Parametric multistage stochastic optimization for day-ahead power scheduling

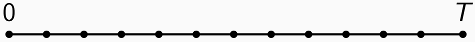
PGMO days 2021 - December 1rst, 2021

Adrien Le Franc, Pierre Carpentier, Jean-Philippe Chancelier, Michel De Lara

École des Ponts ParisTech

A typical power scheduling example

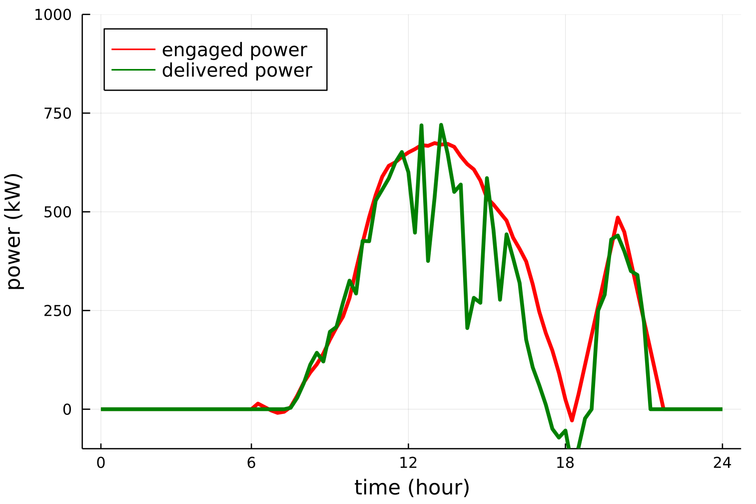
- We operate a solar plant over **one day** with discrete time steps $t \in \{0, 1, \dots, T\}$



- For every operating day
  - In the **day-ahead** stage, we must supply a power production profile $p \in \mathbb{R}^T$
  - In the **intraday** stage, we manage the power plant and deliver a power profile $\tilde{p} \in \mathbb{R}^T$

Engaged power vs delivered power

The delivered power $\tilde{p}$ induces gains
and differences between $\tilde{p}$ and p induce penalties



Contribution of the talk

- **Question**

How can we optimize **day-ahead and intraday decisions**
for operating a solar plant with **uncertain generated power**
at **least expected cost** ?

- **Our contribution**

We introduce
parametric multistage stochastic optimization problems
for day-ahead power scheduling
and study **differentiability properties**
of **parametric value functions**

Outline of the presentation

1. **Parametric multistage stochastic optimization**
2. **Differentiability of parametric value functions**
3. **Numerical example**
4. **Conclusion and Perspectives**

Outline of the presentation

1. **Parametric multistage stochastic optimization**
2. Differentiability of parametric value functions
3. Numerical example
4. Conclusion and Perspectives

Our standard formulation

We consider a **multistage stochastic optimization problem** parametrized by $p \in \mathbb{R}^{n_p \times (T+1)}$ written in standard form as

$$\Phi(p) = \inf_{\mathbf{U}_0, \dots, \mathbf{U}_{T-1}} \mathbb{E}\Big[\sum_{t=0}^{T-1} L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}, p_t) + K(\mathbf{X}_T, p_T)\Big]$$

$$\mathbf{X}_0 = x_0$$

$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) \; , \;\; \forall t \in \{0, \dots, T-1\}$$

$$\mathbf{U}_t \in \mathcal{U}_t(\mathbf{X}_t, p_t) \; , \;\; \forall t \in \{0, \dots, T-1\}$$

$$\sigma(\mathbf{U}_t) \subseteq \sigma(\mathbf{W}_1, \dots, \mathbf{W}_t) \; , \;\; \forall t \in \{0, \dots, T-1\}$$

where $\mathbf{X}_t : \Omega \to \mathbb{R}^{n_x}$, $\mathbf{U}_t : \Omega \to \mathbb{R}^{n_u}$, $\mathbf{W}_t : \Omega \to \mathbb{R}^{n_w}$

Optimal solution via stochastic dynamic programming

Assumption (discrete white noise)

The sequence $\{\mathbf{W}_t\}_{t\in\{1,\dots,T\}}$ *is* **stagewise independent**,
and each noise variable $\mathbf{W}_t$ *has a* **finite support**

