Plenary Lecture

Measuring smart grids[1]

Leo Liberti*
CNRS LIX, Ecole Polytechnique, 91128 Palaiseau, France,
`liberti@lix.polytechnique.fr`

Claudia D'Ambrosio
CNRS LIX, Ecole Polytechnique, 91128 Palaiseau, France,
`dambrosio@lix.polytechnique.fr`

Pierre-Louis Poirion
CNRS LIX, Ecole Polytechnique, 91128 Palaiseau, France,
`poirion@lix.polytechnique.fr`

Sonia Toubaline
CNRS LIX, Ecole Polytechnique, 91128 Palaiseau, France,
`poirion@lix.polytechnique.fr`

Abstract. The SO-grid project aims to develop a smart grid pilot for the French electricity distribution network. Our workpackage is concerned with the efficient placement of measuring devices on the network. We give an overview of our progress so far: combinatorial problems and variants, complexity, inapproximability, polynomial cases, single- and bilevel mathematical programming formulations, row-generation algorithms, and generalization to a large class of bilevel MILPs.

Introduction

The three defining properties of a smart grid are reliability, sustainability and value. Reliability is the main motivation for making electrical grids smart: when it was still possible to increase network capacity to make it reliable, electrical grids were not termed *smart*. As soon as capacity increase became infeasible, we started noticing unstable behaviours, such as large-scale blackouts due to a single point of failure, or to a short but high demand peak. The only possible way out, when capacity is maximum, is to address the demand. By installing an Advanced Metering Infrastructure (AMI), it became possible to define new policies based on Demand Response (DR), mainly by predictive and prescriptive analytics based on the continuing measurements of electrical quantities such as voltage and current. One of the most important DR tools these analytics can help improve is pricing. In parallel, renewable source of energies such and sun and wind power were integrated in the power grid, addressing some sustainability issues. And finally, in capitalistic societies the private sector's only motive for investing into such a costly transformation of the power grid is increased value (another feature which pricing can help leverage).

[1]This work was carried out as part of the SO-grid project (www.so-grid.com), co-funded by the French agency for Environment and Energy Management (ADEME) and developed in collaboration between participating academic and industrial partners.

The type of analytics we discuss here is prescriptive, and concerns the optimal placement of certain measuring devices on the electrical grid. Although we are not at liberty to disclose the exact specification or nature of such devices, for the purposes and the intended audience of this paper we shall assume such devices to be Phasor Measurement Units (PMU), which measure the voltage at the node v of the grid they are installed at, as well as at a varying number of nodes in its neighbouring set $N(v)$ [7]. Ohm's law applied to a link $\{u, v\} \in E$ of the grid (represented by a graph $G = (V, E)$),

$$\mathcal{V}_v - \mathcal{V}_u = R\,\mathcal{I}_{uv},$$

allows the measure of the current along the link if the voltage is known at the endpoints. Moreover, by Kirchoff's law,

$$\forall v \in V \quad \sum_{\{u,v\}\in E} \mathcal{I}_{uv} = 0,$$

it follows that, if $\mathcal{V}_v$ is known and, for all u in the neighbourhood $N(v)$ *except for at most one* (call it w), $\mathcal{V}_u$ is also known, then $\mathcal{V}_u$ is known for all $u \in N(v)$, including w. It suffices to exploit Kirchoff's law to compute the current along the link $\{v, w\}$, and then Ohm's law to compute $\mathcal{V}_w$.

We call *observed* the nodes with known voltage and the links with known currents. As shown above, nodes and links can be observed directly (because of PMUs) or indirectly, through the application of Ohm's and Kirchoff's laws. Since PMUs can be expensive, the general problem we are concerned with is to find the placement of the smallest number of PMUs which allows the grid to achieve the state of *full observability*, i.e. such that all nodes and links of the grid are observed.

PMUs may have a limited number k of observation channels, meaning that if a PMU is placed at $v \in V$, it measures the voltage at v and at up to k nodes in $N(v)$. When $k = 1$, a PMU installed at v can measure the voltage at v and at a single neighbour u. Since, by Ohm's law, the current on $\{u, v\}$ is known, this situation is logically equivalent at placing the PMU on the edge $\{u, v\}$ and measuring the current. In the rest of this paper, we mostly focus on PMUs with $k = 1$.

