A Decomposition Method for Large Scale MILPs, with Performance Guarantees and a Power System Application

Robin Vujanic[a], Peyman Mohajerin Esfahani[a], Paul Goulart[b], Sébastien Mariéthoz[c], Manfred Morari[a]

[a]*Automatic Control Laboratory, ETH Zurich*
[b]*Department of Engineering Science, University of Oxford*
[c]*Bern University of Applied Sciences*

Abstract

Lagrangian duality in mixed integer optimization is a useful framework for problem decomposition and for producing tight lower bounds to the optimal objective. However, in contrast to the convex case, it is generally unable to produce optimal solutions directly. In fact, solutions recovered from the dual may not only be suboptimal, but even infeasible. In this paper we concentrate on large scale mixed–integer programs with a specific structure that appears in a variety of application domains such as power systems and supply chain management. We propose a solution method for these structures, in which the primal problem is modified in a certain way, guaranteeing that the solutions produced by the corresponding dual are feasible for the original unmodified primal problem. The modification is simple to implement and the method is amenable to distributed computation. We also demonstrate that the quality of the solutions recovered using our procedure improves as the problem size increases, making it particularly useful for large scale problem instances for which commercial solvers are inadequate. We illustrate the efficacy of our method with extensive experimentations on a problem stemming from power systems.

Keywords: optimization; decomposition methods; large-scale systems; integer programming; duality; electric vehicles; power-system control

1. Introduction

In this paper we investigate mixed-integer optimization problems in the form

$$\left\{\begin{array}{lll} \underset{x}{\text{minimize}} & \sum_{i\in I} c_i^\top x_i & \\ \text{subject to} & \sum_{i\in I} H_i x_i \leq b & \\ & x_i \in X_i \qquad & \forall i \in I. \end{array}\right. \tag{$\mathcal{P}$}$$

We refer to $b \in \mathbb{R}^m$ as the *resource vector*, and to the sets X_i as the *subsystems*. We assume that each of the sets X_i is a non-empty, compact, mixed-integer polyhedral set that can be written as

$$X_i = \left\{ x \in \mathbb{R}^{r_i} \times \mathbb{Z}^{z_i} \;\middle|\; A_i x \leq d_i \right\},$$

with $A_i \in \mathbb{R}^{m_i \times n_i}$ and $d_i \in \mathbb{R}^{m_i}$. We further assume that the problem $\mathcal{P}$ is feasible and that the total number of subsystems $|I|$ is greater than the length m of the resource vector. Our principal interest is in large-scale optimization problems, i.e. those for which $|I| \gg m$, while remaining finite.

Problem $\mathcal{P}$ can be viewed generically as modeling any problem for which a large number of subproblems defined on the domains X_i, whose description can include integer variables, are coupled through a small number of complicating constraints $\sum_{i\in I} H_i x_i \leq b$. These coupling constraints determine the limits on the available resources to be shared among the subsystems. Simple examples of problems in this form include classical combinatorial programs such as the multidimensional knapsack problem, in which $X_i = \{0, 1\}$, and $c_i \geq 0$, $H_i \geq 0$ [38].

More complicated instances of problems in the form $\mathcal{P}$, with more detailed models for the subsystems X_i, arise in a variety of contexts. In power systems, scheduling the operation of power generation plants [39] is a decision problem in which the subsystems are the generating units, integer variables in the local models arise due to, e.g., start-up and shut-down costs, and the coupling constraints are related to the requirement that generation must match load. In supply chain management, models fitting $\mathcal{P}$ appear in the problem of partial shipments [15, 37]. Portfolio optimization for small investors, for which mixed-integer models have been proposed, is another example application [4]. Finally, some sparse problems that do not naturally possess the structure of $\mathcal{P}$ can be reformulated to fit our framework by appropriately permuting rows and columns of the constraints matrix; [5] proposes a method to automate this procedure.

A direct solution of $\mathcal{P}$ is typically problematic when the problem is very large, since the problem amounts to a mixed-integer linear program of possibly very large size. As a result, the Lagrange dual of $\mathcal{P}$ is often taken as a useful alternative, because

Email address: vujanicr@ethz.ch (Robin Vujanic)

the resulting dual problem is separable in the subsystems despite the presence of the complicating constraints. When this dual problem is solved by an iterative method, e.g. using the subgradient method [7], a candidate (primal) solution to $\mathcal{P}$ can be computed at each iteration.

For problems affected by non-zero duality gap such as $\mathcal{P}$, however, this approach suffers from a major drawback. Namely, any guarantee about the properties of these candidate solutions is lost. Even at the dual optimal solution, the associated candidate primal solutions may be suboptimal and can even be infeasible.

The principal goal of this paper is to propose a new solution method for problem $\mathcal{P}$ that preserves the attractive features of solution via the Lagrange dual, while at the same time protecting the recovered primal solutions from infeasibilty.

Literature. Lagrangian relaxation for mixed integer programs was first introduced by [21], and many of its theoretical properties were described in [19]. Properties of the inner solutions in the convex case are well known [31, Thm. 28.1]. It is also well known that in general these properties are lost in the mixed-integer case [7, Section 5.5.3]. Because of this, primal recovery methods based on Lagrangian duality are often two-phase schemes in which an infeasible solution is found through duality in the first stage, and in the second stage it is rectified into a feasible one using heuristics, see, e.g., [10, 29].

Duality for problems specifically in the form $\mathcal{P}$ has been studied at least as early as in [2], where some of its special features were first characterized. In particular, it was noted that the duality gap for this program structure decreases in relative terms as the problem increases in size, as measured by the cardinality of I. We will show that the mechanism behind this vanishing gap effect can also be used to recover "good" primal solutions for the mixed-integer program $\mathcal{P}$ directly from the dual, in a way that resembles the convex (zero gap) case.

In practical applications, this behaviour of the duality gap has been observed in [10] in the context of unit commitments for power systems. In this case it is exploited in an algorithm that provides solutions to the extended master problem, but no connection to the solutions of the inner problem is provided. It also appears in the multistage stochastic integer programming literature [11, 14], where it is used to gauge the strength of the Lagrangian relaxation, but in which no relations to primal solutions are drawn. Another domain in which diminishing gap has been used is in communications, more precisely in optimization of multicarrier communication systems [40]. However, in this case non-convexity is in the objective function rather than due to the presence of integer variables.

Current Contribution. In this paper we further investigate duality for programs structured as $\mathcal{P}$ and focus on the primal solutions recovered at the dual optimum.

- We provide a new relation between the optimizers of a convexified form of $\mathcal{P}$ and the solutions obtained from the dual problem. This relation holds under mild conditions that are commonly satisfied in practice.
- Using this relation we can bound the magnitude of the constraint violations of the solutions recovered from the dual. In light of this bound, we propose a new solution method which is guaranteed to produce feasible solutions for $\mathcal{P}$. The method is based on an appropriate contraction of the resources b.
- We also provide a performance bound of the solutions recovered, which indicates that their quality improves as the problem size increases. For particular structures, arising e.g. from underlying physical networks, we refine our theoretical results to improve the performance of the method.

