Lecture 21

Continuous Problems

Fréchet Derivatives

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

use adjoint methods to compute

data kernels

Part 1

Review of Last Lecture

a function

$m(x)$

is the continuous analog of a vector

$\mathbf{m}$

a linear operator

$\mathcal{L}$

is the continuous analog of a matrix

$\mathbf{L}$

a inverse of a linear operator

$\mathcal{L}^{-1}$

is the continuous analog of the inverse of a matrix

$\mathbf{L}^{-1}$

a inverse of a linear operator
can be used to solve
a differential equation

if $\mathcal{L}m=f$ then $m=\mathcal{L}^{-1}f$

just as the inverse of a matrix
can be used to solve
a matrix equation

if $\mathbf{Lm}=\mathbf{f}$ then $\mathbf{m}=\mathbf{L}^{-1}\mathbf{f}$

the inner product

$$s = \int_{-\infty}^{+\infty} a(x)b(x)\mathrm{d}x = (a, b)$$

is the continuous analog of dot product

$$s = \mathbf{a}^{\mathrm{T}}\mathbf{b}$$

the adjoint of a linear operator

$$\mathcal{L}^{\dagger}$$

is the continuous analog of the transpose of a matrix

$$\mathbf{L}^{\mathrm{T}}$$

the adjoint can be used to manipulate an inner product

$$(\mathcal{L}a, b) = (a, \mathcal{L}^{\dagger} b)$$

just as the transpose can be used to manipulate the dot product

$$(\mathbf{La})^{T}\mathbf{b} = \mathbf{a}^{T}(\mathbf{L}^{T}\mathbf{b})$$

table of adjoints

$c(x)$	$c(x)$
$\mathrm{d}/\mathrm{d}x$	$-\mathrm{d}/\mathrm{d}x$
d^2/dx^2	d^2/dx^2
$\int_{-\infty}^{x} \mathrm{d}x$	$\int_{x}^{+\infty} \cdot \mathrm{d}x$

Part 2

definition of the Fréchet derivatives

first
rewrite the standard inverse theory equation in terms of perturbations

$$d_i = \int G_i(x)\, m(x)\, \mathrm{d}x = (G_i, m)$$

⇩

$$\delta d_i = \int G_i(x)\, \delta m(x)\, \mathrm{d}x = (G_i, \delta m)$$

a small change in the model
causes a small change in the data

second
compare with the standard formula for a derivative

$$\delta d_i = \int G_i(x)\, \delta m(x)\, \mathrm{d}x = (G_i, \delta m)$$

$$\Delta d_i = \sum_{i=1}^{M} G_{ij}^{(0)}\, \Delta m_j \quad \text{with} \quad G_{ij}^{(0)} = \left. \frac{\partial d_i}{\partial m_j} \right|_{\mathbf{m}^{(0)}}$$

third
identify the data kernel as a kind of derivative

$$\delta d_i = \int G_i(x)\, \delta m(x)\, \mathrm{d}x = (G_i, \delta m)$$

$$G_i(x) = \left.\frac{\delta d_i}{\delta m}\right|_{\mathbf{m}^{(0)}}$$

this kind of derivative is called a Fréchet derivative

definition of a Fréchet derivative

$$\delta d_i = \left(\left. \frac{\delta d_i}{\delta m} \right|_{\mathbf{m}^{(0)}}, \delta m \right)$$

this is mostly lingo
though perhaps it adds a little insight about
what the data kernel is

Part 2

Fréchet derivative of Error

treat the data as a continuous function $d(x)$
then the standard $\mathbf{L_2}$ norm error is

$$E = (d^{obs} - d, d^{obs} - d)$$

let the data $d(x)$ be related to the model $m(x)$ by

$$d = \mathcal{L}m$$

$\mathcal{L}m$ could be the data kernel integral

$$\mathcal{L}m = \int G_i(x)\, m(x)\, \mathrm{d}x$$

because integrals are linear operators

now do a little algebra to relate

δE to δm

a perturbation in the model
causes
a perturbation in the error

if
$m^{(0)}$ implies $d^{(0)}$ with error $E^{(0)}$
then ...

$$\delta E = E - E^{(0)} =$$

$$= (d^{obs} - d, d^{obs} - d) - \left(d^{obs} - d^{(0)}, d^{obs} - d^{(0)}\right) =$$

$$= -2(d, d^{obs}) + (d, d) + 2\left(d^{(0)}, d^{obs}\right) - \left(d^{(0)}, d^{(0)}\right) =$$

$$= -2\left(d^{obs} - d^{(0)}, d - d^{(0)}\right) + \left(d - d^{(0)}, d - d^{(0)}\right) =$$

$$= -2\left(d^{obs} - d^{(0)}, \delta d\right) + (\delta d, \delta d)$$

$$\approx -2\left(d^{obs} - d^{(0)}, \delta d\right)$$

$$= -2\left(d^{obs} - d^{(0)}, \mathcal{L}\delta m\right)$$

if
$m^{(0)}$ implies $d^{(0)}$ with error $E^{(0)}$
then ...

