An Exact Symbolic Reduction of Linear Smart Predict+Optimize (SPO) to Mixed Integer Linear Program (MILP)

Jihwan Jeong[1, 2], Parth Jaggi[1], Andrew Butler[1], Scott Sanner[1, 2]

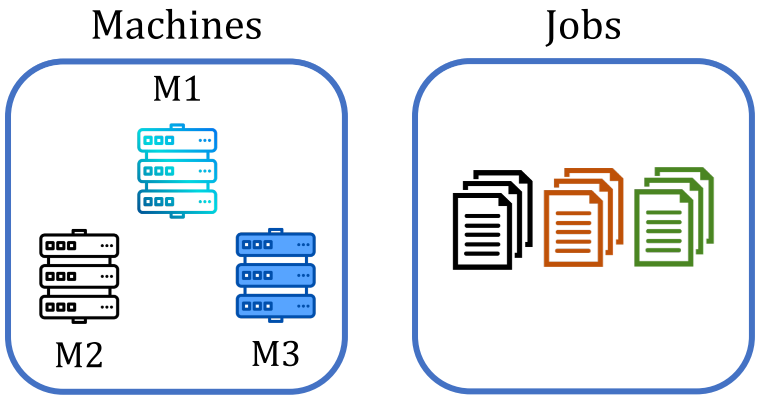
[1]University of Toronto

[2]Vector Institute

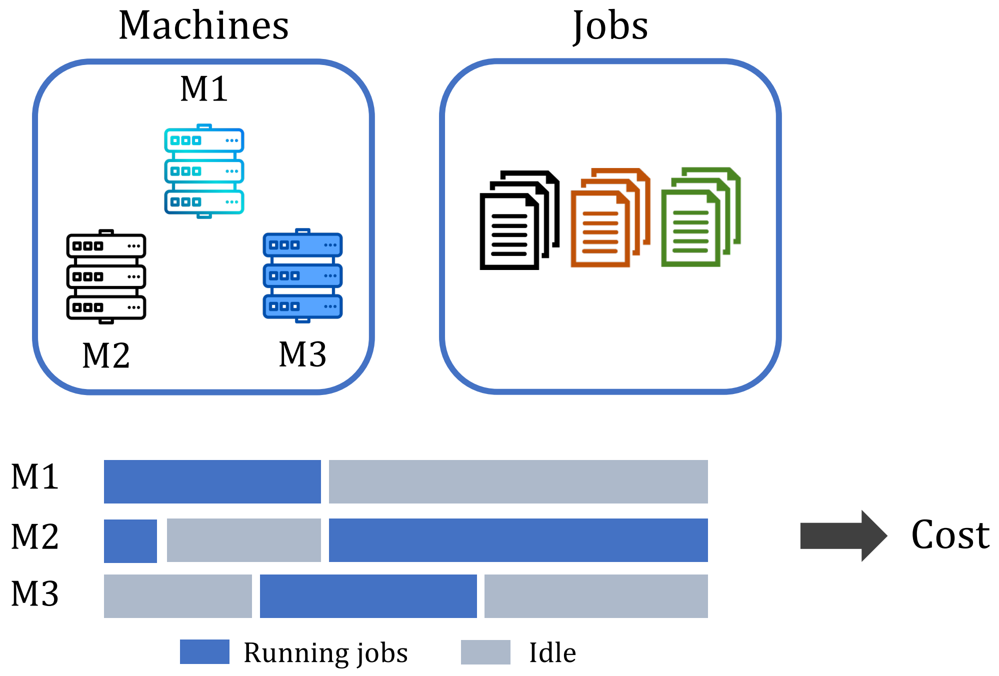
jihwan.jeong@mail.utoronto.ca

UNIVERSITY OF TORONTO

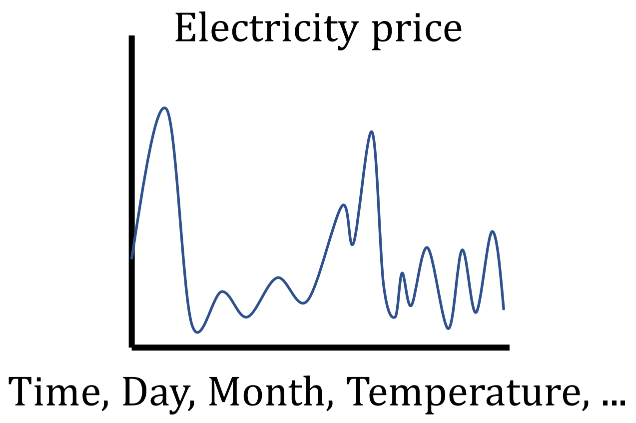
Motivation – Linear Smart "Predict, then Optimize"

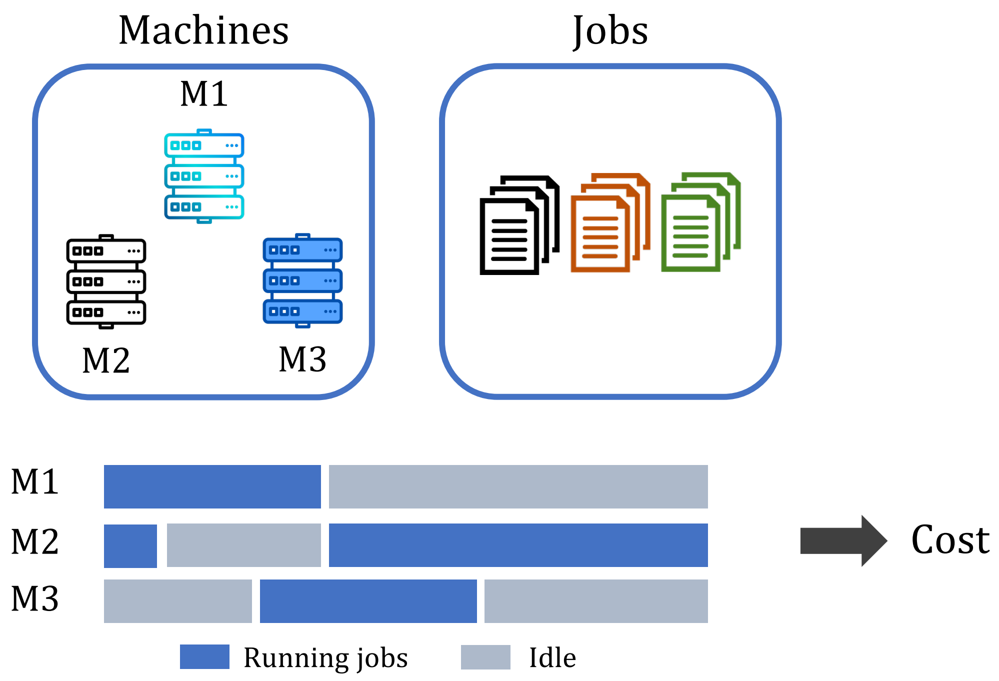


Motivation – Linear Smart "Predict, then Optimize"

