Minimum Norm Construction   I 2.0

- Smallest, flattest models
- 2 data gravity problem
- Including a priori information through norms
- A generic flexible norm

Minimum Norm Construction : Accurate Data

Consider the geophysical experiment that yields data

$$d_j = (g_j, m) \qquad j = 1, N$$

Assume that the data are accurate.

Remarks: We recognize that there are infinitely many solutions. Our goal is to find one.

Some form of prior information must be supplied

Our procedure is to incorporate this prior information into a model norm and then find the solution that has minimum norm.

We'll illustrate several types of prior information and the resulting models.

(i) Smallest (l_2-norm) solution (zero reference model)

(ii) Smallest deviatoric (the roll of the reference model)

(iii) Flattest models

(iv) Weighted models.

Remark: All of the above components eventually will get combined into a generic model objective function that we will use

: We'll work on a simple example. 2-data gravity problem.

: For the present we'll keep everything as analytic as possible. Working with functions and analytic data allows us to explore the fundamental problem of instability that arises

Minimum Norm Construction : Inverse Problem # 1

Consider a geophysical experiment in which the data are given by

$$d_j = (g_j, m) \qquad j = 1, N \qquad (1)$$

These equations form a system of N constraints upon an unknown function m. We recognize that the solution is nonunique and therefore we construct a specific model that is of potential interest to us. Introducing a norm

$$\|m\| = (m, m)^{1/2} \qquad (2)$$

we formulate the problem as finding an m that minimizes (2) subject to the constraints in (1).

The solution can be achieved in two ways:

(i) using calculus of variations

(ii) Projection theorem.

Calculus of Variations

To minimize a function subject to constraints we make a combined functional

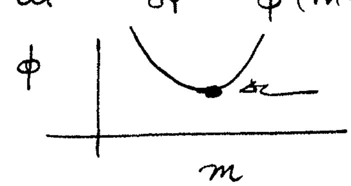
$$\phi(m) = \|m\|^2 + 2 \sum_j \alpha_j \left[d_j - (g_j, m) \right]$$

Comments: (1) Since $\|m\|$ is positive it doesn't matter whether we minimize $\|m\|$ or its square.

(2) The α_j's are Lagrange multipliers. (Factor 2 is arbitrary)

$$\phi(m) = (m, m) + 2 \sum_j \alpha_j \left[d_j - (g_j, m) \right]$$

In a variational approach we introduce an arbitrary perturbation δm and observe the change in the functional ϕ. Then we look at $\delta\phi = \phi(m + \delta m) - \phi(m)$ and set it equal to zero.



$\delta\phi$ is zero here (at an extremum or saddle point)

$$\phi(m+\delta m) = (m+\delta m,\ m+\delta m) + 2\sum_j \alpha_j\left[d_j - (g_j, m+\delta m)\right]$$

$$= (m,m) + 2(m,\delta m) + \underbrace{(\cancel{\delta m}, \delta m)}_{\text{neglect}} + 2\sum_j \alpha_j\left[d_j - (g_j, m) - \underline{(g_j, \delta m)}\right]$$

$$\delta\phi = 2(m,\delta m) - 2\sum_j \alpha_j (g_j, \delta m)$$

$$0 = \left(m - \sum_j \alpha_j g_j,\ \delta m\right) \qquad (3)$$

Now (3) must hold for an arbitrary perturbation δm. The only way that this can be true is if the first term is zero.

ie $$\boxed{m = \sum_{j=1}^{N} \alpha_j g_j} \qquad (4)$$

ie the model which minimizes the norm must be a linear combination of the initial vectors g_j. ie The minimum norm model lies in the activated portion of model space.

The minimum norm model m_0 which fit the data is obtained by computing the coefficients α_j. We require that m_0 fit the data constraints

$$d_j = (g_j, m_0) \qquad j = 1, N$$

Substituting (4)

$$d_j = \left(g_j, \sum_{K=1}^{N} \alpha_K g_K\right)$$

$$= \sum_{K=1}^{N} \alpha_K (g_j, g_K)$$

$$d_j = \sum_K \alpha_K \Gamma_{jK}$$

or $$\boxed{\Gamma \bar{\alpha} = \bar{d}}$$

where $\Gamma_{jK} = (g_j, g_K)$ are elements of the <u>inner product</u> matrix Γ.