$$\delta E = E - E^{(0)} =$$

$$= (d^{obs} - d, d^{obs} - d) - (d^{obs} - d^{(0)}, d^{obs} - d^{(0)}) =$$

$$= -2(d, d^{obs}) + (d, d) + 2(d^{(0)}, d^{obs}) - (d^{(0)}, d^{(0)}) =$$

$$= -2(d^{obs} - d^{(0)}, d - d^{(0)}) + (d - d^{(0)}, d - d^{(0)}) =$$

$$= -2(d^{obs} - d^{(0)}, \delta d) + (\delta d, \delta d)$$

$$\approx -2(d^{obs} - d^{(0)}, \delta d)$$

all this is just algebra

if $m^{(0)}$ implies $d^{(0)}$ with error $E^{(0)}$ then ...

$$\delta E = E - E^{(0)} =$$

$$= (d^{obs} - d, d^{obs} - d) - \left(d^{obs} - d^{(0)}, d^{obs} - d^{(0)}\right) =$$

$$= -2(d, d^{obs}) + (d, d) + 2\left(d^{(0)}, d^{obs}\right) - \left(d^{(0)}, d^{(0)}\right) =$$

$$= -2\left(d^{obs} - d^{(0)}, d - d^{(0)}\right) + \left(d - d^{(0)}, d - d^{(0)}\right) =$$

$$= -2\left(d^{obs} - d^{(0)}, \delta d\right) + (\delta d, \delta d)$$

$$\approx -2\left(d^{obs} - d^{(0)}, \delta d\right)$$

$$= -2\left(d^{obs} - d^{(0)}, \mathcal{L}\delta m\right)$$

use $\delta d = \mathcal{L}\delta m$

if
$m^{(0)}$ implies $d^{(0)}$ with error $E^{(0)}$
then ...

$$\delta E = E - E^{(0)} =$$

$$= (d^{obs} - d, d^{obs} - d) - \left(d^{obs} - d^{(0)}, d^{obs} - d^{(0)}\right) =$$

$$= -2(d, d^{obs}) + (d, d) + 2\left(d^{(0)}, d^{obs}\right) - \left(d^{(0)}, d^{(0)}\right) =$$

$$= -2\left(d^{obs} - d^{(0)}, d - d^{(0)}\right) + \left(d - d^{(0)}, d - d^{(0)}\right) =$$

$$= -2\left(d^{obs} - d^{(0)}, \delta d\right) + (\delta d, \delta d)$$

$$\approx -2\left(d^{obs} - d^{(0)}, \delta d\right)$$

$$= -2\left(d^{obs} - d^{(0)}, \mathcal{L}\delta m\right)$$

$$= \left(-2\mathcal{L}^{\dagger}\left(d^{obs} - d^{(0)}\right), \delta m\right)$$

use adjoint

if

$m^{(0)}$ implies $d^{(0)}$ with error $E^{(0)}$

then ...

$$\delta E = E - E^{(0)} =$$

$$= (d^{obs} - d, d^{obs} - d) - \left(d^{obs} - d^{(0)}, d^{obs} - d^{(0)}\right) =$$

$$= -2(d, d^{obs}) + (d, d) + 2\left(d^{(0)}, d^{obs}\right) - \left(d^{(0)}, d^{(0)}\right) =$$

$$= -2\left(d^{obs} - d^{(0)}, d - d^{(0)}\right) + \left(d - d^{(0)}, d - d^{(0)}\right) =$$

$$= -2\left(d^{obs} - d^{(0)}, \delta d\right) + (\delta d, \delta d)$$

$$\approx -2\left(d^{obs} - d^{(0)}, \delta d\right)$$

$$= -2\left(d^{obs} - d^{(0)}, \mathcal{L}\delta m\right)$$

$$= \left(-2\mathcal{L}^{\dagger}\left(d^{obs} - d^{(0)}\right), \delta m\right)$$

$$\left.\frac{\delta E}{\delta m}\right|_{\mathbf{m}^{(0)}} = -2\mathcal{L}^{\dagger}\left(d^{obs} - d^{(0)}\right)$$

Fréchet derivative of Error

you can use this derivative to solve and inverse problem using the gradient method

example

$$\mathcal{L}\, m(x) = a \frac{d}{\mathrm{d}x} m(x) + b \int_{-\infty}^{x} m(x')\, \mathrm{d}x'$$

example

$$d(x) = \mathcal{L}\, m(x) = a \frac{d}{\mathrm{d}x} m(x) + b \int_{-\infty}^{x} m(x')\, \mathrm{d}x'$$

this is the relationship between model and data

example

$$\mathcal{L}\, m(x) = a \frac{d}{\mathrm{d}x} m(x) + b \int_{-\infty}^{x} m(x')\, \mathrm{d}x'$$

construct adjoint

$$\mathcal{L}^{\dagger}\, d(x) = -a \frac{d}{\mathrm{d}x} d(x) + b \int_{x}^{-\infty} d(x')\, \mathrm{d}x'$$

