Some remarks on computation: three cases where structure counts

Michael C. Ferris

University of Wisconsin, Madison

IPAM Workshop
Los Angeles
January 14, 2016

I: Find a solution to the MOPEC:

$$\underset{q^i}{\text{minimize}} \quad \theta_i(q^i, q^{-i}) \quad \text{subject to} \quad q^i \in D(q^{-i})$$

where

$$\begin{aligned} q &= (q^0, q^1, \ldots, q^{N_a}), \text{ for } i = 1, \ldots, N_a, \\ q^{-i} &= (q^0, q^1, \ldots, q^{i-1}, q^{i+1}, \ldots, q^{N_a}), \\ D(q^{-i}) &= [0, U^i] \cap \{q^i \mid \sum_{j=0}^{N_a} q^j = d\}, \\ \theta_0(q^0, q^{-0}) &= Pq^0 + \sum_{i=1}^{N_a} c^i(q^i) - p\left(\sum_{i=1}^{N_a} q^i\right)\left(\sum_{i=1}^{N_a} q^i\right), \\ \theta_i(q^i, q^{-i}) &= c^i(q^i) - p\left(\sum_{i=1}^{N_a} q^i\right)(q^i), \text{ for } i = 1, \ldots, N_a \end{aligned}$$

Equivalent (polyhedrally constrained) Variational Inequality

Solvable as $\mathrm{VI}(K, F)$, where

$$K = \prod_{i=0}^{N_a} [0, U^i] \cap \{q \mid \sum_{j=0}^{N_a} q^j = 0\}$$

$$F(q) = \left(\nabla_{q^0}\theta_0(q), \nabla_{q^1}\theta_1(q), \ldots, \nabla_{q^{N_a}}\theta_{N_a}(q)\right)^T$$

Key observation: for $q \in K$ we have

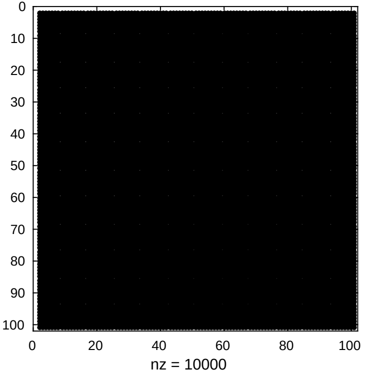
$$d - q^0 = \sum_{i=1}^{N_a} q^i$$

Use this to substitute out expression in F. This defines $\tilde{F}(q)$.

Theorem

q^* is a solution to $\mathrm{VI}(K, F)$ if and only if it is a solution to $\mathrm{VI}(K, \tilde{F})$.

Nonzero patterns of the Jacobian matrix depending on its VI formulation when $n = 100$



Elapsed time of PATH according to formulations

Size (n)	Elapsed time (secs)	
	Original formulation	Reformulation
2,500	48.431	0.696
5,000	570.214	1.408
10,000		2.780
50,000		17.856
100,000		41.440

Summary and issues

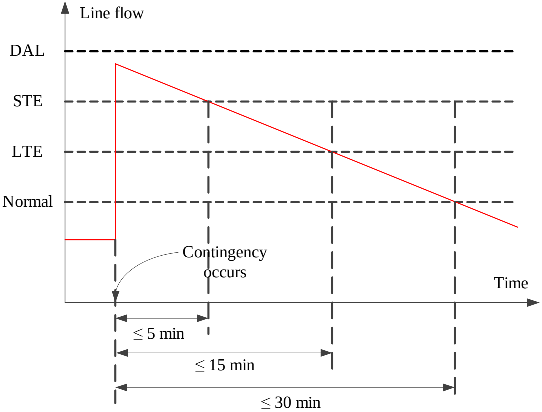
- Can solve much larger instances without need for specialized algorithms
- But, now can extend to stochastic setting (each VI involves Stochastic Program)
  - Model becomes much larger (and sparser)
  - Pivotal method within PATH becomes bottleneck as scenarios increase
- Need for smoothing and/or decomposition - enhanced interface between modeler and solver

