A bilevel framework for decision-making under uncertainty with contextual information

IUMA Day on Mathematical Optimization for Data Science

M. A. Muñoz, S. Pineda, J.M. Morales

OASYS group, University of Málaga (Spain)

March 19, 2021

About Málaga

- Over 300 sundays per year (known as Costa del Sol)
- University of Málaga was stablished in 1972 and currently has 40000 students and 2500 faculty members
- Málaga is becoming the Silicon Valley of the south of Spain
- Andalusia Technology Park includes over 600 companies (Oracle, Ericsson, IBM, TDK, Huawei, Microsoft, Cisco), 20.000 employees and a turnover of 2.000 M€ in 2018

Google To Open A Cybersecurity 'Centre Of Excellence' In Malaga

By **Chris King** - 11 February 2021 @ 21:05

About OASYS

- **O**ptimization and **A**nalytics for **S**ustainable energ**Y** **S**ystems
- Established in 2018
- 2 professors, 2 Postdoc, 3 PhD students, 3 research assistants
- Research topics:
  - Mathematical models for decision-making under uncertainty
  - Use of large amounts of data for Smart Energy Grids
  - Forecasting and optimization for Sustainable Energy Systems
  - Algorithms for the efficient solution of large-scale optimization problems
  - Game theory for the analysis of energy markets
- More info: oasys.uma.es

Framework

General optimization problems can be formulated as follows:

$$z^* = \arg\min_{z \in Z} \; f_0(z; y)$$

where

- z are the decisions variables
- y are parameters
- f_0 is the objective function
- Z is the feasible set

Framework

If some parameters are uncertain, stochastic optimization problems can be formulated as follows:

$$z^* = \arg\min_{z \in Z} \; \mathbb{E}[f_0(z; y)]$$

where

- z are the decisions variables
- y are uncertain parameters following a probability distribution $y \sim Y$
- f_0 is the objective function
- Z is the feasible set
- Stochastic programming and robust optimization are used to solve these problems

Framework

If contextual information is used, conditional stochastic optimization problems can be formulated as follows:

$$z^* = \arg\min_{z \in Z} \ \mathbb{E}[f_0(z; y)|x]$$

where

- z are the decisions variables
- y are uncertain parameters, $y \sim Y$
- x are contextual features, $x \sim X$
- f_0 is the objective function
- Z is the feasible set
- Input: available data $S = \{(x_t, y_t, z_t^*), \forall t \in \mathcal{T}\}$ (training)
- Output: optimal decision for a new period $z_{\tilde{t}}^*$ with known $x_{\tilde{t}}$ (test)

Framework

- At 10 am we have to decide how much ice cream to make (decision z)
- At 10 am we do not know the demand in the afternoon (uncertain parameter y)
- We can use some available information such as the temperature at 10 am (contextual data x)
- Obviously there is a relationship between the morning temperature (x) and the ice cream demand in the afternoon (y).
- We would like to use such relation to make better decisions about ice cream quantity

Approaches

Problem

$$z^* = \arg\min_{z \in Z} \; \mathbb{E}[f_0(z; y) | x]$$

Data

$$S = \{(x_t, y_t, z_t^*), \forall t \in \mathcal{T}\}$$

Approaches

Problem

$$z^* = \arg\min_{z \in Z} \ \mathbb{E}[f_0(z; y)|x]$$

Data

$$S = \{(x_t, y_t, z_t^*), \forall t \in \mathcal{T}\}$$

- Forecasting approach (FO)
    - learns the relation between y and x ignoring f_0 and Z

Approaches

Problem

$$z^* = \arg\min_{z \in Z} \ \mathbb{E}[f_0(z; y)|x]$$

Data

$$S = \{(x_t, y_t, z_t^*), \forall t \in \mathcal{T}\}$$

- Forecasting approach (FO)
  - learns the relation between y and x ignoring f_0 and Z
- Decision rule approach (DR)
  - learns the relation between z^* and x