From a practical point of view, we note that our proposed procedure is straightforward to implement and is amenable to distributed computation. The performance bound indicates that the method is particularly attractive for large problem instances, for which generic purpose solvers may be inadequate. We show that the theoretical results are effective in practice via extensive numerical experiments on difficult problems stemming from the field of power systems control. Our method substantially outperforms commercial solvers on these problems. The limitations of the proposed method, as well as ideas to mitigate them, are also discussed in the paper.

Structure of the Paper. The paper is structured as follows: in Section 2 we review some of the known results concerning duality for the specific structure of $\mathcal{P}$, and we provide a new result related to the primal solutions recovered from the dual. In Section 3 we propose a new method for primal solution recovery, and provide performance bounds for these solutions. We also give some results on how to further improve the solutions' quality in some special cases. In Section 4 we verify the efficacy of our proposed method on a difficult optimization problem stemming from power systems, and in Section 5 we conclude the paper.

Notation. Given some optimization problem $\mathcal{A}$, we denote with $J^\star_{\mathcal{A}}$ its optimal objective and with $J_{\mathcal{A}}(x)$ the performance of the solution x with respect to the objective of $\mathcal{A}$. For a given set X, we denote by $\mathrm{conv}(X)$ its convex hull and by $\mathrm{vert}(X)$ the set of vertices of $\mathrm{conv}(X)$. With "$\geq$" we always intend component-wise inequalities (between vectors or matrices), and with $\otimes$ we indicate the cartesian product of sets. The support of a vector $\mathrm{supp}(x)$ is the set of indexes of the non-zero elements: $\mathrm{supp}(x) = \{i : x_i \neq 0\}$, while $(x)^+$ is the projection of x onto the positive orthant, i.e., $(x)^+ \doteq \max(0, x)$. For the specific structure of $\mathcal{P}$, we use the *overbar* symbol to indicate quantities related to the contracted version of $\mathcal{P}$, as introduced in Section 3. Thus, for instance, $\overline{\mathcal{P}}$ is the contracted form of $\mathcal{P}$ and $\overline{\mathcal{D}}$ is

its dual. We use parenthesis to avoid confusing the sub- and superscripts, e.g., we denote by $(x_{\mathcal{P}})_i$ the part of $x_{\mathcal{P}}$ related to subproblem $i \in I$ of problem $\mathcal{P}$. Finally, we use the superscript H^k to denote the k–th row of matrix H.

2. Duality for Problem $\mathcal{P}$

Consider the dual function $d : \mathbb{R}^m \to \mathbb{R}$ of problem $\mathcal{P}$, defined as

$$d(\lambda) \doteq \min_{x \in X} \Big(\sum_{i \in I} c_i^\top x_i + \lambda^\top (\sum_{i \in I} H_i x_i - b) \Big),$$

and then associate to this function the optimization problem

$$\left\{ \begin{array}{ll} \sup\limits_{\lambda} & -\lambda^\top b + \sum\limits_{i \in I} \min\limits_{x_i \in X_i} \left(c_i^\top x_i + \lambda^\top H_i x_i \right) \\ \text{s.t.} & \lambda \geq 0. \end{array} \right. \tag{$\mathcal{D}$}$$

We call $\mathcal{D}$ the *dual problem* of $\mathcal{P}$, and we refer collectively to the minimizations within $\mathcal{D}$, i.e.,

$$\min_{x_i \in X_i} \left(c_i^\top x_i + \lambda^\top H_i x_i \right), \tag{1}$$

as the *inner problem*. There is substantial practical interest in understanding the properties of the solutions to the inner problem (1) because they are obtained by solving $|I|$ independent (and lower dimensional) minimization problems, in contrast to the single large coupled problem $\mathcal{P}$. Additionally, they are usually obtained as by-products of methods used to solve $\mathcal{D}$ (e.g. the subgradient method). These solutions, in particular those attained at the vertices of $\text{conv}(X_i)$, are the central object of this paper:

Definition 2.1 (inner problem solutions). *For a given multiplier $\lambda \geq 0$, the set $\mathcal{X}_i(\lambda) \subseteq \mathbb{R}^{n_i}$ is defined as the set of inner solutions that are attained at the vertices of X_i, i.e.*

$$\mathcal{X}_i(\lambda) \doteq \text{vert}(X_i) \cap \arg\min_{x_i \in X_i} \left(c_i^\top x_i + \lambda^\top H_i x_i \right). \tag{2}$$

Furthermore, we denote by $x(\lambda)$ any selection from the set $\mathcal{X}(\lambda)$, and refer to it as an inner solution.

Fact 2.2. *The sets $\mathcal{X}_i(\lambda)$, $i \in I$, are non-empty for any $\lambda \geq 0$.*

Proof. See Appendix A.1.

2.1. Bound on Duality Gap

For a general mixed integer linear program, the inner solutions $x(\lambda^\star) \in \mathcal{X}(\lambda^\star)$, in which $\lambda^\star$ is an optimizer of $\mathcal{D}$, do not possess any "nice" property in general: they can be non-unique, suboptimal and even infeasible. In this paper we show that inner solutions for programs structured specifically as $\mathcal{P}$ do acquire some useful properties. Informally speaking, these additional properties arise mainly from the fact that, as $\mathcal{P}$ grows in size, it tends to closely approximate a convex program. One known result of this is that the duality gap between $\mathcal{P}$ and $\mathcal{D}$ vanishes, in relative terms, as $|I|$ increases [6, 9, 2].

Theorem 2.3 (bound on duality gap). *Assume that for any $x_i \in \text{conv}(X_i)$, there exists an $\tilde{x}_i \in X_i$ such that $H_i \tilde{x}_i \leq H_i x_i$. Then*

$$J^\star_{\mathcal{P}} - J^\star_{\mathcal{D}} \leq m \cdot \max_{i \in I} \gamma_i, \quad \gamma_i \doteq \max_{x_i \in X_i} c_i^\top x_i - \min_{x_i \in X_i} c_i^\top x_i. \tag{3}$$

In consideration of Theorem 2.3, let $|I|$ increase, while m remains constant and the sets $\{X_i\}_{i \in I}$ are uniformly bounded. If $J^\star_{\mathcal{P}}$ increases linearly with $|I|$, then

$$\frac{J^\star_{\mathcal{P}} - J^\star_{\mathcal{D}}}{J^\star_{\mathcal{P}}} \to 0 \quad \text{as} \quad |I| \to \infty. \tag{4}$$

An early proof of this result appears in [2], while a more recent version is in [6, Prop. 5.26, p. 374]. The same result also holds for more general problems; see [9, Prop. 5.7.4, p. 223].

Note that while Theorem 2.3 ensures the existence of a primal feasible solution satisfying the performance bound (3), it does not provide an algorithmic way to produce it. An example in which the assumption of Theorem 2.3 is met is the supply chain problem presented in [37]. In this case it is straightforward to observe that the assumption is satisfied since $\forall x_i \in X_i$, $H_i x_i \geq 0$ and $0 \in X_i$. However this assumption may, in general, be restrictive; an example that does not fulfil it is discussed in Section 4, see Remark 4.1. In this work we lift this assumption, at the cost of conservatism and thus performance of the solutions recovered.

