Probing the covariance matrix

Kenneth M. Hanson

T-16, Nuclear Physics; Theoretical Division
Los Alamos National Laboratory

Bayesian Inference and Maximum Entropy Workshop,
Paris, France, July 9-13, 2006

This presentation available at
http://www.lanl.gov/home/kmh/

LA-UR-06-5241

Overview

- Minus-log-probability analogous to a physical potential
- Gaussian approximation near peak of probability density function
- Probing the covariance matrix with an external force
  - deterministic technique to replace stochastic calculations
- Examples
- Potential applications

Physical potential

- Spring produces restoring force proportional to displacement from its equilibrium position

$$F = -k\,x$$

- Potential is integral of force

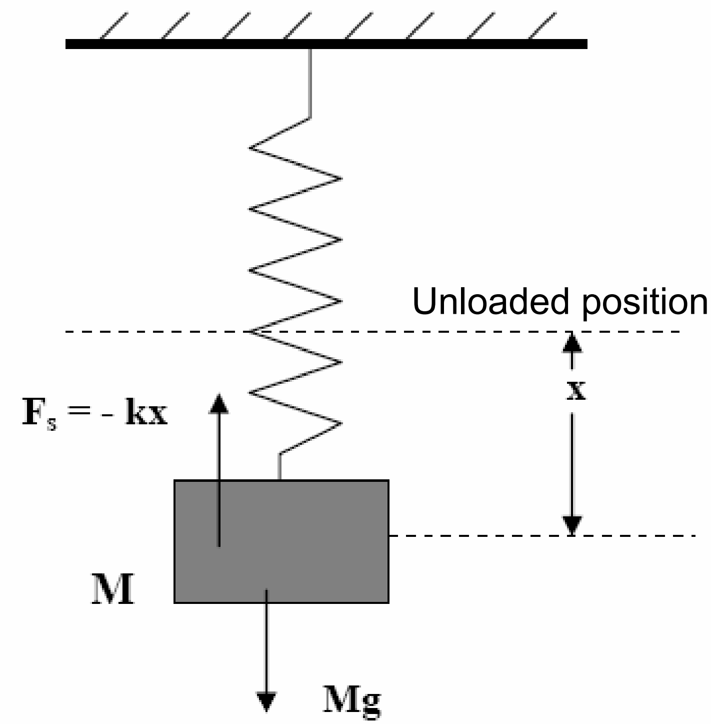
$$\varphi(x) = \int F\,dx = \tfrac{1}{2}k\,x^2$$

  - it is often more useful to think about a physical problem in terms of potentials instead of forces

- Derivative of potential is force

$$\frac{d\varphi(x)}{dx} = F$$

Mass suspended from a spring



At equilibrium
kx = Mg

Analogy to physical system

- Analogy between minus-log-posterior and a physical potential

$$\varphi(\boldsymbol{a}) = -\log p(\boldsymbol{a} \mid \boldsymbol{d}, I)$$

  - $\boldsymbol{a}$ represents parameters
    $\boldsymbol{d}$ represents data
    I represents background information, essential for modeling
- Gradient $\partial_{\boldsymbol{a}}\varphi$ corresponds to forces acting on the parameters
- Maximum *a posteriori* (MAP) estimates parameters $\hat{a}_{\text{MAP}}$
  - condition is $\partial_{\boldsymbol{a}}\varphi = 0$
  - optimized model may be interpreted as mechanical system in equilibrium – net force on each parameter is zero
- This analogy is very useful for Bayesian inference
  - conceptualization
  - developing algorithms

Gaussian approximation

- Posterior distribution is very often well approximated by a Gaussian in the parameters
- Then, φ is quadratic in perturbations in the model parameters from the minimum in φ at $\hat{\boldsymbol{a}}$:

$$\boxed{\varphi(\boldsymbol{a}) = \tfrac{1}{2}(\boldsymbol{a} - \hat{\boldsymbol{a}})^{\mathrm{T}} \boldsymbol{K} (\boldsymbol{a} - \hat{\boldsymbol{a}}) + \varphi_{\min}}$$

where $\boldsymbol{K}$ is the φ curvature matrix (aka *Hessian*);

- Uncertainties in the estimated parameters are summarized by the covariance matrix:

$$\operatorname{cov}(\boldsymbol{a}) \equiv \left\langle (\boldsymbol{a} - \hat{\boldsymbol{a}})(\boldsymbol{a} - \hat{\boldsymbol{a}})^{\mathrm{T}} \right\rangle \equiv \boldsymbol{C} = \boldsymbol{K}^{-1}$$

- Inference process becomes one of finding $\hat{\boldsymbol{a}}$ and $\boldsymbol{C}$

Effect of external force

- Consider applying an constant external force to the parameters
- Effect is to add a linearly increasing term to potential

$$\varphi'(\boldsymbol{a}) = \tfrac{1}{2}(\boldsymbol{a} - \hat{\boldsymbol{a}})^{\mathrm{T}} \boldsymbol{K} (\boldsymbol{a} - \hat{\boldsymbol{a}}) + \varphi_{\min} - \boldsymbol{f}^{\mathrm{T}} \boldsymbol{a}$$

- Gradient of perturbed potential is

$$\frac{\partial \varphi'}{\partial \boldsymbol{a}} = \boldsymbol{K} (\boldsymbol{a} - \hat{\boldsymbol{a}}) - \boldsymbol{f}$$

- At the new minimum, gradient is zero, so

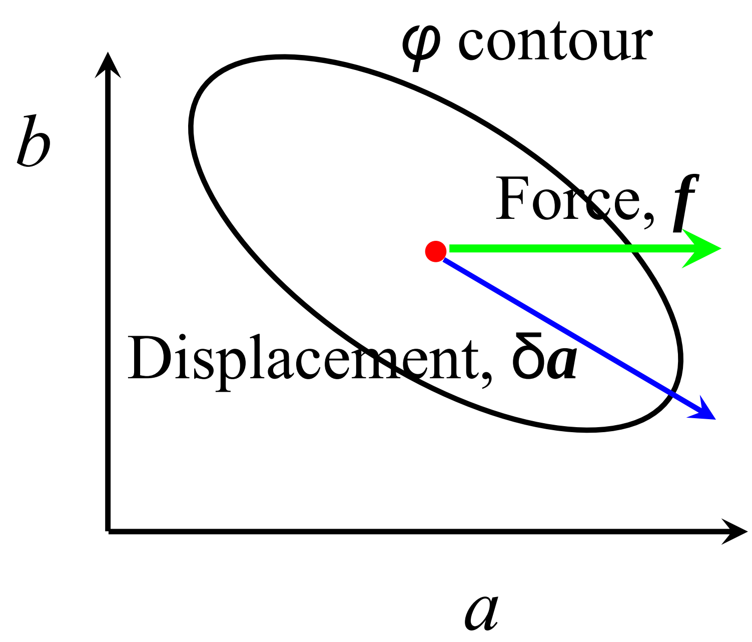
$$\boxed{\delta \boldsymbol{a}_{\min} = \hat{\boldsymbol{a}}' - \hat{\boldsymbol{a}} = \boldsymbol{K}^{-1} \boldsymbol{f} = \boldsymbol{C} \boldsymbol{f}}$$