Approaches

Problem

$$z^* = \arg\min_{z \in Z} \ \mathbb{E}[f_0(z;y)|x]$$

Data

$$S = \{(x_t, y_t, z_t^*), \forall t \in \mathcal{T}\}$$

- Forecasting approach (FO)
  - learns the relation between y and x ignoring f_0 and Z
- Decision rule approach (DR)
  - learns the relation between z^* and x
- Bilevel approach (BL)
  - learns the relation between y and x taking into account f_0 and Z

Forecasting approach (FO)

1) Decide a family of functions to learn the relation between y and x so that $\hat{y} = g^{\mathsf{FO}}(x; w)$.

Forecasting approach (FO)

1) Decide a family of functions to learn the relation between y and x so that $\hat{y} = g^{\mathsf{FO}}(x; w)$.
2) Decide a loss function $l^{\mathsf{FO}}(y, \hat{y})$ to quantify the accuracy of the learning task.

Forecasting approach (FO)

1) Decide a family of functions to learn the relation between y and x so that $\hat{y} = g^{\mathsf{FO}}(x; w)$.

2) Decide a loss function $l^{\mathsf{FO}}(y, \hat{y})$ to quantify the accuracy of the learning task.

3) Adjust parameters w using available information as follows

$$w^{\mathsf{FO}} = \arg\min_{w} \sum_{t \in \mathcal{T}} l^{\mathsf{FO}}(g^{\mathsf{FO}}(x_t; w), y_t)$$

Forecasting approach (FO)

1) Decide a family of functions to learn the relation between y and x so that $\hat{y} = g^{\mathsf{FO}}(x; w)$.

2) Decide a loss function $l^{\mathsf{FO}}(y, \hat{y})$ to quantify the accuracy of the learning task.

3) Adjust parameters w using available information as follows

$$w^{\mathsf{FO}} = \arg\min_{w} \sum_{t \in \mathcal{T}} l^{\mathsf{FO}}(g^{\mathsf{FO}}(x_t; w), y_t)$$

4) Determine the optimal decision of a unseen time period $\tilde{t}$ as

$$z_{\tilde{t}}^{\mathsf{FO}} = \arg\min_{z \in Z} \; f_0(z; g^{\mathsf{FO}}(x_{\tilde{t}}; w^{\mathsf{FO}}))$$

Application: strategic producer

- The price of a good p can be expressed in terms of the demand quantity d using a linear inverse demand function $p = \alpha - \beta q$
- The production cost of the same good can be formulated as a quadratic cost function as follows $c_2 q^2 + c_1 q$
- Produced quantity is usually bounded such that $\underline{q} \leqslant q \leqslant \overline{q}$
- The profit of a producer can be computed as

$$\Pi(q) = (\alpha - \beta q)q - c_2 q^2 - c_1 q = -\beta' q^2 + \alpha' q$$

- A strategic producer maximizes profits by solving

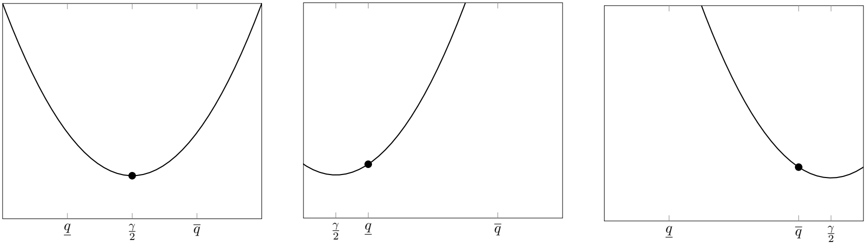
$$q^* = \arg \max_{\underline{q} \leqslant q \leqslant \overline{q}} -\beta' q^2 + \alpha' q = \arg \min_{\underline{q} \leqslant q \leqslant \overline{q}} q^2 - \gamma q$$

