Subdifferentiability in Convex and Stochastic Optimization Applied to Renewable Power Systems

PhD thesis supervised by Michel De Lara

Adrien Le Franc

December 8th, 2021

École des Ponts ParisTech

efficacity

Schneider Electric

Outline of the presentation

1. **EMSx: a numerical benchmark for energy management systems**

2. **Parametric multistage stochastic optimization for day-ahead power scheduling**

3. **Perspectives for numerical methods in generalized convexity**

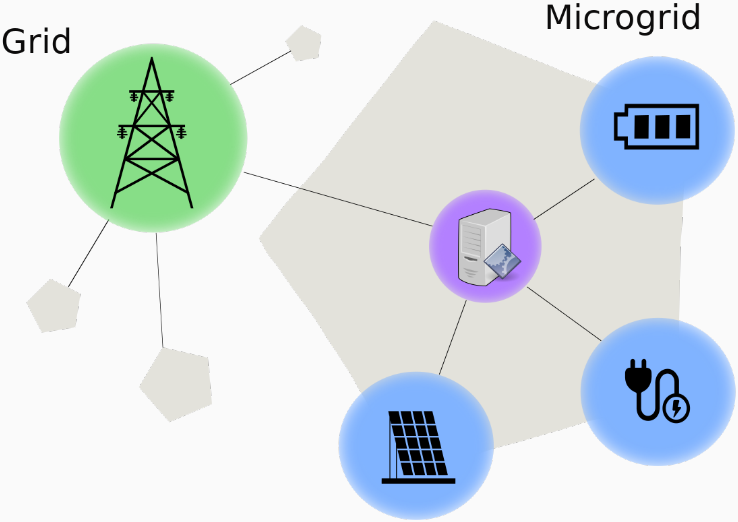
Outline of the presentation

1. **EMSx: a numerical benchmark for energy management systems**

2. Parametric multistage stochastic optimization for day-ahead power scheduling

3. Perspectives for numerical methods in generalized convexity

A PV-battery microgrid



Contribution of the section

- **Question**

How to evaluate an **Energy Management System** (EMS)
designed for operating a microgrid with
uncertain load and production
at **least expected cost**?

- **Our contribution**

We introduce EMSx, a **microgrid controller benchmark**
to compare (deterministic and stochastic) EMS techniques
on an **open** and **diversified** testbed

Outline of the section

1. **EMSx: a numerical benchmark for energy management systems**

   The EMSx dataset

   The EMSx mathematical framework: microgrid model

   The EMSx mathematical framework: controller assessment

   The EMSx software

   Numerical examples of controllers

Examples of daily scenarios from EMSx

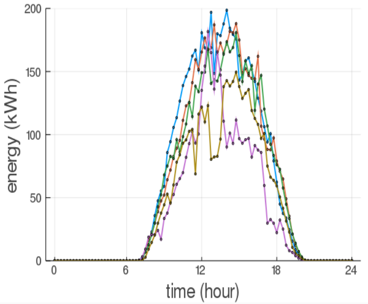


Figure 1: Examples of daily photovoltaic profiles

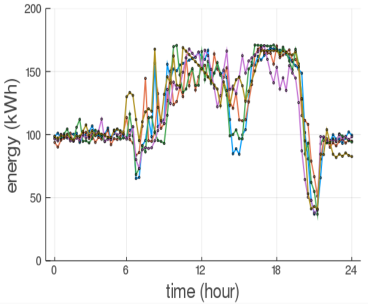


Figure 2: Examples of daily load profiles

Over 1 year of historical observations and forecasts collected by Schneider Electric on 70 industrial sites

Stochasticity of the net demand across sites

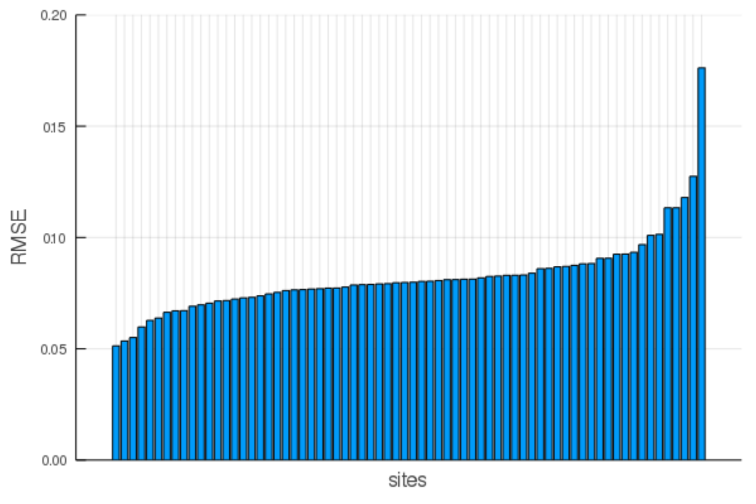


Figure 3: RMSE of the net demand forecasts for each of the 70 sites

Outline of the section

1. **EMSx: a numerical benchmark for energy management systems**

Time scale and variables

We make decisions at time steps $t \in \{0, 1, \ldots, T\}$
over one week ($\Delta_t = 15$ min , $T = 672$)

- $x_t \in [0, 1]$ state of charge of the battery
- $u_t \in [\underline{u}, \overline{u}]$ energy charged ($u_t \geq 0$)
  or discharged ($u_t \leq 0$) over $[t, t+1]$
- $w_{t+1} = (g_{t+1}, d_{t+1})$ generation and demand
  historical data over $[t, t+1]$
- $\hat{w}_{t,t+k} = (\hat{g}_{t,t+k}, \hat{d}_{t,t+k})$, $k \in \{1, \ldots, 96\}$
  generation and demand historical forecast at time t
  over $[t+k-1, t+k]$



Our microgrid management model

- state of charge ruled by the **dynamics**

$$x_{t+1} = f(x_t, u_t) = x_t + \frac{\rho_{\mathrm{c}}}{c} u_t^+ - \frac{1}{\rho_{\mathrm{d}} c} u_t^-$$

- controls restricted to the **admissibility set**

$$\mathcal{U}(x_t) = \{u_t \in \mathbb{R} \mid \underline{u} \leq u_t \leq \overline{u} \text{ and } 0 \leq f(x_t, u_t) \leq 1\}$$

- energy exchanges induce a **cost**

$$L_t(u_t, w_{t+1}) = p_t^{\mathrm{buy}} \cdot (d_{t+1} - g_{t+1} + u_t)^+ - p_t^{\mathrm{sell}} \cdot (d_{t+1} - g_{t+1} + u_t)^-$$

