A fast algorithm to estimate the cost and the right-hand side parameter vectors in inverse linear optimization

EURO 2019, Dublin, Ireland

Juan M. Morales[1], Ricardo Fernández-Blanco[1], Salvador Pineda[1]

[1]OASYS Research Group, University of Malaga, Spain

June 25, 2019

European Research Council
Established by the European Commission

Optimization and Analytics for Sustainable energY Systems

Subject of Study: Inverse linear optimization

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{Minimize}} & (\mathbf{c}^T + \mathbf{s}^T)\mathbf{x} \\ \text{Subject to} & \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{array}$$

Objective: Estimate **c** and **b** from data (observations) $\left\{(\tilde{\mathbf{x}}_t, \mathbf{s}_t)_{t=1}^T\right\}$, i.e.,

Input: $\{\mathbf{s}_t, t = 1, \ldots, T\}$

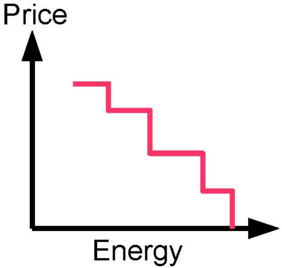
Output: $\{\mathbf{x}_t, t = 1, \ldots, T\}$, but we observe $\{\tilde{\mathbf{x}}_t, t = 1, \ldots, T\}$, as measurements may be contaminated with *noise*

Motivating example[1]

$$\max_{x_b, \forall b} \quad \sum_{b=1}^{B} x_b(u_b - p)$$

$$\text{s. t.} \quad \underline{P} \leq \sum_{b=1}^{B} x_b \leq \overline{P} \qquad (\underline{\lambda}, \overline{\lambda})$$

$$0 \leq x_b \leq E_b \qquad (\underline{\phi}_b, \overline{\phi}_b)$$



Objective: Estimate u_b, $b = 1, \ldots, B$, $\underline{P}$ and $\overline{P}$ from observations $\left\{(\tilde{x}_t, p_t)_{t=1}^{T}\right\}$, i.e.,

Input: $\{p_t, t = 1, \ldots, T\}$

Output: $\{x_{b,t}, b = 1, \ldots, B; t = 1, \ldots, T\}$, but we only observe the (noisy) *aggregate* consumption $\{\tilde{x}_t, t = 1, \ldots, T\}$

[1] J. Saez-Gallego, J. M. Morales, M. Zugno and H. Madsen (2016). A data-driven bidding model for a cluster of price-responsive consumers of electricity. *IEEE Transactions on Power Systems* 31(6): 5001–5011

Motivating example

Time-variant parameters

$$\underline{P}_t = \underline{\mu} + \sum_{r=1}^{R} \underline{\alpha}_r Z_{r,t}$$

$$\overline{P}_t = \overline{\mu} + \sum_{r=1}^{R} \overline{\alpha}_r Z_{r,t}$$

$$u_{b,t} = \mu_b^u + \sum_{r=1}^{R} \alpha_r^u Z_{r,t}$$

Objective: Estimate $\underline{\mu}$, $\overline{\mu}$; μ_b^u, $b = 1, \ldots, B$; $\underline{\alpha}_r$, $\overline{\alpha}_r$, and α_r^u, $r = 1, \ldots, R$ from observations $\left\{(\tilde{x}_t, p_t, Z_{1,t}, Z_{2,t}, \ldots, Z_{R,t})_{t=1}^{T}\right\}$

Estimation: Duality-gap minimization

$$
\begin{aligned}
\underset{\Omega}{\text{minimize}} \quad & \sum_{t=1}^{T} \epsilon_t \\
\text{subject to} \quad & \overline{P}_t \overline{\lambda}_t - \underline{P}_t \underline{\lambda}_t + \sum_{b=1}^{B} E_b \overline{\phi}_{b,t} - \epsilon_t = \sum_{b=1}^{B} x_{b,t}(u_{b,t} - p_t), \quad \forall t \\
& \overline{\phi}_{b,t} - \underline{\phi}_{b,t} + \overline{\lambda}_t - \underline{\lambda}_t = u_{b,t} - p_t, \quad \forall t \\
& \overline{\phi}_{b,t}, \underline{\phi}_{b,t}, \overline{\lambda}_t, \underline{\lambda}_t, \geq 0, \quad \forall t \\
& \epsilon_t \geq 0, \quad \forall t
\end{aligned}
$$

$$
\Omega = \left\{ \left(\epsilon_t, \overline{P}_t, \underline{P}_t, u_{b,t}, \overline{\lambda}_t, \underline{\lambda}_t, \overline{\phi}_{b,t}, \underline{\phi}_{b,t} \right)_{t=1}^{T} \right\}
$$

Estimation: Duality-gap minimization[2]

- The estimation problem is **non-linear and non-convex**.
- We statistically **approximate its solution** by solving two **linear programming problems** instead.

