Optimization of the sum of a convex surrogate and quadratic objective: Example from Electricity Markets

Olivier Huber

WID – University of Wisconsin–Madison

Joint work with: Michael Ferris and Lisa Tang

Collaboration: ISO New England

Outline

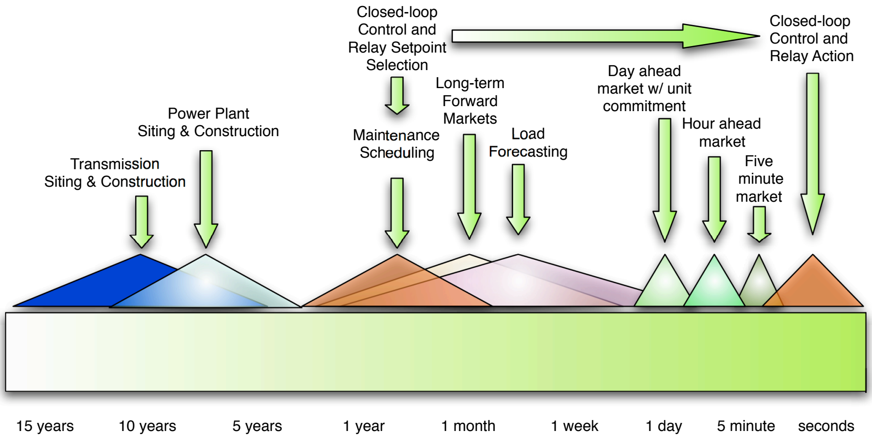
1. Forward Capacity Market
2. MIQP with partially known (closed-loop form) objective function
3. Piecewise-linear approximation
4. Moreau-Yosida approximation to recover differentiability

Principle

- Independence between the grid manager (RTO) and the market participants (companies producing power, retailers, ...)
- One regulator Federal Energy Regulatory Commission (FERC)

Regional Transmission Organization (RTO)

- Manages the network (transmission lines)
- Ensures that the demand is met
- Organise the auction processes
- an RTO has greater responsibility than an Independent System Operator (ISO)


Closed-loop
Control and
Relay Setpoint
Selection
Closed-loop
Control and
Relay Action
Day ahead
market w/ unit
commitment
Long-term
Forward
Markets
Power Plant
Siting & Construction
Hour ahead
market
Maintenance
Scheduling
Load
Forecasting
Transmission
Siting & Construction
Five
minute
market
15 years
10 years
5 years
1 year
1 month
1 week
1 day
5 minute
seconds

Current limitation: limited investments

Capacity shortage

- Few incentives to invest in new facilities or expand/maintain capacity
- Cannot force generators to invest
- There is a high initial investment cost
- Trivia: cannot produce more power than the available capacity

Main issues

- High electricity prices
- Volatility of prices
- Loss of reliability (increased risk of blackout)
- Inability to meet the (future) demand

Forward Capacity Market (FCM)

- "Ensures that the New England power system will have sufficient resources to meet the future demand for electricity"
- provides an incentive for companies to make investments
- the cost is supported by the consumers

Forward Capacity Auction (FCA)

- held annually 3 years in advance
- supply capacity in exchange for market-priced capacity payment
- formulated as an optimization problem

Benefits from an increase in capacity?

ISO's perspective: ICR

- (N)ICR: (Net) Installed Capacity Requirement
- $\approx$ lower bound on the required capacity to meet reliability standards
- criterion for ISONE: "interrupting non-interruptible load, on average, no more than once every 10 years"

Consumer's perspective: EENS minimization

- EENS: Expected Energy Not Served (MWh/year)
- estimate of the demand not met
- depends on the total capacity installed
- computed via Monte-Carlo simulation of scenarios of line and generator failures

Objective function has 2 terms:

$$\underbrace{c^T q}_{\text{Cost of capacity}} \quad + \quad \underbrace{PF \cdot EENS(Q_{ICZ}, Q_{SYS})}_{\text{Cost of lost load}}$$