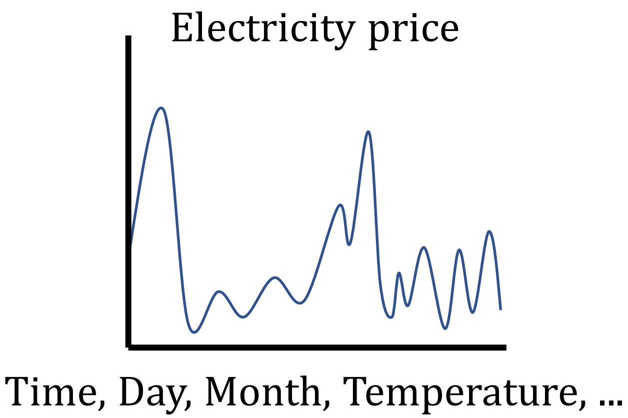


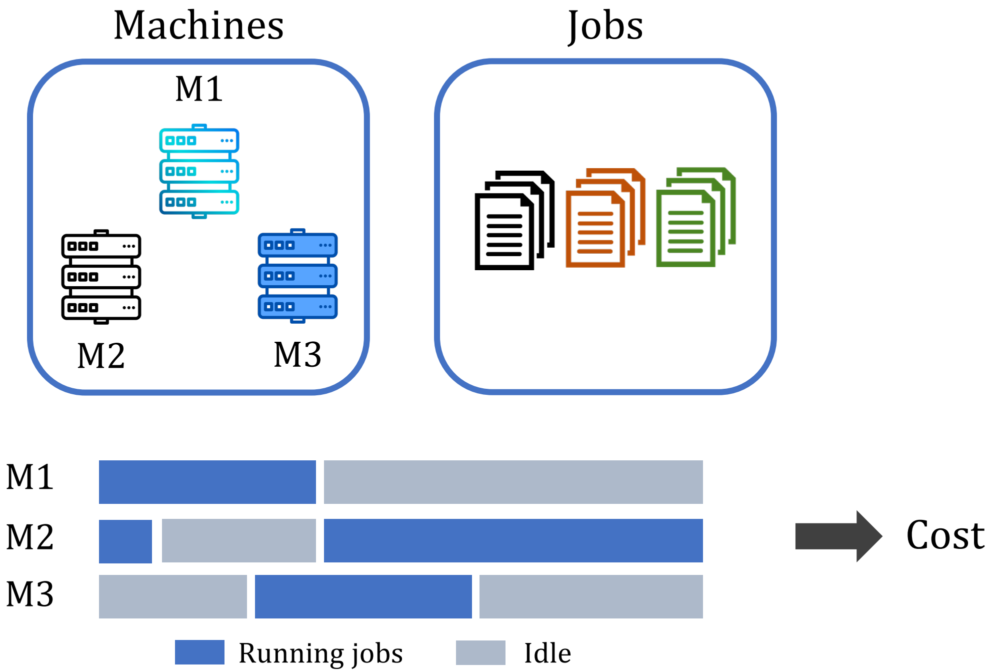
Motivation – Linear Smart "Predict, then Optimize"



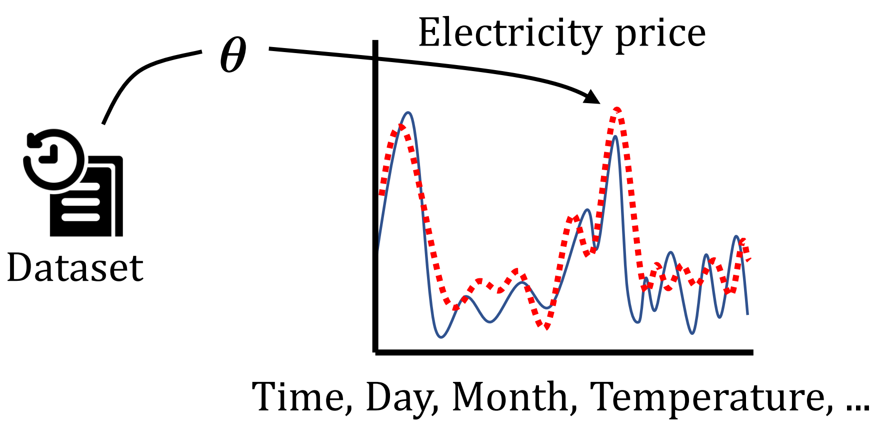


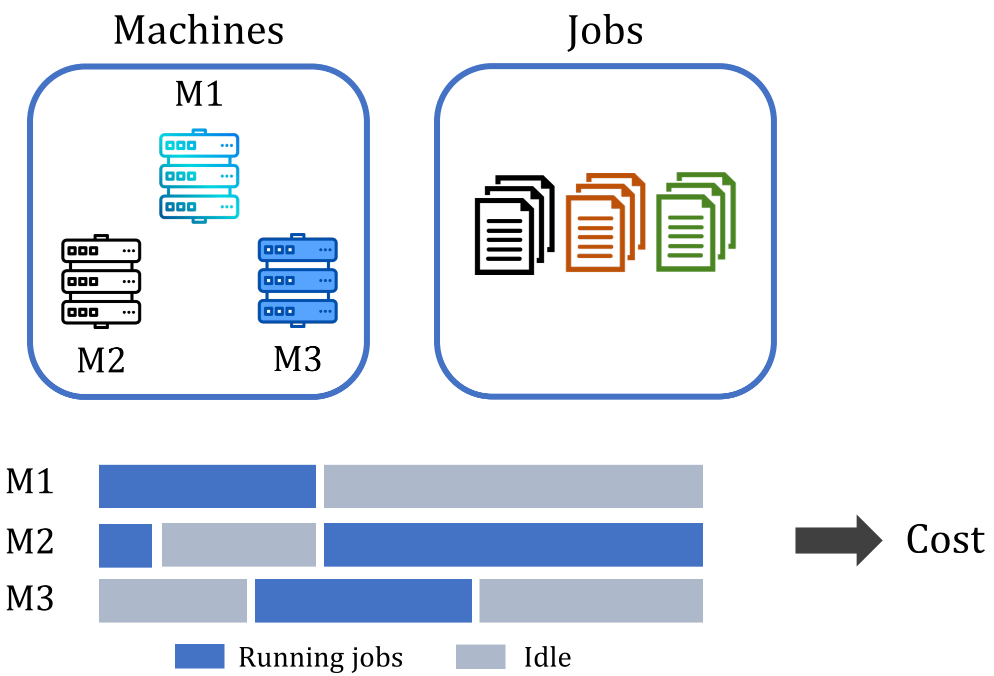
Motivation – Linear Smart "Predict, then Optimize"



