Towards a Combination of Interval and Ellipsoid Uncertainty

Vladik Kreinovich[1], Arnold Neumaier[2], and Gang Xiang[1,3]

[1]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
email vladik@utep.edu
[2]Fakultät für Mathematik
Universität Wien
Nordbergstr. 15
A-1090 Wien, Austria
email Arnold.Neumaier@univie.ac.at
[3]Philips Healthcare Informatics
Business Line RIS
6006 N. Mesa Suite 709
El Paso, TX 79912, USA
email gxiang@acm.org

Abstract

In many real-life situations, we do not know the probability distribution of measurement errors but only upper bounds on these errors. In such situations, once we know the measurement results, we can only conclude that the actual (unknown) values of a quantity belongs to some interval. Based on this interval uncertainty, we want to find the range of possible values of a desired function of the uncertain quantities. In general, computing this range is an NP-hard problem, but in a linear approximation, valid for small uncertainties, there is a linear time algorithm for computing the range. In other situations, we know an ellipsoid that contains the actual values. In this case, we also have a linear time algorithm for computing the range of a linear function.

In some cases, however, we have a combination of interval and ellipsoid uncertainty. In this case, the actual values belong to the intersection of a box and an ellipsoid. In general, estimating the range over the intersection enables us to get a narrower range for the function. In this paper, we provide two algorithms for estimating the range of a linear function over an intersection: a simpler $O(n\,\log(n))$ algorithm and a (somewhat more complex) linear time algorithm. Both algorithms can be extended to the l^p-case, when instead of an ellipsoid we have a set defined by a p-norm.

1 Formulation of the Problem

Interval uncertainty: brief reminder. Measurements are never 100% accurate; hence, the measurement result $\widetilde{x}_i$ is, in general, different from the actual (unknown) value x_i of the corresponding quantity. Traditional engineering approach to processing measurement uncertainty assumes that we know the probability distribution of measurement errors $\Delta x_i := \widetilde{x}_i - x_i$.

In many practical situations, however, we do not know these probability distributions. In particular, in many real-life situations, we only know the upper bound Δ_i on the (absolute value of the) measurement error: $|\Delta x_i| \le \Delta_i$. In such situations, the only information that we get about the actual (unknown) value x_i after the measurement is that x_i belongs to the interval $\boldsymbol{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$; see, e.g., [19].

Data processing under interval uncertainty: brief reminder. In addition to the values of the measured quantities $x_1, \ldots, x_n$, we often need to know the values of other quantities which are related to x_i by a known dependence $y = f(x_1, \ldots, x_n)$. When we know x_i with interval uncertainty, i.e., when we know that $x_i \in \boldsymbol{x}_i$, then the only conclusion about y is that y belongs to the range $\{f(x_1, \ldots, x_n) \,|\, x_1 \in \boldsymbol{x}_1, \ldots, x_n \in \boldsymbol{x}_n\}$ of the function $f(x_1, \ldots, x_n)$ over the box $\boldsymbol{x}_1 \times \ldots \times \boldsymbol{x}_n$.

Data processing: linear approximation. In general, computing this range is NP-hard – even for quadratic functions f; see, e.g., [12]. However, in many practical situations, the measurement errors are small, thus, the intervals $\boldsymbol{x}_i$ are narrow, and so, on the box $\boldsymbol{x}_1 \times \ldots \times \boldsymbol{x}_n$, we can safely replace the original function $f(x_1, \ldots, x_n)$ by the first two terms in its Taylor formula: $f(x_1, \ldots, x_n) = \widetilde{y} + \sum\limits_{i=1}^n c_i\,\Delta x_i$, where $y_0 := f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ and $c_i := \dfrac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$.

For such linear functions, the range is equal to $[\widetilde{y} - \Delta, \widetilde{y} + \Delta]$, where $\Delta = \sum\limits_{i=1}^n |c_i|\,\Delta_i$. The maximum value Δ of the difference $f - \widetilde{y} = \sum\limits_{i=1}^n c_i\,\Delta x_i$ is attained when $\Delta x_i = \Delta_i$ for $c_i \ge 0$ and $\Delta x_i = -\Delta_i$ for $c_i < 0$; correspondingly, the smallest value $-\Delta$ is attained when $\Delta x_i = -\Delta_i$ for $c_i \ge 0$ and $\Delta x_i = \Delta_i$ for $c_i < 0$.

Once we know the derivatives c_i and the bounds Δ_i, $i = 1, \ldots, n$, the value Δ describing the desired range can be computed in linear time $O(n)$.

Comment. To get a guaranteed enclosure for y, we must add to this linear range an interval $[-\delta, \delta]$ which bounds the second and higher order terms in the Taylor expansion; this is, in effect, what is known in interval computations as centered form; see, e.g., [11, 14, 15]. Asymptotically, $\delta = O(\sum \Delta_i^2)$, so we get an asymptotically exact enclosure for the range in linear time.

Ellipsoid uncertainty: a brief reminder. In some cases, the information about the values $\Delta x_1, \ldots, \Delta x_n$ comes not as a bound on the values Δx_i themselves, but rather as a bound $z \leq z_0$ on some quantity $z = g(\Delta x_1, \ldots, \Delta x_n)$ which depends on Δx_i.

