Combining Interval, Probabilistic, and Other Types of Uncertainty in Engineering Applications

Andrew Pownuk

Computational Science Program
University of Texas at El Paso
El Paso, Texas 79968, USA
ampownuk@utep.edu

Part I
Introduction

1. Need for Data Processing

- One of the main objectives of science is to predict future values y of physical quantities:
  - in meteorology, we need to predict future weather;
  - in airplane control, we need to predict the location and the velocity of the plane under current control.
- To make this prediction:
  - we need to know the relation $y = f(x_1, \ldots, x_n)$ between y and related quantities $x_1, \ldots, x_n$;
  - then, we measure or estimate $x_1, \ldots, x_n$;
  - finally, we use the results $\widetilde{x}_i$ of measurement (or estimation) to compute an estimate

$$\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n).$$

- This computation of $\widetilde{y}$ is an important case of *data processing*.

2. Need to Take Uncertainty Into Account When Processing Data

- Measurement are never absolutely accurate: in general,

$$\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i \neq 0.$$

- As a result, the estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is, in general, different from the ideal value $y = f(x_1, \ldots, x_n)$.
- To estimate the accuracy $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$, we need to have some information about the measurement errors Δx_i.
- Traditional engineering approach assumes that we know the probability distribution of each Δx_i.
- Often, $\Delta x_i \sim N(0, \sigma_i)$, and different Δx_i are assumed to be independent.
- In such situations, our goal is to find the probability distribution for Δy.

3. Cases of Interval and Fuzzy Uncertainty

- Often, we only know the upper bound Δ_i: $|\Delta x_i| \le \Delta_i$.
- Then, the only information about the x_i is that

$$x_i \in \mathbf{x}_i \stackrel{\text{def}}{=} [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i].$$

- Different $x_i \in \mathbf{x}_i$ lead, in general, to different

$$y = f(x_1, \ldots, x_n).$$

- We want to find the range $\mathbf{y}$ of possible values of y:

$$\mathbf{y} = \{f(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

- To gauge the accuracy of expert estimates, it is reasonable to use fuzzy techniques, i.e., to describe:
  - for each possible value x_i,
  - the degree $\mu_i(x_i)$ to which x_i is possible.

4. Measurement and Estimation Inaccuracies Are Usually Small

- In many practical situations, the measurement and estimation inaccuracies Δx_i are relatively small.
- Then, we can safely ignore terms which are quadratic (or of higher order) in terms of Δx_i:

$$\Delta y = \widetilde{y} - y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n) =$$

$$\sum_{i=1}^{n} c_i \cdot \Delta x_i, \text{ where } c_i = \frac{\partial f}{\partial x_i}.$$

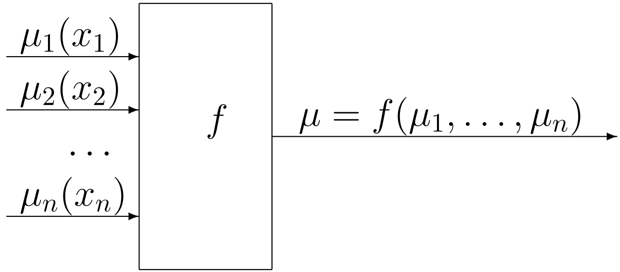
- If needed, the derivative can be estimated by numerical differentiation

$$c_i \approx \frac{f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - \widetilde{y}}{h}.$$

5. Case of Interval Uncertainty

- Let us consider the case when $\Delta y = \sum_{i=1}^{n} c_i \cdot \Delta x_i$.
- In this case, $\mathbf{y} = [\widetilde{y} - \Delta, \widetilde{y} + \Delta]$, where $\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i$.
- Sometimes, we have explicit expressions or efficient algorithms for the partial derivatives c_i.
- Often, however, we use proprietary software in our computations.
- Then, we cannot use differentiation formulas, but we can use numerical differentiation.
- *Problem:* We need $n + 1$ calls to f, to compute $\widetilde{y}$ and n values c_i.
- When f is time-consuming and n is large, we can use a Cauchy-based Monte-Carlo method.

6. Fuzzy Case: A Problem



- *Given:* an algorithm $y = f(x_1, \dots, x_n)$ and n fuzzy numbers $\mu_i(x_i)$.
- *Compute:* $\mu(y) = \max\limits_{x_1,\dots,x_n : f(x_1,\dots,x_n)=y} \min(\mu_1(x_1), \dots, \mu_n(x_n))$.
- *Motivation:* y is a possible value of $Y \leftrightarrow \exists x_1, \dots, x_n$ s.t. each x_i is a possible value of X_i and $f(x_1, \dots, x_n) = y$.
- *Details:* "and" is min, $\exists$ ("or") is max, hence

$$\mu(y) = \max_{x_1,\dots,x_n} \min(\mu_1(x_1), \dots, \mu_n(x_n), t(f(x_1, \dots, x_n) = y)),$$

where $t(\text{true}) = 1$ and $t(\text{false}) = 0$.

7. Fuzzy Case: Reduction to Interval Computations

- *Given:* an algorithm $y = f(x_1, \ldots, x_n)$ and n fuzzy numbers X_i described by membership functions $\mu_i(x_i)$.
- *Compute:* $Y = f(X_1, \ldots, X_n)$, where Y is defined by Zadeh's extension principle:

$$\mu(y) = \max_{x_1, \ldots, x_n : f(x_1, \ldots, x_n) = y} \min(\mu_1(x_1), \ldots, \mu_n(x_n)).$$

- *Idea:* represent each X_i by its α-cuts

$$X_i(\alpha) = \{x_i : \mu_i(x_i) \geq \alpha\}.$$

- *Advantage:* for continuous f, for every α, we have

$$Y(\alpha) = f(X_1(\alpha), \ldots, X_n(\alpha)).$$

- *Resulting algorithm:* for $\alpha = 0, 0.1, 0.2, \ldots, 1$ apply interval computations techniques to compute $Y(\alpha)$.

8. Open Problems

- In engineering applications, we want methods for estimating uncertainty which are:
  - *accurate* – this is most important in most engineering applications;
  - *fast*: this is important in some engineering applications where we need real-time computations,
  - *understandable* to engineers – otherwise, engineers will be reluctant to use them, and
  - sufficiently *general* – so that they can be applied in all kinds of situations.
- It is thus desirable to design more accurate, faster, more understandable, and/or more general methods.
- We also need to make decisions under uncertainty.

9. What We Do in This Dissertation

- First, we show how to make the current methods more accurate.
- Then, we show how to make these methods faster.
- After that, we show how to make these methods more understandable to engineers.
- Third, we analyze how to make these methods more general.
- Finally, we analyze how to make decisions under uncertainty.
- In this presentation, we will focus mainly on the new results, obtained after the Master's thesis.

Part II
How to Get More Accurate Estimates

10. General Structure of This Part

2.1 We show that interval estimates can lead to better accuracy than the traditional statistical ones.

2.2-3 An even higher accuracy can be obtained if we take into account both probabilistic and interval uncertainty.

2.2 This is what we show on the example of providing the value which best represents a given sample.

2.3 We have also shown this in the thesis, on the example of taking model inaccuracy into account.

2.4 Finally, we analyze accuracy of fuzzy-based estimates.

- In this talk, we cover only one section from each chapter.

2.1. In System Identification, Interval Estimates Can Lead to Much Better Accuracy than the Traditional Statistical Ones: General Algorithm and Case Study

11. System Identification: A General Problem

- Often, we are interested in a quantity y which is difficult (or even impossible) to measure directly.
- This difficulty and/or impossibility may be technical:
  - while we can directly measure the distance between the two buildings by simply walking there,
  - there is no easy way to measure the distance to a nearby star by flying there.
- Impossibility may come from predictions – today, we cannot measure tomorrow's temperature.

12. System Identification (cont-d)

- A natural idea is to find easier-to-measure quantities $x_1, \ldots, x_n$ that are related to y by a known dependence

$$y = f(x_1, \ldots, x_n).$$

- Then, we can use the results $\widetilde{x}_i$ of measuring these auxiliary quantities to estimate y as $\widetilde{y} \stackrel{\text{def}}{=} f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.
- *Example:* we can find the distance to a nearby star by measuring the direction to this star in two seasons:
  - when the Earth is at different sides of the Sun, and
  - the angle is thus slightly different.
- To predict tomorrow's temperature T:
  - we can measure the temperature and wind speed and direction at different locations today, and
  - use this data to predict T.

13. System Identification (final)

- In some cases, we know the dependence

$$y = f(x_1, \ldots, x_n).$$

- In other cases, we only know the general form of this dependence

$$y = f(a_1, \ldots, a_m, x_1, \ldots, x_n).$$

- The values a_i must be estimated based on measurement results.
- We have the results $\widetilde{y}_k$ and $\widetilde{x}_{ki}$ of measuring y and x_i in several situations $k = 1, \ldots, K$.
- Estimating a_i is called *system identification.*

14. Need to Take Measurement Uncertainty into Account

- Measurements are not 100% accurate.
- In general, the measurement result $\widetilde{x}$ is different from the actual (unknown) value x: $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x \neq 0$; thus,
  - while for the (unknown) actual values y_k and x_{ki}, we have $y_k = f(a_1, \ldots, a_m, x_{k1}, \ldots, x_{kn})$,
  - the relation between measurement results $\widetilde{y}_k \approx y_k$ and $\widetilde{x}_{ki} \approx x_{ki}$ is approximate:

$$\widetilde{y}_k \approx f(a_1, \ldots, a_m, \widetilde{x}_{k1}, \ldots, \widetilde{x}_{kn}).$$

- It is therefore important to take this uncertainty into account when estimating the values $a_1, \ldots, a_m$.

15. How Can We Describe Uncertainty?

- In all the cases, we should know the bound Δ on the absolute value of the measurement error: $|\Delta x| \le \Delta$.
- This means that only values Δx from the interval $[-\Delta, \Delta]$ are possible.
- If this is the only information we have then:
  - based on the measurement result $\widetilde{x}$,
  - the only information that we have about the actual value x is that $x \in [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.
- Processing data under such interval uncertainty is known as *interval computations*.

16. How Can We Describe Uncertainty (cont-d)

- Ideally, it is also desirable to know how frequent are different values Δx within this interval.
- In other words, it is desirable to know the probabilities of different values $\Delta x \in [-\Delta, \Delta]$.
- The measurement uncertainty Δx often comes from many different independent sources.
- Thus, due to the Central Limit Theorem, the distribution of Δx is close to Gaussian.
- This explains the usual engineering practice of using normal distributions.

17. Two Approximations, Two Options

- Gaussian distribution is that it is *not* located on any interval.
- The probability of measurement error Δx to be in any interval – no matter how far away from Δ – is non-zero.
- From this viewpoint, the assumption that the distribution is Gaussian is an approximation.
- It seems like a very good approximation, since for normal distribution with mean 0 and st. dev. σ:
  - the probability to be outside the 3σ interval $[-3\sigma, 3\sigma]$ is very small, approximately 0.1%, and
  - the probability for it to be outside the 6σ interval is about 10^{-8}, practically negligible.
- Since the difference is small, this should not affect system identification.

18. Two Approximations, Two Options (cont-d)

- At first glance, if we keep the bounds but ignore probabilities, we will do much worse.
- Our results show that the opposite is true:
  - if we ignore the probabilistic information and use only interval (or fuzzy) information,
  - we get much more accurate estimates for a_j than in the statistical case.
- This is not fully surprising: theory shows that asymptotically, interval bounds are better.
- However, the drastic improvement in accuracy was somewhat unexpected.

19. System Identification: Interval Case

- For each pattern $k = 1, \ldots, K$:
  - we know the measurement results $\widetilde{y}_k$ and $\widetilde{x}_{ki}$, and
  - we know the accuracies Δ_k and Δ_{ki} of the corresponding measurements.
- Thus, we know that:
  - the actual (unknown) value y_k belongs to the interval $[\underline{y}_k, \overline{y}_k] = [\widetilde{y}_k - \Delta_k, \widetilde{y}_k + \Delta_k]$; and
  - the actual (unknown) value x_{ki} belongs to the interval $[\underline{x}_{ki}, \overline{x}_{ki}] = [\widetilde{x}_{ki} - \Delta_{ki}, \widetilde{x}_{ki} + \Delta_{ki}]$.
- We need to find $a_1, \ldots, a_m$ for which, for every k, for some $x_{ki} \in [\underline{x}_{ki}, \overline{x}_{ki}]$,

$$f(a_1, \ldots, a_m, x_{k1}, \ldots, x_{kn}) \in [\underline{y}_k, \overline{y}_k].$$

- Specifically, for each j from 1 to m, we would like to find the range $[\underline{a}_j, \overline{a}_j]$ of all possible values of a_j.

20. Analysis of the Problem

- In the statistical case, we use the Least Squares method and find $\widetilde{a}_1, \ldots, \widetilde{a}_m$ that minimize the sum:

$$\sum_{k=1}^{K} (\widetilde{y}_k - f(a_1, \ldots, a_m, \widetilde{x}_{k1}, \ldots, \widetilde{x}_{kn}))^2 \to \min_{a_1, \ldots, a_m}.$$

- The measurement errors Δx_{ki} are usually small.
- Thus, the differences $\Delta a_j = \widetilde{a}_j - a_j$ are also small.
- We can keep only linear terms in the Taylor expansion: $Y_k = y_k - \sum\limits_{j=1}^{m} b_j \cdot \Delta a_j - \sum\limits_{i=1}^{n} b_{ki} \cdot \Delta x_{ki}$, where:

$$Y_k = f(\widetilde{a}_1, \ldots, \widetilde{a}_m, \widetilde{x}_{k1}, \ldots, \widetilde{x}_{kn}), \;\; b_j = \frac{\partial f}{\partial a_j}, \;\; b_{jk} = \frac{\partial f}{\partial x_{ki}}.$$

21. Analysis of the Problem

- For each Δa_j, the min and max values of Y_k are:

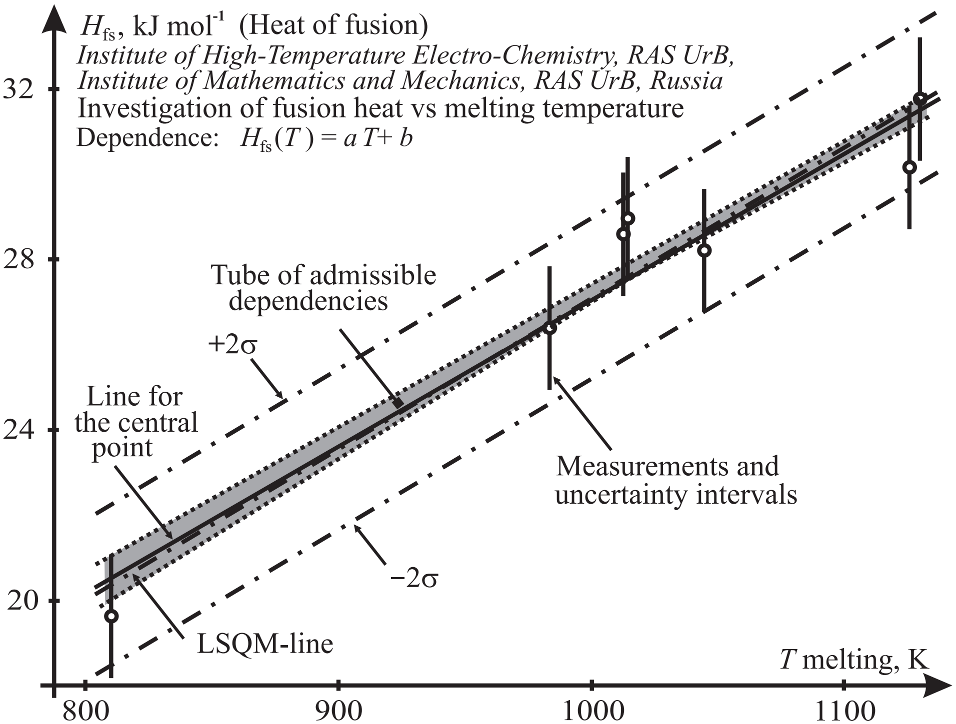
$$\underline{Y}_k = Y_k - \sum_{j=1}^{m} b_j \cdot \Delta a_j - \sum_{i=1}^{n} |b_{ki}| \cdot \Delta_{ki};$$

$$\overline{Y}_k = Y_k - \sum_{j=1}^{m} b_j \cdot \Delta a_j + \sum_{i=1}^{n} |b_{ki}| \cdot \Delta_{ki}.$$

- We want some values $Y_k \in [\underline{Y}_k, \overline{Y}_k]$ to be in $[\underline{y}_k, \overline{y}_k]$, i.e., that $[\underline{Y}_k, \overline{Y}_k] \cap [\underline{y}_k, \overline{y}_k] \neq \emptyset$.
- This is equivalent to $\underline{y}_k \leq \overline{Y}_k$ and $\underline{Y}_k \leq \overline{y}_k$.
- Thus, we need to optimize a linear expression under linear inequalities.
- For such *linear programming* (LP) problems, there are efficient algorithms.

22. Case Study

- One of the important engineering problems is the problem of storing energy:
  - solar power and wind turbines provide access to large amounts of renewable energy,
  - but this energy is not always available – the sun goes down, the wind dies,
  - and storing it is difficult.
- Similarly, electric cars are clean, but we spend a lot of weight on the batteries.
- We want batteries with high energy density.
- One of the most promising directions is using molten salt batteries, including liquid metal batteries.
- Melting energy E linearly depends on temperature T: $E = a \cdot T + b$. What are a and b?



H_{fs}, kJ mol^{-1} (Heat of fusion)
Institute of High-Temperature Electro-Chemistry, RAS UrB,
Institute of Mathematics and Mechanics, RAS UrB, Russia
Investigation of fusion heat vs melting temperature
Dependence: $H_{fs}(T) = a\,T + b$
32
28
24
20
Tube of admissible dependencies
+2σ
Line for the central point
Measurements and uncertainty intervals
−2σ
LSQM-line
T melting, K
800
900
1000
1100

Figure 1:

23. Results of Our Analysis

- We generated two different bounds on y:
  - bounds based on interval estimates, and
  - 2σ-bounds coming from the traditional statistical analysis.
- It turned out that the interval results are an order of magnitude smaller that the statistical ones.
- A similar improvement was observed in other applications ranging from catalysis and to mechanics.