- Parameter γ is usually unknown and the producer has available a set of historical measures $S = \{(x_t, \alpha', \beta'_t), \forall t \in \mathcal{T}\}$, with $x_t \in \mathbb{R}^q$

Application: strategic producer

The solution of the quadratic optimization problem is the following

$$\min_{\underline{q} \leqslant q \leqslant \overline{q}} q^2 - \gamma q$$



$q^* = \frac{\gamma}{2}$ $\quad$ $q^* = \underline{q}$ $\quad$ $q^* = \overline{q}$

Forecasting approach (FO)

- We assume a linear function so that $\hat{\gamma}_t = w^T x_t$ with $w \in \mathbb{R}^q$
- We choose the squared error as the loss function $l^{\mathsf{FO}}(\gamma, \hat{\gamma}) = (\gamma - \hat{\gamma})^2$
- We compute w by solving the quadratic problem

$$w^{\mathsf{FO}} = \arg\min_{w \in \mathbb{R}^q} \sum_{t \in \mathcal{T}} (\gamma_t - w^T x_t)^2$$

- We estimate $\hat{\gamma}$ for a new time period $\tilde{t}$ as

$$\hat{\gamma}_{\tilde{t}} = (w^{\mathsf{FO}})^T x_{\tilde{t}}$$

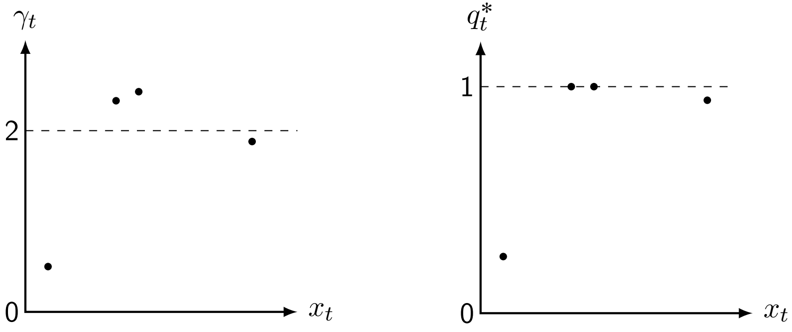
- We determine the produced quantity for a new time period $\tilde{t}$ as

$$q^*_{\tilde{t}} = \arg\min_{\underline{q} \leqslant q_{\tilde{t}} \leqslant \overline{q}} q^2_{\tilde{t}} - \hat{\gamma}_{\tilde{t}} q_{\tilde{t}}$$

Illustrative example

t	x_t	γ_t	$\gamma_t/2$	q_t^*
1	1	0.50	0.25	0.25
2	4	2.33	1.16	1.00
3	5	2.43	1.21	1.00
4	10	1.88	0.94	0.94

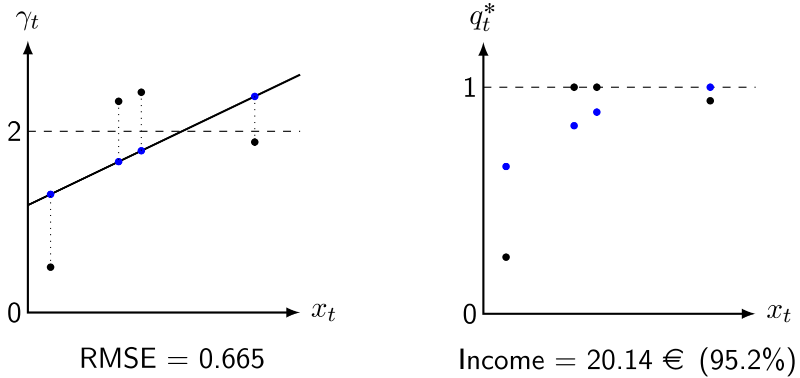
$$0 \leqslant q \leqslant 1$$



Income = 21.16 € (100%)

Illustrative example

Forecasting approach (FO): $\hat{\gamma}_t = 1.184 + 0.12x_t$



Decision-rule approach (DR)

1) Decide a family of functions to learn the relation between z^* and x so that $\hat{z}^* = g^{\mathsf{DR}}(x; w)$.

