Lecture 6

Resolution and Generalized Inverses

Syllabus

Purpose of the Lecture

Introduce the idea of a Generalized Inverse,
the Data and Model Resolution Matrices
and the Unit Covariance Matrix

Quantify the spread of resolution and the size of the covariance

Use the maximization of resolution and/or covariance as the guiding principle for solving inverse problems

Part 1

The Generalized Inverse,

the Data and Model Resolution Matrices

and the Unit Covariance Matrix

all of the solutions

$$\mathbf{m}^{\mathrm{est}} = [\mathbf{G}^{\mathrm{T}}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{d}$$

$$\mathbf{m}^{est} = \mathbf{G}^{\mathrm{T}}[\mathbf{G}\mathbf{G}^{\mathrm{T}}]^{-1}\mathbf{d}$$

$$\mathbf{m}^{\mathrm{est}} = [\mathbf{G}^{\mathrm{T}}\mathbf{G} + \varepsilon^2\mathbf{I}]^{-1}\,\mathbf{G}^{\mathrm{T}}\mathbf{d}$$

$$\mathbf{m}^{\mathrm{est}} = [\mathbf{G}^{\mathrm{T}}\mathbf{W}_e\mathbf{G} + \varepsilon^2\mathbf{W}_m]^{-1}\,[\mathbf{G}^{\mathrm{T}}\mathbf{W}_e\,\mathbf{d} + \varepsilon^2\mathbf{W}_m\langle\mathbf{m}\rangle\,]$$

of the form

$$\mathbf{m}^{\mathrm{est}} = \mathbf{M}\mathbf{d} + \mathbf{v}$$

$$\mathbf{m}^{\text{est}} = \mathbf{Md} + \mathbf{v}$$

let's focus on
this matrix

$$\mathbf{m}^{est} = \mathbf{G}^{-g}\mathbf{d} + \mathbf{v}$$

rename it the "generalized inverse" and use the symbol $\mathbf{G}^{-g}$

(let's ignore the vector $\mathbf{v}$ for a moment)

Generalized Inverse $\mathbf{G}^{-g}$

operates on the data to give an estimate of the model parameters

if

$$\mathbf{d}^{pre} = \mathbf{G}\mathbf{m}^{est}$$

then

$$\mathbf{m}^{est} = \mathbf{G}^{-g}\mathbf{d}^{obs}$$

Generalized Inverse $\mathbf{G}^{-g}$

if $\mathbf{d}^{pre} = \mathbf{G}\mathbf{m}^{est}$ then $\mathbf{m}^{est} = \mathbf{G}^{-g}\mathbf{d}^{obs}$

sort of looks like a matrix inverse

except

$M{\times}N$, not square

and

$\mathbf{G}\mathbf{G}^{-g} \neq \mathbf{I}$ and $\mathbf{G}^{-g}\mathbf{G} \neq \mathbf{I}$

so actually
the generalized inverse is not a
matrix inverse at all

plug one equation into the other

$$\mathbf{d}^{\mathrm{pre}} = \mathbf{G}\mathbf{m}^{\mathrm{est}} \quad \text{and} \quad \mathbf{m}^{\mathrm{est}} = \mathbf{G}^{-g}\mathbf{d}^{\mathrm{obs}}$$

$$\mathbf{d}^{\mathrm{pre}} = \mathbf{G}\mathbf{m}^{\mathrm{est}} = \mathbf{G}\left[\mathbf{G}^{-g}\mathbf{d}^{\mathrm{obs}}\right] = [\mathbf{G}\mathbf{G}^{-g}]\mathbf{d}^{\mathrm{obs}} = \mathbf{N}\mathbf{d}^{\mathrm{obs}}$$

$$\mathbf{d}^{\mathrm{pre}} = \mathbf{N}\mathbf{d}^{\mathrm{obs}} \quad \text{with} \quad \mathbf{N} = \mathbf{G}\mathbf{G}^{-g}$$

"data resolution matrix"

Data Resolution Matrix, $\mathbf{N}$

$$\mathbf{d}^{\mathbf{pre}} = \mathbf{N}\mathbf{d}^{\mathbf{obs}}$$

How much does d_i^{obs} contribute to its own prediction?