- PF penalty factor ($\$/MWh$), c cost vector, q capacities, $q_i = 0$ or $\bar{q}_i$
- $Q_{SYS} = \sum_{i \in \mathcal{I}} q_i$, $Q_{ICZ} = \sum_{i \in \mathcal{J}} q_i, \mathcal{J} \subset \mathcal{I}$

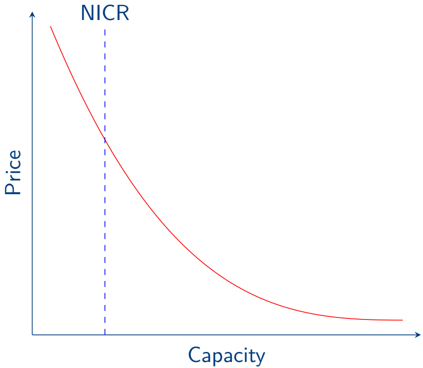
- solution of the optimization problem minimizes this total cost:
  - cost supported by the consumers ($c^T q$)
  - reliability cost
- The penalty factor PF is chosen by ISONE so that the generators have a clear incentive to invest if the capacity is smaller than NICR
- There is a import zone constraint (ICZ)

Economic motivation: benefit associated with increased reliability

$$\text{price offered for a fixed } Q_{SYS}: \quad -PF \cdot \frac{\partial EENS}{\partial Q_{ICZ}}$$

Economic motivation: Investment promotion

- ISONE wants generators to invest in their infrastructure
- Cost is supported by the consumers
- No need to invest when there is already enough capacity



Working hypothesis

Assumptions on the EENS function

- $EENS(Q_{SYS}, Q_{ICZ})$ is a smooth convex function
- Cannot be represented as a quadratic function
- $\dfrac{\partial EENS(Q_{SYS}, Q_{ICZ})}{\partial Q_{ICZ}}$ is a concave function.

Desired properties of the approximate function

- amenable to efficient computation
- preserve the shape of the unknown function
- inherit smoothness property

High-level constraint

- Market participants have to agree on the process beforehand
- Optimization problem has to be solved in a few hours
- Computed price must decrease as the capacity Q_{SYS} increases

Stylised problem

Main optimization problem: MIQP

$$\min_{x,y} \quad f(x) + g(y) \quad \text{s.t.} \quad (x,y) \in P, x_i \in \{0,1\}, i \in \mathcal{I} \tag{1}$$

- f is convex
- P is convex polyhedral
- g is unknown: $g(y)$ is computed by running a long simulation
- y is in a low dimensional space

This problem has to be solved to optimality and in a few hours

Outputs from MIQP (1)

- minimizer pair (x^*, y^*)
- continuous gradient $\nabla g(y^*)$

Proposed procedure

Construct the approximate function $\hat{g}$ (offline part)

- Convex function $\hat{g}(y) := \max_i \hat{g}_i(y)$ with $\hat{g}_i(y) := a_i^T y + b_i$
- Easy to work with (computationally) but no smoothness
- Find $\hat{g}$ via its epigraph by computing an inner approximation of $\operatorname{epi} g$

Solve optimization problem (FCA) (online part)

Compute (x^*, y^*) solution to the MIQP

$$\min_{x,y} \quad f(x) + \hat{g}(y) \quad \text{s.t.} \quad (x,y) \in P, x_i \in \{0,1\}, i \in \mathcal{I} \tag{2}$$

Moreau-Yosida regularisation (online part)

- The subdifferential $\nabla \hat{g}$ is multivalued
- Compute a regularised gradient of $\hat{g}$ at the solution y^* of (2)

Piecewise-Linear (PL) $\hat{g}$: Procedure

Function construction: $\hat{g} \coloneqq \max_i \hat{g}_i$

1. Compute $g(y_i)$ for some $y_i \in Y$
2. Check the convexity assumption (via LP) on $v_i \coloneqq (y_i, g(y_i))$
3. Get the H-representation ($Hx \leq b$) from the V-representation $(\operatorname{conv} v_i)$
4. Extract $\operatorname{epi} \hat{g}$ by removing the hyperplanes forming the "lid" of $\operatorname{conv} v_i$
5. Recover the linear functions $\hat{g}_i$ from H and b.