Decision-rule approach (DR)

1) Decide a family of functions to learn the relation between z^* and x so that $\hat{z}^* = g^{\mathsf{DR}}(x; w)$.

2) Adjust parameters w using available information as follows

$$
\begin{aligned}
w^{\mathsf{DR}} = \arg\min_{w} \sum_{t \in \mathcal{T}} & f_0(g^{\mathsf{DR}}(x_t; w); y_t) \\
\text{s.t. } & g^{\mathsf{DR}}(x_t; w) \in Z \quad \forall t \in \mathcal{T}
\end{aligned}
$$

Decision-rule approach (DR)

1) Decide a family of functions to learn the relation between z^* and x so that $\hat{z}^* = g^{\mathsf{DR}}(x; w)$.

2) Adjust parameters w using available information as follows

$$w^{\mathsf{DR}} = \arg\min_{w} \sum_{t \in \mathcal{T}} f_0(g^{\mathsf{DR}}(x_t; w); y_t)$$
$$\text{s.t. } g^{\mathsf{DR}}(x_t; w) \in Z \ \ \forall t \in \mathcal{T}$$

3) Determine the optimal decision of a unseen time period $\tilde{t}$ as

$$z_{\tilde{t}}^{\mathsf{DR}} = g^{\mathsf{DR}}(x_{\tilde{t}}; w^{\mathsf{DR}})$$

Decision-rule approach (DR)

- We assume a linear function so that $q_t^* = w^T x_t$
- We compute w by solving the quadratic problem

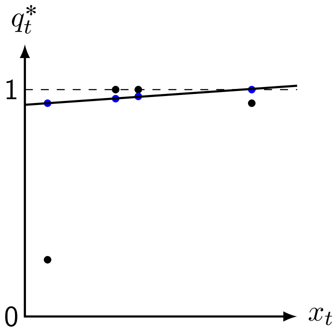
$$
\begin{aligned}
w^{\mathsf{DR}} = \arg\min_{w \in \mathbb{R}^q} \sum_{t \in \mathcal{T}} & \beta_t'(q_t)^2 - \alpha_t' q_t \\
\text{s.t. } & \underline{q} \leqslant q_t \leqslant \overline{q}, \quad \forall t \in \mathcal{T} \\
& q_t = w^T x_t, \quad \forall t \in \mathcal{T}
\end{aligned}
$$

- We determine the produced quantity for a new time period $\tilde{t}$ as

$$q_{\tilde{t}}^* = (w^{\mathsf{DR}})^T x_{\tilde{t}}$$

Illustrative example

Decision-rule approach (DR): $\hat{q}_t = 0.933 + 0.007x_t$



Income = 20.02 € (94.6%)

A bit of bilevel programming

John and Peter have a deal to go to the movies. First John decides the movie theater, and then Peter decides which movie they will watch. John prefers action over terror movies, and terror over romantic movies. Peter prefers romantic over terror, and terror over action movies. Which theater do you think John would choose?

Theater A
Spiderman
Notting Hill

Theater B
The Exorcist
The Matrix

A bit of bilevel programming

Bilevel optimization is a special kind of optimization where one problem is embedded (nested) within another and can be formulated as follows

$$
\begin{aligned}
\min_{x} \quad & F_0(x,y) \\
\text{s.t.} \quad & F_i(x,y) \leqslant 0, \qquad i = 1,\dots,I \\
& H_j(x,y) = 0, \qquad j = 1,\dots,J \\
& \min_{y} \quad f_0(x,y) \\
& \text{s.t.} \quad f_k(x,y) \leqslant 0, \qquad k = 1,\dots,K \\
& \qquad\quad h_l(x,y) = 0, \qquad l = 1,\dots,L
\end{aligned}
$$

Bilevel approach (BL)

1) Decide a family of functions to learn the relation between y and x so that $\hat{y} = g^{\mathsf{BL}}(x; w)$.

