Optimization with uncertain, inexact or unstable data: Linear programming and the interval approach

Michal Černý[1], Milan Hladík[2]

Abstract. This paper accompanies our invited lecture on interval methods in optimization. First, we put the interval approach in context with other approaches to modeling inexactness, instability or imprecision of input data; in particular, we discuss the stochastic and fuzzy approach. Then we turn to interval linear programming. We show some important aspects in which the interval-valued linear programs differ from traditional linear programs. We emphasize how various formulations of the auxiliary linear program, which are equivalent in the traditional setting, differ in the interval setting from the computational point of view. We also consider some natural questions, e.g. weak and strong feasibility and the problem of finding the range of possible optimal values and the catastrophic scenario. We also point out some (subjectively selected) interesting open problems in the field.

Keywords: interval optimization, interval linear programming, inexact data, interval data, open problems

JEL Classification: C02, C61, C63, C65

1 Introduction

Optimization is a well-studied branch of mathematics, which has many important applications in real life: for example, let us mention statistics & data analysis, computer science, operations research, physics and many others. To be more concrete, in computer science there are various applications in data mining, neural networks, artificial intelligence, assignment of tasks to processors, analysis of networks and many others. In operations research, optimization is a tool for solving problems in logistics and transportation problems,

[1] University of Economics, Prague, Czech Republic. Faculty of Computer Science and Statistics, Department of Econometrics. Winston Churchill Square 4, 13067 Prague, Czech Republic. E-mail: cernym@vse.cz.

[2] Charles University, Prague, Czech Republic. Faculty of Mathematics and Physics, Department of Applied Mathematics. Malostranské náměstí 25, 11000 Prague, Czech Republic. E-mail: hladik@kam.mff.cuni.cz.

production planning and scheduling, human resource management and investment & portfolio management.

We have a strong mathematical theory for various kinds of optimization problems – the most fundamental stone is linear programming, and we can continue to discrete optimization, convex nonlinear programming and many other advanced branches. However, the theory available to us usually leans on a strong assumption, which might not be fulfilled in practice: that data of the optimization problem are *known fixed constants*.

In practice it often happens that data suffer from imprecision – they are inexact, estimated, subjective, noisy, affected by errors, or subject to instability or changes. Then, natural questions come to mind: will the optimal solution be stable (or: robust), meaning that a small imprecision or change in data of the optimization problem affects the optimal solution only negligibly? Can a change in data result in a significant change of the optimal value (which often measures costs or profits)?

In this paper we will deal with linear programming, but the questions, answers, ideas and problems of investigation are applicable also in the case of other, more advanced optimization problems, encountered in practice.

There are various approaches to incorporate instability or imprecision of data into optimization problems. We should mention at least three:

- *The stochastic approach*, where data of the optimization problem are treated as random variables. Then, the optimal solution and optimal objective value are random variables. Then it makes sense to ask questions about their distribution, moments, tail probabilities etc.
- *The fuzzy approach*, where data of the optimization problem are treated as fuzzy numbers. Then the problem is known as fuzzy optimization and is extensively studied (see e.g. Ramík, 2006).
- *The interval* (or: *possibilistic*) *approach*, where data of the optimization problem are treated as closed real intervals of possible values, which the unknown/instable/imprecise variable can attain.

This text is devoted to the third approach in case of the foundation stone of optimization – Linear Programming. (The interval approach is also useful in statistics, see e.g. Černý et al., 2013.)

2 Linear Programming with interval data

Given two real matrices $\underline{A} \le \overline{A}$, where the relation "≤" is understood entry-wise, we define the *interval matrix* **A** as a family of matrices $\mathbf{A} = [\underline{A}, \overline{A}] = \{A\colon \underline{A} \le A \le \overline{A}\}$. A similar notation is used for interval vectors (which are one-column interval matrices). Interval quantities (numbers, vectors, matrices) are denoted in boldface.

2.1 Formulations of an interval linear program

An *interval linear program* with data **A**, **b**, **c** is simply a family of all linear programs $\min\{c^{\mathrm{T}}x\colon Ax \le b\}$ satisfying $A \in \mathbf{A}$, $b \in \mathbf{b}$ and $c \in \mathbf{c}$.

