Departmental Technical Reports (CS)   Department of Computer Science

1-2017

Beyond Traditional Applications of Fuzzy Techniques: Main Idea and Case Studies

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Thongchai Dumrongpokaphan
Chiang Mai University, tcd43@hotmail.com

Recommended Citation

Beyond Traditional Applications of Fuzzy Techniques: Main Idea and Case Studies

Vladik Kreinovich, Olga Kosheleva, and Thongchai Dumrongpokaphan

Abstract Fuzzy logic techniques were originally designed to translate expert knowledge – which is often formulated by using imprecise ("fuzzy") from natural language (like "small") – into precise computer-understandable models and control strategies. Such a translation is still the main use of fuzzy techniques. Lately, it turned out that fuzzy methods can help in another class of applied problems: namely, in situations when there are semi-heuristic techniques for solving the corresponding problems, i.e., techniques for which there is no convincing theoretical justification. Because of the lack of a theoretical justification, users are reluctant to use these techniques, since their previous empirical success does not guarantee that these techniques will work well on new problems. In this paper, 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.

1 Introduction

Fuzzy logic techniques (see, e.g., [8, 11, 13]) were originally designed to translate expert knowledge – which is often formulated by using imprecise ("fuzzy") from natural language (like "small") – into precise computer-understandable models and control strategies. Such a translation is still the main use of fuzzy techniques. For example, we want to control a complex plant for which no good control technique is

Vladik Kreinovich and Olga Kosheleva
University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA
e-mail: vladik@utep.edu, olgak@utep.edu

Thongchai Dumrongpokaphan
Department of Mathematics, Faculty of Science, Chiang Mai University, Thailand
e-mail: tcd43@hotmail.com

known, but for which there are experts how can control this plant reasonably well. So, we elicit rules from the experts, and then we use fuzzy techniques to translate these rules into a control strategy.

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

Also, these techniques are usually not perfect, and 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.

In this paper, 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 (i.e., membership function, "and"- and "or"-operations, etc.), but for all combinations of such techniques which are consistent with the corresponding practical problem.

As examples, we provide a reasonably detailed justification of:

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

2 Fuzzy Logic: From Traditional to New Applications

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. A typical example is that 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.

Other situations in which we need help. 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.

An additional Problem of semi-heuristic techniques is that they 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 (described later), when they do not work well, it is not immediately clear what is a reasonable next approximation.

What we show in this paper. 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.

Case studies. As examples, we provide a reasonably detailed justification:

- of sparsity techniques in data and image processing – a very successful hot-topic technique
- whose success is often largely a mystery; and
- of ℓ^p-regularization techniques in solving inverse problems, an empirically successful alternative to smooth approaches which is appropriate for situations when the desired signal or image is not smooth.

Comment. A detailed description of the corresponding case studies can be found in [3, 4, 5, 6].

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

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* (see, e.g., [7]):

- 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 the number of non-zero coefficients – sometimes taken with weights – is the smallest possible:

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

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

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.

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*. The problem is that "reasonable" is not a precise term.

Let us use fuzzy logic. The problem is that 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. For example, 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 $d = 7/10$. There are many other ways to assign these degrees.

Thus, for each i, we can learn to what extent $a_i = 0$ or $a_i \neq 0$ are reasonable.

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 directly ask the expert about all these tuples.

In such situations, we 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 need to be able to solve the following problem:

- 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 $f_\&(a,b)$.

The "and"-estimate is not always exact: an example. It is important to emphasize that the resulting estimate cannot be exact. Let us give two examples.

In the first example, A is "coin falls heads", B is "coin falls tails". 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: $d(A \& B) = 0$.

Let us now consider the second example. 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 examples, we have $d(A') = d(A) = a$ and $d(B') = d(B) = b$, but $d(A \& B) \neq d(A' \& B')$.

Which "and"-operation (t-norm) should we choose. The corresponding function $f_\&(a,b)$ must satisfy some reasonable properties.

For example, 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.