Observability

As mentioned above, the aim is to find the smallest P such that the set of observed nodes Ω is equal to V. If we use PMUs with $k = 1$, by Ohm's law we have:

$$\textbf{R1} \qquad \forall \{u, v\} \in P \ (u, v \in \Omega).$$

Moreover, by Kirchoff's law (as above) we also have:

$$\textbf{R2} \qquad \forall v \in V \ (v \in \Omega \wedge |N(v) \smallsetminus \Omega| \le 1 \to N(v) \subseteq \Omega).$$

Rules **R1** and **R2** are known as *observability rules*, and yield an iterative procedure to achieve full observability of the grid: from a given P, one applies **R1** first (to all links in P) and **R2** for as many times as possible. If $\Omega = V$, P is a *feasible placement*, and otherwise it is *infeasible*.

Combinatorial problems

All problem variants aim at minimizing $|P|$ in order to achieve $\Omega = V$. The name of the problem and its exact definition depends on k. For unlimited k, the problem is known as Power Dominating Set (PDS) [1, 2]. For bounded k, the problem is known as k-PMU placement problem. The 1-PMU problem is also known as Power Edge Set (PES), and the 0-PMU problem is also known as Power Vertex Set (PVS). While PDS, k-PMU and PES can all be applied to the setting of electrical grids, the PVS has an application to influence spreading in social networks [5]. For all variants but the PES, $P \subseteq V$, whereas for PES (which is the case we shall consider) we have $k = 1$ and consequently $P \subseteq E$, as discussed above.

The PDS is known to be **NP**-complete on bipartite graphs, planar bipartite graphs and cographs, and polynomial on trees, meshes, block graphs and bounded treewidth graphs. An $O(\sqrt{n})$ approximation algorithm is known for the PDS (where $n = |V|$), but it is hard to approximate it to within a factor $2^{\log^{1-\varepsilon} n}$.

Hardness and inapproximability

We show that PES is **NP**-hard by reduction from the Vertex Cover problem on 3-regular graphs (3VC). We transform a given graph instance G' of 3VC into a graph G'' where each vertex v of G' is mapped to a 10-vertex gadget in G'': the hardness proof is long but elementary. A similar proof shows hardness of approximation to within a factor $1.12 - \varepsilon$ for any $\varepsilon > 0$ (we can also prove **NP**-hardness for the PVS). Moreover, we show that the PES is polynomial on trees (by reduction to Path Cover on trees) and meshes.

An iteration-indexed Mathematical Program

Our first test towards solving PES instances was to model the iterative observability procedure mentioned above within a Mixed-Integer Linear Program (MILP) using binary indicator variables:

- $s_{uv} = 1$ iff $\{u, v\} \in P$
- $\omega_{vt} = 1$ if $v \in V$ enters the set Ω during iteration t of the observability procedure.

Note that ω_{vt} are iteration-indexed variables, and considerably increase the problem size. The objective function $\sum\limits_{\{u,v\} \in E} s_{uv}$ minimizes $|P|$, and one of the constraints requires that $\Omega = V$ at the last iteration $T_{\max}$ (which is $O(n)$). The constraints implement **R1** and **R2** in a straightforward way. This formulation turned out to be computationally unsatisfactory using CPLEX 12.6 on graphs with up to 30 nodes.

Fixed point reformulation

Although PES observability is naturally modelled by an iterative procedure, it is nonetheless of a monotonic nature. A node is either observed or not; once in Ω, no node can exit the set during the iteration procedure. This idea led us to describe the observability iteration applied to a vertex $v \in V$ by means of the following function of ω, parametrized on s (the indicator vector of the PMU placement):

$$\theta_v^s(\omega) = \max\left(\sum_{u\in N(v)} s_{uv},\ \max_{u\in N(v)} \left\{ \omega_u + \sum_{\substack{w\in N(u)\\ w\neq v}} \omega_w - (|N(u)|-1) \right\} \right). \tag{1}$$

Note that in Eq. (1) ω is independent of the iteration index t: essentially, it describes the evolution of ω at a given iteration: if ω is the indicator vector of Ω before the iteration takes place, $\theta^s(\omega) = (\theta_v^s(\omega) \mid v \in V)$ is the indicator vector of Ω after the iteration. By monotonicity, we can express the end of the iteration procedure by means of a convergence to a fixed point: $\omega = \theta^s(\omega)$.

