Lecture 21

Continuous Problems

Fréchet Derivatives
# Syllabus

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

$L$

is the continuous analog of a matrix

$\mathbf{L}$
a inverse of a linear operator

$L^{-1}$

is the continuous analog of the inverse of a matrix

$L^{-1}$
a inverse of a linear operator can be used to solve a differential equation

\[ \text{if } \mathcal{L}m = f \text{ then } m = \mathcal{L}^{-1}f \]

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

\[ \text{if } Lm = f \text{ then } m = L^{-1}f \]
the inner product

\[ s = \int_{-\infty}^{+\infty} a(x) b(x) \, dx = (a, b) \]

is the continuous analog of dot product

\[ s = a^T b \]
the adjoint of a linear operator

$\mathcal{L}^\dagger$

is the continuous analog of the transpose of a matrix

$L^T$
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

$$(La)^T b = a^T (L^T b)$$
## Table of Adjoints

<table>
<thead>
<tr>
<th>Function</th>
<th>Adjoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c(x) )</td>
<td>( c(x) )</td>
</tr>
<tr>
<td>( \frac{d}{dx} )</td>
<td>( -\frac{d}{dx} )</td>
</tr>
<tr>
<td>( \frac{d^2}{dx^2} )</td>
<td>( \frac{d^2}{dx^2} )</td>
</tr>
<tr>
<td>( \int_{-\infty}^{x} dx )</td>
<td>( \int_{x}^{+\infty} dx )</td>
</tr>
</tbody>
</table>
Part 2

definition of the Fréchet derivatives
first rewrite the standard inverse theory equation in terms of perturbations

\[ d_i = \int G_i(x) \, m(x) \, dx = (G_i, m) \]

\[ \delta d_i = \int G_i(x) \, \delta m(x) \, dx = (G_i, \delta m) \]

a small change in the model causes a small change in the data
second compare with the standard formula for a derivative

\[ \delta d_i = \int G_i(x) \delta m(x) \, dx = (G_i, \delta m) \]

\[ \Delta d_i = \sum_{i=1}^{M} G_{ij}^{(0)} \Delta m_j \quad \text{with} \quad G_{ij}^{(0)} = \left. \frac{\partial d_i}{\partial m_j} \right|_{m^{(0)}} \]
third
identify the data kernel as a kind of derivative

\[ \delta d_i = \int G_i(x) \delta m(x) \, dx = (G_i, \delta m) \]

\[ G_i(x) = \frac{\delta d_i}{\delta m \bigg|_{m^{(0)}}} \]

this kind of derivative is called a Fréchet derivative
definition of a Fréchet derivative

\[
\delta d_i = \left( \frac{\delta d_i}{\delta m} \bigg|_{m(0)} , \delta m \right)
\]

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 $L_2$ 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

\[
\mathcal{L}m = \int G_i(x) \, m(x) \, dx
\]

because integrals are linear operators
now do a little algebra to relate

\[ \delta E \text{ to } \delta m \]

a perturbation in the model
causes
a perturbation in the error
if \( m^{(0)} \) implies \( d^{(0)} \) with error \( E^{(0)} \) then ...

\[
\delta E = E - E^{(0)} =
\]

\[
= (d^{\text{obs}} - d, d^{\text{obs}} - d) - (d^{\text{obs}} - d^{(0)}, d^{\text{obs}} - d^{(0)}) =
\]

\[
= -2(d, d^{\text{obs}}) + (d, d) + 2(d^{(0)}, d^{\text{obs}}) - (d^{(0)}, d^{(0)}) =
\]

\[
= -2(d^{\text{obs}} - d^{(0)}, d - d^{(0)}) + (d - d^{(0)}, d - d^{(0)}) =
\]

\[
= -2(d^{\text{obs}} - d^{(0)}, \delta d) + (\delta d, \delta d)
\approx -2(d^{\text{obs}} - d^{(0)}, \delta d)
= -2(d^{\text{obs}} - d^{\text{(0)}}, L \delta m)\]
if $m^{(0)}$ implies $d^{(0)}$ with error $E^{(0)}$
then ...

\[ \delta E = E - E^{(0)} = \]
\[ = (d_{obs} - d, d_{obs} - d) - (d_{obs} - d^{(0)}, d_{obs} - d^{(0)}) = \]
\[ = -2(d, d_{obs}) + (d, d) + 2(d^{(0)}, d_{obs}) - (d^{(0)}, d^{(0)}) = \]
\[ = -2(d_{obs} - d^{(0)}, d - d^{(0)}) + (d - d^{(0)}, d - d^{(0)}) = \]
\[ = -2(d_{obs} - d^{(0)}, \delta d) + (\delta d, \delta d) \]
\[ \approx -2(d_{obs} - d^{(0)}, \delta d) \]
if $m^{(0)}$ implies $d^{(0)}$ with error $E^{(0)}$ then ...

\[ \delta E = E - E^{(0)} = \]

\[ = (d^{obs} - d, d^{obs} - d) - (d^{obs} - d^{(0)}, d^{obs} - d^{(0)}) = \]

\[ = -2(d, d^{obs}) + (d, d) + 2(d^{(0)}, d^{obs}) - (d^{(0)}, d^{(0)}) = \]

\[ = -2(d^{obs} - d^{(0)}, d - d^{(0)}) + (d - d^{(0)}, d - d^{(0)}) = \]

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

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

\[ = -2(d^{obs} - d^{(0)}, \mathcal{L}\delta m) \]

\[ \text{use } \delta d = \mathcal{L}\delta m \]
if $m^{(0)}$ implies $d^{(0)}$ with error $E^{(0)}$ then ...