24. System Identification: Conclusions

- Traditional engineering techniques assume that the measurement errors are normally distributed.
- In practice, the distribution of measurement errors is indeed often close to normal.
- Often, however, we also have an additional information about measurement uncertainty.
- Namely, we also know the upper bounds Δ on the corresponding measurement errors.
- Based on the measurement result $\widetilde{x}$, the actual value x is in the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.
- We can use interval computations techniques to estimate the accuracy of the result of data processing.
- Example: for linear models, we can use linear programming techniques to compute the corr. bounds.

25. System Identification: Conclusions (cont-d)

- Which approaches leads to more accurate estimates:
  - the traditional approach, when we ignore the upper bounds and only consider the probabilities, or
  - the interval approach, we only take into account the bounds and ignore probabilities?
- When the number of measurements n increases, the interval estimates become more accurate.
- We show that interval techniques indeed lead to much more accurate estimates.
- So, we recommend to try interval techniques: they may lead to more accurate estimates.

Part III
How to Speed Up Computations

26. General Structure of This Part

3.1 In the thesis, we showed how to speed up fuzzy computations for the min t-norm.

3.2 We show how to extend this result beyond min t-norm.

3.3 We describe a speedup for the case of probabilistic uncertainty.

3.4 We also speculate on the possibility of use quantum computing to further speed up data processing.

3.3. How to Speed Up Processing of Probabilistic Data

27. Formulation of the Problem

- We know the probability density functions (pdfs) $\rho_i(\Delta x_i)$ of measurement errors.
- We want to find the pdf for $\Delta y = \sum\limits_{i=1}^{n} c_i \cdot \Delta x_i$.
- It is easy to find pdf of $t_i = c_i \cdot \Delta x_i$.
- To find pdf for $t_1 + t_2 + \ldots$, it is sufficient to compute pdf for $t = t_1 + t_2$: then we compute $t + t_3$, etc.
- We have $\rho(t) = \int \rho_1(t_1) \cdot \rho_2(t - t_1)\, dt_1$.
- Straightforward computation means computing N sums with N terms each: $O(N^2)$ steps.
- A faster computation uses FFT F:
$$\rho = F^{-1}(\chi_1(\omega) \cdot \chi_2(\omega)), \text{ where } \chi_i = F(\rho_i).$$
- For large N, the time N needed for point-wise multiplication is still huge; how can we make it faster?

28. Processing Is Faster for Normal Distributions

- In this case, it is enough to compute mean $\mu = \mu_1 + \mu_2$ and variance $V = V_1 + V_2$.
- We need 2 computational steps instead of $O(N)$.
- Same holds for any *infinitely divisible* distribution $D(\alpha)$, with char. f-n $\chi(\omega) = \exp(\mathrm{i} \cdot \mu \cdot \omega - A \cdot |\omega|^\alpha)$.
- For example, for $\alpha = 1$, we get Cauchy distribution

$$\rho(x) = \frac{1}{\pi \cdot \Delta} \cdot \frac{1}{1 + \frac{(x-\mu)^2}{\Delta^2}}.$$

- In general, $\mu = \mu_1 + \mu_2$ and $A = A_1 + A_2$.

29. Idea

- For some $\alpha_1 < \ldots < \alpha_k$, approximate each t_i by a sum $t_{a,i} = t_{i1} + \ldots + t_{ij} + \ldots + t_{ik}$, where $t_{ij} \sim D(\alpha_j)$, so:
$$\chi_{ij}(\omega) = \exp(\mathrm{i} \cdot \mu_{ij} \cdot \omega - A_{ij} \cdot |\omega|^{\alpha_j}), \text{ and}$$
$$\chi_{a,i}(\omega) = \prod_{j=1}^{k} \chi_{ij}(\omega) = \exp\left(\mathrm{i} \cdot \mu_i \cdot \omega - \sum_{j=1}^{k} A_{ij} \cdot |\omega|^{\alpha_j}\right).$$
- Then, $\chi(\omega) \approx \chi_a(\omega) \stackrel{\text{def}}{=} \chi_{a,2}(\omega) \cdot \chi_{a,2}(\omega) =$
$$\exp\left(\mathrm{i} \cdot \mu \cdot \omega - \sum_{j=1}^{k} A_j \cdot |\omega|^{\alpha_j}\right).$$
- Here, $\mu = \mu_1 + \mu_2$ and $A_j = A_{1j} + A_{2j}$.
- So, we need $k + 1$ arithmetic operations instead of N:
  - one addition to compute μ and
  - k additions to compute k values $A_1, \ldots, A_k$.

30. How Good Is This Algorithm: Example

- t_1 is normally distributed with 0 mean and standard deviation 1.
- As t_2, let us take the Laplace distribution, with probability density $\rho_2(t) = \dfrac{1}{2} \cdot \exp(-|t|)$.
- To approximate both distributions, we used $k = 3$, with $\alpha_1 = 1$, $\alpha_2 = 1.5$, and $\alpha_3 = 2$; then:

$$\mu_1 = 0, \quad A_{11} = A_{12} = 0, \quad \text{and} \quad A_{13} = \frac{1}{2};$$

$$\mu_2 = 0, \; A_{21} = -0.162, \; A_{22} = 1.237, \; \text{and} \; A_{23} = -0.398.$$

- Thus, for $t = t_1 + t_2$ we get:

$$\mu = 0, \quad A_1 = -0.162, \quad A_2 = 1.237, \quad \text{and} \quad A_3 = 0.102.$$

- The mean square approximation error is 1%.

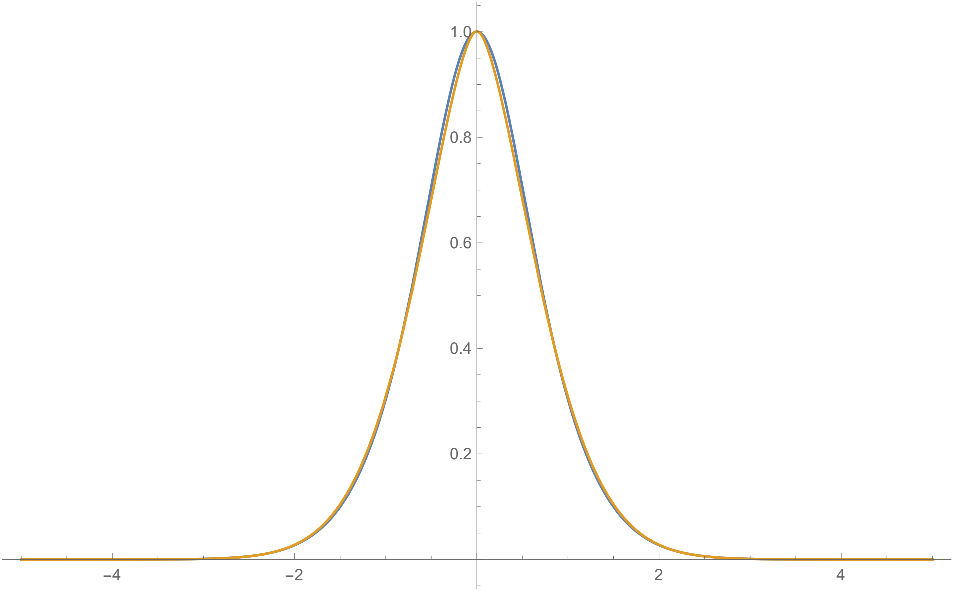


Figure 2: How good is the proposed approximation

Part IV

Towards a Better Understandability of Uncertainty-Estimating Algorithms

31. Structure of This Part

4.1 We provide an intuitive explanation for different types of interval solutions.

4.2 In the thesis, we explained seemingly counter-intuitive interval ideas.

4.3 For probabilistic uncertainty, we explain why it is reasonable to consider mixtures.

4.4 We explain seemingly non-traditional fuzzy ideas.

4.5 We explain the ubiquity of linear dependencies.

4.6 We explain the consequences of deviation from linearity.

4.5. General Case: Explaining Ubiquity of Linear Models

32. Need for Interpolation

- In many practical situations:
  - we know that the value of a quantity y is uniquely determined by the value of some other quantity x,
  - but we do not know the exact form of the corresponding dependence $y = f(x)$.
- To find this dependence, we measure the values of x and y in different situations.
- As a result, we get the values $y_i = f(x_i)$ of the unknown function $f(x)$ for several values $x_1, \ldots, x_n$.
- Based on this information, we would like to predict the value $f(x)$ for all other values x.
- When x is between the smallest and the largest of the values x_i, this prediction is known as the *interpolation*.

33. Why Linear Interpolation?

- Let's consider the case $n = 2$. Let's assume that $f(x)$ is linear on $[x_1, x_2]$; then
$$f(x) = \frac{x - x_1}{x_2 - x_1} \cdot f(x_2) + \frac{x_2 - x}{x_2 - x_1} \cdot f(x_1).$$
- This formula is known as *linear interpolation.*
- The usual motivation for linear interpolation is simplicity: linear functions are the easiest to compute.
- An interesting empirical fact is that in many practical situations, linear interpolation works reasonably well.
- We know that in computational science, often very complex computations are needed.
- So we cannot claim that nature prefers simplicity.
- There should be another reason for the empirical fact that linear interpolation often works well.

34. Reasonable Properties of Interpolation

- We want to be able,
  - given values y_1 and y_2 of the unknown function at points x_1 and x_2, and a point $x \in (x_1, x_2)$,
  - to provide an estimate for $f(x)$.
- Let us denote this estimate by $I(x_1, y_1, x_2, y_2, x)$; what are the reasonable properties of this function?
- If $y_i = f(x_i) \le y$ for both i, it is reasonable to expect that $f(x) \le y$.
- In particular, for $y = \max(y_1, y_2)$, we conclude that

$$I(x_1, y_1, x_2, y_2, x) \le \max(y_1, y_2).$$

- Similarly, if $y \le y_i$ for both i, it is reasonable to expect that $y \le f(x)$.
- In particular, for $y = \min(y_1, y_2)$, we conclude that

$$\min(y_1, y_2) \le I(x_1, y_1, x_2, y_2, x).$$

35. x-Scale-Invariance

- The numerical value of a physical quantity depends:
  - on the choice of the measuring unit and
  - on the starting point.
- If we change the starting point to the one which is b units smaller, then b is added to all the values.
- If we replace a measuring unit by a $a > 0$ times smaller one, then all the values are multiplied by a.
- If we perform both changes, then each original value x is replaced by the new value $x' = a \cdot x + b$.
- For example, if we know the temperature x in C, then the temperature x' in F is $x' = 1.8 \cdot x + 32$.
- The interpolation procedure should not change if we simply re-scale:

  $$I(a \cdot x_1 + b, y_1, a \cdot x_2 + b, y_2, a \cdot x + b) = I(x_1, y_1, x_2, y_2, x).$$

36. y-Scale-Invariance

- Similarly, we can consider different units for y.
- The interpolation result should not change if we simply change the starting point and the measuring unit; so:
  - if we replace y_1 with $a \cdot y_1 + b$ and y_2 with $a \cdot y_2 + b$,
  - then the result of interpolation should be obtained by a similar transformation from the previous one:

$$I(x_1, a \cdot y_1 + b, x_2, a \cdot y_2 + b, x) = a \cdot I(x_1, y_1, x_2, y_2, x) + b.$$

37. Consistency

- When $x_1 \le x'_1 \le x \le x'_2 \le x_2$, the value $f(x)$ can be estimated in two different ways.
- We can interpolate directly from the values $y_1 = f(x_1)$ and $y_2 = f(x_2)$, getting $I(x_1, y_1, x_2, y_2, x)$.
- Or we can:
  - first estimate the values $f(x'_1) = I(x_1, y_1, x_2, y_2, x'_1)$ and $f(x'_2) = I(x_1, y_1, x_2, y_2, x'_2)$, and
  - then use these two estimates to estimate $f(x)$ as

$$I(x_1, f(x'_1), x_2, f(x'_2), x) =$$

$$I(x'_1, I(x_1, y_1, x_2, y_2, x'_1), x'_2, I(x_1, y_1, x_2, y_2, x'_2), x).$$

- It is reasonable to require that these two ways lead to the same estimate for $f(x)$: $I(x_1, y_1, x_2, y_2, x) =$

$$I(x'_1, I(x_1, y_1, x_2, y_2, x'_1), x'_2, I(x_1, y_1, x_2, y_2, x'_2), x).$$

38. Continuity

- Most physical dependencies are continuous.
- Thus, when the two value x and x' are close, we expect the estimates for $f(x)$ and $f(x')$ to be also close.
- Thus, it is reasonable to require that:
  - the interpolation function $I(x_1, y_1, x_2, y_2, x)$ is continuous in x, and
  - that for both $i = 1, 2$, $I(x_1, y_1, x_2, y_2, x)$ converges to $f(x_i)$ when $x \to x_i$.

39. Resulting Definition

A function $I(x_1, y_1, x_2, y_2, x)$ defined for $x_1 < x < x_2$ is called an *interpolation function* if:

- $\min(y_1, y_2) \le I(x_1, y_1, x_2, y_2, x) \le \max(y_1, y_2)$;
- $I(a\cdot x_1 + b, y_1, a\cdot x_2 + b, y_2, a\cdot x + b) = I(x_1, y_1, x_2, y_2, x)$ for all x_i, y_i, x, $a > 0$, and b (x-scale-invariance);
- $I(x_1, a\cdot y_1 + b, x_2, a\cdot y_2 + b, x) = a\cdot I(x_1, y_1, x_2, y_2, x) + b$ for all x_i, y_i, x, $a > 0$, and b (y-scale invariance);
- consistency: $I(x_1, y_1, x_2, y_2, x) =$
$$I(x'_1, I(x_1, y_1, x_2, y_2, x'_1), x'_2, I(x_1, y_1, x_2, y_2, x'_2), x);$$
- continuity:
  - the expression $I(x_1, y_1, x_2, y_2, x)$ is a continuous function of x,
  - $I(x_1, y_1, x_2, y_2, x) \to y_1$ when $x \to x_1$ and $I(x_1, y_1, x_2, y_2, x) \to y_2$ when $x \to x_2$.

40. Main Result

- **Result:** *The only interpolation function satisfying all the properties is the linear interpolation*

$$I(x_1, y_1, x_2, y_2, x) = \frac{x - x_1}{x_2 - x_1} \cdot y_2 + \frac{x_2 - x}{x_2 - x_1} \cdot y_1.$$

- Thus, we have indeed explained that linear interpolation follows from the fundamental principles.
- This may explain its practical efficiency.

41. Proof of Linear Interpolation

- When $y_1 = y_2$, the conservativeness property implies that $I(x_1, y_1, x_2, y_1, x) = y_1$.
- Thus, to complete the proof, it is sufficient to consider two remaining cases: when $y_1 < y_2$ and when $y_2 < y_1$.
- We will consider the case when $y_1 < y_2$.
- The case when $y_2 < y_1$ is considered similarly.
- So, in the following text, without losing generality, we assume that $y_1 < y_2$.

42. Using y-Scale-Invariance

- When $y_1 < y_2$, then $y_1 = a \cdot 0 + b$ and $y_2 = a \cdot 1 + b$ for $a = y_2 - y_1$ and $b = y_1$.
- Thus, the y-scale-invariance implies that

$$I(x_1, y_1, x_2, y_2, x) = (y_2 - y_1) \cdot I(x_1, 0, x_2, 1, x) + y_1.$$

- If we denote $J(x_1, x_2, x) \stackrel{\text{def}}{=} I(x_1, 0, x_2, 1, x)$, then we get

$$I(x_1, y_1, x_2, y_2, x) = (y_2 - y_1) \cdot J(x_1, x_2, x) + y_1 =$$

$$J(x_1, x_2, x) \cdot y_2 + (1 - J(x_1, x_2, x)) \cdot y_1.$$

43. Using x-Scale-Invariance

- Since $x_1 < x_2$, we have $x_1 = a \cdot 0 + b$ and $x_2 = a \cdot 1 + b$, for $a = x_2 - x_1$ and $b = x_1$.
- Here, $x = a \cdot r + b$, where $r = \dfrac{x-b}{a} = \dfrac{x - x_1}{x_2 - x_1}$.
- Thus, the x-scale invariance implies that $J(x_1, x_2, x) = w\left(\dfrac{x - x_1}{x_2 - x_1}\right)$, where $w(r) \stackrel{\text{def}}{=} J(0,1,r)$.
- Thus, the above expression for $I(x_1, y_1, x_2, y_2, x)$ in terms of $J(x_1, x_2, x)$ takes the following simplified form:

$$w\left(\frac{x - x_1}{x_2 - x_1}\right) \cdot y_2 + \left(1 - w\left(\frac{x - x_1}{x_2 - x_1}\right) \cdot y_2\right) \cdot y_1.$$

- To complete our proof, we need to show that $w(r) = r$ for all $r \in (0, 1)$.

44. Using Consistency

- Let us take $x_1 = y_1 = 0$ and $x_2 = y_2 = 1$, then $I(0,0,1,1,x) = w(x)\cdot 1 + (1-w(x))\cdot 0 = w(x)$.
- For $x = 0.25 = \dfrac{0+0.5}{2}$, the value $w(0.25)$ can be obtained by interpolating $w(0) = 0$ and $\alpha \stackrel{\text{def}}{=} w(0.5)$:
$$w(0.25) = \alpha \cdot w(0.5) + (1-\alpha)\cdot w(0) = \alpha^2.$$
- For $x = 0.75 = \dfrac{0.5+1}{2}$, we similarly get:
$$w(0.75) = \alpha \cdot w(1) + (1-\alpha)\cdot w(0.5) = \alpha\cdot 1 + (1-\alpha)\cdot\alpha = 2\alpha - \alpha^2.$$
- $w(0.5)$ can be interpolated from $w(0.25)$ and $w(0.75)$:
$$w(0.5) = \alpha \cdot w(0.75) + (1-\alpha)\cdot w(0.25) =$$
$$\alpha\cdot(2\alpha-\alpha^2) + (1-\alpha)\cdot\alpha^2 = 3\alpha^2 - 2\alpha^3.$$
- By consistency, this estimate should be equal to our original estimate $w(0.5) = \alpha$: $3\alpha^2 - 2\alpha^3 = \alpha$.