Outline of the section

Online information for decision making

- A **partial chronicle** is a sequence $h = (h_0, \ldots, h_{T-1})$ of vectors where for $t \in \{0, \ldots, T-1\}$

$$h_t = \begin{pmatrix} w_t, w_{t-1}, \ldots, w_{t-95} \\ \hat{w}_{t,t+1}, \ldots, \hat{w}_{t,t+96} \end{pmatrix} \in \mathbb{H} = \mathbb{R}^{2\times 96} \times \mathbb{R}^{2\times 96}$$

- A **controller** is a sequence $\phi = (\phi_0, \ldots, \phi_{T-1})$ of mappings where for $t \in \{0, \ldots, T-1\}$

$$\begin{aligned} \phi_t \;:\; & [0,1] \times \mathbb{H} \to \mathbb{R} \\ & (x_t, h_t) \mapsto \phi_t(x_t, h_t) \in \mathcal{U}(x_t) \end{aligned}$$

Management cost of a controller

For each site $i \in I = \{1, \ldots, 70\}$

- A controller ϕ^i applied to a partial chronicle $h \in \mathbb{H}^T$ yields a **management cost**

$$J^i(\phi^i, h) = \sum_{t=0}^{T-1} L_t^i(u_t, w_{t+1})$$

$$x_0 = 0$$

$$x_{t+1} = f^i(x_t, u_t)$$

$$\underbrace{u_t = \phi_t^i(x_t, h_t)}_{\text{nonanticipativity}}$$

- If we allow anticipative decisions we obtain a lower bound for management costs $\underline{J}^i(h) \leq J^i(\phi^i, h)$

Gain of a controller ϕ^i on site $i \in I$

We have a pool of **simulation chronicles** $\mathscr{S}^i \subset \mathbb{H}^T$

- We measure **gains** w.r.t. a **dummy controller** $\phi^{\mathrm{d}} = 0$ (which does not use the battery)

$$G^i(\phi^i) = \frac{1}{|\mathscr{S}^i|} \sum_{h \in \mathscr{S}^i} J^i(\phi^{\mathrm{d}}, h) - J^i(\phi^i, h)$$

- We define the **anticipative gain**

$$\overline{G}^i = \frac{1}{|\mathscr{S}^i|} \sum_{h \in \mathscr{S}^i} J^i(\phi^{\mathrm{d}}, h) - \underline{J}^i(h)$$

- We obtain an upper bound for gains $\overline{G}^i \geq G^i(\phi^i)$

Normalized score of a control technique $\{\phi^i\}_{i \in I}$

- We define the **normalized score** of a control technique $\{\phi^i\}_{i \in I}$

$$\mathcal{G}\left(\{\phi^i\}_{i \in I}\right) = \frac{1}{|I|} \sum_{i \in I} \frac{G^i(\phi^i)}{\overline{G}^i}$$

- A performing control technique gives

$$\underbrace{0 \leq}_{\text{if better that } \phi^{\mathbf{d}}} \quad \mathcal{G}\left(\{\phi^i\}_{i \in I}\right) \quad \underbrace{\leq 1}_{\text{always}}$$

Outline of the section

1. **EMSx: a numerical benchmark for energy management systems**

   The EMSx dataset

   The EMSx mathematical framework: microgrid model

   The EMSx mathematical framework: controller assessment

   The EMSx software

   Numerical examples of controllers

A Julia package: EMSx.jl

```julia
struct Information
    t::Int64
    soc::Float64
    pv::Array{Float64,1}
    forecast_pv::Array{Float64,1}
    load::Array{Float64,1}
    forecast_load::Array{Float64,1}
    price::Price
    battery::Battery
    site_id::String
end
```

The EMSx.jl built-in type Information gathers all the information available to the controller to make a decision

A Julia package: EMSx.jl

```julia
using EMSx

mutable struct DummyController <: EMSx.AbstractController end

EMSx.compute_control(controller::DummyController,
    information::EMSx.Information) = 0.

const controller = DummyController()

EMSx.simulate_sites(controller,
    "home/xxx/path_to_save_folder",
    "home/xxx/path_to_price",
    "home/xxx/path_to_metadata",
    "home/xxx/path_to_simulation_data")
```

Example of the implementation and simulation of a dummy controller with the EMSx.jl package

Outline of the section

1. **EMSx: a numerical benchmark for energy management systems**

    The EMSx dataset

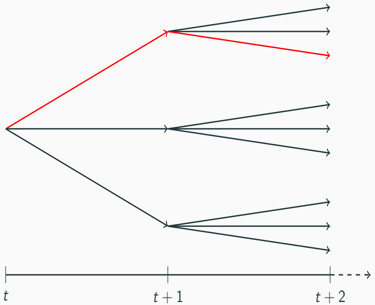
    The EMSx mathematical framework: microgrid model

    The EMSx mathematical framework: controller assessment

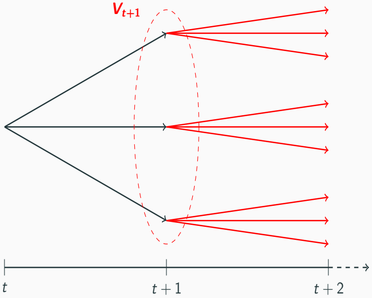
    The EMSx software

    Numerical examples of controllers

Standard controller design techniques



Look-ahead techniques:
MPC, OLFC

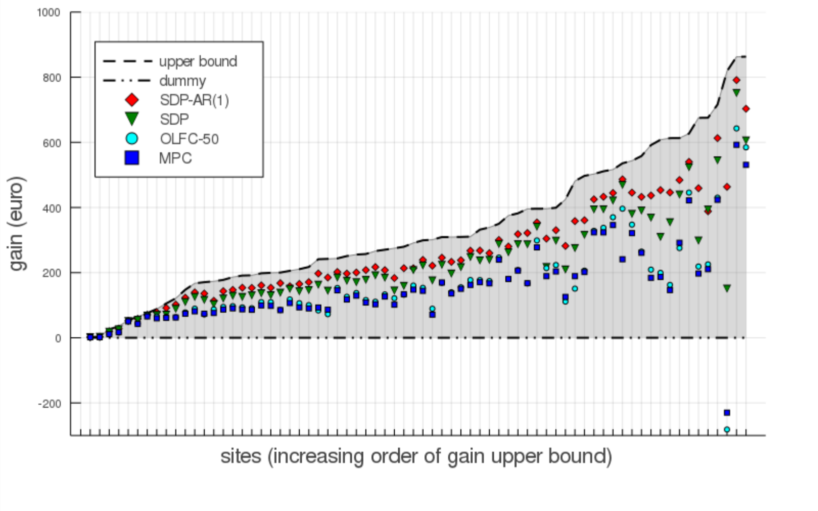


Cost-to-go techniques:
SDP, SDP-AR(k)

