# Need for Expert Knowledge (and Soft Computing) in Cyberinfrastructure-Based Data Processing

**Vladik Kreinovich**

Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@utep.edu
http://www.cs.utep.edu/vladik

# 1. Need for Cyberinfrastructure

- A large amount of data has been collected and stored at different locations.
- Researchers and practitioners need easy and fast access to all the relevant data.
- For example, a geoscientist needs access to:
  - a state geological map (which is usually stored at the state's capital),
  - NASA photos (stored at NASA Headquarters and/or at one of corresponding NASA centers),
  - seismic data stored at different seismic stations, etc.
- An environmental scientist needs access:
  - to satellite radar data,
  - to data from bio-stations,
  - to meteorological data, etc.

## 2. What Is Cyberinfrastructure

- Cyberinfrastructure is a general name for hardware/software tools that facilitate such data transfer/processing.
- Ideally, this data transfer and processing should be as easy and convenient as a google search.
- At present, the main challenges in cyberinfrastructure design are related to the actual development of:
  - the corresponding hardware tools and
  - the corresponding software tools.
- Most existing cyberinfrastructure tools use existing well defined algorithms.
- The results of using cyberinfrastructure are exciting.
- However, there is still room for improvement.

## 3. Cyberinfrastructure: Expert Knowledge Is Needed

- Current cyberinfrastructure results are based only on data processing.
- Some of these results do not make geological sense.
- It is necessary to take into account expert knowledge.
- Specifically, we must incorporate expert knowledge directly into the cyberinfrastructure.
- Some expert knowledge is formulated in precise terms; these types of knowledge are easier to incorporate.
- A large part of expert knowledge is formulated by using *imprecise* (fuzzy) words (like "small").
- To deal with such knowledge, fuzzy techniques have been invented.
- So, to incorporate this knowledge, it is natural to use fuzzy techniques.

## 4. What We Do In This Talk

- In this talk, we describe several problems in which such incorporation is needed.
- These problems come from our experience from geo- and environmental applications of cyberinfrastructure.
- First, we show that expert knowledge is needed even when we "*fuse*" data from different sources.
- Then, we show how expert knowledge can be used in *processing* data.
- Finally, we show how expert knowledge can be used in selecting the best ways of *getting* the data.

## 5. Part 1: Need for Data Fusion

- In many practical situations, we have several results $\widetilde{x}^{(1)}, \ldots, \widetilde{x}^{(n)}$ of measuring the same quantity $x$.
- These results are different since measurements are never 100% accurate.
- It is know that by combining different measurement results, we increase accuracy.
- Simplest case: we use the same measuring instrument for all measurements.
- In this case, an arithmetic average reduces the st. dev. by a factor of $\sqrt{n}$:

$$\widetilde{x} = \frac{\widetilde{x}^{(1)} + \ldots + \widetilde{x}^{(n)}}{n}.$$

- When we fuse measurements of *different* accuracy, we need to use different weights for *different* values $\widetilde{x}^{(i)}$.

## 6. Data Fusion: Challenge

- When we fuse measurements of *different* accuracy, we need to use different weights for *different* values $\widetilde{x}^{(i)}$.
- Sometimes, we can find the actual values and thus, estimate the accuracy of different measurements.
- In other cases – e.g., in geosciences – it is difficult to find the actual density at depth 40 km.
- Hence, in geosciences, it is difficult to gauge the accuracy of seismic, gravity, and other techniques.
- In this case, we need to estimate the accuracies from the observations.
- We will show that in this case, seemingly reasonable statistical methods do not work well.
- Thus, statistical methods need to be supplemented with expert knowledge.

## 7. Traditional Statistical Methods: Reminder

- In many cases, the measurement error is caused by many different causes.
- It is known that the distribution of the sum of many small random variables is $\approx$ normally distributed.
- So, we can conclude that the measurement errors are normally distributed, with probability density

$$\rho(\widetilde{x})=\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(-\frac{(\widetilde{x}-x)^2}{2\sigma^2}\right).$$

- If we have $n$ results $\widetilde{x}^{(i)}$ of independent measurements, then prob. is prop. to $\rho=\prod\limits_{i=1}^{n}\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(-\frac{(\widetilde{x}^{(i)}-x)^2}{2\sigma^2}\right).$
- Maximum Likelihood Method: select most probable $x$ and $\sigma$, for which prob. (hence $\rho$) is the largest.

# 8. Traditional Statistical Methods (cont-d)

- Maximizing $\rho = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot \exp\left(-\frac{(\widetilde{x}^{(i)} - x)^2}{2\sigma^2}\right)$ is equivalent to minimizing
$$\psi = -\ln(\rho) = \text{const} + n \cdot \ln(\sigma) + \sum_{i=1}^{n} \frac{(\widetilde{x}^{(i)} - x)^2}{2\sigma^2}.$$
- W.r.t. $x$, we get the Least Squares method which leads to the arithmetic average $x = \frac{1}{n} \cdot \sum_{i=1}^{n} \widetilde{x}^{(i)}$.
- Differentiating $\psi$ w.r.t. $\sigma$ and equating to 0, we get
$$\frac{n}{\sigma} - \sum_{i=1}^{n} \frac{(\widetilde{x}^{(i)} - x)^2}{\sigma^3} = 0.$$
- So, we get the usual estimate $\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (\widetilde{x}^{(i)} - x)^2$.