The iteration-indexed MILP can now be re-written as the following nonlinear problem (where $m = |E|$):

$$\left.\begin{array}{rrcl} \min\limits_{\substack{s\in\{0,1\}^m\\ \omega\in\{0,1\}^n}} & \sum\limits_{\{u,v\}\in E} s_{uv} & & \\ & \sum\limits_{v\in V} \omega_v & = & n \\ & \theta^s(\omega) & = & \omega, \end{array}\right\} \tag{2}$$

which has fewer decision variables than the MILP described above since there are no iteration indexed variables. Note that the first constraint expresses $\Omega = V$, and the nonlinearity arises because of the function θ^s defined in Eq. (1).

Next, we remark that the fixed point condition on θ^s expresses, by means of the ω variables, the smallest set of vertices obeying rules **R1** (encoded in the left term of the outer maximum in Eq. (1)) and **R2** (encoded in the right term of the outer maximum in Eq. (1)). As such, it can be written as the following MILP, where the variables s are now to be considered as problem parameters:

$$\left.\begin{array}{rrcl} \min\limits_{\omega\in\{0,1\}^n} & \sum\limits_{v\in V} \omega_v & & \\ \forall\{u,v\}\in E & \omega_v & \geq & s_{uv} \\ \forall v\in V, u\in N(v) & \omega_v & \geq & \omega_u + \sum\limits_{\substack{w\in N(u)\\ w\neq v}} \omega_w - (|N(u)|-1). \end{array}\right\} \tag{3}$$

Finally, we replace the fixed point condition in Eq. (2) by the MILP in Eq. (3) to obtain the following bilevel MILP (BMILP):

$$\left.\begin{array}{ll} \min\limits_{s\in\{0,1\}^m} & \sum\limits_{\{u,v\}\in E} s_{uv} \\ & f(s) = n \\ & f(s) = \left\{\begin{array}{rrcl} \min\limits_{\omega\in\{0,1\}^n} & \sum\limits_{v\in V} \omega_v & & \\ \forall\{u,v\}\in E & \omega_v & \geq & s_{uv} \\ \forall v\in V, u\in N(v) & \omega_v & \geq & \omega_u + \sum\limits_{\substack{w\in N(u)\\ w\neq v}} \omega_w - (|N(u)|-1). \end{array}\right. \end{array}\right\} \tag{4}$$

Single-level reformulation

In general, BMILPs are practically even harder to solve than MILPs, so Eq. (4), by itself, is not much of a gain. On the other hand, we can prove, by induction on the steps of the observability iteration procedure, that the integrality constraints in the lower-level problem can be relaxed to $\omega \geq 0$ without changing the optimal solution. This implies that the lower-level problem can be replaced by the KKT conditions [4] of its continuous relaxation to the nonnegative orthant, yielding the following single-level MINLP:

$$\left.\begin{array}{rcl} & \min\limits_{\substack{s\in\{0,1\}^m\\ \lambda,\mu\geq 0}} \sum\limits_{\{u,v\}\in E} s_{uv} & \\ \sum\limits_{\{u,v\}\in E} (s_{uv}\mu_{uv} + (1-|N(u)|)\lambda_{uv}) & \geq & n \\ \forall u\in V \quad \sum\limits_{v\in N(u)} (\mu_{uv} + \lambda_{uv} - \lambda_{vu} - \sum\limits_{\substack{w\in N(u)\\ w\neq v}} \lambda_{wv}) & \leq & 1, \end{array}\right\} \tag{5}$$

where μ, λ are dual variables of the relaxed lower-level problem. We can also prove that the variables μ are bounded above (though the bound M is exponentially large), so the bilinear products $s_{uv}\mu_{uv}$ appearing in Eq. (5) can be linearized exactly using Fortet's reformulation: each product is then replaced by an additional variable p_{uv}. Finally, this yields the following single-level MILP:

$$\left.\begin{array}{rcl} & \min\limits_{\substack{s\in\{0,1\}^m\\ \lambda,\mu\in[0,M]}} \sum\limits_{\{u,v\}\in E} s_{uv} & \\ \sum\limits_{\{u,v\}\in E} (p_{uv} + (1-|N(u)|)\lambda_{uv}) & \geq & n \\ \forall u\in V \quad \sum\limits_{v\in N(u)} (\mu_{uv} + \lambda_{uv} - \lambda_{vu} - \sum\limits_{\substack{w\in N(u)\\ w\neq v}} \lambda_{wv}) & \leq & 1 \\ \forall \{u,v\}\in E \quad p_{uv} & \leq & M s_{uv} \\ \forall \{u,v\}\in E \quad p_{uv} & \leq & \mu_{uv} \\ \forall \{u,v\}\in E \quad p_{uv} & \geq & \mu_{uv} - M(1-s_{uv}). \end{array}\right\} \tag{6}$$

Eq. (6) yields computational improvements of around 1-2 orders of magnitude w.r.t. the iteration-indexed MILP. Its weakest point is that the lower-level problem Eq. (3) cannot easily be changed, lest we should lose its integrality property, which is key to reformulating the bilevel problem to a single level one. The SO-grid application, however, requires us to impose some robustness constraints $\Upsilon\omega \leq \xi$ to the lower level problem. This means we can no longer reformulate Eq. (4) to Eq. (5); and, ultimately, that we need a different solution approach.

Solving the bilevel problem directly

It is well known that every subset S of vertices of the hypercube (in any dimension) can be described linearly, i.e. the convex hull of the vertices in S does not contain any other hypercube vertex aside from those in S. We apply this fact to the feasible region

$$\mathcal{F} = \{s \in \{0,1\}^m \mid f(s) = n\}$$

of the bilevel problem Eq. (4), obtaining the single level problem:

$$\min\big\{ \sum_{\{u,v\}\in E} s_{uv} \mid s \in \mathsf{conv}(\mathcal{F}) \cap \{0,1\}^m \big\}.$$

Since we do not know a polyhedral description of $\mathcal{F}$, we look for a polyhedron $\mathcal{P}$ having the same intersection with the hypercube as $\mathcal{F}$, and aim at solving:

$$\min\{\sum_{\{u,v\}\in E} s_{uv} \mid s \in \mathcal{P} \cap \{0,1\}^m\} \tag{7}$$

using a row generation algorithm [3, 6].

Inequalities for $\mathcal{P}$

Here we write the set $\Omega(s)$ of observed nodes in function of the PMU placement encoded by the indicator vector s. Suppose a placement s is given such that $\Omega(s) \subsetneq V$. To reach full observability we must install a PMU onto an unobserved link, obtain a new s, run the observability iteration procedure, and repeat until $\Omega(s) = V$. This yields a sequence $(s^0, \ldots, s^k)$ with two properties:

1. $\Omega(s^h) \subsetneq |V|$ for $h < k$ and $\Omega(s^k) = V$
2. for $h < k$, s^{h+1} is *minimally larger* than s^h, i.e. $\sum_{\{u,v\}\in E} (s^{h+1}_{uv} - s^h_{uv}) = 1$.

Property 1 implies that $s^0, \ldots, s^{k-1}$ are all infeasible placements for Eq. (4). Our polyhedron $\mathcal{P}$, which approximates the feasibile region $\mathcal{F}$ of Eq. (4), must therefore be defined by inequalities which cut off all of these infeasible placements. Since we can obtain them all very efficiently by repeatedly running observability iterations, we can simply adjoin *no-good cuts* to $\mathcal{P}$ for each s^h for $h < k$. To do this, for all $h < k$ we define $\alpha^h = s^h \text{ XOR } \mathbf{1}$ as the Boolean complement of s^h, so that $\alpha^h s^h = 0$, which implies that:

$$\alpha^h s \geq 1 \tag{8}$$

separates s^h from $\mathcal{P}$, as desired.