II: Security-constrained Economic Dispatch

- Base-case network topology g_0 and line flow x_0.
- If the k-th line fails, line flow jumps to x_k in new topology g_k.
- Ensure that x_k is within limit, for all k.
- SCED model:

$$
\begin{array}{lll}
\min\limits_{u,x_0,\ldots,x_k} & c^T u + \rho(u) & \triangleright \text{ Total cost} \\
\text{s.t.} & 0 \leq u \leq \bar{u} & \triangleright \text{ GEN capacity const.} \\
 & g_0(x_0, u) = 0 & \triangleright \text{Base-case network eqn.} \\
 & -\bar{x} \leq x_0 \leq \bar{x} & \triangleright \text{Base-case flow limit} \\
 & g_k(x_k, u) = 0, \quad k = 1, \ldots, K & \triangleright \text{Ctgcy network eqn.} \\
 & -\bar{x} \leq x_k \leq \bar{x}, \quad k = 1, \ldots, K & \triangleright \text{Ctgcy flow limit}
\end{array}
$$

Reality offers a sweeter deal...



Operating procedure (ISO-NE) requires post-contingency line loadings be:

- $\leq$ STE (short time emergency) rating in 5 minutes;
- $\leq$ LTE (long time emergency) rating in 15 minutes;
- $\leq$ Normal rating in 30 minutes.

What we will contribute

Research issues:

- Corrective actions are not modeled in ISO's dispatch software.
- Because it was "insolvable" due to its large size ($\geq$ 10GB LP).
  - "We looked into SCED with corrective actions before, and were hindered by the computational challenge." – Feng Zhao, senior analyst at ISO-NE, via private correspondence.

Our contributions:

- We **model** the *multi-period* corrective rescheduling in SCED; solutions much better quality
- **Enhance** the Benders' **algorithm** to solve the problem faster
- **Achieve** about 50$\times$ **speedup** compared to traditional approaches

Our model (K contingencies, T periods)

$$
\begin{array}{lll}
\min\limits_{x_0,\ldots,x_k,u_0,\ldots,u_k} & c^T u_0 & \\
\text{s.t.} & g_0(x_0, u_0) = 0 & \\
& h_0(x_0, u_0) \le 0 & \\
& g_k(x_k^t, u_k^t) = 0 & k = 1, \ldots, K,\ t = 0, \ldots, T \\
& h_k(x_k^t, u_k^t) \le 0 & k = 1, \ldots, K,\ t = 0, \ldots, T \\
& |u_k^t - u_k^{t-1}| \le \Delta_t & k = 1, \ldots, K,\ t = 1, \ldots, T \\
& u_k^0 - u_0 = 0 & k = 1, \ldots, K
\end{array}
$$

- Subscript 0 indicates a quantity in the base-case network topology.
- This is a large-scale linear program.
- What special structure does it have?

Model structure

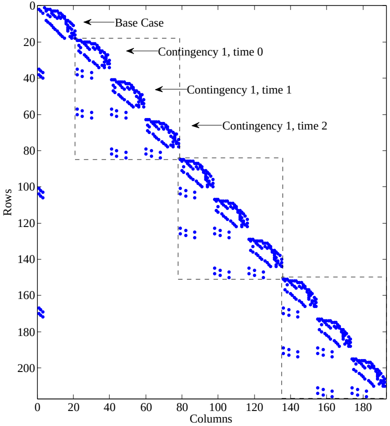


Figure : Sparsity structure of the Jacobian matrix of a 6-bus case, considering 3 contingencies and 3 post-contingency checkpoints.