45. What Is α

- Here, $\alpha = w(0.5) = 0$, $\alpha = 1$, or $\alpha = 0.5$.
- If $\alpha = 0$, then, $w(0.75) = \alpha \cdot w(1) + (1-\alpha) \cdot w(0.5) = 0$.
- By induction, we can show that $\forall n\, (w(1 - 2^{-n}) = 0)$ for each n.
- Here, $1 - 2^{-n} \to 1$, but $w(1 - 2^{-n}) \to 0$, which contradicts to continuity $w(1 - 2^{-n}) \to w(1) = 1$.
- Thus, $\alpha = 0$ is impossible.
- When $\alpha = w(0.5) = 1$, then
$$w(0.25) = \alpha \cdot w(0.5) + (1 - \alpha) \cdot w(0) = 1.$$
- By induction, $w(2^{-n}) = 1$ for each n.
- In this case, $2^{-n} \to 0$, but $w(2^{-n}) \to 1$, which contradicts to continuity $w(2^{-n}) \to w(0) = 0$.
- Thus, $\alpha = 0.5$.

Introduction
How to Get More . . .
How to Speed Up . . .
Towards a Better . . .
How General Can We . . .
Decision Making . . .
Conclusions
Appendix
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 56 of 214
Go Back
Full Screen
Close
Quit

46. Proof: Final Part

- For $\alpha = 0.5$: $w(0) = 0$, $w(0.5) = 0.5$, $w(1) = 1$.
- Let us prove, by induction over q, that for every binary-rational number $r = \dfrac{p}{2^q} \in [0, 1]$, we have $w(r) = r$.
- Indeed, the base case $q = 1$ is proven.
- Let us assume that we have proven it for $q = 1$.
- If p is even $p = 2k$, then $\dfrac{2k}{2^q} = \dfrac{k}{2^{q-1}}$, so the desired equality comes from the induction assumption.
- If $p = 2k + 1$, then $r = \dfrac{p}{2^q} = \dfrac{2k+1}{2^q} =$

$$0.5 \cdot \frac{2k}{2^q} + 0.5 \cdot \frac{2 \cdot (k+1)}{2^q} = 0.5 \cdot \frac{k}{2^{q-1}} + 0.5 \cdot \frac{k+1}{2^{q-1}}.$$

- So $w(r) = 0.5 \cdot w\left(\dfrac{k}{2^{q-1}}\right) + 0.5 \cdot w\left(\dfrac{k+1}{2^{q-1}}\right)$.

47. Proof: Final Part (cont-d)

- By induction assumption, we have
$$w\left(\frac{k}{2^{q-1}}\right) = \frac{k}{2^{q-1}} \text{ and } w\left(\frac{k+1}{2^{q-1}}\right) = \frac{k+1}{2^{q-1}}.$$
- Thus, $w(r) = \alpha \cdot \dfrac{k}{2^{q-1}} + 0.5 \cdot \dfrac{k+1}{2^{q-1}} = \dfrac{2k+1}{2^q} = r.$
- The equality $w(r) = r$ is hence true for all binary-rational numbers.
- Any real number x from the interval $[0,1]$ is a limit of such numbers – truncates of its binary expansion.
- Thus, by continuity, we have $w(x) = x$ for all x.
- Substituting $w(x) = x$ into the above formula for $I(x_1, y_1, x_2, y_2, x)$ leads to linear interpolation. Q.E.D.

4.6. General Case: Non-Linear Effects

48. Compaction Meter Value (CMV)

- It turns out that reasonably good estimates for stiffness can be obtained if we:
  - apply Fourier transform to the signal describing the dependence of acceleration on time, and then
  - evaluate *Compaction Meter Value* (CMV), a ratio A_2/A_1 between the amplitudes corr. to $2f$ and f.
- The relation between CMV and stiffness is an empirical fact.
- However, from the theoretical viewpoint it remains somewhat a mystery.
- In this section, we attempt to provide such a theoretical explanation.

49. Analysis of the Problem

- Let us start our analysis with the extreme situation when there is no stiffness at all.
- In this case, particles forming the soil are completely independent from each other.
- So, the displacement x_i of each particle i is determined by the Newton's equations $\dfrac{d^2 x_i}{dt^2} = \dfrac{1}{m_i} \cdot F_i$.
- For a vibrating compactor, the force F_i is sinusoidal with frequency f.
- Thus, the corresponding accelerations are also sinusoidal with this same frequency.
- So, after the Fourier transform, we will get only one component – corr. to the vibration frequency f.

50. Analysis of the Problem (cont-d)

- Stiffness k means that the acceleration of each particle i is also influence by the locations of other particles x_j:
$$\frac{d^2x_i}{dt^2} = \frac{1}{m_i} \cdot F_i + f_i(k, x_1, \ldots, x_N).$$
- Displacements are usually small. We consider the case when stiffness is also reasonably small.
- It is therefore reasonable to expand this expression in Taylor series and keep only the first few terms:
$$\frac{d^2x_i}{dt^2} = \frac{1}{m_i} \cdot F_i + k \cdot \sum_{j=1}^{N} a_{ij} \cdot x_j + k \cdot \sum_{j=1}^{N} \sum_{\ell=1}^{N} a_{ij\ell} \cdot x_j \cdot x_k.$$
- In general, the solution to the equations (3) depends on the value k: $x_i(t) = x_i(k, t)$.
- When deriving the above equations, we ignored terms which are quadratic (or of higher order) in terms of k.

51. Analysis of the Problem (cont-d)

- It is thus reasonable to also ignore quadratic terms in $x_i(k,t)$: $x_i(k,t)=x_i^{(0)}(t)+k\cdot x_i^{(1)}(t)$; thus:

$$\frac{d^2x_i^{(0)}}{dt^2}+k\cdot\frac{d^2x_i^{(1)}}{dt^2}=\frac{1}{m_i}\cdot F_i+k\cdot\sum_{j=1}^N a_{ij}\cdot x_j^{(0)}+k\cdot\sum_{j=1}^N\sum_{\ell=1}^N a_{ij\ell}\cdot x_j^{(0)}\cdot x_\ell^{(0)}.$$

- This formula should hold for all k, so:
  - terms independent on k should be equal, and
  - terms linear in k should be equal on both sides.
- By equating terms that do not depend on k, we get the linear equation $\dfrac{d^2x_i^{(0)}}{dt^2}=\dfrac{1}{m_i}\cdot F_i$.
- So, for the the sinusoidal force $F_i(t)=A_i\cdot\cos(\omega\cdot t+\Phi_i)$, we get $x_i^{(0)}(t)=a_i\cdot\cos(\omega\cdot t+\varphi_i)$.

52. Analysis of the Problem (cont-d)

- By equating terms linear in k, we get:

$$\frac{d^2 x_i^{(1)}}{dt^2} = \sum_{j=1}^{N} a_{ij} \cdot x_j^{(0)} + \sum_{j=1}^{N} \sum_{\ell=1}^{N} a_{ij\ell} \cdot x_j^{(0)} \cdot x_\ell^{(0)}.$$

- For the sinusoidal $x_i^{(0)}$:
  - linear terms $\sum\limits_{j=1}^{N} a_{ij} \cdot x_j^{(0)}$ in the right-hand side are sinusoidal with the same angular frequency ω;
  - quadratic terms $\sum\limits_{j=1}^{N} \sum\limits_{\ell=1}^{N} a_{ij\ell} \cdot x_j^{(0)} \cdot x_\ell^{(0)}$ are sinusoids with the double angular frequency 2ω.

53. Conclusion of This Section

- Thus, the right-hand side of the equation is the sum of two sinusoids corresponding to frequencies f and $2f$:

$$\frac{d^2x_i}{dt^2} = \frac{d^2x_i^{(0)}}{dt^2} + k\cdot\frac{d^2x_i^{(1)}}{dt^2} = A_i\cdot\cos\left(\omega\cdot t+\Phi_i\right)+$$

$$k\cdot\left(A_i^{(1)}\cdot\cos\left(\omega\cdot t+\Phi_i^{(1)}\right)+A_i^{(2)}\cdot\cos\left(2\omega\cdot t+\Phi_i^{(2)}\right)\right).$$

- The 1st harmonic's amplitude A_1 doesn't depend on k.
- The 2nd harmonic's amplitude A_2 is proportional to k.
- Thus, the stiffness k is indeed proportional to the CMV ratio A_2/A_1.

Part V

How General Can We Go: What Is Computable and What Is Not

54. Structure of This Part

This part was already covered in the thesis

Part VI
Decision Making under Uncertainty

55. Structure of This Part

6.1 Decision making under interval uncertainty: the usual case.

6.2 Decision making under interval uncertainty in conflict situations.

6.3 Decision making under probabilistic uncertainty.

6.4 Decision making under general uncertainty.

6.1. Which Point From an Interval Should We Choose?

56. Formulation of the Practical Problem

- One of the main objectives of science and engineering is to provide an optimal decision in different situations.
- In practice, we often have an algorithm that:
  - provides an optimal decision based
  - under the condition that we know the exact values of the corresponding parameters x.
- In practice, we usually know x with some uncertainty.
- For example, often, we only know an interval $[\underline{x}, \overline{x}]$ that contains the actual (unknown) value x.
- In the case of interval uncertainty, we can implement decisions corresponding to different values $x \in [\underline{x}, \overline{x}]$.
- Which value should we choose?
- Often, practitioners select the midpoint, but is this selection the best choice?

57. Formulation of the Problem in Precise Terms

- In general, we need to make a decision u based on the state x of the system.
- According to decision theory:
  - a rational person selects a decision
  - that maximizes the value of an appropriate function known as *utility*.
- We will consider situations when:
  - for each state x and for each decision u,
  - we know the value of the utility $f(x,u)$ corresponding to us choosing u.
- Then, an optimal decision $u_{\rm opt}(x)$ corresponding to x is the decision for which this utility is the largest:

$$f(x, u_{\rm opt}(x)) = \max_u f(x,u).$$

58. Formulation of the Problem (cont-d)

- In practice, we rarely know the exact state of the system, we usually know this state with some uncertainty.
- Often, we do not know the probabilities of different possible states x, we only know the bounds on x.
- In this section, we will consider the simplest case:
  - when the state is characterized by a single parameter, i.e., when x is a real number, and
  - when a decision is described by a single number u.
- In this case, we only know that the actual (unknown) state x belongs to the interval $[\underline{x}, \overline{x}]$.
- The question is: what decision u should we make in this case?

59. Formulation of the Problem (final)

- We also assume that the uncertainty with which we know x is relatively small.
- So in the corresponding Taylor series, we can only the first few terms in terms of this uncertainty.
- Since we already know how to compute the optimal value $u_{\rm opt}(x)$ corresponding to x, it may be easier:
  - instead of coming up with a new algorithm that describes u as a function of the bounds $\underline{x}$ and $\overline{x}$,
  - to come up with a value s for which $u = u_{\rm opt}(s)$.

60. Decision Making under Interval Uncertainty

- We know the state x with uncertainty.
- Thus, for each possible decision u, we do not know the exact value of the utility.
- We only know that this utility is equal to $f(x, u)$ for some $x \in [\underline{x}, \overline{x}]$.
- Thus, all we know is that this utility value belongs to the interval

$$[f^-(u), f^+(u)] = \left[\min_{x \in [\underline{x}, \overline{x}]} f(x, u), \max_{x \in [\underline{x}, \overline{x}]} f(x, u)\right].$$

- According to decision theory, we should select u for which the following combination is the largest:

$$\alpha \cdot f^+(u) + (1 - \alpha) \cdot f^-(u).$$

61. Decision Making under Interval Uncertainty (cont-d)

- Here, $\alpha = 1$ means that the decision maker is a complete optimist.
- In other words, he/she only takes into account the best-case situations.
- $\alpha = 0$ means that the decision maker is a complete pessimist.
- In other words, he/she only takes into account the worst-case situations.
- $\alpha \in (0, 1)$ means that the decision maker takes into account both worst-case and best-case scenarios.

62. Resulting Formulation of the Problem.

- In these terms, our goal is:
  - given the function $f(x,u)$ and the bounds $\underline{x}$ and $\overline{x}$,
  - to find the value u for which the following objective function takes the largest possible value:

$$\alpha \cdot \max_{x\in[\underline{x},\overline{x}]} f(x,u) + (1-\alpha)\cdot \min_{x\in[\underline{x},\overline{x}]} f(x,u) \to \max_u .$$

- Alternatively, we need to find s for which $u = u_{\rm opt}(s)$ maximizes the above objective function.

63. Description of the Solution

- The solution depends on whether $f(x,u)$ is increasing or decreasing with respect to x.
- If the objective function is an increasing function of x, then we should select a solution corresponding to
$$x = \alpha \cdot \overline{x} + (1-\alpha) \cdot \underline{x}.$$
- If the objective function is a decreasing function of x, then we should select a solution corresponding to
$$x = \alpha \cdot \underline{x} + (1-\alpha) \cdot \overline{x}.$$
- Thus, the usual selection of the midpoint s is only optimal for decision makers for which $\alpha = 0.5$.
- Intuitively, the above solution is in good accordance with the Hurwicz criterion.
- This is, however, not a proof: Hurwicz formula combines utilities, not parameter values.

64. Possible Future Work

- We have provided a solution for the simplest case, when we have:
  - only one parameter x describing the system and
  - only one parameter u describing possible alternatives.
- It's desirable to extend our solution to the case when we have several parameters x and several parameters u.
- What if we also have partial information about the probabilities of different values x from the interval?
- What if we have fuzzy information about x?
- In terms of α-cuts, what if we have different intervals $[\underline{x}, \overline{x}]$ for different levels of certainty α.

Part VII
Conclusions

65. Problem: Reminder

- In many practical application, we process measurement results and expert estimates.
- Measurements and expert estimates are never absolutely accurate.
- Their results are slightly different from the actual (unknown) values of the corresponding quantities.
- It is therefore desirable to analyze:
  - how this measurement and estimation inaccuracy
  - affects the results of data processing.
- There exist numerous methods for estimating the accuracy of the results of data processing.

66. Problem (cont-d)

- These methods work under different models of measurement and estimation inaccuracies:
  - probabilistic,
  - interval, and
  - fuzzy.
- To be useful in engineering applications, these methods:
  - should provide accurate estimate for the resulting uncertainty,
  - should not take too much computation time,
  - should be understandable to engineers, and
  - should be sufficiently general to cover all kinds of uncertainty.

67. What We Did

- In this dissertation, on several case studies, we show how we can achieve these four objectives:
  - We show that we can get more accurate estimates.
  - We show that we can speed up computations.
  - We show that we can make uncertainty-estimating algorithms more understandable.
  - We also analyze how general uncertainty-estimating algorithms can be.
- We also analyze decision making under uncertainty.
- Practical applications include:
  - to the inverse problem in geosciences,
  - to energy storage, and
  - to the problem of pavement compaction.

68. Acknowledgments

I want to express my gratitude to my committee members:

- Vladik Kreinovich, Chair
- Jack Chessa,
- Aaron Velasco, and
- Piotr Wojciechowski.

Appendix.
Part I. Introduction

69. A Faster Method: Cauchy-Based Monte-Carlo

- *Idea:* use Cauchy distribution $\rho_\Delta(x) = \dfrac{\Delta}{\pi} \cdot \dfrac{1}{1 + x^2/\Delta^2}$.
- *Why:* when $\Delta x_i \sim \rho_{\Delta_i}(x)$ are indep., then $\Delta y = \sum\limits_{i=1}^{n} c_i \cdot \Delta x_i \sim \rho_\Delta(x)$, with $\Delta = \sum\limits_{i=1}^{n} |c_i| \cdot \Delta_i$.
- Thus, we simulate $\Delta x_i^{(k)} \sim \rho_{\Delta_i}(x)$; then, $\Delta y^{(k)} \stackrel{\text{def}}{=} \widetilde{y} - f(\widetilde{x}_1 - \Delta x_1^{(k)}, \ldots) \sim \rho_\Delta(x)$.
- Maximum Likelihood method can estimate Δ: $\prod\limits_{k=1}^{N} \rho_\Delta(\Delta y^{(k)}) \to \max$, so $\sum\limits_{k=1}^{N} \dfrac{1}{1 + (\Delta y^{(k)})^2/\Delta^2} = \dfrac{N}{2}$.
- To find Δ from this equation, we can use, e.g., the bisection method for $\underline{\Delta} = 0$ and $\overline{\Delta} = \max\limits_{1 \le k \le N} |\Delta y^{(k)}|$.

70. Monte-Carlo: Successes and Limitations

- *Fact:* for Monte-Carlo, accuracy is $\varepsilon \sim 1/\sqrt{N}$.
- *Good news:* the number N of calls to f depends only the desired accuracy ε.
- *Example:* to find Δ with accuracy 20% and certainty 95%, we need $N = 200$ iterations.
- *Limitation:* this method is *not realistic*; indeed:
  - we know that Δx_i is *inside* $[-\Delta_i, \Delta_i]$, but
  - Cauchy-distributed variable has a high probability to be *outside* this interval.
- *Natural question:* is it a limitation of our method, or of a problem itself?

Part II: How to Get More Accurate Estimates

2.1. In System Identification, Interval Estimates Can lead to Much Better Accuracy than the Traditional Statistical Ones: Details

71. Algorithm for the Interval Case

- We know the expression $f(a_1, \ldots, a_m, x_1, \ldots, x_n)$.
- We know the measurement results $\widetilde{y}_k$ and $\widetilde{x}_{ki}$, and accuracies Δ_k and Δ_{ki}.
- First, we use Least Squares to find $\widetilde{a}_1, \ldots, \widetilde{a}_m$.
- Then, we compute $\underline{y}_k = \widetilde{y}_k - \Delta_k$, $\overline{y}_k = \widetilde{y}_k + \Delta_k$, and the partial derivatives b_j and b_{ki}.
- $\underline{a}_{j_0}$ $(\overline{a}_{j_0})$ is the solution to the following LP problem: minimize (maximize) a_{j_0} under the constraints

$$\underline{y}_k \le Y_k - \sum_{j=1}^{m} b_j \cdot \Delta a_j + \sum_{i=1}^{n} |b_{ki}| \cdot \Delta_{ki}, \quad 1 \le k \le K;$$

$$Y_k - \sum_{j=1}^{m} b_j \cdot \Delta a_j - \sum_{i=1}^{n} |b_{ki}| \cdot \Delta_{ki} \le \overline{y}_k, 1 \le k \le K.$$

72. How to Use These Formulas to Estimate y?

- What if we now need to predict the value y corresponding to given values $x_1, \ldots, x_m$?
- In this case, $y = f(a_1, \ldots, a_m, x_1, \ldots, x_n) =$

$$f(\widetilde{a}_1 - \Delta a_1, \ldots, \widetilde{a}_m - \Delta a_m, x_1, \ldots, x_n) = \widetilde{y} - \sum_{j=1}^{M} B_j \cdot \Delta a_j,$$

where $\widetilde{y} = f(\widetilde{a}_1, \ldots, \widetilde{a}_m, x_1, \ldots, x_n)$, $\displaystyle B_j \stackrel{\text{def}}{=} \frac{\partial f}{\partial a_j}_{|a_k = \widetilde{a}_k, x_i}$.