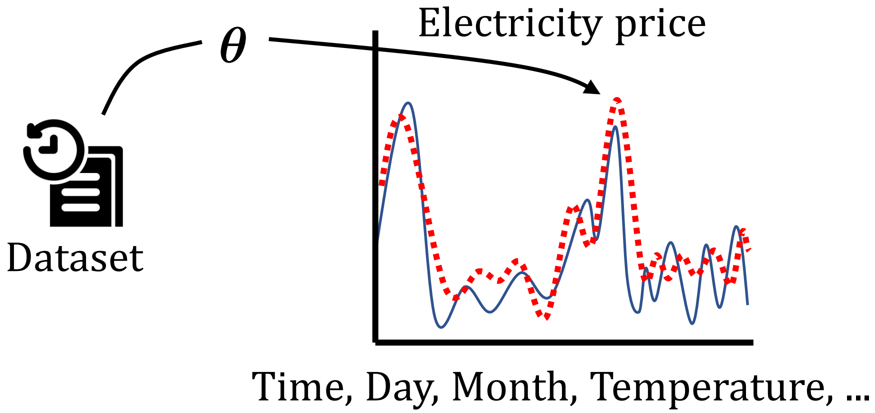


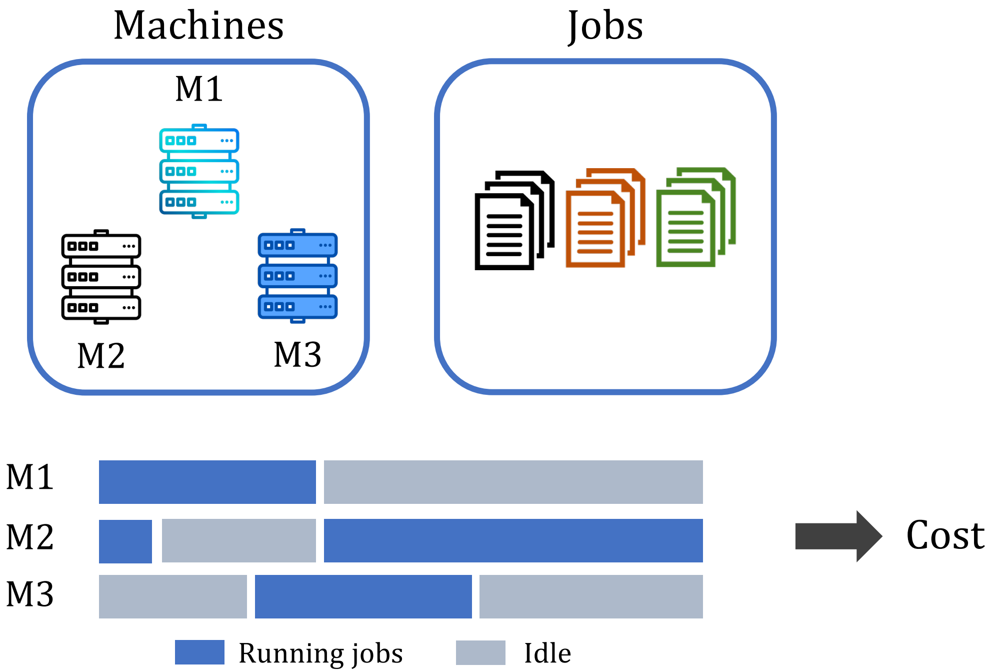
Motivation – Linear Smart "Predict, then Optimize"





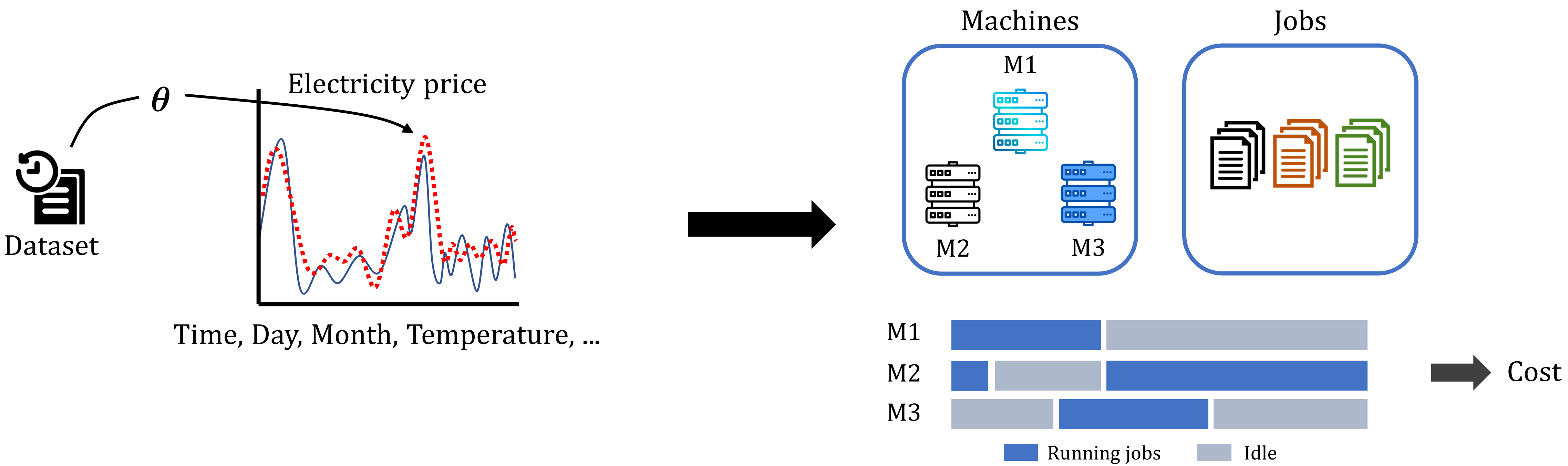
Motivation – Linear Smart “Predict, then Optimize”





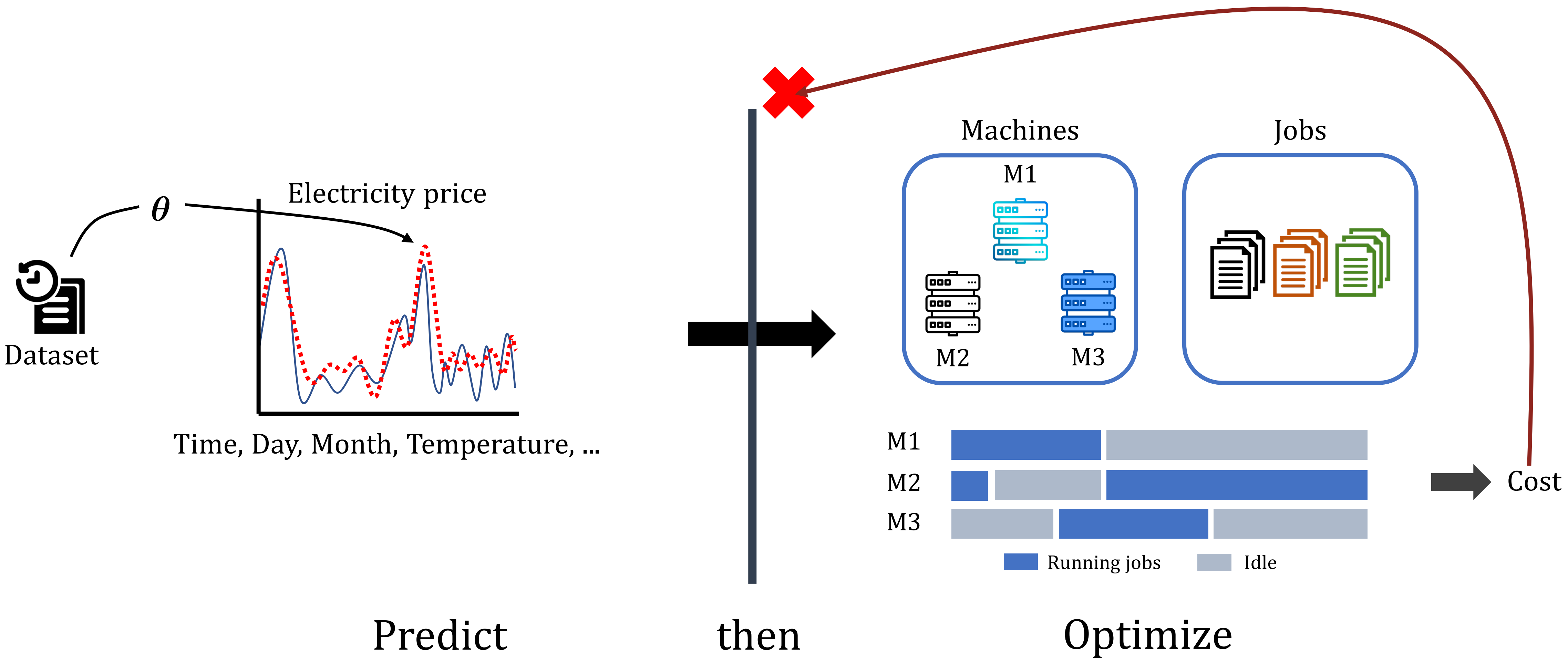
Predict

Motivation – Linear Smart “Predict, then Optimize”

