Interval and p-Box Techniques for Model Validation: on the Example of the Thermal Challenge Problem

Vladik Kreinovich

Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, Texas, 79968, USA

office phone (915) 747-6951
email vladik@utep.edu
http://www.cs.utep.edu/vladik

Realistic Measurement Situations

- Often, the measurement result z depends:
  - not only on the measured value x, but also
  - on the parameters s of the experiment's setting
  - and on the values of some auxiliary quantities y.
- The dependence $z = f(x, s, y)$ is usually known.
- *Ideal case:* we know y, so we find x.
- *Real case:* we know y with some uncertainty.
- *Usually:* uncertainty in y leads to extra measurement error in x.
- *Good news:* often, we can combine multiple measurement results and decrease influence of y's uncertainty.
- We get *sub-noise* measurement accuracy: better than the accuracy with which we know y.

Example: Multi-Spectral Imaging

- We measure $\widetilde{I}(f,\vec{p}) = I(f,\vec{p}) + D(f,\vec{p})$, where:
  - $I(f,\vec{p}) = C(f)\cdot I(\vec{p})$ is the intensity of the source on frequency f at point p;
  - $D(f,\vec{p})$ is the intensity of dust radiation.
- Often, $D \gg I$, so we cannot determine the object's structure.
- We know how D depends on f: $D(f,\vec{p}) = D(\vec{p})\cdot f^{\alpha}$.
- Here, $x = I$, $s = f$, $y = D$, and

$$z = f(x,s,y) = C(s)\cdot x + y\cdot s^{\alpha}.$$

- Based on two observations $z_i = C(s_i)\cdot x + y\cdot s_i^{\alpha}$, we can apply linear algebra ideas to eliminate y:

$$z_1\cdot s_2^{\alpha} - z_2\cdot s_1^{\alpha} = x\cdot (C(s_1)\cdot s_2^{\alpha} - C(s_2)\cdot s_1^{\alpha}).$$

- *Result:* we uncover previously unseen spiral and ring-like structures in distant galaxies.

VLBI Astrometry

- *Very Large Baseline Interferometry (VLBI):* we simultaneously observe a distant radiosource by two (or more) radioantennas i, j.

- *Ideal case:* time delay between the two antennas
$$\tau_{i,j,k} = \frac{1}{c} \cdot \vec{b}_{i,j} \cdot \vec{s}_k.$$

- Synchronization is not perfect ($\Delta t_i \neq 0$), hence
$$\tau_{i,j,k} = \frac{1}{c} \cdot \vec{b}_{i,j} \cdot \vec{s}_k + \Delta t_i - \Delta t_j.$$

- Here, $z = \tau$, $x = \vec{s}_k$, $y = (\vec{b}_{i,j}, \Delta t_i)$.

- Measurement error in τ corresponds to accuracy $\approx 0.001''$, but inaccuracy in Δt_i is much worse.

- *Differential astrometry:*
$$\Delta\tau_{i,j,k,l} = \frac{1}{c} \cdot \vec{b}_{i,j} \cdot \Delta\vec{s}_{k,l},$$
where $\Delta\tau_{i,j,k,l} \stackrel{\text{def}}{=} \tau_{i,j,k} - \tau_{i,j,l}$, drastically improves the accuracy.

VLBI Astrometry: Arc Method

- To get rid of baseline vectors, we need 4 antennas:

$$\Delta\tau_{1,2,k,l} = \frac{1}{c}\cdot \vec{b}_{1,2}\cdot \Delta\vec{s}_{k,l};\quad \Delta\tau_{2,3,k,l} = \frac{1}{c}\cdot \vec{b}_{2,3}\cdot \Delta\vec{s}_{k,l},$$

$$\Delta\tau_{3,4,k,l} = \frac{1}{c}\cdot \vec{b}_{3,4}\cdot \Delta\vec{s}_{k,l}.$$

- For the dual basis $\vec{B}_{i,j}\cdot \frac{1}{c}\cdot \vec{b}_{i,j} = \delta_{(i,j),(i',kj')}$, we get

$$\vec{s}_{k,l} = \Delta\tau_{1,2,k,l}\cdot \vec{B}_{1,2} + \Delta\tau_{2,3,k,l}\cdot \vec{B}_{2,3} + \Delta\tau_{3,4,k,l}\cdot \vec{B}_{3,4}.$$