Bilevel approach (BL)

1) Decide a family of functions to learn the relation between y and x so that $\hat{y} = g^{\mathsf{BL}}(x; w)$.

2) Adjust parameters w by solving the following bilevel problem

$$\begin{aligned} w^{\mathsf{BL}} = \arg\min_{w} & \sum_{t \in \mathcal{T}} f_0(\hat{z}_t; y_t) \\ & \text{s.t. } \hat{z}_t = \arg\min_{z \in Z} f_0(z; g^{\mathsf{BL}}(x_t; w)) \ \ \forall t \in \mathcal{T} \end{aligned}$$

Bilevel approach (BL)

1) Decide a family of functions to learn the relation between y and x so that $\hat{y} = g^{\mathsf{BL}}(x; w)$.

2) Adjust parameters w by solving the following bilevel problem

$$w^{\mathsf{BL}} = \arg\min_{w} \sum_{t \in \mathcal{T}} f_0(\hat{z}_t; y_t)$$

$$\text{s.t. } \hat{z}_t = \arg\min_{z \in Z} f_0(z; g^{\mathsf{BL}}(x_t; w)) \ \ \forall t \in \mathcal{T}$$

3) Determine the optimal decision of a unseen time period $\tilde{t}$ as

$$z_{\tilde{t}}^{\mathsf{BL}} = \arg\min_{z \in Z} \ f_0(z; g^{\mathsf{BL}}(x_{\tilde{t}}; w^{\mathsf{BL}}))$$

Bilevel approach (BL)

We express the feasible set Z with the following set of inequalities and equalities:

$$f_i(z) \leqslant 0, \quad i = 1, ..., I$$
$$h_j(z) = 0, \quad j = 1, ..., J$$

The bilevel problem is then formulated as follows:

$$\begin{aligned}
&\min_w \sum_{t \in \mathcal{T}} f_0(\hat{z}_t; y_t) \\
&\text{s.t.} \quad \hat{z}_t = \arg\min_z f_0(z; g^{\mathsf{BL}}(x_t; w)) \\
&\qquad\qquad\qquad f_i(z) \leqslant 0, \quad i = 1, ..., I \qquad \forall t \in \mathcal{T} \\
&\qquad\qquad\qquad h_j(z) = 0, \quad j = 1, ..., J
\end{aligned}$$

Bilevel approach (BL)

Under convexity assumption of the lower-level problem, it can be replaced by its KKT optimality conditions as follows:

$$\min_{w,\hat{z}_t,\lambda_{it},\upsilon_{it}} \sum_{t\in\mathcal{T}} f_0(\hat{z}_t; y_t)$$

$$\begin{aligned}
\text{s.t. } & \nabla f_0(\hat{z}_t, g^{\mathsf{BL}}(x_t, w)) + \sum_{i=1}^{I} \lambda_{it} \nabla f_i(\hat{z}_t) + \sum_{j=1}^{J} \upsilon_{jt} \nabla h_j(\hat{z}_t) = 0, \ \ \forall t \in \mathcal{T} \\
& f_i(\hat{z}_t) \leqslant 0, \ \ \forall i, \ \ \forall t \in \mathcal{T} \\
& h_j(\hat{z}_t) = 0, \ \ \forall j, \ \ \forall t \in \mathcal{T} \\
& \lambda_{it} \geqslant 0, \ \ \forall i, \ \ \forall t \in \mathcal{T} \\
& \lambda_{it} f_i(\hat{z}_t) = 0, \ \ \forall i, \ \ \forall t \in \mathcal{T}
\end{aligned}$$