2.2. Geometric Properties of the Inner Solutions $x(\lambda^\star)$

Here we present a new connection between the inner solutions $x(\lambda^\star)$ and the optimizers of the following optimization program

$$\left\{ \begin{array}{lll} \underset{x}{\text{minimize}} & \sum\limits_{i \in I} c_i^\top x_i & \\ \text{subject to} & \sum\limits_{i \in I} H_i x_i \leq b & \\ & x_i \in \text{conv}(X_i) & \forall i \in I, \end{array} \right. \tag{$\mathcal{P}_{\text{LP}}$}$$

which amounts to a linear program. We denote by $J^\star_{\mathcal{P}_{\text{LP}}}$ its optimal value, and by $x^\star_{\text{LP}}$ one of its optimizers. The relaxation $\mathcal{P}_{\text{LP}}$ plays a central role in Lagrangian duality for mixed integer programs; it is in fact well known that $\mathcal{P}_{\text{LP}}$ satisfies the (non-obvious) relation $J^\star_{\mathcal{P}_{\text{LP}}} = J^\star_{\mathcal{D}}$ [19, Thm. 1b, p.87]. Accordingly, $\mathcal{P}_{\text{LP}}$ is often used to gain insight into the strength of the relaxation, i.e., the tightness of the lower bounds to $J^\star_{\mathcal{P}}$ provided by the Lagrangian dual. While in most practical cases one cannot solve $\mathcal{P}_{\text{LP}}$ directly since an explicit description of the polyhedral sets $\text{conv}(X_i)$ is required, column generation techniques construct approximations of $\mathcal{P}_{\text{LP}}$ [3, 16, 35]. It must be further emphasized that even though $\mathcal{P}_{\text{LP}}$ is a relaxation of $\mathcal{P}$ and is a linear program, it does *not* coincide with the standard linear relaxation in which the integrality constraints on the discrete variables are relaxed to intervals. In fact, $\mathcal{P}_{\text{LP}}$ is usually tighter; see [19, Thm. 1a].

In consideration of the Shapley–Folkman–Starr theorem [2, p.233], one can expect the vertices of the convexified problem $\mathcal{P}_{\text{LP}}$ to have "structure", i.e. for $(x^\star_{\text{LP}})_i$ to belong to X_i for at least $|I|-m-1$ subproblems, and $(x^\star_{\text{LP}})_i \in \text{conv}(X_i)$ for the remaining $m+1$ ones, see [37, Thm. 1]. This number can be improved to $|I|-m$ using an argument based on simplex tableaus instead of the Shapley–Folkman–Starr theorem. We use this tighter version here, and in the following new result, the crucial technical theorem of the paper, we extend it by establishing that the subproblems for which $(x^\star_{\text{LP}})_i \in X_i$ also "freeze" the corresponding inner solutions $x_i(\lambda^\star)$.

Assumption 2.4 (uniqueness for $\mathcal{P}_{\text{LP}}$ and $\mathcal{D}$). *The programs $\mathcal{P}_{\text{LP}}$ and $\mathcal{D}$ have unique solutions $x^\star_{\text{LP}}$ and $\lambda^\star$, respectively.*

Assumption 2.4 concerns two LPs (see program $\mathcal{D}_{lp}$ in Section A.2 for the LP version of $\mathcal{D}$). Uniqueness of primal and dual optimizers in the LP case is discussed in [27], where necessary and sufficient conditions are provided. The same paper also provides an algorithmic way to verify uniqueness of primal and dual optimizers in the LP setting, see [27, Remark 2]. There are cases for which this assumption may fail, in particular when the problem's data is affected by a high degree of symmetry. These cases, however, can always be avoided by adding negligible perturbations to the cost and resource vectors. These additions are also guaranteed not to affect the optimal values $J^\star_{\mathcal{P}_{\text{LP}}}$ and $J^\star_{\mathcal{D}}$ substantially, since these quantities are continuous with respect to the perturbation size of the cost and resource vectors.

Theorem 2.5 (relation between $x^\star_{\text{LP}}$ and $x(\lambda^\star)$). *Under Assumption 2.4, the solutions $x^\star_{\text{LP}}$ and $x(\lambda^\star)$ differ in at most m subproblem components, for any selection of $x(\lambda^\star) \in \mathcal{X}(\lambda^\star)$. That is, for all $x(\lambda^\star) \in \mathcal{X}(\lambda^\star)$ there exists $I_1 \subseteq I$, with $|I_1| \geq |I| - m$, such that $x_i(\lambda^\star) = (x^\star_{\text{LP}})_i$.*

Proof. See Appendix A.2.

Note that while the structural properties of $x^\star_{\text{LP}}$ appeared in the literature [10, 9, 37], the contribution here is to ensure that, under Assumption 2.4, these advantageous properties are transferred to the inner solutions $x(\lambda^\star)$. This is of substantial practical interest, because it is the inner solutions that one has direct access to when solving the dual. In the following we provide an analytical example that further illustrates the significance of Theorem 2.5. It also includes a counterexample, showing how the desired assertion may fail in absence of Assumption 2.4.

Example 2.6. *Suppose we have to*

$$\left\{\begin{array}{lll} \textit{minimize} & \sum_{i=1}^4 c_i^\top x_i & \\ \textit{s.t.} & \sum_{i=1}^4 H_i x_i \leq 11.1 & \\ & x_i \in X_i & i = 1,\ldots,4 \end{array}\right. \tag{5}$$

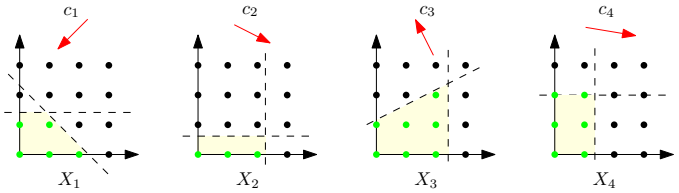


Figure 1: Illustration of the sets X_i in Example 2.6.

and $X_i = \left\{x \in \mathbb{Z}^2_+ \mid A_i x_i \leq d_i\right\}$ *with*

$$A_1 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \quad d_1 = \begin{bmatrix} 1.2 \\ 2.1 \end{bmatrix} \quad c_1^\top = [1,1] \quad H_1 = [1,1]$$
$$A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad d_2 = \begin{bmatrix} 0.6 \\ 2.1 \end{bmatrix} \quad c_2^\top = [-2,1] \quad H_2 = [5,1]$$
$$A_3 = \begin{bmatrix} 1 & 0 \\ -0.5 & 1 \end{bmatrix} \quad d_3 = \begin{bmatrix} 2.2 \\ 1.1 \end{bmatrix} \quad c_3^\top = [0.5,-1] \quad H_3 = [1,1]$$
$$A_4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad d_4 = \begin{bmatrix} 1.2 \\ 2 \end{bmatrix} \quad c_4^\top = [-3,0.5] \quad H_4 = [1,1],$$

see Figure 1. Relaxing the constraint $\sum_{i=1}^4 H_i x_i \leq 11.1$ in this problem leads to the dual function

$$d(\lambda) = \begin{cases} -8 + 0.9\lambda & 0 \leq \lambda \leq 2/5 \\ -4 - 8.9\lambda & 2/5 < \lambda \leq 1 \\ -3 - 9.9\lambda & 1 < \lambda \leq 3 \\ -10.9\lambda & \lambda > 3, \end{cases}$$

so that the dual optimizer is $\lambda^* = 2/5$, and $d(\lambda^\star) = J^\star_{\mathcal{D}} = -7.64$, while the primal optimal objective is $J^\star_{\mathcal{P}} = -7$ (note the duality gap). The corresponding sets of inner solutions are, according to (2),

$$\mathcal{X}_1(\lambda^\star) = \left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right\} \quad \mathcal{X}_2(\lambda^\star) = \left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}\right\}$$
$$\mathcal{X}_3(\lambda^\star) = \left\{\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} \quad \mathcal{X}_4(\lambda^\star) = \left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}.$$