For $t \in \{0, \dots, T\}$ and $x \in \mathbb{R}^{n_x}$
we define the **parametric value functions**

$$V_T(x, p) = K(x, p)$$

$$V_t(x, p) = \inf_{u \in \mathcal{U}_t(x, p_t)} \mathbb{E}\Big[L_t(x, u, \mathbf{W}_{t+1}, p_t) + V_{t+1}\big(f_t(x, u, \mathbf{W}_{t+1}), p\big)\Big]$$

Under the discrete white noise assumption $\quad \Phi(p) = V_0(x_0, p)$

Assumption

Assumption (convex multistage problem)

1. *the cost functions* $\{L_t\}_{t \in \{0,\dots,T-1\}}$ *are* **jointly convex and lsc** *w.r.t.* (x_t, u_t, p_t), *and are* **proper**, *and the final cost* K *is* **convex, proper, lsc**
2. *the dynamics* $\{f_t\}_{t \in \{0,\dots,T-1\}}$ *are* **affine** *w.r.t.* (x_t, u_t)
3. *the set-valued mappings* $\{\mathcal{U}_t\}_{t \in \{0,\dots,T-1\}}$ *are* **closed, convex**, *have* **nonempty domains** *and* **compact ranges**
4. *the problem satisfies a* **relatively complete recourse-like assumption**

Convexity of parametric value functions

Proposition (Le Franc [2021])

Under the **discrete white noise** *assumption*
and the **convex multistage problem** *assumption,*
the parametric value functions $\{V_t\}_{t\in\{0,\ldots,T\}}$ *are* **convex, proper, lsc** *w.r.t.* (x, p)

Outline of the presentation

1. Parametric multistage stochastic optimization
2. **Differentiability of parametric value functions**
3. Numerical example
4. Conclusion and Perspectives

Outline of the section

2. **Differentiability of parametric value functions**

   Smooth parametric value functions

   Lower smooth approximations

   Lower polyhedral approximations

Assumption

Assumption (smoothness)

1. *the cost functions* $\{L_t\}_{t \in \{0,\dots,T-1\}}$ *and* K *are* **differentiable w.r.t.** p_t
2. *for all* $t \in \{0, \dots, T-1\}$*, the set-valued mapping* $\mathcal{U}_t$ **takes the same set value for all** $p_t \in \mathbb{R}^{n_p}$*; in that case, we use the notation* $\mathcal{U}_t(x)$ *instead of* $\mathcal{U}_t(x, p_t)$

Differentiable parametric value functions

Theorem (Le Franc [2021])

Under the **discrete white noise** *assumption,*
the **convex multistage problem** *assumption,*
and the **smoothness** *assumption,*
the value functions $\{V_t\}_{t \in \{0,\dots,T\}}$ *are* **differentiable w.r.t.** p*,*
and their gradients may be computed by backward induction, with

$$\nabla_p V_T(x, p) = \nabla_p K(x, p_T) \ , \ \ \forall (x, p) \in \mathrm{dom}(V_T)$$

and at stage $t \in \{0, \dots, T-1\}$*, for* $(x, p) \in \mathrm{dom}(V_t)$*,*
the solution set $\mathcal{U}_t^*(x, p_t)$ *is nonempty, and for any* $u^* \in \mathcal{U}_t^*(x, p_t)$*,*

$$\nabla_p V_t(x, p) = \mathbb{E}\Big[\nabla_p L_t(x, u^*, \mathbf{W}_{t+1}, p_t) + \nabla_p V_{t+1}\big(f_t(x, u^*, \mathbf{W}_{t+1}), p\big)\Big]$$

Outline of the section

2. Differentiability of parametric value functions

Smooth parametric value functions

Lower smooth approximations

Lower polyhedral approximations

Assumption

We consider a **parameter set** $\mathcal{P} \subseteq \mathbb{R}^{n_p \times (T+1)}$
and define $\mathcal{P}_t = \mathrm{proj}_t(\mathcal{P}) \subseteq \mathbb{R}^{n_p}$, $\forall t \in \{0, \ldots, T\}$

Assumption (parameter set)