- Express $\vec{B}_{i,j}$ as a linear combination of $\vec{s}_{1,2}$, $\vec{s}_{1,3}$, $\vec{s}_{1,4}$.
- For any other source k, we have a similar expression

$$\vec{s}_{k,1} = \vec{s}_k - \vec{s}_1 = \Delta\tau_{1,2,k,1}\cdot \vec{B}_{1,2} + \Delta\tau_{2,3,k,1}\cdot \vec{B}_{2,3} + \Delta\tau_{3,4,k,1}\cdot \vec{B}_{3,4}.$$

- Hence, $\vec{s}_k$ is a linear combinations of $\vec{s}_{1,2}$, $\vec{s}_{1,3}$, $\vec{s}_{1,4}$.
- We have a linear transformation T between the actual and the observed values $\vec{s}_k$.
- Since $\|\vec{s}_k\| = 1$, T is rotation.
- So, we can determine positions modulo rotation.

VLBI Imaging

- *Problem:* find the image $I(\vec{p})$.
- *Solution:* find Fourier transform $F(\vec{b})$ of $I(\vec{p})$.
- *Ideal case:* the phase shift $\widetilde{\varphi}_{i,j}$ between the signals observed by antennas i and j is equal to the phase $\varphi_{i,j}$ of $F(\vec{b}_{ij})$.
- *In reality:* due to synchronization errors $\Delta\varphi_i$,

$$\widetilde{\varphi}_{i,j} = \varphi_{i,j} + \Delta\varphi_i - \Delta\varphi_j.$$

- Here, $z = \widetilde{\varphi}_{i,j}$, $x = \varphi_{i,j}$, $y = \Delta\varphi_i$.
- *Closure phase method* eliminates the effect of the auxiliary parameters by considering the "closure phase" $\widetilde{\varphi}_{ij} + \widetilde{\varphi}_{jk} + \widetilde{\varphi}_{ki}$ for which:

$$\widetilde{\varphi}_{ij} + \widetilde{\varphi}_{jk} + \widetilde{\varphi}_{ki} = \varphi_{ij} + \varphi_{jk} + \varphi_{ki}.$$

Image Georeferencing

- *Problem:* find the relative orientation of geospatial images $I_1(\vec{p})$ and $I_2(\vec{p})$.
- *Problem reformulated:* find shift, rotation angle, and scaling between the images.
- *Difficulty:* to find an angle with accuracy of 1°, we need 360 tests; we need 4 parameters, so we need $360^4 \approx 10^9$ tests – practically impossible.
- *Idea:* separate the problem – find rotation angle and scaling separately from finding the shift.
- *Fact:* in Fourier domain, when $I_2(\vec{p}) = I_1(\vec{p} + \vec{a})$, then $F_2(\vec{\omega}) = F_1(\vec{\omega}) \cdot \exp(\mathrm{i} \cdot \vec{\omega} \cdot \vec{a})$.
- Here, $x = F(\vec{\omega})$, $y = \vec{a}$.
- *Solution:* the shift-independent combination is the absolute value $|F_i(\vec{\omega})|$.

Measuring Strong Electric Currents

- *Problem:* measuring the cable current I at an aluminum plant.
- *Specifics:* I is difficult to measure directly.
- *Specifics:* I is measured by its magnetic field E.
- *Ideal case (single cable):* $E = I/r$, where r is the distance between the sensor and the cable's axis.
- *Real plants:* there is often an auxiliary nearby cable.
- Here, $z = E$, $x = I$, $s =$ sensor locations,
  $y =$ location and current in the auxiliary cable.
- *Difficulty:* $z = f(x, s, y)$ non-linearly depends on the (unknown) location of the auxiliary cable.
- *Solution:* combining the measurements from different sensors eliminates the influence of the auxiliary cable.

Ultrasonic Non-Destructive Testing (in brief)

- *Problem:* find the location and orientation of hidden faults in a plate.
- *Related active measurements:*
  - send ultrasonic Lamb waves to the plate;
  - measure the waves that propagated along the plate.
- *Difficulty:* the resulting signals depend both on the location and on the orientation of the fault.
- *Idea:* separate the effects of location and orientation.
- *Solution:* by appropriately combining sensor readings, we can minimize the effect of location.
- Thus, we can easily determine the fault's orientation.

Formulation of the General Problem

- General problem:
  - *Objective:* we are interested in n_x scalar parameters that form x.
  - *Measurement situation:* each n_z-component measurement result z depends not only on x, but also on n_y components of the auxiliary quantity(-ies) y: $z = f(x, s, y)$.
  - *Desirable objective:* determine x without knowing y precisely.
- Two possible situations:
  - y is fixed (cannot be varied), but we can change s. *Example:* multi-spectral imaging.
  - We cannot change the settings s, but we can use different values of y. *Example:* VLBI astrometry.