Normalized score per design technique

	Normalized score	**Offline time** (seconds)	**Online time** (seconds)
MPC	0.487	0.00	$9.82\ 10^{-4}$
OLFC-10	0.506	0.00	$1.14\ 10^{-2}$
OLFC-50	0.513	0.00	$8.62\ 10^{-2}$
OLFC-100	0.510	0.00	$1.87\ 10^{-1}$
SDP	0.691	2.67	$3.09\ 10^{-4}$
SDP-AR(1)	0.794	38.1	$4.44\ 10^{-4}$
SDP-AR(2)	0.795	468	$5.55\ 10^{-4}$
Upper bound	**1.0**	-	-

Detailed gain over the 70 sites



Outline of the presentation

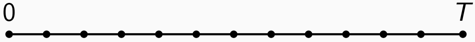
1. EMSx: a numerical benchmark for energy management systems

2. **Parametric multistage stochastic optimization for day-ahead power scheduling**

3. Perspectives for numerical methods in generalized convexity

A typical power scheduling example

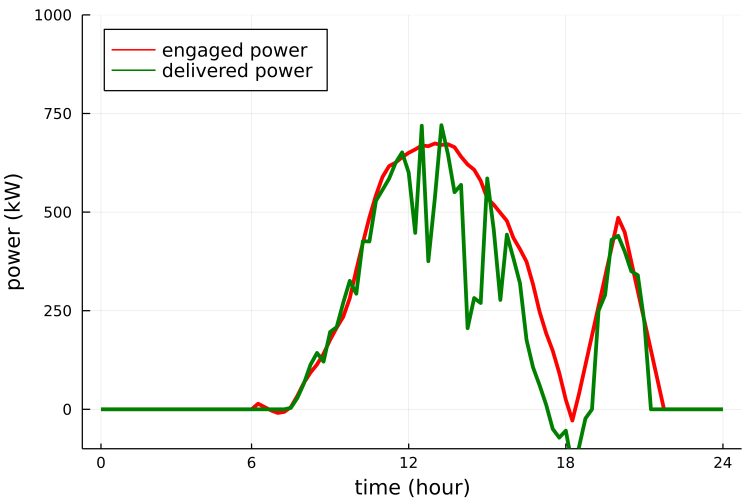
- We operate a solar plant over **one day** with discrete time steps $t \in \{0, 1, \ldots, T\}$



- 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 p and $\tilde{p}$ induce penalties



Contribution of the section

- **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 section

2. Parametric multistage stochastic optimization for day-ahead power scheduling

Parametric multistage stochastic optimization

Differentiability of smooth parametric value functions

Lower smooth approximations

Lower polyhedral approximations

Numerical application

Our standard formulation

We consider a **multistage stochastic optimization problem** parametrized by $p = (p_0, \ldots, p_T) \in \mathbb{R}^{n_p \times (T+1)}$ formulated as

$$\Phi(p) = \inf_{\mathbf{U_0}, \ldots, \mathbf{U}_{T-1}} \mathbb{E}\Big[\sum_{t=0}^{T-1} L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}, p_t) + K(\mathbf{X}_T, p_T)\Big]$$

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

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

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

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

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

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

Outline of the section

2. Parametric multistage stochastic optimization for day-ahead power scheduling

Assumption

Assumption (smoothness)

1. *the cost functions* $\{L_t\}_{t \in \{0,\ldots,T-1\}}$ *and* K
   are ***differentiable w.r.t.*** p_t
2. *for all* $t \in \{0, \ldots, T-1\}$, *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. Parametric multistage stochastic optimization for day-ahead power scheduling

Parametric multistage stochastic optimization

Differentiability of smooth parametric value functions

Lower smooth approximations

Lower polyhedral approximations

Numerical application

Assumption

We consider a **parameter set** $\mathcal{P} \subseteq \mathbb{R}^{n_p \times (T+1)}$

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 $\mathrm{dom}(\mathcal{U}_t) \subseteq \mathbb{R}^{n_x} \times \mathcal{P}_t$

where $\mathcal{P}_t = \mathrm{proj}_t(\mathcal{P}) \subseteq \mathbb{R}^{n_p}$, $\forall t \in \{0, \ldots, T\}$

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

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

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

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

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 $\left\{\underset{\sim}{V}_t^\mu\right\}_{t \in \{0, \dots, T\}}$
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*

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

Outline of the section

2. Parametric multistage stochastic optimization for day-ahead power scheduling

State extension

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

$$\Phi(p) = \inf_{\mathbf{U_0}, \dots, \mathbf{U}_{T-1}} \mathbb{E}\Big[\sum_{t=0}^{T-1} L_t^\sharp(\mathbf{X}_t^\sharp, \mathbf{U}_t, \mathbf{W}_{t+1}) + K^\sharp(\mathbf{X}_T^\sharp) \Big]$$

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

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

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

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

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 $k \in \mathbb{N}^*$ forward-backward passes of the SDDP algorithm the approximation error of the value function $V_0^\sharp$ by the lower polyhedral approximation $\underline{V}_0^k$ is bounded by