- Displacement of minimum in parameters, $\delta \boldsymbol{a}_{\min}$, is proportional to covariance matrix times the force
- With external force, one may "probe" the covariance
  - each applied force probes one column (or average of several)

Effect of external force

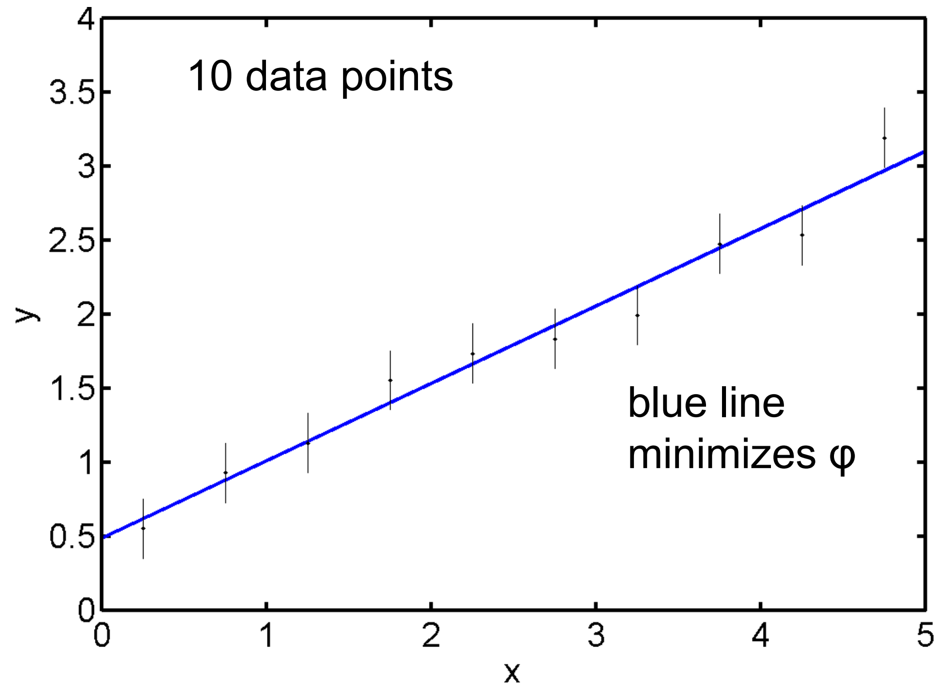
- Displacement of minimizer of φ is in direction other than applied force
- Displacement is controlled by the covariance matrix
  - its direction is determined by correlations
  - its magnitude is proportional to variance (inversely proportional to the curvature or stiffness)

2-D parameter space



Simulated data for straight line

- Linear model: $y = a + bx$
  a is intercept at $x = 0$
  b is slope of line
- Simulate 10 data points, with values: $a = 0.5$ $b = 0.5$
- Add Gaussian noise:
  $\sigma_y = 0.2$



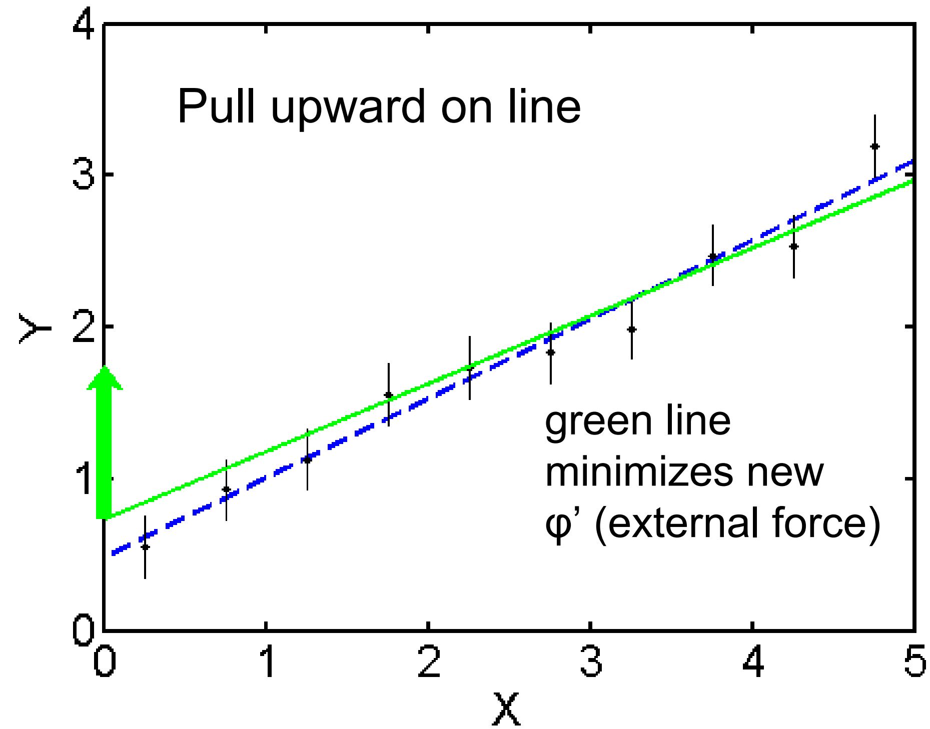
- Find straight line that minimizes

$$\varphi(\boldsymbol{a}) = \tfrac{1}{2}\chi^2 = \tfrac{1}{2}\sum_i \frac{[d_i - y_i(x_i;\boldsymbol{a})]^2}{\sigma_i^2}$$