$\vec{d} = (d_1, d_2, \cdots d_N)^T$

$\vec{\alpha} = (\alpha_1, \alpha_2, \cdots \alpha_N)^T$

The matrix Γ is positive definite and symmetric and therefore invertable. (details later)

$$\Gamma \bar{\alpha} = \bar{d} \qquad\qquad \Gamma : N \times N \text{ matrix}$$

$$\Rightarrow \quad \boxed{\vec{\alpha} = \Gamma^{-1} \bar{d}}$$

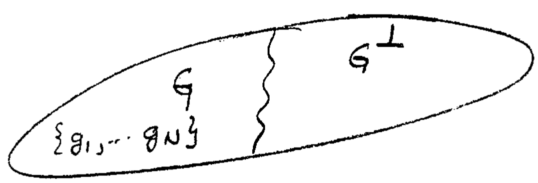
This solves for the coefficients. The minimum norm model is then obtained by

$$\boxed{m_0 = \sum \alpha_j g_j}$$

Actually we have shown that $m = \sum \alpha_j g_j$ is a stationary point. Technically we need to look at other derivatives to ensure that it is not a maximum or saddle point

Method II : Using the Decomposition Theorem.

Our model space can be divided into an activated and unactivated portion



$$m = m^{\parallel} + m^{\perp} \qquad\qquad \begin{array}{l} m^{\parallel} \in G = \operatorname{sp}\{g_1, \cdots g_N\} \\ m^{\perp} \in G^{\perp} \end{array}$$

Every element in $G^\perp$ is perpendicular to every element in G. In particular

$$(g_j, m^\perp) = 0 \qquad j = 1, N$$

Consider the data equations

$$\begin{aligned} d_j &= (g_j, m) \\ &= (g_j, m^{\parallel} + m^{\perp}) = (g_j, m^{\parallel}) + \underbrace{(g_j, m^{\perp})}_{=0} \end{aligned}$$

so $\boxed{d_j = (g_j, m^{\parallel})}$ (something that we knew before. The data are affected only by $m^{\parallel}$.)

Now consider a minimum norm solution

$$
\begin{aligned}
\|m\|^2 &= \|m^{\parallel} + m^{\perp}\|^2 \\
&= (m^{\parallel} + m^{\perp}, m^{\parallel} + m^{\perp}) \\
&= (m^{\parallel}, m^{\parallel}) + 2(m^{\parallel}, m^{\perp}) + (m^{\perp}, m^{\perp}) \\
&= \|m^{\parallel}\|^2 + \|m^{\perp}\|^2
\end{aligned}
$$

Since $m^{\perp}$ has no affect on the data we can include as much or as little of it as desired. From the point of view of minimizing $\|m\|^2$ we choose $m^{\perp} = 0$.

If $m^{\perp} = 0$ then the norm minimizing solution is $m_0 = m^{\parallel}$. Thus

$$\boxed{m_0 = \sum_{j=1}^{N} \alpha_j g_j}$$

Substituting into the data equations yields, as before, a system of equations to be solved:

$$\boxed{\Gamma \bar{\alpha} = \bar{d}} \qquad \Gamma_{ij} = (g_i, g_j)$$

Remark: We now have the algebra for our first inverse problem. Assuming that Γ^{-1} exists, and that we are not worried about computational efficiency, we can compute a minimum norm model that fits the data.

Remark: The algebra presented is completely general and is therefore valid for any Hilbert space. We consider two simple examples for illustration

(i) 2-data gravity problem

(ii) solution to an underdetermined system of equations.

2-Data Gravity Problem

One of the simplest and yet informative problems in linear inverse theory is the two data gravity problem. The goal is to determine the density structure of a spherically symmetric body by measuring its mass and moment of inertia. The appropriate equations are

$$\boxed{\begin{aligned} \frac{\bar{\rho}}{3} &= \int_0^1 r^2 \rho(r)\, dr \\ \frac{\bar{\rho}\gamma}{2} &= \int_0^1 r^4 \rho(r)\, dr \end{aligned}} \qquad (1)$$

where $\bar{\rho} = \dfrac{M_e}{\frac{4}{3}\pi a^3} = 5.5\ Mg/m^3$ 'a' is radius of earth

$\gamma = \dfrac{C}{Ma^2} = .33078$ (C is moment of inertia about the spin axis)

In (1), the radius of the earth has been normalized to unity. So $r=0$ corresponds to the center of the earth, $r=1$ corresponds to the surface.