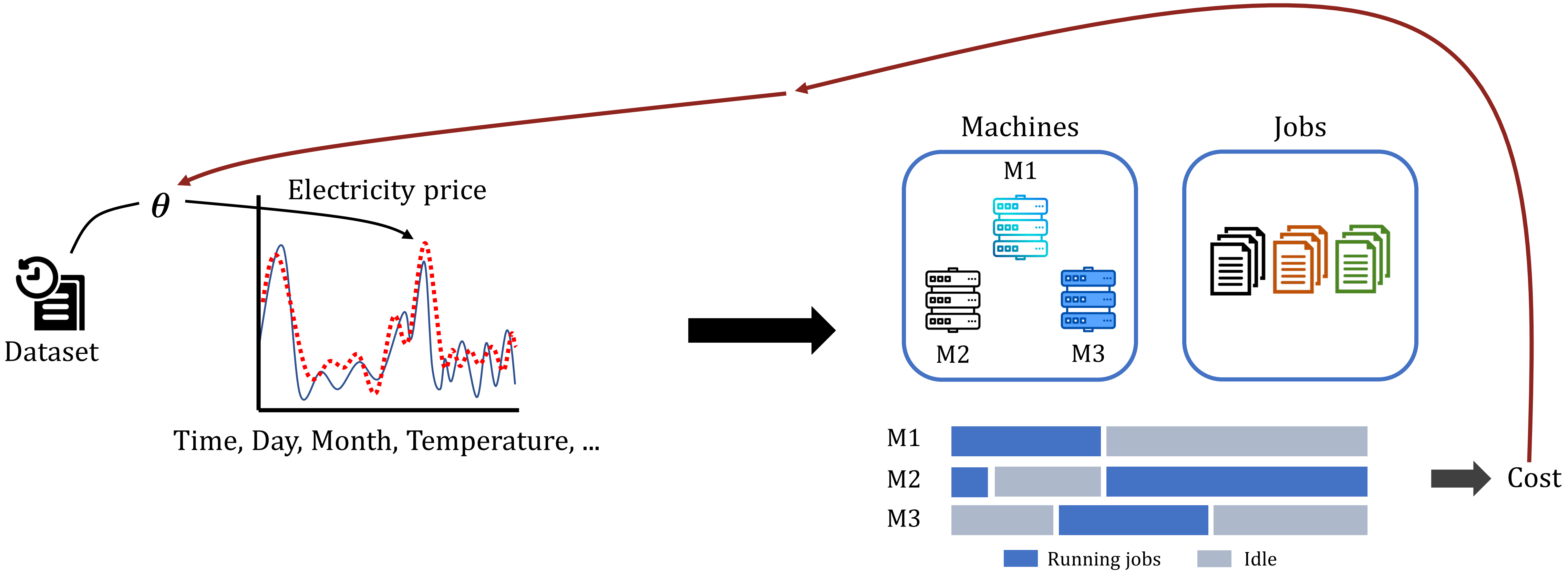


Predict then Optimize

Motivation – Linear Smart "Predict, then Optimize"



Motivation – Linear Smart "Predict, then Optimize"



Smart "Predict then Optimize" (SPO)

Motivation – Linear Smart "Predict, then Optimize"

Smart "Predict then Optimize" (SPO) Loss $\mathcal{L}_{\text{SPO}}(\boldsymbol{\theta})$

- Non-differentiable
- Non-convex

Motivation – Linear Smart "Predict, then Optimize"

Smart "Predict then Optimize" (SPO) Loss $\mathcal{L}_{\mathrm{SPO}}(\boldsymbol{\theta})$

- Non-differentiable
- Non-convex

➢ Sufficient conditions (e.g., IntOpt, QPTL)

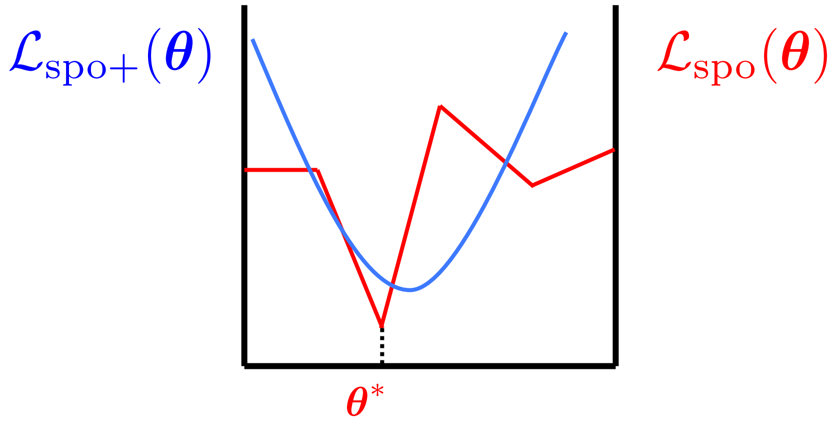
Motivation – Linear Smart "Predict, then Optimize"

Smart "Predict then Optimize" (SPO) Loss $\mathcal{L}_{\text{SPO}}(\boldsymbol{\theta})$

- Non-differentiable
- Non-convex

➢ Sufficient conditions (e.g., IntOpt, QPTL)

➢ Convex surrogate loss (e.g., SPO+)



Motivation – Linear Smart "Predict, then Optimize"

Smart "Predict then Optimize" (SPO) Loss $\mathcal{L}_{\mathrm{SPO}}(\boldsymbol{\theta})$

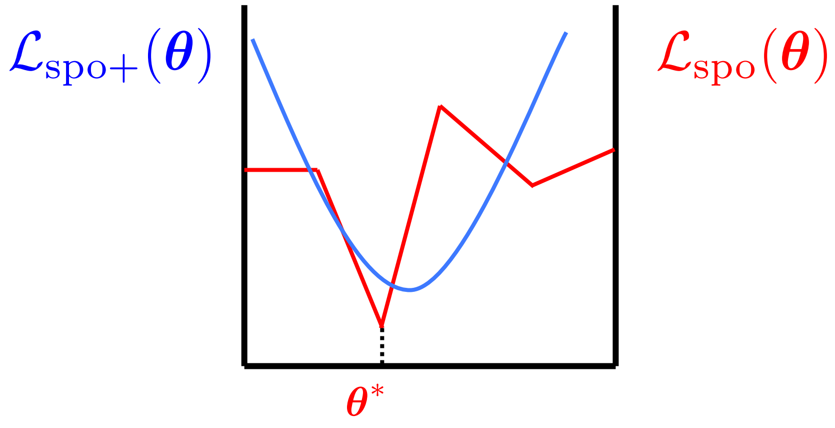
- Non-differentiable
- Non-convex

➢ Sufficient conditions (e.g., IntOpt, QPTL)

➢ Convex surrogate loss (e.g., SPO+)

⇒ None of prior methods "optimally" minimizes

⇒ How good are the existing local solvers?



Motivation – Linear Smart "Predict, then Optimize"

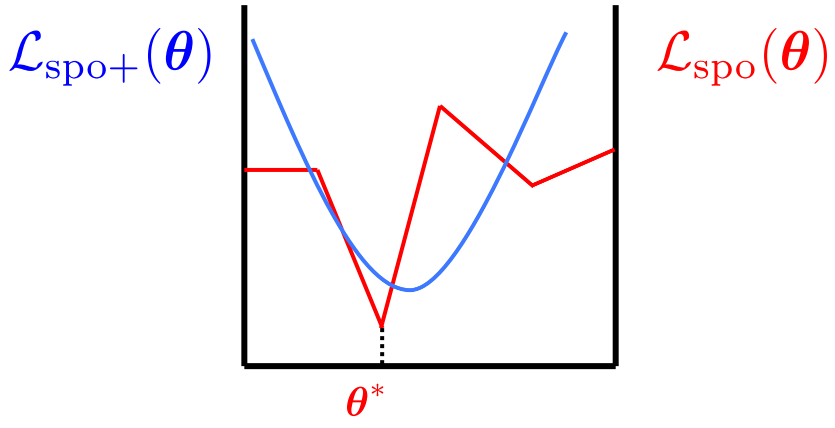
Smart "Predict then Optimize" (SPO) Loss $\mathcal{L}_{\mathrm{SPO}}(\boldsymbol{\theta})$

- Non-differentiable
- Non-convex

➢ Sufficient conditions (e.g., IntOpt, QPTL)
➢ Convex surrogate loss (e.g., SPO+)

⇒ None of prior methods "optimally" minimizes
⇒ How good are the existing local solvers?