On the other hand, $x^\star_{\text{LP}}$ is unique and is given by

$$(x^\star_{\text{LP}})_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (x^\star_{\text{LP}})_2 = \begin{bmatrix} 1.82 \\ 0 \end{bmatrix} \quad (x^\star_{\text{LP}})_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (x^\star_{\text{LP}})_4 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Notice how the relationship $x_i(\lambda^\star) = (x^\star_{\text{LP}})_i$ holds for $i \in \{1,3,4\} = I_1$, and that the cardinality of $|I_1|$ satisfies $|I_1| \geq |I| - m = 4 - 1 = 3$. The validity of Theorem 2.5 is thus verified.

On the other hand, to see how the Theorem may fail in the absence of Assumption 2.4, consider again problem (5), but now with $b = 6$ and with the subsystems determined by

$$A_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad d_i = \begin{bmatrix} 3.2 \\ 1.4 \end{bmatrix} \quad c_i^\top = [-1,1] \quad H_i = [1,1]$$

for $i = 1,\ldots,4$. Notice that all the subsystems are identical and the problem is highly symmetric. The dual function in this case is

$$d(\lambda) = \begin{cases} -12 + 6\lambda & 0 \leq \lambda \leq 1 \\ -6\lambda & \lambda > 1, \end{cases}$$

and the unique dual optimizer is $\lambda^\star = 1$. *However,* $x^\star_{\mathrm{LP}}$ *is not unique. For example*

$$(\bar{x}^\star_{\mathrm{LP}})_1 = \begin{bmatrix}0.7\\0\end{bmatrix} \quad (\bar{x}^\star_{\mathrm{LP}})_2 = \begin{bmatrix}1.6\\0\end{bmatrix} \quad (\bar{x}^\star_{\mathrm{LP}})_3 = \begin{bmatrix}0.6\\0\end{bmatrix} \quad (\bar{x}^\star_{\mathrm{LP}})_4 = \begin{bmatrix}3.1\\0\end{bmatrix}$$

and

$$(\bar{x}^\star_{\mathrm{LP}})_1 = \begin{bmatrix}2\\0\end{bmatrix} \quad (\bar{x}^\star_{\mathrm{LP}})_2 = \begin{bmatrix}2\\0\end{bmatrix} \quad (\bar{x}^\star_{\mathrm{LP}})_3 = \begin{bmatrix}2\\0\end{bmatrix} \quad (\bar{x}^\star_{\mathrm{LP}})_4 = \begin{bmatrix}0\\0\end{bmatrix}$$

are both valid optimizers of $\mathcal{P}_{\mathrm{LP}}$. *Assumption 2.4 is therefore not fulfilled. The sets of inner solutions are*

$$\mathcal{X}_i(\bar{\lambda}^\star) = \left\{\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}3\\0\end{bmatrix}\right\}, \quad i = 1, \dots, 4,$$

and the relationship of Theorem 2.5 is violated. Note, however, that the addition of a small perturbation to the cost vector is sufficient to make the problem fulfil Assumption 2.4 again, and thus recover the relationship of Theorem 2.5.

Remark 2.7 (nonlinear extension). *Theorem 2.5 holds even when the objective and the coupling constraints functions are concave. This is immediate by noticing that, in either case, local solutions are found at the vertices of* X_i, *according to a more general version of the Fundamental Theorem of Linear Programming, see [9, Prop. 2.4.2]. The passage (A.1) in the proof of Lemma 2.2 remains unchanged, and the proof of Theorem 2.5 follows verbatim.*

3. A Distributed Solution Method for $\mathcal{P}$

Theorem 2.5 says that the inner solutions $x(\lambda^\star)$ differ from $x^\star_{\mathrm{LP}}$ in at most m subsystems, where m is the dimension of the coupling constraint. Since $x^\star_{\mathrm{LP}}$ is feasible with respect to the coupling constraints and attains a better objective than $J^\star_{\mathcal{P}}$, one can expect the solutions obtained by solving the dual to be nearly feasible and to attain good objective values. In this section we exploit this result to propose a method aimed at obtaining "good" feasible solutions to problem $\mathcal{P}$ in a distributed fashion.

3.1. Contraction of the Resources

Our proposed method is to contract the resources vector b by an appropriate amount, which is determined by the results of the previous section. We show that any inner solution recovered at the dual optimum $\lambda^\star$ of the contracted problem is a feasible solution for $\mathcal{P}$. We also provide a performance bound for these solutions, which indicates that their quality improves with increasing problem size.

Consider the following modified version of problem $\mathcal{P}$

$$\left\{\begin{array}{lll} \text{minimize} & \sum_{i\in I} c_i^\top x_i & \\ \text{subject to} & \sum_{i\in I} H_i x_i \le \bar{b} & \\ & x_i \in X_i \qquad & \forall i \in I. \end{array}\right. \tag{$\overline{\mathcal{P}}$}$$

The resource vector b has been contracted to $\bar{b} \doteq b - \rho$, where the k-th element of the contraction $\rho \in \mathbb{R}^m$ is given by

$$[\rho]_k = m \cdot \max_{i\in I} \left(\max_{x_i \in X_i} H_i^k x_i - \min_{x_i \in X_i} H_i^k x_i \right), \tag{6}$$

where H_i^k is the k-th row of H_i. Correspondingly, we introduce the problems $\overline{\mathcal{P}}_{\mathrm{LP}}$ and $\overline{\mathcal{D}}$, defined similarly to $\mathcal{P}_{\mathrm{LP}}$ and $\mathcal{D}$, replacing the resource vector b with $\bar{b}$. We next establish that the primal solutions recovered from the dual of $\overline{\mathcal{P}}$ are feasible for $\mathcal{P}$.

Theorem 3.1 (feasible solutions). *If Assumption 2.4 holds for the programs* $\overline{\mathcal{P}}_{\mathrm{LP}}$ *and* $\overline{\mathcal{D}}$, *then any selection* $x(\bar{\lambda}^\star) \in \mathcal{X}(\bar{\lambda}^\star)$ *is feasible for* $\mathcal{P}$, *where* $\bar{\lambda}^\star$ *is the optimal solution of* $\overline{\mathcal{D}}$.

Proof. See Appendix A.3.