## 9. Case of Different Measuring Instruments (MI): Surprising Problem

- *Situation:* for different quantities $x_j$, $j = 1, \ldots, m$, we have measurement results $\widetilde{x}_j^{(i)}$ corr. to diff. MI, w/diff. $\sigma_i$.
- The resulting probability is proportional to

$$\rho = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi} \cdot \sigma_i} \cdot \exp\left(-\frac{(\widetilde{x}_j^{(i)} - x_j)^2}{2\sigma_i^2}\right).$$

- *Seemingly natural idea:* use Maximum Likelihood method, i.e., find $x_j$ and $\sigma_i$ for which $\rho \to \max$.
- *We tried*, and found that at maximum, one of $\sigma_i$ is 0.
- *We then theoretically confirmed:* that maximum $\rho_{\max} = \infty$ is attained:
  - when $\sigma_{i_0} = 0$ for some $i_0$, and
  - when $x_j = \widetilde{x}_j^{(i_0)}$ for all $j$.

## 10. Analysis of the Problem

- *We know:* that all the measuring instruments are imperfect, i.e., $\sigma_i > 0$.
- *From the mathematical viewpoint:* we get $\sigma_{i_0} = 0$ for some $i_0$.
- This mathematical solution is not physically meaningful.
- To avoid this non-physical solution, we need to explicitly add the requirement that $\sigma_i > 0$ for all $i$.
- This *crisp* requirement does not help: by taking smaller and smaller $\sigma_{i_0}$, we can get $\rho$ as large as possible.
- Intuitively, what we need is a *fuzzy* requirement – that all $\sigma_i$ are not too small.
- This fuzzy requirement enables us to avoid non-physical values of $\sigma_i$.

# 11. Towards an Algorithm

- $\rho = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi} \cdot \sigma_i} \cdot \exp\left(-\frac{(\widetilde{x}_j^{(i)} - x_j)^2}{2\sigma_i^2}\right) \to \max.$
- $\psi = -\ln(\rho) = n \cdot \sum_{i=1}^{n} \ln(\sigma_i) + \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(\widetilde{x}_j^{(i)} - x_j)^2}{2\sigma_i^2} \to \min.$
- Differentiating $\psi$ w.r.t. $x_j$ and $\sigma_i$, we get:

$$x_j = \frac{\sum_{i=1}^{n} \sigma_i^{-2} \cdot \widetilde{x}_j^{(i)}}{\sum_{i=1}^{n} \sigma_i^{-2}}; \quad \sigma_i^2 = \frac{1}{m} \cdot \sum_{j=1}^{m} (\widetilde{x}_j^{(i)} - x_j)^2.$$

- We first take $\sigma_i = \text{const}$, then iteratively compute:

  (1) $x_j$ from $\sigma_j$, (2) $\sigma_j$ from $x_i$, (3) $x_j$ from $\sigma_j$, ...
- We stop when one of $\sigma_i$ becomes too small (2-3 cycles).
- We return results of the previous cycle (cf. astrometry).

# 12. Part 2: Use of Expert Knowledge in Actual Data Processing

- We need to reconstruct the values of the quantities of interest from the measurement results.
- Geosciences example: reconstructing density at different depths and different locations.
- Often, several drastically different density distributions are consistent with the same observations.
- Such problems are called "ill-posed".
- Out of all these distributions, we need to select the physically meaningful one(s).
- This is where expert knowledge is needed, to describe what "physically meaningful" means.
- On the example of the above geophysical problem, we show how to use this expert knowledge.

## 13. Determining Earth Structure Is Important

- *Importance:* civilization greatly depends on the things we extract from the Earth: oil, gas, water.
- *Need* is growing, so we must find new resources.
- *Problem:* most easy-to-access mineral resources have been discovered.
- *Example:* new oil fields are at large depths, under water, in remote areas – so drilling is very expensive.
- *Objective:* predict resources before we invest in drilling.
- *How:* we know what structures are promising.
- *Example:* oil and gas concentrate near the top of (natural) underground domal structures.
- *Conclusion:* to find mineral resources, we must determine the structure at diff. depths $z$ and locations $(x, y)$.

# 14. Data that We Can Use to Determine the Earth Structure

- *Available measurement results:* those obtained without drilling boreholes.
- *Examples:*
  - gravity and magnetic measurements;
  - travel-times $t_i$ of seismic ways through the earth.
- *Need for active seismic data:*
  - passive data from earthquakes are rare;
  - to get more information, we make explosions, and measure how the resulting seismic waves propagate.
- *Resulting seismic inverse problem:*
  - we know the travel times $t_i$;
  - we want to reconstruct velocities at different depths.