✓ The first globally optimal solution to the linear SPO
✓ Exact reduction of linear SPO to MILP



EMSPO – Exact MILP Reduction of Linear SPO

$$\min_{\boldsymbol{\theta}} \mathcal{L}_{\text{SPO}}(\boldsymbol{\theta}) \;=\; \min_{\boldsymbol{\theta}} \quad \mathbb{E}_{(\mathbf{c},\varphi)\sim\mathcal{D}}\left[\mathbf{c}^\top \mathbf{x}^*(\bar{\mathbf{c}}) - \mathbf{c}^\top \mathbf{x}^*(\mathbf{c})\right]$$

$$\text{s.t.} \quad \mathbf{x}^*(\bar{\mathbf{c}}) \;=\; \arg\min_{\bar{\mathbf{x}}\in\mathcal{X}} \; (\underbrace{\boldsymbol{\theta}\varphi}_{\bar{\mathbf{c}}})^\top \bar{\mathbf{x}}$$

- Bi-level formulation

EMSPO – Exact MILP Reduction of Linear SPO

$$\min_{\boldsymbol{\theta}} \mathcal{L}_{\text{SPO}}(\boldsymbol{\theta}) \;=\; \min_{\boldsymbol{\theta}} \;\; \mathbb{E}_{(\mathbf{c},\varphi)\sim\mathcal{D}}\left[\mathbf{c}^\top \mathbf{x}^*(\bar{\mathbf{c}}) - \mathbf{c}^\top \mathbf{x}^*(\mathbf{c})\right]$$

$$\text{s.t.} \;\; \mathbf{x}^*(\bar{\mathbf{c}}) \;=\; \arg\min_{\bar{\mathbf{x}}\in\mathcal{X}} \big(\underbrace{\boldsymbol{\theta}\varphi}_{\bar{\mathbf{c}}}\big)^\top \bar{\mathbf{x}}$$

- Bi-level formulation

EMSPO – Exact MILP Reduction of Linear SPO

$$\min_{\boldsymbol{\theta}} \mathcal{L}_{\text{SPO}}(\boldsymbol{\theta}) \;=\; \min_{\boldsymbol{\theta}} \;\; \mathbb{E}_{(\mathbf{c},\varphi)\sim\mathcal{D}}\Big[\mathbf{c}^\top \mathbf{x}^*(\bar{\mathbf{c}}) - \mathbf{c}^\top \mathbf{x}^*(\mathbf{c})\Big]$$

$$\text{s.t.} \;\; \mathbf{x}^*(\bar{\mathbf{c}}) \;=\; \arg\min_{\bar{\mathbf{x}}\in\mathcal{X}} \;\big(\underbrace{\boldsymbol{\theta}\varphi}_{\bar{\mathbf{c}}}\big)^\top \bar{\mathbf{x}}$$

- Bi-level formulation

EMSPO – Exact MILP Reduction of Linear SPO

cost with predicted c̄

$$\min_{\boldsymbol{\theta}} \mathcal{L}_{\mathrm{SPO}}(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \mathbb{E}_{(\mathbf{c},\varphi)\sim\mathcal{D}}\left[\mathbf{c}^\top \mathbf{x}^*(\bar{\mathbf{c}}) - \mathbf{c}^\top \mathbf{x}^*(\mathbf{c})\right]$$

$\mathbf{x}^*$ with true costs

$$\text{s.t.} \quad \mathbf{x}^*(\bar{\mathbf{c}}) = \arg\min_{\bar{\mathbf{x}}\in\mathcal{X}} \big(\underbrace{\boldsymbol{\theta}\varphi}_{\bar{\mathbf{c}}}\big)^\top \bar{\mathbf{x}}$$

- Bi-level formulation
- SPO loss: decision regret compared to the oracle cost

EMSPO – Exact MILP Reduction of Linear SPO

cost with predicted $\bar{\mathrm{c}}$

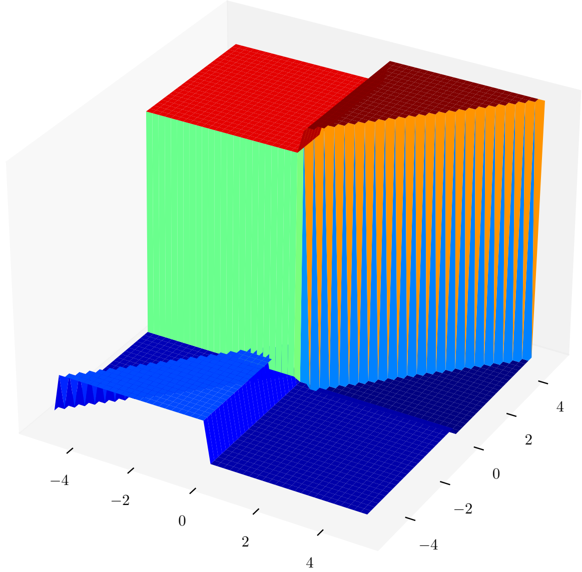
$$\min_{\boldsymbol{\theta}} \mathcal{L}_{\mathrm{SPO}}(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \mathbb{E}_{(\mathbf{c},\varphi)\sim\mathcal{D}}\left[\mathbf{c}^\top \mathbf{x}^*(\bar{\mathbf{c}}) - \mathbf{c}^\top \mathbf{x}^*(\mathbf{c})\right]$$

$\mathbf{x}^*$ with true costs

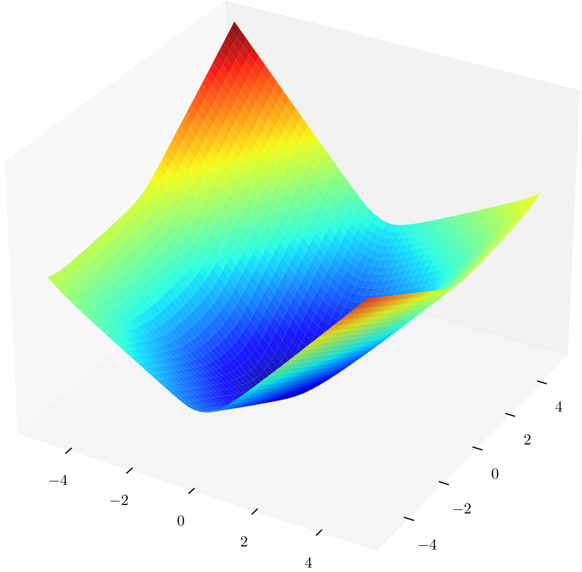
$$\text{s.t.} \quad \mathbf{x}^*(\bar{\mathbf{c}}) = \arg\min_{\bar{\mathbf{x}}\in\mathcal{X}} \big(\underbrace{\boldsymbol{\theta}\varphi}_{\bar{\mathbf{c}}}\big)^\top \bar{\mathbf{x}}$$

- Bi-level formulation

Piece-wise linear surface
⇒ MILP



True SPO loss



Surrogate SPO+ loss

EMSPO – Exact MILP Reduction of Linear SPO

cost with predicted $\bar{c}$

$$\min_{\boldsymbol{\theta}} \mathcal{L}_{\text{SPO}}(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \mathbb{E}_{(\mathbf{c},\varphi)\sim\mathcal{D}}\left[\mathbf{c}^\top \mathbf{x}^*(\bar{\mathbf{c}}) - \mathbf{c}^\top \mathbf{x}^*(\mathbf{c})\right]$$