  - where d_i are the data, y_i are the model values at positions x_i

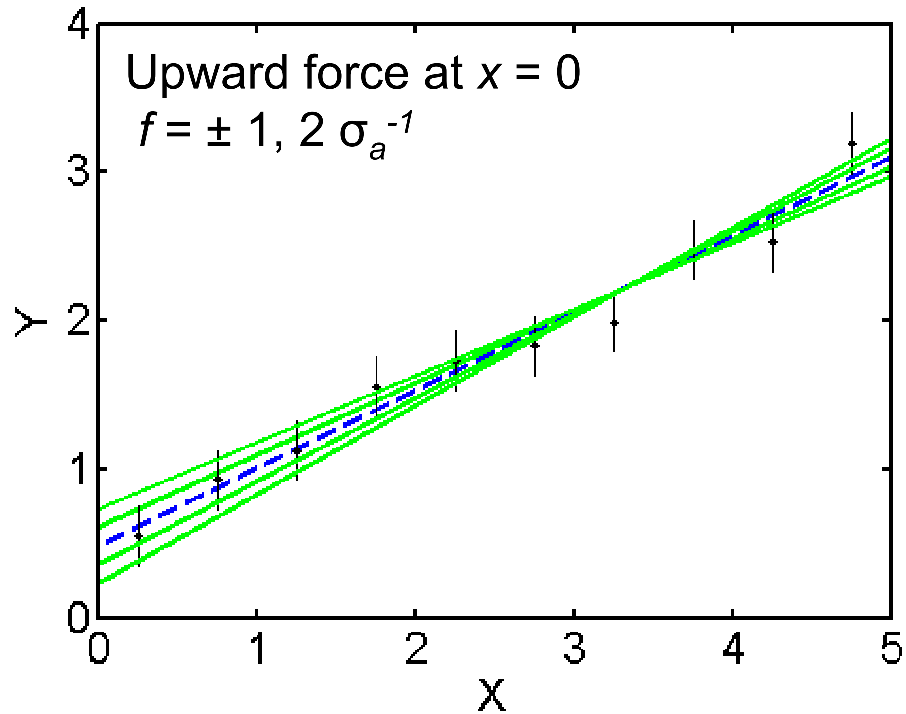
Apply force to solution

- Apply upward force to solution line at $x = 0$ and find new minimum in φ
  - thus, pull only on parameter a
- Effect is to pull line upward at $x = 0$ **and** reduce its slope
  - data constrain solution
- Conclude that a (intercept) and b (slope) are anti-correlated



Apply several forces to solution

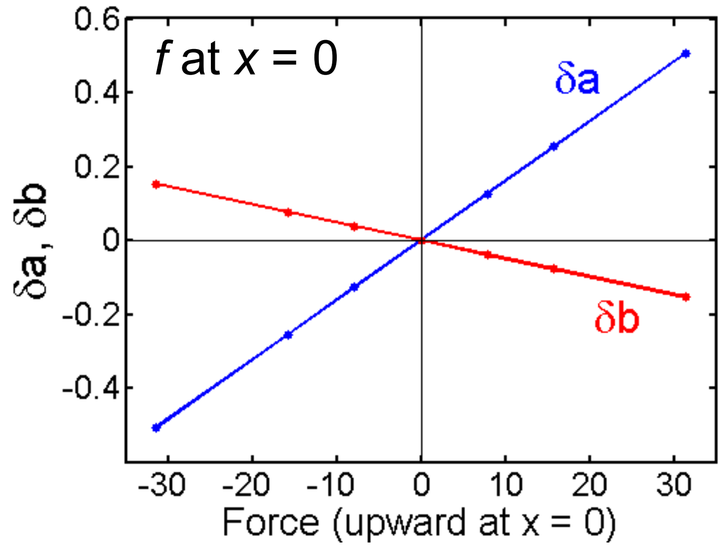
- Family of lines shown for forces applied upward at $x = 0$:
  $f = \pm\, 1, \pm\, 2$ x constant
  - observe proportional displacement of intercept ($x = 0$)

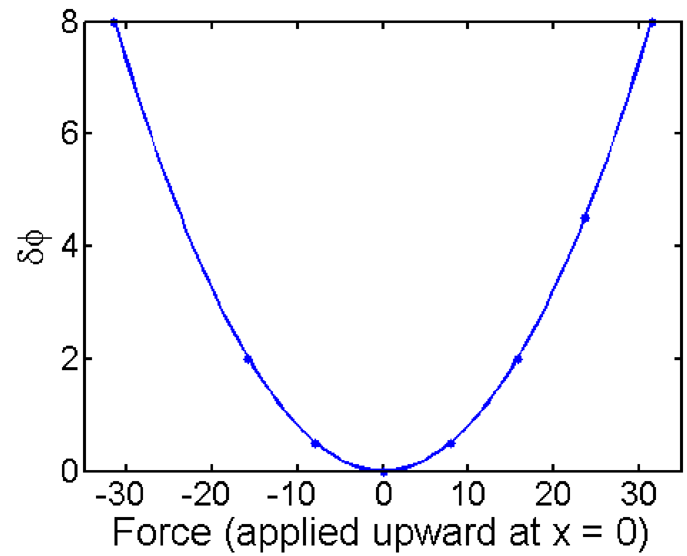


- These results yield quantified estimates of parts of the covariance matrix

Uncertainties in straight line fit

- Plot above shows results for variety of forces applied upward at $x = 0$
  - perturbations in parameters proportional to force
  - slope of $\delta a = \sigma_a^{\,2} = C_{aa} = (0.127)^2$
  - slope of $\delta b = C_{ab} = -\,4.84\mathrm{x}10^{-3}$
- Plot below shows φ (= ½χ^2) is quadratic function of force
  - for force $f = \pm\,\sigma_a^{\,-1} = (0.127)^{-1}$ min φ increases by 0.5 (min χ^2 increases by 1)
- Either dependence provides a way to quantify (co)variance estimates
- C_{bb} not determined