## 15. Algorithm for the Forward Seismic Problem

- *We know:* velocities $v_j$ in each grid cell $j$.
- *We want to compute:* traveltimes $t_i$.
- *First step:* find shortest (in time) paths.
- *Within cell:* path is a straight line.
- *On the border:* between cells with velocities $v$ and $v'$, we have Snell's law $\dfrac{\sin(\varphi)}{v} = \dfrac{\sin(\varphi')}{v'}$.
- *Comment:* if $\sin(\varphi') > 1$, the wave cannot get penetrate into the neighboring cell; it bounces back.
- *Resulting traveltimes:* $t_i = \sum\limits_j \dfrac{\ell_{ij}}{v_j}$, where $\ell_{ij}$ is the length of the part of $i$-th path within cell $j$.
- *Simplification:* use slownesses $s_j \stackrel{\text{def}}{=} \dfrac{1}{v_j}$; $t_i = \sum\limits_j \ell_{ij} \cdot s_j$.

# 16. Algorithm for the Inverse Problem: General Description

- *The most widely used:* John Hole's iterative algorithm.
- *Starting point:* reasonable initial slownesses.
- *On each iteration:* we use current (approximate) slownesses $s_j$ to compute the travel-times $t_i = \sum\limits_j \ell_{ij} \cdot s_j$.
- *Fact:* measured travel-times $\widetilde{t}_i$ are somewhat different: $\Delta t_i \stackrel{\text{def}}{=} \widetilde{t}_i - t_i \neq 0$.
- *Objective:* find $\Delta s_j$ so that $\sum \ell_{ij} \cdot (s_j + \Delta s_j) = \widetilde{t}_i$.
- *Problem:* we have many observations $n$, and computation time $\sim n^3$ – too long, so we need faster techniques.
- *Stopping criterion:* when average error $\dfrac{1}{n}\sum\limits_{i=1}^{n}(\Delta t_i)^2$ is below noise.

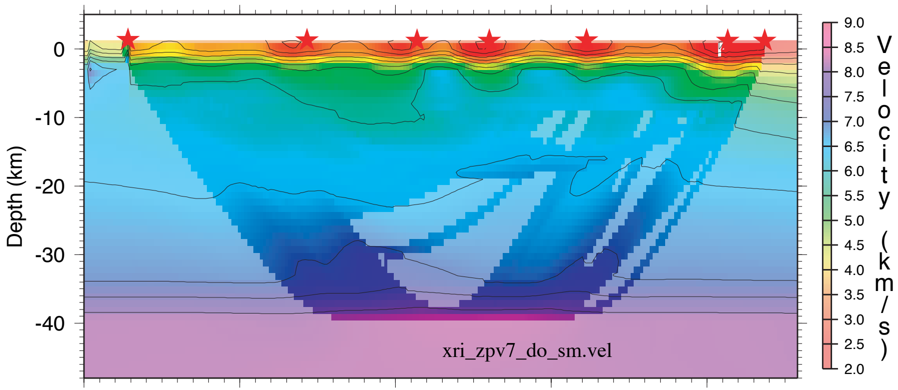
## 17. Algorithm for the Inverse Problem: Details

- *Objective (reminder):* find $\Delta s_j$ s.t. $\sum \ell_{ij} \cdot \Delta s_j = \Delta t_i$.
- *Simplest case:* one path.
- *Specifics:* under-determined system: 1 equation, many unknowns $\Delta s_j$.
- *Idea:* no reason for $\Delta s_j$ to be different: $\Delta s_j \approx \Delta s_{j'}$.
- *Formalization:* minimize $\sum\limits_{j,j'} (\Delta s_j - \Delta s_{j'})^2$ under the constraint $\sum \ell_{ij} \cdot \Delta s_j = \Delta t_i$.
- *Solution:* $\Delta s_j = \dfrac{\Delta t_i}{L_i}$ for all $j$, where $L_i = \sum\limits_j \ell_{ij}$.
- *Realistic case:* several paths; we have $\Delta s_{ij}$ for different paths $i$.
- *Idea:* least squares $\sum\limits_i (\Delta s_j - \Delta s_{ij})^2 \to \min$.
- *Solution:* $\Delta s_j$ is the average of $\Delta s_{ij}$.

# 18. Successes, Limitations, Need for Prior Knowledge

- *Successes:* the algorithm usually leads to reasonable geophysical models.
- *Limitations:* often, the resulting velocity model is not geophysically meaningful.
- *Example:* resulting velocities outside of the range of reasonable velocities at this depth.
- *It is desirable:* incorporate the expert knowledge into the algorithm for solving the inverse problem.