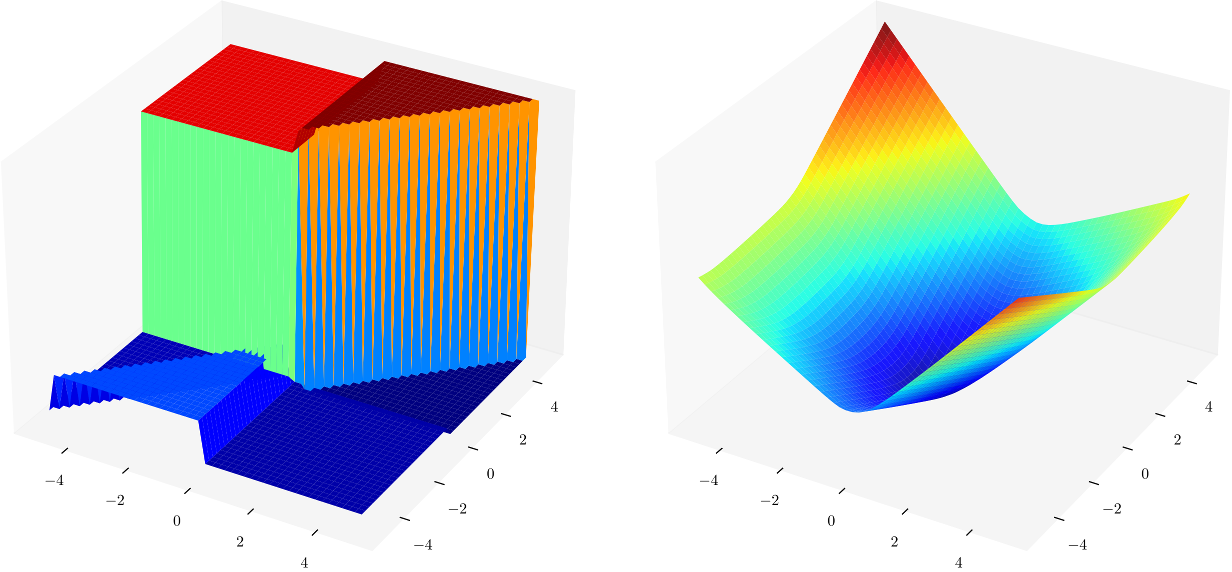
$\mathbf{x}^*$ with true costs

$$\text{s.t.} \quad \mathbf{x}^*(\bar{\mathbf{c}}) = \arg\min_{\bar{\mathbf{x}}\in\mathcal{X}} (\underbrace{\boldsymbol{\theta}\varphi}_{\bar{\mathbf{c}}})^\top \bar{\mathbf{x}}$$

- Bi-level formulation

Piece-wise linear surface
⇒ MILP

Linear SPO ⇒ MILP



True SPO loss   Surrogate SPO+ loss

EMSPO – Exact MILP Reduction of Linear SPO

SPO loss in bi-level form

$$\min_{\boldsymbol{\theta}} \mathcal{L}_{\mathrm{SPO}}(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \ \mathbb{E}_{(\mathbf{c},\varphi)\sim\mathcal{D}}\left[\mathbf{c}^\top \mathbf{x}^*(\bar{\mathbf{c}}) - \mathbf{c}^\top \mathbf{x}^*(\mathbf{c})\right]$$

$$\text{s.t.} \ \ \mathbf{x}^*(\bar{\mathbf{c}}) = \arg\min_{\bar{\mathbf{x}}\in\mathcal{X}} \ \bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

EMSPO – Exact MILP Reduction of Linear SPO

SPO loss in bi-level form

$\Downarrow$

SPO loss in single-level form

$$\min_{\boldsymbol{\theta}} \quad \mathbb{E}_{(\mathbf{c},\varphi)\sim\mathcal{D}}\left[\mathbf{c}^\top \operatorname*{arg\,min}_{\bar{\mathbf{x}}\in\mathcal{X}} \ \bar{\mathbf{c}}^\top \bar{\mathbf{x}} - \mathbf{c}^\top \mathbf{x}^*(\mathbf{c})\right]$$

EMSPO – Exact MILP Reduction of Linear SPO

$$\mathbf{x}^*(\bar{\mathbf{c}}) \; = \; \arg\min_{\bar{\mathbf{x}} \in \mathcal{X}} \; \bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1, \; \bar{x}_2, \; \ldots$ sequentially with annotations

EMSPO – Exact MILP Reduction of Linear SPO

$$\mathbf{x}^*(\bar{\mathbf{c}}) = \arg\min_{\bar{\mathbf{x}} \in \mathcal{X}} \bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

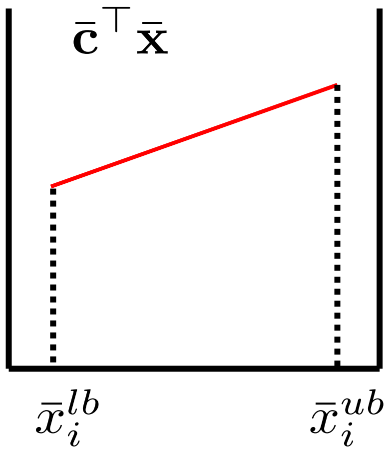
- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1, \ \bar{x}_2, \ \ldots$ sequentially with annotations

$\bar{\mathbf{c}}^\top \bar{\mathbf{x}}$

EMSPO – Exact MILP Reduction of Linear SPO

$$\mathbf{x}^*(\bar{\mathbf{c}}) = \underset{\bar{\mathbf{x}} \in \mathcal{X}}{\arg\min}\ \bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1,\ \bar{x}_2,\ \ldots$ sequentially with annotations

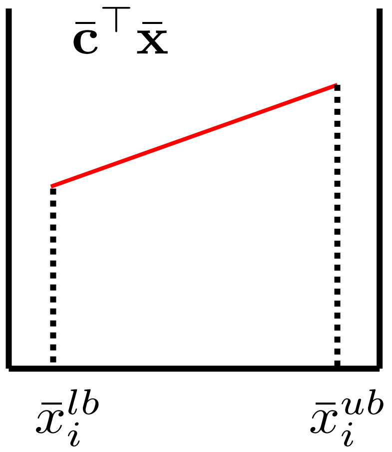


EMSPO – Exact MILP Reduction of Linear SPO

$$\mathbf{x}^*(\bar{\mathbf{c}}) = \underset{\bar{\mathbf{x}} \in \mathcal{X}}{\arg\min} \; \bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1, \; \bar{x}_2, \; \ldots$ sequentially with annotations

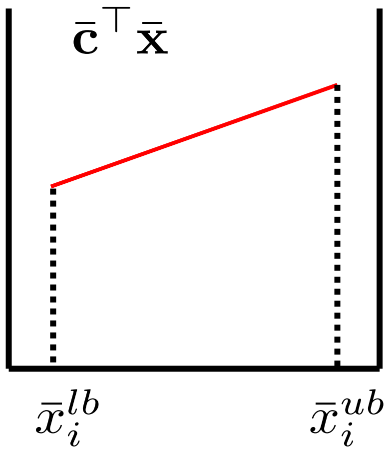
$$\bar{\mathbf{c}}^\top \bar{\mathbf{x}} = \bar{c}_i \bar{x}_i + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$



EMSPO – Exact MILP Reduction of Linear SPO

$$\mathbf{x}^*(\bar{\mathbf{c}}) = \underset{\bar{\mathbf{x}} \in \mathcal{X}}{\arg\min}\ \bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1,\ \bar{x}_2,\ \ldots$ sequentially with annotations



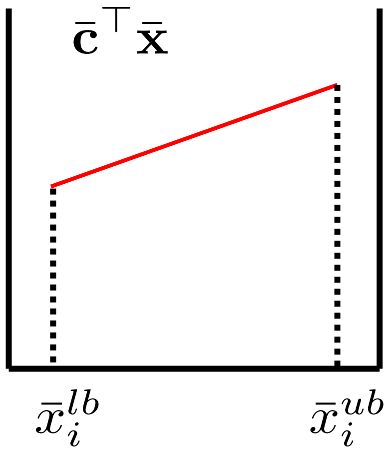
$$\bar{\mathbf{c}}^\top \bar{\mathbf{x}} = \bar{c}_i \bar{x}_i + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

$$\bar{c}_i \bar{x}_i^{lb} + \sum_{j \neq i} \bar{c}_j \bar{x}_j \qquad \bar{c}_i \bar{x}_i^{ub} + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