Variable Settings: Analysis of the Problem

- *Situation:* after we performed the measurement in N_s different settings $s_1, \ldots, s_{N_s}$, we get N_s measurement results $z_1, \ldots, z_{N_s}$.
- *Situation:* we do not know y.
- *Conclusion:* select N_s so that we will be able to uniquely determine both x and y.
- After N_s measurements, we have N_s n_z-component equations $z_i = f(x, s_i, y)$ to determine n_x unknown components of x and n_y unknown components of y.
- *Fact:* # of equations must be $\geq$ # of unknowns.
- We have $N_s \cdot n_z$ scalar equations for $n_x + n_y$ unknowns.
- *Recommendation:* perform the measurements in at least $N_s \geq (n_x + n_y)/n_z$ different settings.

Practical Question: How to Solve the System of Equations?

- *Difficulty:* in general, the dependence $z = f(x, y)$ is non-linear.
- So, we have a system of non-linear equations.
- *What helps:* often, we know good approximations $x^{(0)}$ and $y^{(0)}$ to x and y.
- *How it helps:*
  - We only need to find $\Delta x \stackrel{\text{def}}{=} x - x^{(0)}$ and $\Delta y \stackrel{\text{def}}{=} y - y^{(0)}$.
  - Usually, Δx and Δy are small.
  - So, we can expand $f(x, y)$ in Taylor series in Δx and Δy and ignore 2nd and higher order terms.
  - As a result, to find Δx and Δy, we get an easier-to-solve system of *linear* equations.

Variable Settings: Example

- *Case study:* multi-spectral astronomical imaging.
- *Reminder:* $\widetilde{I}(f, \vec{p}) = C(f) \cdot I(\vec{p}) + D(\vec{p}) \cdot f^{\alpha}$.
- Here, $z = \widetilde{I}$, $x = I$, $s = f$, $y = D$, and

$$z = f(x, s, y) = C(s) \cdot x + y \cdot s^{\alpha}.$$

- *Specifics:* $n_z = 1$, $n_x = 1$, and $n_y = 1$.
- *General recommendation:* we must have at least $(n_x + n_y)/n_z = (1 + 1)/1 = 2$ settings.
- *Confirmation:* we have shown that, based on measurements in two different settings

$$z_1 = C(s_1) \cdot x + y \cdot s_1^{\alpha}, \quad z_2 = C(s_2) \cdot x + y \cdot s_2^{\alpha},$$

we can uniquely determine the desired value x:

$$z_1 \cdot s_2^{\alpha} - z_2 \cdot s_1^{\alpha} = x \cdot (C(s_1) \cdot s_2^{\alpha} - C(s_2) \cdot s_1^{\alpha}).$$

Different Values of y: Analysis

- *General idea:* we measure several (N_x) objects x_i.
- *General idea:* we measure each object under several (N_y) circumstances y_j, $j = 1, \ldots, N_y$.
- Based on the results $z_{i,j} = f(x_i, y_j)$ of these measurements, we must be able to determine x_i and y_j.
- *Example:* in VLBI astrometry example, we observe several sources x_i by using several radiotelescopes y_j.
- After $N_x \cdot N_y$ measurements of z, we get $n_z \cdot N_x \cdot N_y$ scalar equations.
- We must find N_x vectors x_i with n_x components/x.
- We must find N_y vectors y_j with n_y components/y.
- *Recommendation:* select N_x and N_y so that:

$$n_z \cdot N_x \cdot N_y \geq N_x \cdot n_x + N_y \cdot n_y.$$

Different Values of y: Good News and Bad News

- *Recommendation:* $n_z \cdot N_x \cdot N_y \geq N_x \cdot n_x + N_y \cdot n_y$.
- *Good news:* this inequality is true when N_x and N_y are large enough.
- *Good news:* often, we know reasonably good approximations $x_i^{(0)}$ and $y_j^{(0)}$, so we can linearize.
- *Bad news:* sometimes, we cannot uniquely determine x_i and y_j even for large N_x and N_y.
- *Example:* in astrometry, we cannot uniquely determine directions to the sources $\vec{s}_i$.
- *Reason:* if we rotate all the directions $\vec{s}_i$ and $\vec{b}_{i,j}$, we get the same time delays.
- *What we can determine in this case:* coordinates of the sources $\vec{s}_i$ modulo rotations.

