How to Efficiently Process Uncertainty within an Cyberinfrastructure without Sacrificing Privacy and Confidentiality

Luc Longpré, Vladik Kreinovich, Eric Freudenthal

Department of Computer Science
University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA
longpre@utep.edu, vladik@utep.edu, efreudenthal@utep.edu

1. Web Services: What They Do

- *Situation:*
  - large amounts of data are stored in different locations;
  - algorithms for processing this data are also implemented at different locations.
- *Traditional solution:* centralization.
- *Drawback:* excessive workload.
- *New solution:* web services (cyberinfrastructure) allow users to submit requests without worrying about the geographic locations of different computational resources (databases and programs).
- Web services enable the user to receive:
  - the desired data $x_1, \ldots, x_n$ and
  - the results $y = f(x_1, \ldots, x_n)$ of processing this data.

2. Web Services: What They Should Do

- *Known fact:* data x_i usually come from measurements.
- *Uncertainty:* in general, the measured values x_i are different from the actual (unknown) values X_i:

$$x_i \neq X_i$$

- *Result:* the result y of data processing is, in general, different from the actual value Y of the desired quantity:

$$y = f(x_1, \ldots, x_n) \neq Y = f(X_1, \ldots, X_n).$$

- *Problem:* gauge this difference.

3. Traditional Methods of Processing Uncertainty

- *In engineering practice:* $\Delta x_i \stackrel{\text{def}}{=} x_i - X_i$ are independent normally distributed with 0 mean and known σ_i.
- *Corollary:* $\Delta y = y - Y$ is also normally distributed, with

$$\sigma = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \cdot \sigma_i^2}.$$

- *We know:* an algorithm that computes f in time T.
- *Automatic differentiation (AD):* computes all $\dfrac{\partial f}{\partial x_i}$ in time $\le 3T$.
- *Computing σ:* in time $O(T + n)$.
- *Interval case:* we only know upper bounds Δ_i on Δx_i.
- *Interval formula:* the upper bound on Δy is

$$\Delta = \sum_{i=1}^{n} \left|\frac{\partial f}{\partial x_i}\right| \cdot \Delta_i.$$

- *Computing Δ:* AD helps us compute Δ in time $O(T + n)$.

4. Need for Privacy Makes the Problem More Complex: Problem and Solution

- *Privacy concern:* owners of f may not want to disclose its code.
- *Result:* we can only use f as a black box.
- *Computational problem:* AD-based $O(T+n)$ methods are not applicable;
- *Possible solution:* compute $\dfrac{\partial f}{\partial x_i}$ by numerical differentiation.
- *Drawback:* computation time $T \cdot n \gg T + n$.
- *Solution for probabilistic uncertainty:* Monte-Carlo simulations compute σ in time $O(T) \ll T \cdot n$.
- *Solution for interval uncertainty:*
  - use an (artificial) Monte-Carlo simulations in which each Δx_i is Cauchy distributed with parameter Δ_i;
  - then simulated Δy is Cauchy distributed with the desired parameter Δ;
  - result: we compute Δ in time $O(T)$.

5. Acknowledgments

This work was supported in part:

- by NASA under cooperative agreement NCC5-209,
- by NSF grant EAR-0225670,
- by NIH grant 3T34GM008048-20S1,
- by Army Research Lab grant DATM-05-02-C-0046,
- by Star Award from the University of Texas System,
- and by Texas Department of Transportation grant No. 0-5453.