if

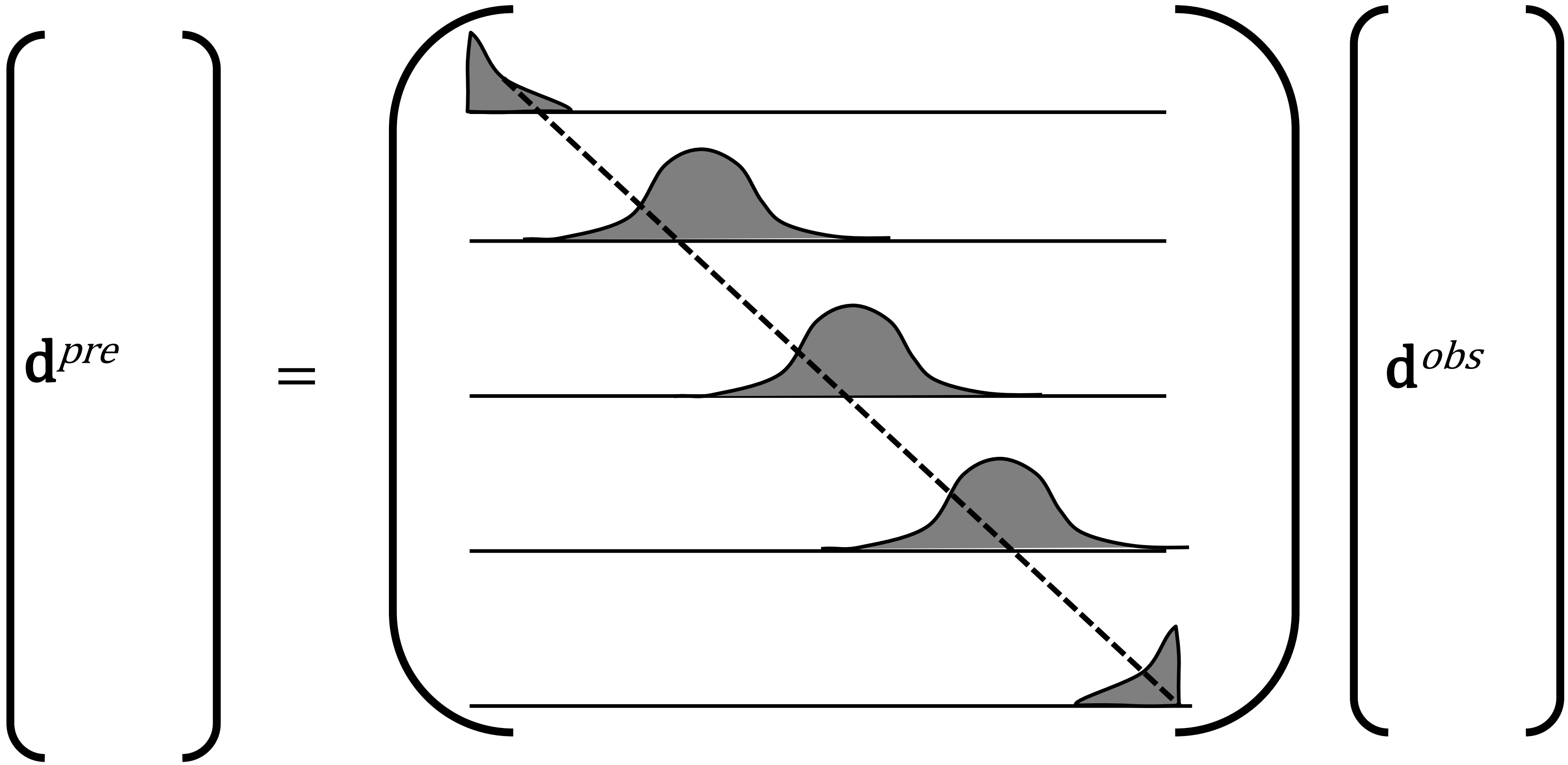
$$\mathbf{N}=\mathbf{I}$$

$$\mathbf{d}^{\mathbf{pre}} = \mathbf{d}^{\mathbf{obs}}$$

$$d_i^{pre} = d_i^{obs}$$

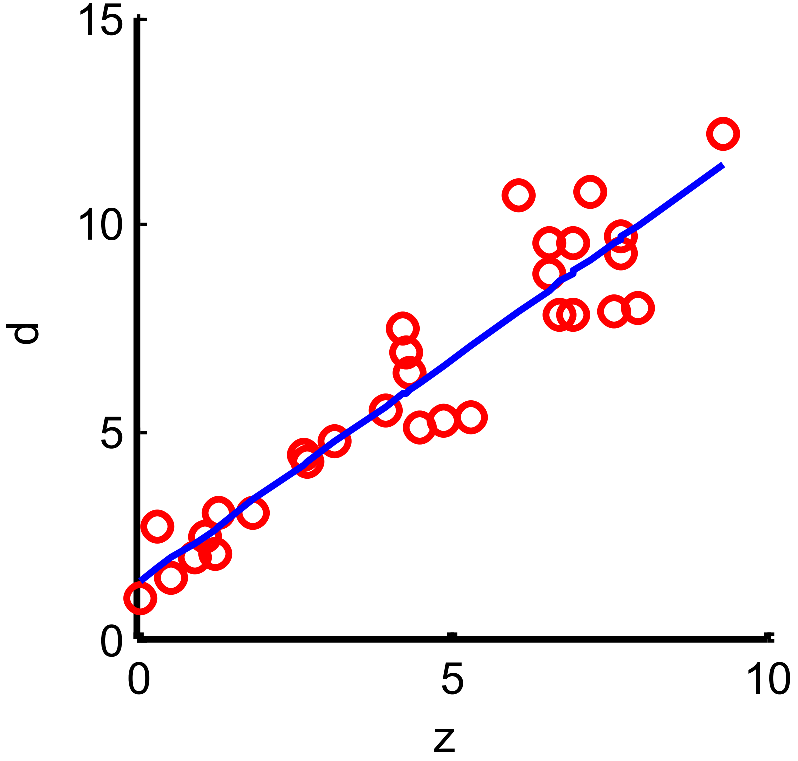
d_i^{obs} completely controls its own prediction