# Hole Tomography Smashed Masked Velocity Models



## 19. Case of Interval Prior Knowledge

- *Idea:* for each cell $j$, a geophysicist provides an interval $[\underline{s}_j, \overline{s}_j]$ of possible values of $s_j$.
- *Hole's code:* along each path $i$, we find corrections $\Delta s_{ij}$ that minimize
$$\sum_{j,j'} (\Delta s_{ij} - \Delta s_{ij'})^2$$
under the constraint
$$\sum_{j=1}^{c} \ell_{ij} \cdot \Delta s_{ij} = \Delta t_i.$$
- *Modification:* we must minimize under the additional constraints
$$\underline{s}_j \le s_j^{(k)} + \Delta s_{ij} \le \overline{s}_j.$$
- *What we designed:* an $O(c \cdot \log(c))$ algorithm for solving this new problem.

# 20. Main Idea of an Algorithm

- *Idea* – method of alternating projections:
  - first, add a correction that satisfy the first constraint,
  - then, the additional correction that satisfies the second constraint,
  - etc.
- *Specifics:*
  - first, add equal values $\Delta s_{ij}$ to minimize $\Delta t_i$;
  - restrict the values to the nearest points from $[\underline{s}_j, \overline{s}_j]$,
  - repeat until converges.
- *Comment:* this way, we can also use other prior knowledge (e.g., probabilistic).

# 21. New Algorithm: For Each Path on Each Iteration

- *Case:* $\Delta t_i > 0$; for $\Delta t_i < 0$, we have similar formulas.
- Compute, for each cell $j$,
$$\underline{\Delta}_j = \underline{s}_j - s_j^{(k-1)} \text{ and } \overline{\Delta}_j = \overline{s}_j - s_j^{(k-1)}.$$
- Sort values $\overline{\Delta}_j$ into
$$\overline{\Delta}_{(1)} \le \overline{\Delta}_{(2)} \le \ldots \le \overline{\Delta}_{(c)}.$$
- For every $p$ from 0 to $c$, compute:

  $A_0 = 0,\ \mathcal{L}_0 = L_i,\ A_p = A_{p-1} + \ell_{i(p)} \cdot \overline{\Delta}_{(p)},\ \mathcal{L}_p = \mathcal{L}_{p-1} - \ell_{i(p)}.$
- Compute $S_p = A_p + \mathcal{L}_p \cdot \Delta_{(p+1)}$, and find $p$ s.t.
$$S_{p-1} \le \Delta t_i < S_p.$$
- Take $\Delta s_{i(j)} = \overline{\Delta}_j$ for $j \le p$, and $\Delta s_{(j)} = \dfrac{\Delta t_i - A_p}{\mathcal{L}_p}$ else.
- Then, average $\Delta s_{ij}$ over paths $i$.

## 22. Explicit Expert Knowledge: Fuzzy Uncertainty

- Experts can usually produce a wide interval of which they are practically 100% certain.
- In addition, experts can also produce narrower intervals about which their degree of certainty is smaller.
- As a result, instead of a *single* interval, we have a *nested* family of intervals corr. to diff. levels of uncertainty.
- In effect, we get a *fuzzy* interval (of which different intervals are $\alpha$-cuts).
- *Previously:* a solution is satisfying or not.
- *New idea:* a satisfaction *degree* $d$.
- *Specifics:* $d$ is the largest $\alpha$ for which all $s_i$ are within the corresponding $\alpha$-cut intervals.

## 23. How We Can Use Fuzzy Uncertainty

- *Objective:* find the largest possible value $\alpha$ for which the slownesses belong to the $\alpha$-cut intervals.
- *Possible approach:*
  - try $\alpha = 0$, $\alpha = 0.1$, $\alpha = 0.2$, etc., until the process stops converging;
  - the solution corresponding to the previous value $\alpha$ is the answer.
- *Comment:*
  - this is the basic straightforward way to take fuzzy-valued expert knowledge into consideration;
  - several researchers successfully used fuzzy expert knowledge in geophysics (Nikravesh, Klir, et al.).

## 24. Part 3: How to Best Acquire the Data?

- The above two applications are related to processing the existing data.
- In many practical situations, the data from the existing instruments is not sufficient.
- So, new measuring instruments are needed.
- E.g.: to get a better understanding of weather and climate processes, we need to place more instruments in Arctic, Antarctic, desert areas.
- Which are the best locations for these new instruments?
- We would like to gain as much information as possible from these new instruments.
- The problem is that we do not know exactly what processes we will observe.

## 25. How to Best Acquire the Data? (cont-d)

- We would like to gain as much information as possible from these new instruments.
- The problem is that we do not know exactly what processes we will observe.
- This uncertainty is what motivates us to build the new stations in the first place.
- Because of this uncertainty, to make a reasonable decision, we need to use expert knowledge.
- NASA faced a similar problem when selecting the Moon landing sites.
- We will used NASA's experience to find the optimal location of meteorological instruments.

