Fuzzy Logic Can Justify and Improve Semi-Heuristic Data and Image Processing Techniques: Main Idea and Case Studies

Vladik Kreinovich
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
vladik@utep.edu
http://www.cs.utep.edu/vladik

1. Traditional Use of Fuzzy Logic

- Expert knowledge is often formulated by using imprecise ("fuzzy") from natural language (like "small").
- Fuzzy logic techniques was originally invented to translate such knowledge into precise terms.
- Such a translation is still the main use of fuzzy techniques.
- *Example:* we want to control a complex plant for which:
  - no good control technique is known, but
  - there are experts how can control this plant reasonably well.
- So, we elicit rules from the experts.
- Then we use fuzzy techniques to translate these rules into a control strategy.

2. Fuzzy Logic Can Help in Other Cases As Well

- Lately, it turned out that fuzzy techniques can help in another class of applied problems: in situations when
  - there are semi-heuristic techniques for solving the corresponding problems, i.e.,
  - techniques for which there is no convincing theoretical justification.
- These techniques lack theoretical justification.
- Their previous empirical success does not guarantee that these techniques will work well on new problems.
- Thus, users are reluctant to use these techniques.

3. Additional Problem of Semi-Heuristic Techniques

- Semi-heuristic techniques are often not perfect.
- Without an underlying theory, it is not clear how to improve their performance.
- For example, linear models can be viewed as first approximation to Taylor series.
- So, a natural next approximation is to use quadratic models.
- However, e.g., for ℓ^p-models:
  - when they do not work well,
  - it is not immediately clear what is a reasonable next approximation.

4. What We Show

- We show that in many such situations, the desired theoretical justification can be obtained if:
  - in addition to known (crisp) requirements on the desired solution,
  - we also take into account requirements formulated by experts in natural-language terms.
- Naturally, we use fuzzy techniques to translate these imprecise requirements into precise terms.
- To make the resulting justification convincing, we need to make sure that this justification works:
  - not only for one specific choice of fuzzy techniques (membership function, t-norm, etc.),
  - but for all techniques which are consistent with the practical problem.

5. Case Studies

As examples, we provide the detailed justification of:

- ℓ^p-regularization techniques in solving inverse problems
  - an empirically successful alternative to Tikhonov regularization
  - which is appropriate for situations when the desired signal or image is not smooth;
- sparsity techniques in data and image processing –
  - a very successful hot-topic technique
  - whose success is often largely a mystery; and
- non-linear empirical models of soil mechanics used in road construction.

Part I

Why ℓ_p-methods in Signal and Image Processing: A Fuzzy-Based Explanation

6. Need for Deblurring

- Cameras and other image-capturing devices are getting better and better every day.
- However, none of them is perfect, there is always some blur, that comes from the fact that:
  - while we would like to capture the intensity $I(x, y)$ at each spatial location (x, y),
  - the signal $s(x, y)$ is influenced also by the intensities $I(x', y')$ at nearby locations (x', y'):

$$s(x, y) = \int w(x, y, x', y') \cdot I(x', y')\, dx'\, dy'.$$

- When we take a photo of a friend, this blur is barely visible – and does not constitute a serious problem.
- However, when a spaceship takes a photo of a distant plant, the blur is very visible – so deblurring is needed.

7. In General, Signal and Image Reconstruction Are Ill-Posed Problems

- The image reconstruction problem is *ill-posed* in the sense that:
  - large changes in $I(x, y)$
  - can lead to very small changes in $s(x, y)$.
- Indeed, the measured value $s(x, y)$ is an average intensity over some small region.
- Averaging eliminates high-frequency components.
- Thus, for $I^*(x, y) = I(x, y) + c \cdot \sin(\omega_x \cdot x + \omega_y \cdot y)$, the signal is practically the same: $s^*(x, y) \approx s(x, y)$.
- However, the original images, for large c, may be very different.

8. Need for Regularization

- To reconstruct the image reasonably uniquely, we must impose additional conditions on the original image.
- This imposition is known as *regularization*.
- Often, a signal or an image is smooth (differentiable).
- Then, a natural idea is to require that the vector $d = (d_1, d_2, \ldots)$ formed by the derivatives is close to 0:

$$\rho(d, 0) \le C \Leftrightarrow \sum_{i=1}^{n} d_i^2 \le c \stackrel{\text{def}}{=} C^2.$$

- For continuous signals, sum turns into an integral:

$$\int (\dot{x}(t))^2\, dt \le c \text{ or } \int \left(\left(\frac{\partial I}{\partial x} \right)^2 + \left(\frac{\partial I}{\partial y} \right)^2 \right) dx\, dy \le c.$$

9. Tikhonov Regularization

- Out of all smooth signals or images, we want to find the best fit with observation: $J \stackrel{\text{def}}{=} \sum_i e_i^2 \to \min$.
- Here, e_i is the difference between the actual and the reconstructed values.
- Thus, we need to minimize J under the constraint

$$\int (\dot{x}(t))^2\,dt \le c \text{ and } \int \left(\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2 \right) dx\,dy \le c.$$

- Lagrange multiplier method reduced this constraint optimization problem to the unconstrained one:

$$J + \lambda \cdot \int \left(\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2 \right) dx\,dy \to \min_{I(x,y)}.$$

- This idea is known as *Tikhonov regularization*.

10. From Continuous to Discrete Images

- In practice, we only observe an image with a certain spatial resolution.
- So we can only reconstruct the values $I_{ij} = I(x_i, y_j)$ on a certain grid $x_i = x_0 + i \cdot \Delta x$ and $y_j = y_0 + j \cdot \Delta y$.
- In this discrete case, instead of the derivatives, we have differences:

$$J + \lambda \cdot \sum_i \sum_j ((\Delta_x I_{ij})^2 + (\Delta_y I_{ij})^2) \to \min_{I_{ij}}.$$

- Here:
  - $\Delta_x I_{ij} \stackrel{\text{def}}{=} I_{ij} - I_{i-1,j}$, and
  - $\Delta_y I_{ij} \stackrel{\text{def}}{=} I_{ij} - I_{i,j-1}$.