  1. A **feasibility** problem (estimation of power bounds).
  2. An **optimality** problem (estimation of marginal utilities).

- A **two-step data-driven estimation procedure** to achieve **optimality** and **feasibility** of $\tilde{\mathbf{x}}$ in a statistical sense.

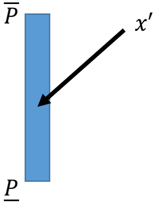
[2] J. Saez-Gallego and J. M. Morales (2018). Short-term forecasting of price-responsive loads using inverse optimization. *IEEE Transactions on Smart Grid* 9(5):4805–4814

Feasibility problem: Estimation of power bounds

$$\underset{\underline{\boldsymbol{P}},\overline{\boldsymbol{P}},\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\alpha}}{\text{Minimize}} \sum_{t=1}^{T} \left((1-\mathrm{K}) \left(\overline{\xi}_{\mathrm{t}}^{+} + \underline{\xi}_{\mathrm{t}}^{+} \right) + \mathrm{K} \left(\overline{\xi}_{\mathrm{t}}^{-} + \underline{\xi}_{\mathrm{t}}^{-} \right) \right)$$

subject to

$$\begin{aligned}
&\overline{P}_t - x'_t = \overline{\xi}_t^{+} - \overline{\xi}_t^{-} && \forall t\\
&x'_t - \underline{P}_t = \underline{\xi}_t^{+} - \underline{\xi}_t^{-} && \forall t\\
&\underline{P}_t \leq \overline{P}_t && \forall t\\
&\underline{P}_t = \underline{\mu} + \sum_{r=1}^{R} \underline{\alpha}_r Z_{r,t} && \forall t\\
&\overline{P}_t = \overline{\mu} + \sum_{r=1}^{R} \overline{\alpha}_r Z_{r,t} && \forall t\\
&0 \leq \overline{\xi}_t^{+}, \overline{\xi}_t^{-}, \underline{\xi}_t^{+}, \underline{\xi}_t^{-} && \forall t
\end{aligned}$$

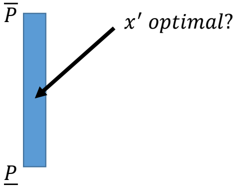


$\widehat{\overline{P}}_t, \widehat{\underline{P}}_t, \hat{\underline{\mu}}, \hat{\overline{\mu}}, \hat{\underline{\alpha}}_r, \hat{\overline{\alpha}}_r$

K controls the percentage of observations falling within $[\underline{P}, \overline{P}]$

Optimality problem: Estimating marginal utilities

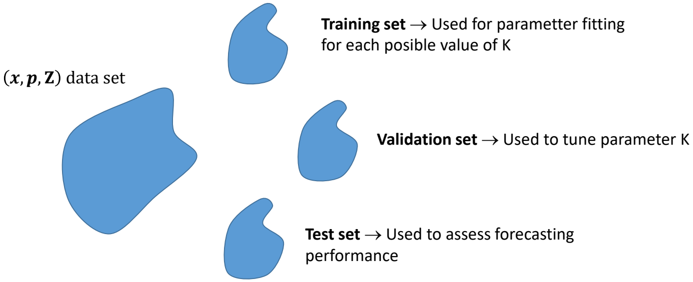
$$
\begin{aligned}
\underset{\Omega}{\text{Minimize}} \; & \sum_{t=1}^{T} \epsilon_t \\
\text{subject to } & \widehat{\overline{P}}_t \overline{\lambda}_t - \underline{\widehat{P}}_t \underline{\lambda}_t + \sum_{b=1}^{B} E_b \overline{\phi}_{b,t} - \epsilon_t = \\
& \qquad \sum_{b=1}^{B} \tilde{x}'_{b,t} \left(u_{b,t} - p_t \right) & \forall t \\
& -\underline{\phi}_{b,t} + \overline{\phi}_{b,t} - \underline{\lambda}_t + \overline{\lambda}_t = u_{b,t} - p_t & \forall b, t \\
& u_{b,t} = \mu^u_b + \sum_{r} \alpha^u_r Z_{r,t} & \forall b, t \\
& \mu^u_b \geq \mu^u_{b+1} & \forall b < B \\
& \mu^u_1 \geq 200 + \mu^u_2 \\
& 0 \leq \overline{\lambda}_t, \underline{\lambda}_t, \underline{\phi}_{b,t}, \overline{\phi}_{b,t}, \epsilon_t & \forall b, t.
\end{aligned}
$$