Under certain assumptions, this problem can be solved using bigM (mixed-integer problem) or regularization techniques (non-linear problem)

Bilevel approach (BL)

- We assume a linear function so that $\hat{\gamma}_t = w^T x_t$ with $w \in \mathbb{R}^q$
- We compute w by solving the bilevel

$$w^{\mathsf{BL}} = \arg\min_{w \in \mathbb{R}^q} \sum_{t \in \mathcal{T}} \beta'_t \hat{q}_t^2 - \alpha'_t \hat{q}_t$$

$$\text{s.t. } \hat{q}_t = \arg\min_{\underline{q} \leqslant q_t \leqslant \overline{q}} q_t^2 - w^T x_t q_t, \ \ \forall t \in \mathcal{T}$$

- We determine the produced quantity for a new time period $\tilde{t}$ as

$$q^*_{\tilde{t}} = \arg\min_{\underline{q} \leqslant q_{\tilde{t}} \leqslant \overline{q}} q_{\tilde{t}}^2 - (w^{\mathsf{BL}})^T x_{\tilde{t}} q_{\tilde{t}}$$

Bilevel approach (BL)

We replace the lower-level problem by its KKT and reformulate the complementarity conditions using Fortuny-Amat to obtain the following single-level mixed-integer quadratic optimization problem

$$
\begin{aligned}
w^{\mathsf{BL}} = \arg\min_{w} \ & \sum_{t\in\mathcal{T}} \beta_t' \hat{q}_t^2 - \alpha_t' \hat{q}_t \\
\text{s.t. } & 2\hat{q}_t - w^T x_t - \lambda_{1t} + \lambda_{2t} = 0, \ \ \forall t \in \mathcal{T} \\
& 0 \leqslant \hat{q}_t - \underline{q} \leqslant (1 - u_{1t}) M^P, \ \ \forall t \in \mathcal{T} \\
& 0 \leqslant \overline{q} - \hat{q}_t \leqslant (1 - u_{2t}) M^P, \ \ \forall t \in \mathcal{T} \\
& 0 \leqslant \lambda_{1t} \leqslant u_{1t} M^D, \ \ \forall t \in \mathcal{T} \\
& 0 \leqslant \lambda_{2t} \leqslant u_{2t} M^D, \ \ \forall t \in \mathcal{T} \\
& u_{1t}, u_{2t} \in \{0, 1\}, \ \ \forall t \in \mathcal{T}
\end{aligned}
$$

where $\lambda_{1t}, \lambda_{2t}$ are dual variables of the lower-level problem and M^P, M^D are large enough constants

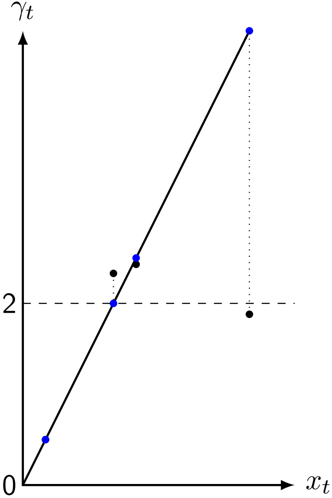
Bilevel approach (BL)

Alternatively, we can use the regularization approach and iteratively solve the non-linear optimization below for decreasing values of parameter ϵ

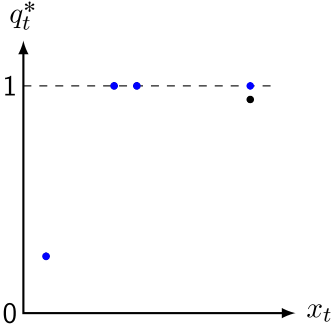
$$
\begin{aligned}
w^{\mathsf{BL}} = \arg\min_{w} \quad & \sum_{t \in \mathcal{T}} \beta_t' \hat{q}_t^2 - \alpha_t' \hat{q}_t \\
\text{s.t.} \quad & 2\hat{q}_t - w^T x_t - \lambda_{1t} + \lambda_{2t} = 0, \quad \forall t \in \mathcal{T} \\
& \underline{q} \leqslant \hat{q}_t \leqslant \overline{q}, \quad \forall t \in \mathcal{T} \\
& \lambda_{1t}, \lambda_{2t} \geqslant 0, \quad \forall t \in \mathcal{T} \\
& \lambda_{1t}(\hat{q}_t - \underline{q}) + \lambda_{2t}(\overline{q} - \hat{q}_t) \leqslant \epsilon
\end{aligned}
$$