In traditional linear programming, it is well-known that various forms of the auxiliary linear program $\min\{c^{\mathrm{T}}x\colon Ax \le b\}$ may be transformed into other equivalent forms, such as $\min/\max\{c^{\mathrm{T}}x\colon Ax = b,\ x \ge 0\}$, or $\min/\max\{c^{\mathrm{T}}x\colon Ax \le b,\ x \ge 0\}$, or more sophisticated formats, like the self-dual embedding (see Roos et al., 1998), Karmarkar Normal Form or Khachiyan Normal Form (see Schrijver, 1988). However, in the interval case, such transformations are not always possible. The problem is that conversion of one form into another might cause serious computational problems. Of course, it depends on which question about the interval linear program we are interested in. Let us give three examples.

Question 1. *Is the interval linear programming problem strongly feasible*? (Strong feasibility means: the linear program is feasible for *all* choices $A \in \mathbf{A}$, $b \in \mathbf{b}$, $c \in \mathbf{c}$.) The following results hold true:

- Both questions "$(\forall A \in \mathbf{A})(\forall b \in \mathbf{b})(\exists x \in \mathrm{P}^n)[Ax \le b]$?" and "$(\forall A \in \mathbf{A})(\forall b \in \mathbf{b})(\exists x \in \mathrm{P}^n)[Ax \le b,\ x \ge 0]$?" are decidable in polynomial time.
- On the other hand: the question "$(\forall A \in \mathbf{A})(\forall b \in \mathbf{b})(\exists x \in \mathrm{P}^n)[Ax = b,\ x \ge 0]$?" is co-**NP**-hard.

Question 2. *Is the interval linear programming problem weakly feasible*? (Weak feasibility means: the linear program is feasible for *some* choice $A \in \mathbf{A}$, $b \in \mathbf{b}$, $c \in \mathbf{c}$.) The following results hold true:

- Both questions "$(\forall A \in \mathbf{A})(\forall b \in \mathbf{b})(\exists x \in \mathrm{P}^n)[Ax = b, x \geq 0]$?" and "$(\forall A \in \mathbf{A})(\forall b \in \mathbf{b})(\exists x \in \mathrm{P}^n)[Ax \leq b, x \geq 0]$?" are decidable in polynomial time.
- On the other hand: the question "$(\forall A \in \mathbf{A})(\forall b \in \mathbf{b})(\exists x \in \mathrm{P}^n)[Ax \leq b]$?" is **NP**-hard.

Question 3. *What is the range of optimal values?* The following results hold true. The numbers

$$\overline{f}^{Ax\leq b, x\geq 0}(\mathbf{A}, \mathbf{b}, \mathbf{c}) := \max\{\min\{c^{\mathrm{T}}x : Ax \leq b, x \geq 0\} : A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}\},$$

$$\underline{f}^{Ax\leq b, x\geq 0}(\mathbf{A}, \mathbf{b}, \mathbf{c}) := \min\{\min\{c^{\mathrm{T}}x : Ax \leq b, x \geq 0\} : A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}\},$$

$$\overline{f}^{Ax\leq b}(\mathbf{A}, \mathbf{b}, \mathbf{c}) := \max\{\min\{c^{\mathrm{T}}x : Ax \leq b\} : A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}\},$$

$$\underline{f}^{Ax=b, x\geq 0}(\mathbf{A}, \mathbf{b}, \mathbf{c}) := \min\{\min\{c^{\mathrm{T}}x : Ax = b, x \geq 0\} : A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}\}$$

are computable in polynomial time. On the other hand, the computation of the numbers

$$\overline{f}^{Ax=b, x\geq 0}(\mathbf{A}, \mathbf{b}, \mathbf{c}) := \max\{\min\{c^{\mathrm{T}}x : Ax = b, x \geq 0\} : A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}\},$$

$$\underline{f}^{Ax\leq b}(\mathbf{A}, \mathbf{b}, \mathbf{c}) := \min\{\min\{c^{\mathrm{T}}x : Ax \leq b\} : A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}\}$$

is an **NP**-hard problem.

2.2 The range of optimal values in detail and the catastrophic scenario

The range of optimal values $[\underline{f}, \overline{f}]$, discussed in Question 3 of Section 2.1, is an extremely important notion in interval linear programming. (To be precise, the range need not be a connected set; however, under quite general assumptions it is (Mostafaee et al., 2013), and we will tacitly assume that this is our case.) When we interpret the objective function as a cost function, the value $\overline{f}$ tells us the highest possible costs, i.e. the worst-case outcome of the problem modeled by the linear program. Therefore, the value $\overline{f}$ can be informally called as the *catastrophic scenario*. And the value $\underline{f}$ shows the best case, i.e. the lowest possible costs that can be achieved by a choice $A \in \mathbf{A}$, $b \in \mathbf{b}$, $c \in \mathbf{c}$.