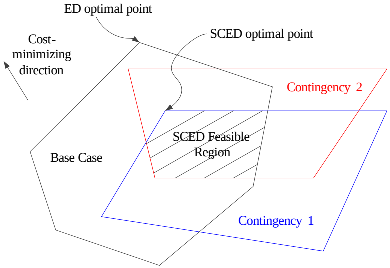


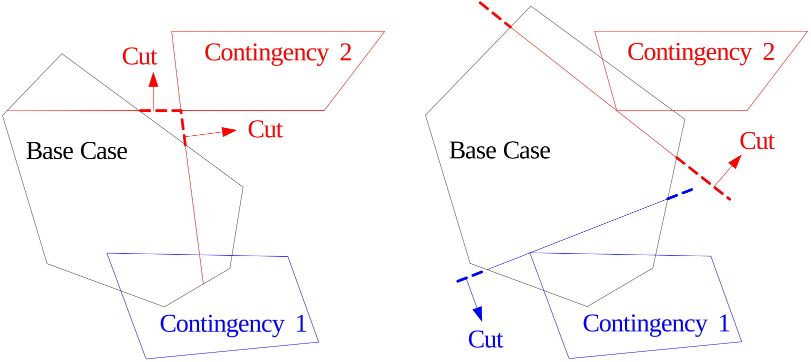
Figure : On the u_0 plane, the feasible region of a SCED is the intersection of K+1 polyhedra.

How we enhanced the Benders' algorithm ...

1. Reduce the number of LPs
2. Solve subproblem LPs faster
3. Parallel computing
4. Add difficult contingencies to master model

Case	Ctgcy	Big LP (time)		**Enhanced Benders**		
		Simplex	Barrier	Iter	LPs	Time
118-bus	183	207.8	13.8	12	755	13.5
2383-bus	20	175.0	205.5	11	60	41.5
2383-bus	50	1403	123.1	11	135	46.5
2383-bus	100	3621	240.6	12	245	79.4
2383-bus	400	-	2354.5	13	879	197.8
2383 wp	2349			21	9529	515.7
2736 sp	2749			4	5500	220.9
2737 sop	2753			1	2753	100.5
2746 wop	2794			1	2794	118.5
2746 wp	2719			14	5558	333.5

Dealing with Infeasibility



(a) Contingency 2 is intrinsically infeasible. Either the corresponding subproblem is infeasible or its Benders' cuts will render the master problem infeasible.

(b) Each individual contingency is feasible, but they are not simultaneously feasible. Their Benders' cuts will render the master problem infeasible.

Figure : Two cases of infeasibility.

Identifying infeasible contingencies in Benders' algorithm

- If a subproblem is infeasible (in the first iteration), the corresponding contingency is intrinsically infeasible. Remove (tabu) it.
  - Typically line failure results in an islanded load node or sub-network.
- Master problem infeasible: solve a modified master model to find the "minimal" set of problematic contingencies using sparse optimization.

$$
\begin{aligned}
\min_{x_0,u_0} \quad & f_0(x_0,u_0) + \sum_{k\in K} M v_k \\
\text{s.t.} \quad & g_0(x_0,u_0) = 0, h_0(x_0,u_0) \le 0 \\
& \bar{w}_k^i + \bar{\lambda}_k^i(u_0 - \bar{u}_0^i) - v_k \le 0, v_k \ge 0 \qquad \forall (k,i) \in \text{CUT}
\end{aligned}
$$

  - Solution of this model indicates the violated cuts.
  - Tabu the contingency that has contributed one or more violated cuts.
- Start a pre-screening daemon in parallel when the Active List size is smaller than L^{fc}.
  - Tabu infeasible ones, and add feasible ones to the master problem.