When the measurement errors are small, we can expand the function g into a Taylor series and keep only the lowest terms in this expansion. In particular, if we keep quadratic terms, we get a quadratic zone $g(\Delta x_1, \ldots, \Delta x_n) \leq z_0$. If this zone is a bounded set, then it describes an ellipsoid. In this case, the only information about the tuple $\Delta x = (\Delta x_1, \ldots, \Delta x_n)$ is that it belongs to this ellipsoid.

Another situation when we get such an ellipsoid uncertainty is when measurement errors are independent normally distributed random variables, with 0 mean and standard deviations σ_i. In this case, the probability density is described by the known formula $\rho(\Delta x) = \text{const} \cdot \exp\left(-\sum\limits_{i=1}^{n} \dfrac{\Delta x_i^2}{2\sigma_i^2}\right)$. This probability density $\rho(\Delta x)$ is everywhere positive; thus, in principle, an arbitrary tuple Δx is possible. In practical statistics, however, tuples with very low probability density $\rho(\Delta x)$ are considered impossible.

For example, in 1-dimensional case, we have a “three sigma” rule: values for which $|\Delta x_1| > 3\sigma_1$ are considered to be almost impossible. In multi-dimensional case, it is natural to choose some threshold $t > 0$, and consider only tuples for which $\rho(\Delta x) \geq t$ as possible ones. This formula is equivalent to $\ln(\rho(\Delta x)) \geq \ln(t)$. For Gaussian distribution, this equality takes the form $\sum\limits_{i=1}^{n} \dfrac{\Delta x_i^2}{\sigma_i^2} \leq r^2$ for some appropriate value r – i.e., the form of an ellipsoid. The sum is $\chi^2(n)$ distributed, with expectation n and standard deviation $\sqrt{n}$, so here, $r^2 = n + O(\sqrt{n})$ is a natural choice. In this paper, we will consider ellipsoids of this type.

Comment. If the measurement errors are small but *not independent*, then we also have an ellipsoid, but with a general positive definite quadratic form in the left-hand side of the inequality.

Ellipsoids are also known to be the *optimal* approximation sets for different problems with respect to several reasonable optimality criteria; see, e.g., [8, 13].

Ellipsoid error estimates are actively used in different applications; see, e.g., [1, 3, 4, 7, 9, 16, 20, 21].

Data processing under ellipsoid uncertainty: linear approximation. The range of a linear function $\sum\limits_{i=1}^n c_i\,\Delta x_i$ over an ellipsoid can be easily computed by using, e.g., the Lagrange multiplier method. First, one can easily check that the maximum of a linear function is attained at the border of the ellipsoid, i.e., when $\sum\limits_{i=1}^n \dfrac{\Delta x_i^2}{\sigma_i^2} = r^2$. Maximizing the linear function $\sum\limits_{i=1}^n c_i\,\Delta x_i$ under the above constraint is equivalent to solving the unconstrained optimization problem $\sum\limits_{i=1}^n c_i\,\Delta x_i + \lambda \sum\limits_{i=1}^n \dfrac{\Delta x_i^2}{\sigma_i^2}$, where λ is the Lagrange multiplier. For every $i = 1,\ldots,n$, differentiating with respect to Δx_i and equating the derivative to 0, we conclude that the maximum value Δ of the linear function is attained when $\Delta x_i = \alpha\, c_i \sigma_i^2$ for $\alpha = -\dfrac{1}{2\lambda}$. Here, the parameter α is determined by the condition that $\sum\limits_{i=1}^n \dfrac{\Delta x_i^2}{\sigma_i^2} = r^2$ – i.e., that $\alpha^2 \sum\limits_{i=1}^n c_i^2\,\sigma_i^2 = r^2$ and $\alpha = r/\sqrt{\sum c_i^2\sigma_i^2}$. The smallest possible value $-\Delta$ of this function is attained when $\Delta x_i = -\alpha\, c_i \sigma_i^2$ for all $i = 1,\ldots,n$.

The corresponding value Δ is equal to $\Delta = r\sqrt{\sum c_i^2\sigma_i^2}$. This value can also be computed in linear time.

Need for combining interval and ellipsoid uncertainty. In some practical cases, we have a combination of interval and ellipsoid uncertainty. For example, in the statistical case, we may have an ellipsoid bound and also the 3 sigma bound $|\Delta x_i| \le 3\sigma_i$ for each measurement error.

In this case, the actual values $(\Delta x_1,\ldots,\Delta x_n)$ belong to the *intersection* of the box $\boldsymbol{x}_1 \times \ldots \times \boldsymbol{x}_n$ and the ellipsoid.

In general, the smaller the set over which we estimate the range of a given function, the narrower the resulting range. It is therefore desirable to be able to estimate the range of a linear function $\sum\limits_{i=1}^n c_i\,\Delta x_i$ over such an intersection.

What we do in this paper: main result. In this paper, we provide two algorithms for estimating the range of a linear function over an intersection in linear time: a simpler $O(n\,\log(n))$ algorithm and a (somewhat more complex) linear time algorithm.

From ellipsoids to generalized ellipsoids. We have mentioned that ellipsoids correspond to normal distributions. In many practical cases, the distribution of the measurement errors is different from normal; see, e.g., [17, 18, 19].

