Why Tensors?

Olga Kosheleva, Martine Ceberio, and Vladik Kreinovich

University of Texas at El Paso
El Paso, Texas 79968, USA
olgak@utep.edu, mceberio@utep.edu
vladik@utep.edu

1. Objective of Science and Engineering

- One of the main objectives: help people select decisions which are the most beneficial to them.
- To make these decisions,
  - we must know people's *preferences*,
  - we must have the information about different *events* – possible consequences of different decisions, and
  - we must also have information about the *degree of certainty*
    * (since information is never absolutely accurate and precise).

2. Partial Orders Naturally Appear in Many Application Areas

- *Reminder:* we need info re *preferences*, *events*, and *degrees of certainty*.
- All these types of information naturally lead to partial orders:
  - For *preferences*, $a < b$ means that b is preferable to a.
    - This relation is used in *decision theory*.
  - For *events*, $a < b$ means that a can influence b.
    - This causality relation is used in *space-time physics*.
  - For *degrees of certainty*, $a < b$ means that a is less certain than b.
    - This relation is used in logics describing uncertainty – such as *fuzzy logic*.

3. Numerical Characteristics Related to Partial Orders

+ An order is a natural way of describing a relation.

– Orders are difficult to process, since most data processing algorithms process *numbers*.

- *Natural idea:* use numerical characteristics to describe the orders.

- *Fact:* this idea is used in all three application areas:

  – in decision making, *utility* describes preferences:

  $$a < b \text{ if and only if } u(a) < u(b);$$

  – in space-time physics, *metric* (and time coordinates) describes causality relation;

  – in logic and soft constraints, numbers from the interval $[0, 1]$ are used to describe degrees of certainty.

4. Need to Combine Numerical Characteristics: Emergence of Polynomial Aggregation Formulas

- In decision making, we need to combine utilities u_1, …, u_n of different participants.
  - Nobelist Josh Nash showed that reasonable conditions lead to $u = u_1 \cdot \ldots \cdot u_n$.
- In space-time geometry, we need to combine coordinates x_i into a metric.
  - Reasonable conditions lead to polynomial metrics

$$s^2 = c^2 \cdot (x_0 - x'_0)^2 - (x_1 - x'_1)^2 - (x_2 - x'_2)^2 - (x_3 - x'_3)^2;$$

$$s^4 = (x_1 - x'_1) \cdot (x_2 - x'_2) \cdot (x_3 - x'_3) \cdot (x_4 - x'_4).$$

- In fuzzy logic, we must combine degrees of certainty d_i in A_i into a degree d for $A_1 \,\&\, A_2$.
  - Reasonable conditions lead to polynomial functions like $d = d_1 \cdot d_2$.

5. Mathematical Observation: Polynomial Formulas Are Tensor-Related

- *Fact:* in many areas, we have a general polynomial dependence

$$f(x_1, \ldots, x_n) = f_0 +$$
$$\sum_{i=1}^n f_i \cdot x_i +$$
$$\sum_{i=1}^n \sum_{j=1}^n f_{ij} \cdot x_i \cdot x_j +$$
$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n f_{ijk} \cdot x_i \cdot x_j \cdot x_k +$$
$$\ldots$$

- *In mathematical terms:* to describe this dependence, we need a finite set of tensors f_0, f_i, f_{ij}, f_{ijk}, ...

6. Towards a General Justification of Polynomial (Tensor) Formulas

- *Fact:* similar polynomials appear in different application areas.
- *Reasonable conclusion:* there must be a common reason behind them.
- *What we do:* we provide such a general reason.

7. Class of Functions

- *Objective:* find a finite-parametric class F of analytical functions $f(x_1, \ldots, x_n)$.
- *Meaning:* $f(x_1, \ldots, x_n)$ approximate the actual complex aggregation function.
- *Reasonable requirement:* this class F is invariant with respect to addition and multiplication by a constant.
- *Conclusion:* the class F is a (finite-dimensional) linear space of functions.
- *Meaning:* invariance w.r.t. multiplication by a constant corresponds to the choice of a measuring unit.
- If we replace the original measuring unit by a one which is λ times smaller, then all the numerical values $\cdot\lambda$:

$$f(x_1, \ldots, x_n) \text{ is replaced with } \lambda \cdot f(x_1, \ldots, x_n).$$

