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

Fernando Cervantes[1], Brian Usevitch[1],
Leobardo Valera[2], and Vladik Kreinovich[2,3]
[1]Department of Electrical and Computer Engineering
[2]Computational Science Program
[3]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
fcervantes@miners.utep.edu, usevitch@utep.edu
leobardovalera@gmail.com, vladik@utep.edu

***Abstract*—In many practical applications, it turned out to be efficient to assume that the signal or an image is *sparse*, i.e., that when we decompose it into appropriate basic functions (e.g., sinusoids or wavelets), most of the coefficients in this decomposition will be zeros. At present, the empirical efficiency of sparsity-based techniques remains somewhat a mystery. In this paper, we show that fuzzy-related techniques can explain this empirical efficiency. A similar explanation can be obtained by using probabilistic techniques; this fact increases our confidence that our explanation is correct.**

I. Formulation of the Problem

Sparsity-based techniques are useful. In many practical applications, it turned out to be efficient to assume that the signal or an image is sparse; see, e.g., [1], [2], [3], [4], [5], [6], [7], [8], [9], [12], [13], [14], [17], [18].

In precise terms, sparsity means that when we decompose the original signal $x(t)$ (or original image) into appropriate basic functions $e_1(t)$, $e_2(t)$, ... (e.g., sinusoids or wavelets), i.e., represent this signal (or image) as a linear combination $x(t) = \sum\limits_{i=1}^{\infty} a_i \cdot e_i(t)$, then most of the coefficients a_i in this decomposition will be zeros.

Moreover, it is usually beneficial to select, among all the signals which are consistent with all the observations (and will all additional knowledge), the signal for which:

- either the number of non-zero coefficients is the smallest possible,
- or, more generally, the "weighted number" of such coefficient is the smallest possible, where the weighted number is defined as $\sum\limits_{i:a_i \neq 0} w_i$, for some weights $w_i > 0$.

Comment. Sparsity can be viewed as a particulate case of the Occam's razor, according to which we should always select the simplest model that fits observation.

But why are sparsity-related techniques useful? At present, the empirical efficiency of sparsity-based techniques remains somewhat a mystery.

What we do in this paper. In this paper, we show that fuzzy-related techniques can explain this empirical efficiency.

We also show that a similar explanation can be obtained by using probabilistic techniques; this fact increases our confidence that our explanation is correct.

II. General Analysis of the Problem

Why do we need data processing in the first place? To better understand why different techniques are more or less empirically successful in data processing, it is important to recall why we need data processing in the first place.

One of the main goals of science and engineering is to predict the future state of the worlds, and to design gadgets and strategies that would make the future state of the world more beneficial for us.

To predict the state of the world, we need to know the current state of the world, and we need to know how this state changes in time. In general, the state of the world can be described by the numerical values of different physical quantities. In these terms, to predict the future state of the world means to predict the future values of the corresponding quantities $y_1, \ldots, y_m$.

In each practical problem, we are usually interested only in a small number of quantities. However, to predict the future values of these quantities, we often need to know the initial values of some auxiliary quantities as well. For example, when we launch a spaceship, we are interested in its location and direction when it leaves the atmosphere, and we are not directly interested in the future values of the winds on different heights. However, these winds affect the spaceship's trajectory, and, as a result, we need to know their initial values

to correctly predict the desired values $y_1, \ldots, y_m$. In general, we need to know $n \gg m$ initial values $x_1, \ldots, x_n$ to make the desired prediction.

The relation between x_i and y_j is often complicated. So, to predict the values $y_1, \ldots, y_m$ based on the inputs $x_1, \ldots, x_n$, we need to apply complex computer-based algorithms, i.e., perform *data processing*.