In many such cases, we have a distribution of the type

$$\rho(\Delta x) = \text{const} \cdot \exp\left(-\sum_{i=1}^{n} \frac{|\Delta x_i|^p}{k\sigma_i^p}\right)$$

for some value $p \neq 2$ [17]. For this distribution, the condition $\rho(\Delta x) \geq t$ takes the form $\sum\limits_{i=1}^{n} \frac{|\Delta x_i|^p}{\sigma_i^p} \leq r^p$ for some value r.

The corresponding l^p-methods have been successfully used in data processing; see, e.g., [6, 22].

It is therefore reasonable to consider such *generalized ellipsoids* as well. For a generalized ellipsoid, the Lagrange approach to maximizing a linear function $\sum\limits_{i=1}^{n} c_i\,\Delta x_i$ leads to

$$\sum_{i=1}^{n} c_i\,\Delta x_i + \lambda \sum_{i=1}^{n} \frac{|\Delta x_i|^p}{\sigma_i^p} \to \max,$$

$$c_i + \lambda p \cdot \text{sign}(\Delta x_i) \frac{|\Delta x_i|^{p-1}}{\sigma_i^p} = 0, \quad i = 1, \ldots, n,$$

and hence, for $p > 1$, to

$$\Delta x_i = \alpha \cdot \text{sign}(c_i)|c_i|^{1/(p-1)} \sigma_i^{p/(p-1)}, \quad i = 1, \ldots, n,$$

for some constant α. Here, the parameter α is determined by the condition that $\sum\limits_{i=1}^{n} \frac{|\Delta x_i|^p}{\sigma_i^p} = r^p$ – i.e., that $\alpha^p \sum\limits_{i=1}^{n} |c_i|^{p/(p-1)} \sigma_i^{p/(p-1)} = r^p$ and

$$\alpha = r / \sqrt[p]{\sum |c_i|^{p/(p-1)} \sigma_i^{p/(p-1)}}.$$

The smallest possible value $-\Delta$ of this function is attained when

$$\Delta x_i = -\alpha \cdot \text{sign}(c_i)|c_i|^{1/(p-1)} \sigma_i^{p/(p-1)}.$$

The corresponding value Δ is equal to

$$\Delta = r \left(\sum_{i=1}^{n} |c_i|^{p/(p-1)} \sigma_i^{p/(p-1)}\right)^{(p-1)/p}.$$

This value can also be computed in linear time.

Need for combining interval and generalized ellipsoid uncertainty. Similarly to the case $p = 2$, it is desirable to estimate the range of a linear function $\sum\limits_{i=1}^{n} c_i\,\Delta x_i$ over an intersection of a box and a generalized ellipsoid. In this paper, we will consider this problem for $p > 1$.

2 Analysis of the Problem: General Form of the Optimal Tuple

In the general case, we want to find the maximum and the minimum of a linear function $\sum\limits_{i=1}^{n} c_i\,\Delta x_i$ over an intersection of generalized ellipsoid and a box. In order to describe an algorithm for computing the maximum and minimum, let us first describe the general properties of the tuples Δx for which these maximum and minimum are attained.

Definition 1 *By a* generalized ellipsoid E*, we mean a set of all the tuples* $\Delta x = (\Delta x_1, \ldots, \Delta x_n)$ *which satisfy the inequality* $\sum\limits_{i=1}^{n} \dfrac{|\Delta x_i|^p}{\sigma_i^p} \le r^p$*, where* p*,* r*, and* σ_i *are positive real numbers.*

We want to find the maximum and the minimum of a linear function on the intersection $I = E \cap B$ of a generalized ellipsoid and a box

$$B = [-\Delta_1, \Delta_1] \times \ldots \times [-\Delta_n, \Delta_n].$$

Without losing generality, we can assume that all the coefficients c_i, $i = 1, \ldots, n$, of a linear function are non-negative. Indeed, if $c_i < 0$ for some i, then we can simply replace the original variable Δx_i with a new variable $\Delta x'_i = -\Delta x_i$. After this replacement, the expressions for the ellipsoid E and for the box B remain the same, but the corresponding coefficient c_i becomes positive.

Under this assumption, one can easily see that the maximum of a linear function $\sum\limits_{i=1}^{n} c_i\,\Delta x_i$ with $c_i \ge 0$ is attained when $\Delta x_i \ge 0$ for all i. We then get the following result.

Proposition 1 *The maximum of a linear function* $\sum\limits_{i=1}^{n} c_i\,\Delta x_i$ *with* $c_i \ge 0$ *over the intersection* $E \cap B$ *of a box* $B = [-\Delta_1, \Delta_1] \times \ldots \times [-\Delta_n, \Delta_n]$ *and a generalized ellipsoid* $E = \left\{\Delta x : \sum\limits_{i=1}^{n} \dfrac{|\Delta x_i|^p}{\sigma_i^p} \le r^p\right\}$ *is attained, for a certain value* α*, at a tuple*

$$\Delta x_i = \min(\Delta_i, \alpha\, c_i^{1/(p-1)} \sigma_i^{p/(p-1)}), \quad i = 1, \ldots, n.$$

Observation. This expression has an interesting relation to the corresponding expressions for the box and for the generalized ellipsoid. Indeed, let us recall that for the box B, the maximum is attained for $\Delta x_i = \Delta_i$, $i = 1, \ldots, n$. For the generalized ellipsoid E, the maximum is attained when for a certain value α_E, we have $\Delta x_i = \alpha_E\, c_i^{1/(p-1)} \sigma_i^{p/(p-1)}$, $i = 1, \ldots, n$. According to Proposition 1, the optimizing tuple for the intersection $E \cap B$ is a component-wise minimum of the two tuples:

- the tuple with components Δ_i, $i = 1, \ldots, n$, which maximizes the linear function on the box B, and
- the tuple with components $\alpha\, c_i^{1/(p-1)} \sigma_i^{p/(p-1)}$, $i = 1, \ldots, n$, which is similar to the tuple that maximizes the linear function on the generalized ellipsoid E.