Remark: Remember, for a sphere if $\rho(r)$ is constant $\Rightarrow \gamma = 0.4$. For the earth $\gamma = .33078 < 0.4$ $\therefore$ The earth must be more dense toward the center.

Let
$$\boxed{\begin{aligned} d_1 &= \frac{\bar{\rho}}{3} = 1.833\ Mg/m^3 \\ d_2 &= \frac{\bar{\rho}\gamma}{2} = .9095\ Mg/m^3 \end{aligned}}$$

Then the two equations in (1) can be written as

$$\boxed{\begin{aligned} &d_j = (g_j, m) \qquad j = 1,2 \\ &\text{where} \quad g_1 = r^2, \quad g_2 = r^4 \end{aligned}}$$

Smallest Model Solution

The minimum norm model, when $\|\rho\| = (\rho,\rho)^{1/2}$ is given by

$$\rho(r) = \sum \alpha_j g_j = \alpha_1 r^2 + \alpha_2 r^4$$

where α's are found by solving $\Gamma \bar{\alpha} = \bar{d}$
where $\Gamma_{ij} = (g_i, g_j)$. Our inner product is

$$(g_i, g_j) = \int_0^1 g_i(r)\, g_j(r)\, dr$$

$$(g_1, g_1) = \int_0^1 r^4 dr = 1/5 \qquad (g_1, g_2) = (g_2, g_1) = \frac{1}{7} \qquad (g_2, g_2) = \frac{1}{9}$$

So

$$\Gamma = \begin{pmatrix} 1/5 & 1/7 \\ 1/7 & 1/9 \end{pmatrix} \qquad \Gamma^{-1} = \frac{2205}{4}\begin{pmatrix} 1/9 & -1/7 \\ -1/7 & 1/5 \end{pmatrix}$$

$$\bar{\alpha} = \Gamma^{-1} d = \frac{2205}{4}\begin{pmatrix} 1/9 & -1/7 \\ -1/7 & 1/5 \end{pmatrix}\begin{pmatrix} 1.833 \\ .9095 \end{pmatrix} = \begin{pmatrix} 40.658 \\ -44.086 \end{pmatrix}$$

So

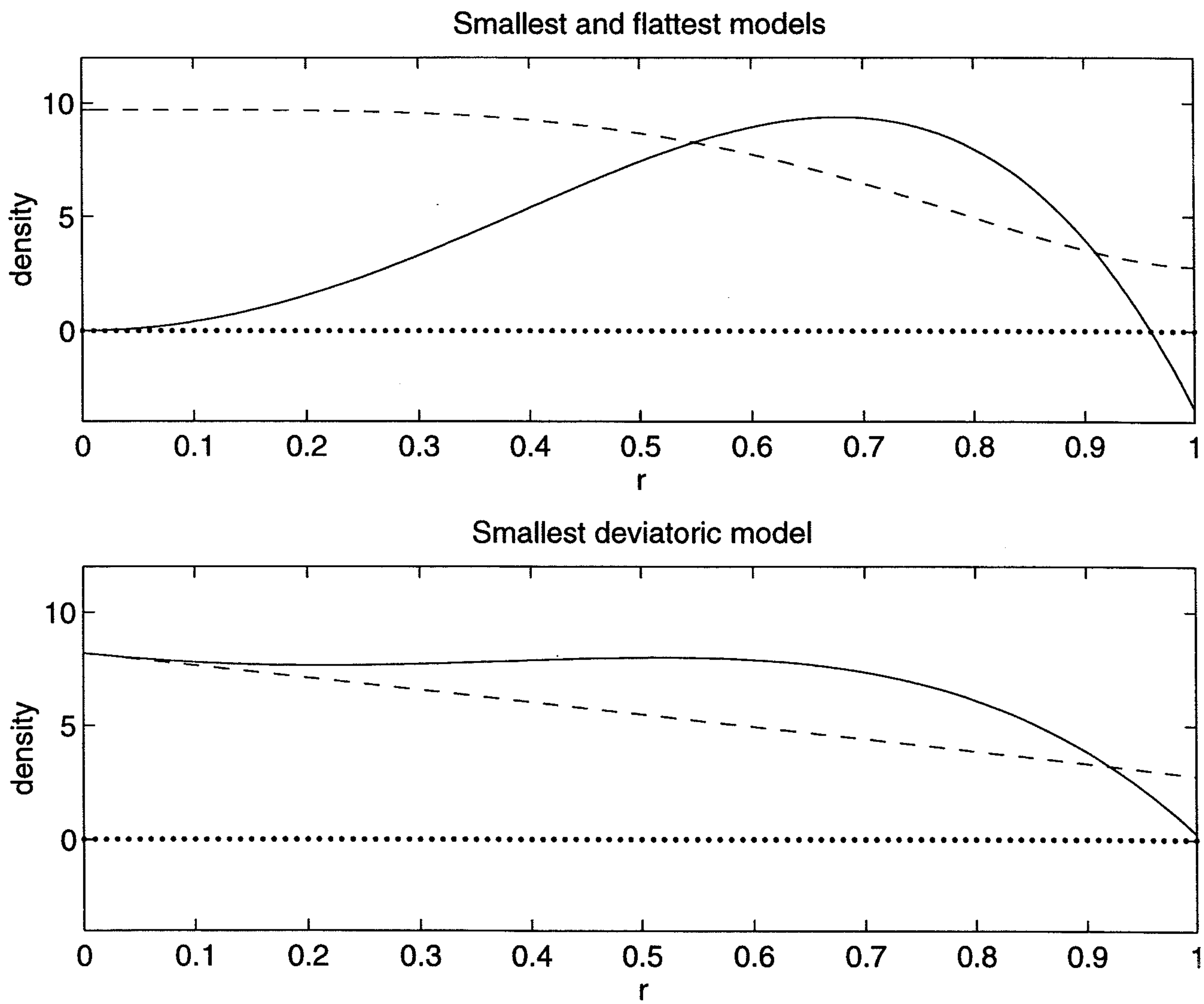
$$\boxed{\rho(r) = 40.658\, r^2 - 44.086\, r^4 \quad \frac{Mg}{m^3}}$$