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

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

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

Outline of the section

2. Parametric multistage stochastic optimization for day-ahead power scheduling

Parametric multistage stochastic optimization

Differentiability of smooth parametric value functions

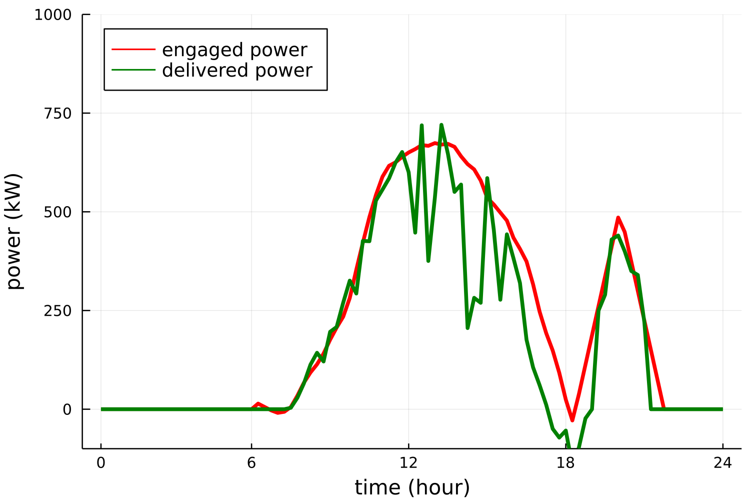
Lower smooth approximations

Lower polyhedral approximations

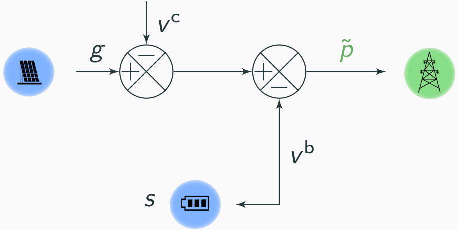
Numerical application

Back to our problem

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

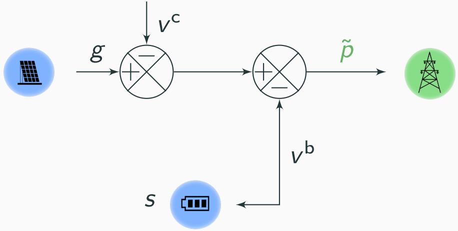


Schematic organization of the solar plant



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

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

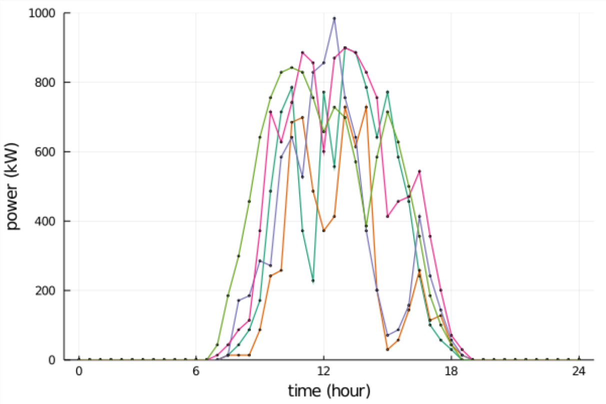
- The state process $\mathbf{X}$ is ruled by the dynamics

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

- The stage costs formulate as

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

Scenarios



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

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

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

Generic method

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

for $n = 1 \dots N$ **do**

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

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**
i.e. several discretizations of $\mathbb{R}^{n_x}, \mathbb{R}^{n_u}$ or several values of k

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

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

$$\underline{V}_T(x) = K(x) \ , \ \ \forall x \in \mathbb{R}^2$$

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

$$\forall x \in \mathbb{R}^2 \ , \ \ \forall t \in \{0, \dots, T-1\}$$

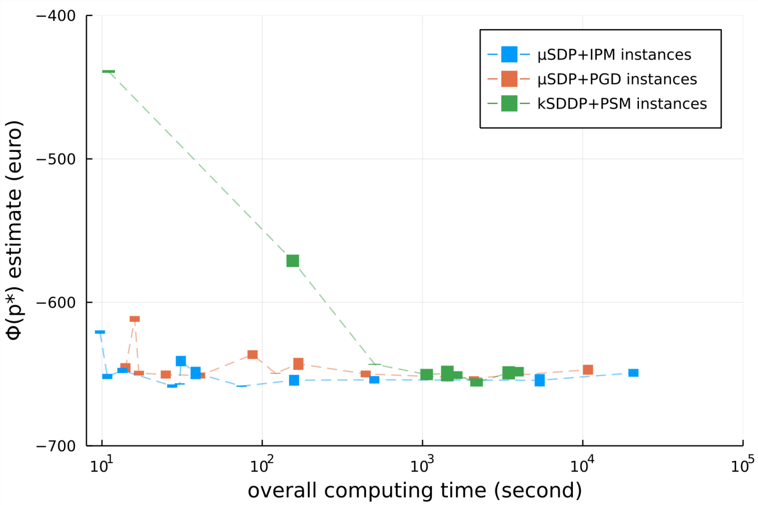
Then, we obtain a policy $\{\underline{\pi}_t\}_{t \in \{0,\dots,T-1\}}$ from $\{\underline{V}_t\}_{t \in \{0,\dots,T-1\}}$ and estimate the expected cost by sampling 25.000 scenarios

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

We deduce

$$\underbrace{\underline{V}_0(x_0) \leq}_{\text{exact}} \ \Phi(p^*) \ \underbrace{\leq \overline{V}_0(x_0)}_{\text{statistical}}$$

Results: cost vs overall computing time



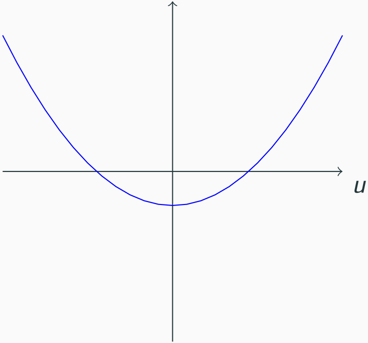
Outline of the presentation

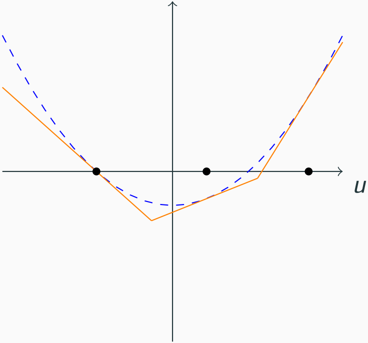
1. EMSx: a numerical benchmark for energy management systems

2. Parametric multistage stochastic optimization for day-ahead power scheduling

3. Perspectives for numerical methods in generalized convexity

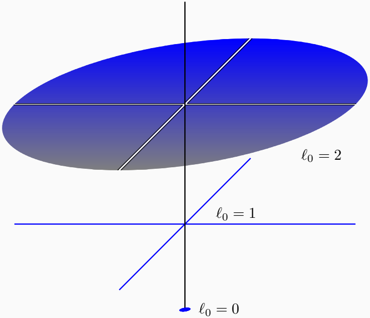
Fenchel conjugate, subdifferential, and polyhedral approximate of a convex lower semicontinuous function

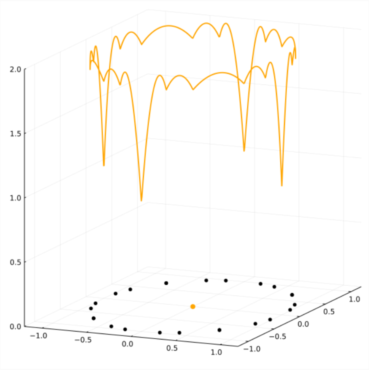




$$f(u) \geq \max_{\substack{v_i \in \partial f(u_i) \\ i \in I}} \Big(\langle u \, , v_i \rangle + \big(-f^\star(v_i) \big) \Big)$$

Beyond convex lower semicontinuous functions...





$$\ell_0(u) \geq \max_{\substack{v_i \in \partial_{\text{¢}} \ell_0(u_i) \\ i \in I}} \left(\text{¢}(u, v_i) + \left(-\ell_0^{\text{¢}}(v_i)\right) \right)$$

Contribution of the section

- **Question**

Can we leverage **generalized convexity** notions
to solve **nonconvex optimization problems**?