$\hat{u}_{b,t}, \hat{\mu}^u_b, \hat{\alpha}^u_r$

Solving the estimation problem

In the **bound estimation problem**, the **penalty parameter K** is statistically tuned through **validation**:



We choose K so that the **out-of-sample prediction error** is minimized

Noiseless case: $\tilde{x}_t = x_t, \quad t = 1, \ldots, T$

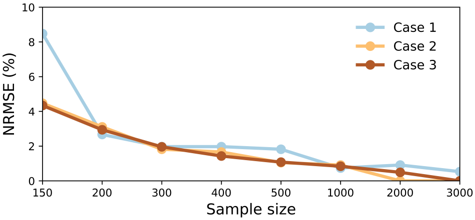
- We assume $p_t \sim N(0.050, 0.0125)$; $\quad \mathbf{Z} = (p_{t-1}, p_{t-2})$
- We analyze 3 cases:
    - Case 1: $(u_{b,t}, \underline{P}_t, \overline{P}_t)$ are time-invariant
    - Case 2: $u_{b,t}$ is time-variant and $(\underline{P}_t, \overline{P}_t)$ are time-invariant
    - Case 3: $(u_{b,t}, \underline{P}_t, \overline{P}_t)$ are time-variant
- Let us assume $u_{1,t} = 0.2$, $\mu_2^u = 0.070$, $\mu_b^u = \mu_{b-1} - 0.0015 \quad \forall b > 2$, $\underline{\mu} = 0$, $\overline{\mu} = 100$

Case	α_r^u		$\underline{\alpha}_r$		$\overline{\alpha}_r$	
	1	2	1	2	1	2
1	0.0	0.0	0.0	0.0	0.0	0.0
2	-0.5	0.3	0.0	0.0	0.0	0.0
3	-0.5	0.3	100	-20	200	-100

- We consider 8 data samples of sizes: 150, 200, 300, 400, 500, 1000, 2000, 3000. $\Omega^{test} = 100$ last periods (identical for all of them); Ω^{tr} / $\Omega^{v} \Rightarrow 70\%/30\%$

Noiseless case: $\tilde{x}_t = x_t, \quad t = 1, \ldots, T$

- Normalized root mean square error

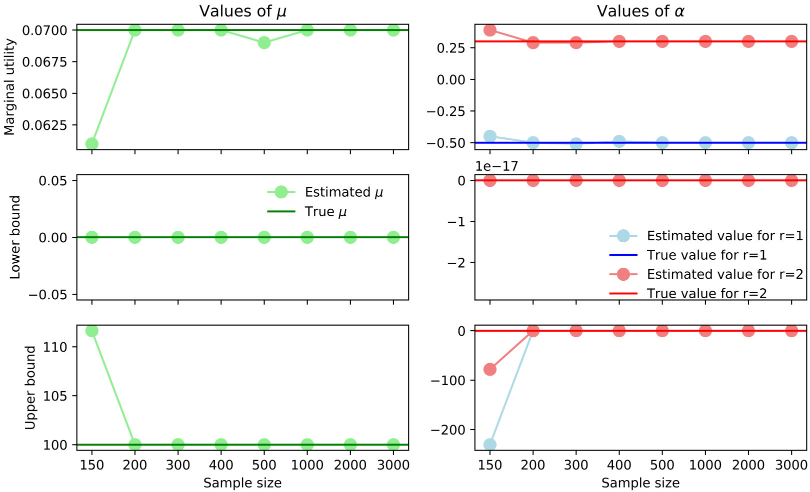


- Optimal hyper-parameter K tends to 1 when increasing the sample size
- Estimated parameter for case 1 (time-invariant!)

