our own Sylvester & Winebrenner re reflection coefficients: promising for this one problem, although limited to 1D with no attenuation

SIAM J. APPL. MATH.
Vol. 56, No. 3, pp. 736–754, June 1996

© 1996 Society for Industrial and Applied Mathematics
003

LAYER STRIPPING FOR THE HELMHOLTZ EQUATION*

JOHN SYLVESTER†, DALE WINEBRENNER‡, AND FRED GYLYS-COLWELL§

A few clarifications

- Note model function $m(.)$ doesn't have to be 1D, but we still use vector **m** :

$$m(x) = \sum_j m_j \, b_j(x)$$

$$m(x,y,z) = \sum_j m_j \, b_j(x,y,z)$$

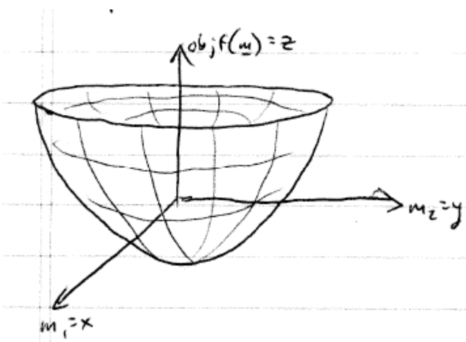
- Rank-deficiency and ill-posedness :

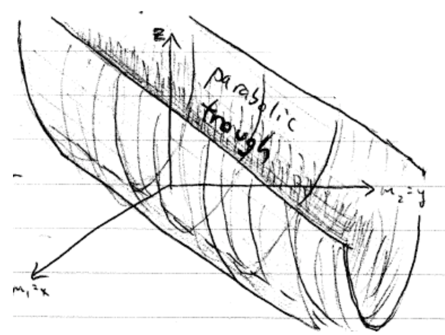
To keep notation clean here, let us assume noise in d_{obs} is $\varepsilon \sim \mathrm{N}(\mathbf{0},\mathbf{I})$

Define objective function:

$$objf(\mathbf{m}) = \|\mathbf{d}_{\mathrm{obs}} - \mathbf{G}\mathbf{m}\|_2^2$$

(i.e. data misfit)





Linear inverse problems

Finding a solution estimate in frequentist approach

…again let $\varepsilon \sim N(\mathbf{0},\mathbf{I})$ here.

- Regularized least squares example (Tikhonov) :

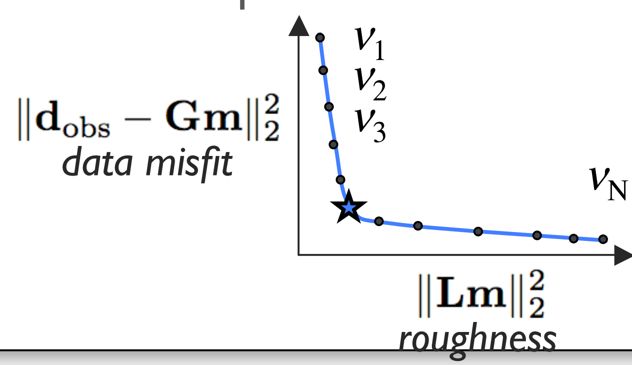
$$\tilde{\mathbf{m}} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}_{\text{obs}} \quad \rightarrow \quad \tilde{\mathbf{m}} = (\mathbf{G}^T\mathbf{G} + \underline{\nu^2\mathbf{L}^T\mathbf{L}})^{-1}\mathbf{G}^T\mathbf{d}_{\text{obs}}$$

"messes up" the estimator
in return for stability (see next slide)

- Occam's regularization - 2nd derivative constraint via "roughness" matrix L :

$$\mathbf{L} = \frac{1}{(\Delta x)^2}\begin{bmatrix} 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & & & \ddots & & \\ & & & & 1 & -2 & 1 \end{bmatrix}$$

- Choosing optimal tradeoff parameter on L-curve :



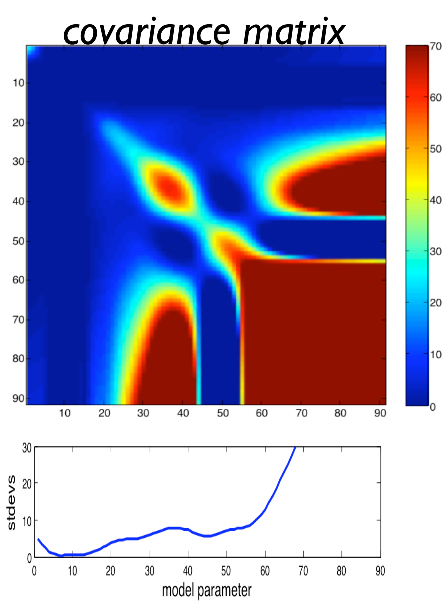
We want to only fit the data to within the noise statistics, then choose the smoothest model fn. But we often don't know the noise statistics, so the choice of best tradeoff param isn't always trivial.