Remark: This density structure is plotted on the following graph. Places of interest are:

$r=0$ $\quad \rho = 0$ (center of the earth has zero density)

$r=1$ $\quad \rho = -3.428\ Mg/m^3$ (surface of the earth has negative density)

✱: We have constructed a mathematical solution to the problem but not a geophysical solution. For this example, the smallest model was not the best norm to be minimized.



Smallest and flattest models
density
10
5
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
r
Smallest deviatoric model
density
10
5
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
r

Smallest deviatoric model.

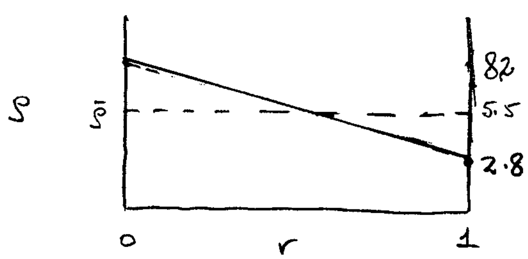
Remark: The density structure produced by minimizing $\|\rho\|^2$ was unacceptable on physical grounds. Does this mean that a smallest model formulation is not appropriate for the 2-data gravity problem?

Remark: In minimizing $\int_0^1 \rho(r)^2 dr$ we have attempted to find a model that is close to zero everywhere. But really, this is not a result that is consistent even with elementary geophysics. Lets exame what we know.

(1) averages density for the earth is $\bar{\rho} = 5.5\ Mg/m^3$

(2) surface rocks (if we think of the "surface" as being a planetary surface and think of the average properties of the top 100 km) the a reasonable estimate for the density at the surface is $\rho_s \sim 2.8\ Mg/m^3$. If the average value is $5.5\ Mg/m^3$ then we would estimate that the density at the center of the earth is greater than $5.5\ Mg/m^3$. This of course is entirely in accordance with our ideas of compaction and compression of rocks as we descend into the earth.

So, a guess for the earth density



Choose $\boxed{\rho_0(r) = 8.2 - 5.4\,r \quad \frac{Mg}{m^3}}$

So $\rho_0(r)$ is a "guess" at what the density structure might be even if its isn't a particularly brilliant guess.