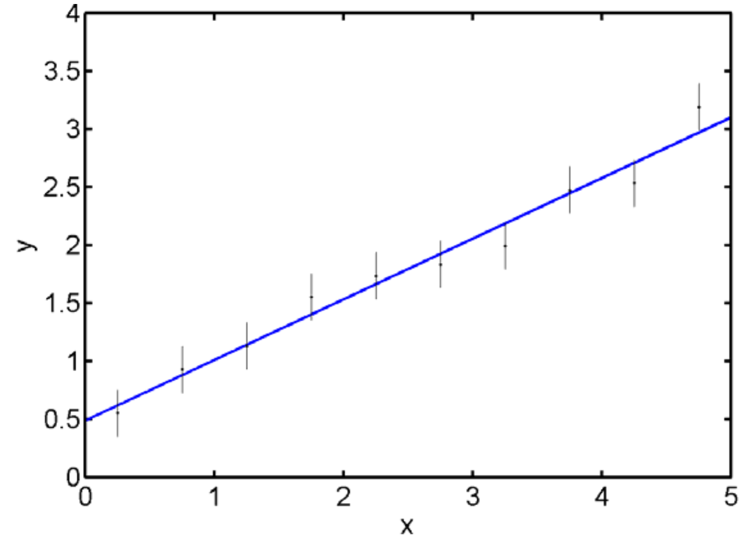
Compare to result from standard minimum-χ^2

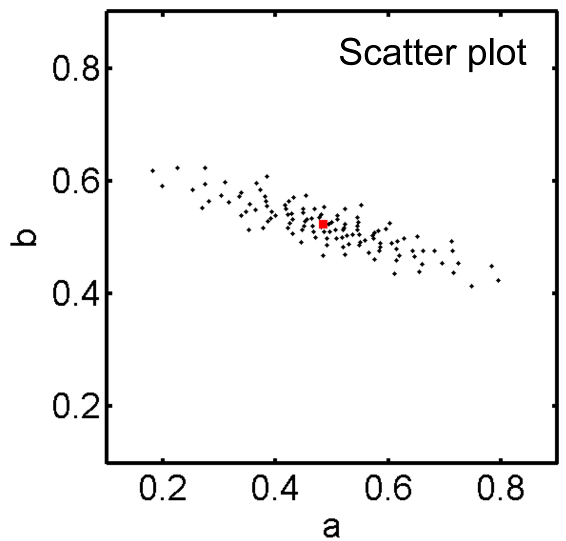
- Fit linear model: $y = a + bx$
- Determine parameters a and b by minimum χ^2 (least-squares) analysis
- Results: $\chi^2_{\min} = 4.04 \quad p = 0.775$

$$\hat{a} = 0.484 \qquad \hat{b} = 0.523$$

$$\sigma_a = 0.127 \qquad \sigma_b = 0.044$$

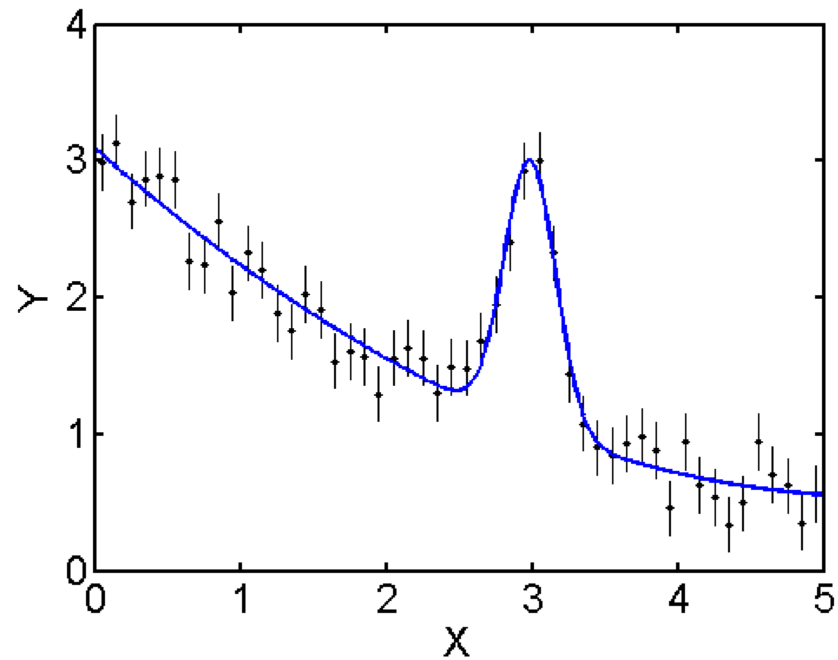
  - correlation: $r_{ab} = -0.867$
- Covariance estimates from these
  - $C_{aa} = \sigma_a^2 = (0.127)^2$
  - $C_{ab} = r_{ab}\,\sigma_a\,\sigma_b = -4.84\times10^{-3}$
  - these are identical to values obtained by applying external force
  - C_{bb} not determined with external force





Simple spectrum

- Simulate simple spectrum with a single peak:
  - Gaussian peak (ampl = 2, *w* = 0.2)
  - quadratic background
  - add random noise (rmsdev = 0.2)
- Minimize φ wrt 6 parameters
  - amplitude, width, position of peak
  - 3 coefficients for quadratic background
- Nonlinear problem
- Suppose quantity of interest is the area under the peak;
  - what force should be applied to parameters?



External force for derived quantities

- Consider a scalar quantity z, which is a function of parameters $\boldsymbol{a}$

$$z = z(\boldsymbol{a})$$

- The small perturbation $\delta\boldsymbol{a}$ results in a perturbation in z

$$\delta z = \boldsymbol{s}_z^{\mathrm{T}}\, \delta\boldsymbol{a}$$

  - where $\boldsymbol{s}_z$ is the sensitivity vector for z (derivative of z wrt $\boldsymbol{a}$)

- The variance in z is

$$\boldsymbol{C}_z = \mathrm{var}(z) \equiv \left\langle \delta z\, \delta z^{\mathrm{T}} \right\rangle = \left\langle \boldsymbol{s}_z^{\mathrm{T}}\, \delta\boldsymbol{a}\, \delta\boldsymbol{a}^{\mathrm{T}} \boldsymbol{s}_z \right\rangle = \boldsymbol{s}_z^{\mathrm{T}}\, \boldsymbol{C}_{\boldsymbol{a}}\, \boldsymbol{s}_z$$