It should be mentioned that the second tuple is not exactly the one that maximizes the linear function over E, since, in general, the value α (corresponding to the maximum over the intersection $E \cap B$) is different from the value α_E corresponding to the maximum over E.

Comment. For general (not necessarily non-negative) coefficients c_i, we get

$$\Delta x_i = \text{sign}(c_i) \cdot \min(\Delta_i, \alpha\, |c_i|^{1/(p-1)} \sigma_i^{p/(p-1)}), \quad i = 1, \ldots, n.$$

Proof. Let $\Delta x = (\Delta x_1, \ldots, \Delta x_n)$ be an optimal (maximizing) tuple.

If there are indices i and j for which $1 \le i, j \le n$, $\Delta x_i < \Delta_i$ and $\Delta x_j < \Delta_j$, then, for sufficiently small real numbers ε_i and ε_j, we can replace Δx_i with $\Delta x_i + \varepsilon_i$, Δx_j with $\Delta x_j + \varepsilon_j$, and still stay within the intervals $[0, \Delta_i]$ and $[0, \Delta_j]$ – i.e., within the box B. Let us select the changes ε_i and ε_j in such a way that the sum $s := \dfrac{|\Delta x_i|^p}{\sigma_i^p} + \dfrac{|\Delta x_j|^p}{\sigma_j^p}$ remain unchanged – then we will stay within the generalized ellipsoid as well.

For small ε_i and ε_j, we have

$$\frac{(\Delta x_i + \varepsilon_i)^p}{\sigma_i^p} + \frac{(\Delta x_j + \varepsilon_j)^p}{\sigma_i^p} =$$

$$\frac{(\Delta x_i)^p}{\sigma_i^p} + \frac{(\Delta x_j)^p}{\sigma_i^p} + \frac{p\, \varepsilon_i\, \Delta x_i^{p-1}}{\sigma_i^p} + \frac{p\, \varepsilon_j\, \Delta x_j^{p-1}}{\sigma_j^p} + o(\varepsilon_i).$$

Thus, to make sure that s does not change, we must select ε_j for which

$$\frac{\varepsilon_i\, \Delta x_i^{p-1}}{\sigma_i^p} + \frac{\varepsilon_j\, \Delta x_j^{p-1}}{\sigma_j^p} = o(\varepsilon_i),$$

i.e.,

$$\varepsilon_j = -\varepsilon_i\, \frac{\Delta x_i^{p-1}}{\Delta x_j^{p-1}}\, \frac{\sigma_j^p}{\sigma_i^p} + o(\varepsilon_i).$$

The resulting change in the maximized linear function is equal to $c_i \varepsilon_i + c_j \varepsilon_j$. Substituting the expression for ε_j in terms of ε_i, we conclude that this change is equal to

$$\varepsilon_i \left(c_i - c_j\, \frac{\Delta x_i^{p-1}}{\Delta x_j^{p-1}}\, \frac{\sigma_j^p}{\sigma_i^p} \right) + o(\varepsilon_i).$$

If the coefficient at ε_i was positive, then we could take a small positive ε_i and further increase the value of the linear function – which contradicts our selection of the tuple Δx_i for which the maximum is attained. Similar, if the coefficient at ε_i was negative, then we could take a small negative ε_i and further increase the value of the linear function. Thus, this coefficient cannot be positive and cannot be negative – hence it must be equal to 0. So,

$$c_i - c_j \frac{\Delta x_i^{p-1}}{\Delta x_j^{p-1}} \frac{\sigma_j^p}{\sigma_i^p} = 0,$$

or, equivalently,

$$\frac{\Delta x_i^{p-1}}{c_i \sigma_i^p} = \frac{\Delta x_j^{p-1}}{c_j \sigma_j^p}.$$

This equality holds for every two indices i and j for which $1 \le i, j \le n$, $\Delta x_i < \Delta_i$, and $\Delta x_j < \Delta_j$. Thus, for all the indices $i = 1, \ldots, n$ for which $\Delta x_i < \Delta_i$, the above ratio has the same value. Let us denote this common ratio by r_0; then, we conclude that for all such indices i, we have $\dfrac{\Delta x_i^{p-1}}{c_i \sigma_i^p} = r_0$ and hence, that

$$\Delta x_i = \alpha\, c_i^{1/(p-1)} \sigma_i^{p/(p-1)},$$

where we denoted $\alpha := r_0^{1/(p-1)}$.