Remark: If $\rho_0(r)$ is our current best estimate for the density then we would like our density structure which fits the two data to be close to $\rho_0(r)$. That is we want to minimize

$$\phi = \|\rho - \rho_0\| \quad \text{subject to data constraints.}$$

but $\rho(r) = \rho_0(r) + \delta\rho(r)$. Minimizing $\|\rho - \rho_0\|$ is the same as minimizing $\|\delta\rho\|$. Write the data equations in terms of $\delta\rho$

$$d_j = (g_j, \rho) = (g_j, \rho_0 + \delta\rho)$$
$$= (g_j, \rho_0) + (g_j, \delta\rho)$$

but ρ_0 is known so (g_j, ρ_0) can be evaluated and taken to the other side.

$$d_j - (g_j, \rho_0) = (g_j, \delta\rho)$$

Introducing new data we have

$$\boxed{\begin{aligned} f_j &= (g_j, \delta\rho) \\ f_j &= d_j - (g_j, \rho_0) \end{aligned}}$$

But this is same type of equation that we had before. The only change is that

(i) new data

(ii) "model" is now $\delta\rho$.

The minimum norm solution, minimizing $\|\delta\rho\|$ is completely appropriate.

I: Compute the new data.

$$(g_1, \rho_0) = \int_0^1 r^2(8.2 - 5.4r)\,dr = \frac{8.2}{3} - \frac{5.4}{4} = 1.383$$

$$(g_2, \rho_0) = \int_0^1 r^4(8.2 - 5.4r)\,dr = \frac{8.2}{5} - \frac{5.4}{6} = .740$$

so
$$f_1 = d_1 - (g_1, \rho_0) = 1.833 - 1.383 = .45$$
$$f_2 = d_2 - (g_2, \rho_0) = .9095 - .740 = .1695$$

$\boxed{\vec{f} = (.45,\ .1695)^T}$ are new data.

The smallest norm solution $\quad \delta\rho = \alpha_1 r^2 + \alpha_2 r^4$

$$\vec{\alpha} = \Gamma^{-1}\vec{f} = \frac{2205}{4}\begin{pmatrix} \frac{1}{9} & -\frac{1}{7} \\ -\frac{1}{7} & \frac{1}{5} \end{pmatrix}\begin{pmatrix} .533 \\ .203 \end{pmatrix} = \begin{pmatrix} 16.66 \\ -19.59 \end{pmatrix}$$

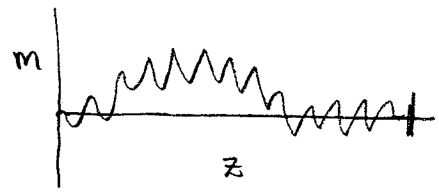
So $\delta r = 16.66\, r^2 - 19.59\, r^4$

and
$$\boxed{\rho(r) = 8.2 - 5.4r + 16.66\, r^2 - 19.59^4 \ \frac{Mg}{m^3}}$$

How does this look? $\left.\begin{array}{l} \rho(0) = 8.2 \ Mg/m^3 \\ \rho(1) = 0.262 \ Mg/m^3 \end{array}\right\}$ See handout.

So we still don't have a good geophysical answer but we're getting close. This is certainly a better result than we had previously. If we had made an estimate with $\rho_0(0)$ to be larger, we would have done better with repect to the surface density.

Remark: The above cases are a bit pessimistic to be generalized. Usually smallest, or smallest deviatoric, models can be constructed which look reasonably geophysical. A characteristic of these models however is that they tend to have fluctuations.



← typical result of trying to construct a model which is as close to zero everywhere (subject to data constraints

This may be undesireable. Is there a way around this. Can a solution be formulated which doesn't suffere from these oscillations.

Remark: Consider integral equations of the form

$$d_j = \int_a^b g_j(x)\, m(x)\, dx \quad = (g_j, m) \qquad j=1,N$$

Integrate these equations by parts. Let

$$h_j(x) = \int_a^x g_j(u)\, du \qquad \boxed{(h_j(x) = g^{(-1)}(x) \text{ in Parker's notat}}$$

be the indefinite integrals of the kernel functions. We obtain

$$d_j = h_j(x)\, m(x) \Big|_a^b - \int_a^b h_j(x)\, m'(x)\, dx$$

$$= h_j(b)\, m(b) - \underbrace{h_j(a)}_{=0} m(a) - \int_a^b h_j(x)\, m'(x)\, dx$$

Suppose that $m(b)$ is known. Then

$$\boxed{h_j(b)\, m(b) - d_j = \int_a^b h_j(x)\, m'(x)\, dx} \qquad (1)$$

So we have a new system of equations

$$\boxed{f_j = (h_j, m') \qquad j=1,N} \qquad (2)$$

where $f_j = h_j(b)\, m(b) - d_j$ are new data
$h_j(x)$: are new kernels
$m'(x)$: new "model".