It is known that each such operation can be approximated, with any given accuracy, by an *Archimedean* t-norm of the type $f_\&(a,b) = f^{-1}(f(a) \cdot f(b))$, for some strictly increasing function $f(x)$; see, e.g., [10].

Thus, without losing generality, we can assume that the actual t-norm is Archimedean.

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 $\neq$, we estimate the degree that this sequence is reasonable as:

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

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., $\varepsilon = (\neq, \ldots, \neq)$.

Thus, we arrive at the following definitions.

Definition 1.

- *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, strictly increasing,* $f(0) = 0$*, and* $f(1) = 1$*.*
- *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* $d^{\neq} = (d_1^{\neq}, \ldots, d_N^{\neq})$ *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* $d(\varepsilon) = \max_{\varepsilon' \in S} d(\varepsilon')$*.*

Now, we can formulate the main result of this section.

Proposition 1. *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* $\sum_{i:\varepsilon_i=\neq} w_i = \sum_{i:a_i \neq 0} w_i$ *is the smallest possible.*

Discussion. Thus, fuzzy-based techniques indeed naturally lead to the sparsity condition.

Proof of Proposition 1. 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 the function $f(x)$ 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 $e_i^= > e_i^{\neq}$.

Maximizing $e(\varepsilon) = \prod\limits_{i=1}^{N} e_i^{\varepsilon_i}$ is equivalent to maximizing $\dfrac{e(\varepsilon)}{c}$, for a constant $c \stackrel{\text{def}}{=}$ $\prod\limits_{i=1}^{N} e_i^{=}$. The ratio $\dfrac{e(\varepsilon)}{c}$ can be reformulated as $\dfrac{e(\varepsilon)}{c} = \prod\limits_{i:\varepsilon_i=\neq} \dfrac{e_i^{\neq}}{e_i^{=}}$.

Since $\ln(x)$ is an increasing function, maximizing this product is equivalent to minimizing minus logarithm of this product:

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

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.

A similar derivation can be obtained in the probabilistic case. Alternatively, 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, thus

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

So, we arrive at the following definition.

Definition 2.

- *Let $p^{=} = (p_1^{=}, \ldots, p_N^{=})$ be a sequence of real numbers from the interval $[0,1]$, and let $p_i^{\neq} \stackrel{\text{def}}{=} 1 - p_i^{=}$.*

- *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 $p(\varepsilon) = \max\limits_{\varepsilon' \in S} p(\varepsilon')$.*

Proposition 2. *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 to each other:*

- *the sequence* $\varepsilon \in S$ *is the most probable,*
- *the sum* $\sum\limits_{i:\varepsilon_i=\neq} w_i$ *is the smallest possible.*

Proof: the proof of this proposition is similar to the proof of Proposition 1.

Discussion. In other words, probabilistic techniques also lead to the sparsity condition.

Fuzzy approach vs. probabilistic approach. The fact that the probabilistic approach leads to the same conclusion as the fuzzy approach makes us more confident that our justification of sparsity is valid.

It should be mentioned, however, that 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. This is an important advantage of the fuzzy-based approach.

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

Need for beblurring. The second case study deals with signal and image processing.

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 planet, the blur is very visible – so deblurring is needed.

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.

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:

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

For continuous signals, sum turns into an integral:

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

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

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

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

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

This idea is known as *Tikhonov regularization*; see, e.g., [12].

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:

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

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

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; see, e.g., [9].

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; see, e.g., [2].

Remaining problem. A big problem is that the ℓ^p-methods are heuristic. For example, there is no convincing explanation of why necessarily we replace the square with a p-th power and not with some other function.

What we show. In this section, 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*, techniques that were designed to transform imprecise intuitive ideas into exact formulas.

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:

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

We have already mentioned, in the previous section, that 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.

Analysis of the problem. We want to select an image with the largest degree d of satisfying the above 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 $g(x)$.