If $\Delta x_i < \Delta_i$ and $\Delta x_j = \Delta_j$, then we can similarly change Δx_i and Δx_j, but only the changes for which $\varepsilon_j < 0$ will keep us inside the box. Since the sign of ε_j is opposite to the sign of ε_i, we thus conclude that we can only take $\varepsilon_i > 0$. Thus, the coefficient at ε_i in the expression for the change in the (linear) objective function cannot be positive – because then, we would be able to further increase this objective function. So, this coefficient must be non-positive, i.e.,

$$c_i - c_j \frac{\Delta x_i^{p-1}}{\Delta x_j^{p-1}} \frac{\sigma_j^p}{\sigma_i^p} \le 0,$$

or, equivalently,

$$\frac{\Delta x_i^{p-1}}{c_i \sigma_i^p} \le \frac{\Delta x_j^{p-1}}{c_j \sigma_j^p}.$$

Since $\Delta x_i < \Delta_i$ for the index i, we have $\dfrac{\Delta x_i^{p-1}}{c_i \sigma_i^p} = r_0$. Thus, we conclude that $\dfrac{\Delta x_j^{p-1}}{c_j \sigma_j^p} \le r_0$, i.e., $\Delta x_j = \Delta_j \le \alpha\, c_j^{1/(p-1)} \sigma_j^{p/(p-1)}$.

Hence,

- when $\Delta x_i < \Delta_i$, we get $\Delta x_i = \alpha\, c_i^{1/(p-1)} \sigma_i^{p/(p-1)}$;

- when $\Delta x_j = \Delta_i$, we get $\Delta x_j = \Delta_j \le \alpha\, c_j^{1/(p-1)}\sigma_j^{p/(p-1)}$.

To complete the proof of our proposition, let us consider two cases.

If $\Delta_i \le \alpha\, c_i^{1/(p-1)}\sigma_i^{p/(p-1)}$, then we cannot have $\Delta x_i < \Delta_i$ – because then we would have $\Delta x_i = \alpha\, c_i^{1/(p-1)}\sigma_i^{p/(p-1)}$ and thus, $\Delta_i > \Delta x_i = \alpha\, c_i^{1/(p-1)}\sigma_i^{p/(p-1)}$ and $\Delta_i > \alpha\, c_i^{1/(p-1)}\sigma_i^{p/(p-1)}$ – which contradicts our assumption. Thus, the only remaining case here is $\Delta x_i = \Delta_i$.

On the other hand, if $\Delta_j > \alpha\, c_j^{1/(p-1)}\sigma_j^{p/(p-1)}$, then we cannot have $\Delta x_j = \Delta_j$ – because otherwise, we would have $\Delta_j \le \alpha\, c_j^{1/(p-1)}\sigma_j^{p/(p-1)}$, which also contradicts our assumption. Thus, in this case, we must have $\Delta x_j < \Delta_j$, and we already know that in this case, $\Delta x_j = \alpha\, c_j^{1/(p-1)}\sigma_i^{p/(p-1)}$. So:

- if $\Delta_i \le \alpha\, c_i^{1/(p-1)}\sigma_i^{p/(p-1)}$ then $\Delta x_i = \Delta_i$;
- if $\Delta_j > \alpha\, c_j^{1/(p-1)}\sigma_j^{p/(p-1)}$ then $\Delta x_j = \alpha\, c_j^{1/(p-1)}\sigma_i^{p/(p-1)}$.

In both cases, we have

$$\Delta x_i = \min(\Delta_i, \alpha\, c_i^{1/(p-1)}\sigma_i^{p/(p-1)}), \quad i = 1, \ldots, n.$$

The proposition is proven.

3 Analysis of the Problem: How to Find α

According to our result, once we know the value of the parameter α, we will be able to find all the values Δx_i, $i = 1, \ldots, n$, from the optimal tuple, and thus, find the largest possible value Δ of the desired linear function $\sum\limits_{i=1}^{n} c_i\, \Delta x_i$.

For each $i = 1, \ldots, n$, writing $z_i := \dfrac{\Delta_i}{|c_i|^{1/(p-1)}\sigma_i^{p/(p-1)}}$, the dependence of $|\Delta x_i|$ on α can be described as follows:

- If $\alpha\,|c_i|^{1/(p-1)}\sigma_i^{p/(p-1)} < \Delta_i$, i.e., if $\alpha < z_i$, then we take $|\Delta x_i| = \alpha\,|c_i|^{1/(p-1)}\sigma_i^{p/(p-1)}$.
- On the other hand, if $\alpha\,|c_i|^{1/(p-1)}\sigma_i^{p/(p-1)} \ge \Delta_i$, i.e., if $\alpha \ge z_i$, then we take $|\Delta x_i| = \Delta_i$.

So, if we sort the indices by the value z_i, into a sequence $z_1 \le z_2 \ldots \le z_n$, then the maximizing tuple have the form

$$\Delta x = (\mathrm{sign}(c_1)\cdot\Delta_1, \ldots, \mathrm{sign}(c_t)\cdot\Delta_t,$$

$$\alpha\,\mathrm{sign}(c_{t+1})\cdot|c_{t+1}|^{1/(p-1)}\sigma_{t+1}^{p/(p-1)}, \ldots, \alpha\,\mathrm{sign}(c_n)\cdot|c_n|^{1/(p-1)}\sigma_n^{p/(p-1)})$$

for some threshold value t for which $z_t \le \alpha < z_{t+1}$.