- **Our contribution**

We focus on **one-sided linear conjugacies**
to extend the **mirror descent algorithm**
and study its aplicability in **sparse optimization**

Outline of the section

3. Perspectives for numerical methods in generalized convexity

Background in one-sided linear (OSL) conjugacies

The mirror descent algorithm

The Capra coupling and the ℓ_0 pseudonorm

Perspectives for sparse optimization

Couplings and generalized Fenchel-Moreau conjugacies

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\} \quad \text{and} \quad (+\infty) \dotplus (-\infty) = (-\infty) \dotplus (+\infty) = -\infty$$

Definition

Two sets $\mathbb{U}$ ("Primal") and $\mathbb{V}$ ("Dual") paired by
a **coupling function** $c : \mathbb{U} \times \mathbb{V} \to \overline{\mathbb{R}}$
give rise to the ***c*-Fenchel-Moreau conjugacy**

$$f \in \overline{\mathbb{R}}^{\mathbb{U}} \mapsto f^c \in \overline{\mathbb{R}}^{\mathbb{V}}$$

$$f^c(v) = \sup_{u \in \mathbb{U}} \Big(c(u,v) \dotplus \big(-f(u)\big) \Big) , \quad \forall v \in \mathbb{V}$$

Example: two vector spaces $\mathbb{U}$ and $\mathbb{V}$ paired with a bilinear form $\langle \cdot \, , \cdot \rangle$
give rise to the classic Fenchel conjugacy $f \in \overline{\mathbb{R}}^{\mathbb{U}} \mapsto f^\star \in \overline{\mathbb{R}}^{\mathbb{V}}$

Generalized c-biconjugate and c-convexity

- We also introduce the c'-**Fenchel-Moreau conjugacy**

$$g \in \overline{\mathbb{R}}^{\mathbb{V}} \mapsto g^{c'} \in \overline{\mathbb{R}}^{\mathbb{U}} , \quad g^{c'}(u) = \sup_{v \in \mathbb{V}} \Big(\underbrace{c(u,v)}_{=c'(v,u)} \dot{+} \big(-g(v)\big) \Big) , \quad \forall u \in \mathbb{U}$$

- This gives rise to the c-**Fenchel-Moreau biconjugate**

$$f \in \overline{\mathbb{R}}^{\mathbb{U}} \mapsto f^{cc'} \in \overline{\mathbb{R}}^{\mathbb{U}} , \quad f^{cc'}(u) = \big(f^{c}\big)^{c'}(u) , \quad \forall u \in \mathbb{U}$$

Definition

A function $f \in \overline{\mathbb{R}}^{\mathbb{U}}$ is c-**convex** if $f = f^{cc'}$, that is

$$f(u) = \sup_{v \in \mathbb{V}} \Big(c(u,v) \dot{+} \big(-f^{c}(v)\big) \Big) , \quad \forall u \in \mathbb{U}$$

Example: a proper function $f \in \overline{\mathbb{R}}^{\mathbb{U}}$ is $\langle\, , \rangle$-convex iff f is convex and lsc

Generalized c-subdifferential

Definition

The ***c*-subdifferential** of a function $f \in \overline{\mathbb{R}}^{\mathbb{U}}$ at $u \in \mathbb{U}$
with respect to the coupling c is the subset $\partial_c f(u) \subseteq \mathbb{V}$
defined equivalently either by

$$v \in \partial_c f(u) \iff f^c(v) = c(u,v) \dotplus \big(-f(u)\big)$$

or by

$$v \in \partial_c f(u) \iff c(u,v) \dotplus \big(-f(u)\big) \geq c(u',v) \dotplus \big(-f(u')\big) \ , \ \ \forall u' \in \mathbb{U}$$

One-sided linear (OSL) couplings

- Let $\mathbb{U}$ and $\mathbb{V}$ be two **vector spaces** paired by a bilinear form $\langle \cdot \, , \cdot \rangle$
- We suppose given a mapping $\theta : \mathbb{W} \to \mathbb{U}$ where $\mathbb{W}$ is **any set**

Definition

We define the **one-sided linear coupling (OSL)**

$$\mathbb{W} \overset{\star_\theta}{\longleftrightarrow} \mathbb{V}$$

between $\mathbb{W}$ and $\mathbb{V}$ by

$$\star_\theta(w, v) = \langle \theta(w) \, , v \rangle \; , \;\; \forall w \in \mathbb{W} \; , \;\; \forall v \in \mathbb{V}$$

Some properties of convex analysis can be extended
to $\star_\theta$-convex functions...

Outline of the section

The standard Bregman divergence

Definition

Let $\mathbb{W}$ and $\mathbb{V}$ be two vector spaces paired by a bilinear form $\langle \cdot \,, \cdot \rangle$
let $\kappa \in \overline{\mathbb{R}}^{\mathbb{W}}$ be a proper, closed, convex and differentiable (divergence generating) function.
We define the **Bregman divergence associated with κ** by

$$D_\kappa(w, w') = \kappa(w) - \kappa(w') - \langle w - w' \,, \nabla\kappa(w') \rangle \,, \quad \forall (w, w') \in \mathbb{W} \times \mathrm{dom}(\nabla\kappa)$$

D_κ is not a distance, but if κ is **strongly convex**

- $D_\kappa(w, w') \geq 0 \,, \;\; \forall (w, w') \in \mathbb{W} \times \mathrm{dom}(\nabla\kappa)$
- $D_\kappa(w, w') = 0 \iff w = w'$
- We have a "triangular inequality" that makes mirror descent work

The Bregman divergence with couplings

Definition

Let $\mathbb{W}$ and $\mathbb{V}$ be two sets and a **finite coupling** $\mathbb{W} \overset{c}{\longleftrightarrow} \mathbb{V}$
let $\kappa \in \overline{\mathbb{R}}^{\mathbb{W}}$ be a proper c-convex (divergence generating) function.
We define the **c-Bregman divergence associated with κ** by

$$D^c_\kappa(w, w', v') = \kappa(w) - \kappa(w') - c(w, v') + c(w', v') ,$$
$$\forall (w, w') \in \mathbb{W} \times \mathrm{dom}(\partial_c \kappa) , \ \forall v' \in \partial_c \kappa(w')$$