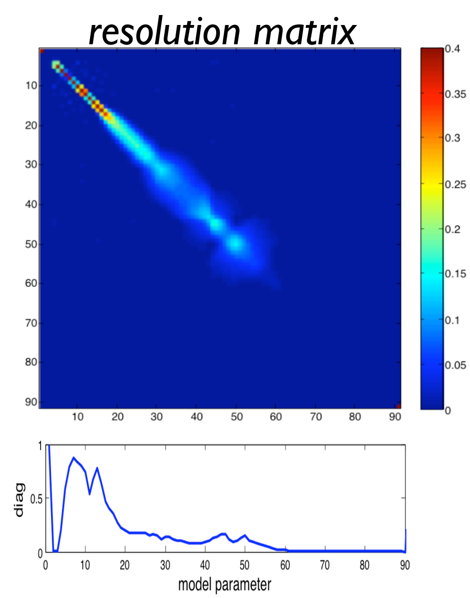
Linear inverse problems

Uncertainty and resolution of the frequentist solution estimate

…again let $\varepsilon \sim N(\mathbf{0},\mathbf{I})$ here.

- Covariance and resolution matrices :
  - "generalized inverse" of G : $\mathbf{G}^{\#} = (\mathbf{G}^T\mathbf{G} + \underline{\nu^2\mathbf{L}^T\mathbf{L}})^{-1}\mathbf{G}^T$
  - compute covariance matrix as : $\mathbf{C}_m = \mathbf{G}^{\#}\mathbf{G}^{\#T}$
  - compute resolution matrix as : $\mathbf{R}_m = \mathbf{G}^{\#}\mathbf{G}$





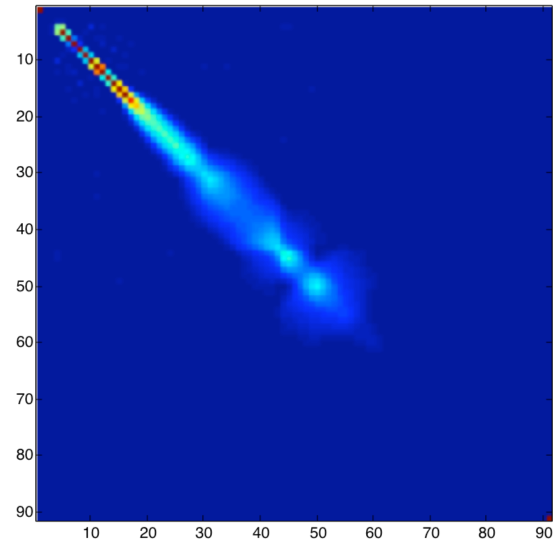
"Messing up" the estimator means its solution becomes a set of weighted averages of the true solution. The resolution matrix contains the weights, and this covariance matrix is that of the weighted averages, NOT of the model parameters themselves. Thus this approach does not provide a probabilistic result.

Resolution analysis for experiment design

Resolution analysis to optimize experiment geometry, before doing the experiment

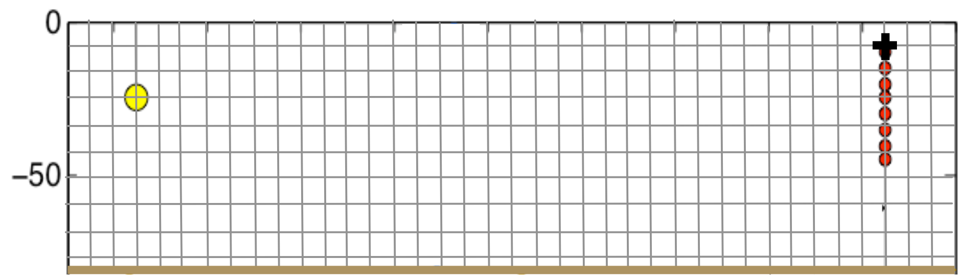
The weights are generally clumped about the parameter of interest, lending an interpretation of limited resolution -- for example, parameter 30 here cannot be resolved independently from parameters 25 - 40. A diagonal resolution matrix has perfect resolution, and increasing spread about the diagonal shows coarsening resolution. (However, note not all resolution matrices are symmetric.)

resolution matrix **R**



tr(**R**) ~ # of parameters resolved by the dataset

Could map a function *q*(*range*,*depth*) where each grid point gets a tr(**R**) value corresponding to a different receiver array location.