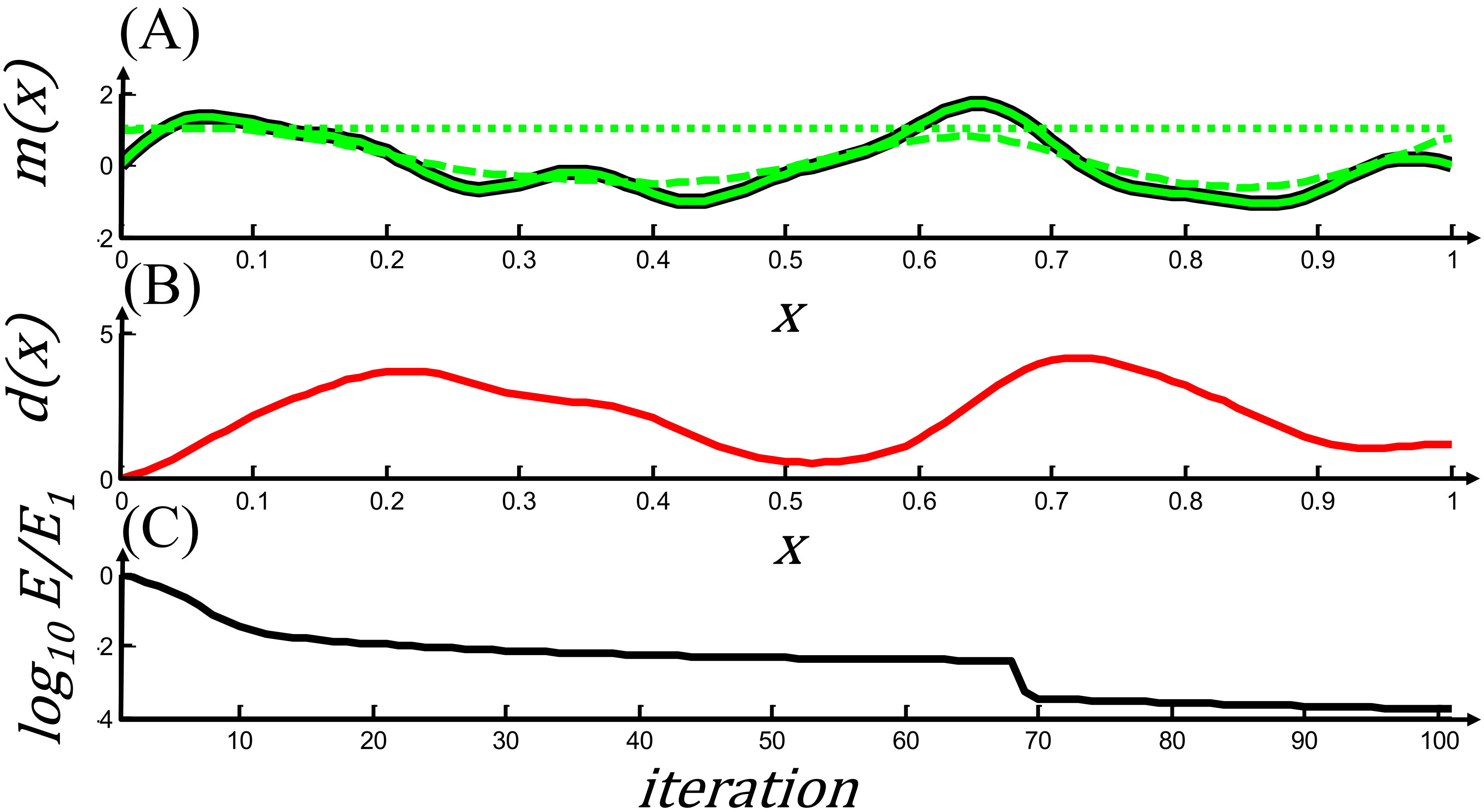
example

$$\mathcal{L}\, m(x) = a\frac{d}{\mathrm{d}x}m(x) + b\int_{-\infty}^{x} m(x')\,\mathrm{d}x'$$

$$\mathcal{L}^{\dagger}\, d(x) = -a\frac{d}{\mathrm{d}x}d(x) + b\int_{x}^{-\infty} d(x')\,\mathrm{d}x'$$

Fréchet derivative of Error

$$\left.\frac{\delta E}{\delta m}\right|_{\mathbf{m}^{(0)}} = 2a\frac{d}{\mathrm{d}x}\left[d^{obs}(x) - d^{(0)}(x)\right] - 2b\int_{x}^{-\infty}\left[d^{obs}(x') - d^{(0)}(x')\right]\mathrm{d}x'$$



(A)
m(x)
2
0
-2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
x
(B)
d(x)
5
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
x
(C)
log10 E/E1
0
-2
-4
10
20
30
40
50
60
70
80
90
100
iteration

Part 3

Backprojection

continuous analog of least squares

$$\left.\frac{\delta E}{\delta m}\right|_{\mathbf{m}^{(0)}} = 0 = -2\mathcal{L}^{\dagger}(d^{obs} - d) = -2\mathcal{L}^{\dagger}(d^{obs} - \mathcal{L}m)$$

or

$$\mathcal{L}^{\dagger}\mathcal{L}m = \mathcal{L}^{\dagger}d^{obs}$$

now define the identity operator $\mathcal{I}$

$$m(x) = \mathcal{I}\, m(x)$$

$$\mathcal{L}^{\dagger}\mathcal{L}m = \mathcal{L}^{\dagger}d^{obs}$$

$$(\mathcal{L}^{\dagger}\mathcal{L} + \mathcal{I} - \mathcal{I})m = \mathcal{L}^{\dagger}d^{obs}$$

$$m = \mathcal{I}m = \mathcal{L}^{\dagger}d^{obs} - (\mathcal{L}^{\dagger}\mathcal{L} - \mathcal{I})\, m$$

view as a recursion

$$m^{(i+1)} = \mathcal{L}^{\dagger} d^{obs} - \left(\mathcal{L}^{\dagger}\mathcal{L} - \mathcal{I}\right) m^{(i)}$$