11. Limitations of Tikhonov Regularization and ℓ^p-Method

- Tikhonov regularization is based on the assumption that the signal or the image is smooth.
- In real life, images are, in general, not smooth.
- For example, many of them exhibit a fractal behavior.
- In such non-smooth situations, Tikhonov regularization does not work so well.
- To take into account non-smoothness, researchers have proposed to modify the Tikhonov regularization:
  - instead of the squares of the derivatives,
  - use the p-th powers for some $p \neq 2$:

$$J + \lambda \cdot \sum_i \sum_j (|\Delta_x I_{ij}|^p + |\Delta_y I_{ij}|^p) \to \min_{I_{ij}}.$$

- This works much better than Tikhonov regularization.

12. Remaining Problem

- *Problem:* the ℓ^p-methods are heuristic.
- There is no convincing explanation of why necessarily we replace the square:
  - with a p-th power and
  - not, for example, with some other function.
- *We show:* that a natural formalization of the corresponding intuitive ideas indeed leads to ℓ^p-methods.
- To formalize the intuitive ideas behind image reconstruction, we use *fuzzy techniques*.
- Fuzzy techniques were designed to transform:
  - imprecise intuitive ideas into
  - exact formulas.

13. Let Us Apply Fuzzy Techniques

- We are trying to formalize the statement that the image is continuous.
- This means that the differences $\Delta x_k \stackrel{\text{def}}{=} \Delta_x I_{ij}$ and $\Delta_y I_{ij}$ between image intensities at nearby points are small.
- Let $\mu(x)$ denote the degree to which x is small, and $f_\&(a,b)$ denote the "and"-operation.
- Then, the degree d to which Δx_1 is small *and* Δx_2 is small, etc., is:

$$d = f_\&(\mu(\Delta x_1), \mu(\Delta x_2), \mu(\Delta x_3), \ldots).$$

- *Known:* each "and"-operation can be approximated, for any $\varepsilon > 0$, by an *Archimedean* one:

$$f_\&(a,b) = f^{-1}(f(a)) \cdot f(b)).$$

- Thus, without losing generality, we can safely assume that the actual "and"-operation is Archimedean.

14. Analysis of the Problem

- We want to select an image with the largest degree of satisfying this condition:

$$d = f^{-1}(f(\mu(\Delta x_1)) \cdot f(\mu(\Delta x_2)) \cdot f(\mu(\Delta x_3)) \cdot \ldots) \to \max .$$

- Since the function $f(x)$ is increasing, maximizing d is equivalent to maximizing

$$f(d) = f(\mu(\Delta x_1)) \cdot f(\mu(\Delta x_2)) \cdot f(\mu(\Delta x_3)) \cdot \ldots$$

- Maximizing this product is equivalent to minimizing its negative logarithm

$$L \stackrel{\text{def}}{=} -\ln(d) = \sum_k g(\Delta x_k), \text{ where } g(x) \stackrel{\text{def}}{=} -\ln(f(\mu(x))).$$

- In these terms, selecting a membership function is equivalent to selecting the related function $g(x)$.

15. Which Function $g(x)$ Should We Select: Idea

- The value $\Delta x_i = 0$ is small, so $\mu(0) = 1$ and $g(0) = -\ln(1) = 0$.
- The numerical value of a difference Δx_i depends on the choice of a measuring unit.
- If we choose a measuring unit (MU) which is a times smaller, then $\Delta x_i \to a \cdot \Delta x_i$.
- It's reasonable to request that the requirement $\sum\limits_k g(\Delta x_k) \to \min$ not change if we change MU.
- For example, if $g(z_1) + g(z_2) = g(z'_1) + g(z'_2)$, then

$$g(a \cdot z_1) + g(a \cdot z_2) = g(a \cdot z'_1) + g(a \cdot z'_2).$$

16. Main Result

- *Reminder:* selecting the most reasonable values of Δx_k ($d \to \max$) is equivalent to $\sum_k g(\Delta x_k) \to \min$.
- *Main condition:* we are looking for a function $g(x)$ for which $g(z_1) + g(z_2) = g(z'_1) + g(z'_2)$, then

$$g(a \cdot z_1) + g(a \cdot z_2) = g(a \cdot z'_1) + g(a \cdot z'_2).$$

- *Main result:* $g(a) = C \cdot a^p + \text{const}$, for some $p > 0$.
- *Fact:* minimizing $\sum_k g(\Delta x_k)$ is equivalent to minimizing the sum $\sum_k |\Delta x_k|^p$.
- *Fact:* minimizing $\sum_k |\Delta x_k|^p$ under condition $J \le c$ is equivalent to minimizing $J + \lambda \cdot \sum_k |\Delta x_k|^p$.
- *Conclusion:* fuzzy techniques indeed justify ℓ^p-method.

17. Proof

- We are looking for a function $g(x)$ for which $g(z_1) + g(z_2) = g(z_1') + g(z_2')$, then
$$g(a \cdot z_1) + g(a \cdot z_2) = g(a \cdot z_1') + g(a \cdot z_2').$$
- Let us consider the case when $z_1' = z_1 + \Delta z$ for a small Δz, and $z_2' = z_2 + k \cdot \Delta z + o(\Delta z)$ for an appropriate k.
- Here, $g(z_1 + \Delta z) = g(z_1) + g'(z_1) \cdot \Delta z + o(\Delta z)$, so $g'(z_1) + g'(z_2) \cdot k = 0$ and $k = -\dfrac{g'(z_1)}{g'(z_2)}$.
- The condition $g(a \cdot z_1) + g(a \cdot z_2) = g(a \cdot z_1') + g(a \cdot z_2')$ similarly takes the form $g'(a \cdot z_1) + g'(z_2) \cdot k = 0$, so
$$g'(a \cdot z_1) - g'(a \cdot z_2) \cdot \frac{g'(z_1)}{g'(z_2)} = 0.$$
- Thus, $\dfrac{g'(a \cdot z_1)}{g'(z_1)} = \dfrac{g'(a \cdot z_2)}{g'(z_2)}$ for all a, z_1, and z_2.