Remark. The results of Question 3 of Section 2.1 show us that the range of optimal values $[\underline{f}, \overline{f}]$ can be computed efficiently only when the auxiliary linear program takes the form $\min\{c^{\mathrm{T}}x\colon Ax \le b, x \ge 0\}$. Though this form is quite general, observe that it is not easy to introduce e.g. an equality. The traditional trick of replacing the equality $d^{\mathrm{T}}x = e$ by two inequalities $d^{\mathrm{T}}x \le e$ and $-d^{\mathrm{T}}x \le -e$ does not work in general. Indeed, it can happen that given $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$, $\mathbf{e}$, we have

$$\begin{aligned}
&\max\{\min\{c^{\mathrm{T}}x\colon Ax \le b,\ d^{\mathrm{T}}x = e,\ x \ge 0\}\colon A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}, d \in \mathbf{d}, e \in \mathbf{e}\} \\
&\quad = \max\{\min\{c^{\mathrm{T}}x\colon Ax \le b,\ d^{\mathrm{T}}x \le e,\ -d^{\mathrm{T}}x \le -e,\ x \ge 0\}\colon A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}, \\
&\qquad\qquad d \in \mathbf{d}, e \in \mathbf{e}\} \\
&\quad \ne \max\{\min\{c^{\mathrm{T}}x\colon Ax \le b,\ d^{\mathrm{T}}x \le e,\ -(d')^{\mathrm{T}}x \le -e',\ x \ge 0\}\colon A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}, \\
&\qquad\qquad d \in \mathbf{d}, d' \in \mathbf{d}, e \in \mathbf{e}, e' \in \mathbf{e}\} \\
&\quad = \max\{\min\{c^{\mathrm{T}}x\colon A'x \le b',\ x \ge 0\}\colon A' \in \mathbf{A}', b' \in \mathbf{b}', c \in \mathbf{c}\}, \qquad (1)
\end{aligned}$$

where $\mathbf{A}' = \begin{pmatrix} \mathbf{A} \\ \mathbf{d} \\ -\mathbf{d} \end{pmatrix}$ and $\mathbf{b}' = \begin{pmatrix} \mathbf{b} \\ \mathbf{e} \\ -\mathbf{e} \end{pmatrix}$. Notwithstanding, the catastrophic scenario can be computed in polynomial time when the form is $Ax \le b$ (though here is the same problem with introduction of equalities). Unfortunately, the catastrophic scenario with the formulation $Ax = b$, $x \ge 0$ cannot be computed efficiently in general.

Remark. For special linear programs, the situation need not be so disappointing. For example, consider the problem of computation of the value of a matrix game in mixed strategies, when the payoff matrix $\mathbf{A}$ is interval. The catastrophic scenario is the value $f^* := \min\{\max\{\gamma\colon Ax \ge \gamma 1,\ 1^{\mathrm{T}}x = 1,\ x \ge 0\}\colon A \in \mathbf{A}\}$, where 1 denotes the all-one vector. An easy observation is that a positive perturbation of the payoff matrix cannot decrease the value of the game. It follows that $f^* = \max\{\gamma\colon \underline{A}x \ge \gamma 1,\ 1^{\mathrm{T}}x = 1,\ x \ge 0\}$, and the last expression is an ordinary linear program.

2.3 Extreme scenarios

Of course, we are interested not only in the interval $[\underline{f}, \overline{f}]$, but we also want to know which choice of data $A \in \mathbf{A}$, $b \in \mathbf{b}$, $c \in \mathbf{c}$ of the interval linear program will lead to the catastrophic scenario and/or to the most optimistic scenario. We restrict ourselves to the formats $Ax \le b$, $x \ge 0$ and $Ax = b$, $x \ge 0$.