$m^{(0)}=0$

$$m^{(1)} = \mathcal{L}^{\dagger} d^{obs}$$

view as a recursion

$$m^{(i+1)} = \mathcal{L}^{\dagger} d^{obs} - (\mathcal{L}^{\dagger}\mathcal{L} - \mathcal{I})\, m^{(i)}$$

$$m^{(0)}=0$$

$$m^{(1)} = \mathcal{L}^{\dagger} d^{obs}$$

using the adjoint as if it were the inverse

example

$$d^{obs}(x) = \mathcal{L}\, m(x) = \int_{-\infty}^{x} m(x')\, \mathrm{d}x'$$

exact

$$m(x) = \mathcal{L}^{-1}\, d^{obs} = \mathrm{d}\, d^{obs} / \mathrm{dx}$$

backprojection

$$m^{(1)}(x) = \mathcal{L}^{\dagger}\, d^{obs}(x) = \int_{x}^{\infty} d^{obs}(x')\, \mathrm{d}x'$$

example

$$d^{obs}(x) = \mathcal{L}\, m(x) = \int_{-\infty}^{x} m(x')\, \mathrm{d}x'$$

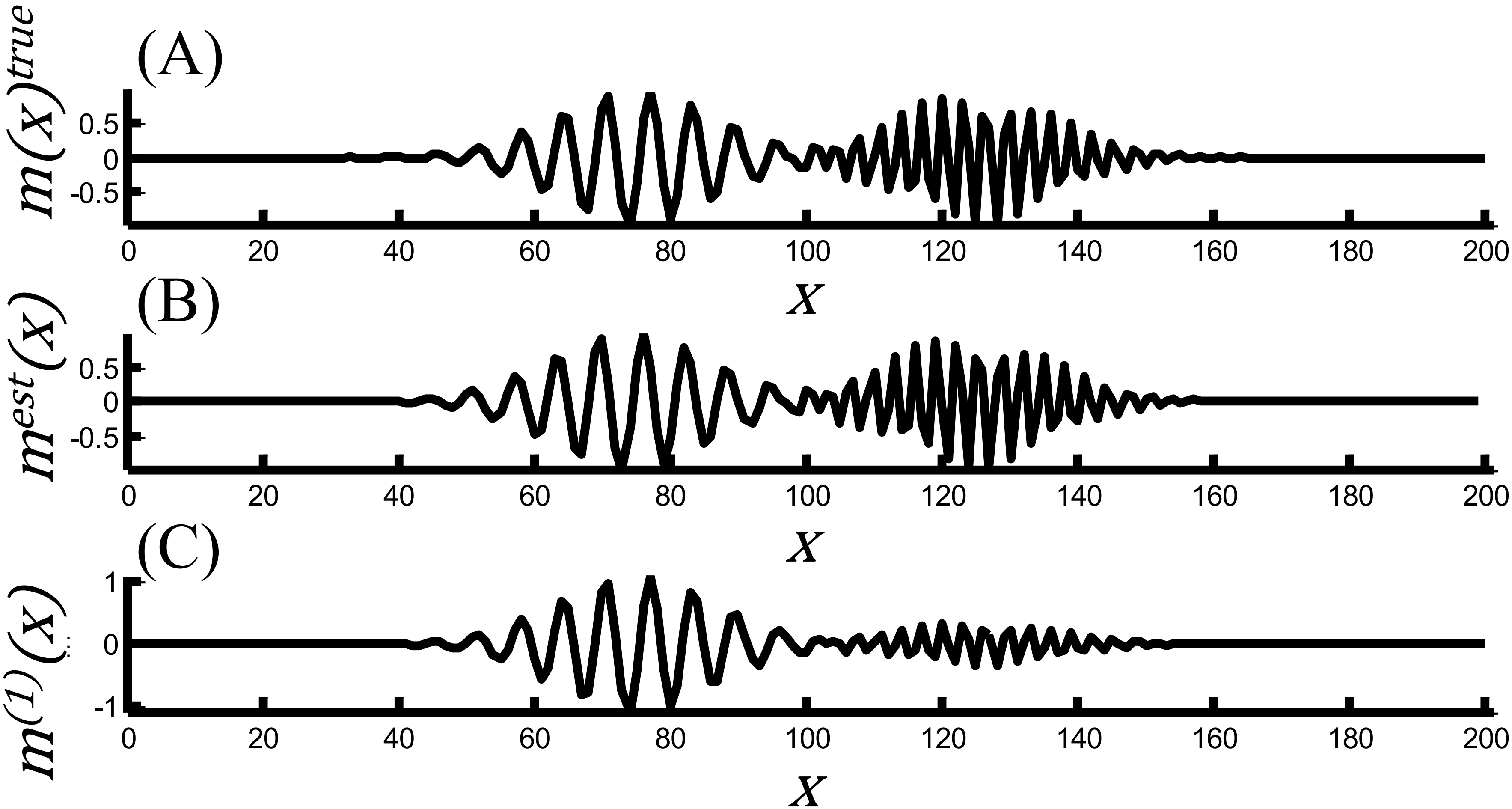
exact

$$m(x) = \mathcal{L}^{-1}\, d^{obs} = \mathrm{d}\; d^{obs} \,/\, \mathrm{dx}$$

backprojection

$$m^{(1)}(x) = \mathcal{L}^{\dagger}\, d^{obs}(x) = \int_{x}^{\infty} d^{obs}(x')\, \mathrm{d}x'$$

crazy!