If moreover

- $\mathbb{V}$ is a vector space
- The coupling c is **OSL**
- The function κ is **c-strongly convex**

We retrieve some properties of the original Bregman divergence

The mirror descent algorithm with OSL couplings

We consider the optimization problem

$$\min_{w \in W} f(w)$$

- We initialize three sequences by

$$\begin{aligned} w_0 &\in W \\ v_0 &\in \partial_{\star_\theta} \left(\kappa + \delta_W\right)(w_0) \\ v_0^f &\in \partial_{\star_\theta} f(w_0) \end{aligned}$$

- We run $N \in \mathbb{N}$ steps with a step size $\alpha_n > 0$ and the update rules

$$\begin{aligned} w_{n+1} &\in \arg\min_{w \in W} \left(\kappa(w) + \left\langle \theta(w), \alpha_n v_n^f - v_n \right\rangle\right) \\ v_{n+1} &= v_n - \alpha_n v_n^f \\ v_{n+1}^f &\in \partial_{\star_\theta} f(w_{n+1}) \end{aligned}$$

Convergence result

Theorem (Le Franc [2021])

Under a suitable choice of divergence generating function κ we can bound the optimality gap by

$$\min_{0 \leq n \leq N-1} \left(f(w_n) - \inf_{w \in W} f(w) \right) \leq \frac{R^2 + \frac{G^2}{4} \sum_{n=0}^{N-1} \alpha_n^2}{\sum_{n=0}^{N-1} \alpha_n}$$

- R and G are constant values determined by the problem and by κ
- We retrieve the same convergence rule as in the standard mirror descent algorithm

Outline of the section

3. Perspectives for numerical methods in generalized convexity

Background in one-sided linear (OSL) conjugacies

The mirror descent algorithm

The Capra coupling and the ℓ_0 pseudonorm

Perspectives for sparse optimization

We introduce the Capra coupling

Definition (Chancelier and De Lara [2020])

Let $\|\!|\cdot|\!\|$ be a norm on $\mathbb{R}^d$ called the **source norm**

we define the **Capra** *coupling* $\mathbb{R}^d \overset{¢}{\longleftrightarrow} \mathbb{R}^d$ by

$$\forall v \in \mathbb{R}^d\ ,\ \ ¢(u,v) = \begin{cases} \frac{\langle u \,, v \rangle}{\|\!|u|\!\|} & \text{if } u \neq 0 \\ 0 & \text{if } u = 0 \end{cases}$$

The coupling Capra is

- **Constant Along Primal RAys (Capra)**
- **OSL** with $¢(u,v) = \langle n(u)\,, v \rangle\ ,\ \ \forall (u,v) \in (\mathbb{R}^d)^2$

$$\text{where} \quad n(u) = \begin{cases} \frac{u}{\|\!|u|\!\|} & \text{if } u \neq 0 \\ 0 & \text{if } u = 0 \end{cases}$$

The ℓ_0 pseudonorm

$$\ell_0(u) = \left|\left\{j \in \{1, \ldots, d\} \,\middle|\, u_j \neq 0\right\}\right| , \ \forall u \in \mathbb{R}^d$$

Proposition (Chancelier and De Lara [2021])

If both the source norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are ***orthant-strictly monotonic****, we have that*

$$\ell_0 = \ell_0^{¢¢'}$$

$$\partial_{¢}\ell_0(u) \neq \emptyset , \ \forall u \in \mathbb{R}^d$$

that is, the pseudonorm ℓ_0 is ***Capra-convex*** *and* ***Capra-subdifferentiable*** *everywhere on $\mathbb{R}^d$*

Examples:
- $||(0,1)||_\infty = ||(1,1)||_\infty = 1$ hence ℓ_∞ is not OSM
- ℓ_2 is OSM

Explicit formulations for the Capra-subdifferential of ℓ_0

Proposition (Le Franc [2021])

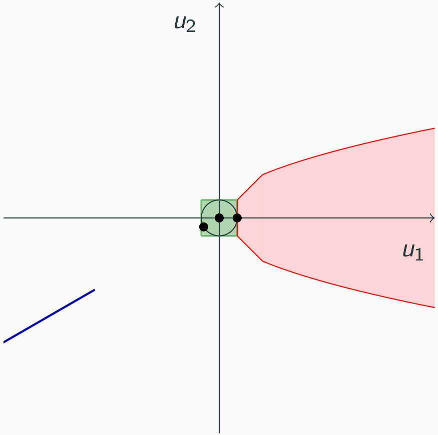
For the source norms $|||\cdot||| = \ell_p$, $p \in]1, \infty[$, *we have that*

$$\partial_{¢}\ell_0(0) = \mathbb{B}_{\ell_\infty}$$

and if $u \neq 0$, $l = \ell_0(u)$, $L = supp(u)$, *for* $v \in \mathbb{R}^d$, $|v_{\nu(1)}| \geq \ldots \geq |v_{\nu(d)}|$, $||v||^{\mathrm{tn}}_{(k,q)} = \left(|v_{\nu(1)}|^q + \ldots + |v^q_{\nu(k)}|\right)^{\frac{1}{q}}$ *and* $\frac{1}{p} + \frac{1}{q} = 1$

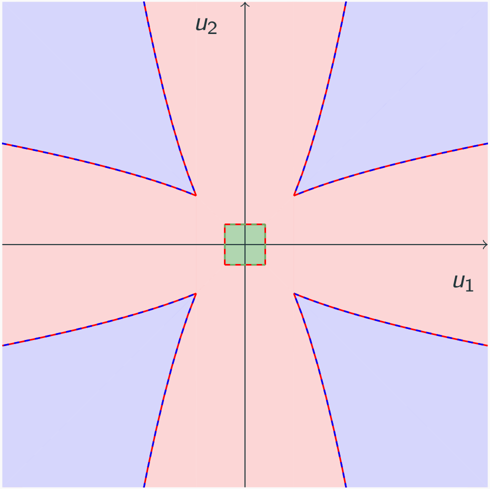
$$v \in \partial_{¢}\ell_0(u) \iff \begin{cases} v_L \in \mathcal{N}_{\mathbb{B}_{||\cdot||_p}}\left(\frac{u}{||u||_p}\right) \\ |v_j| \leq \min_{i \in L} |v_i| \,, \ \ \forall j \notin L \\ |v_{\nu(k+1)}|^q \geq \left(||v||^{\mathrm{tn}}_{(k,q)} + 1\right)^q - \left(||v||^{\mathrm{tn}}_{(k,q)}\right)^q \,, \ \ \forall k < l \\ |v_{\nu(l+1)}|^q \leq \left(||v||^{\mathrm{tn}}_{(l,q)} + 1\right)^q - \left(||v||^{\mathrm{tn}}_{(l,q)}\right)^q \end{cases}$$