Which function $g(x)$ should we select: idea. The value $\Delta x_i = 0$ is absolutely small, so we should have $\mu(0) = 1$ 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 which is a times smaller, then $\Delta x_i \to a \cdot \Delta x_i$. It is reasonable to request that the requirement $\sum_k g(\Delta x_k) \to \min$ not change if we change a measuring unit. For example, if $g(z_1) + g(z_2) = g(z_1') + g(z_2')$, then

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

Which functions $g(z)$ satisfy this property?

Definition 3. *A function $g(z)$ is called* scale-invariant *if it satisfies the following two conditions:*

- $g(0) = 0$ *and*
- *for all z_1, z_2, z_1', z_2', and a, $g(z_1) + g(z_2) = g(z_1') + g(z_2')$ implies*

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

Proposition 3. *A function $g(z)$ is scale-invariant if and only if it has the form $g(a) = c \cdot a^p$, for some c and $p > 0$.*

Discussion. Minimizing $\sum_k g(\Delta x_k)$ is equivalent to minimizing the sum $\sum_k |\Delta x_k|^p$. Minimizing the sum $\sum_k |\Delta x_k|^p$ under condition $J \leq c$ is equivalent to minimizing the expression $J + \lambda \cdot \sum_k |\Delta x_k|^p$. Thus, fuzzy techniques indeed justify the ℓ^p-method.

Proof of Proposition 3. 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 $g'(a \cdot z_1) + g'(a \cdot z_2) \cdot k = 0$, 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.

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, $F(a_1 \cdot a_2) = F(a_1) \cdot F(a_2)$. Continuous solutions of this functional equations are well known (see, e.g., [1]), so we conclude that $F(a) = a^q$ for some real number q. For this function $F(a)$, the equality $\dfrac{g'(a \cdot z_1)}{g'(z_1)} = F(a)$ becomes $g'(a \cdot z_1) = g'(z_1) \cdot a^q$.

In particular, for $z_1 = 1$, we get $g'(a) = C \cdot a^q$, where $C \overset{\text{def}}{=} g'(1)$.

In general, 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 $g(0) = 0$. Thus, this case is impossible, and $q \neq -1$. Integrating, for $q \neq -1$, we get $g(a) = \dfrac{C}{q+1} \cdot a^{q+1} + \text{const}$. The condition $g(0) = 0$ implies that $\text{const} = 0$.

Thus, the proposition is proven, for $p = q + 1$.

5 How to Improve the Existing Semi-Heuristic Technique

What we do in this section. Until now, we have discussed how to justify the existing semi-heuristic techniques. However, often, these techniques are not perfect, so it is desirable to improve them. Let us describe an example of how this can be done.

Blind image deconvolution: formulation of the problem. In general, 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*.

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 $Y(\omega) = \mathscr{F}(y)$;
- we compute $X(\omega) = \dfrac{Y(\omega)}{H(\omega)}$, and
- finally, we compute $x = \mathscr{F}^{-1}(X(\omega))$.

Deconvolution in the presence of noise with known characteristics. Suppose that signal and noise are independent, and we know the power spectral densities

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

Then, 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

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

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

$$\widehat{\Omega} = \arg\max_{\Omega} p(\Omega|y) = \int\int_{x,h} p(\Omega, x, h, y) \, dx \, dh \text{ and}$$

$$(\widehat{x}, \widehat{h}) = \arg\max_{x,h} p(x, h|\widehat{\Omega}, y).$$

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 number of non-zero coefficients:

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

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

State-of-the-art technique for sparsity-based blind deconvolution. Sparsity is the main idea behind the algorithm described in [2] 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, $R_1(x) = \sum_{d \in D} 2^{1-o(d)} \sum_i |\Delta_i^p(x)|^p$, 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 $\sum_i |v_i(x)|^p$ is optimized as $\sum_i \dfrac{(v_i(x^{(k)}))^2}{v_i^{2-p}}$, where $v_i = v_i(x^{(k-1)})$ for x from the previous iteration.

This method results in the best blind image deconvolution.

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.

