Fusing Continuous and Discrete Data, on the Example of Merging Seismic and Gravity Models in Geophysics

Omar Ochoa[1,2], Aaron Velasco[1,3], Vladik Kreinovich[1,2]

[1]Cyber-ShARE Center of Excellence: A Center for Sharing Cyberresources to Advance Science and Education
Departments of [2]Computer Science and [3]Geological Sciences
University of Texas at El Paso, El Paso, TX 79968, USA
omar@miners.utep.edu, velasco@geo.utep.edu, vladik@utep.edu

1. Case Study

- To find the density $\rho(v)$ at different locations and different depths, we can use two types of data:
  - the *seismic data*, i.e., the arrival times of signals from earthquake and test explosions;
  - the *gravity* data.
- Both data provide complementary information:
  - seismic data provides information about a narrow zone around a path;
  - gravity data provides information about the larger area – but with much smaller spatial resolution.
- At present, there are no efficient algorithms for processing both types of data.
- So, we must fuse the results of processing these two types of data: a seismic model and a gravity model.

2. Computational Problem: Need to Fuse Discrete and Continuous Models

- Traditionally, seismic models are *continuous*: the velocity smoothly changes with depth.
- In contrast, the gravity models are *discrete*: we have layers, in each of which the velocity is constant.
- The abrupt transition corresponds to a steep change in the continuous model.
- Both models locate the transition only approximately.
- So, if we simply combine the corresponding values value-by-value, we will have *two* transitions instead of one:
  - one transition where the continuous model has it, and
  - another transition nearby where the discrete model has it.

3. What We Plan to Do

- *We want* to avoid the misleading double-transition models.
- *Idea:* first fuse the corresponding transition locations.
- *In this paper*, we provide an algorithm for such location fusion.
- *Specifically*, first, we formulate the problem in the probabilistic terms.
- *Then*, we provide an algorithm that produces the most probable transition location.
- *We show* that the result of the probabilistic location algorithm is in good accordance with common sense.
- *We also show* how the commonsense intuition can be reformulated in fuzzy terms.

4. Available Data: What is Known and What Needs to Be Determined

- For each location, in the discrete model, we have the exact depth z_d of the transition.
- In contrast, for the continuous model, we do not have the abrupt transition.
- Instead, we have velocity values $v(z)$ at different depths.
- We must therefore extract the corresponding transition value z_c from the velocity values.
- To be more precise, we have values $v_1, v_2, \ldots, v_i, \ldots, v_n$ corresponding to different depths.
- We need to find i for which the transition occurs between the depths i and $i+1$.

5. Probabilistic Approach

- The difference $\Delta v_j \stackrel{\text{def}}{=} v_j - v_{j+1}$ $(j \neq i)$ is caused by many independent factors.
- Due to the Central Limit Theorem, we thus assume that it is normally distributed, with probability density

$$p_j \stackrel{\text{def}}{=} \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \exp\left(-\frac{1}{2 \cdot \sigma^2} \cdot (\Delta v_j)^2\right).$$

- The value Δv_i at the transition depth i is *not* described by the normal distribution.
- We assume that differences corresponding to different depths j are independent, so:

$$L_i = \prod_{j \neq i} p_j = \prod_{j \neq i} \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \exp\left(-\frac{1}{2 \cdot \sigma^2} \cdot (\Delta v_j)^2\right).$$

6. How to Find the Location: The General Idea of the Maximum Likelihood Approach

- *Reminder:* the likelihood of each model is:

$$L_i = \prod_{j\ne i} p_j = \prod_{j\ne i} \frac{1}{\sqrt{2\cdot\pi}\cdot\sigma}\cdot\exp\left(-\frac{1}{2\cdot\sigma^2}\cdot(\Delta v_j)^2\right).$$

- *Natural idea:* select the parameters for which the likelihood of the observed data is the largest.
- The value L_i is the largest if and only if $-\ln(L_i)$ is the smallest: $-\ln(L_i) = \text{const} + \frac{1}{2\cdot\sigma^2}\cdot\sum\limits_{j\ne i}(\Delta v_j)^2 \to \min\limits_i$.
- This sum is equal to $\sum\limits_{j\ne i}(\Delta v_j)^2 = \sum\limits_{j=1}^{n-1}(\Delta v_j)^2 - (\Delta v_i)^2$.
- The first term in this expression does not depend on i.
- Thus, the difference is the smallest $\Leftrightarrow$ the value $(\Delta v_i)^2$ is the largest $\Leftrightarrow$ $|\Delta v_i|$ is the largest.

7. Resulting Location

- *We want:* to select the most probable location of the transition point.
- *We select:* the depth i_0 for which the absolute value $|\Delta v_i|$ of the difference $\Delta v_i = v_{i+1} - v_i$ is the largest.
- This conclusion seems to be very reasonable:
  - the most probable location of the actual abrupt transition between the layers
  - is the depth at which the measured difference is the largest.

8. The Results of the Probabilistic Approach are in Good Accordance with Common Sense

- Intuitively, for each depth i, our confidence that i a transition point depends on the difference $|\Delta v_i|$:
  - the smaller the difference, the less confident we are that this is the actual transition depth, and
  - the larger the difference, the more confident we are that this is the actual transition depth.
- In our probabilistic model, we select a location with the largest possible value $|\Delta v_i|$.
- This shows that the probabilistic model is in good accordance with common sense.
- This coincidence increases our confidence in this result.