  - standard result for propagating covariance

- The force on parameters $\boldsymbol{a}$ needed to probe z is

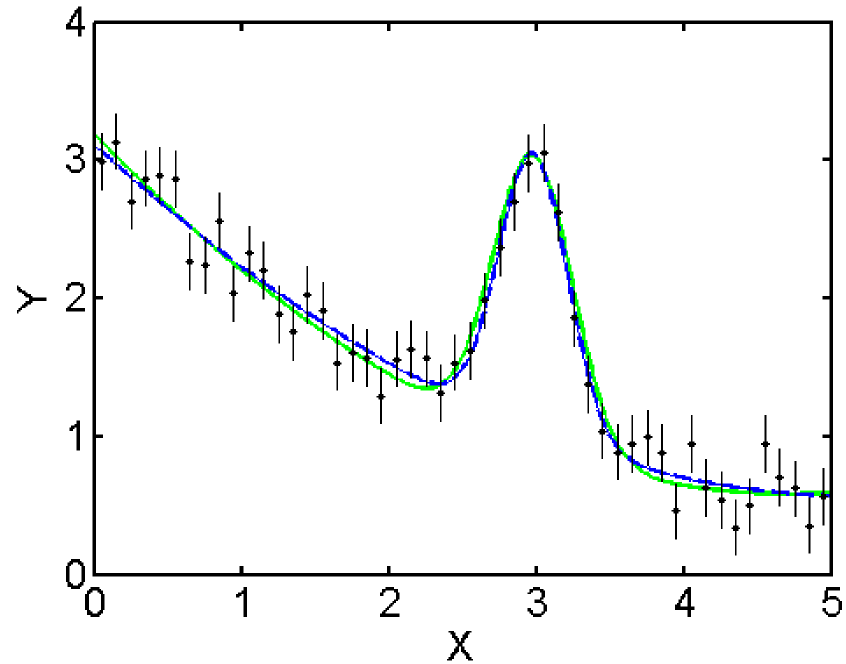
$$\boxed{\boldsymbol{f}_z = \boldsymbol{s}_z = \partial_{\boldsymbol{a}} z}$$

resulting in $\qquad \delta z = \boldsymbol{C}_z\, \boldsymbol{f}_z$

which is the same relation as for $\delta\boldsymbol{a}$

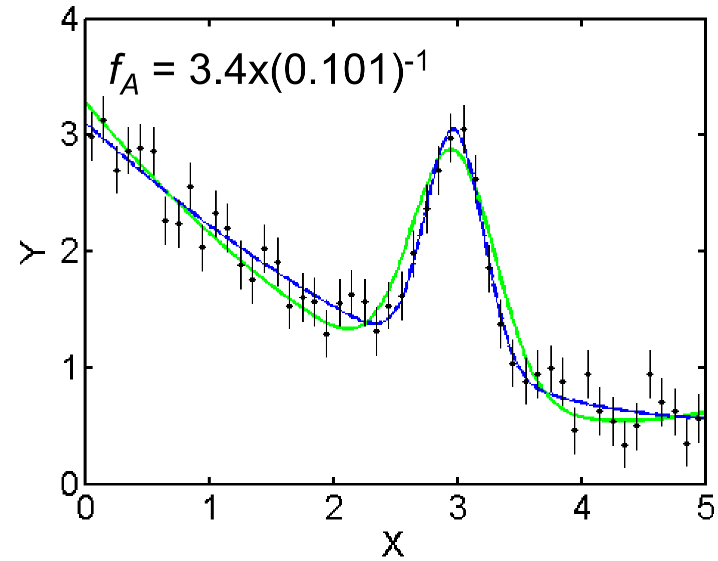
Simple spectrum – apply force to peak area

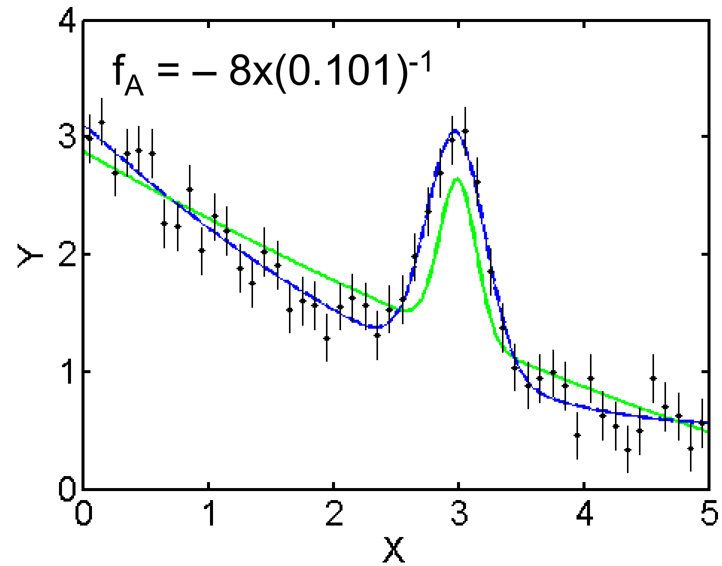
- Area under Gaussian peak;
  a = amplitude, w = rms width:
  $$A = \sqrt{2\pi}\, a w$$
  $$= 0.86$$
- To examine the area, apply force to parameters proportional to derivatives of area wrt parameters,
  $$\frac{\partial A}{\partial a} = \sqrt{2\pi}\, w \qquad \frac{\partial A}{\partial w} = \sqrt{2\pi}\, a$$
- Plot shows result of applying force proportional to these derivatives
  - area of Gaussian increased
  - background altered slightly