How do we find this threshold value t? In principle, it is possible that the optimal solution is attained when $\Delta x_i = \pm\Delta_i$ for all i. In this case, the generalized ellipsoid contains the whole box. In all other cases, the value α must be determined by the condition that the optimal tuple is on the surface of the generalized ellipsoid, i.e., that

$$\sum_{i=1}^{t} \frac{\Delta_i^p}{\sigma_i^p} + \alpha^p \sum_{j=t+1}^{n} |c_i|^{p/(p-1)} \sigma_j^{p/(p-1)} = r^p,$$

or, equivalently,

$$\sum_{i=1}^{n} \frac{(\min(\Delta_i, \alpha\,|c_i|^{1/(p-1)} \sigma_i^{p/(p-1)}))^p}{\sigma_i^p} = r^p.$$

The left-hand side of this equality is an increasing function of α. Thus, to find the proper value of t, it is sufficient to check all the values $\alpha = z_1, \ldots, z_n$.

If for some $k = 1, \ldots, n$, we get

$$\sum_{i=1}^{k} \frac{\Delta_i^p}{\sigma_i^p} + z_k^p \sum_{j=k+1}^{n} |c_j|^{p/(p-1)} \sigma_j^{p/(p-1)} > r^p,$$

this means that we need to decrease α, i.e., that we should have fewer values $\Delta x_i = \pm\Delta_i$ – in other words, this means that $t < k$.

On the other hand, if for some $k = 1, \ldots, n$, we get

$$\sum_{i=1}^{k} \frac{\Delta_i^p}{\sigma_i^p} + z_k^p \sum_{j=k+1}^{n} |c_j|^{p/(p-1)} \sigma_i^{p/(p-1)} \le r^p,$$

this means that $t \ge k$.

So, we can find the desired threshold t as the largest index k for which for $\alpha = z_k$, the left-hand side of the above equality is still less than or equal to r^p; due to monotonicity with respect to α, this value t can be found by bisection.

Once we find this threshold value t, we can then find α from the equation

$$\sum_{i=1}^{t} \frac{\Delta_i^p}{\sigma_i^p} + \alpha^p \sum_{j=t+1}^{n} |c_j|^{p/(p-1)} \sigma_j^{p/(p-1)} = r^p,$$

i.e., $\alpha^p = \dfrac{r^p - E^-}{E^+}$, where $E^- := \sum\limits_{i=1}^{t} \dfrac{\Delta_i^p}{\sigma_i^p}$ and $E^+ := \sum\limits_{j=t+1}^{n} |c_j|^{p/(p-1)} \sigma_j^{p/(p-1)}$. After that, we can uniquely determine the optimal tuple $\Delta x = (\Delta x_1, \ldots, \Delta x_n)$ and thus the desired maximal value $\Delta = \sum\limits_{i=1}^{k} |c_i| \cdot \Delta_i + \alpha \sum\limits_{j=t+1}^{n} |c_j|^{p/(p-1)} \sigma_j^{p/(p-1)}$.

So, we arrive at the following algorithms for computing Δ.

4 A Simpler $O(n\,\log(n))$ Algorithm

Algorithm. First, we check whether the generalized ellipsoid contains the box, i.e., whether $\sum\limits_{i=1}^{n} \dfrac{\Delta_i^p}{\sigma_i^p} \le r^p$. If this is the case, then the desired maximum is equal to $\sum\limits_{i=1}^{n} |c_i|\,\Delta_i$. If this is not the case, then we apply our algorithm.

In this algorithm, we first sort the indices $i = 1, \ldots, n$ in the increasing order by the value of z_i.

After this sorting, we apply the following iterative algorithm. At each iteration of this algorithm, we have two numbers:

- the number i^- such that for all indices $i \le i^-$, we already know that for the optimal tuple Δx, we have $|\Delta x_i| = \Delta_i$;
- the number i^+ of all the indices $j \ge i^+$ for which we already know that for the optimal tuple Δx, we have $|\Delta x_j| < \Delta_j$.

In the beginning, $i^- = 0$ and $i^+ = n + 1$. At each iteration, we also update the value of two auxiliary quantities $E^- := \sum\limits_{i=1}^{i^-} \dfrac{\Delta_i^p}{\sigma_i^p}$ and $E^+ := \sum\limits_{j=i^+}^{n} |c_j|^{p/(p-1)} \sigma_j^{p/(p-1)}$.

In principle, on each iteration, we could compute these sums "from scratch"; however, to speed up computations, on each iteration, we update these auxiliary values in a way that is faster than re-computing the corresponding sums.

Initially, since $i^- = 0$ and $i^+ = n + 1$, we take $E^- = E^+ = 0$.

At each iteration, we do the following:

- first, we compute the midpoint $m = (i^- + i^+)/2$;
- we compute $e^- := \sum\limits_{i=i^-+1}^{m} \dfrac{\Delta_i^p}{\sigma_i^p}$ and $e^+ := \sum\limits_{j=m+1}^{i^+-1} |c_j|^{p/(p-1)} \sigma_j^{p/(p-1)}$;
- if $E^- + e^- + z_m^p\,(E^+ + e^+) > r^p$, then we replace i^+ with $m + 1$ and E^+ with $E^+ + e^+$;
- if $E^- + e^- + z_m^p\,(E^+ + e^+) \le r^p$, then we replace i^- with m and E^- with $E^- + e^-$.