8. Similar Scale-Invariance for the Inputs x_i

- *Similarly:* in all three areas, the numerical values x_i are defined modulo the choice of a measuring unit.
  - If we replace the original measuring unit by a one which is λ times smaller,
  - then all the numerical values get multiplied by this factor λ:

$$x_i \text{ is replaced with } \lambda \cdot x_i.$$

- *Conclusion:* it is reasonable to require that the finite-dimensional linear space F be invariant with respect to such re-scalings:
  - if $f(x_1, \ldots, x_n) \in F$,
  - then for every $\lambda > 0$, the function

$$f_\lambda(x_1, \ldots, x_n) \stackrel{\text{def}}{=} f(\lambda \cdot x_1, \ldots, \lambda \cdot x_n)$$

also belongs to the family F.

9. Definition and the Main Result

Definition. *Let n be an arbitrary integer. We say that a finite-dimensional linear space F of analytical functions of n variables is* scale-invariant *if for every $f \in F$ and for every $\lambda > 0$, the function*

$$f_\lambda(x_1, \ldots, x_n) \stackrel{\text{def}}{=} f(\lambda \cdot x_1, \ldots, \lambda \cdot x_n)$$

also belongs to the family F.

Main result. *For every scale-invariant finite-dimensional linear space F of analytical functions, every element $f \in F$ is a polynomial.*

10. Proof (Part 1)

- Let F be a scale-invariant finite-dimensional linear space F of analytical functions.
- Let $f(x_1, \ldots, x_n)$ be a function from this family F.
- By definition, an analytical function $f(x_1, \ldots, x_n)$ is an infinite series consisting of monomials $m(x_1, \ldots, x_n)$:

$$m(x_1, \ldots, x_n) = a_{i_1 \ldots i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n}.$$

- For each such term, by its *total order*, we will understand the sum $i_1 + \ldots + i_n$.
  - if we multiply each input of this monomial by λ,
  - then the value of the monomial is multiplied by λ^k:

$$m(\lambda \cdot x_1, \ldots \lambda \cdot x_n) = a_{i_1 \ldots i_n} \cdot (\lambda \cdot x_1)^{i_1} \cdot \ldots \cdot (\lambda \cdot x_n)^{i_n} =$$
$$\lambda^{i_1 + \ldots + i_n} \cdot a_{i_1 \ldots i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n} = \lambda^k \cdot m(x_1, \ldots, x_n).$$

11. Proof (Part 2)

- *Reminder:* $f(x_1, \ldots, x_n)$ is a sum of monomials

$$m(x_1, \ldots, x_n) = a_{i_1 \ldots i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n}.$$

- For each monomial, by its order, we will understand the sum $k = i_1 + \ldots + i_n$.
- For each order k, there are finitely many possible combinations of integers $i_1, \ldots, i_n$ for which $i_1 + \ldots + i_n = k$.
- So, there are finitely many possible monomials of the order k.
- Let $P_k(x_1, \ldots, x_n)$ denote the sum of all the monomials of order k in the expansion of $f(x_1, \ldots, x_n)$.
- Then, we have

$$f(x_1, \ldots, x_n) = P_0 + P_1(x_1, \ldots, x_n) + P_2(x_1, x_2, \ldots, x_n) + \ldots$$

12. Proof (Part 3)

- $f(x) = P_0 + P_1(x_1, \ldots, x_n) + P_2(x_1, \ldots, x_n) + \ldots$, where $P_k(x_1, \ldots, x_n)$ is the sum of monomials of order k.
- Some of the sums P_k may be zeros – if the expansion of f has no monomials of the corresponding order.
- Let k_0 be the first index for which the term $P_{k_0}(x_1, \ldots, x_n)$ is not identically 0. Then,

  $$f(x_1, \ldots, x_n) = P_{k_0}(x_1, \ldots, x_n) + P_{k_0+1}(x_1, \ldots, x_n) + \ldots$$

- Since the family F is scale-invariant, it also contains

  $$f_\lambda(x_1, \ldots, x_n) = f(\lambda \cdot x_1, \ldots, \lambda \cdot x_n).$$

- At this re-scaling, each term P_k is multiplied by λ^k.
- Thus, we get

  $$f_\lambda(x) = \lambda^{k_0} \cdot P_{k_0}(x_1, \ldots, x_n) + \lambda^{k_0+1} \cdot P_{k_0+1}(x_1, \ldots, x_n) + \ldots$$

