Lecture 7

Backus-Gilbert Generalized Inverse and the Trade Off of Resolution and Variance

Syllabus

Lecture 01	Describing Inverse Problems
Lecture 02	Probability and Measurement Error, Part 1
Lecture 03	Probability and Measurement Error, Part 2
Lecture 04	The L_2 Norm and Simple Least Squares
Lecture 05	A Priori Information and Weighted Least Squared
Lecture 06	Resolution and Generalized Inverses
Lecture 07	**Backus-Gilbert Inverse and the Trade Off of Resolution and Variance**
Lecture 08	The Principle of Maximum Likelihood
Lecture 09	Inexact Theories
Lecture 10	Nonuniqueness and Localized Averages
Lecture 11	Vector Spaces and Singular Value Decomposition
Lecture 12	Equality and Inequality Constraints
Lecture 13	L_1 , L_∞ Norm Problems and Linear Programming
Lecture 14	Nonlinear Problems: Grid and Monte Carlo Searches
Lecture 15	Nonlinear Problems: Newton's Method
Lecture 16	Nonlinear Problems: Simulated Annealing and Bootstrap Confidence Intervals
Lecture 17	Factor Analysis
Lecture 18	Varimax Factors, Empircal Orthogonal Functions
Lecture 19	Backus-Gilbert Theory for Continuous Problems; Radon's Problem
Lecture 20	Linear Operators and Their Adjoints
Lecture 21	Fréchet Derivatives
Lecture 22	Exemplary Inverse Problems, incl. Filter Design
Lecture 23	Exemplary Inverse Problems, incl. Earthquake Location
Lecture 24	Exemplary Inverse Problems, incl. Vibrational Problems

Purpose of the Lecture

Introduce a new way to quantify the spread of resolution

Find the Generalized Inverse that minimizes this spread

Include minimization of the size of variance

Discuss how resolution and variance trade off

Part 1

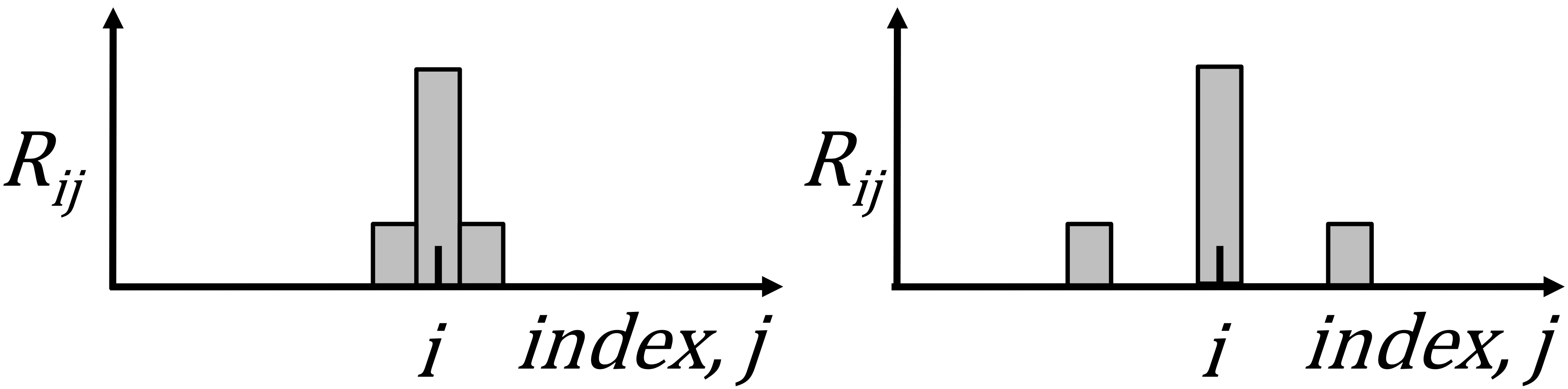
A new way to quantify the spread of resolution

criticism of Direchlet *spread()* functions

when **m** represents *m(x)*

is that they don't capture the sense of
"being localized" very well

These two rows of the model resolution matrix have the same Direchlet spread …



… but the left case is better “localized”

old way

$$\text{spread}(\mathbf{R}) = \sum_{i=1}^{M}\sum_{j=1}^{M}\left[R_{ij} - \delta_{ij}\right]^2$$