- The smallest possible value $\underline{y}$ of y can be found by minimizing $\widetilde{y} - \sum\limits_{j=1}^{m} B_j \cdot \Delta a_j$ under the same constraints.
- The largest possible value $\overline{y}$ of y can be found by maximizing the expression $\widetilde{y} - \sum\limits_{j=1}^{m} B_j \cdot \Delta a_j$.

73. What if We Underestimated the Measurement Inaccuracy?

- In practice, the constraints were often inconsistent.
- So, we underestimated the measurement inaccuracy.
- Since measuring y is the most difficult part, most probably we underestimated the accuracies of measuring y.
- Let's denote the ignored part of y-error by ε.
- Then, we should have $|\Delta y_k| \le \Delta_k + \varepsilon$.
- It's reasonable to look for the smallest $\varepsilon > 0$ s.t. constraints are consistent, i.e., minimize $\varepsilon > 0$ under:

$$\widetilde{y}_k - \Delta_k - \varepsilon \le Y_k - \sum_{j=1}^{m} b_j \cdot \Delta a_j + \sum_{i=1}^{n} |b_{ki}| \cdot \Delta_{ki},$$

$$Y_k - \sum_{j=1}^{m} b_j \cdot \Delta a_j - \sum_{i=1}^{n} |b_{ki}| \cdot \Delta_{ki} \le \widetilde{y}_k + \Delta_k + \varepsilon.$$

74. Simplest Case: Linear Dependence on One Variable $y = a \cdot x + b$

- Let's consider the case $a > 0$ ($a < 0$ is similar).
- In this case, the range of $a \cdot x + b$ is $[a \cdot \underline{x}_k + b, a \cdot \overline{x}_k + b]$.
- This interval intersects with $[\underline{y}_k, \overline{y}_k]$ if

$$a \cdot \underline{x}_k + b \le \overline{y}_k \text{ and } \underline{y}_k \le a \cdot \overline{x}_k + b.$$

- So, once we know a, we have the following lower bounds and upper bounds for b:

$$\underline{y}_k - a \cdot \overline{x}_k \le b \text{ and } b \le \overline{y}_k - a \cdot \underline{x}_k.$$

- Such a value b exists if and only if every lower bound for b is $\le$ every upper bound for b:

$$\underline{y}_k - a \cdot \overline{x}_k \le \overline{y}_\ell - a \cdot \underline{x}_\ell \text{ for all } k \text{ and } \ell.$$

- This is equivalent to $\overline{y}_\ell - \underline{y}_k \ge a \cdot (\underline{x}_\ell - \overline{x}_k)$.

75. Case When $y = a \cdot x + b$ (cont-d)

- We have $\overline{y}_\ell - \underline{y}_k \geq a \cdot (\underline{x}_\ell - \overline{x}_k)$.
- If $\underline{x}_\ell - \overline{x}_k > 0$, $a \leq \dfrac{\overline{y}_\ell - \underline{y}_k}{\underline{x}_\ell - \overline{x}_k}$; if $\underline{x}_\ell - \overline{x}_k < 0$, $a \geq \dfrac{\overline{y}_\ell - \underline{y}_k}{\underline{x}_\ell - \overline{x}_k}$.
- Thus, the range $[\underline{a}, \overline{a}]$ for a goes from the largest of the lower bounds to the smallest of the upper bounds:

$$\underline{a} = \max_{k,\ell:\ \underline{x}_\ell < \overline{x}_k} \frac{\overline{y}_\ell - \underline{y}_k}{\underline{x}_\ell - \overline{x}_k}; \quad \overline{a} = \min_{k,\ell:\ \underline{x}_\ell > \overline{x}_k} \frac{\overline{y}_\ell - \underline{y}_k}{\underline{x}_\ell - \overline{x}_k}.$$

- Similarly, $a \cdot \underline{x}_k + b \leq \overline{y}_k$ and $\underline{y}_k \leq a \cdot \overline{x}_k + b$ is equivalent to: $a \cdot x_k \leq \overline{y}_k - b$ and $\overline{y}_k - b \leq a \cdot \overline{x}_k$.
- If $\underline{x}_k > 0$, $a \leq \dfrac{\overline{y}_k}{\underline{x}_k} - \dfrac{1}{\underline{x}_k} \cdot b$; if $\underline{x}_k < 0$, $\dfrac{\overline{y}_k}{\underline{x}_k} - \dfrac{1}{\underline{x}_k} \cdot b \leq a$.
- If $\overline{x}_k > 0$, $\dfrac{\underline{y}_k}{\overline{x}_k} - \dfrac{1}{\overline{x}_k} \cdot b \leq a$; if $\overline{x}_k > 0$, $a \leq \dfrac{\underline{y}_k}{\overline{x}_k} - \dfrac{1}{\overline{x}_k} \cdot b$.

76. Case When $y = a \cdot x + b$ (cont-d)

- Inequalities $A_p + B_p \cdot b \le a$, $a \le C_q + D_q \cdot b$ are consistent if every lower bound $\le$ every upper bound:

$$A_p + B_p \cdot b \le C_q + D_q \cdot b \Leftrightarrow (D_q - B_p) \cdot b \ge A_p - C_q.$$

- So, similarly to the a-case, we get:

$$\underline{b} = \max_{p,q:\ D_q > B_p} \frac{A_p - C_q}{D_q - B_p}; \quad \overline{b} = \max_{p,q:\ D_q < B_p} \frac{A_p - C_q}{D_q - B_p}.$$

- If we underestimated the measurement inaccuracy, we get the new bounds $\underline{y}_k - \varepsilon$ and $\overline{y}_k + \varepsilon$.

- So, if $\underline{x}_\ell > \overline{x}_k$, we get $a \le \dfrac{\overline{y}_\ell - \underline{y}_k}{\underline{x}_\ell - \overline{x}_k} + \dfrac{2}{\underline{x}_\ell - \overline{x}_k} \cdot \varepsilon$, else $a \ge \dfrac{\overline{y}_\ell - \underline{y}_k}{\underline{x}_\ell - \overline{x}_k} + \dfrac{2}{\underline{x}_\ell - \overline{x}_k} \cdot \varepsilon.$

77. What If We Underestimate Measurement Uncertainty

- If $\underline{x}_\ell > \overline{x}_k$, we get $a \le \dfrac{\overline{y}_\ell - \underline{y}_k}{\underline{x}_\ell - \overline{x}_k} + \dfrac{2}{\underline{x}_\ell - \overline{x}_k} \cdot \varepsilon$, else $a \ge \dfrac{\overline{y}_\ell - \underline{y}_k}{\underline{x}_\ell - \overline{x}_k} + \dfrac{2}{\underline{x}_\ell - \overline{x}_k} \cdot \varepsilon$.
- Inequalities $A_p + B_p \cdot \varepsilon \le a$ and $a \le C_q + D_q \cdot \varepsilon$ are consistent if every lower bound $\le$ every upper bound:

$$A_p + B_p \cdot \varepsilon \le C_q + D_q \cdot \varepsilon \Leftrightarrow (D_q - B_p) \cdot \varepsilon \ge A_p - C_q.$$

- So, the desired lower bound for ε for b is equal to the largest of the lower bounds:

$$\varepsilon = \max_{p,q:\ D_q > B_p} \frac{A_p - C_q}{D_q - B_p}.$$

2.2. Which Value $\widetilde{x}$ Best Represents a Sample $x_1, \ldots, x_n$: Utility-Based Approach Under Interval Uncertainty

78. Need to Combine Several Estimates

- In many practical situations, we have several estimates $x_1, \ldots, x_n$ of the same quantity x.
- In such situations, it is often desirable to combine this information into a single estimate $\widetilde{x}$.
- Sometimes, we know the probability distribution of the corresponding estimation errors $x_i - x$.
- Then, we can use known statistical techniques to find $\widetilde{x}$.
- E.g., we can use the Maximum Likelihood Method.
- In many cases, however, we do not have any information about the corresponding probability distribution.
- How can we then find $\widetilde{x}$?

79. Utility-Based Approach

- According to the decision theory, decisions of a rational person are $\Leftrightarrow$ maximizing *utility value* u.
- Let us thus find the estimate $\widetilde{x}$ for which the utility $u(\widetilde{x})$ is the largest.
- We use a single value $\widetilde{x}$ instead of all n values x_i; for each i:
  - if the actual estimate is x_i and we use a different value $\widetilde{x} \neq x_i$ instead,
  - then we are not doing an optimal thing.
- For example:
  - if the optimal speed at which the car needs the least amount of fuel is x_i,
  - and we instead run it at a speed $\widetilde{x} \neq x_i$, we thus waste some fuel.

80. Utility-Based Approach (cont-d)

- For each i, the disutility d comes from the fact that the difference $\widetilde{x} - x_i$ is different from 0.
- There is no disutility if we use the actual value, so $d = d(\widetilde{x} - x_i)$ for some function $d(y)$.
- Here $d(0) = 0$ and $d(y) > 0$ for $y \neq 0$.
- The estimates are usually reasonably accurate, so the difference $x_i - \widetilde{x}$ is small.
- So, we can expand the function $d(y)$ in Taylor series and keep only the first few terms in this expansion:
$$d(y) = d_0 + d_1 \cdot y + d_2 \cdot y^2 + \ldots$$
- From $d(0) = 0$ we conclude that $d_0 = 0$.
- From $d(y) > 0$ for $y \neq 0$ we conclude that $d_1 = 0$ (else we would have $d(y) < 0$ for small y) and $d_2 > 0$, so
$$d(y) = d_2 \cdot y^2 = d_2 \cdot (\widetilde{x} - x_i)^2.$$

81. Utility-Based Approach (final)

- The overall disutility $d(\widetilde{x})$ of using $\widetilde{x}$ instead of each of the values $x_1, \ldots, x_n$ can be computed as the sum:

$$d(\widetilde{x}) = \sum_{i=1}^{n} d(\widetilde{x} - x_i)^2 = d_2 \cdot \sum_{i=1}^{n} (\widetilde{x} - x_i)^2.$$

- $u(\widetilde{x}) \stackrel{\text{def}}{=} -d(\widetilde{x}) \to \max \Leftrightarrow d(\widetilde{x}) \to \min.$
- Since $d_2 > 0$, minimizing disutility is equivalent to minimizing re-scaled disutility:

$$D(\widetilde{x}) \stackrel{\text{def}}{=} \frac{d(\widetilde{x})}{d_2} = \sum_{i=1}^{n} (\widetilde{x} - x_i)^2.$$

- Equating the derivative to 0, we get the well-known sample mean: $\widetilde{x} = \dfrac{1}{n} \cdot \displaystyle\sum_{i=1}^{n} x_i.$

82. Case of Interval Uncertainty

- In many practical situations, we only know the intervals $[\underline{x}_i, \overline{x}_i]$ that contain the unknown values x_i.
- For different $x_i \in [\underline{x}_i, \overline{x}_i]$, we get, in general, different values of utility

$$U(\widetilde{x}, x_1, \ldots, x_n) = -D(\widetilde{x}, x_1, \ldots, x_n), \text{ where}$$

$$D(\widetilde{x}, x_1, \ldots, x_n) = \sum_{i=1}^{n} (\widetilde{x} - x_i)^2.$$

- Thus, all we know is that the actual (unknown) value of the utility belongs to the interval

$$[\underline{U}(\widetilde{x}), \overline{U}(\widetilde{x})] = [-\overline{D}(\widetilde{x}), -\underline{D}(\widetilde{x})], \text{ where}$$

$$\underline{D}(\widetilde{x}) = \min D(\widetilde{x}, x_1, \ldots, x_n), \ \overline{D}(\widetilde{x}) = \max D(\widetilde{x}, x_1, \ldots, x_n).$$

- In such situations, decision theory recommends using Hurwicz optimism-pessimism criterion, i.e., maximize:

$$U(\widetilde{x}) \stackrel{\text{def}}{=} \alpha \cdot \overline{U}(\widetilde{x}) + (1 - \alpha) \cdot \underline{U}(\widetilde{x}).$$

83. Case of Interval Uncertainty (cont-d)

- According to Hurwicz criterion, we maximize:

$$U(\widetilde{x}) \stackrel{\text{def}}{=} \alpha \cdot \overline{U}(\widetilde{x}) + (1-\alpha) \cdot \underline{U}(\widetilde{x}).$$

- The parameter $\alpha \in [0,1]$ describes the decision maker's degree of optimism.
- For $U = -D$, this is equivalent to minimizing the expression

$$D(\widetilde{x}) = -U(\widetilde{x}) = \alpha \cdot \underline{D}(\widetilde{x}) + (1-\alpha) \cdot \overline{D}(\widetilde{x}).$$

- In this section, we describe an efficient algorithm for computing such $\widetilde{x}$.

84. Analysis of the Problem

- Each term $(\widetilde{x}-x_i)^2$ in the sum $D(\widetilde{x}, x_1, \ldots, x_n)$ depends only on its own variable x_i. Thus, with respect to x_i:
  - the sum is the smallest when each term is min, and
  - the sum is the largest when each term is the largest.
- When $x_i \in [\underline{x}_i, \overline{x}_i]$, the max of $(\widetilde{x} - x_i)^2$ is attained:
  - at $x_i = \underline{x}_i$ when $\widetilde{x} \geq \widetilde{x}_i \stackrel{\text{def}}{=} \dfrac{\underline{x}_i + \overline{x}_i}{2}$ and
  - at $x_i = \overline{x}_i$ when $\widetilde{x} < \widetilde{x}_i$.
- Thus, $\overline{D}(\widetilde{x}) = \sum\limits_{i:\widetilde{x}<\widetilde{x}_i} (\widetilde{x} - \overline{x}_i)^2 + \sum\limits_{i:\widetilde{x}\geq\widetilde{x}_i} (\widetilde{x} - \underline{x}_i)^2$.
- Similarly, the minimum of the term $(\widetilde{x}-x_i)^2$ is attained:
  - for $x_i = \widetilde{x}$ when $\widetilde{x} \in [\underline{x}_i, \overline{x}_i]$ (in this case, min $= 0$);
  - for $x_i = \underline{x}_i$ when $\widetilde{x} < \underline{x}_i$; and
  - for $x_i = \overline{x}_i$ when $\widetilde{x} > \overline{x}_i$.

85. Analysis of the Problem (cont-d)

- Thus, $\underline{D}(\widetilde{x}) = \sum\limits_{i:\widetilde{x}>\overline{x}_i} (\widetilde{x}-\overline{x}_i)^2 + \sum\limits_{i:\widetilde{x}<\underline{x}_i} (\widetilde{x}-\underline{x}_i)^2.$
- So, for $D(\widetilde{x}) = \alpha\cdot\underline{D}(\widetilde{x}) + (1-\alpha)\cdot\overline{D}(\widetilde{x})$, we get

$$D(\widetilde{x}) = \alpha\cdot\sum_{i:\widetilde{x}>\overline{x}_i} (\widetilde{x}-\overline{x}_i)^2 + \alpha\cdot\sum_{i:\widetilde{x}<\underline{x}_i} (\widetilde{x}-\underline{x}_i)^2 +$$

$$(1-\alpha)\cdot\sum_{i:\widetilde{x}<\widetilde{x}_i} (\widetilde{x}-\overline{x}_i)^2 + (1-\alpha)\cdot\sum_{i:\widetilde{x}\geq\widetilde{x}_i} (\widetilde{x}-\underline{x}_i)^2.$$

86. Towards an Algorithm

- The terms depends on the relation between $\widetilde{x}$ and the values
$$\underline{x}_i, \overline{x}_i, \text{ and } \widetilde{x}_i.$$
- Let us sort these $3n$ values into a sequence
$$s_1 \le s_2 \le \ldots \le s_{3n}.$$
- Then on each interval $[s_j, s_{j+1}]$, the function $D(\widetilde{x})$ is simply a quadratic function of $\widetilde{x}$.
- A quadratic function attains min on an interval:
  - either at one of its midpoints,
  - or at a point when the derivative is equal to 0 (if this point is inside the given interval).

87. Towards an Algorithm (cont-d)

- Equating the derivative $D(\widetilde{x})$ to 0, we get:

$$(\alpha\cdot\#\{i:\widetilde{x}<\underline{x}_i \text{ or } \widetilde{x}>\overline{x}_i\}+1-\alpha)\cdot\widetilde{x}=$$

$$\alpha\cdot\sum_{i:\widetilde{x}>\overline{x}_i}\overline{x}_i+\alpha\cdot\sum_{i:\widetilde{x}<\underline{x}_i}\underline{x}_i+(1-\alpha)\cdot\sum_{i:\widetilde{x}<\widetilde{x}_i}\overline{x}_i+(1-\alpha)\cdot\sum_{i:\widetilde{x}\ge\widetilde{x}_i}\underline{x}_i.$$

- s_j is a listing of all thresholds values $\underline{x}_i$, $\overline{x}_i$, and $\widetilde{x}_i$.
- So, for $\widetilde{x}\in(s_j,s_{j+1})$, $\widetilde{x}<\underline{x}_i \Leftrightarrow s_{j+1}\le\underline{x}_i$.
- Similarly, the inequality $\widetilde{x}>\underline{x}_i$ is equivalent to $s_j\ge\overline{x}_i$.
- In general, for $\widetilde{x}\in(s_j,s_{j+1})$, we get:

$$(\alpha\cdot\#\{i:\widetilde{x}<\underline{x}_i \text{ or } \widetilde{x}>\overline{x}_i\}+1-\alpha)\cdot\widetilde{x}=$$

$$\alpha\cdot\sum_{i:s_j\ge\overline{x}_i}\overline{x}_i+\alpha\cdot\sum_{i:s_{j+1}\le\underline{x}_i}\underline{x}_i+(1-\alpha)\cdot\sum_{i:s_{j+1}\le\widetilde{x}_i}\overline{x}_i+(1-\alpha)\cdot\sum_{i:s_j\ge\widetilde{x}_i}\underline{x}_i.$$

- We can thus find $\widetilde{x}$ at which the derivative is 0.
- Thus, we arrive at the following algorithm.