The reason why this idea works well in practice is that, by Property 2, we can prove that $\alpha^{k-1}s \geq 1$ dominates all other no-good cuts Eq. (8) for $h < k - 1$. We can intuitively see why this holds by asking when the inequality Eq. (8) is tightest whilst still being valid. This occurs when the LHS is (nontrivially) as small as possible, i.e. when as many as possible of the components of α^h are zeros, which, by definition of Boolean complement, means that as many as possible of the components of s^h must be ones. Note that the algorithm which generates the placements s^h is:

1. iterate observability using **R1**, **R2** from s^h
2. if $\Omega(s) \subsetneq V$, pick $\{u, v\} \notin P$ such that installing a PMU on $\{u, v\}$ yields a larger $|\Omega(s)|$, set $s^h_{uv} = 1$, increase h and repeat from Step 1.

Hence, by Property 2, the iteration index h corresponding to the infeasible placement with as many ones as possible is the last before termination, namely $h = k - 1$.

A row generation algorithm

Note that the no-good cuts of Eq. (8) are defined by means of PMU placements $\bar{s}$ that are infeasible (Property 1 above) and $\leq$-*maximally dominant* (Property 2 above). Let $\bar{\mathcal{F}}$ be the set $\{0,1\}^m \smallsetminus \mathcal{F}$ of placements that are infeasible w.r.t. the constraint $f(s) = n$, $\bar{\mathcal{F}}_{\mathsf{max}}$ the set of $\leq$-maximal elements of $\bar{\mathcal{F}}$, and $\zeta(\bar{s}) = \{e \in E \mid \bar{s}_e = 0\}$ be the set of component indices where $\bar{s}$ is zero. We define:

$$\mathcal{P} = \big\{s \in [0,1]^m \mid \forall \bar{s} \in \bar{\mathcal{F}}_{\mathsf{max}} \sum_{e\in\zeta(\bar{s})} s_e \geq 1\big\}.$$

The row generation algorithm we employ dynamically generates facets of $\mathcal{P}$ using the combinatorial algorithm described above in order to find maximally infeasible placements s^{k-1}, and then solves the PMU minimization subproblem defined on $\mathcal{P}^\ell \cap \{0,1\}^m$, where $\mathcal{P}^\ell$ is the MILP relaxation of $\mathcal{P}$ consisting of all the cuts (Eq. (8)) found up to the ℓ-th iteration. This algorithm terminates when the cut generation procedure is unable to find a new cut. The current placement s is then optimum because of the minimization direction of the subproblems. Note that every iteration requires the solution of a MILP, which, in practice, we can solve fairly efficiently using CPLEX 12.6.

Compared to solving the two MILPs described above with CPLEX, this row generation approach is the only one which is able to solve the PES over the medium scale networks required by the SO-grid project (up to around 120 nodes and 180 links).

Generalization to arbitrary bilevel MILPs

It turns out that the algorithmic framework described above is generalizable to a large class of bilevel MILPs for which practically viable solution methods do not appear to have been found yet:

$$\left.\begin{array}{ll} \min\limits_{x\in\{0,1\}^n} & \chi x \\ & Ax \geq b \\ & f(x) \geq c - \gamma x \\ & f(x) = \left\{\begin{array}{ll} \min\limits_{y\in\mathcal{Y}} & \beta(x) y \\ & y \in \Theta(x), \end{array}\right. \end{array}\right\} \quad (9)$$

where $\chi, \gamma \in \mathbb{Q}^n$, $A \in \mathbb{Q}^{mn}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}$, $\beta : \mathbb{Q}^n \to \mathbb{Q}^q$, $\Theta(x)$ is a polyhedron in $\mathbb{R}^q$ for all $x \in \mathbb{R}^n$, and $\mathcal{Y}$ is a mixed-integer set in $\mathbb{R}^q$. Note that the bilevel formulation Eq. (4) of the PES is the special case of Eq. (9) given by $A = 0$, $b = 0$, $\gamma = 0$, $c = n$, $f(\cdot) = |\Omega(\cdot)|$, $\chi, \beta = \mathbf{1}$, and $\Theta(\cdot)$ given by the θ^s function defined in Eq. (1).