new way

$$\text{spread}(\mathbf{R}) = \sum_{i=1}^{M}\sum_{j=1}^{M} w(i,j)\left[R_{ij} - \delta_{ij}\right]^2$$

old way

$$\text{spread}(\mathbf{R}) = \sum_{i=1}^{M}\sum_{j=1}^{M}\left[R_{ij} - \delta_{ij}\right]^2$$

add this function

new way

$$\text{spread}(\mathbf{R}) = \sum_{i=1}^{M}\sum_{j=1}^{M} w(i,j)\left[R_{ij} - \delta_{ij}\right]^2$$

old way

$$\text{spread}(\mathbf{R}) = \sum_{i=1}^{M}\sum_{j=1}^{M}\left[R_{ij} - \delta_{ij}\right]^2$$

new way

Backus-Gilbert Spread Function

$$\text{spread}(\mathbf{R}) = \sum_{i=1}^{M}\sum_{j=1}^{M} w(i,j)\left[R_{ij} - \delta_{ij}\right]^2$$

grows with **physical distance** between model parameters

zero when i=j

$$\text{spread}(\mathbf{R}) = \sum_{i=1}^{M} \sum_{j=1}^{M} w(i,j) \left[R_{ij} - \delta_{ij} \right]^2$$

if $w(i,i)=0$ then

$$\text{spread}(\mathbf{R}) = \sum_{i=1}^{M}\sum_{j=1}^{M} w(i,j)\left[R_{ij} - \delta_{ij}\right]^2 = \sum_{i=1}^{M}\sum_{j=1}^{M} w(i,j)R_{ij}^2$$

for one spatial dimension

$\mathbf{m}$ is discretized version of $m(x)$

$$\mathbf{m} = [\, m(\Delta x), m(2\,\Delta x), \ldots\, m(M\,\Delta x) \,]^{\mathrm{T}}$$

$$w(i,j) = (i\text{-}j)^2$$

would work fine

for two spatial dimension

m is discretized version of $m(x,y)$

on $K \times L$ grid

$$\mathbf{m} = [\,m(x_1, y_1), m(x_1, y_2), \ldots\, m(x_K, y_L)\,]^T$$

$$w(i,j) = (x_i - x_j)^2 + (y_i - y_j)^2$$

would work fine

Part 2

Constructing a Generalized Inverse

under-determined problem

find $\mathbf{G}^{-g}$ that minimizes spread($\mathbf{R}$)

under-determined problem

find $\mathbf{G}^{-g}$ that minimizes spread($\mathbf{R}$)

with the constraint

$$\sum_{j=1}^{M} R_{ij} = [1]_i$$

under-determined problem

find $\mathbf{G}^{-g}$ that minimizes spread($\mathbf{R}$)

with the constraint

$$\sum_{j=1}^{M} R_{ij} = [1]_i$$

(since R_{ii} is not constrained by spread function) once again, solve for each row of $\mathbf{G}^{-g}$ separately

spread of kth row of resolution matrix $\mathbf{R}$

$$J_k = \sum_{l=1}^{M} w(l,k)\, R_{kl} R_{kl}$$

$$= \sum_{l=1}^{M} w(l,k) \left[\sum_{i=1}^{N} G_{ki}^{-g} G_{il} \right] \left[\sum_{j=1}^{N} G_{kj}^{-g} G_{jl} \right]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ki}^{-g} G_{kj}^{-g} \sum_{l=1}^{M} w(l,k) G_{il} G_{jl}$$

symmetric in i,j

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ki}^{-g} G_{kj}^{-g} [S_{ij}]_k \quad \text{with} \quad [S_{ij}]_k = \sum_{l=1}^{M} w(l,k) G_{il} G_{jl}$$

for the constraint

$$[1]_k = \sum_{k=1}^{M} R_{ik} = \sum_{k=1}^{M} \left[\sum_{j=1}^{N} G_{ij}^{-g} G_{jk} \right] = \sum_{j=1}^{N} G_{ij}^{-g} \sum_{k=1}^{M} G_{jk} = \sum_{j=1}^{N} G_{ij}^{-g} u_j$$

with $\quad u_j = \sum_{k=1}^{M} G_{jk}$

Lagrange Multiplier Equation

$$\Phi = \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ki}^{-g} G_{kj}^{-g} [S_{ij}]_k + 2\lambda \sum_{j=1}^{N} G_{kj}^{-g} u_j$$