88. Resulting Algorithm

- We want to minimize the expression

$$D(\widetilde{x}) = -U(\widetilde{x}) = \alpha \cdot \underline{D}(\widetilde{x}) + (1-\alpha) \cdot \overline{D}(\widetilde{x}), \text{ where}$$

$$\underline{D}(\widetilde{x}) = \alpha \cdot \sum_{i:\widetilde{x}>\overline{x}_i} (\widetilde{x} - \overline{x}_i)^2 + \alpha \cdot \sum_{i:\widetilde{x}<\underline{x}_i} (\widetilde{x} - \underline{x}_i)^2 +$$

$$(1-\alpha) \cdot \sum_{i:\widetilde{x}<\widetilde{x}_i} (\widetilde{x} - \overline{x}_i)^2 + (1-\alpha) \cdot \sum_{i:\widetilde{x}\ge\widetilde{x}_i} (\widetilde{x} - \underline{x}_i)^2.$$

- First, for each interval $[\underline{x}_i, \overline{x}_i]$, we compute its midpoint

$$\widetilde{x}_i = \frac{\underline{x}_i + \overline{x}_i}{2}.$$

- Then, we sort the $3n$ values $\underline{x}_i$, $\overline{x}_i$, and $\widetilde{x}_i$ into an increasing sequence $s_1 \le s_2 \le \ldots \le s_{3n}$.
- To cover the whole real line, to these values, we add $s_0 = -\infty$ and $s_{3n+1} = +\infty$.

89. Algorithm (cont-d)

- We compute the value of the objective function on each of the endpoints $s_1, \ldots, s_{3n}$.
- Then, for each interval (s_i, s_{j+1}), we compute $\widetilde{x}$ as:

$$\frac{\alpha \sum\limits_{i:s_j \ge \overline{x}_i} \overline{x}_i + \alpha \sum\limits_{i:s_{j+1} \le \underline{x}_i} \underline{x}_i + (1-\alpha) \sum\limits_{i:s_{j+1} \le \widetilde{x}_i} \overline{x}_i + (1-\alpha) \sum\limits_{i:s_j \ge \widetilde{x}_i} \underline{x}_i}{\alpha \cdot \#\{i : \widetilde{x} < \underline{x}_i \text{ or } \widetilde{x} > \overline{x}_i\} + 1 - \alpha}.$$

- If $\widetilde{x}$ is within (s_i, s_{j+1}), we compute $D(\widetilde{x})$.
- After that:
  - out of all the values $\widetilde{x}$ for which we computed $D(\widetilde{x})$,
  - we return $\widetilde{x}$ for which $D(\widetilde{x})$ is the smallest.

90. What Is the Computational Complexity of This Algorithm

- Sorting $3n = O(n)$ values $\underline{x}_i$, $\overline{x}_i$, and $\widetilde{x}_i$ takes time

$$O(n \cdot \ln(n)).$$

- Computing each value $D(\widetilde{x})$ of the objective function requires $O(n)$ computational steps.
- We compute $d(\widetilde{x})$:
  - for $3n$ endpoints and
  - for $\le 3n+1$ values at which the derivative is 0 at each of the intervals (s_j, s_{j+1}).
- Overall, we compute $D(\widetilde{x})$ at $O(n)$ values.
- Thus, overall, we need $O(n \cdot \ln(n)) + O(n) \cdot O(n) = O(n^2)$ computation steps.
- Hence, our algorithm runs in quadratic time.

2.4. How to Gauge the Accuracy of Fuzzy Control Recommendations: A Simple Idea

91. Need to Gauge Accuracy of Fuzzy Recommendations

- Fuzzy logic has been successfully applied to many different application areas, e.g., in control.
- A natural question is: with what accuracy do we need to implement this recommendation?
- In many applications, this is an important question:
  - it is often much easier to implement the control value approximately,
  - but maybe a more accurate actuator is needed?
- To answer this question, we must be able to gauge the accuracy of the corresponding recommendations.

92. Such Gauging Is Possible for Probabilistic Uncertainty

- Probabilistic uncertainty means that instead of the exact value x, we only know a probability distribution.
- This distribution can be described, e.g., by the probability density $\rho(x)$.
- If we need to select a single value x, a natural idea is to select, e.g., the mean value $\overline{\mathrm{x}} = \int x \cdot \rho(x)\,dx$.
- A natural measure of accuracy is the mean square deviation from the mean, known as the standard deviation:

$$\sigma \stackrel{\text{def}}{=} \sqrt{\int (x - \overline{\mathrm{x}})^2\,dx}.$$

- We need a similar formula for the fuzzy case.

93. How We Elicit Fuzzy Degrees: A Brief Reminder

- For each possible value x of the corresponding quantity, we ask the expert to mark:
  - on a scale from 0 to 1,
  - his/her degree of confidence that x satisfies the given property.
- For example, we ask the expert to specify the degree to which the value x is small.
- In some cases, this is all we need.
- However, in many other cases, we get a *non-normalized* membership function, for which $\max\limits_x \mu(x) < 1$.
- Most fuzzy techniques assume that the membership function is normalized.

94. How We Elicit Fuzzy Degrees (cont-d)

- So, we sometimes need to perform an additional step to get an easy-to-process membership function.
- Namely, we *normalize* the original values $\mu(x)$ by dividing them by the largest of the values $\mu(y)$:

$$\mu'(x) \stackrel{\text{def}}{=} \frac{\mu(x)}{\max\limits_{y} \mu(y)}.$$

- Sometimes, the experts have some subjective probabilities $\rho(x)$ assigned to different values x.
- In this case, when asked to indicate their degree of certainty, they list $\mu(x) = \rho(x)$.
- After normalizing this $\mu(x)$, we get the membership function $\mu(x) = \dfrac{\rho(x)}{\max\limits_{y} \rho(y)}$.

95. Let Us Use This Idea to Gauge the Accuracy of Fuzzy Recommendations

- We assign, to each probability density function $\rho(x)$, a membership function $\mu(x) = \dfrac{\rho(x)}{\max\limits_y \rho(y)}$.
- Vice versa, if we know that $\mu(x)$ was obtained by normalizing some $\rho(x)$, we can uniquely reconstruct $\rho(x)$:
$$\rho(x) = \frac{\mu(x)}{\int \mu(y)\,dy}.$$
- Our idea is then to use the probabilistic formulas corresponding to this artificial distribution.
- At first glance, this does not make sense.
- The probabilistic measure of accuracy is based on the assumption that we use the mean.
- But don't we use something else in fuzzy?

96. Let Us Use This Idea to Gauge the Accuracy of Fuzzy Recommendations (cont-d)

- Don't we use something else in fuzzy?
- Actually, not really.
- The mean of the distribution $\rho(x) = \dfrac{\mu(x)}{\int \mu(y)\,dy}$ is

$$\overline{\mathrm{x}} = \int x \cdot \rho(x)\,dx = \frac{\int x \cdot \mu(x)\,dx}{\int \mu(x)\,dx}.$$

- This is the centroid defuzzification – one of the main ways to transform $\mu(x)$ into a control recommendation.
- Since the above idea makes sense, let us use it to gauge the accuracy of the fuzzy control recommendation.

97. Resulting Recommendation

- For a given membership function $\mu(x)$, we usually generate the result $\overline{\mathrm{x}}$ of its centroid defuzzification.
- We should also generate, as a measure of the accuracy of this recommendation, the following value σ:

$$\sigma^2 = \int (x - \overline{\mathrm{x}})^2 \cdot \rho(x)\,dx = \frac{\int (x-\overline{\mathrm{x}})^2 \cdot \mu(x)\,dx}{\int \mu(x)\,dx} =$$

$$\frac{\int x^2 \cdot \mu(x)\,dx}{\int \mu(x)\,dx} - \left(\frac{\int x \cdot \mu(x)\,dx}{\int \mu(x)\,dx}\right)^2.$$

98. But What Should We Do in the Interval-Valued Fuzzy Case?

- Often, experts cannot tell us the exact values $\mu(x)$.
- Instead, for each x, they tell us the interval $[\underline{\mu}(x), \overline{\mu}(x)]$ of possible value of degree of confidence $\mu(x)$.
- For different functions $\mu(x) \in [\underline{\mu}(x), \overline{\mu}(x)]$, we get different values σ^2.
- It is desirable to find the range of possible values σ^2 when $\mu(x) \in [\underline{\mu}(x), \overline{\mu}(x)]$.

99. Resulting Algorithm

- For all possible pairs $\underline{x} < \overline{x}$, we compute $\sigma^2(\mu^-)$ and $\sigma^2(\mu^+)$, where:
  - $\mu^+(x) = \overline{\mu}(x)$ when $x < \underline{x}$ or $x > \overline{x}$, and $\mu^+(x) = \underline{\mu}(x)$ when $\underline{x} < x < \overline{x}$;
  - $\mu^-(x) = \underline{\mu}(x)$ when $x < \underline{x}$ or $x > \overline{x}$, and $\mu^-(x) = \overline{\mu}(x)$ when $\underline{x} < x < \overline{x}$.
- As the upper bound for σ^2, we take the maximum of the values $\sigma^2(\mu^+)$ corresponding to different pairs $\underline{x} < \overline{x}$.
- As the lower bound for σ^2, we take the minimum of the values $\sigma^2(\mu^-)$ corresponding to different pairs $\underline{x} < \overline{x}$.

100. The Algorithm in Section 2.4 Is Correct

- According to calculus, when $f(z)$ attains max on $[\underline{z}, \overline{z}]$ at $z_0 \in [\underline{z}, \overline{z}]$, then we have one of the three cases:
  - we can have $z_0 \in (\underline{z}, \overline{z})$, in which case $\dfrac{df}{dz}(z_0) = 0$;
  - we can have $z_0 = \underline{z}$, in which case $\dfrac{df}{dz}(z_0) \le 0$, or
  - we can have $z_0 = \overline{z}$, in which case $\dfrac{df}{dz}(z_0) \ge 0$.
- Similarly, when $f(z)$ attains min on $[\underline{z}, \overline{z}]$ at $z_0 \in [\underline{z}, \overline{z}]$, then we have one of the three cases:
  - we can have $z_0 \in (\underline{z}, \overline{z})$, in which case $\dfrac{df}{dz}(z_0) = 0$;
  - we can have $z_0 = \underline{z}$, in this case, in which case $\dfrac{df}{dz}(z_0) \ge 0$, or
  - we can have $z_0 = \overline{z}$, in which case $\dfrac{df}{dz}(z_0) \le 0$.

101. Proof (cont-d)

- Let us apply this to the dependence of σ on $\mu(a)$.
- Here, since $\int \mu(x)\,dx \approx \sum \mu(x_i)\cdot \Delta x_i$, we get:

$$\frac{\partial(\int \mu(x)\,dx)}{\partial(\mu(a))} = \Delta x, \quad \frac{\partial(\int x\cdot \mu(x)\,dx)}{\partial(\mu(a))} = a\cdot \Delta x \text{ and}$$

$$\frac{\partial(\int x^2\cdot \mu(x)\,dx)}{\partial(\mu(a))} = a^2\cdot \Delta x.$$

- By using the usual rules for differentiating the ratio, for the composition, and for the square, we get:

$$\frac{\partial(\sigma^2)}{\partial(\mu(a))} = \Delta x\cdot S(a), \text{ where } S(a) \stackrel{\text{def}}{=} \frac{a^2}{\int \mu(x)\,dx} -$$

$$\frac{\int x^2\cdot \mu(x)\,dx}{\left(\int \mu(x)\,dx\right)^2} - 2\cdot \overline{\mathrm{x}}\cdot \left(\frac{a}{\int \mu(x)\,dx} - \frac{\int x\cdot \mu(x)\,dx}{\left(\int \mu(x)\,dx\right)^2}\right).$$

102. Proof (cont-d)

- We are only interested in the sign of the derivative $\dfrac{\partial(\sigma^2)}{\partial(\mu(a))}$, and this is exactly the sign of $S(a)$.
- Similarly, the sign of $S(a)$ is the same as the sign of $s(a) \stackrel{\text{def}}{=} S(a) \cdot \int \mu(y)\,dy$, for which:

$$s(a) = a^2 - ((\overline{\mathrm{x}})^2 + \sigma^2) - 2 \cdot \overline{\mathrm{x}} \cdot (a - \overline{\mathrm{x}}).$$

- If we know the roots $\underline{x} < \overline{x}$ of this quadratic expression, we can conclude that this quadratic expression $s(a)$ is:
  - positive when $a < \underline{x}$ and
  - negative when $a > \overline{x}$.
- Here, the value $a = \overline{\mathrm{x}}$ is between $\underline{x}$ and $\overline{x}$, since for this value a, we have $s(\overline{\mathrm{x}}) = -\sigma^2 < 0$.

103. Proof (final)

- Thus, due to calculus, when $a < \underline{x}$ or $a > \overline{x}$:
  - to find $\overline{\sigma^2}$, we must take $\mu(a) = \overline{\mu}(a)$ and
  - to find $\underline{\sigma^2}$, we must take $\mu(a) = \underline{\mu}(a)$.
- When $\underline{x} < a < \overline{a}$, then, vice versa:
  - we need to take $\mu(a) = \underline{\mu}(a)$ to find $\overline{\sigma^2}$ and
  - we must take $\mu(a) = \overline{\mu}(a)$ to find $\underline{\sigma^2}$.
- This mathematical conclusion makes perfect sense:
  - to get the largest σ^2, we concentrate the distribution as much as possible on values far from $\overline{\mathrm{x}}$;
  - to get the smallest σ^2, we concentrate it as much as possible on values close to the mean $\overline{\mathrm{x}}$.
- Thus, we arrive at the above algorithm.

Part III: How to Speed Up Computations

3.2. Fuzzy Data Processing Beyond Min t-Norm

104. Need to Go Beyond min t-Norm

- Usually, Zadeh's extension principle is applied to the situations in which we use the min "and"-operation

$$f_\&(a, b) = \min(a, b).$$

- However, in many practical situations, other "and"-operations more adequately describe expert reasoning.
- It is therefore desirable to consider the general case

$$m(y) = \sup\{f_\&(m_1(x_1), m_2(x_2), \ldots : u = f(x_1, \ldots, x_n)\}.$$

- We are interested in the linearized case, when

$$\Delta y = \sum_{i=1}^{n} c_i \cdot \Delta x_i.$$

- How can we speed up computations in this general case?

105. We Can Reduce This Problem to Computing the Sum of Two Fuzzy Quantities

- We know that $\Delta y = \sum\limits_{i=1}^{n} y_i$, where $y_i \stackrel{\text{def}}{=} c_i \cdot \Delta x_i$.
- Once we know the membership f-n $m_i(\Delta x_i)$, we can easily find the membership f-n for y_i as $m_i(y_i/c_i)$.
- If we know how to find the membership f-n for the sum of two fuzzy quantities, then we can:
  - find the membership f-n for $s_2 = y_1 + y_2$;
  - then, find membership f-n for
  $$s_3 = s_2 + y_2 \ (= y_1 + y_2 + y_3),$$
  - etc, until we reach $s_n = y$.
- For fuzzy quantities with membership f-ns $n_1(x_1)$ and $n_2(x_2)$, the membership f-n for $y = x_1 + x_2$ is
$$n(y) = \max_{x_1} f_{\&}(n_1(x_1), n_2(y - x_1)).$$

106. Reduction to the Case of Product t-Norm

- We want to compute

$$n(y) = \max_{x_1} f_\&(n_1(x_1), n_2(y - x_1)).$$

- Some t-norms are *Archimedean*; for them:

$$f_\&(a, b) = f^{-1}(f(a) \cdot f(b)) \text{ for some function } f(x).$$

- Archimedean t-norms are universal approximators:
  - for every $\varepsilon > 0$,
  - every t-norm $f_\&$ is ε-close to some Archimedean.
- Thus, from the practical viewpoint, we can safely assume that $f_\&$ is Archimedean.
- Then, for $q_i(x_i) \stackrel{\text{def}}{=} f(n_i(x_i))$ and $q(y) \stackrel{\text{def}}{=} f(n(y))$:

$$q(y) = \max_{x_1} q_1(x_1) \cdot q_2(y - x_1).$$

107. Reduction to Informal Convolution

- We want to compute $q(y) = \max\limits_{x_1} q_1(x_1) \cdot q_2(y - x_1)$.
- For $\ell_i(x_i) \stackrel{\text{def}}{=} -\ln(q_i(x_i))$ and $\ell(y) \stackrel{\text{def}}{=} -\ln(q(y))$:
$$\ell(y) = \min_{x_1}(\ell_1(x_1) + \ell_2(y - x_1)).$$
- This operation is known as *infimal convlution*, and denoted by $\ell = \ell_1 \,\Box\, \ell_2$.
- It is known that $\ell_1 \Box \ell_2 = (\ell_1^* + \ell_2^*)^*$ for *Legendre transform*
$$\ell^*(s) \stackrel{\text{def}}{=} \sup_x (s \cdot x - \ell(x)).$$
- There exists a linear-time algorithm for computing Legendra transform.
- Thus, we can compute $\ell_1 \,\Box\, \ell_2$ in linear time.

108. Resulting Linear Time Algorithm for Fuzzy Data Processing Beyond min t-Norm

- We are given:
  - a function $f(x_1, \ldots, x_n)$;
  - n membership functions $m_1(x_1)$, ..., $m_n(x_n)$; and
  - an "and"-operation $f_\&(a, b)$.
- We want to compute a new membership function

$$m(y) = \max\{f_\&(m_1(x_1), \ldots, m_n(x_n)) : f(x_1, \ldots, x_n) = y\}.$$

- First, we represent $f_\&$ in the Acrhiemdean form $f_\&(a, b) = f^{-1}(f(a) \cdot f(b))$ for an appropriate $f(x)$.
- We thus assume that we have algorithms for computing $f(x)$ and the inverse function $f^{-1}(x)$.
- Then, for each i, we find the value $\widetilde{x}_i$ for which $m_i(x_i)$ attains its largest possible value $m_i(\widetilde{x}_i) = 1$.