Piecewise-Linear (PL) $\hat{g}$: Procedure

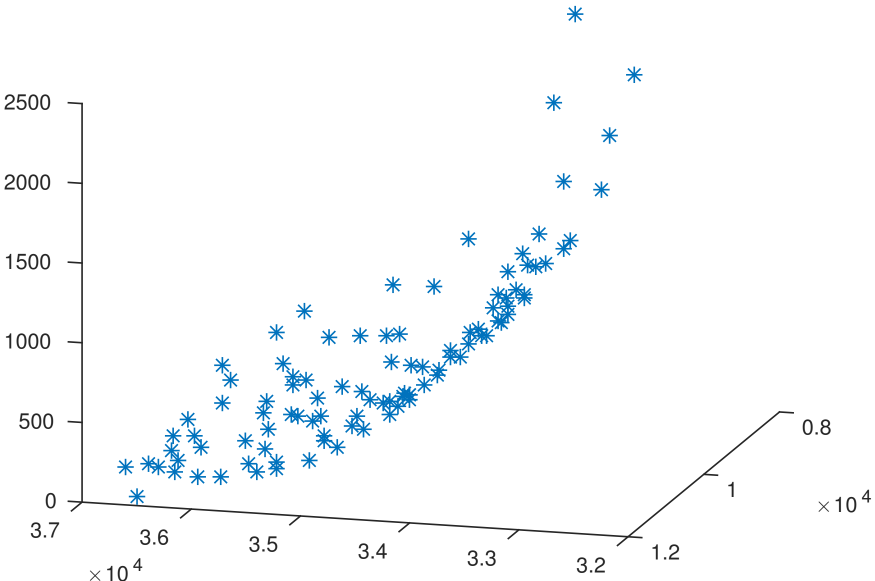
Function construction: $\hat{g} := \max_i \hat{g}_i$

1. Compute $g(y_i)$ for some $y_i \in Y$
2. Check the convexity assumption (via LP) on $v_i := (y_i, g(y_i))$
3. Get the H-representation $(Hx \leq b)$ from the V-representation $(\operatorname{conv} v_i)$
4. Extract $\operatorname{epi} \hat{g}$ by removing the hyperplanes forming the "lid" of $\operatorname{conv} v_i$
5. Recover the linear functions $\hat{g}_i$ from H and b.

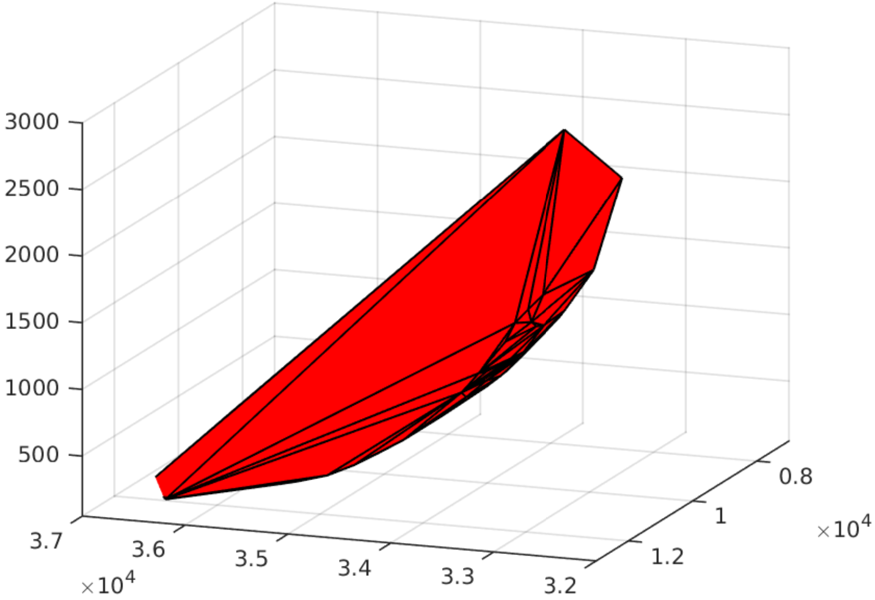
Hyperplane separation LP (Fukuda's online FAQ)