But these equations are precisely the same as those we solved previously for a minimum norm model. (minimize $\phi = (m', m$

The solution is therefore

$$\boxed{m'(x) = \sum_{j=1}^{N} \beta_j \, h_j(x)} \qquad (3)$$

where

$$\boxed{\Gamma \vec{\beta} = \vec{f}} \qquad \text{and} \qquad \Gamma_{ij} = (h_i, h_j)$$

Once $m'(x)$ is found, we can integrate to obtain $m(x)$.

$$\boxed{m(x) = \int^{x} m'(u)\, du + c} \qquad (4)$$

The constant is evaluated by requiring that the known boundary condition at $x=b$ is satisfied. ($m(b)$ was presumed to be known in order evaluate the new data f_j)

* (Note $m(x) = \int^{x} \sum_{j=1}^{N} \beta_j h_j(u)\, du + c = \sum_{j=1}^{N} \beta_j g^{(-2)}(x) + c$) : So the model is made of doubly smoothed ker

Remark: The above procedure required that a value of the model be supplied at the right hand end point. Suppose however that $m(a)$ was known rather than $m(b)$. Is there a way to alter the formulation?

Fundamental Theorem of Calculus.

$$m(b) - m(a) = \int_a^b m'(x)\, dx$$

So we could alter (1) by substituting for $m(b)$ to get

$$h_j(b)\, m(b) - d_j = \int_a^b h_j(x)\, m'(x)\, dx$$

$$h_j(b) \left\{ m(a) + \int_a^b m'(x)\, dx \right\} - d_j = \int h_j(x)\, m'(x)\, dx$$

$$h_j(b)\, m(a) - d_j = \int \left(h_j(x) - h_j(b) \right) m'(x)\, dx$$

or

$$\boxed{f_j = d_j - h_j(b)\, m(a) = \int_a^b \left[h_j(b) - h_j(x) \right] m'(x)\, dx}$$

Remark: This process of integration by parts may be continued indefinitely so long as values of the function and its derivatives are known at a boundary.

eg: with the next integration by parts we obtain equations of the form

$$p_j = (r_j, m'') \qquad r_j = g_j^{(-2)} \quad \text{(double indefinite integral)},$$

The smallest norm model minimizing $\phi = (m'', m'')$ is called the "smoothest" model (at least by me!)

Flattest model for the 2-data gravity problem.

The information we had about the surface rocks was that $\rho(1) = 2.8 \frac{Mg}{m^3}$
Use a flattest model formulation that requires knowledge of the model at its right hand end point.

$$h_j(1)\rho(1) - d_j = \int_0^1 h_j(r)\rho'(r)\,dr$$

$$h_1(r) = \int_0^r u^2\,du = \frac{r^3}{3} \qquad h_2(r) = \int_0^r u^4\,du = \frac{r^5}{5}$$

$$h_1(1) = \frac{1}{3} \qquad h_2(1) = 1/5$$

$$\rho(1) = 2.8 \frac{Mg}{m^3} \qquad \bar{\rho} = 5.5\, Mg/m^3$$

$$f_1 = h_1(1)\rho(1) - d_1 = \frac{\rho(1)}{3} - \frac{\bar{\rho}}{3} = \frac{1}{3}(\rho(1) - \bar{\rho})$$

$$f_2 = h_2(1)\rho(1) - d_2 = \frac{\rho(1)}{5} - \frac{\bar{\rho}\gamma}{2} = \frac{1}{5}\left(\rho(1) - \frac{5}{2}\bar{\rho}\gamma\right)$$

$$\cancel{\tfrac{1}{3}}(\rho(1) - \bar{\rho}) = \int_0^1 \frac{r^3}{\cancel{3}}\rho'(r)\,dr$$

$$\cancel{\tfrac{1}{5}}\left(\rho(1) - \frac{5}{2}\bar{\rho}\gamma\right) = \int_0^1 \frac{r^5}{\cancel{5}}\rho'(r)\,dr$$