109. Algorithm (cont-d)

- We then compute the values $c_0 = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, $c_i = \dfrac{\partial f}{\partial x_i}_{|x_i = \widetilde{x}_i}$, and $a_0 = c_0 - \sum\limits_{i=1}^{n} c_i \cdot \widetilde{x}_i$.
- Then, $f(x_1, \ldots, x_n) \approx a_0 + \sum\limits_{i=1}^{n} c_i \cdot x_i$.
- After that, e compute the membership f-ns $n_1(s_1) = m_1((s_1 - a_0)/c_1)$ and $n_i(y_i) = m_i(y_i/c_i)$ for $i > 2$.
- In terms of the variables $s_1 = a_0 + c_1 \cdot x_1$ and $y_i = c_i \cdot x_i$, the desired quantity y has the form $y = s_1 + y_2 + \ldots + y_n$.
- We compute the minus logarithms of the resulting functions: $\ell_i(y_i) = -\ln(n_i(y_i))$.
- For each i, we then use the Fast Legendre Transform algorithm to compute ℓ_i^*.

110. Algorithm (final)

- Then, we add all these Legendre transforms and apply the Fast Legendre Transform once again, getting:

$$\ell = (\ell_1^* + \ldots + \ell_n^*)^*.$$

- This function $\ell(y)$ is equal to $\ell(y) = -\ln(n(y))$, so we can reconstruct $\nu(y)$ as $n(y) = \exp(-\ell(y))$.
- Finally, we can compute the desired membership function $m(u)$ as $m(y) = f^{-1}(n(y))$.

111. Conclusion for this Section

- To process fuzzy data, we need to use Zadeh's extension principle.
- In principle, this principle can be used for any t-norm.
- However, usually, it is only used for the min t-norm.
- Reason: only for this t-norm, an efficient (linear-time) algorithm for fuzzy data processing was known.
- In many practical situations, other t-norms are more adequate in describing expert's reasoning.
- We have shown that similar efficient linear-time algorithms can be designed for an arbitrary t-norm.
- Thus, it is possible to use a more adequate t-norm – and keep fuzzy data processing efficient.

Part IV: Towards a Better Understanding of Uncertainty-Estimating Algorithms

4.1. Case of Interval Uncertainty: Practical Need for Algebraic (Equality-Type) Solutions of Interval Equations and for Extended-Zero Solutions

112. Sometimes, We do not Know the Exact Dependence

- So far, we assumed that when we know the exact dependence $y_i = f_i(x_1, \ldots, x_n)$ between y_i and x_j.
- In practice, often, we do not know the exact dependence.
- Instead, we know that the dependence belongs to a finite-parametric *family* of dependencies, i.e., that

$$y_i = f_i(x_1, \ldots, x_n, a_1, \ldots, a_k) \text{ for some parameters } a_1, \ldots, a_k.$$

- *Example:* y_i is a linear function of x_j, i.e.,
  $y_i = c_i + \sum\limits_{j=1}^{n} c_{ij} \cdot x_j$ for some c_i and c_{ij}.
- The presence of these parameters complicates the corresponding data processing problem.
- Depending on what we know about the parameters, we have different situations.

113. Specific Case: Control Solution

- Sometimes, we can *control* the values a_ℓ, by setting them to any values within certain intervals $[\underline{a}_\ell, \overline{a}_\ell]$.
- By setting the appropriate values of the parameters, we can change the values y_i.
- We would like the values y_i to be within some given ranges $[\underline{y}_i, \overline{y}_i]$.
- For example, we would like the temperature to be within a comfort zone.
- So, we need to find x_j for which, by applying controls $a_i \in [\underline{a}_\ell, \overline{a}_\ell]$, we can place each y_i within $[\underline{y}_i, \overline{y}_i]$:

$$X = \{x : \exists a_\ell \in [\underline{a}_\ell, \overline{a}_\ell]\, \forall i\, f_i(x_1, \ldots, x_n, a_1, \ldots, a_k) \in [\underline{y}_i, \overline{y}_i]\}.$$

- This set is known as the *control solution* to the corresponding interval system of equations $f(x, a) = y$.

114. Situation When We Need to Find the Parameters from the Data

- Sometimes, we do not know these values a_ℓ, we must determine these values from the measurements.
- After each cycle c of measurements, we conclude that:
  - the actual (unknown) value of $x_j^{(c)}$ is in the interval $[\underline{x}_j^{(c)}, \overline{x}_j^{(c)}]$ and
  - the actual value of $y_i^{(c)}$ is in the interval $[\underline{y}_i^{(c)}, \overline{y}_i^{(c)}]$.
- We want to find the set A of all the values a for which $y^{(c)} = f(x^{(c)}, a)$ for some $x^{(c)}$ and $y^{(c)}$:

$$A = \{a : \forall c\, \exists x_j^{(c)} \in [\underline{x}_j^{(c)}, \overline{x}_j^{(c)}]\, \exists y_i^{(c)} \in [\underline{y}_i^{(c)}, \overline{y}_i^{(c)}]\, (f(x^{(c)}, a) = y^{(c)})\}.$$

- This set A is known as the *united solution* to the interval system of equations.

115. Comment About Notations

- In general, in our description:
  - y denotes the desired quantities,
  - x denote easier-to-measure quantities, and
  - a denote parameters of the dependence between these quantities.
- In some cases, we have some information about a, and we need to know x – case of the control solution.
- In other cases, we have some information about x, and we need to know a – case of the united solution.
- As a result, sometimes x's are the unknowns, and sometimes a's are the unknowns.

116. What Can We Do Once We Have Found the Range of Possible Values of a

- Once we have found the set A of possible values of a, we can find the range of possible values of y_i:

$$\{f_i(x_1, \ldots, x_n, a) : x_j \in [\underline{x}_j, \overline{x}_j] \text{ and } a \in A\}.$$

- This is a particular case of the main problem of interval computations.
- Often, we want to make sure that each value y_i lies within the given bounds $[\underline{y}_i, \overline{y}_i]$.
- Then we must find the set X of possible values of x for which $f_i(x, a) \in [\underline{y}_i, \overline{y}_i]$ for all $a \in A$:

$$X = \{x : \forall a \in A\, \forall i\, (f_i(x, a) \in [\underline{y}_i, \overline{y}_i])\}.$$

- This set is known as the *tolerance solution* to the interval system of equations.

117. Sometimes, the Values a May Change

- Up to now, we consider the cases when the values a_ℓ are either fixed, or can be changed by us.
- In practice, these values may change in an unpredictable way.
- For example, these parameters may represent some physical processes that influence y_i's.
- We therefore do not know the exact values of a_ℓ, only the bounds $[\underline{a}_\ell, \overline{a}_\ell]$.
- So, the set A of all possible combinations $a = (a_1, \ldots, a_k)$ is contained in a box:

$$A \subseteq [\underline{a}_1, \overline{a}_1] \times \ldots \times [\underline{a}_k, \overline{a}_k].$$

- For example, the set A can be an ellipsoid.

118. Sometimes, the Values a May Change (cont-d)

- In this case, we can still solve the same two problems whose solutions we described earlier.
- We can solve the main problem of interval computations – the problem of computing the range.
- This way we find the set Y of possible values of y.
- We can also solve the corresponding tolerance problem.
- This way, we find the set of values x that guarantee that each y_i is within the desired interval.

119. Is This All There Is?

- There are also more complex problems.
- However, most interval computation packages support the above four problems:
  - range estimation,
  - finding a control solution,
  - finding a united solution, and
  - finding a tolerance solution.
- We show: in practice, we need to use a different notion of an *algebraic* (equality-type) solution.
- This notion:
  - has been previously proposed and analyzed
  - but is not usually included in interval computations packages.

120. How to Find the Set A?

- We considered the case when the values of the parameter a can change.
- We assumed that we know the set A of possible values of the corresponding parameter vector a.
- But how do we find this set?
- All information comes from measurements.
- The only relation between the parameters a and measurable quantities is the formula $y = f(x, a)$.
- Thus, to find the set A of possible values of a, we need to measure x and y many times; so, we get:
  - the set X of possible values of the vector x and
  - the set Y of possible values of the vector y.
- Based on the sets X and Y, we need to find the A.

121. Independence

- It is reasonable to assume that x and a are *independent* in some reasonable sense.
- Independence notion is well known for probabilities: the probability of x does not depend on a:

$$P(x\,|\,a) = P(x\,|\,a') \text{ for all } a, a'.$$

- In the interval case, we do not know the probabilities, we only know which pairs (x, a) are possible.
- We have a *set* $S \subseteq X \times A$ of possible pairs (x, a).
- So, we arrive at the following definition:
- x and a are *independent* if the set $S_a = \{x : (x, a) \in S\}$ of possible values of x does not depend on a: $S_a = S_{a'}$.

122. What We Can Now Conclude About the Dependence Between A, X, and Y

- **Proposition.** *x and a are independent if and only if S is a Cartesian product, i.e.,*

$$S = s_x \times s_a \text{ for some } s_x \subseteq X \text{ and } s_a \subseteq A.$$

- Thus, the set Y is equal to the range of $f(x, a)$ when $x \in X$ and $a \in A$.
- So, we look for sets A for which

$$Y = f(X, A) \stackrel{\text{def}}{=} \{f(x, a) : x \in X \text{ and } a \in A\}.$$

- This set A is known as an *algebraic* (*formal*, *equality-type*) solution to the interval system of equations.
- This notion was introduced and studied by Nickel, Ratschek, Shary, et al.

123. Proof of the Independence Result

- **Proposition.** *x and a are independent if and only if S is a Cartesian product, i.e.,*

$$S = s_x \times s_a \textit{ for some } s_x \subseteq X \textit{ and } s_a \subseteq A.$$

- If $S = s_x \times s_a$, then $S_a = s_x$ for each a and thus, $S_a = S_{a'}$ for all $a, a' \in A$.
- Vice versa, let us assume that x and a are independent.
- Let us denote the common set $S_a = S_{a'}$ by s_x.
- Let us denote by s_a, the set of all possible values a, i.e., the set of all a for which $(x, a) \in S$ for some x.
- Let us prove that in this case, $S = s_x \times s_a$.
- Indeed, if $(x, a) \in S$, then, by definition of s_x, $x \in S_a = s_x$, and, by definition of s_a, $a \in s_a$.

124. Proof of the Independence Result (cont-d)

- We have shown that $x \in s_x$ and $a \in s_a$.
- Thus, by the definition of the Cartesian product $B \times C$ as the set of all pairs (b, c), $b \in B$, $c \in C$, we have

$$(x, a) \in s_x \times s_a.$$

- Vice versa, let $(x, a) \in s_x \times s_a$, i.e., let $x \in s_x$ and $a \in s_a$.
- By definition of the set s_x, we have $S_a = s_x$, thus $x \in S_a$.
- By definition of the set S_a, this means that $(x, a) \in S$.
- The proposition is proven.

125. What If There Is No Algebraic Solution

- Sometimes, the corresponding problem has no solutions.
- For example, for $f(x, a) = x + a$, with $Y = [-1, 1]$ and $X = [-2, 2]$, there is no solution.
- The width $w(X + A)$ of $X + A$ is always $\geq w(X) = 4$ of X and thus, cannot be equal to $w(Y) = 2$.
- What shall we do in this case?
- Of course, this would not happen if we had the *actual* ranges X and Y.
- So, the fact that we cannot find A means something is wrong with these estimates.
- To find out what can be wrong, let us recall how the ranges can be obtained from the experiments.

126. How Ranges Can Be Obtained From Experiments?

- For example, in the 1-D case, we perform several measurements of the quantity x_1 in different situations.
- Based on the measurement results $x_1^{(c)}$, we conclude that the set of possible values includes

$$[\underline{x}_1^{\approx}, \overline{x}_1^{\approx}], \text{ where } \underline{x}_1^{\approx} \stackrel{\text{def}}{=} \min_c x_1^{(c)} \text{ and } \overline{x}_1^{\approx} \stackrel{\text{def}}{=} \max_c x_1^{(c)}.$$

- Of course, we can also have some values outside this interval.
- Example: for a uniform distribution on $[0,1]$, the interval $[\underline{x}^{\approx}, \overline{x}^{\approx}]$ is narrower than $[0,1]$.
- The fewer measurement we take, the narrower this interval.
- So, to estimate the actual range, we *inflate* the interval $[\underline{x}_1^{\approx}, \overline{x}_1^{\approx}]$.

127. Back to Our Problem: What If There Is No Formal Solution

- That we have a mismatch between X and Y means that one of the intervals was not inflated enough.
- X corresponds to easier-to-measure quantities.
- We can thus measure x many times.
- So, even without inflation, get pretty accurate estimates of the actual range X.
- On the other hand, the values y are difficult to measure.
- For these values, we do not have as many measurement results and thus, there is a need for inflation.
- So, we can safely assume that the range for X is reasonably accurate, but the range of Y needs inflation.
- To make this idea precise, let us formalize what is an inflation.

128. What Is an Inflation: Analysis of the Problem

- We want to define a mapping I that transforms each non-degenerate interval $\mathbf{x} = [\underline{x}, \overline{x}]$ into a wider interval
$$I(\mathbf{x}) \supset \mathbf{x}.$$
- What are the natural properties of this transformation?
- The numerical value x of the corresponding quantity depends:
  - on the choice of the measuring unit,
  - on the choice of the starting point, and
  - sometimes, on the choice of direction.
- Example: we can measure temperature t_C in Celsius,
- We can also use a different measuring unit and a different starting point, and get $t_F = 1.8 \cdot t_C + 32$.

129. What Is an Inflation (cont-d)

- We can use the usual convention and consider the usual signs of the electric charge.
- We could also use the opposite signs – then an electron would be a positive electric charge.
- It is reasonable to require that the result of the inflation transformation does not change if we simply:
  - change the measuring units,
  - change the starting point, and/or
  - change the sign.
- Changing the starting point leads to a new interval $[\underline{x}, \overline{x}] + x_0 = [\underline{x} + x_0, \overline{x} + x_0]$ for some x_0.
- Changing the measuring unit leads to $\lambda \cdot [\underline{x}, \overline{x}] = [\lambda \cdot \underline{x}, \overline{x}]$ for some $\lambda > 0$.
- Changing the sign leads to $-[\underline{x}, \overline{x}] = [-\overline{x}, -\underline{x}]$.

130. What Is an Inflation: Resulting Definition and the Main Result

- So, an *inflation* is a mapping from non-degenerate intervals $\mathbf{x} = [\underline{x}, \overline{x}]$ to $I(\mathbf{x}) \supseteq \mathbf{x}$ such that:
  - for every x_0, we have $I(\mathbf{x} + x_0) = I(\mathbf{x}) + x_0$;
  - for every $\lambda > 0$, we have $I(\lambda \cdot \mathbf{x}) = \lambda \cdot I(\mathbf{x})$; and
  - we have $I(-\mathbf{x}) = -I(\mathbf{x})$.
- **Proposition.** *Every inflation operation has the form*

  $[\widetilde{x} - \Delta, \widetilde{x} + \Delta] \to [\widetilde{x} - \alpha \cdot \Delta, \widetilde{x} + \alpha \cdot \Delta]$ *for some* $\alpha > 1$.
- So how do we find A?
- We want to make sure that $f(X, A)$ is equal to the result of a proper inflation of Y.
- How can we tell that an interval Y' is the result of a proper inflation of Y?

131. Proof of the Inflation Result

- **Proposition.** *Every inflation operation has the form* $[\widetilde{x}-\Delta,\widetilde{x}+\Delta]\to[\widetilde{x}-\alpha\cdot\Delta,\widetilde{x}+\alpha\cdot\Delta]$ *for some* $\alpha>1$.
- It is easy to see that the above operation satisfies all the properties of an inflation.
- Let us prove that, vice versa, every inflation has this form.
- Indeed, for intervals $\mathbf{x}$ of type $[-\Delta,\Delta]$, we have $-\mathbf{x}=\mathbf{x}$, thus $I(\mathbf{x})=I(-\mathbf{x})$.
- On the other hand, due to the sign-invariance, we should have $I(-\mathbf{x})=-I(\mathbf{x})$.
- Thus, for the interval $[\underline{v},\overline{v}]\stackrel{\text{def}}{=}I(\mathbf{x})$, we should have $-[\underline{v},\overline{v}]=[-\overline{v},-\underline{v}]=[\underline{v},\overline{v}]$ and thus, $\underline{v}=-\overline{v}$.
- So, we have $I([-\Delta,\Delta])=[-\Delta'(\Delta),\Delta'(\Delta)]$ for some Δ' depending on Δ

132. Proof of the Inflation Result (cont-d)

- Since we should have $[-\Delta, \Delta] \subset I([-\Delta, \Delta])$, we must have $\Delta'(\Delta) > \Delta$.
- Let us denote $\Delta'(1)$ by α.
- Then, $\alpha > 1$ and $I([-1, 1]) = [-\alpha, \alpha]$.
- By applying scale-invariance, with $\lambda = \Delta$, we can then conclude that

$$I([-\Delta, \Delta]) = [-\alpha \cdot \Delta, \alpha \cdot \Delta].$$

- By applying shift-invariance, with $x_0 = \widetilde{x}$, we get the desired equality

$$I([\widetilde{x} - \Delta, \widetilde{x} + \Delta]) = [\widetilde{x} - \alpha \cdot \Delta, \widetilde{x} + \alpha \cdot \Delta].$$

- The proposition is proven.

133. So How Do We Find A?

- How can we tell that an interval Y' is the result of a proper inflation of Y?
- One can check that this is equivalent to the fact that the difference $Y' - Y$ is a symmetric interval $[-u, u]$.
- Such intervals are known as *extended zeros*; thus:
  - if we cannot find the set A for which $Y = f(X, A)$,
  - we should look for the set A for which the difference $f(X, A) - Y$ is an extended zero.
- What if we have several variables, i.e., $m > 1$?
- In this case, we may have different inflations for different components Y_i of the set Y.
- So, we should look for the set A for which, for all i, the difference $f_i(X, A) - Y_i$ is an extended zero.

4.4. Case of Fuzzy Uncertainty: Every Sufficiently Complex Logic Is Multi-Valued Already

134. A Gap Between Fuzzy Logic and the Traditional 2-Valued Fuzzy Logic

- One of the main ideas behind fuzzy logic is that:
  - in contrast to the traditional 2-valued logic, in which every statement is either true or false,
  - in fuzzy logic, we allow intermediate degrees.
- In other words, fuzzy logic is an example of a *multi-valued* logic.
- This led to a misunderstanding between researchers in fuzzy and traditional logics.
- Fuzzy logic books claim that the 2-valued logic cannot describe intermediate degrees.
- On the other hand, 2-valued logicians criticize fuzzy logic for using "weird" intermediate degrees.