This map can then show optimal location(s) for the receiver array.

Introducing Bayesian inversion

- **Frequentists** define probability in terms of frequency of repeatable events. So one can't know anything about model before the event/experiment.
- **Bayesians** define probability in terms of degree of belief. So one *can* know about the model before the event/experiement.

- Bayes' Rule: (from definition of conditional probability)

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m})}{p(\mathbf{d})} \propto p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m})$$

"posterior" distribution of model parameters — $p(\mathbf{m}|\mathbf{d})$

"data likelihood" function — $p(\mathbf{d}|\mathbf{m})$

"prior" distribution of model parameters — $p(\mathbf{m})$

for linear problem + gaussian dists, frequentist & Bayesian solutions look similar…

Comparison of Bayesian and frequentist inversion for linear problems

…again here $\varepsilon \sim N(\mathbf{0},\mathbf{I})$ in $\mathbf{d}_{obs} = \mathbf{f}(\mathbf{m}) + \varepsilon$

- Linear problems + gaussian distributions → **same ML/MAP estimate**.
  (for same "prior info", eg roughness regularization).
  - Freq : $\tilde{\mathbf{m}}_{ML} = (\mathbf{G}^T\mathbf{G} + \nu^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{G}^T\mathbf{d}_{\text{obs}}$ (for $\mathbf{m}_{\text{preferred}}=\mathbf{0}$)
  - Bayes : $\tilde{\mathbf{m}}_{MAP} = (\mathbf{G}^T\mathbf{G} + \underline{\mathbf{C}_\nu^{-1}})^{-1}\mathbf{G}^T\mathbf{d}_{\text{obs}}$ (for $\mathbf{m}_{\text{prior}}=\mathbf{0}$)

    prior cov parameterized by ν

- But the solution **covariances differ** due to the different philosophies :
  - Freq : $\tilde{\mathbf{C}} = \mathbf{G}^\#\mathbf{C}_\varepsilon\mathbf{G}^{\#T} = \mathbf{G}^\#\mathbf{G}^{\#T}$ *(since $\mathbf{C}_\varepsilon=\mathbf{I}$)*

    $= (\mathbf{G}^T\mathbf{G} + \nu^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{G}^T\mathbf{G}(\mathbf{G}^T\mathbf{G} + \nu^2\mathbf{L}^T\mathbf{L})^{-1}$
  - Bayes : $\tilde{\mathbf{C}} = (\mathbf{G}^T\mathbf{G} + \mathbf{C}_\nu^{-1})^{-1}$

Comparison of Bayesian and frequentist inversion for linear problems

…again here $\varepsilon \sim N(\mathbf{0},\mathbf{I})$ in $\mathbf{d}_{obs} = \mathbf{f}(\mathbf{m}) + \varepsilon$

A linear resolution matrix can be defined for both cases (although they're not identically the same quantity because the philosophies/covariances differ) :

- Freq : $$\mathbf{G}^{\#} = (\mathbf{G}^T\mathbf{G} + \nu^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{G}^T$$ $$\mathbf{R} = \mathbf{G}^{\#}\mathbf{G}$$
- Bayes : $$\mathbf{R} = \mathbf{I} - \tilde{\mathbf{C}}\ \mathbf{C}_\nu^{-1}$$

Unlike frequentist inversion which must estimate weighted averages of model parameters, Bayesian inversion is a fully probabilistic description of the model parameters themselves. So no weighted averages, and all resolution information is contained in the probability distributions.

Tarantola, 2005

Weakly nonlinear inverse problems

Finding a solution estimate via local linearization in either frequentist or Bayesian inv.