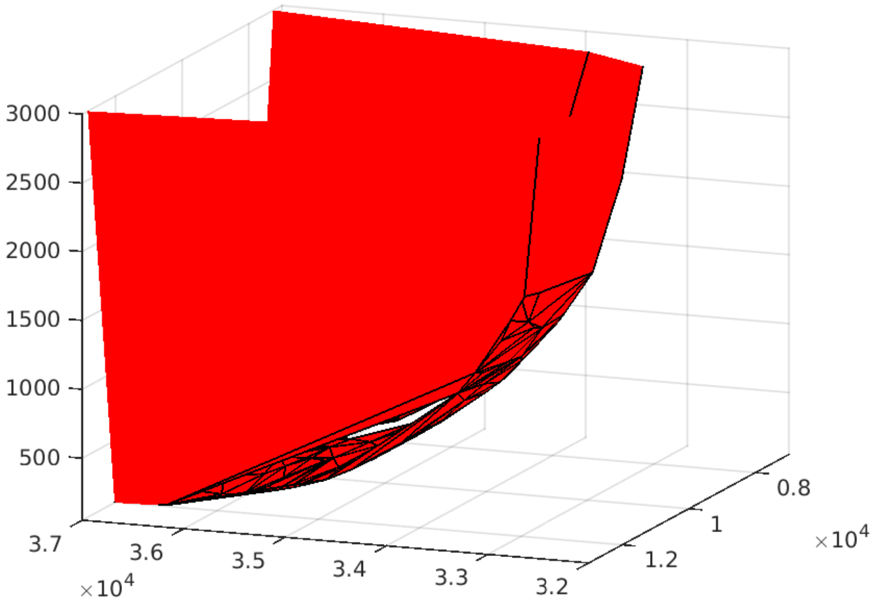
$$\begin{aligned}
\max_{h \in \mathbb{R}^{m+1},\, h_0 \in \mathbb{R}} \quad & h^T v_k - h_0 \\
\text{s.t.} \quad & h^T v_i - h_0 \leq 0 && \forall i \neq k \\
& h^T v_k - h_0 \leq 1 && \text{(boundedness of the objective value)} \\
& h^T \tilde{v}_k - h_0 \leq 0 && \tilde{v}_k := (y_k, 2g_{max}) \text{ and } g_{max} := \max_i g(y_i)
\end{aligned}$$

PL $\hat{g}$ construction: Vertices



PL $\hat{g}$ construction: $\operatorname{conv} v_i$




3000
2500
2000
1500
1000
500
3.7
3.6
3.5
3.4
3.3
3.2
$\times 10^4$
0.8
1
1.2
$\times 10^4$

Moreau-Yosida approximation: Basics

Function "level"