How Can we Describe Such Non-General Situations? Enter Transformation Groups

- *Problem:*
  - we measure all the objects x for all the values y,
  - we cannot determine all the values x and y.
- *Reformulation:*
  - even when we know all the values $f(x, y)$,
  - there exist values $T_x(x) \neq x$ and $T_y(y) \neq y$ for which the measurement results are exactly the same:

$$f(x, y) = f(T_x(x), T_y(y)).$$

- Such pairs of transformations form a *group* G.
- We can only find x modulo transformations $\in G$.
- *Example:* in astrometry, we have rotations group.

Thermal Challenge Problem: In Brief

- *Objective:* make sure that:
  - for a manufacturing-related distribution of thermal properties k and ρC_p (as given by samples),
  - for given time t, thickness L, and heat flux q,
  - the probability P that a temperature T exceeds a given threshold T_0 should be $\geq 1 - p_0$ (=0.99).
- *We know:* an approximate model $T \approx f(k, \rho C_p, t, L, q)$.
- *Complexity:* it is difficult to measure T for high q.
- *We have performed:*
  - several experiments for smaller q, and
  - one extra (accreditation) experiment for a large q.
- *Problem:* use the known data to check whether

$$P \stackrel{\text{def}}{=} \text{Prob}(T \leq T_0) \geq 1 - p_0.$$

Thermal Challenge Problem

- *How this problems fits into our general framework:*
  - measured quantity z: temperature $z = T$;
  - known auxiliary quantity: time $s_1 = t$;
  - unknown auxiliary quantities: $y_1 = k$, $y_2 = \rho C_p$;
  - we know the $\approx$ dependence $z_1 \approx f(s_1, y_1, y_2)$.
- *Additional complexity:* the model is only approximate:

$$\left|z^{(k)} - f(s_1^{(k)}, y_1^{(k)}, y_2^{(k)})\right| \le \varepsilon$$

for some (unknown) accuracy ε.

- *Natural idea:* once, for a sample, we know $z^{(k)} = T$ for different moments $t = s^{(k)}$, we find y_1 and y_2 for which $\varepsilon \to \min$, where:

$$\left|z^{(k)} - f(s_1^{(k)}, y_1, y_2)\right| \le \varepsilon.$$

How to Implement the Above Idea

- *Linearizable case:* we know approximate values $y_1^{(0)}$ and $y_2^{(0)}$ such that the differences $\Delta y_i \stackrel{\text{def}}{=} y_i - y_i^{(0)}$ are small (hence quadratic terms can be ignored).
- *Resulting solution:* solve a linear programming problem

$$\varepsilon \to \min$$

under the conditions

$$-\varepsilon \le z^{(k)} - f(s^{(k)}, y_1^{(0)}, y_2^{(0)}) - \frac{\partial f}{\partial y_1} \cdot \Delta y_1 - \frac{\partial f}{\partial y_2} \cdot \Delta y_2 \le \varepsilon.$$

- *General case*– use Newton's approach:
  - we solve a linearized system, find Δy_i; then
  - we take $y_i^{(0)} + \Delta y_i$ as a new initial approximation;
  - repeat until the process converges.

Solving the Thermal Challenge Problem: First Approximation

- *Objective:* check that for given s, y_1, and y_2, we have $z \le z_0$ with probability $\ge 1 - p_0$ (=0.99).
- *Preliminary analysis:* for each object v, we use the records $T_v(t)$ to find $y_1 = k$, $y_2 = \rho C_p$, and ε_v.
- *Gauging the model's accuracy:* we take $\varepsilon \stackrel{\text{def}}{=} \max\limits_v \varepsilon_v$ as the measure of the model's accuracy.
- *Reformulating the objective:* check that

$$P_0 \stackrel{\text{def}}{=} \text{Prob}(f(s, y_1, y_2) \le z_0 - \varepsilon) \ge 1 - p_0.$$

- *Assumption:* y_1, y_2 are independent normally distributed; we find means and st. dev. from given data.
- *Resulting approach:* for these normal distributions, we check whether $P_0 \ge 1 - p_0$ by using linearization (when z is also normal) or Monte-Carlo simulations.

Towards More Accurate Description

- *Fact:*
  - for some values of the parameters s_i, measurements are easier;
  - for some, they are more difficult.
- *Example:* for the thermal challenge problem, this parameter is the thermal flow $s_2 = q$.
- *Consequence:* we have more data for easier-to-measure values.
- *Consequence:* the model is more accurate for easier-to-measure values of the parameters
- *How to take this fact into account:*
  - instead of a single measure ε of the model's accuracy ε,
  - we explicitly consider the dependence $\varepsilon(s_2, \ldots)$.