(A)
$m(x)^{true}$
0.5
0
-0.5
0
20
40
60
80
100
120
140
160
180
200
x
(B)
$m^{est}(x)$
0.5
0
-0.5
0
20
40
60
80
100
120
140
160
180
200
x
(C)
$m^{(1)}(x)$
1
0
-1
0
20
40
60
80
100
120
140
160
180
200
x

interpretation as tomography

$$d^{obs}(x) = \int_{-\infty}^{x} m(x')\,\mathrm{d}x'$$

m is slowness
d is travel time of a ray from –∞ to x

backprojection

$$m^{(1)}(x) = \int_{x}^{\infty} d^{obs}(x')\,\mathrm{d}x'$$

integrate (=add together) the travel times of all rays that pass through the point x

discrete analysis

$$\mathbf{Gm=d}$$

$$\mathbf{G}= \mathbf{U\Lambda V}^{T} \qquad \mathbf{G}^{-g}= \mathbf{V\Lambda}^{-1}\mathbf{U}^{T} \qquad \mathbf{G}^{T}= \mathbf{V\Lambda U}^{T}$$

if $\mathbf{\Lambda}^{-1} \approx \mathbf{\Lambda}$ then $\mathbf{G}^{-g} \approx \mathbf{G}^{T}$

backprojection works when the singular values are all roughly the same size suggests scaling

$\mathbf{Gm}=\mathbf{d} \rightarrow \mathbf{WGm}=\mathbf{Wd}$

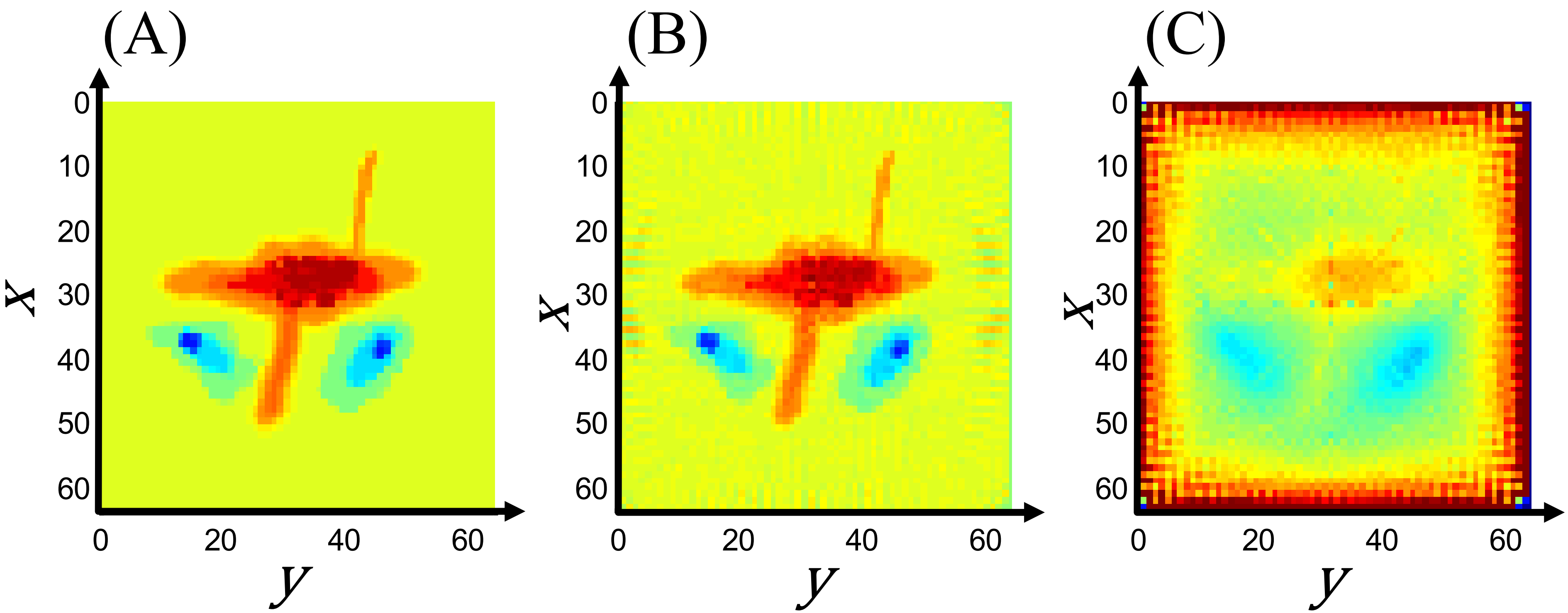
where **W** is a diagonal matrix chosen to make the singular values more equal in overall size

Traveltime tomography:

$$W_{ii} = \text{(length of ith ray)}^{-1}$$

so $[\mathbf{Wd}]_i$ has interpretation of the average slowness along the ray *i*.

Backprojection now adds together the average slowness of all rays that interact with the point *x*.