EMSPO – Exact MILP Reduction of Linear SPO

$$\mathbf{x}^*(\bar{\mathbf{c}}) \;=\; \underset{\bar{\mathbf{x}} \in \mathcal{X}}{\arg\min}\; \bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1,\ \bar{x}_2,\ \ldots$ sequentially with annotations



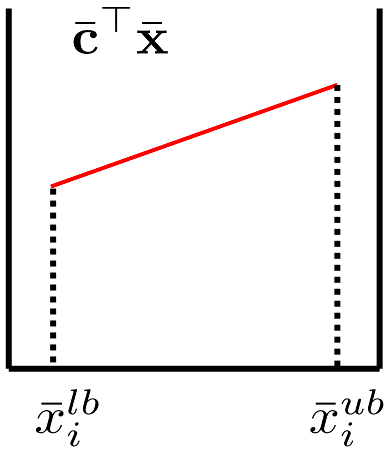
$$\bar{\mathbf{c}}^\top \bar{\mathbf{x}} = \bar{c}_i \bar{x}_i + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

$$\bar{c}_i \bar{x}_i^{lb} + \sum_{j \neq i} \bar{c}_j \bar{x}_j \qquad\qquad \bar{c}_i \bar{x}_i^{ub} + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

EMSPO – Exact MILP Reduction of Linear SPO

$$\mathbf{x}^*(\bar{\mathbf{c}}) = \underset{\bar{\mathbf{x}} \in \mathcal{X}}{\arg\min}\ \bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1,\ \bar{x}_2,\ \ldots$ sequentially with annotations



$$\bar{\mathbf{c}}^\top \bar{\mathbf{x}} = \bar{c}_i \bar{x}_i + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

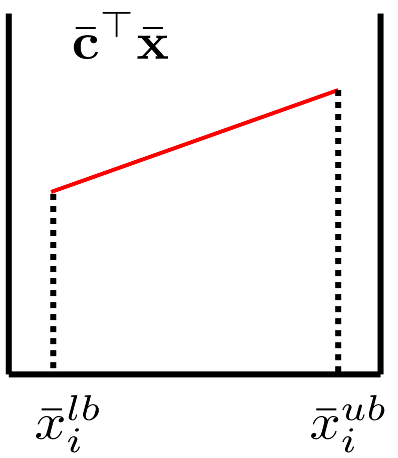
$$\bar{c}_i \bar{x}_i^{lb} + \sum_{j \neq i} \bar{c}_j \bar{x}_j \qquad \bar{c}_i \bar{x}_i^{ub} + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

$$\min\left(\bar{c}_i \bar{x}_i^{lb} \quad , \quad \bar{c}_i \bar{x}_i^{ub} \right)$$

EMSPO – Exact MILP Reduction of Linear SPO

$$\mathbf{x}^*(\bar{\mathbf{c}}) = \arg\min_{\bar{\mathbf{x}} \in \mathcal{X}} \bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1, \ \bar{x}_2, \ \ldots$ sequentially with annotations



$$\bar{\mathbf{c}}^\top \bar{\mathbf{x}} = \bar{c}_i \bar{x}_i + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

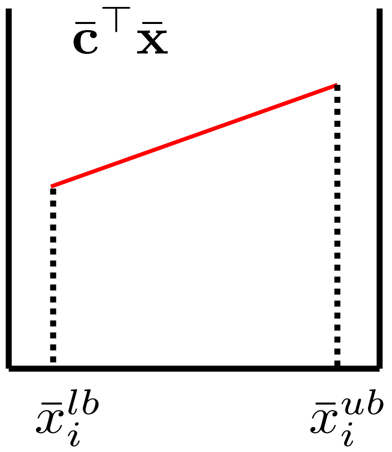
$$\bar{c}_i \bar{x}_i^{lb} + \sum_{j \neq i} \bar{c}_j \bar{x}_j \qquad \bar{c}_i \bar{x}_i^{ub} + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

$$\min \left(\bar{c}_i \bar{x}_i^{lb} \ , \ \bar{c}_i \bar{x}_i^{ub} \right) \longrightarrow \begin{cases} \bar{c}_i \leq 0 : \ \bar{c}_i \bar{x}_i^{ub} \\ \bar{c}_i \geq 0 : \ \bar{c}_i \bar{x}_i^{lb} \end{cases}$$

EMSPO – Exact MILP Reduction of Linear SPO

$$\mathbf{x}^*(\bar{\mathbf{c}}) \;=\; \arg\min_{\bar{\mathbf{x}} \in \mathcal{X}} \;\bar{\mathbf{c}}^\top \bar{\mathbf{x}}$$

- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1,\ \bar{x}_2,\ \ldots$ sequentially with annotations



$$\bar{\mathbf{c}}^\top \bar{\mathbf{x}} = \bar{c}_i \bar{x}_i + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

$$\bar{c}_i \bar{x}_i^{lb} + \sum_{j \neq i} \bar{c}_j \bar{x}_j \qquad\qquad \bar{c}_i \bar{x}_i^{ub} + \sum_{j \neq i} \bar{c}_j \bar{x}_j$$

$$\arg\min_{\bar{x}_i} \;\bar{\mathbf{c}}^\top \bar{\mathbf{x}} \;\rightarrow\; \begin{cases} \bar{c}_i \leq 0 : & \bar{x}_i^{ub} \\ \bar{c}_i \geq 0 : & \bar{x}_i^{lb} \end{cases}$$

EMSPO – Exact MILP Reduction of Linear SPO

$$\underset{\bar{x}_i}{\arg\min} \quad \bar{\mathbf{c}}^\top \bar{\mathbf{x}} \longrightarrow \begin{cases} \bar{c}_i \leq 0 : & \bar{x}_i^{ub} \\ \bar{c}_i \geq 0 : & \bar{x}_i^{lb} \end{cases}$$

Can be generalized... See our paper ☺

- Symbolic variable elimination (SVE) of $\bar{\mathbf{x}}$
- *Eliminate* $\bar{x}_1, \ \bar{x}_2, \ \ldots$ sequentially with annotations
- To generalize
  - Need efficient data structure since cases will blow up quickly

"Bi-level SPO Loss" ⇒ "Mixed-Integer Linear Program"

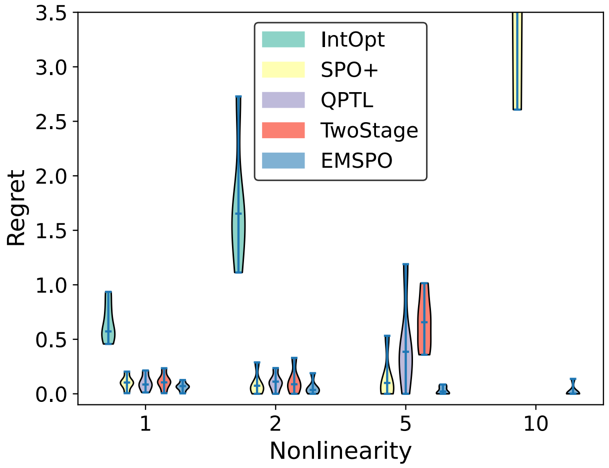
Experiments

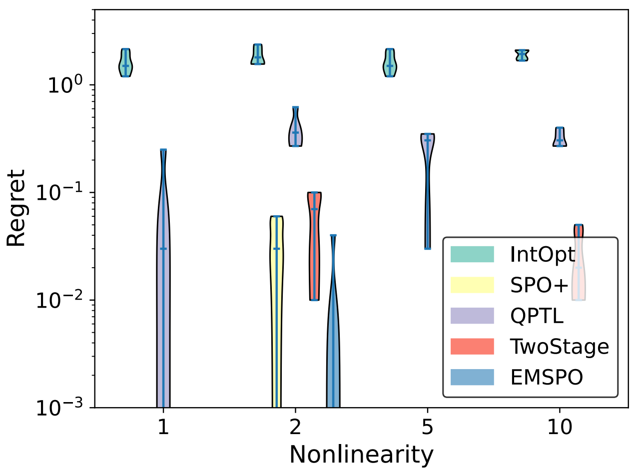
- Metric
  - Decision regret
- Compared methods
  - SPO+, IntOpt, QPTL, Two-stage