The above description captures the main reason for data processing, but it is somewhat oversimplified, since it assumes that we know the values $x_1, \ldots, x_n$ of the original quantities. For some quantities, this is indeed true, since we can directly measure their values. However, there are many other quantities which are difficult to measure directly. For example, when we are trying to predict the state of the engine, it is desirable to know the current temperature inside; however, this temperature is difficult to measure directly. If we cannot directly measure a certain value x_i, a natural idea is to find easier-to-measure auxiliary quantities $z_1, \ldots, z_p$ which are related to x_i by a known dependence, and then use this known dependence to reconstruct x_i. The corresponding computations may be complex, so we have another reason why data processing is needed.

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 $x_1, \ldots, x_N$ that describe the current state of the world 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, before we decide how to transform the inputs x_i into the desired output, we first need to check which inputs are actually useful. This checking is a very important stage of data processing – if we do not filter out unnecessary quantities x_i, we will waste time and resources measuring and processing these unnecessary quantities.

III. Analysis of the Problem: Let Us Use Fuzzy-Related Techniques

Description of our problem in natural-language terms. 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. In this formula, the basis functions $e_i(t)$ are known, and the coefficients a_i need to be determined. Based on measurement results and prior knowledge, we need to estimate the values a_i.

This reconstruction problem is, of course, a particular case of a general data processing problem. As we have mentioned in the previous section, a natural way to approach data processing problems in general is that:

- first, we find out which quantities are important for this particular problem, and
- then, we find the values of the important quantities.

In the above data processing problem, the quantities are the coefficients a_i. The quantity a_i is irrelevant if it does not affect the resulting signal, i.e., if $a_i = 0$. When $a_i \neq 0$, this means that this quantity affects the resulting signal $x(t) = \sum_{i=1}^{\infty} a_i \cdot e_i(t)$ and is, therefore, relevant. Thus, for our problem, the above two-stage data processing process takes the following form:

- first, we decided which values a_i are zeros and which are non-zeros, and
- then, we use an appropriate data processing algorithm to estimate the numerical values of non-zero coefficients a_i.

On the first stage, we can make several different decisions, all of which are consistent with the measurements and with the prior knowledge. For example, if in one decision, we take $a_i = 0$, then taking a_i to be very small but still different from 0 will still make this slightly modified signal consistent with all the measurement results. Out of all such possible decisions, we need to select *the most reasonable one*.

"Reasonable" is not a precise term. So, to be able to solve the problem, we need to translate this imprecise natural-language description into precise terms.

Fuzzy techniques can translate this natural-language description into a precisely formulated problem. In order to translate the above natural-language problem into precise terms, it is reasonable to use techniques specifically designed for such translations – namely, the techniques of *fuzzy logic*; see, e.g., [11], [16], [19].

In fuzzy logic, the meaning of each imprecise ("fuzzy") natural-language statement $P(x)$ about a quantity x is described by assigning, to each possible value x, the degree $\mu_P(x) \in [0, 1]$ to which we are sure that x satisfies this property P. For simple properties, we can determine these values, e.g., by simply asking the experts to mark, on a scale from 0 to 10, how much they are sure that P holds for x; if an expert marks the number 7, we take $\mu_P(x) = 7/10$.

This can be done for properties that depend on a single quantity. However, for properties like "reasonable", that depend on many values $a_1, \ldots, a_n, \ldots$, it is not feasible to ask the expert for the degrees corresponding to all possible combinations of the values a_i. In such situations, we can use the fact that from the commonsense viewpoint, a sequence $(a_1, a_2, \ldots)$ is reasonable if and only if a_1 is reasonable *and* a_2 is reasonable, *and* ... For each of the quantities a_i, we can elicit, from the expert, degree to which different values of a_i are reasonable.

Since this is all the information that we have, we need to estimate the degree to which a_1 is reasonable and a_2 is reasonable, and ..., based on the degrees to which a_1 is reasonable, to which a_2 is reasonable, etc. 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 $A \,\&\, B$.

It is worth mentioning that this *estimate* cannot be always *exact*, because our degree of belief in a composite statement $A \,\&\, B$ depends not only on our degrees of belief in A and B,

it also depends on the (usually unknown) dependence between A and B. Let us give an example.