18. Proof (cont-d)

- *Reminder:* $\dfrac{g'(a \cdot z_1)}{g'(z_1)} = \dfrac{g'(a \cdot z_2)}{g'(z_2)}$ for all z_1 and z_2.
- This means that the ratio $\dfrac{g'(a \cdot z_1)}{g'(z_1)}$ does not depend on z_i: $\dfrac{g'(a \cdot z_1)}{g'(z_1)} = F(a)$ for some $F(a)$.
- For $a = a_1 \cdot a_2$, we have

$$F(a) = \frac{g'(a \cdot z_1)}{g'(z_1)} = \frac{g'(a_1 \cdot a_2 \cdot z_1)}{g'(z_1)} =$$

$$\frac{g'(a_1 \cdot (a_2 \cdot z_1))}{g'(a_2 \cdot z_1)} \cdot \frac{g'(a_2 \cdot z_1)}{g'(z_1)} = F(a_1) \cdot F(a_2).$$

- So, $F(a_1 \cdot a_2) = F(a_1) \cdot F(a_2)$, thus $F(a) = a^q$ for some real number q.
- $\dfrac{g'(a \cdot z_1)}{g'(z_1)} = F(a)$ becomes $g'(a \cdot z_1) = g'(z_1) \cdot a^q$.

19. Proof (final part)

- *Reminder:* we have $g'(a \cdot z_1) = g'(z_1) \cdot a^p$.
- For $z_1 = 1$, we get $g'(a) = C \cdot a^q$, where $C \stackrel{\text{def}}{=} g'(1)$.
- We could have $q = -1$ or $q \neq -1$.
- For $q = -1$, we get $g(a) = C \cdot \ln(a) + \text{const}$, which contradicts to $g(0) = 0$.
- Integrating, for $q \neq -1$, we get

$$g(a) = \frac{C}{q+1} \cdot a^{q+1} + \text{const}.$$

- The main result is proven.

Part II

Why Sparse? Fuzzy Techniques Explain Empirical Efficiency of Sparsity-Based Data- and Image-Processing Algorithms

20. Sparsity Is Useful, But Why?

- In many practical applications, it turned out to be efficient to assume that the signal or an image is *sparse*:
  - when we decompose the original signal $x(t)$ (or image) into appropriate basic functions $e_i(t)$:

$$x(t) = \sum_{i=1}^{\infty} a_i \cdot e_i(t),$$

  - then most of the coefficients a_i in this decomposition will be zeros.
- It is often beneficial to select, among all the signals consistent with the observations, the signal for which

$$\#\{i : a_i \neq 0\} \to \min \quad \text{or} \quad \sum_{i:a_i \neq 0} w_i \to \min .$$

- At present, the empirical efficiency of sparsity-based techniques remains somewhat a mystery.

21. Before We Perform Data Processing, We First Need to Know Which Inputs Are Relevant

- In general, in data processing, we:
  - estimate the value of the desired quantity y_j based on
  - the values of the known quantities $x_1, \ldots, x_n$ that describe the current state of the world.
- In principle, all possible quantities $x_1, \ldots, x_n$ could be important for predicting some future quantities.
- However, for each specific quantity y_j, usually, only a few of the quantities x_i are actually useful.
- So, we first need to check which inputs are actually useful.
- This checking is an important stage of data processing: else we waste time processing unnecessary quantities.

22. Analysis of the Problem

- We are interested in a reconstructing a signal or image $x(t) = \sum_{i=1}^{\infty} a_i \cdot e_i(t)$ based on:
  - the measurement results and
  - prior knowledge.
- First, we find out which quantities a_i are relevant.
- The quantity a_i is irrelevant if it does not affect the resulting signal, i.e., if $a_i = 0$.
- So, first, we decide which values a_i are zeros and which are non-zeros.
- Out of all such possible decisions, we need to select *the most reasonable one.*
- *Problem:* "reasonable" is not a precise term.

23. Let Us Use Fuzzy Logic

- *Reminder:* we want the most reasonable decision, but "reasonable" is not a precise term.
- So, to be able to solve the problem, we need to translate this imprecise description into precise terms.
- Let's use fuzzy techniques which were specifically designed for such translations.
- In fuzzy logic, we assign, to each statement S, our degree of confidence d in S.
- E.g., we ask experts to mark, on a scale from 0 to 10, how confident they are in S.
- If an expert marks the number 7, we take $d = 7/10$.
- Thus, for each i, we can learn to what extent $a_i = 0$ or $a_i \neq 0$ are reasonable.

24. Need for an "And"-Operation

- We want to estimate, for each tuple of signs, to which extent this tuple is reasonable.
- There are 2^n such tuples, so for large n, it is not feasible to ask about all of them.
- We thus need to estimate:
  - the degree to which a_1 is reasonable *and* a_2 is reasonable ...
  - based on individual degrees to which a_i are reasonable.
- In other words:
  - we know the degrees of belief $a = d(A)$ and $b = d(B)$ in statements A and B, and
  - we need to estimate the degree of belief in the composite statement $A \,\&\, B$, as $f_\&(a, b)$.

25. The "And"-Estimate Is Not Always Exact: an Example

- First case:
  - A is "coin falls heads", B is "coin falls tails", then for a fair coin, degrees a and b are equal: $a = b$.
  - Here, $A \,\&\, B$ is impossible, so our degree of belief in $A \,\&\, B$ is zero: $d(A \,\&\, B) = 0$.
- Second case:
  - If we take $A' = B' = A$, then $A' \,\&\, B'$ is simply equivalent to A.
  - So we still have $a' = b' = a$ but this time $d(A' \,\&\, B') = a > 0$.
- In these two cases:
  - we have $d(A') = d(A) = a$ and $d(B') = d(B) = b$,
  - but $d(A \,\&\, B) \neq d(A' \,\&\, B')$.

26. Which "And"-Operation (t-Norm) Should We Choose

- The corresponding function $f_\&(a,b)$ must satisfy some reasonable properties: e.g.,
  - since $A \,\&\, B$ means the same as $B \,\&\, A$, this operation must be commutative;
  - since $(A \,\&\, B) \,\&\, C$ is equivalent to $A \,\&\, (B \,\&\, C)$, this operation must be associative, etc.
- *Known result:* each such operation can be approximated, with any given accuracy,
  - by an *Archimedean* t-norm

$$f_\&(a,b) = f^{-1}(f(a) \cdot f(b)),$$

  - for some strictly increasing function $f(x)$.
- Thus, without losing generality, we can assume that the actual t-norm is Archimedean.