The method is simple to implement and determining the amount of contraction required only necessitates local computations, which can be parallelized. Furthermore, knowledge of the system is sometimes sufficient to determine an upper bound for ρ without having to explicitly carry out the optimizations in (6); in Section 4 we discuss an example in which this is the case. Finally, well established methods exist for solving the dual problem (e.g., the subgradient [33], cutting planes [32] and bundle methods [24]) and they can be applied directly here.

The critical assumption of Theorem 3.1 is that the resources available should be sufficiently abundant, such that the problem remains feasible after the contraction has been applied. In Section 3.2 we discuss practical cases in which it is possible to safely decrease the necessary resource reduction.

In the next Theorem we assess the performance of the solutions $x(\bar{\lambda}^\star)$. In order to obtain an explicit bound, we first make the following assumption.

Assumption 3.2 (Slater point with increasing slack). *There exist* $\zeta > 0$ *and* $\hat{x}_i \in \mathrm{conv}(X_i)$ *for all* $i \in I$ *such that*

$$\sum_{i\in I} H_i \hat{x}_i \le \bar{b} - \zeta |I| \mathbb{1}. \tag{7}$$

Theorem 3.3 (performance guarantee). *Suppose that the programs* $\overline{\mathcal{P}}_{\mathrm{LP}}$ *and* $\overline{\mathcal{D}}$ *satisfy Assumption 2.4 and Assumption 3.2 holds. Then any solution* $x(\bar{\lambda}^\star) \in \mathcal{X}(\bar{\lambda}^\star)$ *recovered satisfies*

$$J_{\mathcal{P}}(x(\bar{\lambda}^\star)) - J^\star_{\mathcal{P}} \le (m + \|\rho\|_\infty/\zeta) \cdot \max_{i\in I} \gamma_i, \tag{8}$$

where γ_i *and* ρ *are as defined in (3) and (6), respectively.*

Proof. See Appendix A.4.

In view of Theorem 3.3, if the sets $\{X_i\}_{i\in I}$ are uniformly bounded and $J^\star_{\mathcal{P}}$ grows linearly in terms of $|I|$, then

$$\frac{J(x(\bar{\lambda}^\star)) - J^\star_{\mathcal{P}}}{J^\star_{\mathcal{P}}} \to 0 \quad \text{as} \quad |I| \to \infty. \tag{9}$$

Accordingly, the quality of the solutions recovered increases the larger the problem becomes, as the optimality gap decreases at a "$1/|I|$" rate. In Section 3.3 we will discuss Assumption 3.2 and show this asymptotic behavior can be expected even in the absence of a Slater point.

Theorem 3.1 and 3.3 provide a systematic way to produce solutions that are guaranteed to be feasible and that satisfy the performance bound (8). This bound resembles (3), where the additional term "$\|\rho\|_\infty/\zeta$" may be viewed as the price to ensure feasibility and to lift the assumption required by Theorem 2.3.

3.2. Reducing Conservatism

The contraction proposed in Theorem 3.1 can be interpreted as a robustification of problem $\mathcal{P}$ toward alterations of m local solutions x_i. In this section we take a closer look at the coupling constraints matrix $H \doteq [H_1, H_2, \ldots, H_{|I|}]$ and discuss some special cases in which its structure can be exploited to safely reduce the necessary contraction.

Suppose that the matrix H has block structure, as depicted in Figure 2(a). As illustrated, we introduce the set I_k as the index set of the subsystems contributing to the k-th coupling constraint, i.e., for which $H_i^k \neq 0$. We furthermore define the submatrix $[H_i]_{i\in I_k}$, obtained by collecting the columns of H related to the subsystems in I_k.

Such a block structured H may arise in applications in which the resources present a hierarchical structure, or when the optimization is over tree or tree-star networks, as shown on Figure 2(b). In this case, the uniform contraction proposed in Theorem 3.1 can be safely reduced.

Theorem 3.4 (refinement for block structure). *Theorem 3.1 holds with the contraction (6) substituted by*

$$[\rho]_k = \operatorname{rank}([H_i]_{i\in I_k}) \cdot \max_{i\in I_k} \left(\max_{x_i\in X_i} H_i^k x_i - \min_{x_i\in X_i} H_i^k x_i \right). \tag{10}$$

Proof. See Appendix A.5.

This theorem implies, as a special case, that we can generally substitute m with $\operatorname{rank}(H)$ in (6), independently of whether the problem has block structure. This is important when the vectors determining the coupling constraints are linearly dependent. An example exploiting this result is discussed in Section 4.

Furthermore, instead of immunizing against $\operatorname{rank}([H_i]_{i\in I_k})$ times the largest subproblem budget consumption change, it is sufficient to immunize against the $\operatorname{rank}([H_i]_{i\in I_k})$ largest ones, i.e.,

Remark 3.5. *The contraction (6) can be safely substituted by*

$$[\rho]_k = \max_{\substack{\tilde I\subseteq I_k \\ |\tilde I|=\operatorname{rank}([H_i]_{i\in I_k})}} \left(\sum_{i\in\tilde I} \max_{x_i\in X_i} H_i^k x_i - \min_{x_i\in X_i} H_i^k x_i \right). \tag{11}$$

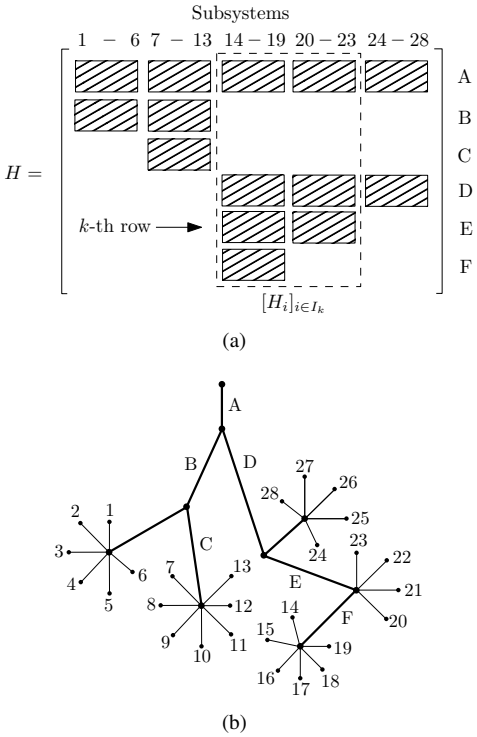


Figure 2: (a) block structure considered in Theorem 3.4; hatched boxes indicate non-zero submatrices, while the dashed box contains the submatrix $[H_i]_{i\in I_k}$. (b) an example network that would give rise to such a block structured H. In this illustrative Figure there are 28 subsystems, and 6 sets of coupling constraints determined by constraints on the network links A–F.

Finally, an important subclass of problems for which we can suppress the necessary contraction to $\rho = 0$ is the following.

Remark 3.6. *If $H_i x_i \geq 0$ for all $x_i \in X_i$, and $\mathbf{0} \in X_i$, then one can obtain the same performance bound as in (8) while setting $\rho = 0$, and a feasible solution can be recovered by setting $x_i(\lambda^\star) = \mathbf{0}$ for at most m subsystem solutions.*

This is for instance the case for the (multidimensional) knapsack problem and some of its variants. Namely, a feasible solution can be obtained by removing at most m items from the knapsacks.