Computational Results

Table : Solution for big cases on opt-a006, 80 threads, $L^{fc} = 5$

Case	Ctgcy	Iter	LPs	Time	To Master	Tabu
2383 wp	2896	15	7694	522.1	6	547
2736 sp	3269	4	6020	252.9	1	520
2737 sop	3269	4	6023	242.2	0	516
2746 wop	3307	4	6102	280.2	0	513
2746 wp	3279	8	6053	334.3	4	560
2383 wp	2353	16	7156	460.6	6	4
2736 sp	2749	4	5498	245.9	1	0
2737 sop	2753	1	2753	110.8	0	0
2746 wop	2794	1	2794	131.7	0	0
2746 wp	2719	14	5558	354.4	4	0

- Upper: all lines are in the Contingency List (N-1 security).
- Lower: all pre-screened lines are in the Contingency List.

SCED with SDP subproblems

- Economic dispatch is a short-term planning problem, so a “DC” model is OK.
- Contingency response is an operational problem, and should be studied on full AC network representation.
- But AC power flow gives a nonconvex problem, which cannot generate valid cuts from a Benders’ subproblem.

Idea

Relaxing the AC feasibility problem using semi-definite programming (SDP) to obtain a convex subproblem.

Goal

Producing a base-case dispatch solution such that contingencies are “really” controllable in the AC context.

SDP relaxation of AC feasibility problem

Model ACF-SDP:

$$
\begin{array}{lll}
\min\limits_{W \succeq 0} & A_0 \bullet W & \\
\text{s.t.} & \sum\limits_{g \in \mathcal{G}_i} \underline{G}_g^{\text{real}} - D_i^{\text{real}} \leq A_{1i} \bullet W \leq \sum\limits_{g \in \mathcal{G}_i} \bar{G}_g^{\text{real}} - D_i^{\text{real}} & \forall i \in BUS \\
& \sum\limits_{g \in \mathcal{G}_i} \underline{G}_g^{\text{imag}} - D_i^{\text{imag}} \leq A_{2i} \bullet W \leq \sum\limits_{g \in \mathcal{G}_i} \bar{G}_g^{\text{imag}} - D_i^{\text{imag}} & \forall i \in BUS \\
& -\bar{F}_{i,j} \leq A_{3ij} \bullet W \leq \bar{F}_{i,j} & \forall (i,j) \in LINE \\
& (\underline{V}_i)^2 \leq A_{4i} \bullet W \leq (\bar{V}_i)^2 & \forall i \in BUS \\
& \sum\limits_{g \in \mathcal{G}_i} (G_g^0 - \Delta_g) \leq A_{5i} \bullet W \leq \sum\limits_{g \in \mathcal{G}_i} (G_g^0 + \Delta_g) & \forall i \in BUS
\end{array}
$$

- It is a convex optimization problem and weak (strong) duality holds.
- It is a relaxation because the requirement that W has rank 1 is dropped.

Experiments

Case	Cont	Solution				Performance		
		Model	Tabu	Cost	Time	IF	FS	FT
14	20	LP	0	13253.3	4.2	12	12	0
		SDP	6	16065.8	18.4	6	0	0
		SDP0	6	16003.4	11.9	6	0	0
30	40	LP	0	582.0	4.0	1	1	0
		SDP	1	585.0	20.1	1	0	0
		SDP0	1	600.5	22.1	1	0	0
57	20	LP	0	12508.0	1.9	1	1	0
		SDP	1	12508.0	13.2	1	0	0
		SDP0	1	12560.0	50.9	1	0	0
118	15	LP	0	139716.8	54.0	16	16	0
		SDP	0	141372.2	2414.3	1	1	0
		SDP0	0	144220.1	11951.1	0	0	0

- SDP subproblem is “exact” in contingency response, no False Secure, no False Tabu.
- It takes longer time to solve (with room for improvement).

Summary

1. SCED is a million-dollar problem for system operators.
2. SCED with corrective actions can save money, but is hard to solve.
3. Our algorithmic enhancements yield significant speedup.
4. Potential for practical deployment.
5. SDP extension allows for more accurate operational modeling.