27. Let Us Use Fuzzy Logic

- Let $d_i^{=} \stackrel{\text{def}}{=} d(a_i = 0)$ and $d_i^{\neq} \stackrel{\text{def}}{=} d(a_i \neq 0)$.
- So, for each sequence $(\varepsilon_1, \varepsilon_2, \ldots)$, where ε_i is $=$ or $\neq$:

$$d(\varepsilon) = f_\&(d_1^{\varepsilon_1}, d_2^{\varepsilon_2}, \ldots).$$

- *Problem:*
  - out of all sequences ε which are consistent with the measurements and with the prior knowledge,
  - we must select the one for which this degree of belief is the largest possible.
- If we have no information about the signal, then the most reasonable choice is $x(t) = 0$, i.e.,

$$a_1 = a_2 = \ldots = 0 \text{ and } \varepsilon = (=, =, \ldots).$$

- Similarly, the least reasonable is the sequence in which we take all the values into account, i.e., $\varepsilon = (\neq, \ldots, \neq)$.

28. Definitions

- By a *t-norm*, we mean $f_\&(a,b) = f^{-1}(f(a)\cdot f(b))$, where $f:[0,1]\to[0,1]$ is continuous, $\uparrow$, $f(0)=0$, $f(1)=1$.
- By a *sequence*, we mean a sequence $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_N)$, where each symbol ε_i is equal either to $=$ or to $\neq$.
- Let $d^= = (d^=_1,\ldots,d^=_N)$ and $d^{\neq} = (d^{\neq}_1,\ldots,d^{\neq}_N)$ be sequences of real numbers from the interval $[0,1]$.
- For each sequence ε, we define its *degree of reasonableness* as $d(\varepsilon)\stackrel{\text{def}}{=} f_\&(d_1^{\varepsilon_1},\ldots,d_N^{\varepsilon_N})$.
- We say that the sequences $d^=$ and $d^{\neq}$ *properly describe reasonableness* if the following two conditions hold:
  - for $\varepsilon_= \stackrel{\text{def}}{=} (=,\ldots,=)$, $d(\varepsilon_=) > d(\varepsilon)$ for all $\varepsilon\neq\varepsilon_=$,
  - for $\varepsilon_{\neq} \stackrel{\text{def}}{=} (\neq,\ldots,\neq)$, $d(\varepsilon_{\neq}) < d(\varepsilon)$ for all $\varepsilon\neq\varepsilon_{\neq}$.
- For each set S of sequences, we say that a sequence $\varepsilon\in S$ *is the most reasonable* if $d(\varepsilon)=\max\limits_{\varepsilon'\in S} d(\varepsilon')$.

29. Main Result

- **Proposition.**
  - *Let us assume that the sequences* $d^{=}$ *and* $d^{\neq}$ *properly describe reasonableness.*
  - *Then, there exist weights* $w_i > 0$ *for which, for each set* S*, the following two conditions are equivalent:*
    - *the sequence* $\varepsilon \in S$ *is the most reasonable,*
    - *the sum* $\sum\limits_{i:\varepsilon_i = \neq} w_i = \sum\limits_{i:a_i \neq 0} w_i$ *is the smallest possible.*
- **Discussion:** thus, fuzzy-based techniques indeed naturally lead to the sparsity condition.

30. A Similar Derivation Can Be Obtained in the Probabilistic Case

- Reasonableness can be described by assigning a *probability* $p(\varepsilon)$ to each possible sequence ε.
- Let $p_i^=$ be the probability that $a_i = 0$, and let $p_i^{\neq} = 1 - p_i^=$ be the probability that $a_i \neq 0$.
- We do not know the relation between the values ε_i and ε_j corresponding to different coefficients $i \neq j$.
- So, it makes sense to assume that the corresponding random variables ε_i and ε_j are independent, so

$$p(\varepsilon) = \prod_{i=1}^{N} p_i^{\varepsilon_i}.$$

- So, we arrive at the following definitions.

31. Probabilistic Case: Definitions

- Let $p^= = (p^=_1, \ldots, p^=_N)$ be a sequence of real numbers from the interval $[0,1]$, and let $p^{\neq}_i \stackrel{\text{def}}{=} 1 - p^=_i$.
- For each sequence ε, its *probability* is $p(\varepsilon) \stackrel{\text{def}}{=} \prod\limits_{i=1}^N p_i^{\varepsilon_i}$.
- We say that the sequence $p^=$ *properly describes reasonableness* if the following two conditions are satisfied:
  - the sequence $\varepsilon_= \stackrel{\text{def}}{=} (=, \ldots, =)$ is more probable than all others, i.e., $p(\varepsilon_=) > p(\varepsilon)$ for all $\varepsilon \neq \varepsilon_=$,
  - the sequence $\varepsilon_{\neq} \stackrel{\text{def}}{=} (\neq, \ldots, \neq)$ is less probable than all others, i.e., $p(\varepsilon_{\neq}) < p(\varepsilon)$ for all $\varepsilon \neq \varepsilon_{\neq}$.
- For each set S of sequences, we say that a sequence $\varepsilon \in S$ *is the most probable* if $p(\varepsilon) = \max\limits_{\varepsilon' \in S} p(\varepsilon')$.

32. Probabilistic Case: Main Result

- **Proposition.**
  - *Let us assume that the sequence* $p^=$ *properly describes reasonableness.*
  - *Then, there exist weights* $w_i > 0$ *for which, for each set* S*, the following two conditions are equivalent:*
    - *the sequence* $\varepsilon \in S$ *is the most probable,*
    - *the sum* $\sum\limits_{i:\varepsilon_i = \neq} w_i$ *is the smallest possible.*
- **Discussion.** In other words, probabilistic techniques also lead to the sparsity condition.

33. Fuzzy Approach vs. Probabilistic Approach

- *Fact:* the probabilistic approach leads to the same conclusion as the fuzzy approach.
- *First conclusion:* this makes us more confident that our justification of sparsity is valid.
- *Observation:*
  - the probability-based result is based on the assumption of independence, while
  - the fuzzy-based result can allow different types of dependence – as described by different t-norms.
- *Second conclusion:* this is an important advantage of the fuzzy-based approach.