Dominance assumptions

In order to simplify the presentation, we make two assumptions, which we shall show how to relax later on:

1. $\gamma \geq 0$
2. $\forall x' \leq x''$ satisfying $Ax \geq b$, we have $\Theta(x') \supseteq \Theta(x'')$.

Note that these two assumptions are satisfied by the bilevel formulation Eq. (4) of PES: $\gamma = 0$ and, for a PMU placement s'' dominating a given placement s', the **R1** constraints $\omega_v \geq s_{uv}$ ensure that the feasible region of the lower level problem in Eq. (3) becomes smaller.

Mutatis mutandis[2], the method outlined above for the PES formulation goes through essentially unchanged, with

$$\begin{aligned}\mathcal{F} &= \{x \in \{0,1\}^n \mid Ax \geq b \wedge f(x) \geq c - \gamma x\}\\ \bar{\mathcal{F}} &= \{x \in \{0,1\}^n \mid Ax \geq b \wedge f(x) < c - \gamma x\},\end{aligned}$$

aside from the cut generation step, since the one discussed above is based on the structure of the PES problem. Next, we outline the generalization of the cut generation subproblem.

The generalized cut generation subproblem

Recall the cut generation step in the PES setting: we start from an infeasible placement s, compute the observed nodes, then add one PMU to s and repeat until all nodes are observed; we then take the maximal infeasible placement s^{k-1} at the last iteration, before full observability is achieved, and separate it with a no-good cut. In the generalized setting we lack the combinatorial relationship between ω and s given by $\Omega(s)$ and defined using the propagation rules **R1** and **R2**. However, given an infeasible solution x' of the upper level problem, we can find the set X^* of all x that $\leq$-minimally dominate x', replace x' by all of the elements of X^* (in turn) that are also in $\bar{\mathcal{F}}$, stopping when $X^* \subseteq \mathcal{F}$. When this happens, x' is in $\bar{\mathcal{F}}_{\mathsf{max}}$ (i.e. $\leq$-maximally infeasible), and can be used to generate a facet of $\mathcal{P}$ similar to Eq. (8): if α' is the complement of x' in $\{0,1\}^n$, the cut is $\alpha' x \geq 1$.

Given $x' \in \bar{\mathcal{F}}$, we compute X^* by finding all solutions to the following cut generation MILPs (for all $j \leq n$):

$$\mathsf{CG}_j = \left\{ \begin{array}{rrcll} \min\limits_{x\in\{0,1\}^n} & \sum\limits_{i\leq n} & x_i & & \\ & Ax & \geq & b & \\ & x & \geq & x' & (\dagger) \\ & x_j & \geq & 1, & (\ddagger) \end{array}\right. \tag{10}$$

where $(\dagger)$ enforces nonstrict domination w.r.t. x', with strictness on at least one coordinate direction $\mathbf{e}_j$ being enforced by $(\ddagger)$ and the objective function.

Relaxing the dominance assumptions

The first dominance assumption $\gamma \geq 0$ is mostly a technical detail which we shall not discuss here; its relaxation follows because of the new methods employed to deal with the second assumption, which intuitively states that, given $x' \in \bar{\mathcal{F}}$, in the direction of the negative orthant pointed at x' we can never find a feasible solution. We can relax this assumption by simply replacing the negative orthant with an *infeasible cone* pointed at x'. For this purpose, we write $\Theta(x)$ explicitly as:

$$\Theta(x) = \{y \in \mathbb{R}^q \mid By \geq d + Cx\}$$

for appropriately sized B, d, C, and define a (C,γ)-dominance by means of a cone $\mathcal{C}(x') = \{x \in \mathbb{R}^n \mid Cx \leq Cx' \wedge \gamma x \leq \gamma x'\}$ pointed at x', which allows us to state:

$$x \leq_{\mathcal{C}} x' \quad \Leftrightarrow \quad x \in \mathcal{C}(x').$$

[2] The Italian translation *avendo cambiato le mutande* is a courtesy of a few friends from high school — those very same friends who created the eternal motto of our high school banners, namely *schola mutanda est*.

One complication introduced by this generalized dominance is that it prevents us from using no-good cuts, since a no-good for a maximally infeasible solutions might fail to dominate no-goods from lesser infeasible solutions; moreover, no-good cuts may be invalid in this setting.