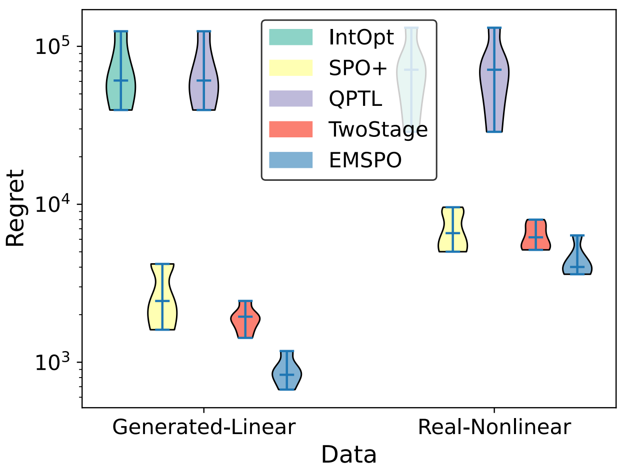
Experiments

- Metric
  - Decision regret
- Compared methods
  - SPO+, IntOpt, QPTL, Two-stage
- Nonlinearity
- Dataset size

Experiments



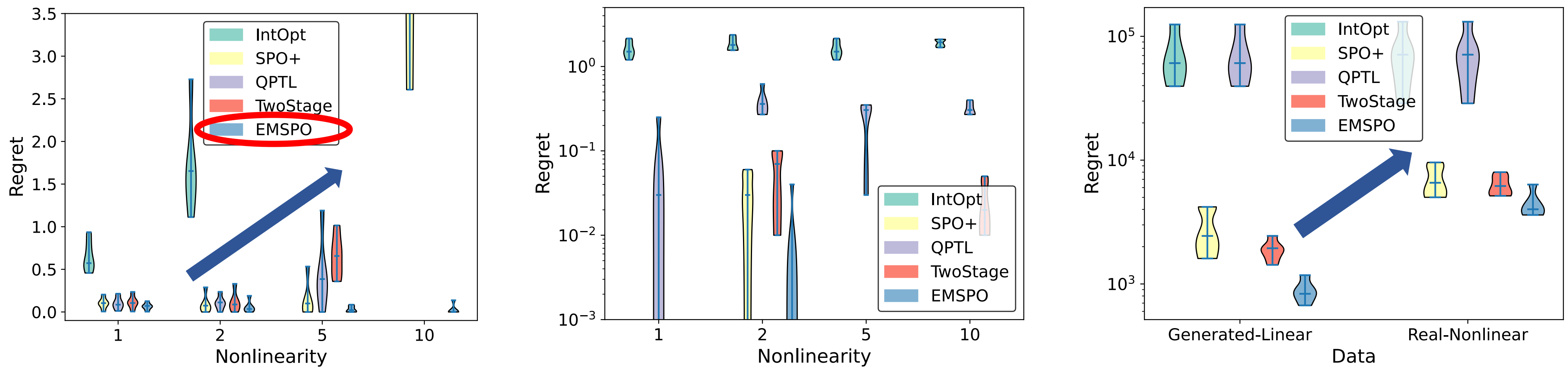




Nonlinearity ↑

- Decision regret vs. nonlinearity in the data

Experiments

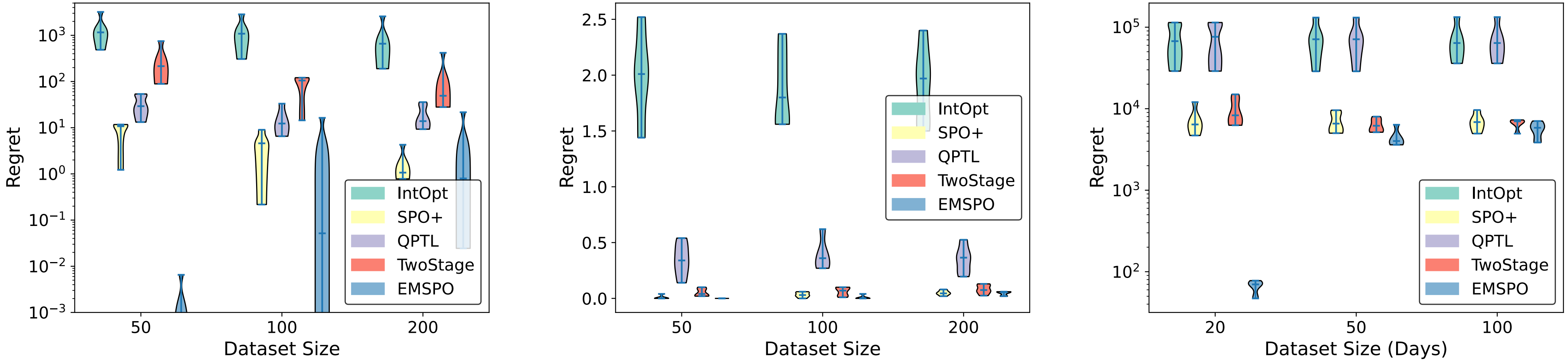


Nonlinearity ↑

- Decision regret vs. nonlinearity in the data

Nonlinearity ↑ ⇒ Decision regret ↑

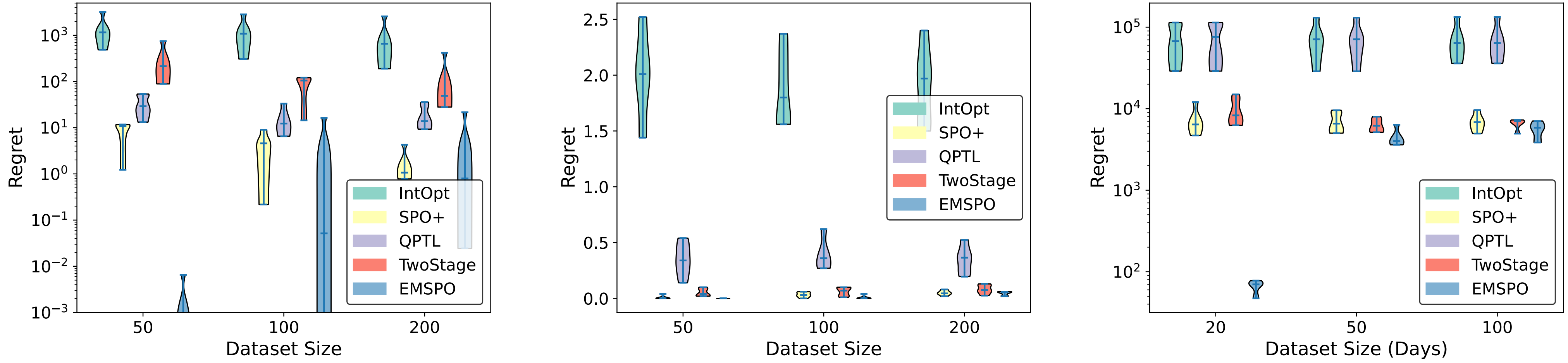
Experiments



Dataset size ↑

- Decision regret vs. Dataset size

Experiments



Dataset size ↑

- Decision regret vs. Dataset size

Dataset size ↑ ⇒ Decision regret ↓

SPO+ performs robustly across the board

Conclusion

- First globally optimal solver of linear SPO problems subject to mild assumptions
  - Symbolic bilinear argmin solver that solves the lower optimization
  - Overall problem reduces to MILP

Conclusion

- First globally optimal solver of linear SPO problems subject to mild assumptions
  - Symbolic bilinear argmin solver that solves the lower optimization
  - Overall problem reduces to MILP
- Benchmarked existing approximate solvers
  - Found that there is big room for improvement