Part III

How to Improve the Existing Semi-Heuristic Technique

34. Blind Image Deconvolution: Formulation of the Problem

- The measurement results y_k differ from the actual values x_k dues to additive noise and blurring:

$$y_k = \sum_i h_i \cdot x_{k-i} + n_k.$$

- From the mathematical viewpoint, y is a *convolution* of h and x: $y = h \star x$.
- Similarly, the observed image $y(i,j)$ differs from the ideal one $x(i,j)$ due to noise and blurring:

$$y(i,j) = \sum_{i'} \sum_{j'} h(i-i', j-j') \cdot x(i',j') + n(i,j).$$

- It is desirable to reconstruct the original signal or image, i.e., to perform *deconvolution*.

35. Ideal No-Noise Case

- In the ideal case, when noise $n(i,j)$ can be ignored, we can find $x(i,j)$ by solving a system of linear equations:

$$y(i,j)=\sum_{i'}\sum_{j'} h(i-i',j-j')\cdot x(i',j').$$

- However, already for 256×256 images, the matrix h is of size 65,536×65,536, with billions entries.
- Direct solution of such systems is not feasible.
- A more efficient idea is to use Fourier transforms, since $y=h\star x$ implies $Y(\omega)=H(\omega)\cdot X(\omega)$; hence:
  - we compute $Y(\omega)=\mathcal{F}(y)$;
  - we compute $X(\omega)=\dfrac{Y(\omega)}{H(\omega)}$, and
  - finally, we compute $x=\mathcal{F}^{-1}(X(\omega))$.

36. Deconvolution in the Presence of Noise with Known Characteristics

- Suppose that signal and noise are independent, and we know the power spectral densities

$$S_I(\omega) = \lim_{T\to\infty} E\left[\frac{1}{T}\cdot |X_T(\omega)|^2\right], S_N(\omega) = \lim_{T\to\infty} E\left[\frac{1}{T}\cdot |N_T(\omega)|^2\right]$$

- We minimize the expected mean square difference

$$d \stackrel{\text{def}}{=} \lim_{T\to\infty} \frac{1}{T}\cdot E\left[\int_{-T/2}^{T/2} (\widehat{x}(t) - x(t))^2\, dt\right].$$

- Minimizing d leads to the known Wiener filter formula

$$\widehat{X}(\omega_1,\omega_2) = \frac{H^*(\omega_1,\omega_2)}{|H(\omega_1,\omega_2)|^2 + \dfrac{S_N(\omega_1,\omega_2)}{S_I(\omega_1,\omega_2)}}\cdot Y(\omega_1,\omega_2).$$

37. Blind Image Deconvolution in the Presence of Prior Knowledge

- Wiener filter techniques assume that we know the blurring function h.
- In practice, we often only have partial information about h.
- Such situations are known as *blind deconvolution*.
- Sometimes, we know a joint probability distribution $p(\Omega, x, h, y)$ corresponding to some parameters Ω:
$$p(\Omega, x, h, y) = p(\Omega) \cdot p(x|\Omega) \cdot p(h|\Omega) \cdot p(y|x, h, \Omega).$$
- In this case, we can find
$$\widehat{\Omega} = \arg\max_{\Omega} p(\Omega|y) = \int\int_{x,h} p(\Omega, x, h, y)\, dx\, dh \text{ and}$$
$$(\widehat{x}, \widehat{h}) = \arg\max_{x,h} p(x, h|\widehat{\Omega}, y).$$

38. Blind Image Deconvolution in the Absence of Prior Knowledge: Sparsity-Based Techniques

- In many practical situations, we do not have prior knowledge about the blurring function h.
- Often, what helps is *sparsity* assumption: that in the expansion $x(t) = \sum_i a_i \cdot e_i(x)$, most a_i are zero.
- In this case, it makes sense to look for a solution with the smallest value of

$$\|a\|_0 \stackrel{\text{def}}{=} \#\{i : a_i \neq 0\}.$$

- The function $\|a\|_0$ is not convex and thus, difficult to optimize.
- It is therefore replaced by a close *convex* objective function $\|a\|_1 \stackrel{\text{def}}{=} \sum_i |a_i|$.

39. State-of-the-Art Technique for Sparsity-Based Blind Deconvolution

- Sparsity is the main idea behind the algorithm described in (Amizic et al. 2013) that minimizes

$$\frac{\beta}{2}\cdot\|y-\mathbf{W}a\|_2^2+\frac{\eta}{2}\cdot\|\mathbf{W}a-\mathbf{H}x\|_2^2+\tau\cdot\|a\|_1+\alpha\cdot R_1(x)+\gamma\cdot R_2(h).$$

- Here, $R_1(x)=\sum\limits_{d\in D} 2^{1-o(d)}\sum_i |\Delta_i^p(x)|^p$, where $\Delta_i^p(x)$ is the difference operator, and
- $R_2(h)=\|\mathbf{C}h\|^2$, where $\mathbf{C}$ is the discrete Laplace operator.
- The ℓ^p-sum $\sum\limits_i |v_i(x)|^p$ is optimized as $\sum\limits_i \frac{(v_i(x^{(k)}))^2}{v_i^{2-p}}$, where $v_i=v_i(x^{(k-1)})$ for x from the previous iteration.
- This method results in the best blind image deconvolution.

40. Need for Improvement

- The current technique is based on minimizing the sum $|\Delta_x I|^p + |\Delta_y I|^p$.
- This is a discrete analog of the term $\left|\dfrac{\partial I}{\partial x}\right|^p + \left|\dfrac{\partial I}{\partial y}\right|^p$.
- For $p = 2$, this is the square of the length of the gradient vector and is, thus, rotation-invariant.
- However, for $p \neq 2$, the above expression is not rotation-invariant.
- Thus, even if it works for some image, it may not work well if we rotate this image.
- To improve the quality of image deconvolution, it is thus desirable to make the method rotation-invariant.
- We show that this indeed improves the quality of deconvolution.

