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

Fernando Cervantes[1], Bryan Usevitch[1],
Leobardo Valera[2], and Vladik Kreinovich[2]

[1]Department of Electrical and Computer Engineering
[2]Computational Science Program
University of Texas at El Paso, El Paso, TX 79968, USA
fcervantes@miners.utep.edu, usevitch@utep.edu
leobardovalera@gmail.com, vladik@utep.edu

1. 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

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

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

2. 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.

3. Analysis of the Problem

- We are interested in a reconstructing a signal or image $x(t) = \sum\limits_{i=1}^{\infty} a_i \cdot e_i(t)$ 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.

4. 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.

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

6. 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 $d(A \,\&\, B) \neq d(A' \,\&\, B')$.

7. 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.

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

9. 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 $d(\varepsilon) = \max\limits_{\varepsilon' \in S} d(\varepsilon')$.

10. 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.

11. 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.

12. Probabilistic Case: Definitions

- 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 $p(\varepsilon) \stackrel{\text{def}}{=} \prod_{i=1}^{N} p_i^{\varepsilon_i}$.
- 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 $p(\varepsilon) = \max\limits_{\varepsilon' \in S} p(\varepsilon')$.

13. 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.

14. 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.

15. 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.

16. Proof

- By definition of the t-norm, we have
$$d(\varepsilon) = f_\&(d_1^{\varepsilon_1}, \ldots, d_N^{\varepsilon_N}) = f^{-1}(f(d_1^{\varepsilon_1}) \cdot \ldots \cdot f(d_N^{\varepsilon_N})).$$
- 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 $e_i^= > e_i^{\neq}$.

17. Proof (cont-d)

- 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 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.

18. Bibliography: Sparse Methods Are Efficient

- 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.
- 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.
- 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.

19. Sparse Bibliography (cont-d)

- E. J. Candès and T. Tao, "Decoding by linear programming", *IEEE Transactions on Information Theory*, 2005, Vol. 51, No. 12, pp. 4203–4215.
- E. J. Candès and M. B. Wakin, "An Introduction to compressive sampling", *IEEE Signal Processing Magazine*, 2008, Vol. 25, No. 2, pp. 21–30.
- D. L. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory*, 2005, Vol. 52, No. 4, pp. 1289–1306.
- 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.

20. Sparse Bibliography (cont-d)

- 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.
- M. Elad, *Sparse and Redundant Representations*, Springer Verlag, 2010.
- 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.
- 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.

21. Sparse Bibliography (end)

- B. K. Natarajan, "Sparse approximate solutions to linear systems", *SIAM Journal on Computing*, 1995, Vol. 24, pp. 227–234.
- Y. Tsaig and D. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory*, 2006, Vol. 52, No. 4, pp. 1289–1306.
- 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.

Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 22 of 100
Go Back
Full Screen
Close
Quit

22. Bibliography: Fuzzy Techniques

- G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
- 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.
- H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
- L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.