(A)

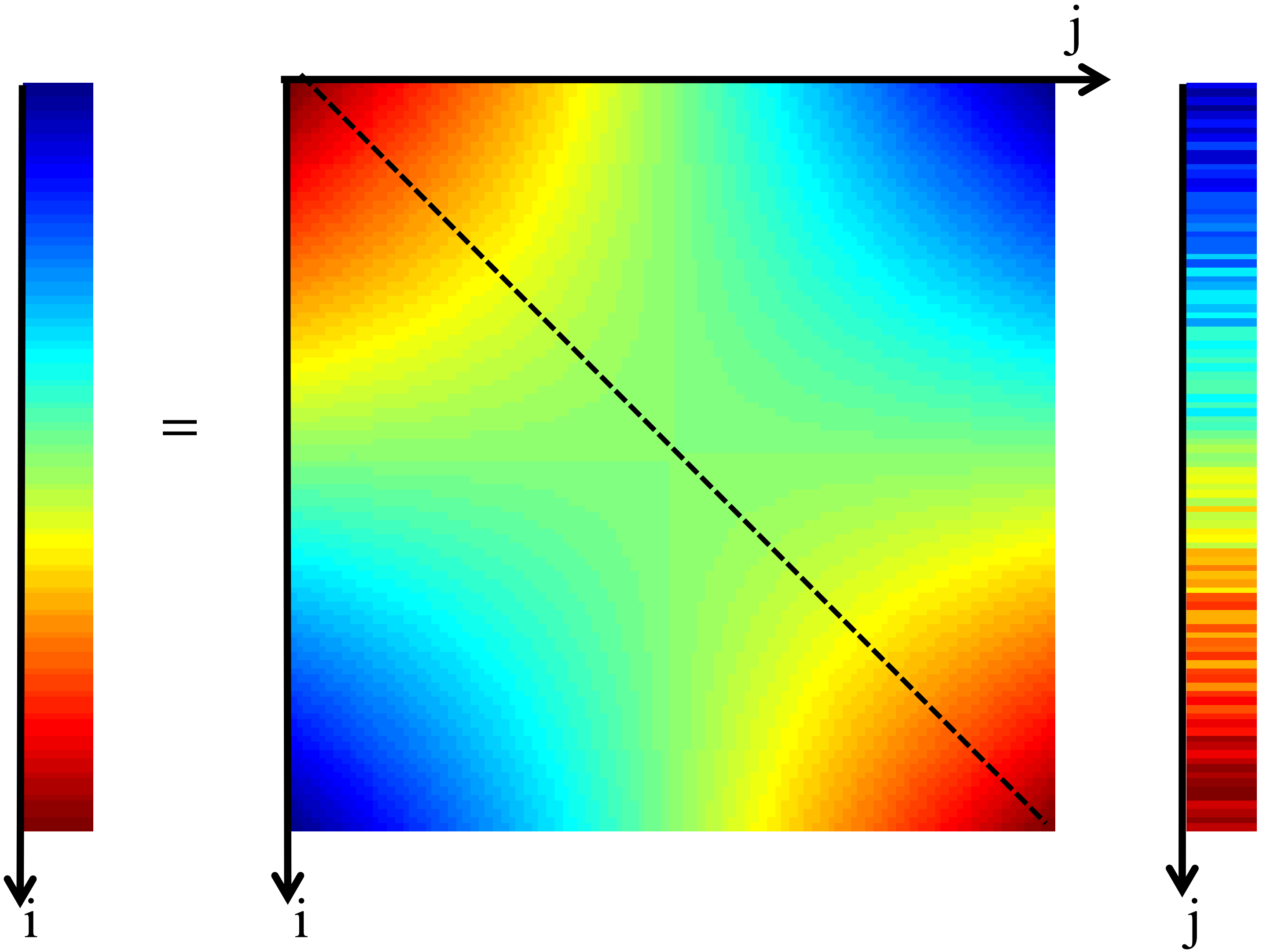


The closer $\mathbf{N}$ is to $\mathbf{I}$, the more d_i^{obs} controls its own prediction

straight line problem



$$\mathbf{d}^{\mathrm{pre}} = \mathbf{N} \; \mathbf{d}^{\mathrm{obs}}$$



only the data at the ends control their own prediction plug one equation into the other

$$\mathbf{d}^{\text{obs}} = \mathbf{G}\mathbf{m}^{\text{true}} \quad \text{and} \quad \mathbf{m}^{\text{est}} = \mathbf{G}^{-g}\mathbf{d}^{\text{obs}}$$

$$\mathbf{m}^{\text{est}} = \mathbf{G}^{-g}\mathbf{d}^{\text{obs}} = \mathbf{G}^{-g}[\mathbf{G}\mathbf{m}^{\text{true}}] = [\mathbf{G}^{-g}\mathbf{G}]\mathbf{m}^{\text{true}} = \mathbf{R}\mathbf{m}^{\text{true}}$$

$$\mathbf{m}^{\text{est}} = \mathbf{R}\mathbf{m}^{\text{true}} \quad \text{with} \quad \mathbf{R} = \mathbf{G}^{-g}\mathbf{G}$$