For any $x \in \{0,1\}^n$, we define a distance $\Delta_{x'}(x)$ from x to the infeasible cone $\mathcal{C}(x')$:

$$\forall x \in \mathbb{R}^n \quad \Delta_{x'}(x) = \begin{cases} \min\limits_{e,f \in \mathbb{R}^n_+} & \mathbf{1}\,(e+f) \\ C(x-e+f) & \leq \quad Cx' \\ \gamma(x-e+f) & \leq \quad \gamma x', \end{cases}$$

which is obtained as the sum of slacks needed to make the above Linear Program (LP) feasible, and replace the no-good cut of Eq. (8) by the nonlinear cut $\Delta_{x'}(x) > 0$. Since strict inequalities are not allowed in MP, we need to solve an auxiliary MILP to find the maximum scalar $\delta_{x'}$ such that

$$\Delta_{x'}(x) \geq \delta_{x'} \tag{11}$$

is a valid cut for x'. The cut in Eq. (11), however, is only nonlinear because $\Delta_{x'}(x)$ is computed via an LP. This LP can simply be replaced by its KKT conditions, which are a nonlinear system of equations and inequalities in function of primal and dual variables. As with the single-level reformulation Eq. (5), the nonlinearities are bilinear products between binary and continuous variables, which can be linearized exactly. This ultimately yields feasibility-only MILP which we shall refer to as $\mathsf{RC}(x')$.

Since $\mathsf{RC}(x')$ is feasibility only, it simply consists of new constraints and new variables, which can be appended to any MP. We shall therefore take it as the output of the cut generation subproblem in this generalized setting without assumptions. Due to the introduction of the dual variables at each iteration, the row generation algorithm now becomes a *row-and-column generation algorithm*, where the block $\mathsf{RC}(x')$ is added to the master problem at each itereation.

The cut generation subproblem is somewhat complicated by the replacement of the negative orthant with the infeasible cone. More precisely, we can no longer define auxiliary subproblems for each coordinate direction $\mathbf{e}_j$ for $j \leq n$ as in CG_j, since these span an orthant. Instead, we heuristically find a few rays of a Hilbert basis H [8] of the intersection of the negative of the infeasible cone with an appropriately chosen orthant, and define an equivalent subproblem (not altogether dissimilar from CG_j) for each ray in H, which, taken together, replace the coordinate directions $\mathbf{e}_j$. Of course Hilbert bases can have exponential size, so we use this tool heuristically: we shall not find facets of $\mathcal{P}$, but simply cuts.

Incredibly, for all this heavy use of worst-case exponential time algorithms, the whole scheme appears to be surprisingly efficient in practice. These techniques allow us to solve bilevel MILPs with matrices in the order of 1 to 10 rows and 20 to 30 columns in a few seconds to a few minutes of CPU time of a modern laptop.

References

[1] A. Aazami and M. Stilp. Approximation algorithms and hardness for domination with propagation. In M. Charikar, K. Jansen, O. Reingold, and J. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques*, volume 4627 of *LNCS*, pages 1–15, New York, 2007. Springer.

[2] D. Brueni and L. Heath. The PMU placement problem. *SIAM Journal on Discrete Mathematics*, 162(1):744–761, 2005.

[3] B. Colson, P. Marcotte, and G. Savard. Bilevel programming: a survey. *4OR*, 3:87–107, 2005.

[4] S. Dempe. *Foundations of bilevel programming.* Kluwer, Dordrecht, 2002.

[5] P. Domingos and M. Richardson. Mining the network value of customers. In F. Provost and R. Srikant, editors, *International Conference on Knowledge Discovery and Data Mining*, volume 7 of *ACM-SIGKDD*, pages 57–66, New York, 2001. ACM.

[6] A. Lodi, T. Ralphs, and G. Woeginger. Bilevel programming and the separation problem. *Mathematical Programming A*, to appear.

[7] N. Manousakis, G. Korres, and P. Georgilakis. Taxonomy of PMU placement methodologies. *IEEE Transactions on Power Systems*, 27(2):1070–1077, 2012.

[8] A. Schrijver. *Theory of Linear and Integer Programming.* Wiley, Chichester, 1986.