# 26. Case Study: Detailed Description

- *Objective:* select the best location of a sophisticated multi-sensor meteorological tower.
- *Constraints:* we have several criteria to satisfy.
- *Example:* the station should not be located too close to a road.
- *Motivation:* the gas flux generated by the cars do not influence our measurements of atmospheric fluxes.
- *Formalization:* the distance $x_1$ to the road should be larger than a threshold $t_1$: $x_1 > t_1$, or $y_1 \stackrel{\text{def}}{=} x_1 - t_1 > 0$.
- *Example:* the inclination $x_2$ at the tower's location should be smaller than a threshold $t_2$: $x_2 < t_2$.
- *Motivation:* otherwise, the flux determined by this inclination and not by atmospheric processes.

## 27. General Case

- *In general:* we have several differences $y_1, \ldots, y_n$ all of which have to be non-negative.
- For each of the differences $y_i$, the larger its value, the better.
- Our problem is a typical setting for *multi-criteria optimization.*
- A most widely used approach to multi-criteria optimization is *weighted average*, where
  - we assign weights $w_1, \ldots, w_n > 0$ to different criteria $y_i$ and
  - select an alternative for which the weighted average

$$w_1 \cdot y_1 + \ldots + w_n \cdot y_n$$

attains the largest possible value.

## 28. Limitations of the Weighted Average Approach

- *In general:* the weighted average approach often leads to reasonable solutions of the multi-criteria problem.
- *In our problem:* we have an additional requirement – that all the values $y_i$ must be positive. So:
  - when selecting an alternative with the largest possible value of the weighted average,
  - we must only compare solutions with $y_i > 0$.
- *We will show:* under the requirement $y_i > 0$, the weighted average approach is not fully satisfactory.
- *Conclusion:* we need to find a more adequate solution.

## 29. Limitations of the Weighted Average Approach: Details

- The values $y_i$ come from measurements, and measurements are never absolutely accurate.
- The results $\widetilde{y}_i$ of the measurements are not exactly equal to the actual (unknown) values $y_i$.
- *If:* for some alternative $y = (y_1, \ldots, y_n)$
  - we measure the values $y_i$ with higher and higher accuracy and,
  - based on the measurement results $\widetilde{y}_i$, we conclude that $y$ is better than some other alternative $y'$.
- *Then:* we expect that the actual alternative $y$ is indeed better than $y'$ (or at least of the same quality).
- Otherwise, we will not be able to make any meaningful conclusions based on real-life measurements.

## 30. The Above Natural Requirement Is Not Always Satisfied for Weighted Average

- *Simplest case:* two criteria $y_1$ and $y_2$, w/weights $w_i > 0$.
- If $y_1, y_2, y'_1, y'_2 > 0$, and $w_1 \cdot y_1 + w_2 \cdot y_2 > w_1 \cdot y'_1 + w_2 \cdot y'_2$, then $y = (y_1, y_2) \succ y' = (y'_1, y'_2)$.
- If $y_1 > 0$, $y_2 > 0$, and at least one of the values $y'_1$ and $y'_2$ is non-positive, then $y = (y_1, y_2) \succ y' = (y'_1, y'_2)$.
- Let us consider, for every $\varepsilon > 0$, the tuple $y(\varepsilon) \stackrel{\text{def}}{=} (\varepsilon, 1 + w_1/w_2)$, and $y' = (1, 1)$.
- In this case, for every $\varepsilon > 0$, we have

$$w_1 \cdot y_1(\varepsilon) + w_2 \cdot y_2(\varepsilon) = w_1 \cdot \varepsilon + w_2 + w_2 \cdot \frac{w_1}{w_2} = w_1 \cdot (1 + \varepsilon) + w_2$$

and $w_1 \cdot y'_1 + w_2 \cdot y'_2 = w_1 + w_2$, hence $y(\varepsilon) \succ y'$.

- However, in the limit $\varepsilon \to 0$, we have $y(0) = \left(0, 1 + \dfrac{w_1}{w_2}\right)$, with $y(0)_1 = 0$ and thus, $y(0) \prec y'$.

## 31. Heuristic Idea Motivated by Fuzzy Logic

- *Problem:* the first criterion must satisfied *and* the second criterion must be satisfied, ...
- *Fuzzy logic approach:*
  - First, we estimate the degrees $d_1, \ldots, d_n$ to which each of the constraints is satisfied.
  - Then, we use a t-norm (fuzzy analogue of "and") to combine these degrees into a single degree $d$.
- *Simplest membership functions:* triangular, for which $d_i(y_i) = k_i \cdot y_i$, with $k_i > 0$ (when $y_i > 0$).
- *Selecting a t-norm:* the simplest is min, but it is not smooth hence tough to optimize; next simplest is $a \cdot b$.
- *Result:* maximize $d = \prod_{i=1}^{n} (k_i \cdot y_i) \Leftrightarrow$ maximize $\prod_{i=1}^{n} y_i$.
- *This approach is indeed better than weighted average:* e.g., if $y'(\varepsilon) \succ y$ and $y'(\varepsilon) \to y'(0)$, then $y'(0) \succeq y$.