135. What We Do in This Section

- We show that the mutual criticism is largely based on a misunderstanding.
- It *is* possible to describe intermediate degrees in the traditional 2-valued logic.
- However, such a representation is complicated.
- The main advantage of fuzzy techniques is that they provide a simply way of doing this.
- And simplicity is important for applications.
- We also show that the main ideas of fuzzy logic are consistent with the 2-valued foundations.
- Moreover, they naturally appear in these foundations if we try to adequately describe expert knowledge.
- We hope to help researchers from both communities to better understand each other.

136. Source of Multi-Valuedness in Traditional Logic: Gödel's Theorem

- A naive understanding of the 2-valued logic assumes that every statement S is either true or false.
- This is possible in simple situations.
- However, Gödel's showed that this not possible for complex theories.
- Gödel analyzed arithmetic – statements obtained
  - from basic equalities and inequalities between polynomial expressions
  - by propositional connectives &, $\vee$, $\neg$, and quantifiers over natural numbers.
- He showed that it is not possible to have a theory T in which for every statement S, either $T \models S$ or $T \models \neg S$.

137. We Have, in Effect, at Least Three Different Truth Values

- Due to Gödel's theorem, there exist statements S for which $T \not\models S$ and $T \not\models \neg S$. So:
  - while, legally speaking, the corresponding logic is 2-valued,
  - in reality, such a statement S is neither true nor false.
- Thus, we have more than 2 possible truth values.
- At first glance, we have 3 truth values: "true", "false", and "unknown".
- However, different "unknown" statements are not necessarily provably equivalent to each other.
- So, we may have more than 3 truth values.

138. How Many Truth Values Do We Actually Have

- It is reasonable to consider the following equivalence relation between statements A and B:

$$\models (A \Leftrightarrow B)$$

- Equivalence classes with respect to this relation can be viewed as the actual truth values.
- The set of all such equivalence classes is known as the *Lindenbaum-Tarski algebra*.
- Lindenbaum-Tarski algebra shows that any sufficiently complex logic is, in effect, multi-valued.
- However, this multi-valuedness is different from the multi-valuedness of fuzzy logic.
- We show that there is another close-to-fuzzy aspect of multi-valuedness of the traditional logic.

139. Need to Consider Several Theories

- In the previous section, we considered the case when we have a single theory T.
- Gödel's theorem states that:
  - for every given theory T that includes formal arithmetic,
  - there is a statement S that can neither be proven nor disproven in this theory.
- This statement S can neither be proven not disproven based on the axioms of theory T.
- So, a natural idea is to consider additional reasonable axioms that we can add to T.

140. Need to Consider Several Theories (cont-d)

- This is what happened in geometry with the V-th postulate P – that
  - for every line ℓ in a plane and for every point P outside this line,
  - there exists only one line ℓ' which passes through P and is parallel to ℓ.
- It turned out that neither P not $\neg P$ can be derived from all other (more intuitive) axioms of geometry.
- So, a natural solution is to explicitly add this statement as a new axiom.
- If we add its negation, we get Lobachevsky geometry – historically the first non-Euclidean geometry.

141. Need to Consider Several Theories (cont-d)

- A similar thing happened in set theory, with the Axiom of Choice and Continuum Hypothesis.
- They cannot be derived or rejected based on the other (more intuitive) axioms of set theory.
- Thus, they (or their negations) have to be explicitly added to the original theory.
- The new – extended – theory covers more statements that the original theory T.
- However, the same Gödel's theory still applies to the new theory:
  - there are statements that
  - can neither be deduced nor rejected based on this new theory.
- Thus, we need to add one more axiom, etc.

142. We Have a Family of Theories

- So, instead of a *single* theory, it makes sense to consider a *family* of theories $\{T_\alpha\}_\alpha$.
- In the above description, we end up with a family which is *linearly ordered* in the sense that:
  - for every two theories T_α and T_β,
  - either $T_\alpha \models T_\beta$ or $T_\beta \models T_\alpha$.
- However, it is possible that on some stage, different groups of researchers select two different axioms.
- In this case, we will have two theories which are not derivable from each other.
- Thus, we have a family of theories which is not linearly ordered.

143. How's This Applicable to Expert Knowledge?

- We can select only the statements in which experts are 100% sure, and we get one possible theory.
- We can add statements S for which the expert's degree of confidence $d(S)$ exceeds a certain threshold α:

$$\{S : d(S) \geq \alpha\}.$$

- For different α, we get different theories T_α.
- For example, if we select $\alpha = 0.7$, then:
  - For every x for which $\mu_{\text{small}}(x) \geq 0.7$, we consider $S(x)$ ("x is small") to be true.
  - For all other objects x, we consider $S(x)$ to be false.
- Similarly, we only keep "if-then" rules for which the expert's degree of confidence in these rules is ≥ 0.7.

144. Once We Have a Family of Theories, How Can We Describe the Truth of a Statement?

- If we have a single theory T, then for every S:
  - either $T \models S$, i.e., the statement S is true in the theory T,
  - or $T \not\models \neg S$, i.e., S is not true in the theory T.
- In general:
  - to describe whether a statement S is true or not,
  - we should consider the values corresponding to all the theories T_α.
- So, we should consider the whole set

$$\deg(S) \stackrel{\text{def}}{=} \{\alpha : T_\alpha \models S\}.$$

- This set is our degree of belief that S is true – i.e., in effect, the truth value of the statement S.

145. Logical Operations on the New Truth Values

- If a theory T_α implies both S and S', then this theory implies their conjunction $S \,\&\, S'$ as well.
- So, the truth value of the conjunction includes the intersection of truth value sets corresponding to S and S':

$$\deg(S \,\&\, S') \supseteq \deg(S) \cap \deg(S').$$

- Similarly, if a theory T_α implies either S or S', then this theory also implies the disjunction $S \vee S'$.
- Thus, the truth value of the disjunction includes the union of truth value sets corresponding to S and S':

$$\deg(S \vee S') \supseteq \deg(S) \cup \deg(S').$$

146. What Happens in the Simplest Case, When the Theories Are Linearly Ordered?

- If the theories T_α are linearly ordered, then, once $T_\alpha \models S$ and $T_\beta \models T_\alpha$, we also have $T_\beta \models S$.
- Thus, with every T_α, the truth value $\deg(S) = \{\alpha : T_\alpha \models S\}$ includes:
  - with each index α,
  - the indices of all the stronger theories – i.e., all the theories T_β for which $T_\beta \models T_\alpha$.
- In particular, for a finite family of theories, each degree is equal to $D_{\alpha_0} \stackrel{\text{def}}{=} \{\alpha : T_\alpha \models T_{\alpha_0}\}$ for some α_0.
- In terms of the linear order $\alpha \le \beta \Leftrightarrow T_\alpha \models T_\beta$, this degree takes the form $D_{\alpha_0} = \{\alpha : \alpha \le \alpha_0\}$.
- We can thus view α_0 as the degree of truth of the statement S: $\mathrm{Deg}(S) \stackrel{\text{def}}{=} \alpha_0$.

147. Linearly Ordered Case (cont-d)

- In case of expert knowledge, this means that we consider the smallest degree of confidence d for which:
  - we can derive the statement S
  - if we allow all the expert's statements whose degree of confidence is at least d.
- These sets D_α are also linearly ordered: one can easily show that $D_\alpha \subseteq D_\beta \Leftrightarrow \alpha \le \beta$.
- The intersection of sets D_α and D_β simply means that we consider the set $D_{\min(\alpha,\beta)}$.
- The union of sets D_α and D_β simply means that we consider the set $D_{\max(\alpha,\beta)}$.
- Thus, the statements about & and $\vee$ take the form:

$$\mathrm{Deg}(S \,\&\, S') \ge \min(\mathrm{Deg}(S), \mathrm{Deg}(S'));$$

$$\mathrm{Deg}(S \vee S') \ge \max(\mathrm{Deg}(S), \mathrm{Deg}(S')).$$

148. Relation to Fuzzy

- We have shown that:

$$\mathrm{Deg}(S \,\&\, S') \geq \min(\mathrm{Deg}(S), \mathrm{Deg}(S'));$$

$$\mathrm{Deg}(S \vee S') \geq \max(\mathrm{Deg}(S), \mathrm{Deg}(S')).$$

- The above formulas are very similar to the formulas of the fuzzy logic corresponding to min and max.
- The only difference is that we get $\geq$ instead of $=$.
- Thus, fuzzy logic ideas can be indeed naturally obtained in the classical 2-valued environment.
- They can be interpreted as a particular case of the same general idea as the Lindenbaum-Tarski algebra.

Part VI: Decision Making under Uncertainty

6.1. Which Point from an Interval Should We Choose: Proof

149. Proof from Section 6.1

- Let us denote the midpoint $\dfrac{\underline{x}+\overline{x}}{2}$ of the interval $[\underline{x},\overline{x}]$ by x_0.
- Then, each point x from this interval can be represented as $x = x_0+\Delta x$, where we denoted $\Delta x \stackrel{\text{def}}{=} x-x_0$.
- The range of possible values of Δx is $[\underline{x}-x_0,\overline{x}-x_0] = [-\Delta,\Delta]$, where we denoted $\Delta \stackrel{\text{def}}{=} \dfrac{\overline{x}-\underline{x}}{2}$.
- The difference Δx is small, so we should be able to keep only the few first terms in Δx.
- When x is known exactly, the optimal decision is $u_{\text{opt}}(x)$.
- Uncertainty is assumed to be small.
- Thus, the optimal decision u should be close to the optimal decision $u_0 \stackrel{\text{def}}{=} u_{\text{opt}}(x_0)$ corresponding to x_0.

150. Proof (cont-d)

- So, the difference $\Delta u \stackrel{\text{def}}{=} u - u_0$ should also be small.
- In terms of Δu, the original value u has the form $u = u_0 + \Delta u$.
- Substituting $x = x_0 + \Delta x$ and $u = u_0 + \Delta u$ into $f(x, u)$ and keeping quadratic terms, we get:

$$f(x, u) = f(x_0 + \Delta x, u_0 + \Delta u) =$$

$$f(x_0, u_0) + f_x \cdot \Delta x + f_u \cdot \Delta u +$$

$$\frac{1}{2} \cdot f_{xx} \cdot (\Delta x)^2 + f_{xu} \cdot \Delta x \cdot \Delta u + \frac{1}{2} \cdot f_{uu} \cdot (\Delta u)^2,$$

$$\text{where } f_x \stackrel{\text{def}}{=} \frac{\partial f}{\partial x}(x_0, u_0), \quad f_u \stackrel{\text{def}}{=} \frac{\partial f}{\partial u}(x_0, u_0),$$

$$f_{xx} \stackrel{\text{def}}{=} \frac{\partial^2 f}{\partial x^2}(x_0, u_0), \quad f_{xu} \stackrel{\text{def}}{=} \frac{\partial^2 f}{\partial x \partial u}(x_0, u_0),$$

$$f_{uu} \stackrel{\text{def}}{=} \frac{\partial^2 f}{\partial u^2}(x_0, u_0).$$

151. Proof (cont-d)

- To find an explicit expression for the objective function, we need to find:
  - the maximum and the minimum of the utility function $f(x, u)$
  - when $x \in [\underline{x}, \overline{x}]$, i.e., when $\Delta x \in [-\Delta, \Delta]$.
- To find the maximum and the minimum of a function of an interval, it is useful to compute its derivative.
- For our utility function we have

$$\frac{\partial f}{\partial x} = f_x + f_{xx} \cdot \Delta x + f_{xu} \cdot \Delta u.$$

- In general, the value f_x is different from 0; we will ignore a possible degenerate case when $f_x = 0$.
- Since Δx and Δu are small, their linear combination is smaller than $|f_x|$.

152. Proof (cont-d)

- Thus, on the whole interval $\Delta x \in [-\Delta, \Delta]$, the sign of the derivative $\dfrac{\partial f}{\partial x}$ is the same as $s_x \stackrel{\text{def}}{=} \text{sign}(f_x)$.
- When $f_x > 0$ and $s_x = +1$, then:
  - the function $f(x, u)$ is an increasing function of x;
  - its maximum is attained when x is attained its largest possible values $\overline{x}$, i.e., when $\Delta x = \Delta$, and
  - its minimum is attained when $\Delta x = -\Delta$.
- When $f_x < 0$ and $s_x = -1$, then:
  - the function $f(x, u)$ is an decreasing function of x;
  - its maximum is attained when x is attained its smallest possible values $\underline{x}$, i.e., when $\Delta x = -\Delta$,
  - and its minimum is attained when $\Delta x = \Delta$.

- In both cases, the maximum of $f(x,u)$ is attained when $\Delta x = s_x \cdot \Delta$ and its minimum when $\Delta x = -s_x \cdot \Delta$:

$$\max_{x\in[\underline{x},\overline{x}]} f(x,u) = f(x_0 + s_x \cdot \Delta, u_0 + \Delta u) =$$

$$f(x_0,u_0) + f_x \cdot s_x \cdot \Delta + f_u \cdot \Delta u +$$

$$\frac{1}{2} \cdot f_{xx} \cdot (\Delta)^2 + f_{xu} \cdot s_x \cdot \Delta \cdot \Delta u + \frac{1}{2} \cdot f_{uu} \cdot (\Delta u)^2;$$

$$\min_{x\in[\underline{x},\overline{x}]} f(x,u) = f(x_0 - s_x \cdot \Delta, u_0 + \Delta u) =$$

$$f(x_0,u_0) - f_x \cdot s_x \cdot \Delta + f_u \cdot \Delta u +$$

$$\frac{1}{2} \cdot f_{xx} \cdot (\Delta)^2 - f_{xu} \cdot s_x \cdot \Delta \cdot \Delta u + \frac{1}{2} \cdot f_{uu} \cdot (\Delta u)^2.$$

- Therefore, our objective function takes the form

$$\alpha \cdot \max_{x\in[\underline{x},\overline{x}]} f(x,u) + (1-\alpha) \cdot \min_{x\in[\underline{x},\overline{x}]} f(x,u) =$$

$$f(x_0,u_0) + (2\alpha - 1) \cdot f_x \cdot s_x \cdot \Delta + f_u \cdot \Delta u +$$

$$\frac{1}{2} \cdot f_{xx} \cdot (\Delta)^2 + (2\alpha - 1) \cdot f_{xu} \cdot s_x \cdot \Delta \cdot \Delta u + \frac{1}{2} \cdot f_{uu} \cdot (\Delta u)^2.$$

153. Proof (cont-d)

- Let us differentiate this expression with respect to u and equate the derivative to 0:

$$f_u + (2\alpha - 1) \cdot f_{xu} \cdot s_x \cdot \Delta + f_{uu} \cdot \Delta u_{\max} = 0.$$

- Thus, $\Delta u_{\max} = -\dfrac{f_u + (2\alpha - 1) \cdot f_{xu} \cdot s_x \cdot \Delta}{f_{uu}}.$
- Let us now take into account that for each x, $f(x, u)$ attains its maximum at the known value $u_{\text{opt}}(x)$.
- Differentiating $f(x, u)$ with respect to u and equating the derivative to 0, we get:

$$f_u + f_{xu} \cdot \Delta x + f_{uu} \cdot \Delta u = 0.$$

- For $x = x_0$, i.e., when $\Delta x = 0$, this maximum is attained when $u = u_0$, i.e., when $\Delta u = 0$.

154. Proof (cont-d)

- Substituting $\Delta x = 0$ and $\Delta u = 0$ into the above formula, we conclude that $f_u = 0$, and thus,

$$\Delta u_{\max} = -\frac{(2\alpha - 1) \cdot f_{xu} \cdot s_x \cdot \Delta}{f_{uu}}.$$

- In general, we can similarly expand $u_{\text{opt}}(x)$ in Taylor series and keep only a few first terms in this expansion:

$$u_{\text{opt}}(x) = u_{\text{opt}}(x_0 + \Delta x) = u_0 + u_x \cdot \Delta x, \text{ where } u_x \stackrel{\text{def}}{=} \frac{\partial u_{\text{opt}}}{\partial x}.$$

- Thus, for the optimal decision, $\Delta u = u_{\text{opt}}(x) - u_0 = u_x \cdot \Delta x$.
- Substituting this expression and $f_u = 0$ into the above formula, we get $f_{xu} \cdot \Delta x + f_{uu} \cdot u_x \cdot \Delta x = 0$ for all Δx.
- Thus, $f_{xu} + f_{uu} \cdot u_x = 0$, and $\dfrac{f_{xu}}{f_{uu}} = -u_x$.

155. Proof (cont-d)

- Substituting this expression for u_x into the formula for $\Delta u_{\max}$, we conclude that:

$$\Delta u_{\max} = (2\alpha - 1) \cdot u_x \cdot s_x \cdot \Delta.$$

- Let us describe this solution in terms of the value $s \in [\underline{x}, \overline{x}]$ for which $u(s) = u_0 + \Delta u_{\max}$.
- Since $s \in [\underline{x}, \overline{x}]$, we can represent s as $s = x_0 + \Delta s$, where $\Delta s \stackrel{\text{def}}{=} s - x_0$.
- Thus, $u(s) = u(x_0 + \Delta s) = u_0 + u_x \cdot \Delta s$.
- Equating $u(s)$ and $u_0 + \Delta u_{\max}$ and using the above formula for $\Delta u_{\max}$, we conclude that

$$u_x \cdot \Delta s = (2\alpha - 1) \cdot u_x \cdot s_x \cdot \Delta.$$

156. Proof (final)

- We get: $u_x \cdot \Delta s = (2\alpha - 1) \cdot u_x \cdot s_x \cdot \Delta$.
- Thus, $\Delta s = (2\alpha - 1) \cdot s_x \cdot \Delta$, where $\Delta = \dfrac{\overline{x} - \underline{x}}{2}$.
- Here, $s = x_0 + \Delta s = \dfrac{\underline{x} + \overline{x}}{2} + \Delta s$.
- If $f_x > 0$ and $s_x = +1$, then

$$s = x_0 + \Delta s = \frac{\underline{x} + \overline{x}}{2} + (2\alpha - 1) \cdot \frac{\overline{x} - \underline{x}}{2} =$$

$$\alpha \cdot \overline{x} + (1 - \alpha) \cdot \underline{x}.$$

- If $f_x < 0$, then we similarly get

$$s = \alpha \cdot \underline{x} + (1 - \alpha) \cdot \overline{x}. \text{ Q.E.D.}$$

6.2. What Decision to Make In a Conflict Situation under Interval Uncertainty: Efficient Algorithms for the Hurwicz Approach

157. How Conflict Situations Are Usually Described

- In many practical situations – e.g., in security – we have conflict situations.
- For example, a terrorist group wants to attack one of our assets, while we want to defend them.
- A *zero-sum game* is when gain of one side is the loss of another side.
- For each possible pair of strategies (i, j), let u_{ij} be the gain of the first side (negative if this is a loss).
- Then, the gain of the second side is $v_{ij} = -u_{ij}$.
- While zero-sum games are a useful approximation, they are not always a perfect description of the situation.