- If A is "coin falls heads", and 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: $d(A \,\&\, B) = 0$.
- However, 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, $d(A') = d(A)$, $d(B') = d(B)$, but $d(A \,\&\, B) \neq d(A' \,\&\, B')$.

In general, let $f_{\&}(a, b)$ be the estimate for $d(A \,\&\, B)$ based on the known values $a = d(A)$ and $b = d(B)$. 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.

Operations with these properties are known as *"and"-operations*, or, alternatively, *t-norms*.

Let us apply an appropriate t-norm to our problem. In our case, for each variable a_i, we only need to find the degrees of belief in two situations: that $a_i = 0$ and that $a_i \neq 0$. Let us denote the degree to which it is reasonable to believe that $a_i = 0$ by $d_i^{=}$, and the degree to which it is reasonable to believe that $a_i \neq 0$ by $d_i^{\neq}$. Thus, we arrive at the following formulation of the first stage of data processing.

Resulting precise formulation of the first stage of data processing in precise terms. Our goal is to select a sequence $(\varepsilon_1, \varepsilon_2, \ldots)$, where each ε_i is equal either to $=$ or to $\neq$. If ε_i is $=$, this means that we have decided that $a_i = 0$, and if ε_i is $\neq$, this means that we have decided that $a_i \neq 0$.

For each such sequence $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots)$, we can determine the degree $d(\varepsilon)$ to which this sequence is reasonable, by applying the selected t-norm $f_{\&}(a, b)$ to the degrees $d_i^{\varepsilon_i}$ to which we belief that each choice ε_i is reasonable:

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

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.

An additional fact that we can use. If we have no information about the signal, i.e., in other words, if there is no evidence that there is a non-zero signal, then the most reasonable choice is to select $x(t) = 0$, i.e., to select a signal for which $a_1 = a_2 = \ldots = 0$.

In other words, if we do not have any way to impose restrictions on the sequence ε, then the most reasonable should be the sequence $(=, =, \ldots)$.

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

A comment about t-norms. In principle, there are many possible t-norms. However, it is known (see, e.g., [15]) that an arbitrary continuous t-norm can be approximated, with an arbitrary accuracy, by an *Archimedean* t-norm, i.e., by a t-norm of the type $f_{\&}(a, b) = f^{-1}(f(a) \cdot f(b))$, for some continuous strictly increasing function $f(x)$. Thus, without losing generality, we can assume that the actual t-norm is Archimedean.

Now, we are ready to formulate and solve the corresponding problem.

IV. Definitions and the Main Result: Fuzzy-Related Techniques Explain Sparsity

Definition 1.

- *By a* t-norm*, we means a function* $f_{\&} : [0, 1] \times [0, 1] \to [0, 1]$ *of the form* $f_{\&}(a, b) = f^{-1}(f(a) \cdot f(b))$*, where* $f : [0, 1] \to [0, 1]$ *is a continuous strictly increasing function for which* $f(0) = 0$ *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 are satisfied:*
  - *the sequence* $\varepsilon_{=} \stackrel{\text{def}}{=} (=, \ldots, =)$ *is more reasonable than all others, i.e.,* $d(\varepsilon_{=}) > d(\varepsilon)$ *for all* $\varepsilon \neq \varepsilon_{=}$*, and*
  - *the sequence* $\varepsilon_{\neq} \stackrel{\text{def}}{=} (\neq, \ldots, \neq)$ *is less reasonable than all others, i.e.,* $d(\varepsilon_{\neq}) < d(\varepsilon)$ *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 its degrees of reasonableness is the largest possible, i.e., if* $d(\varepsilon) = \max_{\varepsilon' \in S} d(\varepsilon')$.

Proposition 1. *Let us assume that the sequences* $d^{=}$ *and* $d^{\neq}$ *properly describe reasonableness. Then, there exist weights* $w_i > 0$ *for which, within each set* S*, a sequence* $\varepsilon \in S$ *is the most reasonable if and only if for this sequence, the sum* $\sum_{i:\varepsilon_i = \neq} w_i$ *is the smallest possible.*

Discussion. In other words, a sequence is the most reasonable if and only if the sum $\sum_{i:a_i \neq p} w_i$ attains the smallest possible value. Thus, fuzzy-based techniques indeed naturally lead to the sparsity condition.