(A)
(B)
(C)
0
10
20
30
40
50
60
x
0
20
40
60
y

Part 4

Fréchet Derivative

involving a differential equation

Part 4

Fréchet Derivative

involving a differential equation

seismic wave equation
Navier-Stokes equation of fluid flow
etc

field u is related to model parameters m via a differential equation

$$\mathcal{L}\, u(x) = m(x)$$

data d is related to field u via an inner product

$$d_i = \big(h_i(x), u(x)\big)$$

write in terms of perturbations

perturbation δu is related to perturbation δm via a differential equation

$$\mathcal{L}\,\delta u(x) = \delta m(x)$$

pertrubation δd is related to perturbation δu via an inner product

$$\delta d_i = \big(h_i(x), \delta u(x)\big)$$

what’s the data kernel ?

$$\delta d_i = \int G_i(x)\, \delta m(x)\, \mathrm{d}x = (G_i, \delta m)$$

easy using adjoints

$\delta d_i = (h_i, \delta u)$ data inner product with field

easy using adjoints

$$\delta d_i = (h_i, \delta u)$$ data is inner product with field

$$= (h_i, \mathcal{L}^{-1} \delta m)$$ field satisfies $\mathcal{L}\delta u = \delta m$

easy using adjoints

$\delta d_i = (h_i, \delta u)$ data is inner product with field

$= (h_i, \mathcal{L}^{-1}\delta m)$ field satisfies $\mathcal{L}\delta u = \delta m$

$= ((\mathcal{L}^{-1})^{\dagger} h_i, \delta m)$ employ adjoint

easy using adjoints

$$\delta d_i = (h_i, \delta u)$$ data is inner product with field

$$= (h_i, \mathcal{L}^{-1}\delta m)$$ field satisfies $\mathcal{L}\delta u = \delta m$

$$= ((\mathcal{L}^{-1})^\dagger h_i, \delta m)$$ employ adjoint

$$= ((\mathcal{L}^\dagger)^{-1} h_i, \delta m)$$ inverse of adjoint is adjoint of inverse

easy using adjoints

$\delta d_i = (h_i, \delta u)$ data is inner product with field

$= (h_i, \mathcal{L}^{-1}\delta m)$ field satisfies $\mathcal{L}\delta u = \delta m$

$= ((\mathcal{L}^{-1})^{\dagger} h_i, \delta m)$ employ adjoint

$= ((\mathcal{L}^{\dagger})^{-1} h_i, \delta m)$ inverse of adjoint is adjoint of inverse

data kernel

$$G_i(x) = (\mathcal{L}^{\dagger})^{-1} h_i(x)$$

easy using adjoints

$\delta d_i = (h_i, \delta u)$ data is inner product with field

$= (h_i, \mathcal{L}^{-1}\delta m)$ field satisfies $\mathcal{L}\delta u = \delta m$

$= ((\mathcal{L}^{-1})^{\dagger} h_i, \delta m)$ employ adjoint

$= ((\mathcal{L}^{\dagger})^{-1} h_i, \delta m)$ inverse of adjoint is adjoint of inverse

data kernel

$$G_i(x) = (\mathcal{L}^{\dagger})^{-1} h_i(x)$$

data kernel satisfies "adjoint differential equation

$$\mathcal{L}^{\dagger} G_i(x) = h_i(x)$$

most problem involving differential equations are solved numerically

so instead of just solving

$$\mathcal{L}\,\delta u(x) = \delta m(x)$$

you must solve

$$\mathcal{L}\,\delta u(x) = \delta m(x) \quad \text{and} \quad \mathcal{L}^{\dagger} G_i(x) = h_i(x)$$

so there’s more work

but the same sort of work

example

time t instead of position x

field solves a Newtonian-type heat flow equation where u is temperature and m is heating

$$\mathcal{L}u(t) = \left\{\frac{d}{dt} + c\right\} u(t) = m(t)$$

data is concentration of chemical whose production rate is proportional to temperature

$$d_i = P(t_i) = b \int_0^{t_i} u(t)\, \mathrm{d}t$$

example

time t instead of position x

field solves a Newtonian-type heat flow equation where u is temperature

$$\mathcal{L}u(t) = \left\{\frac{d}{dt} + c\right\} u(t) = m(t)$$

data is concentration of chemical whose production rate is proportional to temperature

$$d_i = P(t_i) = b \int_0^{t_i} u(t)\,\mathrm{d}t \;=\; (bH(t_i - t), u)$$

so $h_i = bH(t_i - t)$

we will solve this problem
analytically
using Green functions

in more complicated cases
the differential equation
must be solved numerically

Newtonian equation

$$\mathcal{L}u(t) = \left\{\frac{d}{dt} + c\right\} u(t) = m(t)$$

its Green function

$$F(t,\tau) = H(t-\tau)\ \exp\{-c(t-\tau)\}$$

adjoint equation

$$\mathcal{L}^\dagger u(t) = \left\{-\frac{d}{dt} + c\right\} g_i(t) = h_i(t)$$

its Green function

$$Q(t,\tau) = H(\tau - t)\ \exp\{+c(t-\tau)\}$$

note that the adjoint Green function

$$Q(t,\tau) = H(\tau - t)\ \exp\{+c(t-\tau)\}$$

is the original Green function

$$F(t,\tau) = H(t - \tau)\ \exp\{-c(t-\tau)\}$$

backward in time

that's a fairly common pattern
whose significance will be pursued in a homework problem

