Fuzzy Techniques Provide a Theoretical Explanation for the Heuristic ℓ_p-Regularization of Signals and Images

Fernando Cervantes[1], Bryan Usevitch[1]
Leobardo Valera[2], Vladik Kreinovich[2], and Olga Kosheleva[2]
[1]Department of Electrical and Computer Engineering
[2]Computational Science Program
University of Texas at El Paso
El Paso, TX 79968, USA
contact email vladik@utep.edu

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 some 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 Study

As an example, 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;

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:

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

- For continuous signals, sum turns into an integral:

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

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 $\sum\limits_k g(\Delta x_k) \to \min$.
- *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 $\sum\limits_k g(\Delta x_k)$ is equivalent to minimizing the sum $\sum\limits_k |\Delta x_k|^p$.
- *Fact:* minimizing $\sum\limits_k |\Delta x_k|^p$ under condition $J \le c$ is equivalent to minimizing $J + \lambda \cdot \sum\limits_k |\Delta x_k|^p$.
- *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 $g'(a\cdot z_1) = g'(z_1)\cdot a^p$.

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.

20. What Next?

- So far, we considered the simplest case, when all the membership functions form a 1-D family.
- A natural next step is to consider situations when they form a 2-D family, then a 3-D family, etc.
- In the above proof, we showed that:
  - from the fact the set of the corresponding membership functions is closed under multiplication,
  - we can conclude that that the set of its logarithms forms a linear space.
- In general, each n-dimensional space is formed by linear combinations of n basis functions $f_1(x), \ldots, f_n(x)$.
- Scale-invariance means that the functions $f_i(\lambda \cdot x)$ belongs to the same linear space.

21. What Next: Analysis of the Problem

- So, we conclude that that: $f_i(\lambda \cdot x) = \sum\limits_{j=1}^{n} c_{ij}(\lambda) \cdot f_j(x)$ for some $c_{ij}(\lambda)$.
- Differentiating both sides relative to λ and taking $\lambda = 1$, we get $x \cdot \dfrac{df_i(x)}{dx} = \sum\limits_{j=1}^{n} c'_{ij}(1) \cdot f_j(x)$.
- Here, $\dfrac{dx}{x} = dz$ for $z = \ln(x)$.
- Thus, for $F_i(z) \stackrel{\text{def}}{=} f_i(\exp(z))$, we get a system of linear differential equations with constant coefficients:

$$\frac{dF_i(z)}{dz} = \sum_{j=1}^{n} c'_{ij}(1) \cdot F_j(z).$$

22. Analysis (cont-d)

- *Reminder:* we get a system of linear differential equations with constant coefficients.
- Solutions to such systems are known: they are linear combinations of functions of the type

$$z^k \cdot \exp(a \cdot z) \cdot \sin(\omega \cdot z + \varphi), \text{ where:}$$

  - $k \geq 0$ is a natural number and
  - $a + \omega \cdot \mathrm{i}$ is an eigenvalue of the corresponding matrix.
- Thus, the functions $F_i(x)$ are linear combinations of the functions of the type

$$z^k \cdot \exp(a \cdot z) \cdot \sin(\omega \cdot z + \varphi).$$

- Substituting here $z = \ln(x)$, we conclude that $f_i(x) = -\ln(\mu(x))$ is a linear combination of functions

$$(\ln(x))^k \cdot x^a \cdot \sin(\omega \cdot \ln(x) + \varphi).$$

23. What Next: Towards Conclusions

- *Reminder:* $f_i(x) = -\ln(\mu(x))$ is a linear combination of functions

$$(\ln(x))^k \cdot x^a \cdot \sin(\omega \cdot \ln(x) + \varphi).$$

- Thus, each membership function takes the form $\exp(-f_i(x))$ for such functions $f_i(x)$.
- For a 1-D real-valued matrix, the eigenvalue is a real number, so $\omega = 0$, $k = 0$, and we have $f(x) = x^a$.
- This is exactly what we showed in our main result.
- In the 2-D case:
  - we can have two different real eigenvalues,
  - or we can have double real value,
  - or we can have two mutually conjugate complex eigenvalues.

24. What Next: Conclusions

- For the complex eigenvalues, we do not have monotonicity, so this case has to be dismissed.
- Thus, for the 2-D case, only two options are left.
- In the case of two different eigenvalues, the membership function is equal to $\exp(-a \cdot |d|^p - a' \cdot |d|^{p'})$.
- Thus, regularization is equivalent to the constraint

$$\sum_i |d_i|^p + a \cdot \sum_i |d_i|^{p'} \le C \text{ for some } a \text{ and } p'.$$

- In the case of a double eigenvalue, the membership function is equal to $\exp(-a \cdot |d|^p - a' \cdot |d|^p \cdot \ln(|d|))$.
- Thus, regularization is equivalent to the constraint

$$\sum_i |d_i|^p + a \cdot \sum_i |d_i|^p \cdot \ln(|d_i|) \le C.$$

25. 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.