now set $\partial \Phi / \partial G^{-g}{}_{kp}$

$$\partial \Phi / \partial G_{kp}^{-g} = 2 \sum_{i=1}^{N} [S_{pi}]_k G_{ki}^{-g} + 2\lambda u_p = 0$$

$$\mathbf{S}\,\mathbf{g}^{(k)} + \lambda \mathbf{u} = 0$$

with $\mathbf{g}^{(k)T}$ the *k*th row of $\mathbf{G}^{-g}$

solve simultaneously with

$$\mathbf{u}^{T}\,\mathbf{g}^{(k)} = 1$$

putting the two equations together

$$\begin{bmatrix} \mathbf{S}^{(\mathbf{k})} & \mathbf{u} \\ \mathbf{u}^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{g}^{(\mathrm{k})} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$$

construct inverse

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^{\mathrm{T}} & c \end{bmatrix} \begin{bmatrix} \mathbf{S}^{(\mathbf{k})} & \mathbf{u} \\ \mathbf{u}^{\mathrm{T}} & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{S}^{(\mathbf{k})} + \mathbf{b}\mathbf{u}^{\mathrm{T}} & \mathbf{A}\mathbf{u} \\ \mathbf{b}^{\mathrm{T}}\mathbf{S}^{(\mathbf{k})} + \mathbf{c}\mathbf{u}^{\mathrm{T}} & \mathbf{b}^{\mathrm{T}}\mathbf{u} \end{bmatrix}$$

A, **b**, c unknown

construct inverse

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^{\mathrm{T}} & c \end{bmatrix} \begin{bmatrix} \mathbf{S}^{(\mathrm{k})} & \mathbf{u} \\ \mathbf{u}^{\mathrm{T}} & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{S}^{(\mathrm{k})} + \mathbf{b}\mathbf{u}^{\mathrm{T}} & \mathbf{A}\mathbf{u} \\ \mathbf{b}^{\mathrm{T}}\mathbf{S}^{(\mathrm{k})} + c\mathbf{u}^{\mathrm{T}} & \mathbf{b}^{\mathrm{T}}\mathbf{u} \end{bmatrix}$$

$$\mathbf{A}\mathbf{S}^{(\mathrm{k})} + \mathbf{b}\mathbf{u}^{\mathrm{T}} = \mathbf{I} \quad \text{so that} \quad \mathbf{A} = \left[\mathbf{S}^{(\mathrm{k})}\right]^{-1}\left[\mathbf{I} - \mathbf{b}\mathbf{u}^{\mathrm{T}}\right]$$

$$\mathbf{A}\mathbf{u} = \mathbf{O} \quad \text{so that} \quad \left[\mathbf{S}^{(\mathrm{k})}\right]^{-1}\mathbf{u} = \mathbf{b}\mathbf{u}^{\mathrm{T}}\mathbf{S}^{(\mathrm{k})}\mathbf{u}$$

$$\text{so} \quad \mathbf{b} = \frac{\left[\mathbf{S}^{(\mathrm{k})}\right]^{-1}\mathbf{u}}{\mathbf{u}^{\mathrm{T}}[\mathbf{S}^{(\mathrm{k})}]^{-1}\mathbf{u}}$$

$$\mathbf{b}^{\mathrm{T}}\mathbf{S}^{(\mathrm{k})} + c\mathbf{u}^{\mathrm{T}} = 0 \quad \text{so that} \quad c = \frac{-1}{\mathbf{u}^{\mathrm{T}}[\mathbf{S}^{(\mathrm{k})}]^{-1}\mathbf{u}}$$

$$\begin{bmatrix} \mathbf{g}^{(\mathrm{k})} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^{\mathrm{T}} & c \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$$

$$\mathbf{g}^{(\mathrm{k})} = \mathbf{b} = \frac{\left[\mathbf{S}^{(\mathrm{k})}\right]^{-1}\mathbf{u}}{\mathbf{u}^{\mathrm{T}}\left[\mathbf{S}^{(\mathrm{k})}\right]^{-1}\mathbf{u}}$$

$$\mathbf{g}^{(\mathrm{k})\mathrm{T}} = \frac{\mathbf{u}^{\mathrm{T}}\left[\mathbf{S}^{(\mathrm{k})}\right]^{-1}}{\mathbf{u}^{\mathrm{T}}\left[\mathbf{S}^{(\mathrm{k})}\right]^{-1}\mathbf{u}}$$