## 32. Natural Next Idea: Use Hedges

- *Idea:* different criteria have different importance:
  - for some criteria, it is sufficient to have them somewhat satisfied;
  - for others, they must be very satisfied.
- So, instead of combing degrees $d_i$, we combined hedged degrees $h_i(d_i)$.
- *Natural requirement:* e.g., "very $(a \,\&\, b)$" should mean the same as "very $a$ and very $b$".
- Thus, $h(a \cdot b) = h(a) \cdot h(b)$ and hence, $h(a) = a^\alpha$.
- *Conclusion:* we combine $h(d_i) = d_i^{\alpha_i}$, i.e., we optimize the product $\prod\limits_{i=1}^{n} y_i^{\alpha_i}$.
- *What we prove:* this fuzzy-motivated expression is the *only* expression that satisfies reasonable properties.

# 33. Towards a Precise Description

- Each alternative is characterized by a tuple of $n$ positive values $y = (y_1, \ldots, y_n)$.
- Thus, the set of all alternatives is the set $(R^+)^n$ of all the tuples of positive numbers.
- For each two alternatives $y$ and $y'$, we want to tell whether
  - $y$ is better than $y'$ (we will denote it by $y \succ y'$ or $y' \prec y$),
  - or $y'$ is better than $y$ ($y' \succ y$),
  - or $y$ and $y'$ are equally good ($y' \sim y$).
- *Natural requirement:* if $y$ is better than $y'$ and $y'$ is better than $y''$, then $y$ is better than $y''$.
- The relation $\succ$ must be transitive.

## 34. Towards a Precise Description (cont-d)

- *Reminder:* the relation $\succ$ must be transitive.
- Similarly, the relation $\sim$ must be transitive, symmetric, and reflexive ($y \sim y$), i.e., be an *equivalence relation.*
- *An alternative description:* a transitive *pre-ordering* relation $a \succeq b \Leftrightarrow (a \succ b \vee a \sim b)$ s.t. $a \succeq b \ \vee \ b \succeq a$.
- Then, $a \sim b \Leftrightarrow (a \succeq b) \,\&\, (b \succeq a)$, and

$$a \succ b \Leftrightarrow (a \succeq b) \,\&\, (b \not\succeq a).$$

- *Additional requirement:*
  - *if* each criterion is better,
  - *then* the alternative is better as well.
- *Formalization:* if $y_i > y'_i$ for all $i$, then $y \succ y'$.

# 35. Scale Invariance: Motivation

- *Fact:* quantities $y_i$ describe completely different physical notions, measured in completely different units.
- *Examples:* wind velocities measured in m/s, km/h, mi/h; elevations in m, km, ft.
- Each of these quantities can be described in many different units.
- A priori, we do not know which units match each other.
- Units used for measuring different quantities may not be exactly matched.
- It is reasonable to require that:
  - if we simply change the units in which we measure each of the corresponding $n$ quantities,
  - the relations $\succ$ and $\sim$ between the alternatives $y = (y_1, \ldots, y_n)$ and $y' = (y'_1, \ldots, y'_n)$ do not change.

# 36. Scale Invariance: Towards a Precise Description

- *Situation:* we replace:
  - a unit in which we measure a certain quantity $q$
  - by a new measuring unit which is $\lambda > 0$ times smaller.
- *Result:* the numerical values of this quantity increase by a factor of $\lambda$: $q \to \lambda \cdot q$.
- *Example:* 1 cm is $\lambda = 100$ times smaller than 1 m, so the length $q = 2$ becomes $\lambda \cdot q = 2 \cdot 100 = 200$ cm.
- Then, scale-invariance means that for all $y, y' \in (R^+)^n$ and for all $\lambda_i > 0$, we have
  - $y = (y_1, \ldots, y_n) \succ y' = (y'_1, \ldots, y'_n)$ implies $(\lambda_1 \cdot y_1, \ldots, \lambda_n \cdot y_n) \succ (\lambda_1 \cdot y'_1, \ldots, \lambda_n \cdot y'_n)$,
  - $y = (y_1, \ldots, y_n) \sim y' = (y'_1, \ldots, y'_n)$ implies $(\lambda_1 \cdot y_1, \ldots, \lambda_n \cdot y_n) \sim (\lambda_1 \cdot y'_1, \ldots, \lambda_n \cdot y'_n)$.

## 37. Formal Description

- By a *total pre-ordering relation* on a set $Y$, we mean
  - a pair of a transitive relation $\succ$ and an equivalence relation $\sim$ for which,
  - for every $y, y' \in Y$, exactly one of the following relations hold: $y \succ y'$, $y' \succ y$, or $y \sim y'$.
- We say that a total pre-ordering is *non-trivial* if there exist $y$ and $y'$ for which $y \succ y'$.
- We say that a total pre-ordering relation on $(R^+)^n$ is:
  - *monotonic* if $y'_i > y_i$ for all $i$ implies $y' \succ y$;
  - *continuous* if
    - whenever we have a sequence $y^{(k)}$ of tuples for which $y^{(k)} \succeq y'$ for some tuple $y'$, and
    - the sequence $y^{(k)}$ tends to a limit $y$,
    - then $y \succeq y'$.