"model resolution matrix"

Model Resolution Matrix, **R**

$$\mathbf{m}^{\mathbf{est}} = \mathbf{R}\mathbf{m}^{\mathbf{true}}$$

How much does m_i^{true} contribute to its own estimated value?

if

$$\mathbf{R}=\mathbf{I}$$

$$\mathbf{m}^{\mathbf{est}} = \mathbf{m}^{\mathbf{true}}$$

$$m_i^{est} = m_i^{true}$$

m_i^{est} reflects m_i^{true} only

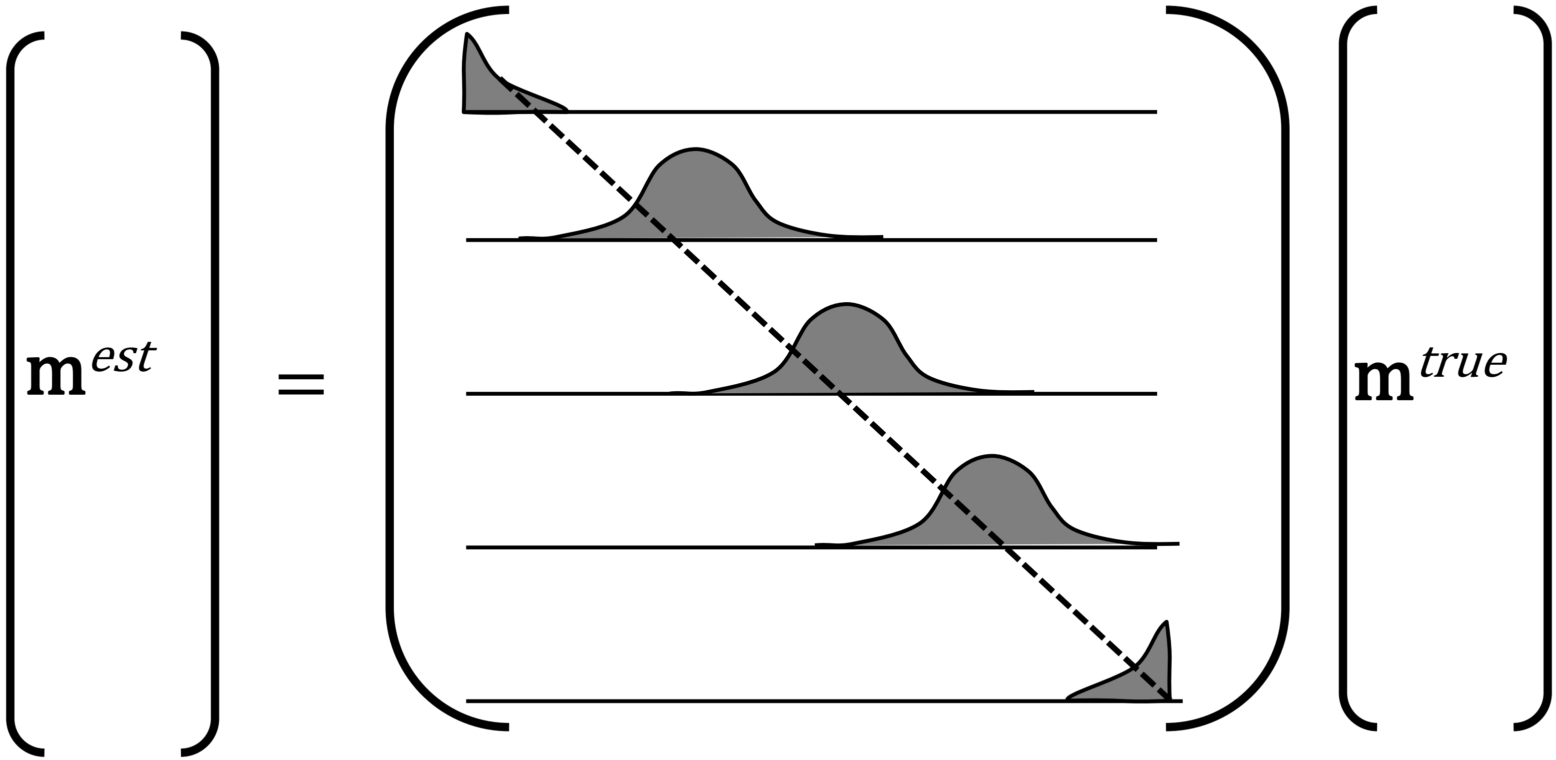
else if

R≠I

$$m_i^{est} =$$

$$\ldots + R_{i,i-1} m_{i-1}^{true} + R_{i,i} m_i^{true} + R_{i,i+1} m_{i+1}^{true} + \ldots$$

m_i^{est} is a **weighted average** of all the elements of $\mathbf{m}^{\mathrm{true}}$