Sample size	μ_2^u	$\underline{\mu}$	$\overline{\mu}$
150	0.062	21.05	100.0
200	0.067	5.26	100.0
300	0.070	0.00	100.0
True	0.070	0.00	100.0

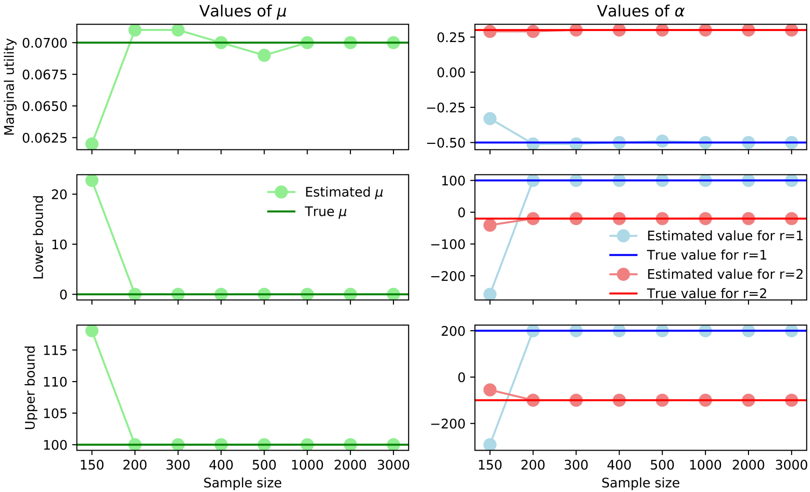
Noiseless case: $\tilde{x}_t = x_t, \quad t = 1, \ldots, T$

- Results for case 2: $u_{b,t}$ is time-variant and $(\underline{P}_t, \overline{P}_t)$ are time-invariant



Noiseless case: $\tilde{x}_t = x_t, \quad t = 1, \ldots, T$

- Results for case 3: $(u_{b,t}, \underline{P}_t, \overline{P}_t)$ are time-variant



Noiseless case: $\tilde{x}_t = x_t, \quad t = 1, \ldots, T$

Takeaway message: Our two-step estimation procedure performs satisfactorily in a noiseless setup, but...

What if the observations $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_T)$ are corrupted with some measurement i.i.d. noise $\omega_t \sim N(0, \sigma)$?

Noisy case: $\tilde{x}_t = x_t + \omega_t, \quad t = 1, \ldots, T; \quad \omega_t \sim N(0, \sigma)$

Consider the following illustrative linear program[3]:

$$\begin{aligned} \min_{x} \quad & (\theta + p) \cdot x \\ \text{s. t} \quad & -1 \leq x \leq 1 \end{aligned}$$

Objective: Estimate θ from observations $(\tilde{x}_t, p_t)$ where:

$p_1, p_2, \ldots, p_t, \ldots, p_T$ i.i.d, with $p_t \sim U(-2, 1)$

$\tilde{x}_t = x_t + \omega_t$, with $\omega_1, \ldots, \omega_t, \ldots, \omega_T$ i.i.d with $\omega_t \sim N(0, 1)$

[3] A. Aswani, Z.-J. Shen and A. Siddiq (2018). Inverse optimization with noisy data. *Operations Research* 66(3):870–892

Noisy case: $\tilde{x}_t = x_t + \omega_t, \quad t = 1, \ldots, T; \quad \omega_t \sim N(0, \sigma)$

Optimality problem:

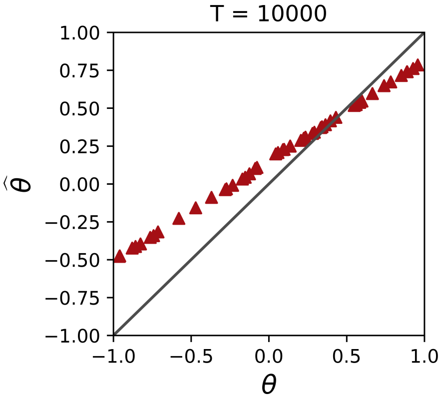
$$\min_{\widehat{\theta}, \epsilon_t, \overline{\beta}_t, \underline{\beta}_t} \quad \sum_{t=1}^{T} \epsilon_t$$

$$\text{s. t.} \quad \left(\widehat{\theta} + p_t\right) \cdot \widetilde{x}_t - \epsilon_t = -\overline{\beta}_t - \underline{\beta}_t, \quad \forall t$$

$$\underline{\beta}_t - \overline{\beta}_t = \theta + u_t, \quad \forall t$$

$$\overline{\beta}_t, \underline{\beta}_t \geq 0, \quad \forall t$$

$$\epsilon_t \geq 0, \quad \forall t$$



Unable to recover the true parameters!

Noisy case: $\tilde{x}_t = x_t + \omega_t, \quad t = 1, \dots, T; \;\; \omega_t \sim N(0, \sigma)$

Any idea to preserve statistical consistency?