Examples of sets $\partial_{\text{¢}}\ell_0(u)$ in $\mathbb{R}^2$ with the source norm $|||\cdot||| = \ell_2$



$$\partial_{\text{¢}}\ell_0(0,0)\ ,\ \ \partial_{\text{¢}}\ell_0(1,0)\ ,\ \ \partial_{\text{¢}}\ell_0(u_1,u_2)$$

Vizualisation of $\partial_{\c{c}}\ell_0$ in $\mathbb{R}^2$ with the source norm $|||\cdot||| = \ell_2$



$$\partial_{\c{c}}\ell_0(0) \bigcup \left\{ \bigcup_{\ell_0(u)=1} \partial_{\c{c}}\ell_0(u) \right\} \bigcup \left\{ \bigcup_{\ell_0(u)=2} \partial_{\c{c}}\ell_0(u) \right\}$$

Outline of the section

3. Perspectives for numerical methods in generalized convexity

Background in one-sided linear (OSL) conjugacies

The mirror descent algorithm

The Capra coupling and the ℓ_0 pseudonorm

Perspectives for sparse optimization

Capra-convex sparse optimization problems

We consider problems of type

$$\min_{u \in U} \ell_0(u)$$

and we look for constraint sets $U \subseteq \mathbb{U}$ for which
we have a **Capra-convex (sparse) optimization problem**

Definition

We say that **the set $U \subseteq \mathbb{U}$ is Capra-convex** if the indicator function **δ_U is a Capra-convex function**

Which sets are Capra-convex ?

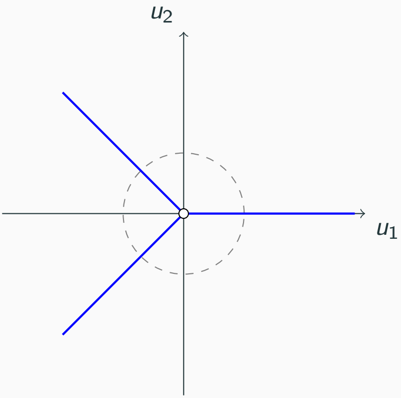
Characterization of Capra-convex sets

Proposition (Le Franc [2021])

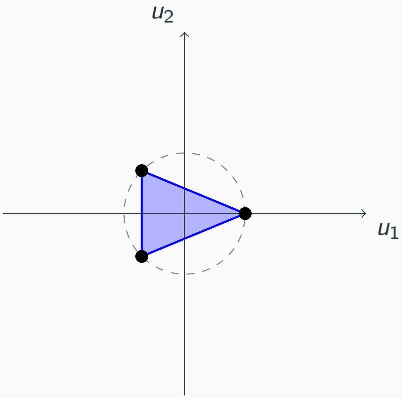
Let the source norm $|||\cdot||| = ||\cdot||_p$, $p \in]1, \infty[$ *and* $U \subseteq \mathbb{U}$ *be a nonempty set*

$$U \text{ is Capra-convex} \iff \begin{cases} U \text{ is a cone} \\ U \cup \{0\} \text{ is closed} \\ U \cap \{0\} = \overline{\mathrm{co}}\big(n(U)\big) \cap \{0\} \end{cases}$$