With $\hat{g}$ a convex function, its Moreau-Yosida approximation is defined as

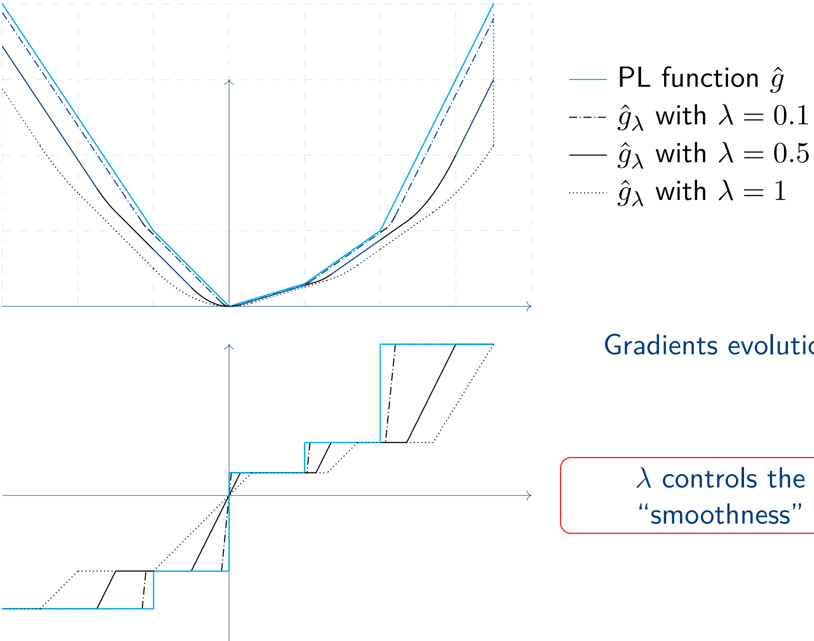
$$\tilde{g}(y) := \min_{z} \hat{g}(z) + \frac{1}{2\lambda}\|z - y\|_2^2 \tag{3}$$

- z^* unique solution to (3) is the *proximal point*
- $\tilde{g}$ is at least C^1
- $\tilde{g}$ is also convex
- Proximal point algorithm: x^{k+1} is the proximal point

Operator (subgradient) "level"

The subdifferential $\partial\hat{g}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone ($\hat{g}$ is convex)

- The regularised gradient $\nabla\tilde{g}$ is single-valued maximal monotone
- $\nabla\tilde{g} := (\lambda I + (\partial\hat{g})^{-1})^{-1}$
- $\nabla\tilde{g}(y) = \frac{1}{\lambda}(y - z^*)$



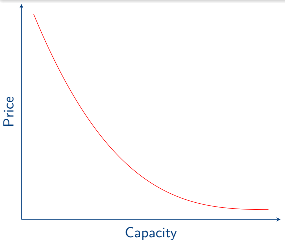
Gradients evolution

λ controls the "smoothness"

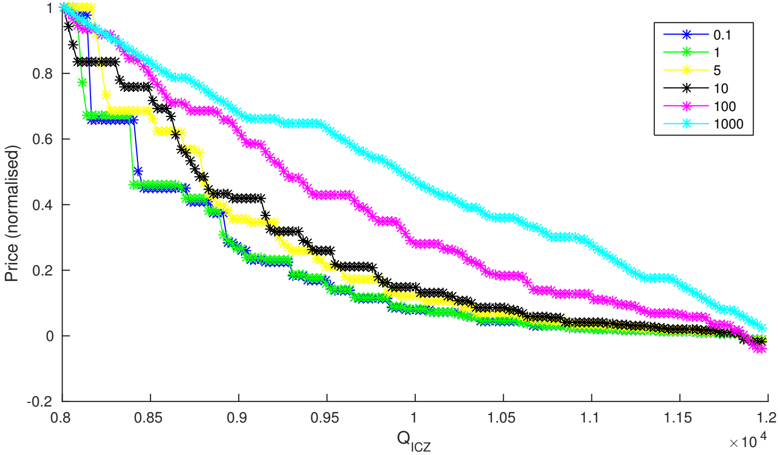
A few observations

- The parameter λ "controls" how far the proximal point will be from the point of interest.
- The gradient $\nabla\tilde{g}$ is Lipschitz: $\|\nabla\tilde{g}(y_1) - \tilde{g}(y_2)\| \leq \lambda^{-1}\|y_1 - y_2\|$. Hence the smoothing effect grows with λ.
- $\tilde{g} \equiv \min_y \hat{g}(y)$ when $\lambda \to \infty$
- With large λ, the shape of $\tilde{g}$ is close to a quadratic