Vajda's Theorem. We have $\underline{f}^{Ax\le b, x\ge 0}(\mathbf{A}, \mathbf{b}, \mathbf{c}) = \min\{\underline{c}^{\mathrm{T}}x : \underline{A}x \le \overline{b}, x \ge 0\}$ and $\overline{f}^{Ax\le b, x\ge 0}(\mathbf{A}, \mathbf{b}, \mathbf{c}) = \min\{\overline{c}^{\mathrm{T}}x : \overline{A}x \le \underline{b}, x \ge 0\}$.

Rohn's Theorem. We have $\underline{f}^{Ax=b, x\ge 0}(\mathbf{A}, \mathbf{b}, \mathbf{c}) = \min\{\underline{c}^{\mathrm{T}}x : \underline{A}x \le \overline{b}, \overline{A}x \ge \underline{b}, x \ge 0\}$ and

$$\overline{f}^{Ax=b, x\ge 0}(\mathbf{A}, \mathbf{b}, \mathbf{c})$$

$$= \max_{s\in\{0,1\}^n} \min\{\overline{c}^{\mathrm{T}}x :\ (A^C - \mathrm{diag}(s)A^\Delta)x = (b^C + \mathrm{diag}(s)b^\Delta),\ x \ge 0\}, \quad (2)$$

where $A^C := \frac{1}{2}(\overline{A} + \underline{A})$ is the *center* of $\mathbf{A}$, $A^\Delta := \frac{1}{2}(\overline{A} - \underline{A})$ is the *radius* of $\mathbf{A}$ and diag$(s_1, ..., s_n)$ is the diagonal matrix with diagonal elements $s_1, ..., s_n$.

Vajda's Theorem tells that the catastrophic scenario is attained with the choice $\overline{A} \in \mathbf{A}$, $\underline{b} \in \mathbf{b}$, $\overline{c} \in \mathbf{c}$. Observe that it does not depend on $\underline{A}, \overline{b}, \underline{c}$.

Rohn's Theorem suggests an exponential-time method for determining the catastrophic scenario – in general we must solve 2^n linear programs, one for each choice of s (i.e., for each orthant of P^n). This is not surprising since from Question 3 of Section 2.1 we know that the problem is **NP**-hard. When s is the maximizer of (1), then the catastrophic scenario is attained for the choice $A^C - diag(s)A^\Delta \in \mathbf{A}$, $b^C + diag(s)b^\Delta \in \mathbf{b}$, $\overline{c} \in \mathbf{c}$.

3 Further problems

The results of Sections 2.1 – 2.3 give only an essence of problems appearing in interval linear programming. We could see that many usual tricks, applicable in traditional linear programming, cannot be applied in the interval setting: for example, we cannot simply rewrite an equation as two inequalities. We have also illustrated some natural questions applicable in interval linear programming: how to compute the range of optimal values? For which choices of $A \in \mathbf{A}$, $b \in \mathbf{b}$, $c \in \mathbf{c}$ are they attained? There are many more interesting questions; the answers to some of them are known, some of them are still subject to further research. Here we list some of them, which are particularly interesting in our opinion.

Problem 1. Given (A, b, c), let $F(A, b, c)$ denote the feasible polyhedron of the linear program with data A, b, c (in either of the three formats mentioned above). Given interval data $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c})$, define the *weakly feasible set* and *strongly feasible set*, respectively, as

$$F^* = \bigcup_{A\in\boldsymbol{A},\, b\in\boldsymbol{b},\, c\in\boldsymbol{c}} F(A,b,c), \quad F^{**} = \bigcap_{A\in\boldsymbol{A},\, b\in\boldsymbol{b},\, c\in\boldsymbol{c}} F(A,b,c).$$

The problem is to describe or approximate the sets F^* and F^{**}. Such an approximation may take, for example, the form of a tight interval enclosure. A *tight interval enclosure* is an interval vector $\boldsymbol{z} \supseteq F^*$ such that no interval vector $\boldsymbol{z}' \subseteq \boldsymbol{z}$ with $\boldsymbol{z}' \neq \boldsymbol{z}$ satisfies $\boldsymbol{z}' \supseteq F^*$. (A similar task is interesting for F^{**}.) The approximation may also take the form of a tight inscribed interval vector, tight circumscribed ellipsoid etc.

Problem 2. An interval linear programming problem with data $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$ is called *weakly unbounded* if the linear program with data A, b, c it is unbounded for some choice $A \in \mathbf{A}$, $b \in \mathbf{b}$, $c \in \mathbf{c}$. Prove that testing weak unboundedness is either (co-)**NP**-hard, or solvable in polynomial time. Positive results: the problem is polynomial-time solvable for the format $Ax \le b$, $x \ge 0$.