158. Describing Conflict Situations (cont-d)

- For example, the main objective of the terrorists may be publicity, so:
  - a small attack in the country's capital may not cause much damage but bring media attention,
  - a serious attack in a remote are may be more damaging but not as media-attractive.
- To take this difference into account, we need, for each pair of strategies (i, j), to describe both:
  - the gain u_{ij} of the first side and
  - the gain v_{ij} of the second side.
- In general, we do not necessarily have $v_{ij} = -u_{ij}$.

159. It Is Often Beneficial to Act Randomly

- If we only one security person and two objects to protect, then we can:
  - post this person at the first objects and
  - post him/her at the second object.
- If we follow one of these strategies, then the adversary will attack the other (unprotected) object.
- It is thus more beneficial to assign the security person to one of the objects at random.
- This way, for each object of attack, there will be a 50% probability that this object will be defended.
- In general, the first side's strategy can be described by the probabilities $p_1, \ldots, p_n$ of selecting an arrangement:

$$\sum_{i=1}^{n} p_i = 1.$$

160. Toward Precise Formulation of the Problem

- Similarly, the second side selects probabilities $q_1, \ldots, q_m$ for which $\sum\limits_{j=1}^{m} q_j = 1$.
- The expected gains of the two sides are:

$$g_1(p,q) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i \cdot q_j \cdot u_{ij} \text{ and } g_2(p,q) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i \cdot q_j \cdot v_{ij}.$$

- Once the 1st side selects the probabilities p_i, the 2nd side knows them – simply by observing the past history.
- So, the 2nd side selects a strategy q for which its gain is the largest possible:

$$g_2(p, q(p)) = \max_q g_2(p,q).$$

- Similarly, the 2nd side select a strategy q for which

$$g_2(p(q), q) \to \max_q, \text{ where } p(q) \stackrel{\text{def}}{=} \arg\max_p g_1(p,q).$$

161. Towards an Algorithm for Solving this Problem

- Once the strategy p is selected, the 2nd side selects q that maximizes $g_2(p,q)$.
- The expression $g_2(p,q)$ is linear in terms of q_j.
- Thus, $g_2(p,q)$ is the convex combination of gains corr. to deterministic strategies:

$$g_2(p,q)=\sum_{j=1}^{m} q_j\cdot q_{2j}(p),\ \text{ where } g_{2j}(p)\stackrel{\text{def}}{=}\sum_{i=1}^{n} p_i\cdot v_{ij}.$$

- So, the largest possible gain is attained when q is a deterministic strategy.
- The j-th strategy is selected it is better than others:

$$\sum_{i=1}^{n} p_i\cdot v_{ij}\ge\sum_{i=1}^{n} p_i\cdot v_{ik} \text{ for all } k\ne j.$$

162. Towards an Algorithm (cont-d)

- For strategies p for which the second side selects the j-th response, the gain of the 1st side is $\sum\limits_{i=1}^{n} p_i \cdot u_{ij}$.
- Among all strategies p with this "j-property", we select the one with max expected gain of the 1st side.
- This can be found by optimizing a linear function under constraints which are linear inequalities.
- It is known that for such *linear programming* problems, there are efficient algorithms.
- Then, we find j for which the gain is the largest.

163. An Algorithm for Solving the Problem

- For each j from 1 to m, we solve the following linear programming (LP) problem:

$$\sum_{i=1}^{n} p_i^{(j)} \cdot u_{ij} \to \max_{p_i^{(j)}} \text{ under the constraints}$$

$$\sum_{i=1}^{n} p_i^{(j)} = 1, \;\; p_i^{(j)} \geq 0, \;\; \sum_{i=1}^{n} p_i^{(j)} \cdot v_{ij} \geq \sum_{i=1}^{n} p_i^{(j)} \cdot v_{ik} \text{ for all } k \neq j.$$

- We then select $p^{(j)} = \left(p_1^{(j)}, \ldots, p_n^{(j)}\right)$, $1 \leq j \leq m$ for which the value $\sum\limits_{i=1}^{n} p_i^{(j)} \cdot u_{ij}$ is the largest.

- *Comment.* Solution is simpler in zero-sum situations, where we only need to solve one LP problem.

164. Need for Parallelization

- When each side has a small number of strategies, the corresponding problem is easy to solve.
- However, e.g., when we assign air marshals to different international flights, the number of strategies is huge.
- Then, the only way to solve the problem is to perform at least some computations in parallel.
- Good news: all m linear programming problems can be solved on different processors.
- Not so good news: programming problems are P-hard, i.e., provably the hardest to parallelize.

165. Need to Take Uncertainty into Account

- In practice, we rarely know the exact gains u_{ij} and v_{ij}.
- At best, we know the *bounds* on these gains, i.e., we know:
  - the interval $[\underline{u}_{ij}, \overline{u}_{ij}]$ that contains the actual (unknown) values u_{ij}, and
  - the interval $[\underline{v}_{ij}, \overline{v}_{ij}]$ that contains the actual (unknown) values v_{ij}.
- It is therefore necessary to decide what to do in such situations of interval uncertainty.

166. How Interval Uncertainty is Taken into Account Now

- In the above description of a conflict situation, we mentioned that:
  - when we select the strategy p,
  - we maximize the worst-case situation, i.e., the smallest possible gain $\min\limits_q g_1(p,q)$.
- It seems reasonable to apply the same idea to the case of interval uncertainty.
- So, we maximize the smallest possible gain $g_1(p,q)$:
  - over all possible strategies q of the 2nd side *and*
  - over all possible values $u_{ij} \in [\underline{u}_{ij}, \overline{u}_{ij}]$.
- Efficient algorithms for such worst-case formulation have indeed been proposed.

167. Need for a More Adequate Solution

- The adversary wants to minimize our gain, so the worst-case approach makes sense.
- For interval uncertainty, the most adequate idea is to select the alternative a that maximizes:

$$u^H(a) \stackrel{\text{def}}{=} \alpha \cdot \overline{u}(a) + (1 - \alpha) \cdot \underline{u}(a).$$

- Here $\alpha \in [0, 1]$ describes the decision maker's attitude.
- This expression was first proposed by Polish-American Nobelist Leonid Hurwicz.
- For $\alpha = 0$, we optimize the worst-case value $\underline{u}(a)$.
- For other values α, we have different optimization problems.

168. Analysis of the Problem

- Once both sides select strategies p and q, the gain of the 2nd side can take any value from
$$\underline{g}_2(p,q)=\sum_{i=1}^n\sum_{j=1}^m p_i\cdot q_j\cdot \underline{v}_{ij} \text{ to } \overline{g}_2(p,q)=\sum_{i=1}^n\sum_{j=1}^m p_i\cdot q_j\cdot \overline{v}_{ij}.$$
- According to Hurwicz's approach, the 2nd side selects a strategy q that maximizes
$$g_2^H(p,q)\stackrel{\text{def}}{=}\alpha_v\cdot\overline{g}_2(p,q)+(1-\alpha_v)\cdot\underline{g}_2(p,q).$$
- This expression can be represented as $g_2^H(p,q)=\sum\limits_{i=1}^n\sum\limits_{j=1}^m p_i\cdot q_j\cdot v_{ij}^H$, where $v_{ij}^H\stackrel{\text{def}}{=}\alpha_v\cdot\overline{v}_{ij}+(1-\alpha_v)\cdot\underline{v}_{ij}$.
- Under the above strategy $q=q(p)$ of the 2nd side, the 1st side gains the value $g_1(p,q(p))=\sum\limits_{i=1}^n\sum\limits_{j=1}^m p_i\cdot q_j\cdot u_{ij}$.

169. Analysis of the Problem (cont-d)

- We do not know the exact values u_{ij}, we only know the bounds $\underline{u}_{ij} \le u_{ij} \le \overline{u}_{ij}$.
- So, all we know is that this gain will be between

$$\underline{g}_1(p, q(p)) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i \cdot q_j \cdot \underline{u}_{ij} \text{ and } \overline{g}_1(p, q(p)) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i \cdot q_j \cdot \overline{u}_{ij}.$$

- According to Hurwicz's approach, the 1st side should select a strategy p that maximizes

$$g_1^H(p, q) \stackrel{\text{def}}{=} \alpha_u \cdot \overline{g}_1(p, q(p)) + (1 - \alpha_u) \cdot \underline{g}_1(p, q(p)).$$

- This expression has the form $g_1^H(p, q) = \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} p_i \cdot q_j \cdot u_{ij}^H$, where $u_{ij}^H \stackrel{\text{def}}{=} \alpha_u \cdot \overline{u}_{ij} + (1 - \alpha_u) \cdot \underline{u}_{ij}$.
- The resulting optim. problem is the same as in the no-uncertainty case, with u_{ij}^H, v_{ij}^H instead of u_{ij}, v_{ij}.

170. What Is Known: Reminder

- For every deterministic strategy i of the 1st side and for every deterministic strategy j of the 2nd side:
  - we know the interval $[\underline{u}_{ij}, \overline{u}_{ij}]$ of the possible values of the gain of the 1st side, and
  - we know the interval $[\underline{v}_{ij}, \overline{v}_{ij}]$ of the possible values of the gain of the 2nd side.
- We also know the parameters α_u and α_v characterizing decision making of each side under uncertainty.

171. Algorithm for Solving Conflict Situation under Hurwicz-Type Interval Uncertainty

- First, we compute the values

$$u^H_{ij} \stackrel{\text{def}}{=} \alpha_u \cdot \overline{u}_{ij} + (1-\alpha_u)\cdot \underline{u}_{ij} \text{ and } v^H_{ij} \stackrel{\text{def}}{=} \alpha_v \cdot \overline{v}_{ij} + (1-\alpha_v)\cdot \underline{v}_{ij}.$$

- For each j, we solve the following linear programming problem:

$$\sum_{i=1}^n p_i^{(j)} \cdot u^H_{ij} \to \max_{p_i^{(j)}} \text{ under the constraints}$$

$$\sum_{i=1}^n p_i^{(j)} = 1,\ \ p_i^{(j)} \ge 0,\ \ \sum_{i=1}^n p_i^{(j)} \cdot v^H_{ij} \ge \sum_{i=1}^n p_i^{(j)} \cdot v^H_{ik} \text{ for all } k \ne j.$$

- We select a solution $p^{(j)} = \left(p_1^{(j)}, \ldots, p_n^{(j)}\right)$ that maximizes $\sum\limits_{i=1}^n p_i^{(j)} \cdot u^H_{ij}$.

172. Zero-Sum Case

- In the no-uncertainty case, zero-sum games are easier to process.
- Let us consider situations in which possible values v_{ij} are exactly values $-u_{ij}$ for possible u_{ij}:

$$[\underline{v}_{ij}, \overline{v}_{ij}] = \{-u_{ij} : u_{ij} \in [\underline{u}_{ij}, \overline{u}_{ij}]\}.$$

- In this case, $\underline{v}_{ij} = -\overline{u}_{ij}$ and $\overline{v}_{ij} = -\underline{u}_{ij}$.
- Then, $v_{ij}^H = \alpha_v \cdot \overline{v}_{ij} + (1 - \alpha_v) \cdot \underline{v}_{ij}$ and $u_{ij}^H = \alpha_u \cdot \overline{u}_{ij} + (1 - \alpha_u) \cdot \underline{u}_{ij}$.
- One can check that the resulting game is zero-sum (i.e., $v_{ij}^H = -u_{ij}^H$) only when $\alpha_u = 1 - \alpha_v$.
- In all other cases, the general algorithm will be needed, without a zero-sum simplification.

173. Conclusion to this Section

- In this section, we show how to take interval uncertainty into account when solving conflict situations.
- Such algorithms are known when each side of the conflict maximizes its worst-case expected gain.
- A general Hurwicz approach provides a more adequate description of decision making under uncertainty.
- In this approach, each side maximizes the convex combination of the worst-case and the best-case gains.
- We describe how to resolve conflict situations under the general Hurwicz approach to interval uncertainty.

6.3. Why Unexpectedly Positive Experiences Make Decision Makers More Optimistic: An Explanation

174. Formulation of the Problem

- Experiments show that unexpectedly positive experiences make decision makers more optimistic.
- This was first observed on rats: rats like being ticked, and tickled rats became more optimistic.
- Several later papers showed that the same phenomenon holds for other decision making situations as well.
- Similarly, decision makers who had an unexpectedly negative experiences became more pessimistic.
- There seems to be no convincing explanation for this experimental fact.
- We show that this phenomenon can be explained in the traditional utility-based decision theory.

175. What Does Optimism Mean?

- Traditional decision theory assumes that we know the probabilities of all possible consequences of each action.
- Then, a rational decision maker maximizes the expected value $u(a)$ of a special function called *utility*.
- In this case, there is no such thing as optimism or pessimism: we just select the best alternative a.
- In practice, we often have only *partial* information about these probabilities.
- In such situations, there are several possible probability distributions consistent with our knowledge.
- For different distributions, we have, in general, different values of the expected utility.
- As a result, for each alternative a, we have an *interval* $[\underline{u}(a), \overline{u}(a)]$ of possible values of $u(a)$.

176. What Does Optimism Mean (cont-d)

- In this case, we should select an alternative a that maximizes $u(a) = \alpha \cdot \overline{u}(a) + (1 - \alpha) \cdot \underline{u}(a)$.
- This idea was proposed by the Nobelist Leo Hurwicz.
- The selection of α, depends on the person.
- The value $\alpha = 1$ means that the decision maker only takes into account the best possible consequences.
- In other words, the values $\alpha = 1$ corresponds to complete optimism.
- Similarly, the value $\alpha = 0$ corresponds to complete pessimism.
- The larger α, the close this decision maker to complete optimism.
- The *optimism-pessimism index* α is a numerical measure of the decision maker's optimism.

177. What Does Optimism Mean (cont-d)

- The *optimism-pessimism index* α is a numerical measure of the decision maker's optimism.
- Thus, the phenomenon to-be-explained takes the following precise meaning:
  - if a decision maker has unexpectedly positive experiences, then this decision maker's α increases;
  - if a decision maker has unexpectedly negative experiences, then this decision maker's α decreases.

178. α Can Be Interpreted as the Subjective Probability of Positive Outcome

- The decision maker selects an alternative a that maximizes $\alpha \cdot \overline{u}(a) + (1 - \alpha) \cdot \underline{u}(a)$.
- Here, $\overline{u}(a)$ corresponds to the positive outcome, and $\underline{u}(a)$ corresponds to negative outcome.
- For simplicity, let us consider the situation when we have only two possible outcomes:
  - the positive outcome, with utility $\overline{u}(a)$, and
  - the negative outcome, with utility $\underline{u}(a)$.
- A traditional approach to decision making assumes that we know the probabilities of different outcomes.
- In this case of uncertainty, we do not know the actual (objective) probabilities.

179. α Can Be Interpreted as the Subjective Probability of Positive Outcome (cont-d)

- In the case of uncertainty, we do not know the actual (objective) probabilities.
- However, we can always come up with estimated (subjective) ones.
- Let us denote the subjective probability of the positive outcome by p_+.
- Then, the subjective probability of the negative outcome is equal to $1 - p_+$.
- The expected utility is equal to $p_+ \cdot \overline{u}(a) + (1 - p_+) \cdot \underline{u}(a)$.
- This is exactly what we optimize when we use Hurwicz's approach, with $\alpha = p_+$.
- Thus, the value α can be interpreted as the subjective probability of the positive outcome.

180. A New Reformulation of Our Problem

- Unexpectedly positive experiences increase the subjective probability of a positive outcome.
- Unexpectedly negative experiences decrease the subjective probability of a positive outcome.
- To explain this phenomenon, let us recall where subjective probabilities come from.
- If we observe an event in n out of N cases, our estimate is n/N.
- Example: if a coin fell heads 6 times out of 10, we estimate the probability of it falling heads as $6/10$.
- Let us show that this leads to the desired explanation.

181. Resulting Explanation

- Suppose that a decision maker had n positive experiences in the past N situations.
- Then, the decision maker's subjective probability of a positive outcome is $p_+ = n/N$.
- Unexpectedly positive experiences means that:
  - we have a series of new experiments,
  - in which the fraction of positive outcomes was higher than the expected frequency p_+.
- In other words, unexpectedly positive experiences means that $n'/N' > p$, where:
  - N' is the overall number of new experiences, and
  - n' is the number of those new experiences in which the outcome turned out to be positive.

182. Resulting Explanation (cont-d)

- The new subjective probability p'_+ is equal to the new ratio $p'_+ = \dfrac{n+n'}{N+N'}$.
- Here, by definition of p_+, we have $n = p_+ \cdot N$.
- Due to unexpected positiveness of new experiences, we have $n' > p_+ \cdot N'$.
- By adding this inequality and the previous equality, we conclude that $n + n' > p_+ \cdot (N + N')$, i.e., that

$$p'_+ = \frac{n+n'}{N+N'} > p_+.$$

- In other words, unexpectedly positive experiences increase the subjective probability of a positive outcome.

183. Resulting Explanation (final)

- The subjective probability of the positive outcome is exactly the optimism-pessimism coefficient α.
- Thus, $p'_+ > p_+$ means that $\alpha' > \alpha$.
- So, unexpectedly positive experiences make the decision maker more optimistic.
- Similarly, if we had unexpectedly negative experiences, i.e., $n' < p_+ \cdot N'$, then $p'_+ = \dfrac{n+n'}{N+N'} < p_+$ and $\alpha' < \alpha$.
- So, we conclude that unexpectedly negative experiences make the decision maker less optimistic.
- This is also exactly what we observe.
- So, we have the desired explanation.