## 38. Main Result

**Theorem.** *Every non-trivial monotonic scale-inv. continuous total pre-ordering relation on* $(R^+)^n$ *has the form:*

$$y' = (y'_1, \ldots, y'_n) \succ y = (y_1, \ldots, y_n) \Leftrightarrow \prod_{i=1}^{n} (y'_i)^{\alpha_i} > \prod_{i=1}^{n} y_i^{\alpha_i};$$

$$y' = (y'_1, \ldots, y'_n) \sim y = (y_1, \ldots, y_n) \Leftrightarrow \prod_{i=1}^{n} (y'_i)^{\alpha_i} = \prod_{i=1}^{n} y_i^{\alpha_i},$$

*for some constants* $\alpha_i > 0$.

*Comment:* Vice versa,

- for each set of values $\alpha_1 > 0, \ldots, \alpha_n > 0$,
- the above formulas define a monotonic scale-invariant continuous pre-ordering relation on $(R^+)^n$.

## 39. Practical Conclusion

- *Situation:*
  - we need to select an alternative;
  - each alternative is characterized by characteristics $y_1, \ldots, y_n$.
- *Traditional approach:*
  - we assign the weights $w_i$ to different characteristics;
  - we select the alternative with the largest value of $\sum\limits_{i=1}^n w_i \cdot y_i$.
- *New result:* it is better to select an alternative with the largest value of $\prod\limits_{i=1}^n y_i^{w_i}$.
- *Equivalent reformulation:* select an alternative with the largest value of $\sum\limits_{i=1}^n w_i \cdot \ln(y_i)$.

40. Acknowledgment

This work was supported in part:

- by the National Science Foundation grants:
  - HRD-0734825 (Cyber-ShARE Center of Excellence) and
  - DUE-0926721,
- and by Grant 1 T36 GM078000-01 from the National Institutes of Health.

# 41. Pt. 2: Case of Probabilistic Prior Knowledge

- *Description:* from prior observations, we know $\widetilde{s}_j \approx s_j$, and we know the st. dev. $\sigma_j$ of this value.
- *Minimize:* $\sum\limits_{j,j'} (\Delta s_{ij} - \Delta s_{ij'})^2$ s.t. $\sum\limits_{j=1}^{c} \ell_{ij} \cdot \Delta s_{ij} = \Delta t_i$ and

$$\frac{1}{n} \cdot \sum_{j=1}^{c} \frac{((s_j^{(k)} + \Delta s_{ij}) - \widetilde{s}_j)^2}{\sigma_j^2} = 1.$$

- *Solution* (Lagrange multipliers): $\overline{\Delta s} \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum\limits_{j=1}^{c} \Delta s_{ij}$,

$$\frac{2}{n} \cdot \Delta s_{ij} - \frac{2}{n} \cdot \overline{\Delta s} + \lambda \cdot \ell_{ij} + \frac{2\mu}{n \cdot \sigma_j^2} \cdot (s_j^{(k)} + \Delta s_{ij} - \widetilde{s}_j) = 0.$$

- *Fact:* $\Delta s_{ij}$ is an explicit function of $\lambda$, $\mu$, $\overline{\Delta s}$.
- *Algorithm:* solve 3 non-linear equations (above one + 2 constraints) with unknowns $\lambda$, $\mu$, $\overline{\Delta s}$.

# 42. Combination of Different Types of Prior Knowledge

- *Need:* we often have both:
  - prior measurement results – i.e., *probabilistic* knowledge, and
  - expert estimates – i.e., *interval* and *fuzzy* knowledge.
- *Minimize:* $\sum\limits_{j,j'}(\Delta s_{ij}-\Delta s_{ij'})^2$ s.t. $\sum\limits_{j=1}^{c}\ell_{ij}\cdot\Delta s_{ij}=\Delta t_i$,

$$\frac{1}{n}\cdot\sum_{j=1}^{c}\frac{((s_j^{(k)}+\Delta s_{ij})-\widetilde{s}_j)^2}{\sigma_j^2}\le 1,$$

and $\underline{s}_j\le s_j^{(k)}+\Delta s_{ij}\le\overline{s}_j$.

- *Idea:* we minimize a convex function under convex constraints; efficient algorithms are known.

# 43. Combination of Different Types of Prior Knowledge: Algorithm

- *Idea* – method of alternating projections:
  - first, add a correction that satisfy the first constraint,
  - then, the additional correction that satisfies the second constraint,
  - etc.
- *Specifics:*
  - first, add equal values $\Delta s_{ij}$ to minimize $\Delta t_i$;
  - restrict the values to the nearest points from $[\underline{s}_j, \overline{s}_j]$,
  - find the extra corrections that satisfy the probabilistic constraint,
  - repeat until converges.