We obtain

$$-2.7 = \int_0^1 r^3 \rho'(r)\, dr$$

$$-1.7482 = \int_0^1 r^5 \rho'(r)\, dr$$

$$\Gamma = \begin{pmatrix} \frac{1}{7} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{11} \end{pmatrix} \qquad \Gamma^{-1} = 1559.25 \begin{pmatrix} \frac{1}{11} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{1}{7} \end{pmatrix}$$

$$\vec{\beta} = \Gamma^{-1} f = -1559.25 \begin{pmatrix} \frac{1}{11} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{1}{7} \end{pmatrix} \begin{pmatrix} -2.7 \\ -1.7482 \end{pmatrix} = \begin{pmatrix} -79.849 \\ 78.363 \end{pmatrix}$$

So

$$\rho'(r) = \beta_1 r^3 + \beta_2 r^5$$

$$\Rightarrow \quad \rho(r) = \frac{\beta_1 r^4}{4} + \frac{\beta_2 r^6}{6} + C$$

$$\boxed{\rho(r) = C - 19.962\, r^4 + 13.06\, r^6}$$

Adjust the constant of integration C to make a surface density $\rho(1) = 2.8 \quad \Rightarrow \quad C = 9.702.$

$$\therefore \quad \boxed{\rho(r) = 9.702 - 19.962\, r^4 + 13.06\, r^6}$$

This curve monotonically decreases from the core to the surface.

$$\rho(0) = 9.7 \text{ Mg/m}^3$$

Remark: This is definitely a geophysically acceptable density distribution.

Summary: 2-data gravity problem

(1) We can rigorously generate models that fit the data

(2) Depending upon what objective function is minimized we get models with different character.

- some intuitively simple mathematical norms (eg $\|m\|^2$) don't produce physically appealing models
- others, like minimizing $\|m'\|^2$ or $\|m - m_{ref}\|^2$ produce more reasonable models.

Clearly the approach we are establishing is to first "design" the right kind of objective f^{cn} for the problem. That is, the model that minimizes $\phi_m(m)$ should have the "right character" and be consistent with any a-priori information we have about the model.

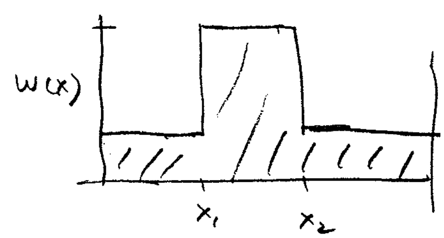
We already have some components for building up a general objective f^{cn} but we need one more.

Weighted model objective f^{cns}

Let
$$\phi_m = \int_a^b w(x)\, m^2(x)\, dx \quad \text{or} \quad \int_a^b w(x)\left(m(x) - m_{ref}(x)\right)^2 dx$$

where $\quad w(x) > 0 \quad w \in [a, b]$

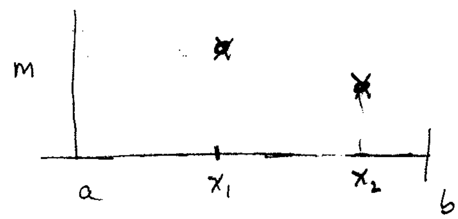
Suppose



This weighting discriminates against $m_c(x)$ (constructed model) to have energy in this region.

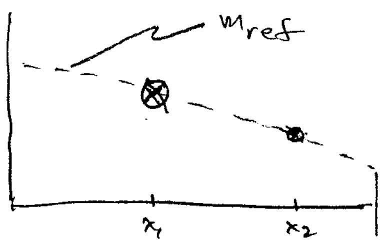
Potential uses of the weighting f^{cn}

(1) Construct a model which includes local knowledge obtained from previous investigations

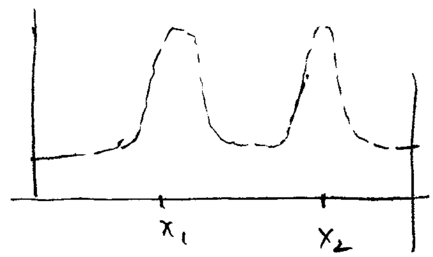
eg 

Estimates of $m(x)$ are obtained at x_1, x_2

We want the constructed model to be close to these values at these locations.