41. Rotation-Invariant Modification: Description and Results

- We want to replace the expression $\left|\dfrac{\partial I}{\partial x}\right|^p + \left|\dfrac{\partial I}{\partial y}\right|^p$ with a rotation-invariant function of the gradient.
- The only rotation-invariant characteristic of a vector a is its length $\|a\| = \sqrt{\sum\limits_i a_i^2}$.
- Thus, we replace the above expression with

$$\left(\left|\frac{\partial I}{\partial x}\right|^2 + \left|\frac{\partial I}{\partial y}\right|^2\right)^{p/2}.$$

- Its discrete analog is $((\Delta_x I)^2 + (\Delta_y I)^2)^{p/2}$.
- This modification leads to a 10% improvement in reconstruction accuracy $\|\widehat{x} - x\|_2$, from 1360 to 1190.

Part IV

How to Estimate Resilient Modulus for Unbound Aggregate Materials: A Theoretical Explanation of an Empirical Formula

42. Formulation of the Problem

- It is important to make sure that all the pavement layers have reached a certain stiffness level.
- To characterize stiffness of unbound pavement materials, transportation engineers use *resilient modulus* M_r.
- M_r is actually an estimate of its modulus of elasticity E, i.e., of ratio of stress by strain.
- The difference is that:
  - E corresponds to a *slowly* applied load, while
  - M_r characterizes the effect of *rapidly* applied loads – like those experienced by pavements.
- In the usual (*linear*) elastic materials, the modulus does not depend on the stress value.

43. Formulation of the Problem (cont-d)

- In contrast, pavement materials are usually *non-linear*: M_r depends on the stress eigenvalues σ_i.
- Several empirical formulas have been proposed to describe this dependence.
- Experimental comparison shows that the best description is provided by the formula

$$M_r = k'_1 \cdot \left(\frac{\theta}{P_a} + 1\right)^{k'_2} \cdot \left(\frac{\tau_{\text{oct}}}{P_a} + 1\right)^{k'_3}, \text{ where}$$

$$\theta = \sigma_1 + \sigma_2 + \sigma_3 \text{ and}$$

$$\tau_{\text{oct}} = \frac{1}{3} \cdot \sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2}.$$

- In this talk, we provide a theoretical explanation for the above empirical formula.

44. Main Idea: Fundamental Physical Formulas Should Not Depend on the Choice of the Starting Point or of the Measuring Unit

- Computers process numerical values of different quantities.
- A numerical value of a quantity depends:
  - on the choice of a measuring unit and
  - in many cases – also on the choice of the starting point.
- For example, we can describe the height of the same person as 1.7 m or 170 cm.
- 14.00 by El Paso time is 15.00 by Austin time.
- Reason: the starting points – midnights (00.00) – differ by an hour.
- The choice of a measuring unit is rather arbitrary.

45. Main Idea (cont-d)

- It is reasonable to require that the fundamental physical formulas not depend
  - on the choice of a measuring unit and
  - if appropriate – on the choice of the starting point.
- We do not expect that, e.g., Newton's laws look differently if we use meters or feet.
- If we change the units, then we may need to adjust units of related quantities.
- For example, if we replace m with cm, then we need to replace m/sec with cm/sec when measuring velocity.
- However, once the appropriate adjustments are made, we expect the formulas to remain the same.

46. Not All Physical Quantities Allow Both Changes

- Some quantities have a fixed starting point.
- Examples:
  - while we can choose an arbitrary starting point for time,
  - for distance, 0 distance seems to be a reasonable starting point.
- As a result:
  - while the change of a measuring unit makes sense for most physical quantities,
  - the change of a starting point only makes sense for some of them.
- A physics-based analysis is needed to decide whether this change makes physical sense.

47. Description in Precise Terms

- *Case:* we replace the original measuring unit with a new unit which is a times smaller.
- Then all numerical values of the measured quantity get multiplied by a: $x' = a \cdot x$.
- *Example:* 1.7 m is $x' = a \cdot x = 100 \cdot 1.7 = 170$ cm.
- *Case:* we replace the original starting point by a new one which is b earlier (or smaller).
- Then to all numerical values of the measured quantity the value b is added: $x' = x + b$.
- 14 hr in El Paso is $x' = x + b = 14 + 1 = 15$ in Austin.
- *General case:* we can change both the measuring unit and the starting point.
- Then, $x \to a \cdot x + b$.

48. How Resilient Modulus M_r Depends on the Bulk Stress θ

- For M_r, there is a clear starting point $M_r = 0$, in which strain does not cause any stress.
- So, for M_r, only a change in a measuring unit makes physical sense.
- For θ, a change in starting point is also possible:
  - we can only count the external stress,
  - or we can explicitly take atmospheric pressure into account.
- So, the requirement that the dependence $M_r(\theta)$ does not change means that:

$$\forall a\, \forall b\, \exists c(a, b)\, (M_r(a \cdot \theta + b) = c(a, b) \cdot M_r(\theta)).$$

- *Problem:* only $M_r(\theta) = \text{const}$ satisfies this functional equation, and we know that $M_r(\theta) \neq \text{const}$.

49. Since We Cannot Require All the Invariances, Let Us Require Only Some of Them

- If:
  - a formula does not change when we apply each transformation,
  - it will also not change if we apply them one after another (i.e., apply a superposition).
- Each shift can be represented as a superposition of many small (infinitesimal) shifts $\theta \to \theta + B \cdot dt$.
- Similarly, each re-scaling can be represented as a superposition of many small (infinitesimal) re-scalings

$$\theta \to (1 + A \cdot dt) \cdot \theta.$$

- Thus, it is sufficient to consider invariance with respect to an infinitesimal transformation

$$\theta \to \theta' = (1 + A \cdot dt) \cdot \theta + B \cdot dt.$$

50. How M_r Depends on θ

- Invariance relative to an infinitesimal transformation means that
$$M_r((1+A\cdot dt)\cdot\theta+B\cdot dt)=(1+C\cdot dt)\cdot M_r(\theta).$$
- Here, $M_r(\theta+q\cdot dt)=M_r(\theta)+M_r'(\theta)\cdot q\cdot dt$, so
$$(A\cdot\theta+B)\cdot\frac{dM_r}{d\theta}(\theta)=C\cdot M_r(\theta).$$
- Thus, $\dfrac{dM_r}{M_r}=C\cdot\dfrac{d\theta}{A\cdot\theta+b}$.
- Cases when $A=0$ can be approximated, with any given accuracy, by cases when A is small but non-zero.
- So, w.l.o.g., we can safely assume $A\neq 0$.
- In this case, for $x\stackrel{\text{def}}{=}\theta+k$, where $k\stackrel{\text{def}}{=}\dfrac{B}{A}$, we have $\dfrac{dM_r}{M_r}=c\cdot\dfrac{dx}{x}$, where $c\stackrel{\text{def}}{=}\dfrac{C}{A}$.