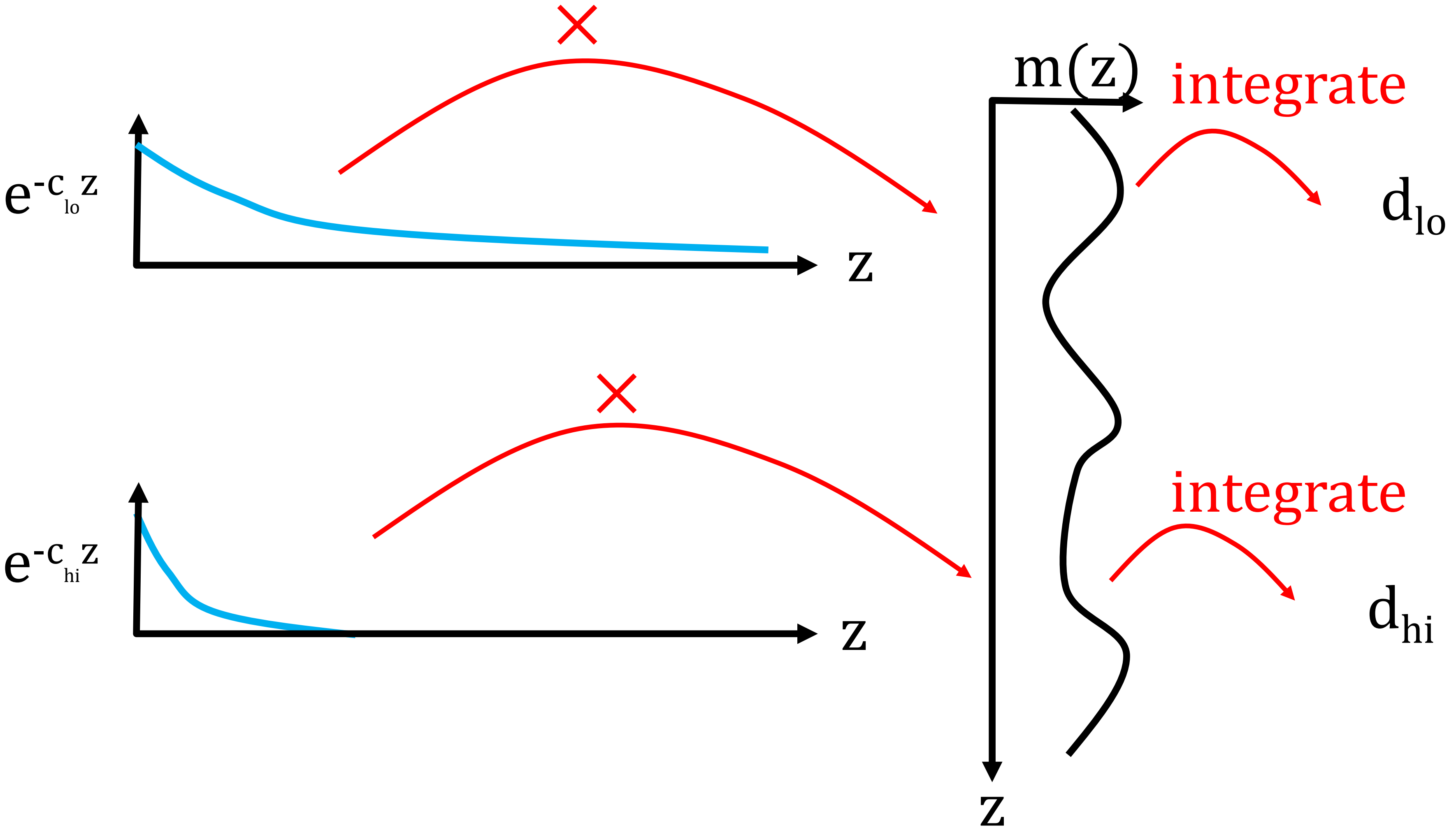


The closer $\mathbf{R}$ is to $\mathbf{I}$, the more m_i^{est} reflects only m_i^{true}