Simple spectrum – apply force to peak area

- Examples of sizable +/– forces applied to area



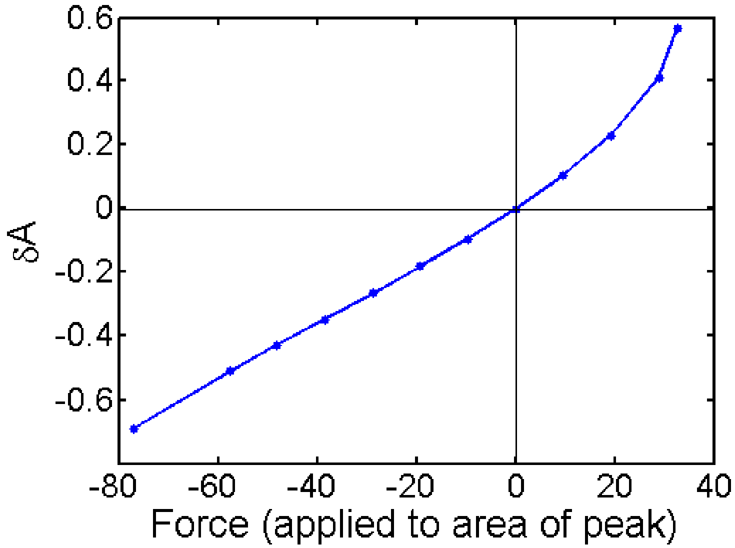


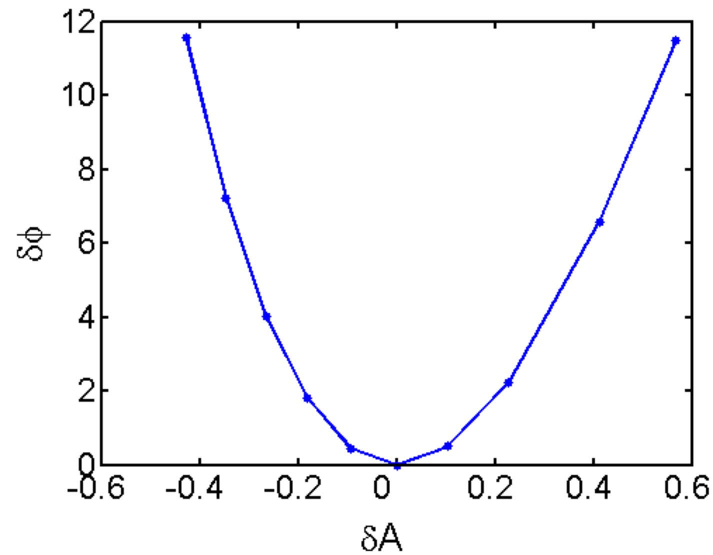
Simple spectrum – apply force to peak area

- Plot shows nonlinear response, but approximately linear for small f
- Plot below shows $\delta\varphi$ as function of displacement δA
- $\delta\varphi$ has quadratic form for small δA

$$\delta\varphi = \tfrac{1}{2}\left[\frac{\delta A}{\sigma_A}\right]^2$$

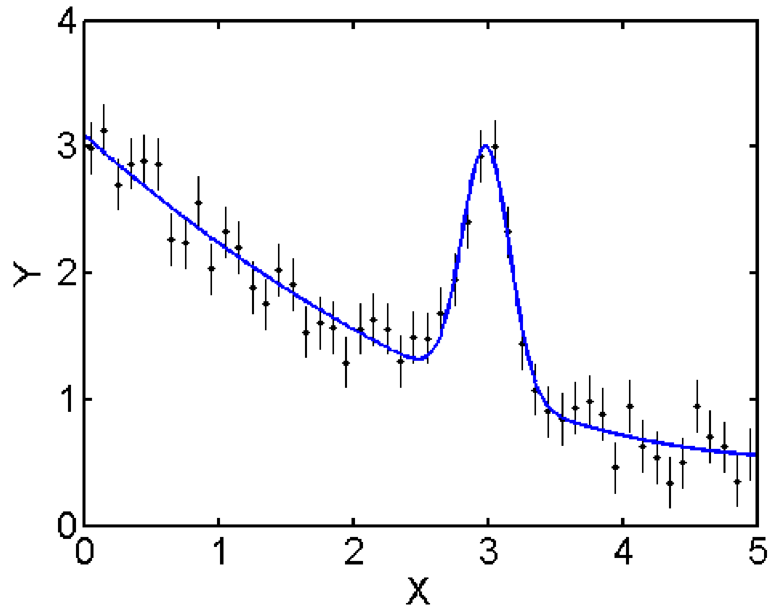
  - this relation allows one to estimate σ_A from a displacement produced by single small applied force:
  - $\sigma_A = 0.098$ (– side); 0.104 (+ side)





Compare to minimum-χ^2 result

- Minimum χ^2 fit
- Fit involves 6 parameters
  - nonlinear problem
  - results: $\chi^2_{\min} = 34.32 \quad p = 0.852$
    ampl. $\hat{a} = 1.948 \quad \sigma_a = 0.149$
    width $\hat{w} = 0.1759 \quad \sigma_w = 0.0165$
  - correlation: $r_{aw} = -0.427$



- From these, standard error in area

$$\sigma_A = \sqrt{2\pi}\left[w^2\sigma_a^2 + a^2\sigma_w^2 - r_{aw}aw\sigma_a\sigma_w\right]^{1/2} = 0.093$$

  - this result agrees fairly well with external force estimate, considering nonlinearity

Summary of steps to estimate variance

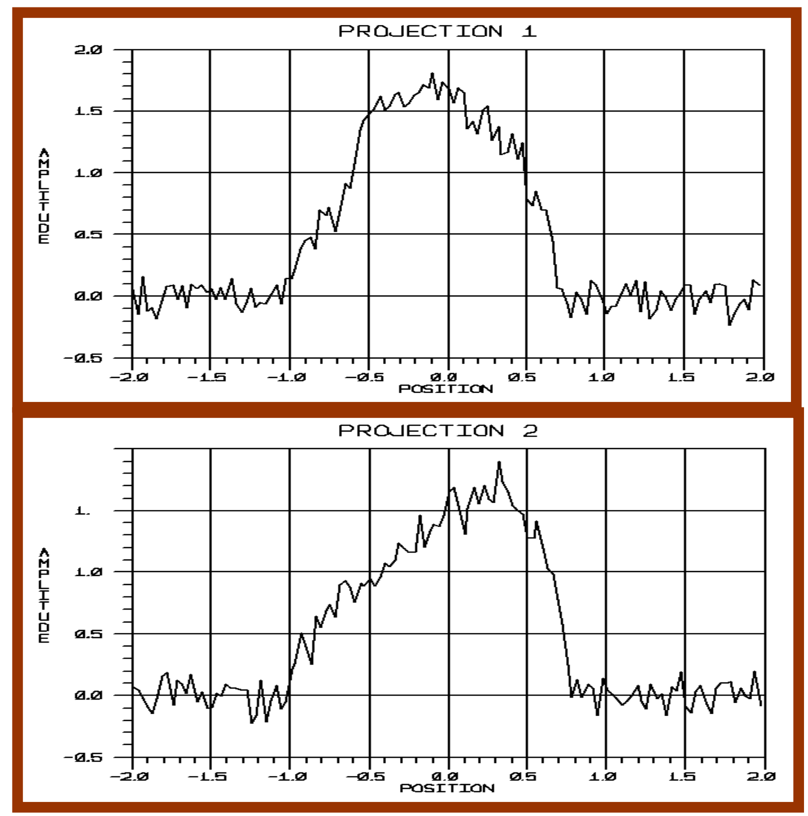
- Find values of model parameters $\boldsymbol{a}$ that minimize $\boldsymbol{\varphi}$
- Decide on quantity of interest z
- If z is not one of parameters, calculate $\mathbf{s}_z = \partial_{\boldsymbol{a}} z$
- Find parameter values that minimize $\boldsymbol{\varphi}' = \boldsymbol{\varphi} - k\, \mathbf{s}_z^{\mathrm{T}}\, \boldsymbol{a}$,
  for some scaling factor k (appropriate value is about σ_z^{-1})
- Check that change in $\boldsymbol{\varphi}$ is around 0.5; if not adjust k and minimize $\boldsymbol{\varphi}'$ again
- From perturbations in parameters, estimate standard error in z by either formula:
  - $\sigma_z^2 = \dfrac{\delta z}{k}$ or $\sigma_z = \dfrac{\delta z}{\sqrt{2\,\delta\varphi}}$
- Further diagnostics may be helpful, if more calculations feasible