*k*th row of $\mathbf{G}^{-g}$

Backus-Gilbert Generalized Inverse

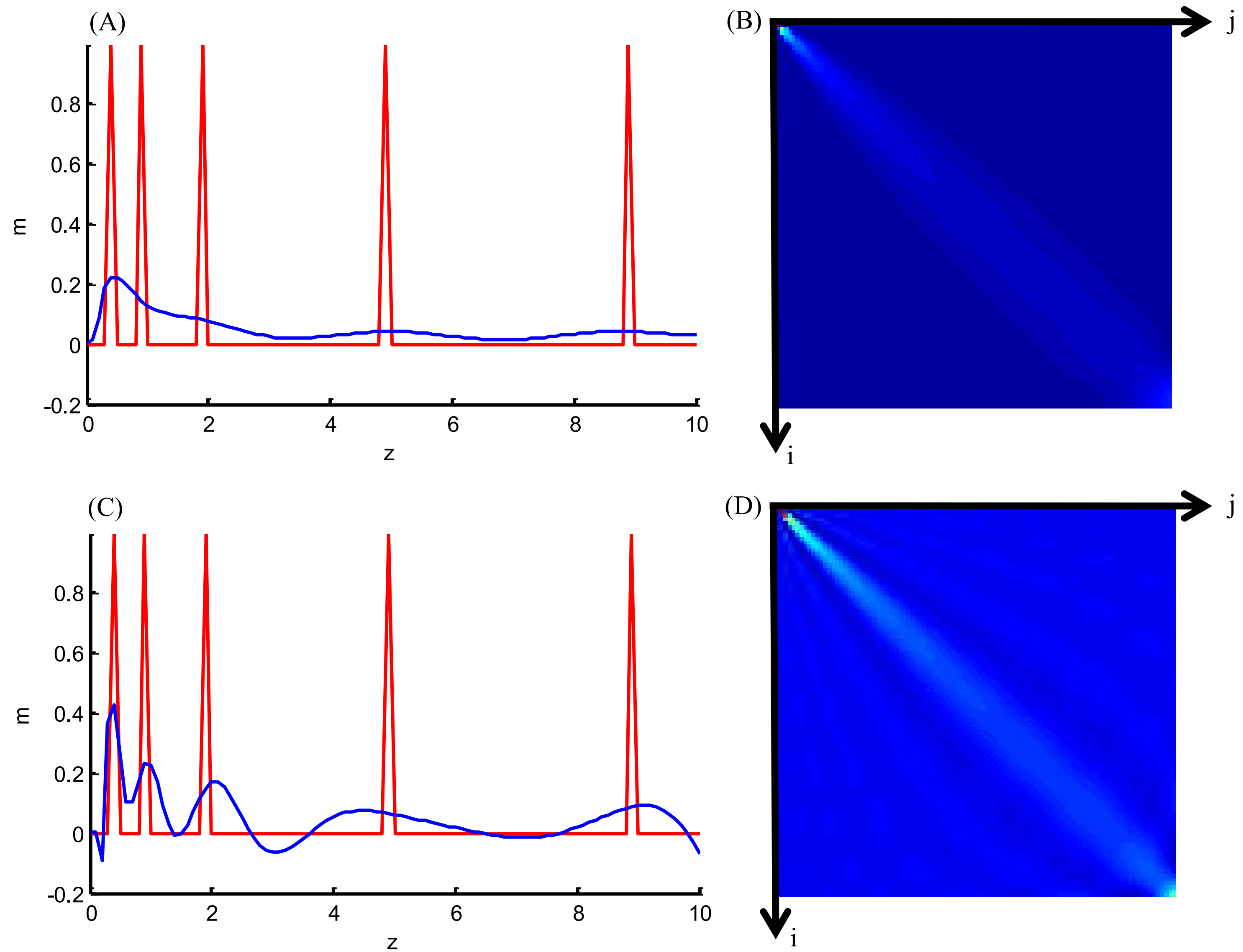
(analogous to Minimum Length Generalized Inverse)
written in terms of components

$$G_{kl}^{-g} = \frac{\sum_{i=1}^{N} [S_{il}]_k^{-1} u_i}{\sum_{i=1}^{N} \sum_{j=1}^{N} u_i [S_{il}]_k^{-1} u_j}$$

with

$$[S_{ij}]_k = \sum_{l=1}^{M} w(l,k) G_{il} G_{jl}$$

$$u_j = \sum_{k=1}^{M} G_{jk}$$



(A)
m
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
z
(B)
j
i
(C)
m
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
z
(D)
j
i