Let us assume asymptotic regime, i.e., $T \upuparrows$:

Weak duality: $\left(\widehat{\theta} + p_t\right) \cdot x_t \geq -\overline{\beta}_t - \underline{\beta}_t$

Strong duality: $(\theta + p_t) \cdot x_t = -\overline{\beta}_t(\theta) - \underline{\beta}_t(\theta)$

$$\left(\widehat{\theta} + p_t\right) \cdot \widetilde{x}_t - \epsilon_t = -\overline{\beta}_t - \underline{\beta}_t \Rightarrow \left(\widehat{\theta} + p_t\right) \cdot (x_t + \omega_t) - \epsilon_t = -\overline{\beta}_t - \underline{\beta}_t \Rightarrow$$

$$\Rightarrow \left(\widehat{\theta} + p_t\right) \cdot x_t + \widehat{\theta}\omega_t + p_t\omega_t - \epsilon_t = -\overline{\beta}_t - \underline{\beta}_t$$

Since $\frac{1}{T}\sum_{t=1}^{T} \omega_t \to 0$ and $\frac{1}{T}\sum_{t=1}^{T} p_t\omega_t \to 0$ when $T \upuparrows$, we have

$$\sum_{t=1}^{T} \epsilon_t = \sum_{t=1}^{T} \left(\widehat{\theta} + p_t\right) \cdot x_t + \overline{\beta}_t + \underline{\beta}_t \geq 0 \text{ and } = 0 \text{ iff } \widehat{\theta} = \theta$$

Noisy case: $\tilde{x}_t = x_t + \omega_t, \quad t = 1, \ldots, T; \;\; \omega_t \sim N(0, \sigma)$

Recast optimality problem

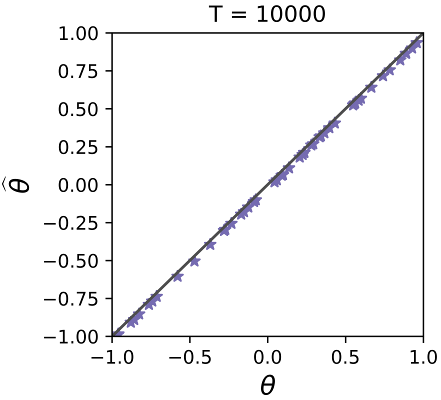
$$\min_{\theta, \epsilon_t, \overline{\beta}_t, \underline{\beta}_t} \quad \sum_{t=1}^{T} \epsilon_t$$

$$\text{s. t.} \quad (\theta + p_t) \cdot \widetilde{x}_t - \epsilon_t = -\overline{\beta}_t - \underline{\beta}_t, \quad \forall t$$

$$\underline{\beta}_t - \overline{\beta}_t = \theta + u_t, \quad \forall t$$

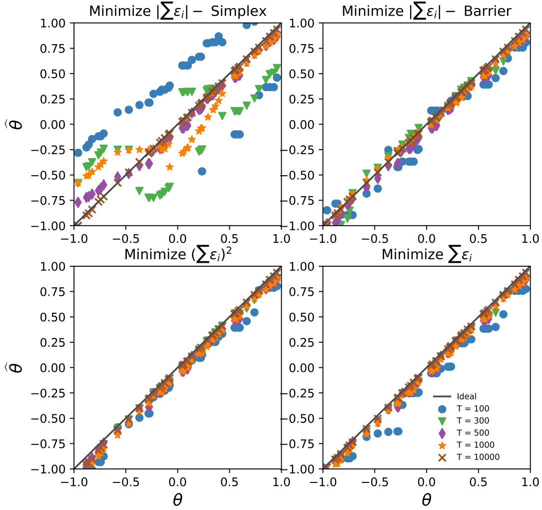
$$\overline{\beta}_t, \underline{\beta}_t \geq 0, \quad \forall t$$

$$\epsilon_t \geq 0, \quad \forall t : \tilde{x}_t \in [-1, 1]$$



Minimizing $\left|\sum_{t=1}^{T} \epsilon_i\right|$ or $\left(\sum_{t=1}^{T} \epsilon_i\right)^2$ should be theoretically equivalent when $T \upuparrows$, but not in practice (finite-sample regime)

Noisy case: $\tilde{x}_t = x_t + \omega_t, \quad t = 1, \ldots, T; \ \ \omega_t \sim N(0, \sigma)$