- Too much information is lost with a large value of λ



Normalised price evolution for different λ



- $\frac{1}{2\lambda}\|z-y\|_2^2$ can be seen as a penalisation term
- Use a different distance, which may change with y

Bregman distance

- $h\colon D \to \mathbb{R}$, strictly convex and C^2
- $D_h(z,y) := h(z) - h(y) - \nabla h(y)^T(z-y)$ is a distance
- Example: if $h = \frac{1}{2}\|\cdot\|_2^2$, then $D_h(z,y) = \frac{1}{2}\|z-y\|_2^2$
- With additional assumptions on D_h, the function defined as

$$\tilde{g}(y) := \min_z \hat{g}(z) + \frac{1}{\lambda} D_h(z,y)$$

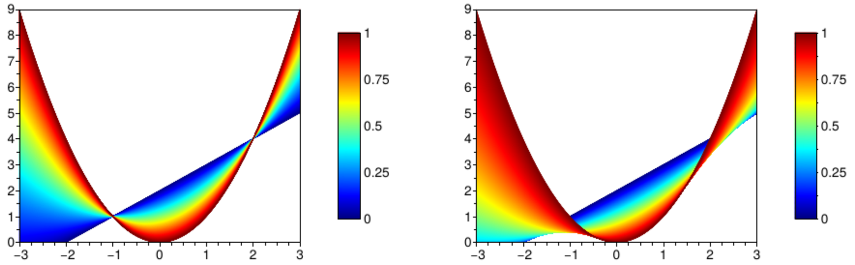
- For proximal point algorithm with Bregman distances, see [Censor & Zenios, 1992], [Eckstein 90's], [Eckstein & Silva], ...

— Give more flexibility, may better capture the shape of the function

— Quite adhoc and the problem is usually nonlinear

Proximal average: Homotopy between epigraphs

- Proximal average $\mathcal{P}(f_0, f_1, \mu)$ is a continuous transformation between 2 convex functions f_0 and f_1
- $\mathcal{P}(f_0, f_1, \mu)(x) := -\min_z -\mu\tilde{f}_0(z) - (1-\mu)\tilde{f}_1(z) + \frac{1}{2\lambda}\|z - x\|_2^2$
- With $\tilde{f}_0(z)$ and $\tilde{f}_1(z)$ the Moreau envelopes with parameter λ



Averages of $f_0(x) = x + 2$ and the quadratic function $f_1(x) = x^2$: Arithmetic (left) and proximal (right). [Bauschke, Lucet, Trienis, 2007]

Moreau-Yosida approximation: proximal average

Motivations

- If we have an under estimator $\hat{g}$ and over estimator $\bar{g}$, the function g is "in between".
- Also use this information in the regularisation

Procedure

- $\hat{g}$ computed as before as a subset of $\operatorname{epi} g$
- Compute $\bar{g}$ via an outer approximation of $\operatorname{epi} g$: supporting hyperplanes at vertices
- Compute proximal average instead of the Moreau-Yosida approximation

Perspectives

Noisy evaluation of g

- Function values $g(y_i)$ may be noisy but are completely trusted
- May loose convexity, gives back wrong gradients
- Noise effects are mostly local, except for points on the boundaries of Y
- Idea: smoothing via local convex quadratic fit

Couple the MIQP and the Moreau-Yosida approximation

- Currently an MIQP is solved and then the regularised gradient is computed
- The two could be merged (with classical Moreau-Yosida approximation)

Extension to other instances

- Apply this approach to other types of problem
- Use this penalty based method to get an approximated solution vs solving exact problem

Conclusion

Context

- "Nice" (convex) optimization problem but with partially known objective function
- Computationally effective and retain lots of features (convexity, shape)

Approach presented

- Use a convex piecewise-linear function $\hat{g}$ because it is convex and the fitting is easy
- The differentiability property is obtained afterward via the Moreau-Yosida approximation