3.3. Further Discussion on the Performance Bound

One of the key factors contributing to the optimality gap identified in Theorem 3.3 is the performance loss due to the contraction ρ, determined by $[J^\star_{\overline{\mathcal{P}}_{\mathrm{LP}}} - J^\star_{\mathcal{P}_{\mathrm{LP}}}]$; see the proof of Theorem 3.3, in particular the term (ii), in Section A.4. In Theorem 3.3, Assumption 3.2 allows us to establish an explicit bound on this term. Here we show that this performance loss can be characterized by the data of only m subsystems, which explains why one may expect a behavior for the optimality gap similar to (9) even in the absence of Assumption 3.2.

Proposition 3.7. *Consider the perturbed version of the pro-*

gram $\mathcal{P}_{\mathrm{LP}}$

$$\begin{cases} \textit{minimize} & \sum_{i\in I} c_i^\top x_i \\ \textit{subject to} & \sum_{i\in I} H_i x_i \le b + \varepsilon \mathbb{1} \\ & x_i \in \mathrm{conv}(X_i) \qquad i \in I, \end{cases} \qquad (\mathcal{P}_{\mathrm{LP}}(\varepsilon))$$

whose optimal value is denoted by $J^\star_{\mathcal{P}_{LP}}(\varepsilon)$. *Let* $\mathfrak{D}i \doteq (X_i, H_i, c_i)$ *be the tuple representing the data of the* i^{th} *subsystem, where the sets* X_i *are all compact. Then, there exist a partition* $I = I_1 \cup I_2$ *and a constant* $L(I_2) \doteq L((\mathfrak{D}i)_{i\in I_2})$, *only depending on the data of subsystems indexed by* I_2, *such that* $|I_2| \le m$ *and*

$$0 \le J^\star_{\mathcal{P}_{LP}}(0) - J^\star_{\mathcal{P}_{LP}}(\varepsilon) \le L(I_2)\varepsilon, \qquad \forall \varepsilon \in \mathbb{R}_+.$$

Proof. The proof, along with some preliminaries, is in Appendix A.6.

This result allows us to provide the following performance bound on the optimality gap for the recovered solutions.

Theorem 3.8 (performance without Slater). *Suppose the programs* $\overline{\mathcal{P}}_{\mathrm{LP}}$ *and* $\overline{\mathcal{D}}$ *satisfy Assumption 2.4. Then, any solution* $x(\bar\lambda^\star) \in \mathcal{X}(\bar\lambda^\star)$ *recovered satisfies*

$$J_{\mathcal{P}}(x(\bar\lambda^\star)) - J^\star_{\mathcal{P}} \le m \cdot \max_{i\in I} \gamma_i + \max_{\substack{I_2\subset I \\ |I_2|\le m}} L(I_2) \cdot \|\rho\|_\infty \qquad (12)$$

where γ_i *and* ρ *are as defined in* (3) *and* (6), *respectively, and* $L(I_2)$ *is the constant determined by subsystems indexed by* I_2 *as introduced in Proposition 3.7.*

The proof of Theorem 3.8 essentially follows the same analysis of Section A.4. In light of this theorem, it is then clear that if $\{\gamma_i\}_{i\in I}$ and $\{L(I_2)\}_{I_2\subset I}$ are uniformly bounded in $|I|$, and $J^\star_{\mathcal{P}}$ grows linearly with $|I|$, we reach the same conclusion on the optimality gap behavior as in (9). These uniform bounds are satisfied if the diversity of the subsystems added to the problem, when we increase its size, is limited.

4. Application Example: Charging of Plug-in Electric Vehicles (PEVs)

We consider a fleet of $|I|$ Plug-in (Hybrid) Electric Vehicles (PEVs) that must be charged by drawing power from the same electricity distribution network. As the number of PEVs increases, it becomes necessary to manage their charging pattern in order to avoid excessive stresses on the lines and transformers of the network. The role of interfacing the fleet of PEVs with the network operators is taken over by a so-called aggregator.

In this Section we take the perspective of such an aggregator. Its control task is to assign charging slots to each individual PEV under its authority. The charging schedules have to be compatible with the local requirements (e.g., a desired final state of charge SoC), as well as global, network wide constraints.

4.1. Model

We will only consider the problem of establishing a feasible overnight charging schedule, since this is the period when most charging will occur [34]. We will also assume that at the time when the schedule is to be decided (e.g., midnight), all PEVs are connected and their local charging requirements (initial and final required SoC) have been communicated to the aggregator. Both of these assumptions can be easily relaxed by buffering newly connected PEVs, and recomputing every 20 minutes a charging schedule with the new population information, in a receding horizon fashion similar to [17]. Further, we assume that charging can be interrupted and resumed, but in order to avoid excessive switching, once charging starts it must continue for at least 20 minutes. This is a reasonable way of charging Lithium-Ion batteries, the most common in PEVs, because they do not present memory effect [30]. Non-interruptible charging is not discussed here as it is uncommon in practice, but those applications for which it may be necessary (e.g., Nickel-Cadmium batteries) can be readily incorporated in our proposed framework with an appropriate design of the local constraints. We thus split the overnight period in intervals of 20 minutes each, and assume that the aggregator has authority to flag, for each individual PEV, the available charging time slots.

For each PEV $i \in I$, charging at the time step k is allowed when $u_i[k] = 1$, otherwise $u_i[k] = 0$. We will also consider as a separate case the situation in which discharging (or vehicle-to-grid V2G) is possible. Then, the discharge requests are modelled using $v_i[k] \in \{0, 1\}$. Charging and discharging rates P_i are assumed to be constant, as done in [17, 13, 23, 18, 36] and reflecting the charging station protocol IEC 61851[1].

The objective of the aggregator is to maximize the profit while satisfying the local charging requirements of each individual PEV and the network constraints, which are established by the network operator. The optimization problem model we work with is as follows.

- **Subsystems model**. The subsystems controlled are the PEV batteries. Battery's i charge level is denoted by $e_i[k]$, its initial state of charge is E_i^{init}, which by the end of the charging period has to attain at least E_i^{ref}. The charging conversion efficiency is $\zeta_i^u \doteq 1 - \zeta_i$, while the discharging efficiency is $\zeta_i^v \doteq 1 + \zeta_i$,[2] where $\zeta_i > 0$ encodes the conversion losses. We denote by $E_i^{\min}$ and $E_i^{\max}$ the battery's

[1] This is particularly true in case of stations with low power ratings. More generally, smart charging stations compatible with the IEC 61851 standard could operate in a semi-continuous fashion, i.e., with a minimum current output when charging, that can be then modulated in a certain band. This requirement results in disjunctive models of the corresponding subsystems, which requires discrete variables and thus fits our proposed framework. However we do not consider this aspect in the model.