Part 3

Include minimization of the size of variance

minimize

$$\alpha\ \mathrm{spread}(\mathbf{R}) + (1-\alpha)\ \mathrm{size}([\mathrm{cov_u}\ \mathbf{m}])$$

$$= \alpha \sum_{i=1}^{M} \sum_{j=1}^{M} w(i,j) R_{ij}^2 + (1-\alpha) \sum_{i=1}^{M} [\mathrm{cov_u}\ \mathbf{m}]_{ii}$$

new version of J_k

$$J'_k = \alpha \sum_{l=1}^{M} w(k,l) R_{kl}^2 + (1-\alpha)[\mathrm{cov_u}\ \mathbf{m}]_{kk}$$

$$J'_k = \alpha \sum_{l=1}^{M} w(k,l) R_{kl}^2 + (1-\alpha)[\mathrm{cov_u}\ \mathbf{m}]_{kk}$$

$$= \alpha \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ki}^{-g} G_{kj}^{-g} [S_{ij}]_k + \ (1-\alpha) \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ki}^{-g} G_{kj}^{-g} [\mathrm{cov_u}\ \mathbf{d}]_{ij}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ki}^{-g} G_{kj}^{-g} [S'_{ij}]_k$$

with $[S'_{ij}]_k = \alpha [S_{ij}]_k + (1-\alpha)[\mathrm{cov_u}\ \mathbf{d}]_{ij}$

$$J'_k = \alpha \sum_{l=1}^{M} w(k,l) R_{kl}^2 + (1-\alpha)[\mathrm{cov_u}\ \mathbf{m}]_{kk}$$

same

$$= \alpha \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ki}^{-g} G_{kj}^{-g} [S_{ij}]_k + (1-\alpha) \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ki}^{-g} G_{kj}^{-g} [\mathrm{cov_u}\ \mathbf{d}]_{ij}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ki}^{-g} G_{kj}^{-g} [S'_{ij}]_k$$

with $[S'_{ij}]_k = \alpha [S_{ij}]_k + (1-\alpha)[\mathrm{cov_u}\ \mathbf{d}]_{ij}$

so adding size of variance
is just a small modification to **S**

Backus-Gilbert Generalized Inverse
(analogous to Damped Minimum Length)
written in terms of components

$$G_{kl}^{-g} = \frac{\sum_{i=1}^{N} [S'_{il}]_k^{-1} u_i}{\sum_{i=1}^{N} \sum_{j=1}^{N} u_i [S'_{il}]_k^{-1} u_j}$$

with

$$[S'_{ij}]_k = \alpha [S_{ij}]_k + (1-\alpha)[\mathrm{cov_u}\, \mathbf{d}]_{ij}$$

$$[S_{ij}]_k = \sum_{l=1}^{M} w(l,k) G_{il} G_{jl} \qquad u_j = \sum_{k=1}^{M} G_{jk}$$

In MatLab

```
GMG = zeros(M,N);
u = G*ones(M,1);
for k = [1:M]
  S = G * diag(([1:M]-k).^2) * G';
  Sp = alpha*S + (1-alpha)*eye(N,N);
  uSpinv = u'/Sp;
  GMG(k,:) = uSpinv / (uSpinv*u);
end
```

$20 Reward!

to the first person who sends me MatLab code that computes the BG generalized inverse without a **`for`** loop

(but no creation of huge 3-indexed quantities, please. Memory requirements need to be similar to my code)