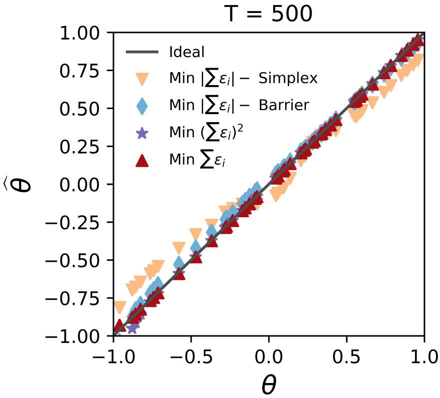


Noisy case: $\tilde{x}_t = x_t + \omega_t, \;\; t = 1, \ldots, T; \;\; \omega_t \sim N(0, \sigma)$

Statistical consistency is a theoretical nice property, but not that useful in practice, where we normally work under a finite-sample regime (that is, T not big enough)

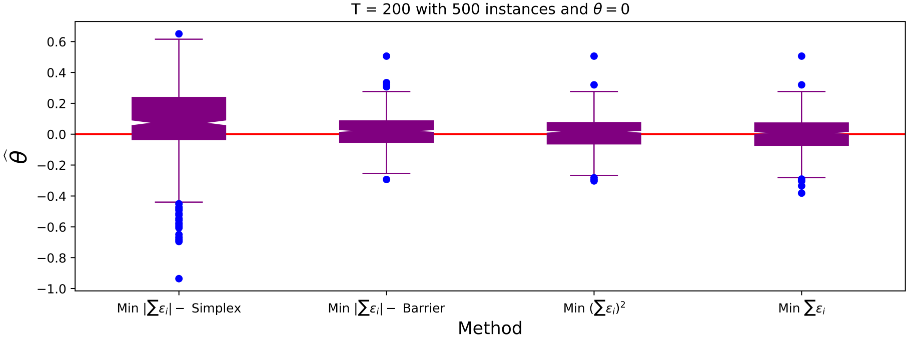
What about the bias and variance of the estimated parameters?

Noisy case: $\tilde{x}_t = x_t + \omega_t, \quad t = 1, \ldots, T; \ \ \omega_t \sim N(0, \sigma)$



Checking for **bias**: Estimator average over 100 instances of $T = 500$

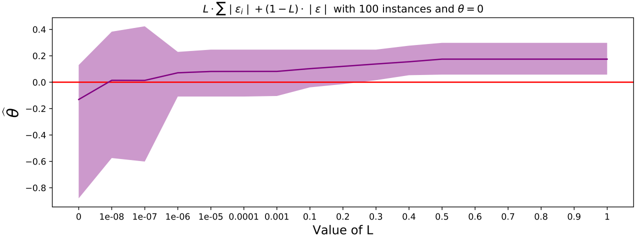
Noisy case: $\tilde{x}_t = x_t + \omega_t, \quad t = 1, \ldots, T; \quad \omega_t \sim N(0, \sigma)$

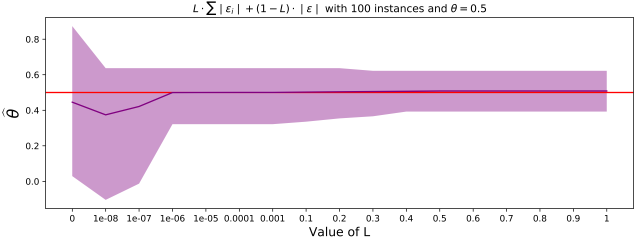


Checking for **bias** and **variance**: Box plot of the parameter estimator computed for 500 instances of $T = 200$ (true value $\theta = 0$)

Noisy case: $\tilde{x}_t = x_t + \omega_t, \;\; t = 1, \ldots, T; \;\; \omega_t \sim N(0, \sigma)$

Bias-variance trade-off: Minimize $L \sum_t^T |\epsilon_t| + (1 - L) \left| \sum_{t=1}^T \epsilon_i \right|$, with L being a control parameter ($L \to 0$ as $T \uparrow$)





Conclusions

- A fast method for inverse linear optimization able to recover the true parameters under a noiseless setup (even in the case of unknown right-hand side parameter vector)
- Different alternatives to preserve statistical consistency in the presence of noise in the measurements
- These alternatives exhibit different behaviors in terms of bias and variance

Future Work

- Need for theoretical insight
- Extension to the case of unknown parameters in the constraints
- Extension to inverse convex optimization

Contacts

Any questions?

European Research Council
Established by the European Commission

Optimization and Analytics for Sustainable energY Systems

Juan Miguel Morales
juan.morales@uma.es
OASYS Webpage: oasys.uma.es