Proof. By definition of the t-norm, we have

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

i.e.,

$$d(\varepsilon) = f_{\&}(d_1^{\varepsilon_1}, \ldots, d_N^{\varepsilon_N}) = f^{-1}(e_1^{\varepsilon_1} \cdot \ldots \cdot e_N^{\varepsilon_N}),$$

where we denoted $e_i^{\varepsilon_i} \stackrel{\text{def}}{=} f(d_i^{\varepsilon_i})$.

Since the continuous function $f(x)$ is strictly increasing, its inverse $f^{-1}(x)$ is also strictly increasing. Thus, maximizing

$d(\varepsilon)$ is equivalent to maximizing the function $e(\varepsilon) \stackrel{\text{def}}{=} f(d(\varepsilon))$. This function has the form

$$e(\varepsilon) = f(d(\varepsilon)) = f(f^{-1}(e_1^{\varepsilon_1} \cdot \ldots \cdot e_N^{\varepsilon_N})),$$

i.e., the form

$$e(\varepsilon) = e_1^{\varepsilon_1} \cdot \ldots \cdot e_N^{\varepsilon_N}.$$

From the condition that the sequences $d^=$ and $d^{\neq}$ properly describe reasonableness, it follows, in particular, that 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

$$\prod_{j=1}^{N} e_j^= > \left(\prod_{j \neq i} e_j^=\right) \cdot e_i^{\neq}.$$

The values $e_j^=$ corresponding to $j \neq i$ cannot be equal to 0, since otherwise, both products would be equal to 0s. Thus, these values are non-zeros. Dividing both sides of the inequality by all these values, we conclude that $e_i^= > e_i^{\neq}$.

Similarly, from the condition that the sequences $d^=$ and $d^{\neq}$ properly describe reasonableness, it also follows, in particular, that for each i, we have $d(\varepsilon_{\neq}) < d(\varepsilon_{\neq}^{(i)})$, where

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

This inequality is equivalent to $e(\varepsilon_{\neq}) > e(\varepsilon_{\neq}^{(i)})$.

Since the values $e(\varepsilon)$ are simply the products, we thus conclude that

$$\prod_{j=1}^{N} e_j^{\neq} < \left(\prod_{j \neq i} e_j^{\neq}\right) \cdot e_i^=.$$

The values $e_j^{\neq}$ corresponding to $j \neq i$ cannot be equal to 0, since otherwise, both products would be equal to 0s.

Thus, for all i, we have $e_i^= > e_i^{\neq} > 0$.

Now, in general, maximizing the product $e(\varepsilon) = \prod_{i=1}^{N} d_i^{\varepsilon_i}$ is equivalent to maximizing the same product divided by a constant $c \stackrel{\text{def}}{=} \prod_{i=1}^{N} d_i^{\neq}$. The ratio $\dfrac{e(\varepsilon)}{c}$ can be equivalently reformulated as $\dfrac{e(\varepsilon)}{c} = \prod_{i:\varepsilon_i=\neq} \dfrac{e_i^=}{e_i^{\neq}}$.

Since logarithm is a strictly increasing function, maximizing this product is, in its turn, equivalent to maximizing its logarithm, i.e., the value

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

where we denoted $w_i \stackrel{\text{def}}{=} \ln\left(\dfrac{e_i^=}{e_i^{\neq}}\right)$. Since $e_i^= > e_i^{\neq} > 0$, we have $\dfrac{e_i^=}{e_i^{\neq}} > 1$ and thus, $w_i > 0$. The proposition is proven.

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

Towards a probabilistic reformulation of the problem. In the probabilistic approach, reasonableness can be described by assigning a prior probability $p(\varepsilon)$ to each possible sequences ε. In this case, out of each set of sequences, we should select the most probable one, i.e., the one with the largest value of the prior probability.