## 44. Part 3, Proof: Part 1

- Due to scale-invariance, for every $y_1, \ldots, y_n, y'_1, \ldots, y'_n$, we can take $\lambda_i = \dfrac{1}{y_i}$ and conclude that

$$(y'_1, \ldots, y'_n) \sim (y_1, \ldots, y_n) \Leftrightarrow \left(\frac{y'_1}{y_1}, \ldots, \frac{y'_n}{y_n}\right) \sim (1, \ldots, 1).$$

- Thus, to describe the equivalence relation $\sim$, it is sufficient to describe $\{z = (z_1, \ldots, z_n) : z \sim (1, \ldots, 1)\}$.
- Similarly,

$$(y'_1, \ldots, y'_n) \succ (y_1, \ldots, y_n) \Leftrightarrow \left(\frac{y'_1}{y_1}, \ldots, \frac{y'_n}{y_n}\right) \succ (1, \ldots, 1).$$

- Thus, to describe the ordering relation $\succ$, it is sufficient to describe the set $\{z = (z_1, \ldots, z_n) : z \succ (1, \ldots, 1)\}$.
- Similarly, it is also sufficient to describe the set

$$\{z = (z_1, \ldots, z_n) : (1, \ldots, 1) \succ z\}.$$

## 45. Proof: Part 2

- *To simplify:* take logarithms $Y_i = \ln(y_i)$, and sets

$$S_\sim = \{Z : z = (\exp(Z_1), \ldots, \exp(Z_n)) \sim (1, \ldots, 1)\},$$
$$S_\succ = \{Z : z = (\exp(Z_1), \ldots, \exp(Z_n)) \succ (1, \ldots, 1)\};$$
$$S_\prec = \{Z : (1, \ldots, 1) \succ z = (\exp(Z_1), \ldots, \exp(Z_n))\}.$$

- Since the pre-ordering relation is total, for $Z$, either $Z \in S_\sim$ or $Z \in S_\succ$ or $Z \in S_\prec$.
- *Lemma:* $S_\sim$ is closed under addition:
  - $Z \in S_\sim$ means $(\exp(Z_1), \ldots, \exp(Z_n)) \sim (1, \ldots, 1)$;
  - due to scale-invariance, we have

$$(\exp(Z_1+Z_1'), \ldots) = (\exp(Z_1)\cdot\exp(Z_1'), \ldots) \sim (\exp(Z_1'), \ldots);$$

  - also, $Z' \in S_\sim$ means $(\exp(Z_1'), \ldots) \sim (1, \ldots, 1)$;
  - since $\sim$ is transitive,

$$(\exp(Z_1 + Z_1'), \ldots) \sim (1, \ldots) \text{ so } Z + Z' \in S_\sim.$$

# 46. Proof: Part 3

- *Reminder:* the set $S_\sim$ is closed under addition;
- Similarly, $S_\prec$ and $S_\succ$ are closed under addition.
- *Conclusion:* for every integer $q > 0$:
  - if $Z \in S_\sim$, then $q \cdot Z \in S_\sim$;
  - if $Z \in S_\succ$, then $q \cdot Z \in S_\succ$;
  - if $Z \in S_\prec$, then $q \cdot Z \in S_\prec$.
- Thus, if $Z \in S_\sim$ and $q \in N$, then $(1/q) \cdot Z \in S_\sim$.
- We can also prove that $S_\sim$ is closed under $Z \to -Z$:
  - $Z = (Z_1, \ldots) \in S_\sim$ means $(\exp(Z_1), \ldots) \sim (1, \ldots)$;
  - by scale invariance, $(1, \ldots) \sim (\exp(-Z_1), \ldots)$, i.e., $-Z \in S_\sim$.
- Similarly, $Z \in S_\succ \Leftrightarrow -Z \in S_\prec$.
- So $Z \in S_\sim \Rightarrow (p/q) \cdot Z \in S_\sim$; in the limit, $x \cdot Z \in S_\sim$.

## 47. Proof: Final Part

- *Reminder:* $S_\sim$ is closed under addition and multiplication by a scalar, so it is a linear space.
- *Fact:* $S_\sim$ cannot have full dimension $n$, since then all alternatives will be equivalent to each other.
- *Fact:* $S_\sim$ cannot have dimension $< n-1$, since then:
  - we can select an arbitrary $Z \in S_\prec$;
  - connect it w/$-Z \in S_\succ$ by a path $\gamma$ that avoids $S_\sim$;
  - due to closeness, $\exists \gamma(t^*)$ in the limit of $S_\succ$ and $S_\prec$;
  - thus, $\gamma(t^*) \in S_\sim$ – a contradiction.
- Every $(n-1)$-dim lin. space has the form $\sum\limits_{i=1}^{n} \alpha_i \cdot Y_i = 0$.
- Thus, $Y \in S_\succ \Leftrightarrow \sum \alpha_i \cdot Y_i > 0$, and

$$y \succ y' \Leftrightarrow \sum \alpha_i \cdot \ln(y_i/y'_i) > 0 \Leftrightarrow \prod y_i^{\alpha_i} > \prod {y'_i}^{\alpha_i}.$$