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Fréchet derivative of Error
you can use this derivative to solve and inverse problem using the gradient method
$\mathcal{L} \ m(x) = a \frac{d}{dx} m(x) + b \int_{-\infty}^{x} m(x') \, dx'$
example

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

this is the relationship between model and data
Example

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

\[ \mathcal{L}^\dagger d(x) = -a \frac{d}{dx} d(x) + b \int_{x}^{-\infty} d(x') \, dx' \]
example

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

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

Fréchet derivative of Error

\[ \left. \frac{\delta E}{\delta m} \right|_{m^{(0)}} = 2a \frac{d}{dx} \left[ d^{obs}(x) - d^{(0)}(x) \right] - 2b \int_{x}^{-\infty} \left[ d^{obs}(x') - d^{(0)}(x') \right] \, dx' \]
Part 3

Backprojection
continuous analog of least squares

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

or

$$\mathcal{L}^\dagger\mathcal{L}m = \mathcal{L}^\dagger d^{obs}$$
now define the identity operator $I$

$$m(x) = I \, m(x)$$

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

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

$$m = Im = \mathcal{L}^\dagger d^{obs} - (\mathcal{L}^\dagger \mathcal{L} - I) \, m$$
view as a recursion

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

\[ m^{(0)} = 0 \]

\[ m^{(1)} = \mathcal{L}^\dagger d^{obs} \]
view as a recursion

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

\[ m^{(0)} = 0 \]

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

using the adjoint as if it were the inverse
example

\[ d^{obs}(x) = \mathcal{L} m(x) = \int_{-\infty}^{x} m(x') \, dx' \]

exact

\[ m(x) = \mathcal{L}^{-1} d^{obs} = d \frac{d^{obs}}{dx} \]

backprojection

\[ m^{(1)}(x) = \mathcal{L}^\dagger d^{obs}(x) = \int_{x}^{\infty} d^{obs}(x') \, dx' \]
example

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

exact

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

backprojection

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

crazy!
interpretation as tomography

\[ d^{obs}(x) = \int_{-\infty}^{x} m(x') \, dx' \]

\( m \) is slowness
\( d \) is travel time of a ray from \(-\infty\) to \( x \)

backprojection

\[ m^{(1)}(x) = \int_{x}^{\infty} d^{obs}(x') \, dx' \]

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

\[ Gm = d \]

\[ G = U\Lambda V^T \quad G^{-g} = V\Lambda^{-1}U^T \quad G^T = V\Lambda U^T \]

if \( \Lambda^{-1} \approx \Lambda \) then \( G^{-g} \approx G^T \)

backprojection works when the singular values are all roughly the same size
suggests scaling
\[ Gm = d \rightarrow WGm = Wd \]
where \( W \) is a diagonal matrix chosen to make the singular values more equal in overall size.

Traveltime tomography:
\[ W_{ii} = (\text{length of } \text{ith ray})^{-1} \]

so \([Wd]_i\) 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 \( x \).
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

\[
d_i = (h_i(x), u(x))
\]
write in terms of perturbations

perturbation $\delta u$ is related to perturbation $\delta m$ via a differential equation

$$L \, \delta u(x) = \delta m(x)$$

perturubation $\delta d$ is related to perturbation $\delta u$ via an inner product

$$\delta d_i = (h_i(x), \delta u(x))$$
what’s the data kernel?

\[ \delta d_i = \int G_i(x) \delta m(x) \, dx = (G_i, \delta m) \]
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

data is inner product with field

$$
\delta d_i = (h_i, \delta u) \\
= (h_i, L^{-1} \delta m) \\
= ((L^{-1})^\dagger h_i, \delta m)
$$

field satisfies $$L \delta u = \delta m$$

employ adjoint
easy using adjoints

\[ \delta d_i = (h_i, \delta u) \quad \text{data is inner product with field} \]
\[ = (h_i, \mathcal{L}^{-1} \delta m) \quad \text{field satisfies } \mathcal{L} \delta u = \delta m \]
\[ = ((\mathcal{L}^{-1})^\dagger h_i, \delta m) \quad \text{employ adjoint} \]
\[ = ((\mathcal{L}^\dagger)^{-1} h_i, \delta m) \quad \text{inverse of adjoint is adjoint of inverse} \]
easy using adjoints

\[ \delta d_i = (h_i, \delta u) \quad \text{data is inner product with field} \]

\[ = (h_i, \mathcal{L}^{-1} \delta m) \quad \text{field satisfies } \mathcal{L} \delta u = \delta m \]

\[ = ((\mathcal{L}^{-1})^\dagger h_i, \delta m) \quad \text{employ adjoint} \]

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

\[ G_i(x) = (\mathcal{L}^\dagger)^{-1} h_i(x) \quad \text{data kernel} \]
easy using adjoints

\[ \delta d_i = (h_i, \delta u) \quad \text{data is inner product with field} \]

\[ = (h_i, \mathcal{L}^{-1} \delta m) \quad \text{field satisfies } \mathcal{L} \delta u = \delta m \]

\[ = ((\mathcal{L}^{-1})^\dagger h_i, \delta m) \quad \text{employ adjoint} \]

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

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

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

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

\[ d_i = P(t_i) = b \int_0^{t_i} u(t) \, dt \]
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

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

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

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

\[ G_i(t) = \int_0^\infty Q(t, \tau) h_i(\tau) \, d\tau = \]

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

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

\[ = b \int_t^{t_i} \exp\{-c(\tau - t)\} \, d\tau \]
\[ G_i(t) = b \int_t^{t_i} \exp\{-c(\tau - t)\} \ d\tau \]

\[ = \begin{cases} 0 & t_i \leq t \\ -\frac{b}{c} [\exp\{-c(t_i - t)\} - 1] & t_i > t \end{cases} \]
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 function 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

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

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

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

perturbation to parameter acts as an unknown forcing

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