[2] The discharging efficiency must be greater than 1. This correctly encodes the fact that the amount of energy fed back to the network is smaller than the battery's energy content decrease.

capacity limits. We thus have

$$e_i[0] = E_i^{\text{init}} \tag{13a}$$
$$e_i[k+1] = e_i[k] + P_i \Delta T(\zeta_i^u u_i[k] - \zeta_i^v v_i[k]) \tag{13b}$$
$$e_i[N] \geq E_i^{\text{ref}} \tag{13c}$$
$$E_i^{\min} \leq e_i[k] \leq E_i^{\max} \tag{13d}$$
$$u_i[k] + v_i[k] \leq 1 \tag{13e}$$
$$u_i, v_i \in \{0,1\}^N . \tag{13f}$$

Condition (13e) removes the possibility of charging and discharging simultaneously.

- **Coupling constraints**. Within a distribution system, network congestions typically occur on the lines departing from the substation, since the power flow at that point is the sum of all the power loads in the network, and thus largest [25]. We therefore model congestion avoidance as a limit on the global aggregate charging and discharging power flow,

$$P^{\min}[k] \leq \sum_{i \in I} P_i(u_i[k] - v_i[k]) \leq P^{\max}[k]. \tag{14}$$

In cases when other network points are susceptible to congestions, similar coupling constraints have to be added, in which the sum is over a smaller subset of PEVs. Then Theorem 3.4 can be used to limit the necessary contraction.

- **Objective function**. The objective function encodes the cost the aggegator incurs to charge its fleet,

$$\underset{u_u, v_i}{\text{minimize}} \sum_{i \in I} \sum_{k=0}^{N-1} P_i \cdot (C^u[k] u_i[k] - C^v[k] v_i[k]) \tag{15}$$

where C^u and C^v are, respectively, the price vector for electricity consumption and injection. We allow for time varying and possibly different charging and discharging prices. In the simulations we assume a 10% markup on injection pricing, i.e., $C^v = 1.1 \cdot C^u$, which the system operator pays to the aggregator in order to incentivize PEVs to make the V2G functionality available.

We can write the complete optimization program (13) – (15) as

$$\begin{cases} \underset{e,u,v}{\text{minimize}} & \sum_{i \in I} P_i \left(C^u \cdot u_i - C^v \cdot v_i\right) \\ \text{subject to} & P^{\min} \leq \sum_{i \in I} P_i (u_i - v_i) \leq P^{\max} \\ & (e_i, u_i, v_i) \in X_i \end{cases} \tag{16}$$

with

$$X_i = \left\{ \begin{bmatrix} e_i \\ u_i \\ v_i \end{bmatrix} \in \mathbb{R}^N \times \mathbb{Z}^{2N} \;\middle|\; \text{Eq. (13)} \right\}. \tag{17}$$

Remark 4.1. *Note that the assumption in Theorem 2.3 does not apply to this model. To see this, we consider the charge–only case. According to (16), $H_i x_i = P_i u_i$, and a fractional $x_i \in \text{conv}(X_i)$ implies that in at least one time step, charge is happening at a partial rate. To rectify it, one has to either increase it to the fixed charge rate or decrease it to 0. In the latter case it may however be necessary to increase charging at another time step, in order to satisfy the energy requirement of the EV (13c). Since any such rectification will cause an increase of resources used at some time, the assumption cannot be met.*

4.2. Solution Method

We apply the method proposed in Theorem 3.1 to problem (16), which we consider under two different scenarios: in the first, only charging is allowed ($v = 0$), while in the second, both charging and V2G controls are enabled. This allows us to illustrate how the method can be adapted in two cases in which the combinatorial structure of the subsystems is substantially different.

In both cases, the number of coupling constraints is $2N$. However, in consideration of Theorem 3.4 we can reduce this number to N. One generally needs to solve the optimization problems in (6) to establish ρ; however, these computations can be simplified in this example by noting that the corresponding objectives only entail one individual time step k. Namely, we have that

$$\begin{aligned} [\rho_{\text{V2G}}]_k &= N \max_{i \in I} \left(\max_{e_i, u_i, v_i \in X_i} P_i(u_i[k] - v_i[k]) - \min_{e_i, u_i, v_i \in X_i} P_i(u_i[k] - v_i[k]) \right) \\ &= 2N \max_{i \in I} P_i \\ \left[\rho_{\text{charge}}\right]_k &= N \max_{i \in I} P_i. \end{aligned} \tag{18}$$

Dualizing the complicating constraints leads to the dual problem

$$\begin{aligned} \sup_{\lambda, \mu} \quad & \sum_{i \in I} \min_{(e_i, u_i, v_i) \in X_i} P_i \left((C^u + \delta_i^u - \lambda + \mu) u_i - (C^v + \delta_i^v - \lambda + \mu) v_i \right) \\ & + \left(\lambda \bar{P}^{\min} - \mu \bar{P}^{\max} \right) \\ \text{s.t.} \quad & \lambda, \mu \geq 0, \end{aligned} \tag{19}$$

in which $\lambda[k]$ is the dual variable associated with the lower power rating constraint $\bar{P}^{\min}[k] \doteq P^{\min}[k] + \rho$, and $\mu[k]$ is the variable for $\bar{P}^{\max}[k] \doteq P^{\max}[k] - \rho$. We note that the cost vector for the subsystems is highly symmetric – every PEV receives the same price profile. In order to ensure that Assumption 2.4 is satisfied, we introduce small additive perturbation terms δ_i^u and δ_i^v to the costs C^u and, respectively, C^v.

For the outer (maximization) problem in (19) we use a subgradient method [1] with a constant stepsize rule, which we decrease every 20 – 30 iterations.

The inner (minimization) problem, on the other hand, is decomposed into $|I|$ decoupled subproblems which are optimal control problems of 1-dimensional systems. For the sole charging case, the optimal local strategy can be proven to be greedy[3]: the least

[3] Optimality of the greedy strategy can be shown using a Dynamic Programming argument, but since it is straightforward we omit it for brevity.

number of charging steps is performed, and those are selected at times of "lowest local prices" (i.e., taking into account λ and μ as well). The local optimizations are thus computationally inexpensive in this case. For the V2G case, on the other hand, the optimal charging and discharging strategy is not as immediate, so it must be solved either as a generic optimization problem, or by applying the Dynamic Programming (DP) algorithm, see e.g [8, p.23]. In our tests we apply DP.

4.3. Simulation Setup

We compare the performance of our proposed method with the results provided by CPLEX 12.5 [22]. For each fleet size considered, we generate 10 random instances based on the parameters provided in Table 1 in Appendix B. In order to ensure a fair comparison, since CPLEX is generally unable to find exact solutions to the model (16), we first run our proposed algorithm on each problem instance, record the optimality gap (we get a tight lower bound for free as a by-product of our method), and then run CPLEX up to the same optimality gap. Furthermore, the perturbation δ_i is added to the objective function, and the perturbed problem is the one on which we deploy both our method as well as CPLEX. This ensures that both methods are exposed to exactly the same problem. All our tests are performed on a Desktop PC with 8GB of RAM and a 3.10 GHz processor.