9. It May Be Useful to Formulate the Common Sense Description in Fuzzy Terms

- Fuzzy logic is known to be a useful way to formalize imprecise commonsense reasoning.
- Common sense: the degree of confidence d_i that i is a transition point is $f(|\Delta v_i|)$, for some monotonic $f(z)$.
- It is reasonable to select a value i for which our degree of confidence is the largest $d_i = f(|\Delta v_i|) \to \max$.
- Since $f(z)$ is increasing, this is equivalent to

$$|\Delta v_i| \to \max.$$

- Of course, to come up with this conclusion, we do not need to use the fuzzy logic techniques.
- However, this description may be useful if we also have other expert information.

10. How Accurate Is This Location Estimate?

- *Reminder:* the likelihood has the form

$$L_i = \prod_{j \neq i} p_j = \prod_{j \neq i} \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \exp\left(-\frac{1}{2 \cdot \sigma^2} \cdot (\Delta v_j)^2\right).$$

- *We have found* the most probable transition i_0 as the value for which L_i is the largest.
- *Similarly:* we can find σ for which L_i is the largest:

$$\sigma^2 = \frac{1}{n-2} \cdot \sum_{j \neq i_0} (\Delta v_j)^2.$$

- The probability P_i that the transition is at location i is proportional to L_i: $P_i = c \cdot L_i$.
- The coefficient c can be determined from the condition that the total probability is 1: $1 = \sum\limits_i P_i = c \cdot \sum\limits_{i=1}^{n} L_i$.
- So, $c = \left(\sum L_i\right)^{-1}$.

11. Accuracy of the Location Estimate (cont-d)

- The mean square deviation σ_0^2 of the actual transition depth from our estimate i_0 is defined as

$$\sigma_0^2 = \sum_{i=1}^{n-1} (i - i_0)^2 \cdot P_i.$$

- We know that $P_i = c \cdot L_i$, and we have formulas for computing L_i and c, so we can compute σ_0.
- We applied this algorithm to the seismic model of El Paso area, and got $\sigma_0 \approx 1.5$ km.
- This value is of the same order (1-2 km) as the difference between:
  - the border depth estimates coming from the seismic data and
  - the border depth coming from the gravity data.

12. How to Fuse the Depth Estimates

- Now, we have two estimates for the transition depth:
  - the estimate i_d from the discrete (gravity) model;
  - the estimate i_0 from the continuous (seismic) model.
- The estimate i_d comes from a standard statistical analysis, so we know standard deviation σ_d.
- For i_0, we already know the standard deviation σ_0.
- It is reasonable to assume that both differences $i_d - i$ and $i_0 - i$ are normally distributed and independent:

$$p_i = \exp\left(-\frac{(i_d - i_f)^2}{2\cdot\sigma_d^2}\right)\cdot\exp\left(-\frac{(i_0 - i_f)^2}{2\cdot\sigma_0^2}\right).$$

- The most probable location i is when $p_i \to \max$, i.e.:

$$i_f = \frac{i_d\cdot\sigma_d^{-2} + i_0\cdot\sigma_0^{-2}}{\sigma_d^{-2}+\sigma_0^{-2}}.$$

13. Towards Fusing Actual Maps

- In the discrete model:
  - values $i < i_d$ correspond to the upper zone;
  - values $i > i_d$ correspond to the lower zone.
- Similarly, in the continuous model:
  - values $i < i_0$ correspond to the upper zone;
  - values $i > i_0$ correspond to the lower zone.
- So, for depths $i \le \min(i_0, i_d)$ and $i \ge \max(i_0, i_d)$, both models correctly describe the zone.
- For these depths, we can simply fuse the values from both models.
- We can fuse them similarly to how we fused the depths.
- For intermediate depths, we need to adjust the models: e.g., by taking the nearest value from the correct zone.

14. How to Fuse the Actual Maps: First Stage

- *First:* we adjust both models so that they both have a transition at depth i_f.
- *Adjusting the discrete model* is easy: we replace
  - the original depth i_d
  - with the new (more accurate) fused value i_f.
- *Adjusting the continuous model:*
  - when $i_f < i_0$, the values at depths i between i_f and i_0 are erroneously assigned to the the upper zone;
  - these values v_i must be replaced by the the value of the nearest point at the lower zone v_{i_0+1};
  - when $i_f > i_0$, the values at depths i between i_0 and i_f are erroneously assigned to the the lower zone;
  - these values v_i must be replaced by the the value of the nearest point at the upper zone v_{i_0}.

15. How to Merge the Adjusted Models

- For each depth i, we now have two adjusted values v_i' and v_i'' corresponding to two adjusted models.
- Let σ' and σ'' be the corresponding standard deviations.
- It is reasonable to assume that both differences $v_i' - v_i$ and $v_i'' - v_i$ are normally distributed and independent:

$$p(v_i) = \exp\left(-\frac{(v_i' - v_i)^2}{2 \cdot (\sigma')^2}\right) \cdot \exp\left(-\frac{(v_i'' - v_i)^2}{2 \cdot (\sigma'')^2}\right).$$

- The most probable value $\widetilde{v}_i$ is when $p(v_i) \to \max$, i.e.:

$$\widetilde{v}_i = \frac{v_i' \cdot (\sigma')^{-2} + v_i'' \cdot (\sigma'')^{-2}}{(\sigma')^{-2} + (\sigma'')^{-2}}.$$

16. Acknowledgment

This work was supported in part

- by the National Science Foundation grants HRD-0734825 (Cyber-ShARE Center) and DUE-0926721,
- by Grant 1 T36 GM078000-01 from the National Institutes of Health,
- by Grant MSM 6198898701 from MŠMT of Czech Republic, and
- by Grant 5015 "Application of fuzzy logic with operators in the knowledge based systems" from the Science and Technology Centre in Ukraine (STCU), funded by European Union.

The authors are thankful to Dr. Musa Hussein for his help and to the anonymous referees for their suggestions.