13. Proof (Part 4)

- *Proven:* $f_\lambda(x) = \lambda^{k_0} \cdot P_{k_0}(x) + \lambda^{k_0+1} \cdot P_{k_0+1}(x) + \ldots \in F.$
- Since F is a linear space, it also contains a function

$$\lambda^{-k_0} \cdot f_\lambda(x) = P_{k_0}(x) + \lambda \cdot P_{k_0+1}(x) + \ldots$$

- Since F is finite-dimensional, it is closed under turning to a limit.
- In the limit $\lambda \to 0$, we conclude that the term $P_{k_0}(x)$ also belongs to the family F: $P_{k_0}(x) \in F$.
- Since F is a linear space, this means that the difference

$$f(x) - P_{k_0}(x) = P_{k_0+1}(x) + P_{k_0+2}(x) + \ldots \in F.$$

- Let k_1 be the first index $k_1 > k_0$ for which the term $P_{k_1}(x)$ is not identically 0.
- Then we can similarly conclude that the term $P_{k_1}(x)$ also belongs to the family F, etc.

14. Proof (Conclusion)

- We can therefore conclude that:
  - for every index k for which $P_k(x) \not\equiv 0$,
  - this term $P_k(x)$ also belongs to the family F.
- *Fact:* monomials of different total order are linearly independent:
  - if there were infinitely many non-zero terms P_k in the expansion of the function $f(x)$,
  - we would have infinitely many linearly independent function in the family F
  - which contradicts to our assumption that the family F is a finite-dimensional linear space.
- So, there are only finitely many non-zero P_k.
- Hence, $f(x)$ is a sum of finitely many monomials – i.e., a polynomial.

15. Towards an Alternative Justification Based on Optimality

- *Idea:* we would like to select the *optimal* finite-dimensional family of analytical functions F.
- *What is an optimality criterion:* when we can decide
  - whether F is better than F' (denoted $F' \prec F$)
  - or F' is better than F ($F \prec F'$)
  - or F' is of the same quality as F (denoted $F \equiv F'$).
- *E.g.:* numerical criterion $F \prec F' \Leftrightarrow J(F) < J(F')$.
- *More general case:*
  - when $J(F) = J(F')$, e.g., for average approximation accuracy $J(F)$,
  - we can still choose between F and F' based on some other criteria J' (e.g., computational simplicity).

16. Towards General Description of Optimality

- *Reminder:*
  - when $J(F) = J(F')$, e.g., for average approximation accuracy $J(F)$,
  - we can still choose between F and F' based on some other criteria J' (e.g., computational simplicity).
- The resulting criterion is non-numerical:

  $$F \prec F' \Leftrightarrow J(F) < J(F') \vee (J(F) = J(F') \,\&\, J'(F) < J'(F')).$$
- *General definition:* a (pre)-ordering relation $\preceq$.
- *Natural requirement:* which operation is better should be not depend on the choice of measuring unit:

  $$F \prec F' \Leftrightarrow F_\lambda \prec F'_\lambda,$$

  where $F_\lambda = \{f_\lambda : f \in F\}$.

17. Optimization Approach: Definitions

- We consider the set A of all finite-dimensional spaces of analytical functions.
- By an *optimality criterion*, we mean a *pre-ordering* (i.e., a transitive, reflexive relation) $\preceq$ on the set A.
- An optimality criterion $\preceq$ on the class of all finite-dimensional is called *scale-invariant* if
  - for all F, F', and $\lambda \neq 0$,
  - $F \preceq F'$ implies $F_\lambda \preceq F'_\lambda$.
- An optimality criterion $\preceq$ is called *final* if there exists
  - one and only one space F
  - that is preferable to all the others, i.e., for which $F' \preceq F$ for all $F' \neq F$.

18. Why Final Criterion: Motivations

- *Reminder:* an optimality criterion $\preceq$ is *final* if there exists one and only one optimal space F.
- If no space is optimal relative to some criterion, then this criterion is useless.
- If the criterion selects several spaces F as equally good, then we can also optimize something else.
- *Example:*
  - if F and F' have the same average approximation accuracy,
  - we can select, among them, the one which is easier to compute.
- Thus, such criteria can be adjusted.
- So, for the final criterion, the optimal space is unique.