Example with $\|\cdot\| = \ell_2$: a non Capra-convex cone

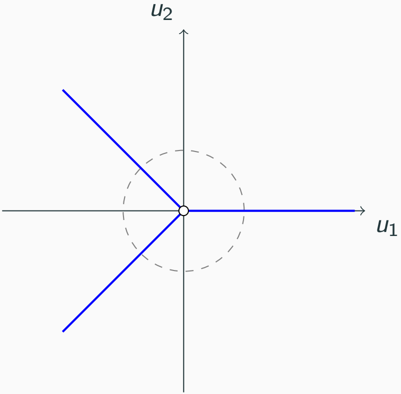


K

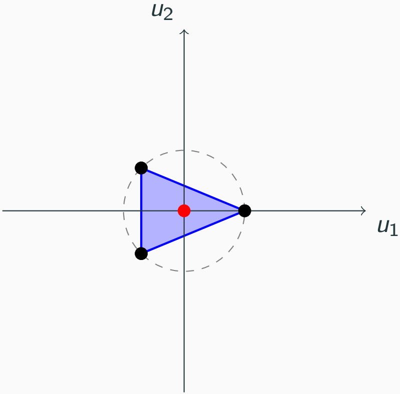


$\overline{\mathrm{co}}(n(K))$

Example with $\|\cdot\| = \ell_2$: a non Capra-convex cone



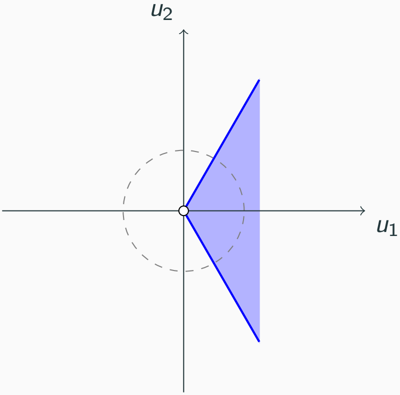
K



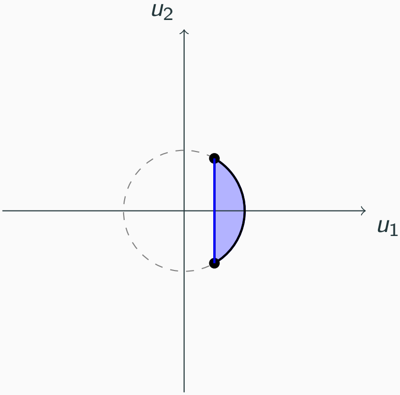
$\overline{\mathrm{co}}(n(K))$

$$K \cap \{0\} \neq \overline{\mathrm{co}}\big(n(K)\big) \cap \{0\}$$

Example with $\|\cdot\| = \ell_2$: a Capra-convex cone

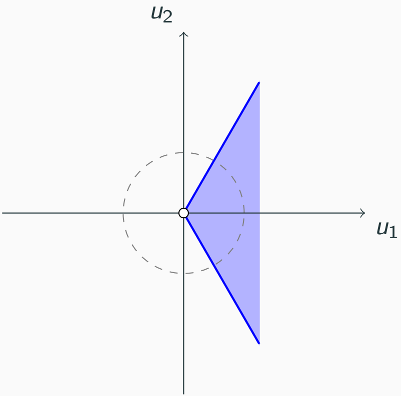


K

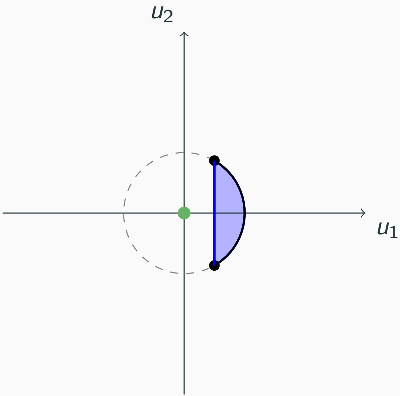


$\overline{\mathrm{co}}(n(K))$

Example with $\|\cdot\| = \ell_2$: a Capra-convex cone



K

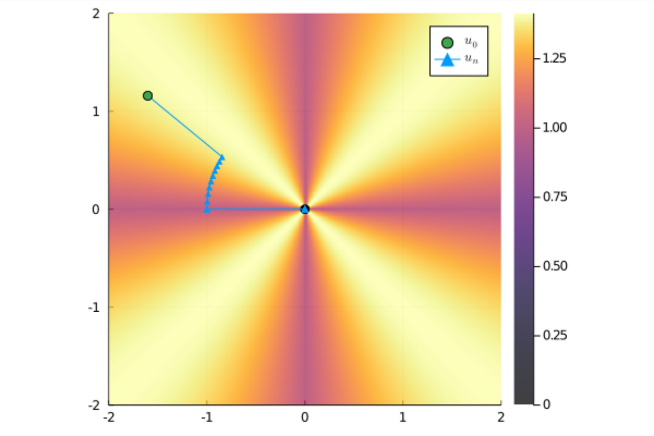


$\overline{\mathrm{co}}(n(K))$

$$K \cap \{0\} = \overline{\mathrm{co}}\big(n(K)\big) \cap \{0\}$$

Application of the mirror descent algorithm

Only a (trivial) unconstrained test so far: $\min_{u \in \mathbb{R}^2} \frac{||u||_1}{||u||_2}$



Perspectives to handle constraint sets

Turning to e.g. $\min_{u \in K} \frac{||u||_1}{||u||_2}$?

- We need to identify suitable **divergence generating functions** κ such that $\kappa + \delta_K$ **is Capra-strongly convex**
- We need to make sure that we can **compute efficiently**

$$u_{n+1} \in \arg\min_{u \in K} \big(\kappa(u) + ¢(u, \alpha_n v_n^f - v_n)\big)$$

Work in progress...

Conclusion

1.
- We have introduced a **dataset**, a **mathematical framework** and a **software** to compare microgrid controller techniques on a **publicly available** benchmark
- The EMSx benchmark is further detailed in [Le Franc et al., 2021]

2.
- We have introduced a class of **parametric multistage stochastic optimization problems** to model **day-ahead power scheduling**
- We have presented **differentiability properties** of parametric value functions and derived efficient optimization methods

3.
- We have extended **the mirror descent algorithm to OSL couplings**
- We have explicited **the Capra-subdifferential of ℓ_0**
- We have identified **Capra-convex sets** and **Capra-convex sparse optimization problems**

Perspectives

- We look forward to applications of our methods in parametric multistage stochastic optimization to several **concrete use cases in energy markets**

- We plan to study further applications of the mirror descent algorithm to solve **Capra-convex problems**

References I

Jean-Philippe Chancelier and Michel De Lara. Constant along primal rays conjugacies and the l0 pseudonorm. *Optimization*, 0(0):1–32, 2020. URL `https://doi.org/10.1080/02331934.2020.1822836`.

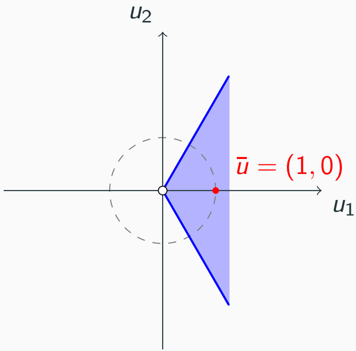
Jean-Philippe Chancelier and Michel De Lara. Capra-convexity, convex factorization and variational formulations for the ℓ_0 pseudonorm. *Set-Valued and Variational Analysis*, pages 1–23, 2021.

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

Adrien Le Franc, Pierre Carpentier, Jean-Philippe Chancelier, and Michel De Lara. Emsx: a numerical benchmark for energy management systems. *Energy Systems*, pages 1–27, 2021.

Appendix: an example where the subdifferential of the sum...

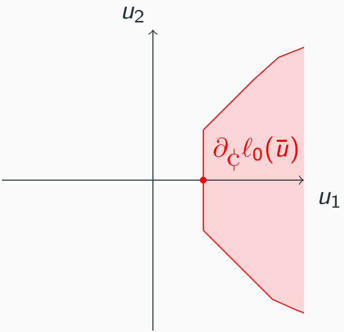
$$|||\cdot||| = \ell_2$$



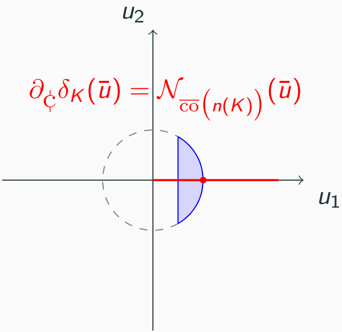
$\bar{u} \in \arg\min_K \ell_0 \implies 0 \in \partial_{\text{ç}}\big(\ell_0 + \delta_K\big)(\bar{u})$ (a property of OSL couplings)

…is not the sum of the subdifferentials

Let $v' \in \partial_{\text{Ç}} \ell_0(\bar{u})$ and $v'' \in \partial_{\text{Ç}} \delta_K(\bar{u})$



$v'_1 \geq 1$



$v''_1 \geq 0$

$0 \notin \partial_{\text{Ç}} \ell_0(\bar{u}) + \partial_{\text{Ç}} \delta_K(\bar{u})$ hence $\partial_{\text{Ç}} \ell_0(\bar{u}) + \partial_{\text{Ç}} \delta_K(\bar{u}) \subsetneq \partial_{\text{Ç}} \left(\ell_0 + \delta_K\right)(\bar{u})$