51. How M_r Depends on θ and on $\tau_{\rm oct}$

- We have $\dfrac{dM_r}{M_r} = c \cdot \dfrac{dx}{x}$.
- Integration leads to $\ln(M_r) = c \cdot \ln(x) + C_0$ for some constant C_0, thus $M_r = C_1 \cdot x^c$ for $C_1 \stackrel{\rm def}{=} \exp(C_0)$, i.e.,

$$M_r(\theta) = C_1 \cdot (\theta + k)^c.$$

- If we represent $\theta + k$ as $k \cdot \left(\dfrac{\theta}{k} + 1\right)$, then we get the desired dependence of M_r on θ:

$$M_r = C_2 \cdot \left(\frac{\theta}{k} + 1\right)^c, \text{ where } C_2 \stackrel{\rm def}{=} C_1 \cdot k^c.$$

- Similarly, for dependence on sheer stress $\tau_{\rm oct}$, we get:

$$M_r = C_2' \cdot \left(\frac{\tau_{\rm oct}}{k'} + 1\right)^{c'}.$$

52. Need to Combine the Formulas $M_r(\theta)$ and $M_r(\tau_{\text{oct}})$ into a Formula for $M_r(\theta, \tau_{\text{oct}})$

- We have used the invariance ideas to derive formulas $M_r(\theta)$ and $M_r(\tau_{\text{oct}})$.
- Let us now use the same ideas to combine these two formulas into a single formula for $M_r(\theta, \tau_{\text{oct}})$.
- Based on the previous analysis, for each pair $(\theta, \tau_{\text{oct}})$, we know two values of the modulus M_r:
  - the value $M_1 \stackrel{\text{def}}{=} M_r(\theta)$ that we obtain if we ignore sheer stress τ_{oct}; and
  - the value $M_2 \stackrel{\text{def}}{=} M_r(\tau_{\text{oct}})$ that we obtain if ignore the bulk stress θ.
- Based on these two values M_1 and M_2, we would like to compute an estimate $M(M_1, M_2)$.

53. How to Combine the Formulas $M_r(\theta)$ and $M_r(\tau_{\rm oct})$ into a Formula for $M_r(\theta, \tau_{\rm oct})$

- All three values M, M_1, and M_2 represent modulus.
- Thus, for all three values, only scaling is possible.
- So, the invariance requirement takes the following form: for every p and q,
  - if we apply re-scalings $M_1 \to p\cdot M_1$ and $M_2 \to q\cdot M_2$,
  - then, after an appropriate re-scaling by some parameter $c(p,q)$ depending on p and q,
  - the resulting dependence $M(p\cdot M_1, q\cdot M_2)$ has the same form as the original dependence $M(M_1, M_2)$.
- So, for every p and every q, there exists a $c(p,q)$ for which, for all M_1 and M_2, we have

$$M(p\cdot M_1, q\cdot M_2) = c(p,q)\cdot M(M_1, M_2).$$

54. Main Result

- For every p and every q, there exists a $c(p, q)$ for which, for all M_1 and M_2, we have

$$M(p \cdot M_1, q \cdot M_2) = c(p, q) \cdot M(M_1, M_2).$$

- From the above invariance requirements, we can conclude that the dependence of $M_r(\theta, \tau_{\text{oct}})$ has the form

$$M(\theta, \tau_{\text{oct}}) = k_1 \cdot \left(\frac{\theta}{k} + 1\right)^{k_2} \cdot \left(\frac{\tau_{\text{oct}}}{k'} + 1\right)^{k_3}.$$

- Thus, we indeed get a theoretical explanation for the empirical dependence.

55. Acknowledgment

This work was supported in part:

- by the National Science Foundation grants:
  - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
  - DUE-0926721, and
- by an award from Prudential Foundation.

56. Proof for Sparsity

- By definition of the t-norm, we have

$$d(\varepsilon) = f_\&(d_1^{\varepsilon_1}, \ldots, d_N^{\varepsilon_N}) = f^{-1}(f(d_1^{\varepsilon_1}) \cdot \ldots \cdot f(d_N^{\varepsilon_N})).$$

- So, $d(\varepsilon) = f_\&(d_1^{\varepsilon_1}, \ldots, d_N^{\varepsilon_N}) = f^{-1}(e_1^{\varepsilon_1} \cdot \ldots \cdot e_N^{\varepsilon_N})$, where we denoted $e_i^{\varepsilon_i} \stackrel{\text{def}}{=} f(d_i^{\varepsilon_i})$.
- Since $f(x)$ is increasing, maximizing $d(\varepsilon)$ is equivalent to maximizing $e(\varepsilon) \stackrel{\text{def}}{=} f(d(\varepsilon)) = e_1^{\varepsilon_1} \cdot \ldots \cdot e_N^{\varepsilon_N}$.
- We required that the sequences $d^=$ and $d^{\neq}$ properly describe reasonableness.
- Thus, for each i, we have $d(\varepsilon_=) > d(\varepsilon_=^{(i)})$, where

$$\varepsilon_=^{(i)} \stackrel{\text{def}}{=} (=, \ldots, =, \neq \text{ (on } i\text{-th place)}, =, \ldots, =).$$

- This inequality is equivalent to $e(\varepsilon_=) > e(\varepsilon_=^{(i)})$.
- Since the values $e(\varepsilon)$ are simply the products, we thus conclude that $e_i^= > e_i^{\neq}$.