Towards More Accurate Description: Specific Implementation

- *Selecting a model for $\varepsilon(q)$:* due to scale-invariance, we take $\varepsilon(q) = \varepsilon_0 \cdot q^\alpha$ for some ε_0 and α.
- *Preliminary analysis:* for each experimentally tested q, based on all samples with given q, we find
$$\varepsilon(q) = \max_{v: q(v) = q} \varepsilon(v).$$
- *Estimating parameters of the $\varepsilon(q)$ model:* we must find ε_0 and α for which $\varepsilon(q) \approx \varepsilon_0 \cdot q^\alpha$.
- *Algorithm:* we use the Least Squares method (LSM) to solve a system of linear equations
$$\ln(\varepsilon(q)) \approx \ln(\varepsilon_0) + \alpha \cdot \ln(q)$$
with unknowns $\ln(\varepsilon_0)$ and α.
- *Final step:* we use the accreditation experiment to improve the accuracy of the $\varepsilon(q)$ model.

Additional Idea: How to Simplify Computations

- *Fact:* in the given formula

$$T(x,t) = T_i + \frac{q\cdot L}{k}\cdot\left[\frac{(k/\rho C_p)\cdot t}{L^2} + \frac{1}{3} - \frac{x}{L} + \frac{1}{2}\cdot\left(\frac{x}{L}\right)^2 - \right.$$

$$\left.\frac{2}{\pi^2}\cdot\sum_{n=1}^{6}\frac{1}{n^2}\cdot e^{-n^2\cdot\pi^2\cdot\frac{(k/\rho C_p)\cdot t}{L^2}}\cdot\cos\left(n\cdot\pi\cdot\frac{x}{L}\right)\right]$$

ρC_p always appears in a ratio $\dfrac{k/\rho C_p}{L^2}$.

- *Resulting idea:*
  - instead of $y_1 = k$ and $y_2 = \rho C_p$,
  - we should use $y_1 = \dfrac{q\cdot L}{k}$ and $y_2 = \dfrac{k/\rho C_p}{L^2}$:

$$T(x,t) = T_i + y_1\cdot\left[y_2\cdot t + \frac{1}{3} - x_0 + \frac{1}{2}\cdot x_0^2 - \right.$$

$$\left.\frac{2}{\pi^2}\cdot\sum_{n=1}^{6}\frac{1}{n^2}\cdot e^{-n^2\cdot\pi^2\cdot y_2\cdot t}\cdot\cos(n\cdot\pi\cdot x_0)\right],$$

where $x_0 \stackrel{\text{def}}{=} \dfrac{x}{L}$.

From Validating a Model to Improving a Model

- *Assumption:* the formula assumes that $y_1 = k$ and $y_2 = \rho C_p$ are constants.
- *Fact:* the average value $\bar{k}$ of $y_1 = k$ grows with temperature T:

T	20	250	500	750	1000
$\bar{k}$	0.49	0.59	0.63	0.69	0.75

- *Natural conclusion:* y_1 is a function of T; example:

$$y_1 \approx a + b \cdot T; \quad \text{LSM:} \;\; a \approx 0.63, \quad b \approx \frac{0.06}{250}.$$

- *Resulting idea:* plug in $y_1(T) = y_1(20) + b \cdot T$ into the original formula and hope for the better fit.
- *Another idea:* try to match the difference between z and $f(s, y_1, y_2)$ by an empirical model.
- *Example:* try a linear dependence for this difference.

Acknowledgments

This work was supported in part:

- by NASA under cooperative agreement NCC5-209;
- by NSF grants NSF grants EAR-0112968, EAR-0225670, and EIA-0321328;
- 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, grant F49620-00-1-0365;
- by the Army Research Laboratories grant DATM-05-02-C-0046.

References

- S. Ferson, *RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers*, CRC Press, Boca Raton, Florida, 2002.
- L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis*, Springer-Verlag, London, 2001.
- V. Kreinovich et al., Interval computations website http://www.cs.utep.edu/interval-comp
- R. Osegueda, V. Kreinovich, et al., "Towards a General Methodology for Designing Sub-Noise Measurement Procedures", *Proc. 10th IMEKO TC7 Int'l Symp. on Advances of Measurement Science*, St. Petersburg, Russia, June 30–July 2, 2004, Vol. 1, pp. 59–64; http://www.cs.utep.edu/vladik/2004/tr04-12.pdf