- Local linearization (using analytical derivs, finite diffs, or alg diff for the $\mathbf{F}(\mathbf{m}_i)$) :

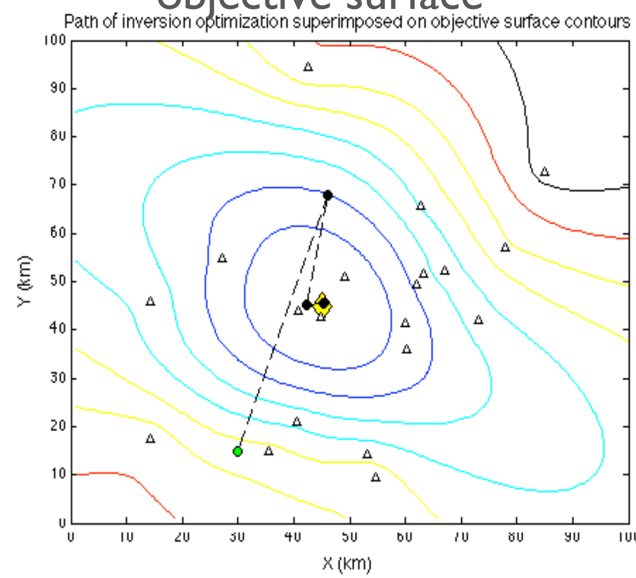
$$\mathbf{d} = \mathbf{f}(\mathbf{m}) = \mathbf{f}(\mathbf{m}_i) + \underline{\mathbf{F}(\mathbf{m}_i)}\delta\mathbf{m}_i + \cdots$$

(i.e. Taylor series expansion about $\mathbf{m}_i$)

matrix of partial derivs

truncate at linear

objective surface



- Observed data : $\mathbf{d}_{\text{obs}} = \mathbf{f}(\mathbf{m}) + \boldsymbol{\varepsilon}$

Let's say $\varepsilon \sim N(\mathbf{0},\mathbf{I})$ to keep notation clean below…

- Gauss-Newton method :

start with initial est. $\mathbf{m}_0$, then...

$\mathbf{G} = \mathbf{F}(\mathbf{m}_i)$ ← *compute local derivatives*

$\delta\mathbf{d}_{\text{obs}} = \mathbf{d}_{\text{obs}} - \mathbf{f}(\mathbf{m}_i)$ ← *compute local residuals*

$\delta\mathbf{m}_i = (\mathbf{G}^T\mathbf{G} + \nu^2\mathbf{L}^T\mathbf{L})^{-1}(\mathbf{G}^T\mathbf{d}_{\text{obs}} - \nu^2\mathbf{L}^T\mathbf{L}\mathbf{m}_i)$ ← *solve for model perturbation step*

$\mathbf{m}_{i+1} = \mathbf{m}_i + \delta\mathbf{m}_i$ ← *add on the new model perturbation step*

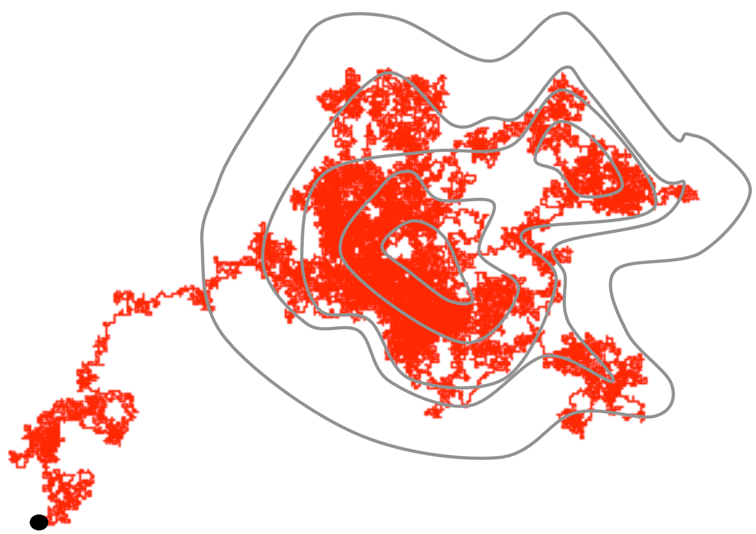
For more strongly nonlinear problems…

Markov Chain Monte Carlo (MCMC) sampling of the Bayesian posterior distribution

Only proportionality necessary for modifying random walk (saves computation):

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m})$$

Metropolis/Hastings : random walk that prefers higher posterior probabilities:



- probability information is in sample density rather than in the p(m|d) values of the samples
- then compute marginal distributions and so on for the parameters
- burn-in, sample independence, and when to stop sampling (Raftery & Lewis, 1996)

Filters and smoothers - stringing together a series of Bayesian inverse problems

"Conjugate prior": for a given data likelihood p(d|m), one that produces a posterior distribution of the same form as the prior, i.e. parameterized the same (e.g. Gaussian or other exponential)

time 0 : $p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m})$

Set next prior equal to previous posterior…

time 1 : $p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m})$

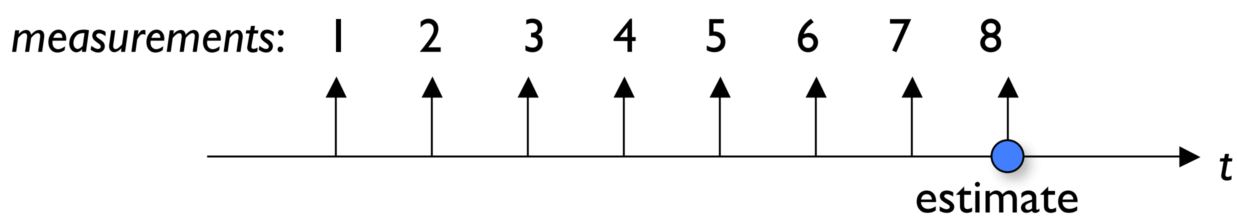
Set next prior equal to previous posterior…

time 2 : $p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m})$

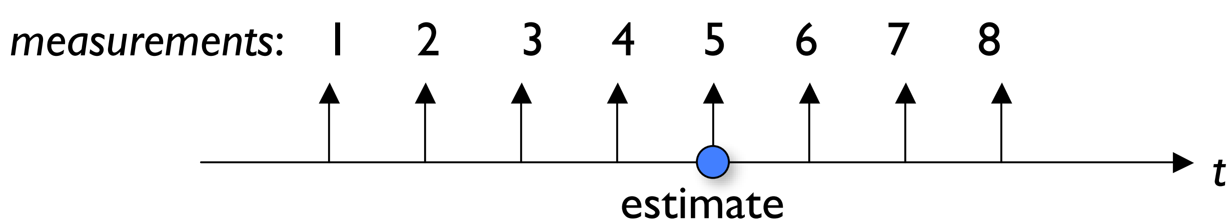
etc…

Filters and smoothers - stringing together a series of Bayesian inverse problems

Filter - estimates a vector function $\mathbf{m}(t)$ or sequence $\mathbf{m}_k$ (k=1..N) at last measurement point:

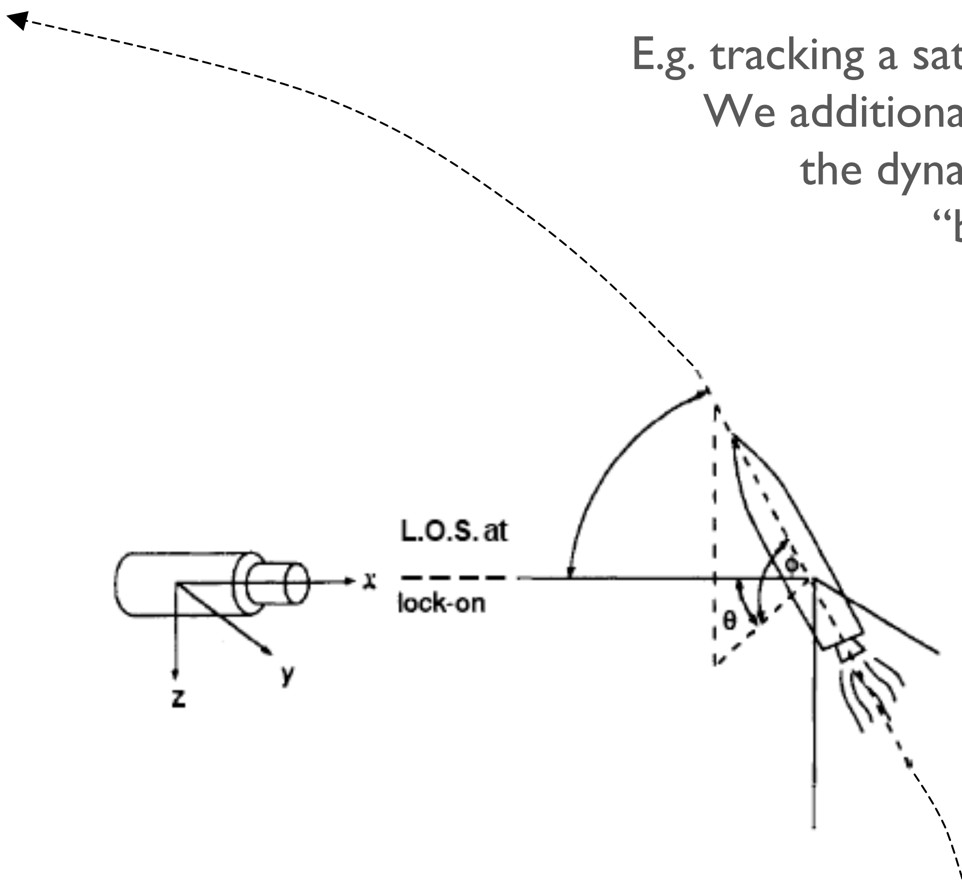


Smoother - estimates a vector function $\mathbf{m}(t)$ or sequence $\mathbf{m}_k$ (k=1..N) in middle of measurements



A common application of filters and smoothers

E.g. tracking a satellite or missile.
We additionally regularize by
the dynamics to find the
"best" trajectory.