Tomographic reconstruction from two views

- Problem - reconstruct uniform-density object from two projections
  - 2 orthogonal, parallel projections (128 samples in each)
  - Gaussian noise added

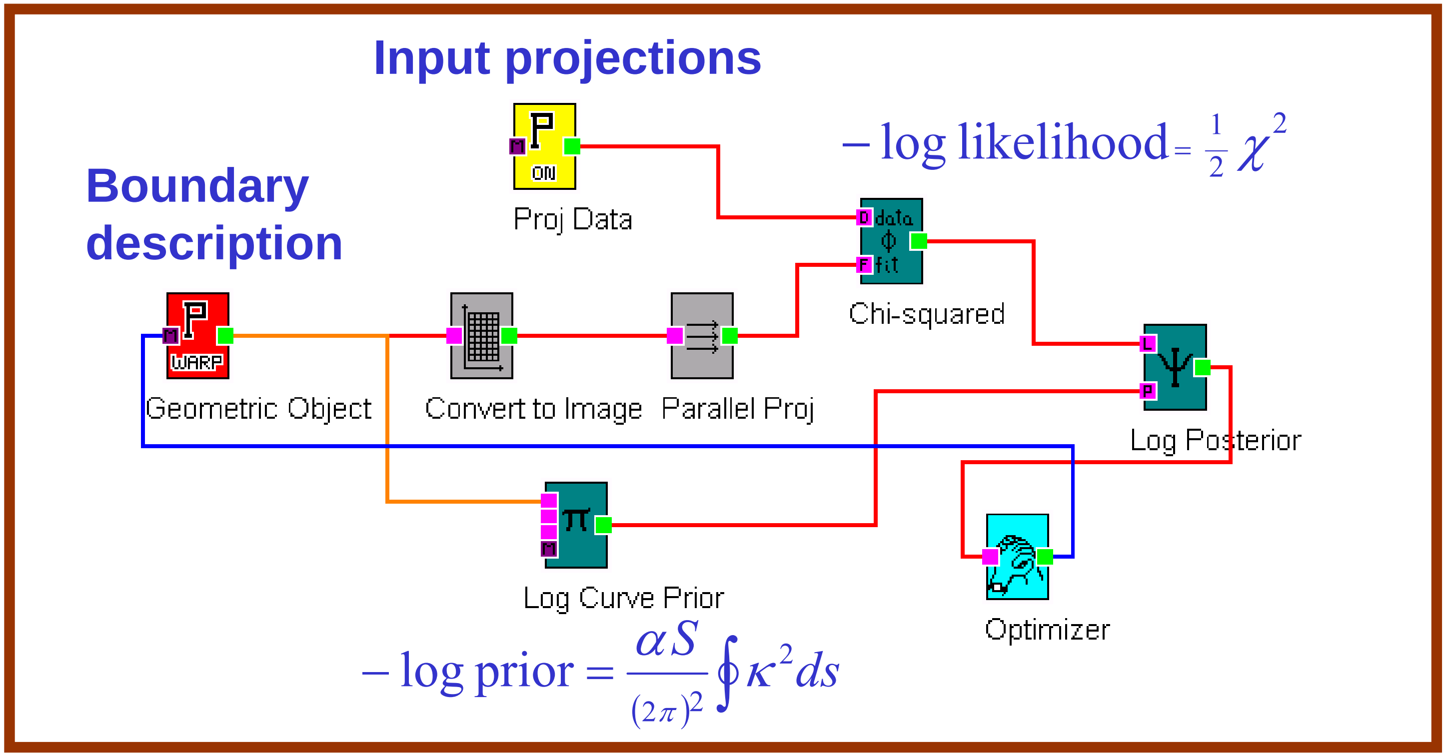
Original object

Two orthogonal projections with 5% rms noise



The Bayes Inference Engine

- BIE data-flow diagram to find max. a posteriori (MAP) solution

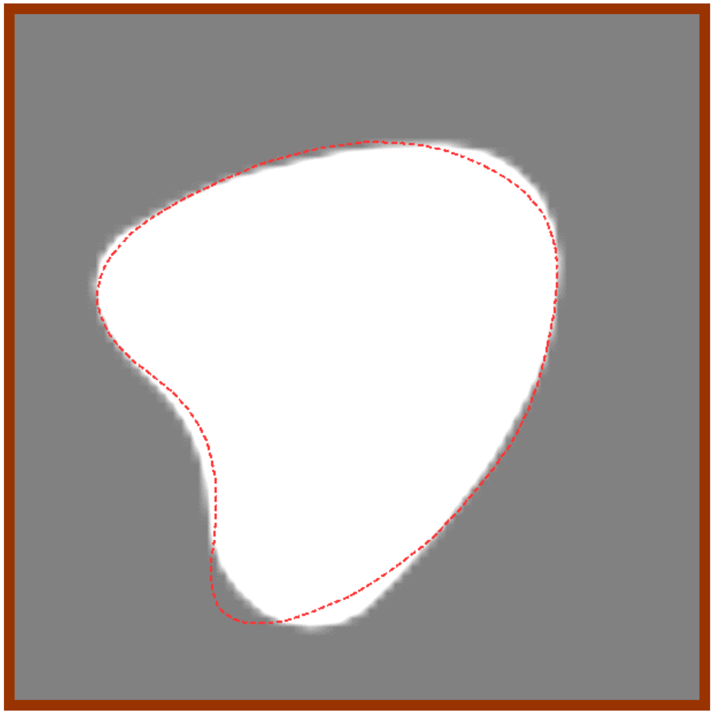


  - 0ptimizer uses gradients that are efficiently calculated by **adjoint differentiation**, a key capability of the BIE

MAP reconstruction – two views

- Model object in terms of:
  - deformable polygonal boundary with 50 vertices
  - boundary smoothness constraint
  - constant interior density
- Determine boundary that maximizes posterior probability
- Reconstruction not perfect, but very good for only two projections
- Question is:
  How do we quantify uncertainty in reconstruction?