we must perform a Green function integral to compute the data kernel

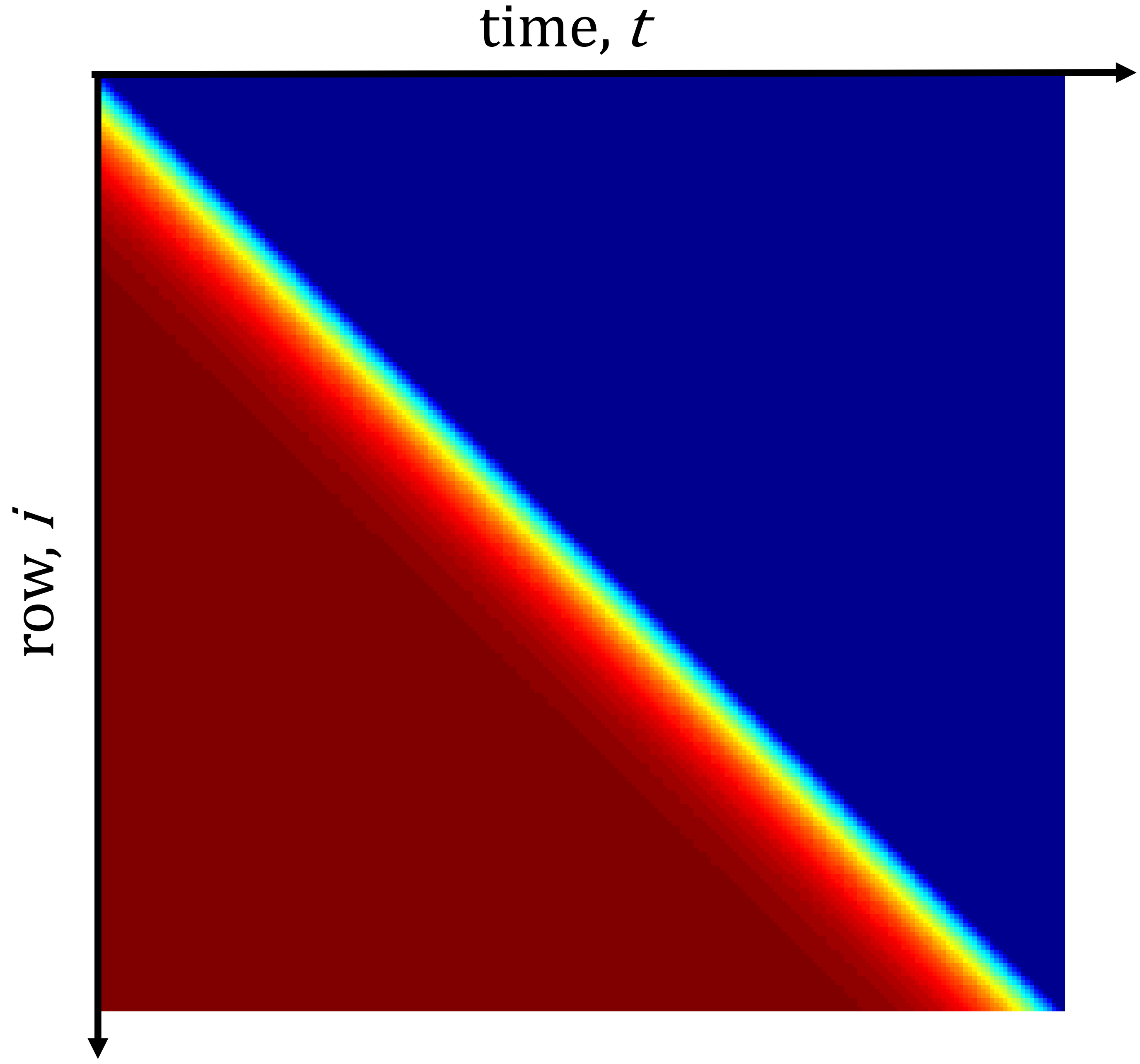
$$G_i(t) = \int_0^{\infty} Q(t,\tau)\, h_i(\tau)\, d\tau =$$