1. *the parameter set* $\mathcal{P}$ *is* **nonempty, convex and compact**
2. *for all* $t \in \{0, \ldots, T-1\}$,
   the domain of the set-valued mapping $\mathcal{U}_t$
   is such that $\mathrm{dom}(\mathcal{U}_t) \subseteq \mathbb{R}^{n_x} \times \mathcal{P}_t$

Moreau envelopes of cost functions

Given values $(x, u, w) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_w}$
and a **regularization parameter** $\mu \in \mathbb{R}_+^*$, we introduce

$$L_t^\mu(x, u, w, p_t) = \inf_{p'_t \in \mathbb{R}^{n_p}} \Big(L_t(x, u, w, p'_t) + \delta_{\mathrm{gr}(\mathcal{U}_t)}(x, u, p'_t) + \delta_{\mathcal{P}_t}(p'_t) + \frac{1}{2\mu} || p_t - p'_t ||_2^2 \Big) \ , \ \ \forall t \in \{0, \dots, T-1\} \ , \ \ \forall p_t \in \mathbb{R}^{n_p}$$

$$K^\mu(x, p_T) = \inf_{p'_T \in \mathbb{R}^{n_p}} \Big(K(x, p'_T) + \delta_{\mathcal{P}_T}(p'_T) + \frac{1}{2\mu} || p_T - p'_T ||_2^2 \Big) \ , \ \ \forall p_T \in \mathbb{R}^{n_p}$$

Lower smooth parametric value functions

$$\underset{\sim}{V}_T^{\mu}(x, p) = K^{\mu}(x, p_T) \ , \ \ \forall (x, p) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_p \times (T+1)}$$

$$\underset{\sim}{V}_t^{\mu}(x, p) = \inf_{u \in \mathrm{range}(\mathcal{U}_t)} \mathbb{E}\Big[L_t^{\mu}(x, u, \mathbf{W}_{t+1}, p_t) + \underset{\sim}{V}_{t+1}^{\mu}\big(f_t(x, u, \mathbf{W}_{t+1}), p\big) \Big]$$

$$\forall (x, p) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_p \times (T+1)} \ , \ \ \forall t \in \{0, \ldots, T-1\}$$

Proposition (Le Franc [2021])

Under the **discrete white noise** *assumption,*
the **convex multistage problem** *assumption,*
and the **parameter set** *assumption,*
the lower smooth parametric value functions $\{\underset{\sim}{V}_t^{\mu}\}_{t \in \{0, \ldots, T\}}$
are **differentiable w.r.t.** p*, and their gradients may be computed by backward induction*

Convergence result

$$\Phi^* = \inf_{p \in \mathcal{P}} \Phi(p)$$

Proposition (Le Franc [2021])

Under the same assumptions, if the sequence of regularization parameters $\{\mu_n\}_{n\in\mathbb{N}} \in (\mathbb{R}_+^*)^{\mathbb{N}}$ *is nonincreasing and such that* $\lim_{n\to+\infty} \mu_n = 0$*, then for any initial state* $x_0 \in \mathbb{R}^{n_x}$*, we have that*

$$\inf_{p\in\mathcal{P}} \underset{\sim}{V}_0^{\mu_n}(x_0, p) \leq \Phi^* \;, \;\; \forall n \in \mathbb{N} \;, \;\; \textit{and} \quad \inf_{p\in\mathcal{P}} \underset{\sim}{V}_0^{\mu_n}(x_0, p) \xrightarrow[n\to+\infty]{} \Phi^*$$

Outline of the section

State extension

$$x_t^\sharp = \begin{pmatrix} x_t \\ p \end{pmatrix} \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_p \times (T+1)} \;, \;\; \forall t \in \{0, \dots, T\}$$

$$\Phi(p) = \inf_{\mathbf{U}_0, \dots, \mathbf{U}_{T-1}} \mathbb{E}\Big[\sum_{t=0}^{T-1} \; L_t^\sharp(\mathbf{X}^\sharp{}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) + K^\sharp(\mathbf{X}^\sharp{}_T) \Big]$$

$$\mathbf{X}^\sharp{}_0 = \begin{pmatrix} x_0 \\ p \end{pmatrix}$$

$$\mathbf{X}^\sharp{}_{t+1} = f_t^\sharp(\mathbf{X}^\sharp{}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) \;, \;\; \forall t \in \{0, \dots T-1\}$$

$$\mathbf{U}_t \in \mathcal{U}_t^\sharp(\mathbf{X}^\sharp{}_t) \;, \;\; \forall t \in \{0, \dots, T-1\}$$

$$\sigma(\mathbf{U}_t) \subseteq \sigma(\mathbf{W}_1, \dots, \mathbf{W}_t) \;, \;\; \forall t \in \{0, \dots, T-1\}$$