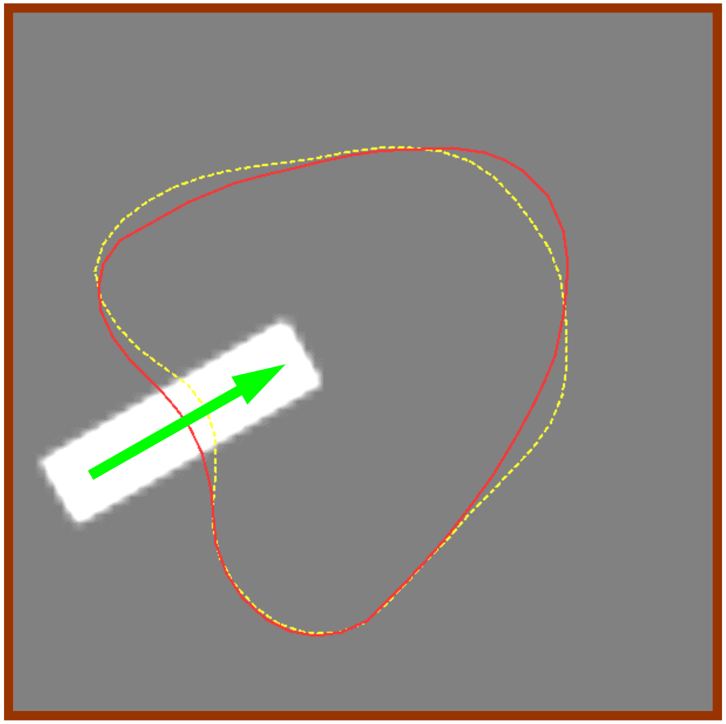
Reconstructed boundary (gray-scale) compared with original object (red line)



Tomographic reconstruction from two views

- Stiffness of model proportional to curvature of φ
- Displacement obtained by applying a force to MAP model and re-minimizing φ is proportional to a column of the **covariance matrix**
- Displacement divided by force
  - at position of force, it is proportional to variance there
  - elsewhere, it is proportional to covariance
- This approach may be efficient alternative to MCMC

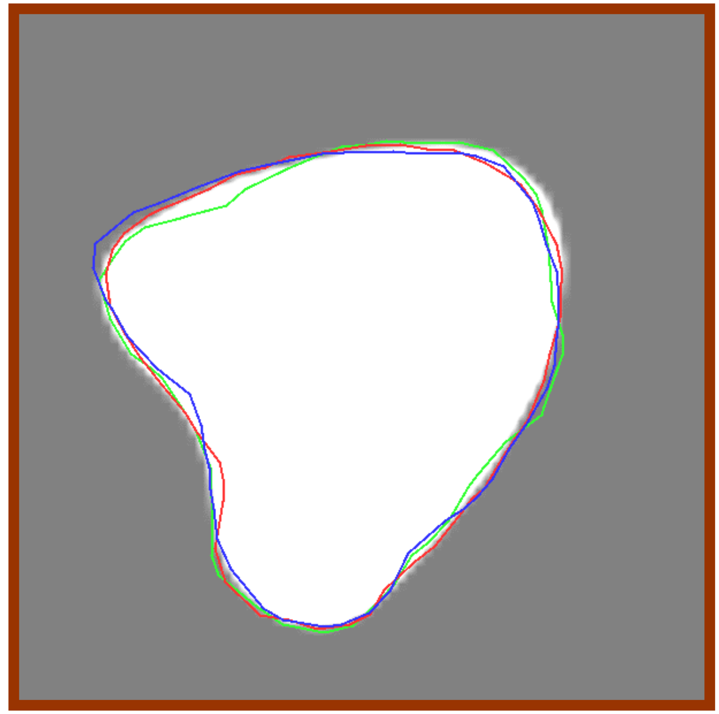
Applying force (white bar) to MAP boundary (red) moves it to new location (yellow-dashed)



Covariance using MCMC

- Use MCMC to draw samples from posterior
- Parameters consist of 50 vertices defining object boundary
- MCMC (Metropolis) 150,000 steps; display three selected boundaries
- Advantage: obtain full covariance matrix
- Disadvantage: calculation takes over 2000 times longer than technique of probing posterior

3 boundaries from 150,000 MCMC steps



compared uncertainties to MAP estimated object

Situations where probing covariance useful

- Technique will be most useful when
  - interest is in uncertainty in one or a few parameters or derived quantities out of many parameters
  - covariance matrix is not known (nor desired)
  - posterior can be well approximated by Gaussian pdf in parameters
  - optimization easy to do
  - gradient calculation (for optimization) can be done efficiently, e.g. by adjoint differentiation of the forward simulation code
- Technique may also be useful for exploring and quantifying
  - non-Gaussian posterior pdfs, including situations with inequality constraints, e.g., non-negativity
  - general pdfs; in contexts other than probabilistic inference
  - pdfs of self-optimizing natural systems (populations, bacteria, traffic)

Summary

- Technique has been presented that
  - is based on interpreting minus-log-posterior as physical potential energy
  - allows one to directly probe a specified component of covariance matrix by applying force to estimated model
  - replaces a stochastic calculation (e.g., MCMC) by a deterministic one
  - may efficiently provide uncertainty estimates in computational situations

Bibliography

- "The hard truth," K. M. Hanson and G. S. Cunningham, *Maximum Entropy and Bayesian Methods*, J. Skilling and S. Sibisi, eds., pp. 157-164 (Kluwer Academic, Dordrecht, 1996)
- Uncertainty assessment for reconstructions based on deformable models," K. M. Hanson et al., *Int. J. Imaging Syst. Technol.* **8**, pp. 506-512 (1997)
- "Operation of the Bayes Inference Engine," K. M. Hanson and G. S. Cunningham, *Maximum Entropy and Bayesian Methods*, W. von der Linden et al., eds., pp. 309-318 (Kluwer Academic, Dordrecht, 1999)
- *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, A. Griewank (SIAM, 2000)

This presentation available at http://www.lanl.gov/home/kmh/