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 $\|a\| = \sqrt{\sum_i a_i^2}$.

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 indeed leads to a statistically significant improvement in reconstruction accuracy $\|\widehat{x} - x\|_2$.

Specifically, to compare the new methods with the original method from [2], we applied each of the two algorithms 30 times, and for each application, we computed the reconstruction accuracy. To make the results of the comparison more robust, for each of the algorithms, we eliminated the smallest and the largest value of this distance, and got a list of 28 values. For the original algorithm, the average of these values is 1195.21. For the new method, the average is 1191.01, which is smaller than the average distance corresponding to the original algorithm. To check whether this difference is statistically significance, we applied the t-test for two independent means. The t-test checks whether the null hypothesis – that both samples comes from the populations with same mean – can be rejected. For the two samples, computations lead to rejection with $p = 0.002$. This is much smaller than the p-values 0.01 and 0.05 normally used for rejecting the null hypothesis. So, we can conclude that the null hypothesis can be rejected, and that, therefore, the modified algorithm is indeed statistically significantly better than the original one (see [3] for details).

How can we go beyond ℓ^p-methods? While ℓ^p-methods are efficient, they are not always perfect. A reasonable idea is to try to improve the quality of signal and image reconstruction by using functions $g(z)$ more general than $g(z) = C \cdot |z|^p$. For example, instead of considering only functions from this 1-parametric family, we can consider a more general 2-parametric family of functions

$$g(z) = C \cdot |z|^p + C_1 \cdot g_1(z).$$

Which function $g_1(z)$ should we use?

In [6], we used the same ideas of scale-invariance – that are used above to justify ℓ^p-techniques – to show that the best choice is to use functions $g_1(z) = |z|^p \cdot \ln(z)$ or $g_1(z) = |z|^{p_1}$ for some p_1. The same approach also helps to decide which functions to use if we consider 3- and more-parametric families instead of 2-parametric ones [6].

Acknowledgments

References

1. J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 2008.
2. B. Amizic, L. Spinoulas, R. Molina, and A. K. Katsaggelos, "Compressive blind image deconvolution", *IEEE Transactions on Image Processing*, 2013, Vol. 22, No. 10, pp. 3994–4006.
3. F. Cervantes, B. Usevitch, and V. Kreinovich, "Rotation-invariance can further improve state-of-the-art blind deconvolution techniques", *Proceedings of International IEEE Conference on Systems, Man, and Cybernetics SMC'2016*, Budapest, Hungary, October 9–12, 2016.
4. F. Cervantes, B. Usevitch, and V. Kreinovich, "Why ℓ^p-methods in signal and image processing: a fuzzy-based explanation", *Proceedings of the 2016 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2016*, El Paso, Texas, October 31 – November 4, 2016.
5. F. Cervantes, B. Usevitch, L. Valera, and V. Kreinovich, "Why sparse? fuzzy techniques explain empirical efficiency of sparsity-based data- and image-processing algorithms", *Proceedings of the 2016 World Conference on Soft Computing*, Berkeley, California, May 22–25, 2016, pp. 165–169.
6. F. Cervantes, B. Usevitch, L. Valera, V. Kreinovich, and O. Kosheleva, "Fuzzy techniques provide a theoretical explanation for the heuristic ℓ^p-regularization of signals and images", *Proceedings of the 2016 IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2016*, Vancouver, Canada, July 24–29, 2016.
7. M. Elad, *Sparse and Redundant Representations*, Springer Verlag, 2010.
8. G. Klir and B. Yuan, "Fuzzy Sets and Fuzzy Logic", Prentice Hall, Upper Saddle River, New Jersey, 1995.
9. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, California, 1983.
10. H. T. Nguyen, V. Kreinovich, and P. Wojciechowski, "Strict Archimedean t-Norms and t-Conorms as Universal Approximators", *International Journal of Approximate Reasoning*, 1998, Vol. 18, Nos. 3–4, pp. 239–249.
11. H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
12. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, Washington, DC, 1977.
13. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.