Extension

1. Algorithm is deployed at ISO-NE (using DC with loss adjustment).
2. Need enhancements to SDP solvers to make ACOPF version practical.

III: Forward Capacity Expansion

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

ISONE's response: Forward capacity

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
- ≈ 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

FCA optimization problem

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)

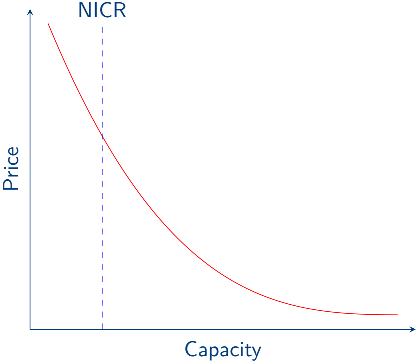
Price formation

Economic motivation: benefit associated with increased reliability

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

Main optimization problem: MIQP

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

- 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

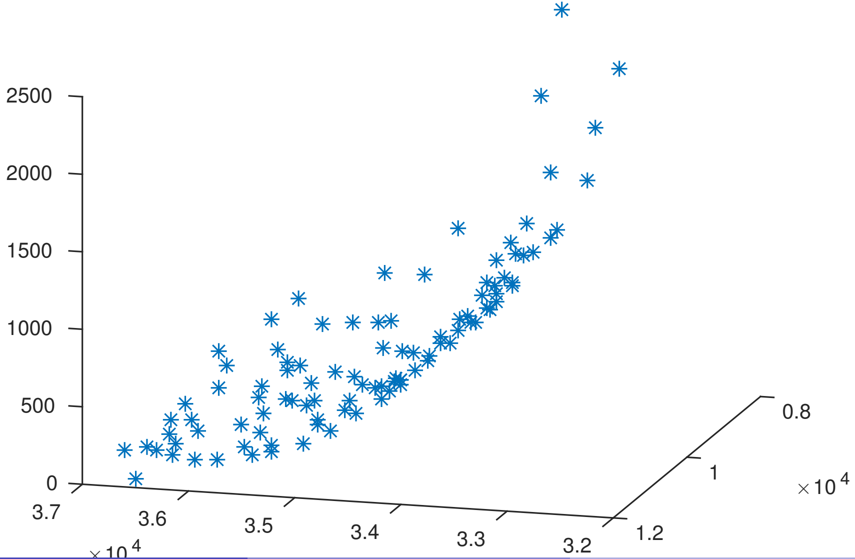
Function construction: $\hat{g} := \max_i \hat{g}_i$

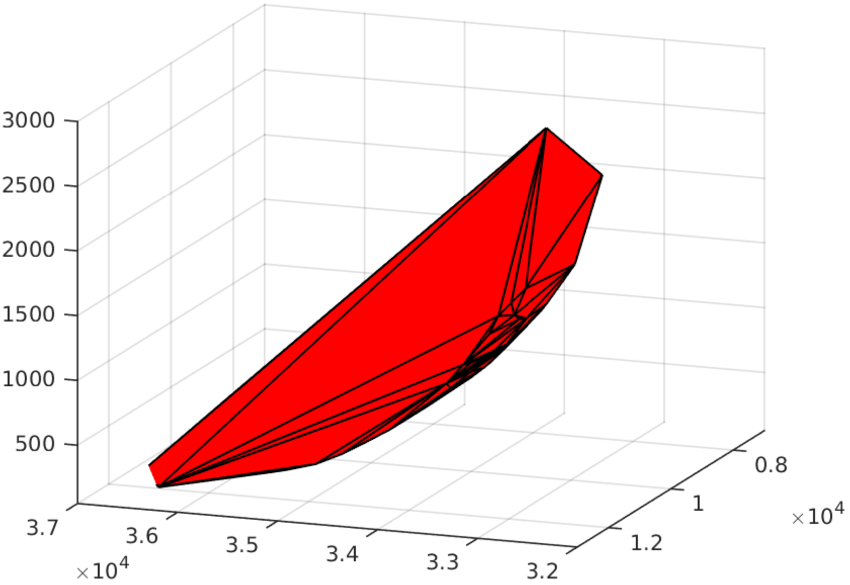
1. Compute $g(y_i)$ for some $y_i \in Y$
2. Check the convexity assumption (via LP) on $v_i := (y_i, g(y_i))$
3. Get the H-representation ($Hx \leq b$) from the V-representation ($\mathrm{conv}\, v_i$)
4. Extract $\mathrm{epi}\hat{g}$ by removing the hyperplanes forming the "lid" of $\mathrm{conv}\,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