Lower polyhedral value functions

- We introduce the state value functions

$$V_T^\sharp(x^\sharp) = K^\sharp(x^\sharp)\ , \ \ \forall x^\sharp \in \left(\mathbb{R}^{n_x} \times \mathbb{R}^{n_p \times (T+1)}\right)$$

$$V_t^\sharp(x^\sharp) = \inf_{u \in \mathcal{U}_t^\sharp(x^\sharp)} \mathbb{E}\Big[L_t^\sharp(x^\sharp, u, \mathbf{W}_{t+1}) + V_{t+1}^\sharp\big(f_t^\sharp(x^\sharp, u, \mathbf{W}_{t+1})\big)\Big]$$

$$\forall x^\sharp \in \left(\mathbb{R}^{n_x} \times \mathbb{R}^{n_p \times (T+1)}\right)\ , \ \ \forall t \in \{0, \dots, T-1\}$$

- We compute **polyhedral lower approximations** $\{\underline{V}_t^k\}_{t \in \{0,\dots,T\}}$ of $\{V_t^\sharp\}_{t \in \{0,\dots,T\}}$ by running $k \in \mathbb{N}$ forward-backward passes of the **SDDP algorithm**

- Since $\underline{V}_0^k$ is polyhedral, **linear programming** gives us a **subgradient** $(y, q) \in \partial \underline{V}_0^k\big((x_0, p)\big)$

Convergence result

Proposition (Le Franc [2021])

Let $(x_0, p) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_p \times (T+1)}$, *if after* $\bar{k} \in \mathbb{N}^*$ *forward-backward passes of the SDDP algorithm the approximation error of the value function* $V_0^\sharp$ *by the lower polyhedral approximation* $\underline{V}_0^{\bar{k}}$ *is bounded by*

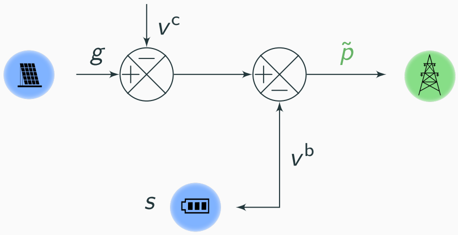
$$V_0^\sharp\big((x_0, p)\big) - \underline{V}_0^{\bar{k}}\big((x_0, p)\big) \leq \varepsilon$$

for some $\varepsilon \in \mathbb{R}_+$, *then if we compute*

$$\begin{cases} \phi = \underline{V}_0^{\bar{k}}\big((x_0, p)\big) \\ (y, q) \in \partial \underline{V}_0^{\bar{k}}\big((x_0, p)\big) \end{cases} \quad \textit{we have that} \quad \begin{cases} |\Phi(p) - \phi| \leq \varepsilon \\ q \in \partial_\varepsilon \Phi(p) \end{cases}$$

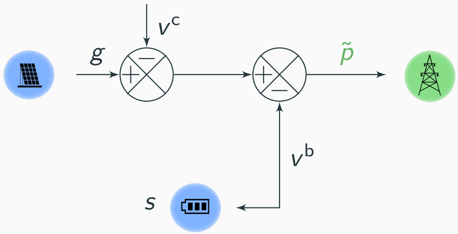
Outline of the presentation

Schematic organization of the solar plant



- $g \in [0, \bar{p}]^T$ generated power (uncertainty)
- $v^c \in [0, g]^T$ curtailed power (control)
- $s \in [0, \bar{s}]^{T+1}$ state of charge (state)
- $v^b \in [\underline{v}, \overline{v}]^T$ battery power (control)
- $\tilde{p} = g - v_b - v_c$ **delivered power**