Problem 3. Develop a theory for the case with dependencies. *Dependency* is a problem occurring when a data point has more occurrences in the formulation of the auxiliary linear program. For example, the interval linear program

$$\min\{c^{\mathrm{T}}(x-y)\colon Ax - Ay \le b,\, x \ge 0,\, y \ge 0\} \ \text{ with } \ A \in \mathbf{A},\, b \in \mathbf{b},\, c \in \mathbf{c}$$

suffers from dependencies since the data matrix A occurs in its formulation twice. Dependencies can be defined more generally as additional (linear) restrictions among the data of the linear program, such as

$$\min\{c^{\mathrm{T}}(x-y)\colon Ax - A'y \le b,\, x \ge 0,\, y \ge 0\}$$
$$\text{with } \ A \in \mathbf{A},\, A' \in \mathbf{A},\, b \in \mathbf{b},\, c \in \mathbf{c} \ \text{ s.t. } \underbrace{A = A'}_{\text{restrictions}}.$$

Recall that we have already encountered dependencies in (1).

Problem 4. *Inverse interval linear programming*. Let the interval data **A**, **b**, **c** be given, together with a value $f_0 \in [\underline{f}, \overline{f}]$. Find $A_0 \in$ **A**, $b_0 \in$ **b**, $c_0 \in$ **c** such that the optimal value of the linear program with data A_0, b_0, c_0 is f_0. If there are more solutions, suggest suitable criteria for classification which of the solutions is "better" than another one and propose a method for finding the "better" ones.

Remark to Problem 4. Problem 4 is interesting when the data of a linear program can be interpreted as controlling variables. For example, consider the problem of network flows – the inverse interval linear programming problem asks how to design capacities of edges of the network to achieve a given maximal flow, when the capacities can be selected from given intervals.

Problem 5. *Weak basis optimality*. Let the interval data **A**, **b**, **c** be given, together with a basis B. Decide whether the basis is optimal for some choice $A \in$ **A**, $b \in$ **b**, $c \in$ **c**. Either find a polynomial algorithm, or prove (co)**NP**-hardness.

Acknowledgments

Michal Černý was supported by the Czech Science Foundation Grants P403/12/1947 and P402/12/G097. Milan Hladík was supported by the Czech Science Foundation Grants P402-13-10660S and P403/12/1947. We are also obliged to the project F4/11/2013 of University of Economics, Prague, in pursuance of which the invitation to SMSIS originated.

References

Beeck, H. (1978) *Linear Programming with Interval Data*. Technical Report TUM-ISU-7830, Technical University of Munich, 1978.

Černý, M., Antoch, J. and Hladík, M. (2013) 'On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data'. *Information Sciences*, to appear. http://dx.doi.org/10.1016/j.ins.2013.04.035.

Chinneck, J. W. and Ramadan, K. (2000) 'Linear programming with interval coefficients'. *Operations Research Letters*, vol. 52, no. 2, pp. 209–220.

Hansen, E. R. (1992) *Global Optimization Using Interval Analysis*. New York: Marcel Dekker.

Hladík, M. (2009) 'Optimal value range in interval linear programming'. *Fuzzy Optimization and Decision Making*, vol. 8, no. 3, pp. 283–294.

Hladík, M. (2012) ‘Interval linear programming. A survey’. In: Mann, Z. (ed.), *Linear Programming. New Frontiers in Theory and Applications*, pp. 85–120. New York: Nova Science Publishers.

Mostafaee, A., Hladík, M. and Černý, M. (2013) ‘Inverse linear programming problem with interval coefficients’. *Omega*, submitted.

Ramík, J. (2006) ‘Fuzzy linear optimization’. In: Fiedler, M. et al. (eds.), *Linear Optimization Problems with Inexact Data*, pp. 117–164. New York: Springer.

Rohn, J. (2006) ‘Interval linear programming’. In: Fiedler, M. et al. (eds.), *Linear Optimization Problems with Inexact Data*, pp. 79–100. New York: Springer.

Roos, C., Vial, J.-P, and Terlaky, T. (1998) *Interior Point Methods for Linear Optimization.* New York: Springer.

Schrijver, A. (1988) *Theory of Linear and Integer Programming.* Chichester: Wiley.