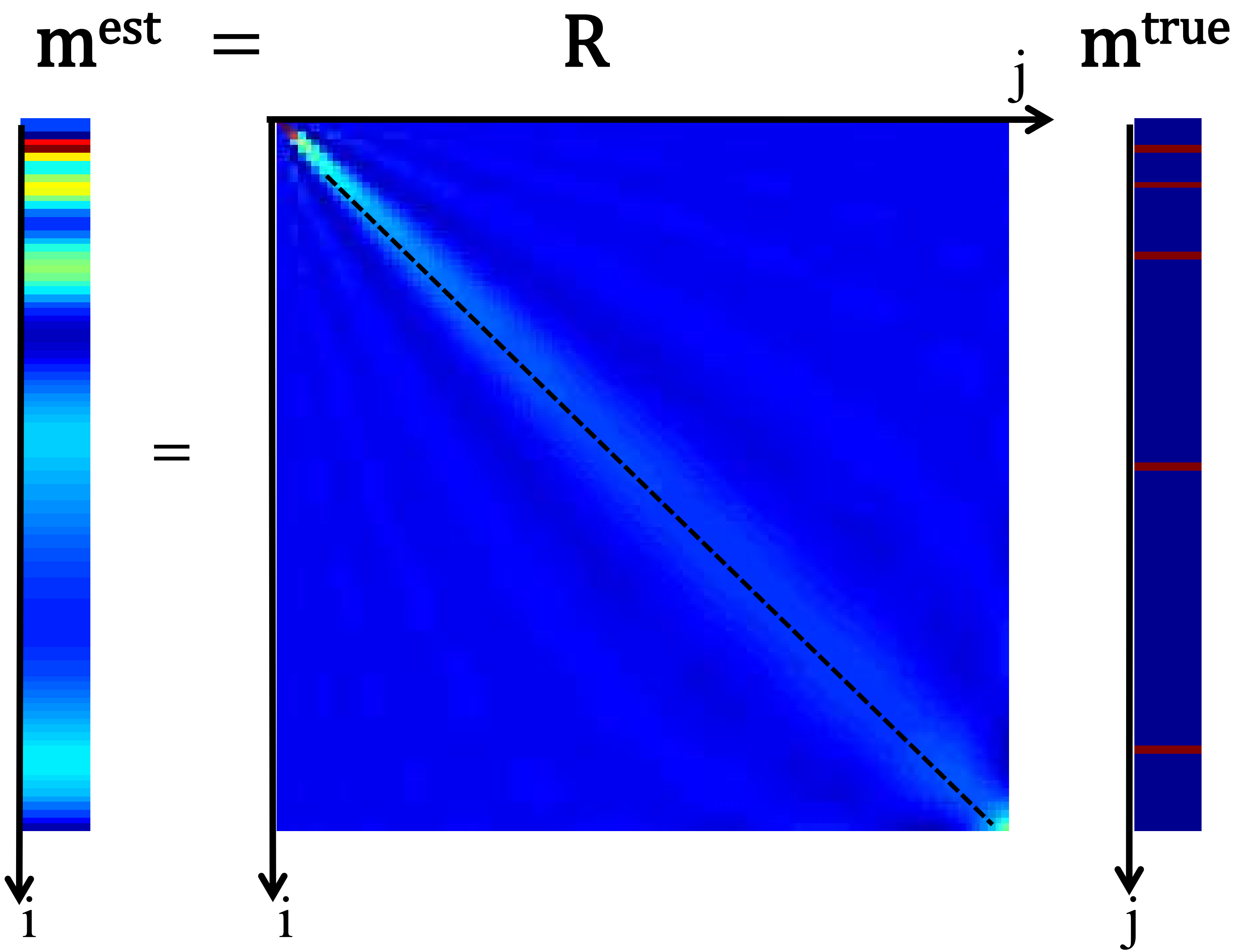
Discrete version of Laplace Transform

$$d(c) = \int_0^\infty \exp(-cz)\, m(z)\, dz \longrightarrow d_i = \sum_{j=1}^{M} \exp(-c_i z_j)\, m_j$$

large c: d is "shallow" average of m(z)
small c: d is "deep" average of m(z)



$e^{-c_{lo}z}$
z
$e^{-c_{hi}z}$
z
m(z)
z
integrate
d_{lo}
integrate
d_{hi}



the shallowest model parameters are “best resolved”

Covariance associated with the Generalized Inverse

"unit covariance matrix"
divide by σ^2 to remove effect of the overall magnitude of the measurement error

$$[\mathrm{cov}_u\ \mathbf{m}] = \sigma^{-2}\mathbf{G}^{-g}[\mathrm{cov}\ \mathbf{d}]\mathbf{G}^{-gT} = \mathbf{G}^{-g}\mathbf{G}^{-gT}$$

unit covariance for straight line problem

$$[\mathrm{cov}_\mathrm{u}\ \mathbf{m}] = \frac{1}{N\sum z_i^2 - (\sum z_i)^2}\begin{bmatrix} N & -\sum z_i \\ -\sum z_i & \sum z_i^2 \end{bmatrix}$$

model parameters uncorrelated when this term zero
happens when data are centered about the origin