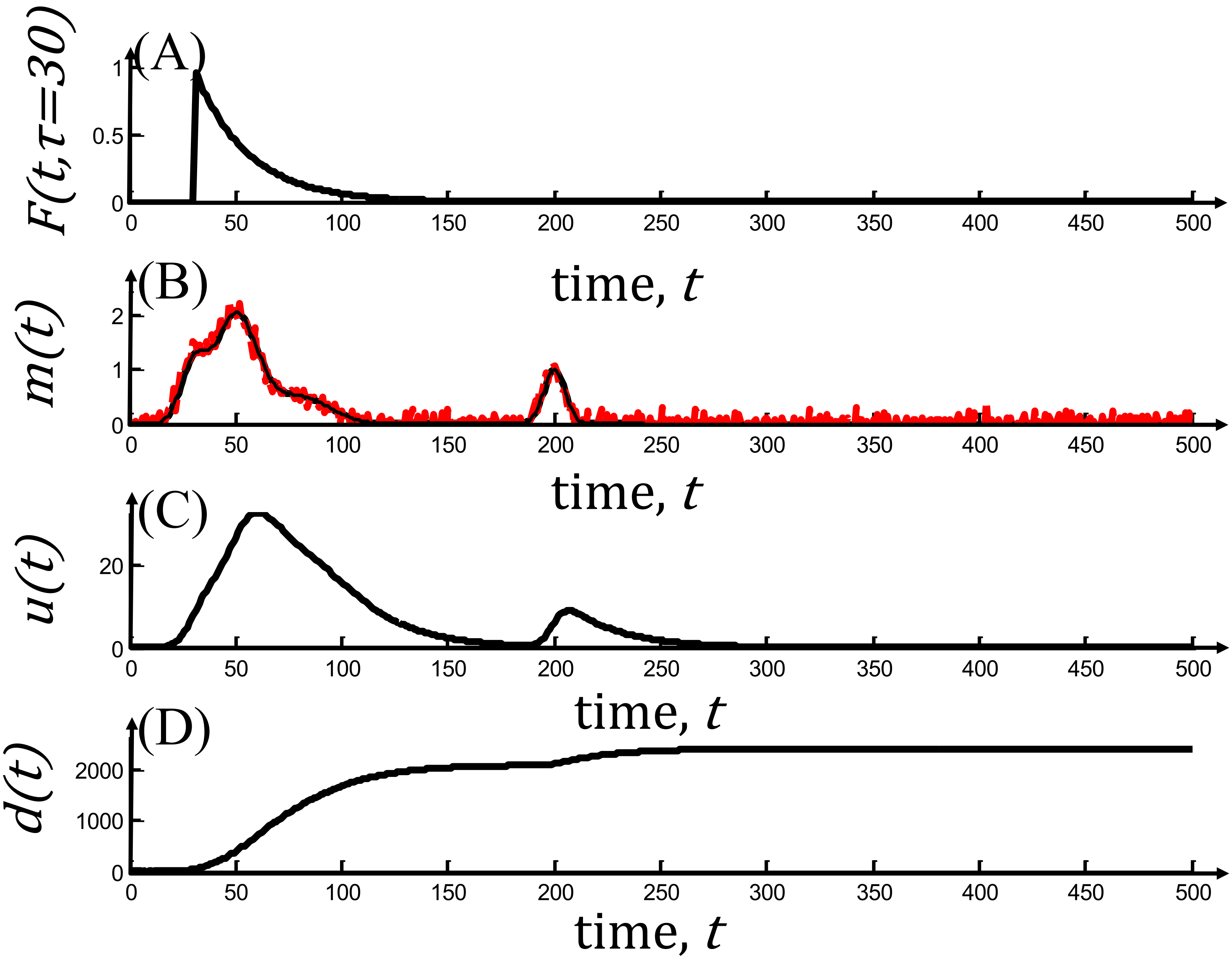
$$= \int_0^{\infty} H(\tau - t)\ \exp\{c(t-\tau)\}\ bH(t_i - \tau)\, d\tau =$$

$$= b\int_0^{\infty} H(\tau - t)\, H(t_i - \tau) \exp\{c(t-\tau)\}\ d\tau =$$

$$= b\int_t^{t_i} \exp\{-c(\tau - t)\}\ d\tau$$

$$
\begin{aligned}
G_i(t) &= b\int_{t}^{t_i} \exp\{-c(\tau - t)\}\, d\tau \\
&= \begin{cases} 0 & t_i \le t \\ -\dfrac{b}{c}\left[\exp\{-c(t_i - t)\} - 1\right] & t_i > t \end{cases}
\end{aligned}
$$


time, t
row, i


(A)
F(t,τ=30)
1
0.5
0
0
50
100
150
200
250
300
350
400
450
500
time, t
(B)
m(t)
2
1
0
0
50
100
150
200
250
300
350
400
450
500
time, t
(C)
u(t)
20
0
0
50
100
150
200
250
300
350
400
450
500
time, t
(D)
d(t)
2000
1000
0
0
50
100
150
200
250
300
350
400
450
500
time, t

Part 4

Fréchet Derivative

involving a parameter in differential equation

Part 4

Fréchet Derivative

involving a parameter in differential equation

previous example

$$\mathcal{L}u(t) = \left\{\frac{d}{dt} + c\right\}u(t) = m(t)$$

unknown
functi on is
"forcing"

another possibility

$$\left\{\frac{d}{dt} + c(t)\right\}u(t) = f(t)$$

parameter is
unknown

forcing
is known

linearize around a simpler equation

$$c(t) = c^{(0)} + \delta c(t)$$

$$u(t) = u^{(0)}(t) + \delta u(t),$$

and assume you can solve this equation

$$\left\{\frac{d}{dt} + c^{(0)}\right\} u^{(0)}(t) = f(t)$$

the perturbed equation is

$$\left\{\frac{d}{dt} + c^0 + \delta c(t)\right\}\left\{u^{(0)}(t) + \delta u(t)\right\} = f(t)$$

$$\left\{\frac{d}{dt} + c^0\right\}u^{(0)}(t) + \left\{\frac{d}{dt} + c^0\right\}\delta u(t) + \delta c(t)u^{(0)}(t) + \delta c(t)\delta u(t) = f(t)$$

subtracting out the unperturbed equation,
ignoring second order terms, and rearranging gives ...

then approximately

$$\left\{\frac{d}{dt} + c^0\right\} \delta u(t) = -\delta c(t)\, u^{(0)}(t)$$

pertubation to parameter acts as an unknown forcing

so it is back to the form of a forcing
and the previous methodology can be applied