Illustrative example

Bilevel approach (BL): $\hat{\gamma}_t = 0.000 + 0.5x_t$



RMSE = 0.745



Income = 21.13 € (99.9%)

Illustrative example

	RMSE	Income
Forecasting (FO)	0.665	95.2%
Decision-rule (DR)	-	94.6%
Bilevel (BL)	0.745	99.9%

- FO minimizes forecast error, but yields suboptimal decisions
- DR simplifies decision-making, but also yields suboptimal decisions
- BL relates uncertain parameters and contextual information to derive the best decisions

Case study

- Real data from Iberian electricity market is used to approximate the inverse demand function
- Wind and solar power forecasts is used as contextual information
- Three different generation technologies: base (nuclear), medium (carbon) and peak (gas)
- 43 sets of 200 hours (160 hours as training and 40 hours as test) are used to compute average results

Case study

	Base	Medium	Peak
Relative income FO	96.0%	77.3%	41.6%
Relative income DR	94.6%	62.6%	18.9%
Relative income BL	96.3%	80.0%	58.7%
Infeasible cases DR	4.9%	1.7%	0.1 %

- All methods provide similar incomes for the base unit since it is at full capacity most of the time
- The uncertainty of the inverse demand function significantly affects the operation of medium and peak units
- The proposed BL approach obtains the highest incomes for the three generating technologies
- DR approach lead to a significant number of infeasible cases

Conclusions

- Forecasting approach (FO)
  - learns the relation between y and x ignoring f_0 and Z
  - 👍 wide variety of learning techniques can be applied
  - 👎 obtained decisions may be suboptimal
- Decision rule approach (DR)
  - learns the relation between z^* and x
  - 👍 decisions are quickly obtained wihtout solving an optimization problem
  - 👎 obtained decisions may be suboptimal and infeasible
- Bilevel approach (BL)
  - learns the relation between y and x taking into account f_0 and Z
  - 👍 best possible decisions using available contextual information
  - 👎 bilevel problem can be only solved under certain assumptions

Thanks for the attention!! Questions??

A bilevel framework for decision-making under uncertainty with contextual information

M. A. Muñoz, S. Pineda, J. M. Morales

OASYS Group, University of Malaga, Malaga, Spain

Abstract

In this paper we propose a novel approach for data-driven decision-making under uncertainty in the presence of contextual information. Given a finite collection of observations of the uncertain parameters and potential explanatory variables (i.e., the contextual information), our approach fits a parametric model to those data that is specifically tailored to maximizing the decision value, while accounting for possible feasibility constraints. From a mathematical point of view, our framework translates into a bilevel program, for which we provide both a fast regularization procedure and a big-M-based reformulation to aim for a local and a global optimal solution, respectively. We showcase the benefits of moving from the traditional scheme for model estimation (based on statistical quality metrics) to decision-guided prediction using the problem of a strategic producer competing *la Cournot* in a market for an homogeneous product. In particular, we include a realistic case study, based on data from the Iberian electricity market, whereby we compare our approach with alternative ones available in the technical literature and analyze the conditions (in terms of the firm's cost structure and production capacity) under which our approach proves to be more advantageous to the producer.

OASYS

Optimization and Analytics for Sustainable energY Systems

More info: oasys.uma.es

Email: spineda@uma.es