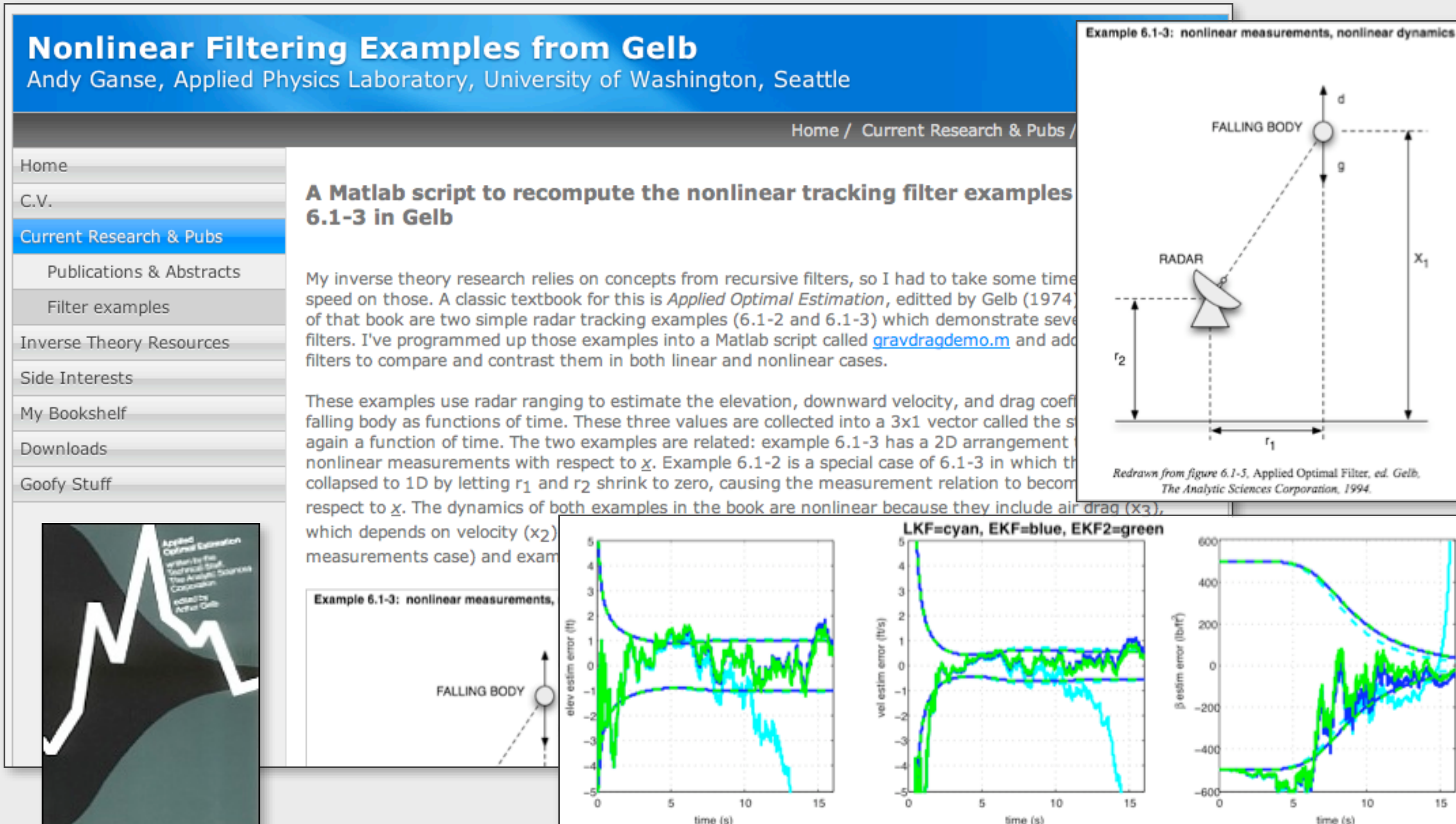


G. Schmidt, *J. Guidance, Cntl., & Dyn.*, 1993

Filter theory tutorial code on my APL website

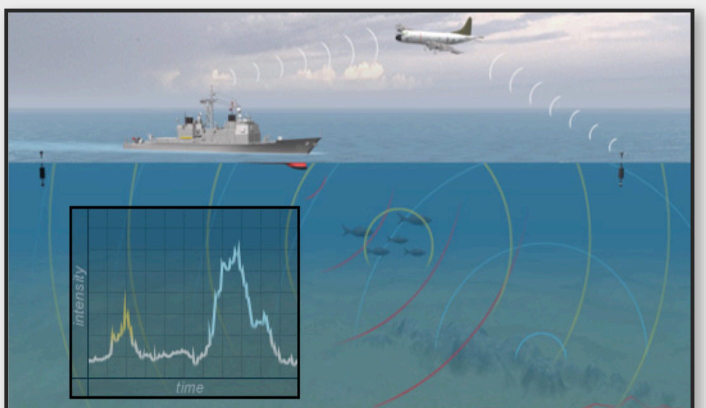
http://staff.washington.edu/aganse

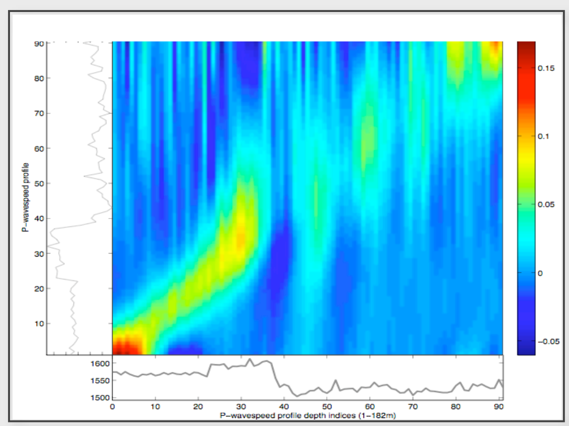
(also linked via ESS and APL directory pages)

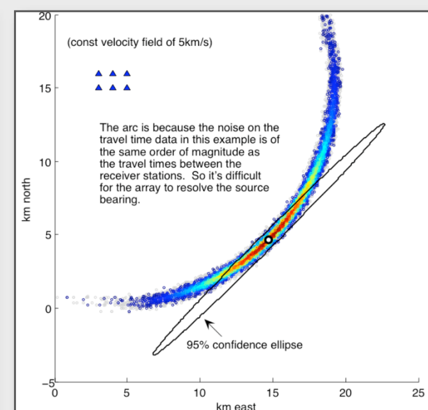


My own PhD work

- Resolution and uncertainty analyses for a shallow water ocean bottom inverse problem:
  - optimize experiment geometry via resolution analysis, and design experiments in which much more information about the bottom is obtained
  - validating linear approximation of uncertainty and resolution via Monte Carlo and nonlinear filter approaches. Ideally the linearization approach is preferrable since fastest, but must check validity.







Summary

- Introduced both frequentist and Bayesian inversion, and filters
- Linear, weakly nonlinear, and more strongly nonlinear problems
- Discussion regarding geophysical vs. analytical inversion
- Some shameless plugs for material on my website ;-)
- And a brief mention of my own PhD research (not the focus today).
- This talk paves the way for Ken Creager's (ESS-UW) talk next time on results of specific geophysical inversion work, including 3D crustal tomography and locating seismic "tremor" signals.

THANK YOU