Hyperplane separation LP

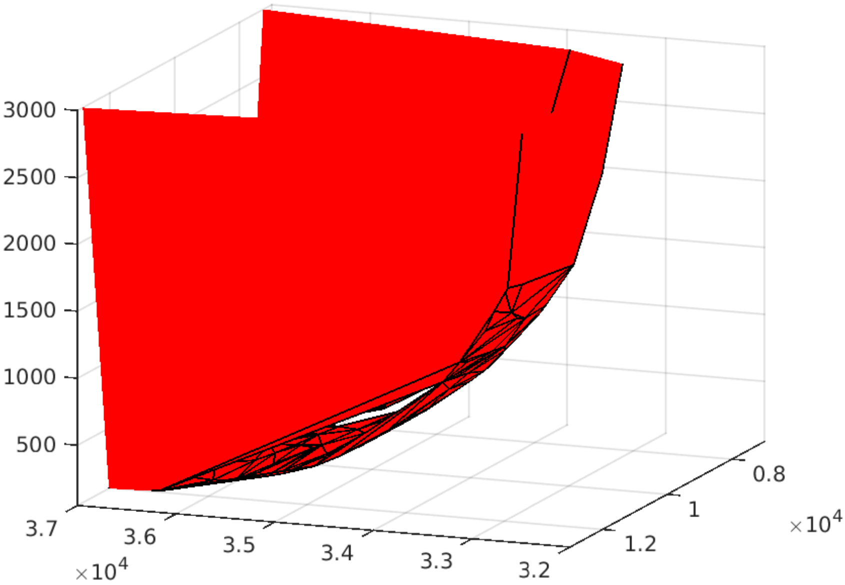
$$
\begin{aligned}
\max_{h \in \mathbb{R}^{m+1},\, h_0 \in \mathbb{R}} \quad & h^T v_k - h_0 \\
\text{s.t.} \quad & h^T v_i - h_0 \leq 0 && \forall i \neq k \\
& h^T v_k - h_0 \leq 1 && \text{(boundedness of the objective value)} \\
& h^T \tilde{v}_k - h_0 \leq 0 && \tilde{v}_k := (y_k, 2g_{max}) \text{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

PL $\hat{g}$ construction: Vertices




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

PL $\hat{g}$ construction: $\operatorname{epi} \hat{g}$



Function "level"