19. Optimization Approach: Main Result

- *Condition:* $F_{\rm opt}$ is optimal w.r.t. some scale-invariant and final optimality criterion.
- *Conclusion:* all elements of $F_{\rm opt}$ are polynomials.
- *Proof:*
  - optimality means $F \preceq F_{\rm opt}$ for all $F \in A$;
  - in particular, $F_{\lambda^{-1}} \preceq F_{\rm opt}$ for all $F \in A$;
  - due to scale-invariance of $\preceq$, we have $F \preceq (F_{\rm opt})_\lambda$ for all $F \in A$;
  - thus, $(F_{\rm opt})_\lambda$ is optimal;
  - since there is only one optimal space, we have
    $$(F_{\rm opt})_\lambda = F_{\rm opt};$$
  - thus, the space $F_{\rm opt}$ is scale-invariant;
  - we already know that in this case, all $f \in F_{\rm opt}$ are polynomials.

20. What If $f(x_1, \ldots, x_n)$ Is Only Smooth?

Definition. *Let n be an arbitrary integer. We say that a finite-dimensional linear space F of smooth functions of n variables is* affine-invariant *if for every $f \in F$ and for every linear transormation $T : R^n \to R^n$, the function*

$$f_T(x) \stackrel{\text{def}}{=} f(Tx)$$

also belongs to the family F.

Main result. *For every affine-invariant finite-dimensional linear space F of smooth functions, every element $f \in F$ is a polynomial.*

21. Proof: Main Ideas

- Let $f_1(x), \ldots, f_m(x)$ be the basis of F.
- For every $i \le m$, for every variable x_j and for every $\lambda > 0$, we have

$$f_i(x_1, \ldots, x_{j-1}, \lambda \cdot x_j, x_{j+1}, \ldots, x_n) \in F.$$

- Since f_i form a basis, for some $c_{ik}(\lambda)$, we have

$$f_i(x_1, \ldots, x_{j-1}, \lambda \cdot x_j, x_{j+1}, \ldots, x_n) =$$

$$\sum_{k=1}^{m} c_{ik}(\lambda) \cdot f_k(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n).$$

- Differentiating both sides by λ, we get

$$x_j \cdot \frac{\partial f_i}{\partial x_j} = \sum_{k=1}^{m} c_{jk} \cdot f_k.$$

22. Proof (cont-d)

- *Reminder:* $x_j \cdot \dfrac{\partial f_i}{\partial x_j} = \sum\limits_{k=1}^{m} c_{jk} \cdot f_k.$
- For $X_j \stackrel{\text{def}}{=} \ln(x_j)$, we have $\dfrac{\partial f_i}{\partial X_j} = \sum\limits_{k=1}^{m} c_{ik} \cdot f_k.$
- In terms of X_j, we have a system of linear ODEs with constant coefficients.
- A general solution to such a system is a linear combination of terms
  - $\exp(\alpha \cdot X_j) = x_j^\alpha$ (with possible complex α) and
  - $X_j^p \cdot \exp(\alpha \cdot X_j) = x_j^\alpha \cdot \ln^p(x_j)$.
- A general linear transformation leads to different terms – except when we have x_j^α for integer $\alpha \geq 0$.
- Thus, every $f \in F$ is a polynomial in each variable – hence a polynomial.

23. Acknowledgments

This work was supported in part

- by the National Science Foundation grants HRD-0734825 and DUE-0926721,
- by Grant 1 T36 GM078000-01 from the National Institutes of Health,
- by Grant MSM 6198898701 from MŠMT of Czech Republic,
- by Grant 5015 “Application of fuzzy logic with operators in the knowledge based systems” from the Science and Technology Centre in Ukraine (STCU), funded by European Union, and
- by the International Finsler Foundation.

24. References Related to Other Application Areas

- Luce, R. D., Raiffa, R.: *Games and decisions: introduction and critical survey*, Dover, New York, 1989.
- Nguyen H. T., Kosheleva O., Kreinovich, V.: "Decision Making Beyond Arrow's Impossibility Theorem", *International Journal of Intelligent Systems*, 24(1), 27–47 (2009).
- Nguyen, H. T., Walker, E. A.: *A First Course in Fuzzy Logic*, Chapman & Hall/CRC Press, Boca Raton, Florida, 2006.