Let $p_i^=$ be the prior probability that $a_i = 0$, and let $p_i^{\neq} = 1 - p_i^=$ be the probability that $a_i \neq 0$. A priori 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.

This assumption is in perfect agreement with the maximum entropy idea (also known as the Laplace's indeterminacy principle), according to which, out of all probability distributions which are consistent with our observations, we should select the one for which the entropy $-\sum p_i \cdot \ln(p_i)$ is the largest possible; see, e.g., [10]. Indeed, if we only know marginal distributions, then the maximum entropy idea implies that, according to the joint distribution, all the random variables are independent.

Under this assumption, $p(\varepsilon) = \prod_{i=1}^{N} p_i^{\varepsilon_i}$. Thus, 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, we define its* prior probability *as*

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

- *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 $\varepsilon \neq \varepsilon_=$, and*
  - *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 its prior probability is the largest possible, i.e., if $p(\varepsilon) = \max_{\varepsilon' \in S} p(\varepsilon')$.*

Proposition 2. *Let us assume that the sequence $p^=$ properly describes reasonableness. Then, there exist weights $w_i > 0$ for which, within each set S, a sequence $\varepsilon \in S$ is the most probable if and only if for this sequence, the sum $\sum_{i:\varepsilon_i=\neq} w_i$ is the smallest possible.*

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

Proof is similar to the Proof of Proposition 1.

Comments.

- 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.
- The comparison of the above two derivations shows an important advantage of fuzzy-based approach in situations like this, when we have a large amount of uncertainty:
  - 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.

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 "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

REFERENCES

[1] 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.

[2] E. J. Candès, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements", *Comm. Pure Appl. Math.*, 2006, Vol. 59, pp. 1207–1223.

[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Transactions on Information Theory*, 2006, Vol. 52, No. 2, pp. 489–509.

[4] E. J. Candès and T. Tao, "Decoding by linear programming", *IEEE Transactions on Information Theory*, 2005, Vol. 51, No. 12, pp. 4203–4215.

[5] E. J. Candès and M. B. Wakin, "An Introduction to compressive sampling", *IEEE Signal Processing Magazine*, 2008, Vol. 25, No. 2, pp. 21–30.

[6] D. L. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory*, 2005, Vol. 52, No. 4, pp. 1289–1306.

[7] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, 2008, Vol. 25, No. 2, pp. 83–91.

[8] T. Edeler, K. Ohliger, S. Hussmann, and A. Mertins, "Super-resolution model for a compressed-sensing measurement setup", *IEEE Transactions on Instrumentation and Measurement*, 2012, Vol. 61, No. 5, pp. 1140–1148.

[9] M. Elad, *Sparse and Redundant Representations*, Springer Verlag, 2010.

[10] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.

[11] G. Klir and B. Yuan, "Fuzzy Sets and Fuzzy Logic", Prentice Hall, Upper Saddle River, New Jersey, 1995.

[12] J. Ma and F.-X. Le Dimet, "Deblurring from highly incomplete measurements for remote sensing", *IEEE Transactions on Geosciences Remote Sensing*, 2009, Vol. 47, No. 3, pp. 792–802.

[13] L. McMackin, M. A. Herman, B. Chatterjee, and M. Weldon, "A high-resolution swir camera via compressed sensing", *Proceedings of SPIE*, 2012, Vol. 8353, No. 1, p. 8353-03.

[14] B. K. Natarajan, "Sparse approximate solutions to linear systems", *SIAM Journal on Computing*, 1995, Vol. 24, pp. 227–234.

[15] 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.

[16] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.

[17] Y. Tsaig and D. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory*, 2006, Vol. 52, No. 4, pp. 1289–1306.

[18] L. Xiao, J. Shao, L. Huang, and Z. Wei, "Compounded regularization and fast algorithm for compressive sensing deconvolution", *Proceedings of the 6th International Conference on Image Graphics*, 2011, pp. 616–621.

[19] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.