Construct m_{ref}



at x_1, x_2 the weight $w(x)$ is large $\Rightarrow$ $m_c(x)$ will tend to be close to $m_{ref}(x_1)$, $m_{ref}(x_2)$

But away from those locations our confidence in m_{ref} is diminished so we are not distressed if our reconstructed model differs substantially.

Remark: As we shall see, there are other ways to include this type of information, but the above scenario is valid.

(2) Specific appraisal analysis

Suppose $m_c(x)$ 

Question: is this feature really there (?). Or is it in the right location (maybe it should be shallower or deeper)
Design an objective f^{cn} to discriminate against the feature or put it in a different location.

(3) General exploration of model space.

Generic Model Objective function

If we combine all of the previous ideas we obtain one (of many) useful forms for a model objective f^{cn}. It's one that we will use here

$$\boxed{\phi_m = \alpha_s \int w_s(x)(m - m_{ref})^2 dx + \alpha_x \int w_x \left(\frac{d}{dx}(m - m_{ref})\right)^2 dx}$$

α_s, α_x : positive constants

$w_s(x)$: weighting f^{cn} for the smallest model component

$w_x(x)$: weighting f^{cn} for the derivative.

m_{ref} : reference model (often leave it out of the second term)

Remark: This form of objective f^{cn} is chosen because of its simplicity and yet great functionality. By varying a relatively few constants or functions, one can generate quite different kinds of models, put in a priori information and explore model space.

Remark: We will later add even more flexibility to the problem by
(i) replacing the l_2-norms with more general l_p norms

So, we should be all ready to solve our inverse problem.
Let's consider a simple, but typical example (One related to the NMR experiment and also, more generally, a Laplace Transform)
(This example is one that you'll do for an assignment so I don't want to do everything for you here, but we do need to illustrate the effects of instability or ill-posedness.

Problem: Given data $d_j = (g_j, m)$

$$\boxed{g_j(x) = e^{-jx}} \qquad x \in [0,1] \quad j = 0, 20$$

Generate true data using a model $\boxed{m(x) = 1 - \frac{1}{2}\cos 2\pi x}$

Compute the smallest model $(\phi_m = (m,m))$ that reproduces the data.
Solution is $m_c(x) = \sum \alpha_j g_j(x)$

$$\Gamma \alpha = d$$

$$\Gamma_{ij} = (g_i, g_j)$$

Do everything in double precision arithmetic. Data and inner product matrix is generated from analytic expressions. (You'll derive them in your homework)

Examples: (1) Data, true + recovered model. (few data, lots of data)
(2) Perturbed data

Summary: (numerical test example)

We have found a way to overcome the fundamental problem of non-uniqueness (namely to generate a model with certain characteristics) but there is a further problem of numerical instability.

The inverse problem is ill-posed because of this instability.

Basically, for our calculation, we had

$$\boxed{\Gamma \alpha = d}$$

Γ is symmetric and positive definite (so long as the kernels are linearly independent), so

$$\Gamma = R \Lambda R^T \qquad \Gamma \text{ is } N \times N$$

where $R = \begin{pmatrix} | & & | \\ r_1 & \cdots & r_N \\ | & & | \end{pmatrix}$ is an orthonormal matrix

$\Lambda = \text{diag}(\lambda_1, \ldots \lambda_N)$ $\qquad$ $\lambda_1 > \lambda_2 \ldots \lambda_N$ are the eigenvalues of $\Gamma \vec{r}_i = \lambda_i \vec{r}_i$

So

$$R \Lambda R^T \alpha = d$$
$$R^T \alpha = \Lambda^{-1} R^T d$$
$$\alpha = R \Lambda^{-1} R^T d \qquad \left(\alpha = \sum_{i=1}^{N} \frac{(r_i^T d)}{\lambda_i} \vec{r}_i \right)$$

$$\boxed{m(x) = \sum_{i=1}^{N} \alpha_i g_i(x)}$$

So the numerical difficulties stem from the small eigenvalues of the inner product matrix. Effectively Γ is almost singular ($\det \Gamma = \prod_{i=1}^{N} \lambda_1 \lambda_2 \cdots \lambda_N \sim 0$

Note that difficulties will arise because of numerical imprecision as the number of data grow! There will also be problem if the data vector d is incorrect.