TRACS: Center for Theoretical Research and its Applications in Computer Science (An Overview)

Faculty supervisors:

Vladik Kreinovich, Luc Longpré, Francois Modave, G. Randy Keller, Scott A. Starks, and Martine Ceberio

University of Texas at El Paso

contact email vladik@cs.utep.edu

Applications Areas (Case Studies)

- Processing geospatial data (PACES, GEON);
- environmental analysis (PACES);
- non-destructive testing of structural integrity of aerospace structures (FAST);
- electronic money, security, and privacy;
- development of methodology to characterize the response of military systems to battlefield threats (ARL);
- nuclear energy safety (ASIC).

Theory: General Overview

- Main objectives of science and engineering:
  - describe the state of the world (data processing);
  - change the state of the world (decision making, control, design).
- First problem: data processing:
  - we are interested in the value of a difficult-to-measure quantity y (e.g., distance to star);
  - to learn y, we measure or estimate the values of auxiliary quantities $x_1, \ldots, x_n$ that are related to y;
  - then, we reconstruct y from these measurement results (estimates) $\widetilde{x}_i$.
- The need for data processing is the main reason why computers are needed.

Data Processing: Problems

- Data processing – reminder:
  - we know the relation between x_i and y;
  - we measure or estimate x_i;
  - we use the estimates $\widetilde{x}_i$ to estimate y.
- Problems:
  - relation between x_i and y is often implicit;
  - measurements and estimates of x_i are never 100% accurate;
  - some measurement and estimates may be outliers;
  - computations are sometimes too slow;
  - to get a better estimate for y, we collect a lot of data, but we do not want to share all this data (privacy).

Data Processing: Problem 1

- *Formulation of the problem:*
  - relation between x_i and y is often implicit:

    $$P_j(x_1, \ldots, x_n, y) = 0, \quad P_k(x_1, \ldots, x_n, y) \geq 0;$$

  - we need an explicit algorithm $x_1, \ldots, x_n \to y$.
- *Solution:* constraint propagation, constraint satisfaction (Ceberio).
- *Application example 1:* outlier detection in gravity data (Beck, Wen); constraint: if $d(x, x') \leq r$, then $|f(x) - f(x')| \leq d$.
- *Application example 2:* detecting and deleting duplicates in gravity and geospatial data (Torres); constraint: if $d(x, x') \leq r_0$, then $x = x'$.
- *Application example 3:* new geometric algorithms for ultrasound testing (E. Roldan, R. Mares).

Data Processing: Problem 2

- *Formulation of the Problem:*
  - measurements are never 100% accurate: $\widetilde{x}_i \neq x_i$;
  - we can have probabilistic, interval, fuzzy uncertainty, and their combinations;
  - $\widetilde{x}_i \neq x_i$ leads to $\widetilde{y} \neq y$.
- *Solution 1:* statistical analysis, interval computations, fuzzy arithmetic, combination of intervals and probabilities (Ceberio, Kreinovich, Longpré).
- *Application example:* environmental analysis – several layers of filtering or pollution.
- *Solution 2:* using data hierarchy – if subsets have intersections, we can decrease error.
- *Application example:* referencing images (Araiza), merging geospatial databases (Torres).

Data Processing: Problem 3

- *Formulation of the problem:* some measurements and estimates $\widetilde{x}_i$ may be outliers.
- *Traditional solution:* $x_i \notin [a - 3\sigma, a + 3\sigma]$.
- *First difficulty:* the estimates for a and σ are affected by outliers.
- *Solution:* iterative estimation: exclude outliers and re-estimate.
- *Application examples:* detecting faults (Seelam, Mulupuru); gravity data (Wen).
- *Second difficulty:* x_i are known with uncertainty.
- *Solution:* new algorithms for detecting outliers under interval uncertainty.
- *Application examples:* fault detection (possible outliers), medical data (guaranteed outliers)

Data Processing: Problem 4

- *Formulation of the problem:* computations are sometimes too slow.
- *Algorithmic solutions:*
  - faster algorithms (e.g., FAST geometric algorithms);
  - real-time algorithms (Longpré);
  - anytime algorithms.
- *Hardware solutions:*
  - parallelization (Torres);
  - quantum computing (Kreinovich, Longpré).

Data Processing: Problem 5

- *Formulation of the problem:*
  - the more information (x_i) we gather, the better our estimate for y;
  - however, from this extensive information, we can extract a lot of things that we may not want to share.
- *Example:* average salary of CS faculty born in 1957.
- *Solution:* for statistical databases, keep only intervals (e.g., 40–50 years old).
- *Related difficulty:* new algorithms are needed to process this interval data.
- *Solution:* new interval-based algorithms (Longpré).
- *Open problems:* multi-D case.

Decision Making

- *Traditional utility approach:* we assume that:
  - the user knows exactly what he wants,
  - and what the situation is.
- *Decision making under incomplete information:*
  - *Interval uncertainty:* division into geological areas (Coblentz).
  - *General unc.:* order-based decision making (Wojciechowski) leads to multi-D probabilities (ARL).
- *Multi-criterion decision making:*
  - *Interval uncertainty:* Instead of (additive) probability measure, $\underline{p}$ and $\overline{p}$ form an non-additive measure (Modave).
  - *General unc.:* order-based decision making (Wojciechowski) leads to multi-D utilities (ARL).

Acknowledgments

This work was supported in part:

- by NASA grants NCC5-209 and NCC2-1232;
- by NSF grants CDA-9522207, EAR-0112968, EAR-0225670, and 9710940 Mexico/Conacyt;
- by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems,
- FAST Center was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, grants F49620-95-1-0518 and F49620-00-1-0365;
- by the Air Force Research Laboratory, Air Force Material Command, USAF, grant F33615-99-C-5211.