At each iteration, the set of undecided indices is divided in half. Iterations continue until all indices are decided, after which we compute α from the condition that $E^- + \alpha^p E^+ = r^p$, i.e., as $\alpha^p := \dfrac{r^p - E^-}{E^+}$. Once we know α, we compute the maximizing tuple $|\Delta x_i| = \min(\Delta_i, \alpha\,|c_i|^{1/(p-1)} \sigma_i^{p/(p-1)})$ and then, the desired maximum $\sum\limits_{i=1}^{n} |c_i|\,|\Delta x_i|$.

Computational complexity of the above algorithm. Sorting requires time $O(n\ \log(n))$; see, e.g., [5].

After this, at each iteration, all the operations with indices from i^- to i^+ require time T linear in the number of such indices: $T \le C \cdot (i^+ - i^-)$ for some C. We start with the set of indices of full size n; on the next iteration, we have a set of size $n/2$, then $n/4$, etc. Thus, after sorting, the overall computation time is $\le C \cdot (n + n/2 + n/4 + \ldots) \le C \cdot 2n$, i.e., linear in n. So, the overall computation time is indeed $O(n\ \log(n)) + O(n) = O(n\ \log(n))$.

Comment. This algorithm works for an even more general case, when there exist a function $\rho_0(x)$ for which for every $i = 1, \ldots, n$, the probability density function $\rho_i(\Delta x_i)$ of the i-th measurement error has the form $\rho_i(\Delta x_i) = \rho_0\left(\dfrac{|\Delta x_i|}{\sigma_i}\right)$ for some σ_i, and the measurement errors are independent, i.e., $\rho(\Delta x) = \rho_1(\Delta x_1) \ldots \rho_n(\Delta x_n)$. In this case, similar arguments lead to a generalized ellipsoid of the type $\sum\limits_{i=1}^{n} \psi\left(\dfrac{|\Delta x_i|}{\sigma_i}\right) \le r_0$, where $\psi(x) := -\ln(\rho_0(x))$. The above algorithm can be extended to the case of strictly convex smooth functions $\psi(x)$ for which both this function, its derivative, and the corresponding inverse functions can be computed in polynomial time. This class includes the l^p-functions $\psi(x) = |x|^p$ with $p > 1$ as particular cases.

5 Linear-Time Algorithm

Main idea behind the linear time algorithm. Our second algorithm is similar to the above $O(n\ \log(n))$ algorithm. In that algorithm, the only non-linear-time part was sorting. To avoid sorting, in the second algorithm, we use the known fact that we can compute the median of a set of n elements in linear time (see, e.g., [5]). (Our use of median is similar to the one from [2, 10].)

Our linear time algorithm is only efficient to large n. It is worth mentioning that while asymptotically, the linear time algorithm for computing the median is faster than sorting, this median computing algorithm is still rather complex – so, for small n, sorting is faster than computing the median.

This is the reason why in this paper, we present two different algorithms – both algorithms are practically useful:

- for large n, the linear time algorithm is faster;
- however, for small n, the $O(n\ \log(n))$ algorithm is faster.

Let us now describe the linear time algorithm.

Algorithm. First, we check whether the generalized ellipsoid contains the box, i.e., whether $\sum\limits_{i=1}^{n} \dfrac{\Delta_i^p}{\sigma_i^p} \le r^p$. If this is the case, then the desired maximum is equal to $\sum\limits_{i=1}^{n} c_i\,\Delta_i$. If this is not the case, then we perform the following iterations.

At each iteration, we have three sets:

- the set I^- of all the indices i from 1 to n for which we already know that for the optimal tuple Δx, we have $|\Delta x_i| = \Delta_i$;
- the set I^+ of all the indices j from 1 to n for which we already know that for the optimal tuple Δx, we have $|\Delta x_j| < \Delta_j$;
- the set $I = \{1, \ldots, n\} - I^- - I^+$ of the indices i for which we are still undecided.

In the beginning, $I^- = I^+ = \emptyset$ and $I = \{1, \ldots, n\}$. At each iteration, we also update the value of two auxiliary quantities $E^- := \sum\limits_{i \in I^-} \dfrac{\Delta_i^p}{\sigma_i^p}$ and $E^+ := \sum\limits_{j \in I^+} |c_j|^{p/(p-1)} \sigma_j^{p/(p-1)}$.

In principle, we could compute this value by computing this sum of squares, but to speed up computations, on each iteration, we update this auxiliary value in a way that is faster than re-computing the corresponding sum.

Initially, since $I^- = I^+ = \emptyset$, we take $E^- = E^+ = 0$.