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

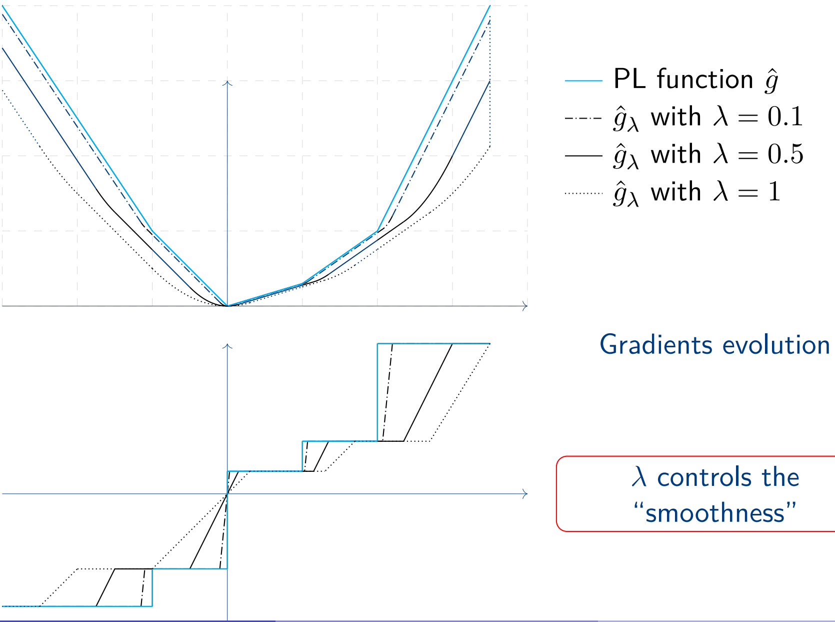
$$\tilde{g}(y) := \min_z \hat{g}(z) + \frac{1}{2\lambda}\|z - y\|_2^2 \tag{3}$$

- 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"

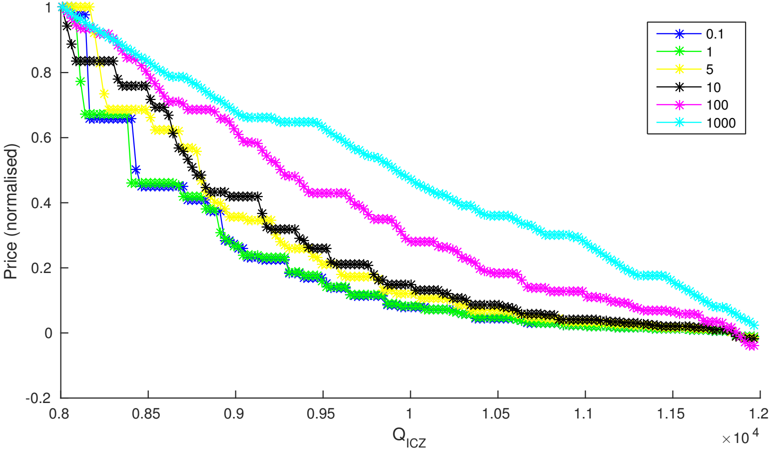
The subdifferential $\partial\hat{g}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone ($\hat{g}$ is convex)

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



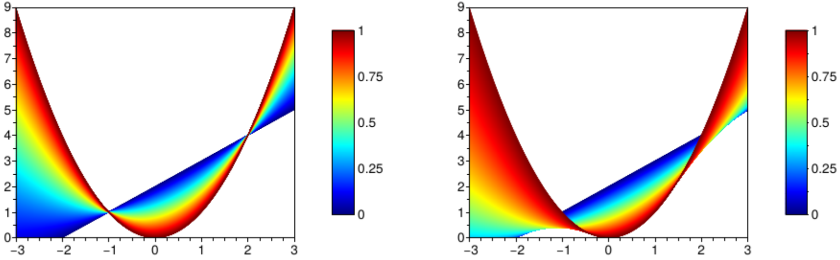
λ controls the "smoothness"

Normalised price evolution for different λ



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 $\mathrm{epi} g$
- Compute $\bar{g}$ via an outer approximation of $\mathrm{epi} g$: supporting hyperplanes at vertices
- Compute proximal average instead of the Moreau-Yosida approximation

Conclusions and remarks

Use of simulations

- Build appropriate approximating models (understand how and where they are to be used)
- Can use adaptive (derivative free) codes for similar problems but limited to small scale problems in design space

Future needs

- Move from black-box approaches to structure exploiting approaches with "subproblems" linked by good theory
- Enhance "subsolvers" to facilitate large scale, global, robust solutions
- Facilitate domain specific expertise to enhance solution efficiency: starting point generation, homotopy
- Build tools to allow algorithmic development in higher level modeling systems