Schematic organization of the solar plant



- $g \in [0, \bar{p}]^T$ generated power (uncertainty) $\rightarrow$ AR(1) process
- $v^c \in [0, g]^T$ curtailed power (control)
- $s \in [0, \bar{s}]^{T+1}$ state of charge (state)
- $v^b \in [\underline{v}, \overline{v}]^T$ battery power (control)
- $\tilde{p} = g - v_b - v_c$ **delivered power**

Stochastic optimal control framework

- We introduce the the **state**, **control** and **noise** variables

$$x = \begin{pmatrix} s \\ g \end{pmatrix} \ , \ \ u = \begin{pmatrix} v^{\mathrm{b}} \\ v^{\mathrm{c}} \end{pmatrix} \ , \ \ w = \epsilon$$

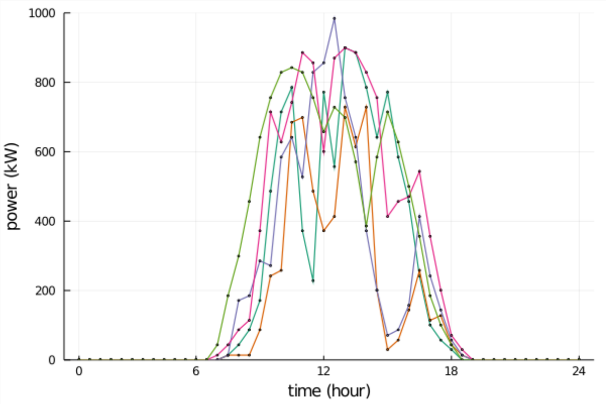
- The state process $\mathbf{X}$ is ruled by the dynamics

$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) = \begin{pmatrix} \mathbf{S}_t + \rho_c {\mathbf{V}_t^{\mathrm{b}}}^+ - \frac{1}{\rho_d} {\mathbf{V}_t^{\mathrm{b}}}^- \\ \alpha_t \mathbf{G}_t + \beta_t + \boldsymbol{\epsilon}_{t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{t+1} \\ \mathbf{G}_{t+1} \end{pmatrix}$$

- The stage costs formulate as

$$L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}, {\color{red} p_t}) = \underbrace{-c_t \tilde{\mathbf{P}}_{t+1}}_{\text{delivery gain}} + \underbrace{\lambda c_t |\tilde{\mathbf{P}}_{t+1} - {\color{red} p_t}|}_{\text{penalty}}$$

Scenarios



We use one year of power data from Ausgrid
to calibrate the weights (α_t, β_t) and the law of $\boldsymbol{\epsilon}_{t+1}$
for the generated power $\mathbf{G}_t$

Methods to compute an optimal profile $p^* \in \mathbb{R}^T$

We want to compute $\quad p^* \in \arg\min_{p \in \mathcal{P}} \Phi(p)$

Generic method

---

input: $p^0 \in \mathcal{P}$

for $k = 1 \ldots K$ **do**

- call a **a first order oracle** to estimate
  - $\rightarrow \Phi(p^k)$
  - $\rightarrow q^k$ as a (sub)gradient of Φ at p^k
- use an **iterative update rule** to compute
  p^{k+1} from $(p^k, q^k, \mathcal{P})$ and a step size $\alpha_k \in \mathbb{R}_+$

end

output: p^*

We define a method as **a first order oracle** + **an iterative algorithm**

Instances of methods

We have three methods

- μ**SDP+IPM:** $\begin{cases} \text{Lower smooth oracle} \\ \text{Interior Points Method} \end{cases}$ $\rightarrow$ the discretization of $\mathbb{R}^{n_x}, \mathbb{R}^{n_u}$ is critical
- k**SDDP+PSM:** $\begin{cases} \text{Lower polyhedral oracle} \\ \text{Projected Subgradient Method} \end{cases}$ $\rightarrow$ the value of $k \in \mathbb{N}$ is critical
- μ**SDP+PGD:** $\begin{cases} \text{Lower smooth oracle} \\ \text{Projected Gradient Descent} \end{cases}$ $\rightarrow$ same as μSDP+IPM

for **each method**, we try **several instances**

Evaluate a profile $p^* \in \mathbb{R}^T$

Given a profile $p^* \in \mathbb{R}^T$, we run the SDDP algorithm to compute

$$\underline{V}_T(x) = K(x)\ , \ \ \forall x \in \mathbb{R}^2$$
$$\underline{V}_t(x) = \inf_{u \in \mathcal{U}_t(x)} \mathbb{E}\Big[L_t(x, u, \mathbf{W}_{t+1}, p_t^*) + \underline{V}_{t+1}\big(f_t(x, u, \mathbf{W}_{t+1})\big)\Big]$$
$$\forall x \in \mathbb{R}^2\ , \ \ \forall t \in \{0, \dots, T-1\}$$