At each iteration, we do the following:

- first, we compute the median m of the set I (median in terms of sorting by z_i);
- then, by analyzing the elements of the undecided set I one by one, we divide them into two subsets $P^- = \{i : z_i \le z_m\}$ and $P^+ = \{j : z_j > z_m\}$;
- we compute $e^- = \sum\limits_{i \in P^-} \dfrac{\Delta_i^p}{\sigma_i^p}$ and $e^+ := \sum\limits_{j \in P^+} |c_j|^{p/(p-1)} \sigma_j^{p/(p-1)}$;
- if $E^- + e^- + z_m^p\,(E^+ + e^+) > r^p$, then we replace I^+ with $I^+ \cup P^+$, I with P^-, and E^+ with $E^+ + e^+$;
- if $E^- + e^- + z_m^p\,(E^+ + e^+) \le r^p$, then we replace I^- with $I^- \cup P^-$, I with P^+, and E^- with $E^- + e^-$.

At each iteration, the set of undecided indices is divided in half. Iterations continue until all indices are decided, after which we compute α from the condition that $E^- + \alpha^p E^+ = r^p$, i.e., as $\alpha^p := \dfrac{r^p - E^-}{E^+}$. Once we know α, we compute the maximizing tuple $|\Delta x_i| = \min(\Delta_i, \alpha\,|c_i|^{1/(p-1)}\sigma_i^{p/(p-1)})$, $i = 1, \ldots, n$, and then, the desired maximum $\sum\limits_{i=1}^{n} |c_i|\,|\Delta x_i|$.

Computational complexity of the above algorithm. Let us show that this algorithm indeed requires linear time. Indeed, at each iteration, computing median requires linear time, and all other operations with I require time T linear in the number of elements $|I|$ of I: $T \le C \cdot |I|$ for some C. We start with the set I of size n; on the next iteration, we have a set of size $n/2$, then $n/4$, etc. Thus, the overall computation time is $\le C \cdot (n + n/2 + n/4 + \ldots) \le C \cdot 2n$, i.e., linear in n.

Acknowledgments

This work was supported in part by NSF grants EAR-0225670 and DMS-0532645 and by Texas Department of Transportation grant No. 0-5453. The authors are thankful to the participants of the Second International Workshop on Reliable Engineering Computing (Savannah, Georgia, February 22–24, 2006) and the International Conference on Finite Element Methods in Engineering and Science FEMTEC'2006 (El Paso, Texas, December 11–15, 2006) for valuable discussions.

We are also thankful to the anonymous referees for important editorial suggestions.

References

[1] G. Belforte and B. Bona, "An improved parameter identification algorithm for signal with unknown-but-bounded errors", *Proceeding of the 7th IFAC Symposium on Identification and Parameter Estimation*, York, U.K., 1985.

[2] P. van der Broek and J. Noppen, "Fuzzy weighted average: alternative approach", *Proceedings of the 25th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2006*, Montreal, Quebec, Canada, June 3–6, 2006.

[3] F. L. Chernousko, *Estimation of the Phase Space of Dynamic Systems*, Nauka publ., Moscow, 1988 (in Russian).

[4] F. L. Chernousko, *State Estimation for Dynamic Systems*, CRC Press, Boca Raton, FL, 1994.

[5] Th. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 2001.

[6] D. I. Doser, K. D. Crain, M. R. Baker, V. Kreinovich, and M. C. Gerstenberger "Estimating uncertainties for geophysical tomography", *Reliable Computing*, 1998, Vol. 4, No. 3, pp. 241–268.

[7] A. F. Filippov, "Ellipsoidal estimates for a solution of a system of differential equations", *Interval Computations*, 1992, Vol. 2, No. 2(4), pp. 6–17.

[8] A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry, error estimation, and other applications of set-valued analysis", *ACM SIGNUM Newsletter*, 1996, Vol. 31, No. 4, pp. 3–25.

[9] E. Fogel and Y. F. Huang, "On the value of information in system identification. Bounded noise case", *Automatica*, 1982, Vol. 18, No. 2, pp. 229–238.

[10] P. Hansen, M. V. P. de Aragao, and C. C. Ribeiro, "Hyperbolic 0-1 programming and optimization in information retrieval", *Math. Programming*, 1991, Vol. 52, pp. 255–263.

[11] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. *Applied Interval Analysis.* Springer-Verlag, London, 2001.

[12] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1998.

[13] S. Li, Y. Ogura, and V. Kreinovich, *Limit Theorems and Applications of Set Valued and Fuzzy Valued Random Variables*, Kluwer Academic Publishers, Dordrecht, 2002.

[14] A. Neumaier, *Interval Methods for Systems of Equations*, Cambridge University Press, Cambridge, 1990.

[15] A. Neumaier, *Introduction to Numerical Analysis*, Cambridge Univ. Press, Cambridge, 2001.

[16] J. P. Norton, "Identification and application of bounded parameter models", *Proceeding of the 7th IFAC Symposium on Identification and Parameter Estimation*, York, U.K., 1985.

[17] P. V. Novitskii and I. A. Zograph, *Estimating the Measurement Errors*, Energoatomizdat, Leningrad, 1991 (in Russian).

[18] A. I. Orlov, "How often are the observations normal?", *Industrial Laboratory*, 1991, Vol. 57, No. 7, pp. 770–772.

[19] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer-Verlag, New York, 2005.

[20] F. C. Schweppe, “Recursive state estimation: unknown but bounded errors and system inputs”, *IEEE Transactions on Automatic Control*, 1968, Vol. 13, p. 22.

[21] F. C. Schweppe, *Uncertain Dynamic Systems*, Prentice Hall, Englewood Cliffs, NJ, 1973.

[22] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam, 1987.