Part 2

The spread of resolution and the size of the covariance

a resolution matrix has small spread if only its main diagonal has large elements

it is close to the identity matrix

“Dirichlet” Spread Functions

$$\text{spread}(\mathbf{N}) = \|\mathbf{N} - \mathbf{I}\|_2^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \left[N_{ij} - \delta_{ij}\right]^2$$

$$\text{spread}(\mathbf{R}) = \|\mathbf{R} - \mathbf{I}\|_2^2 = \sum_{i=1}^{M} \sum_{j=1}^{M} \left[R_{ij} - \delta_{ij}\right]^2$$

a unit covariance matrix has small size
if its diagonal elements are small

error in the data corresponds to only
small error in the model parameters

(ignore correlations)

$$\mathrm{size}([\mathrm{cov}_{\mathrm{u}}\mathbf{m}]) = \left\|[\mathrm{var}_{\mathrm{u}}\ \mathbf{m}]^{1/2}\right\|_2^2 = \sum_{i=1}^{M}[\mathrm{cov}\ _{\mathrm{u}}\mathbf{m}]_{ii}$$

Part 3

minimization of
spread of resolution
and/or
size of covariance
as the guiding principle
for creating a generalized inverse

over-determined case

note that for
simple least squares

$$\mathbf{G}^{-g} = [\mathbf{G}^{T}\mathbf{G}]^{-1}\mathbf{G}^{T}$$

model resolution

$$\mathbf{R}=\mathbf{G}^{-g}\mathbf{G} = [\mathbf{G}^{T}\mathbf{G}]^{-1}\mathbf{G}^{T}\mathbf{G}=\mathbf{I}$$

always the identify matrix

suggests that we try to minimize the spread of the data resolution matrix, **N**

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

spread of the k-th row of **N**

$$J_k = \sum_{i=1}^{N} (N_{ki} - \delta_{ki})^2 = \sum_{i=1}^{N} {N_{ki}}^2 - 2\sum_{i=1}^{N} N_{ki}\delta_{ki} + \sum_{i=1}^{N} \delta_{ki}^2$$

now compute

$$\partial J_k / \partial G_{qr}^{-g} = 0$$

first term

$$\frac{\partial}{\partial G_{qr}^{-g}}\left[\sum_{i=1}^{N}\left[\sum_{j=1}^{M} G_{kj} G_{ji}^{-g}\right]\left[\sum_{p=1}^{M} G_{kp} G_{pi}^{-g}\right]\right]=$$

$$\frac{\partial}{\partial G_{qr}^{-g}}\left[\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{p=1}^{M} G_{ji}^{-g} G_{pi}^{-g} G_{kj} G_{kp}\right]=$$

$$2\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{p=1}^{M} \delta_{jq}\delta_{ir} G_{pi}^{-g} G_{kj} G_{kp}=$$

$$2\sum_{p=1}^{M} G_{pr}^{-g} G_{kq} G_{kp}$$

second term

$$-2\frac{\partial}{\partial G_{qr}^{-g}}\sum_{i=1}^{N}\sum_{j=1}^{M}G_{kj}G_{ji}^{-g}\delta_{ki} = \sum_{i=1}^{N}\sum_{j=1}^{M}G_{kj}\delta_{jq}\delta_{kr}\delta_{ki} = -2G_{kq}\delta_{kr}$$

third term is zero

putting it all together

$$\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{-g} = \mathbf{G}^{\mathrm{T}}$$

which is just
simple least squares

$$\mathbf{G}^{-g} = [\mathbf{G}^{\mathrm{T}}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}$$

the simple least squares solution
minimizes the spread of data resolution
and
has zero spread of the model resolution

under-determined case

note that for
minimum length solution

$$\mathbf{G}^{-g} = \mathbf{G}^{T}\,[\mathbf{G}\mathbf{G}^{T}]^{-1}$$

data resolution

$$\mathbf{N} = \mathbf{G}\mathbf{G}^{-g} = \mathbf{G}\,\mathbf{G}^{T}\,[\mathbf{G}\mathbf{G}^{T}]^{-1} = \mathbf{I}$$

always the identify matrix

suggests that we try to minimize the spread of the model resolution matrix, **R**

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

$$[\mathbf{G}\mathbf{G}^{\mathrm{T}}]\mathbf{G}^{-g} = \mathbf{G}^{\mathrm{T}}$$

which is just
minimum length solution

$$\mathbf{G}^{-g} = \mathbf{G}^{\mathrm{T}}\,[\mathbf{G}\mathbf{G}^{\mathrm{T}}]^{-1}$$

the minimum length solution
minimizes the spread of model resolution
and
has zero spread of the data resolution

general case

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

leads to

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

general case

$$\text{Minimize: } \alpha_1 \text{ spread}(\mathbf{N}) + \alpha_2 \text{ spread}(\mathbf{R}) + \alpha_3 \text{ size}([\text{cov}_u \mathbf{m}])$$

leads to

$$\alpha_1 \, [\mathbf{G}^T \mathbf{G}] \, \mathbf{G}^{-g} + \mathbf{G}^{-g} \Big[\alpha_2 \, [\mathbf{G}\mathbf{G}^T] + \alpha_3 [\text{cov}_u \mathbf{d}] \Big] = [\alpha_1 + \alpha_2] \mathbf{G}^T$$

a Sylvester Equation, so explicit solution in terms of matrices

special case #1

$$\text{Minimize: } \alpha_1 \text{ spread}(\mathbf{N}) + \alpha_2 \text{ spread}(\mathbf{R}) + \alpha_3 \text{ size}([\text{cov}_u \mathbf{m}])$$