57. Proof for Sparsity (cont-d)

- Maximizing $e(\varepsilon)=\prod_{i=1}^{N} e_i^{\varepsilon_i}$ is equivalent to maximizing $\dfrac{e(\varepsilon)}{c}$, for a constant $c \stackrel{\text{def}}{=} \prod_{i=1}^{N} e_i^{=}$.
- The ratio $\dfrac{e(\varepsilon)}{c}$ can be reformulated as $\dfrac{e(\varepsilon)}{c}=\prod_{i:\varepsilon_i=\neq} \dfrac{e_i^{\neq}}{e_i^{=}}$.
- Since $\ln(x)$ is increasing, maximizing this product is equivalent to minimizing minus logarithm

$$L(\varepsilon) \stackrel{\text{def}}{=} -\ln\left(\frac{e(\varepsilon)}{c}\right)=\sum_{i:\varepsilon_i=\neq} w_i, \text{ where } w_i \stackrel{\text{def}}{=} -\ln\left(\frac{e_i^{\neq}}{e_i^{=}}\right).$$

- Since $e_i^{=}>e_i^{\neq}>0$, we have $\dfrac{e_i^{\neq}}{e_i^{=}}<1$ and thus, $w_i>0$.
- The proposition is proven.

58. Proof That Only $M_r(\theta) = \text{const}$ Satisfies the Corresponding Functional Equation

- We want $M_r(a \cdot \theta + b) = c(a, b) \cdot M_r(\theta)$ for all a and b.
- From the physical viewpoint, small changes in θ should lead to small changes in M_r.
- In mathematical terms, the dependence $M_r(\theta)$ should be continuous.
- It is known that every continuous function can be approximated,
  - with any given accuracy,
  - by a differentiable function (e.g., by a polynomial).
- Thus, without losing generality, we can safely assume that the dependence $M_r(\theta)$ is differentiable.
- Thus, the function $c(a, b) = \dfrac{M_r(a \cdot \theta + b)}{M_r(\theta)}$ is also differentiable, as a ratio of two differentiable functions.

59. Proof That Only $M_r(\theta) = \text{const}$ Satisfies the Functional Equation (cont-d)

- For $a = 1$, $M_r(a \cdot \theta + b) = c(a, b) \cdot M_r(\theta)$ becomes
$$M_r(\theta + b) = c(1, b) \cdot M_r(\theta).$$
- Differentiating both sides by b and setting $b = 0$, we get $\dfrac{dM_r}{d\theta} = c \cdot M_r$, hence $\dfrac{dM_r}{M_r} = c \cdot d\theta$.
- So, $\ln(M_r) = c \cdot \theta + C_0$ for some C_0.
- Thus, $M_r = A \cdot \exp(c \cdot \theta)$, where $A \stackrel{\text{def}}{=} \exp(C_0)$.
- For $b = 0$ and $a \neq 1$, we get $M_r(a \cdot \theta) = c(a, 0) \cdot M_r(\theta)$, i.e.,
$$A \cdot \exp(c \cdot a \cdot \theta) = c(a, 0) \cdot \exp(c \cdot \theta).$$
- When $c \neq 0$, the two sides grow with θ at a different speed.
- So, $c = 0$ and $M_r(\theta) = \text{const}$.

60. Proof of the Main Result

- If we re-scale only one of the inputs, e.g., M_1, we get

  $$M(p \cdot M_1, M_2) = c_1(p) \cdot M(M_1, M_2), \text{ where } c_1(p) \stackrel{\text{def}}{=} c(p, 1).$$

- If we first re-scale by p and then by p', then this is equivalent to one re-scaling by $p \cdot p'$.

- In the first case, we get

  $$M((p \cdot p') \cdot M_1, M_2) = M(p' \cdot (p \cdot M_1), M_2) =$$

  $$c_1(p') \cdot M(p \cdot M_1, M_2) = c_1(p') \cdot c_1(p) \cdot M(M_1, M_2).$$

- In the second case, we get

  $$M((p \cdot p') \cdot M_1, M_2) = c_1(p \cdot p') \cdot M(M_1, M_2).$$

- Since the left-hand sides of these equalities are equal, their right-hand sides must be equal as well.

61. Proof of the Main Result (cont-d)

- Dividing the resulting equality by $M(M_1, M_2)$, we conclude that $c_1(p \cdot p') = c_1(p) \cdot c_1(p')$.
- Differentiating this equality by p' and taking $p' = 1$, we get $p \cdot c_1'(p) = c_0 \cdot c_1(p)$, where $c_0 \stackrel{\text{def}}{=} c_1'(1)$.
- Thus, $\dfrac{dc_1}{c_1} = c_0 \cdot \dfrac{dp}{p}$, so integration leads to $\ln(c_1) = c_0 \cdot \ln(p) + \text{const}$, and $c_1(p) = \text{const} \cdot p^{c_0}$.
- For $M_1 = 1$, invariance means

$$M(p, M_2) = \text{const} \cdot p^{c_0} \cdot M(1, M_2).$$

- Renaming the variable,

$$M(M_1, M_2) = \text{const} \cdot M_1^{c_0} \cdot M(1, M_2).$$

- Similarly, we have

$$M(M_1, M_2) = \text{const}' \cdot M_2^{c_0'} \cdot M(M_1, 1), \text{ for some constants.}$$

62. Proof of the Main Result (end)

- *Reminder:* $M(M_1, M_2) = \text{const}' \cdot M_2^{c_0'} \cdot M(M_1, 1)$.
- In particular, for $M_1 = 1$, the formula takes the form

$$M(1, M_2) = \text{const}' \cdot M_2^{c_0'} \cdot M(1, 1).$$

- *Reminder:* $M(M_1, M_2) = \text{const} \cdot M_1^{c_0} \cdot M(1, M_2)$.
- Substituting the expression for $M(1, M_2)$ into this formula, we get

$$M(M_1, M_2) = \text{const} \cdot M_1^{c_0} \cdot \text{const}' \cdot M_2^{c_0'} \cdot M(1, 1).$$

- *Reminder:* $M_1 = C_2 \cdot \left(\frac{\theta}{k} + 1\right)^c$, $M_2 = C_2' \cdot \left(\frac{\tau_{\text{oct}}}{k'} + 1\right)^{c'}$.
- Substituting expressions for M_i into this formula for $M(M_1, M_2)$, we come up with the desired conclusion.