The Direchlet analog of the

Backus-Gilbert Generalized Inverse

is the

Damped Minimum Length
Generalized Inverse

special case #2

$$\text{Minimize:}\ \ \underbrace{\alpha_1}_{0}\ \text{spread}(\mathbf{N}) + \underbrace{\alpha_2}_{1}\ \text{spread}(\mathbf{R}) + \underbrace{\alpha_3}_{\varepsilon^2}\ \text{size}([\text{cov}_\text{u}\mathbf{m}])$$

$$\alpha_1\ [\mathbf{G}^\text{T}\mathbf{G}]\ \mathbf{G}^{-\text{g}} + \ \mathbf{G}^{-\text{g}}\Big[\alpha_2\ [\mathbf{G}\mathbf{G}^\text{T}] + \alpha_3 \overbrace{[\text{cov}_\text{u}\mathbf{d}]}^{\mathbf{I}}\Big] = \ [\alpha_1 + \alpha_2]\mathbf{G}^\text{T}$$

$$\mathbf{G}^{-\text{g}}[\mathbf{G}\mathbf{G}^\text{T}+\varepsilon^2\mathbf{I}] = \mathbf{G}^\text{T}$$

$$\mathbf{G}^{-\text{g}} = \mathbf{G}^\text{T}\ [\mathbf{G}\mathbf{G}^\text{T}+\varepsilon^2\mathbf{I}]^{-1}$$

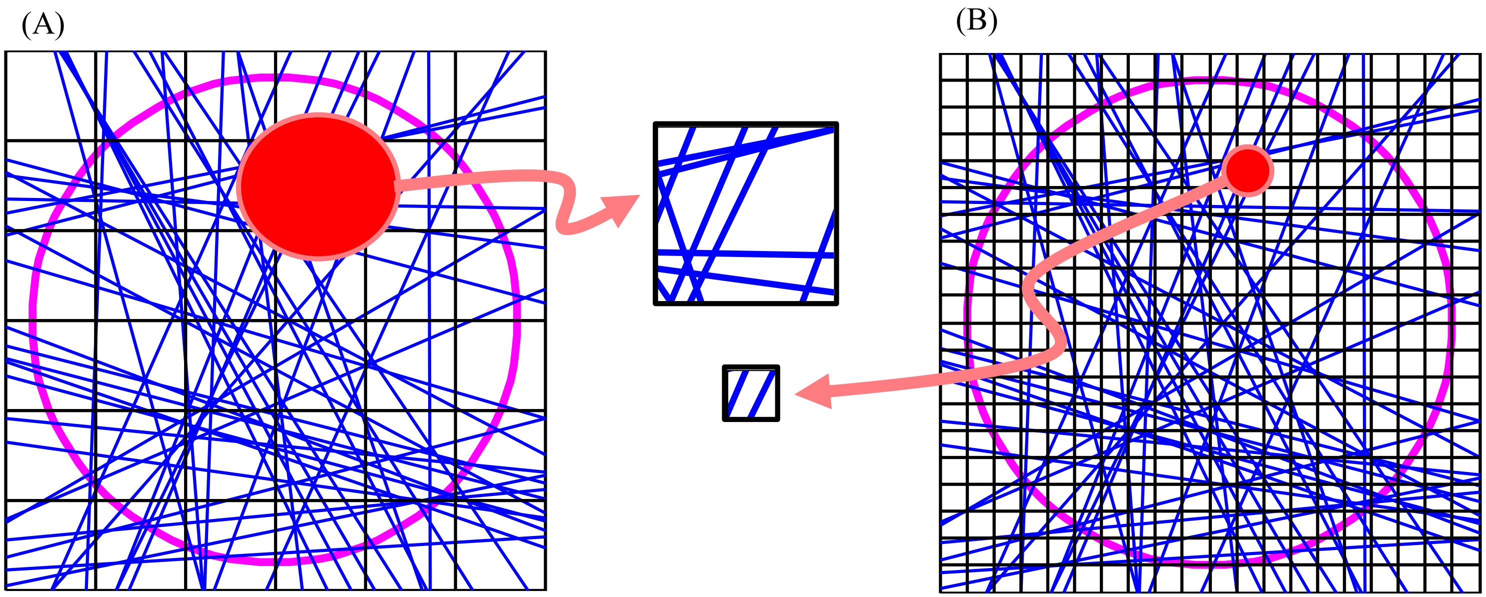
damped minimum length

Part 4

the trade-off of resolution and variance

the value of
a localized average with small spread
is controlled by few data
and so has large variance

the value of
a localized average with large spread
is controlled by many data
and so has small variance


(A)
(B)

small spread
large variance

large spread
small variance

α near 1

small spread

large variance

α near 0

large spread

small variance

Trade-Off Curves