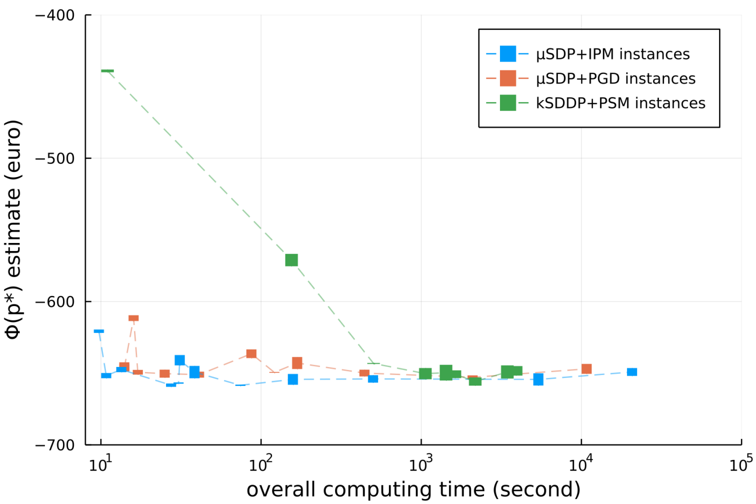
Then, we obtain a policy $\{\underline{\pi}_t\}_{t \in \{0,\dots,T-1\}}$ from $\{\underline{V}_t\}_{t \in \{0,\dots,T-1\}}$ and estimate the expected cost by sampling 25.000 scenarios

$$\overline{V}_0(x_0) = \mathbb{E}\Big[\sum_{t=0}^{T-1} L_t\big(\mathbf{X}_t, \underline{\pi}_t(\mathbf{X}_t), \mathbf{W}_{t+1}, p_t^*\big) + K(\mathbf{X}_T)\Big]$$

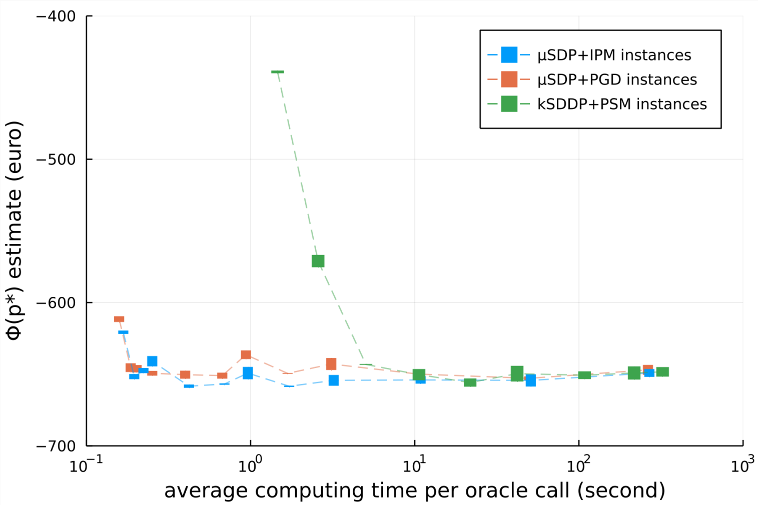
We deduce

$$\underline{V}_0(x_0) \leq \Phi(p^*) \leq \overline{V}_0(x_0)$$

Results: cost vs overall computing time



Results: cost vs time per oracle call



Outline of the presentation

Conclusion and perspectives

- We have introduced a class of **parametric multistage stochastic optimization problems**
  to model **day-ahead power scheduling**
- We have presented the **differentiability properties**
  of parametric value functions
- We have presented **efficient numerical methods**
  to solve such problems
- Our main perspective lies in the application of our methods
  to several **concrete use cases in energy markets**

References I

Adrien Le Franc. Subdifferentiability in convex and stochastic optimization applied to renewable power systems. 2021.