1 0 ε^2

$\mathbf{I}$

$$\alpha_1 \, [\mathbf{G}^T\mathbf{G}] \, \mathbf{G}^{-g} + \mathbf{G}^{-g}\Big[\alpha_2 \, [\mathbf{G}\mathbf{G}^T] + \alpha_3 [\text{cov}_u \mathbf{d}]\Big] = [\alpha_1 + \alpha_2]\mathbf{G}^T$$

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

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

damped least squares

special case #2

Minimize: $\underset{0}{\alpha_1}$ spread($\mathbf{N}$) + $\underset{1}{\alpha_2}$ spread($\mathbf{R}$) + $\underset{\varepsilon^2}{\alpha_3}$ size([$\mathrm{cov_u}\mathbf{m}$])

I

$$\alpha_1 \, [\mathbf{G}^T\mathbf{G}] \, \mathbf{G}^{-g} + \mathbf{G}^{-g}\Big[\alpha_2 \, [\mathbf{G}\mathbf{G}^T] + \alpha_3[\mathrm{cov_u}\mathbf{d}]\Big] = [\alpha_1 + \alpha_2]\mathbf{G}^T$$

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

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

damped minimum length

so

no new solutions have arisen …

… just a reinterpretation of previously-derived solutions

reinterpretation

instead of solving for estimates of the model parameters

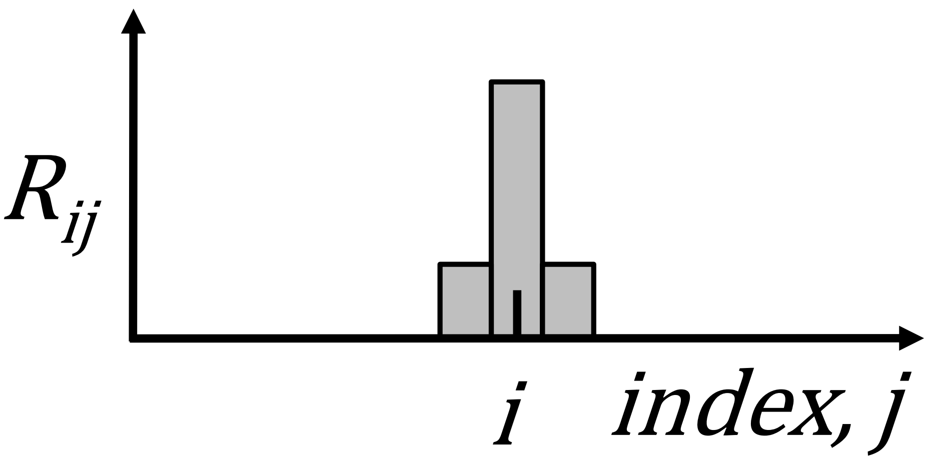
We are solving for estimates of **weighted averages** of the model parameters,

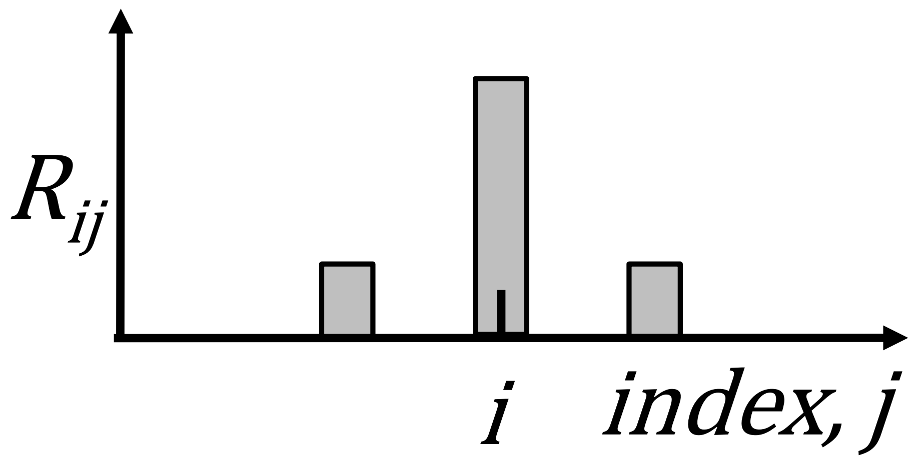
where the weights are given by the model resolution matrix criticism of Direchlet *spread()* functions

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 spread …





… but the left case is better “localized” we will take up this issue in the next lecture