4.4. Results

Figure 3 illustrates the optimality gap of the recovered solutions (min, max and average); note that the vertical axis is logarithmic. The asymptotic behaviour (9) is confirmed. The a–priori performance bound provided by Theorem 3.3 is reported on Figure 3 for comparison, which suggests that the bound is conservative only by a constant factor, i.e., the order of convergence rate $1/|I|$ is indeed tight. It should be noted that while the main purpose of our theoretical performance bound is to ensure good asymptotic behaviour, thanks to our particular experimental setup we can estimate it explicitly. Namely, while it may be generally difficult to determine a numeric value for the Slater constant ζ defined in Assumption 3.2, in the setup of this section we can establish a conservative estimate. This is achieved by considering a worst-case scenario in which all EVs in the population require the highest possible amount of charging energy that is allowed by numeric value of the parameters in Table 1, and that this energy is acquired at the highest rate possible. Further note that since the Slater point only needs to be feasible, the discharging capability does not need to be utilized in the V2G case. Using the numbers from Table 1, we see that the worst-case amount of energy drawn over the period amounts to 12.8kWh − 3.2kWh = 9.6kWh, which at $P_i = 5$kW results in $1.92h$ worth of charging, leading to $6 = N^{\max}$ charging steps per EV. Thus, by allocating a band of average power consumption equal to $(N^{\max}/N) \cdot \max_{i\in I} P_i = 6/24 \cdot 5$kW to each individual EV, we can ensure that the total amount of energy that has to be transferred to the fleet, when distributed uniformly over the entire horizon, leaves a gap to $P^{\max}$ of at least $(N^{\max}/N) \cdot \max_{i\in I} P_i \cdot |I| - P^{\max} = 1.75 \cdot |I|$ kW. Be definition of the Slater point (7), this leads to a value of $\zeta = 1.75$ kW.

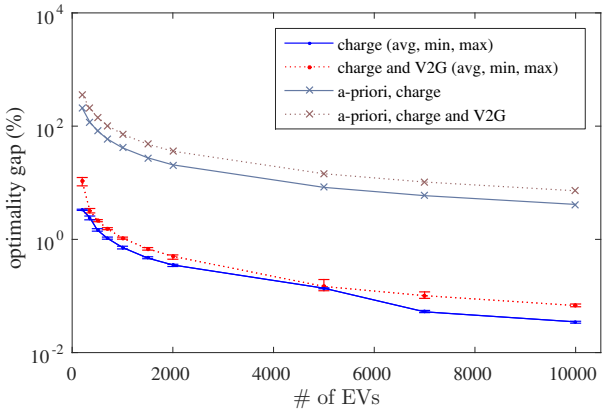
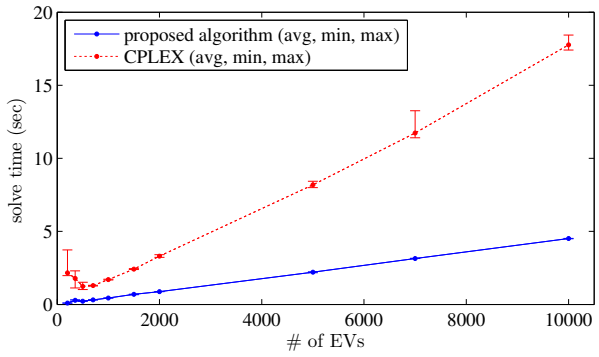
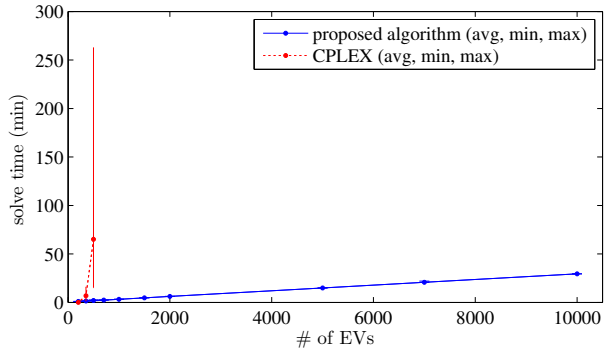


Figure 3: Optimality gap of the solutions recovered using the proposed method.



(a) Solve time (charge only)

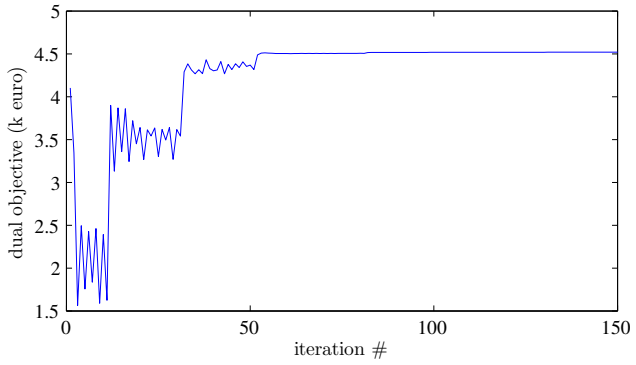


(b) Solve time (charge and V2G)

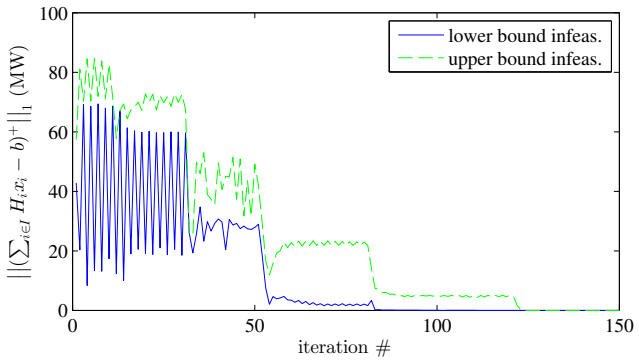
Figure 4: Solve times.

Solution times are shown on Figure 4. Owing to the greedy subproblem structure when only charging is allowed (discussed in the previous Section 4.2), computation times in this case are fast: using our method, the largest instances are consistently solved within 5 seconds, see Figure 4(a). CPLEX is comparably fast. Figure 4(b) shows solve times when the discharging

functionality is enabled. V2G introduces a much more complicated combinatorial subproblem structure – the optimal local control is not greedy anymore. In this case solution via CPLEX is impractical, because solve times are generally long and affected by substantial variances. For the case with 500 PEVs, solution times vary from 15 minutes to 4 and a half hours, and up to 6 hours on the two instances that CPLEX wasn't able to solve before running out of memory. Our proposed method has the advantage of providing consistent solution times across different instances, and the solution times substantially outperform CPLEX also on those instances in which CPLEX provides a solution at all. It should be emphasized that the computations are carried out on a single processor, so that solve times can be reduced substantially by exploiting parallelism.



(a) Dual objective



(b) Infeasibility

Figure 5: Dual function value and feasibility violations at each iteration.

Figure 5(a) and 5(b) show the typical convergence behavior for the dual objective and the coupling constraints violations. Note that inner solutions are feasible starting from iteration $\sim$ 120, while one may have interrupted the dual method already at iteration $\sim$ 60 given the dual objective behaviour. We also remark that the inclusion of a small perturbation, to ensure satisfaction of Assumption 2.4, has the beneficial side effect of accelerating dual convergence, by breaking the symmetry in the problem and reducing oscillations in the iterates, a phenomenon also noted in [26].

Finally, Figure 6 depicts the local charging behaviour of one

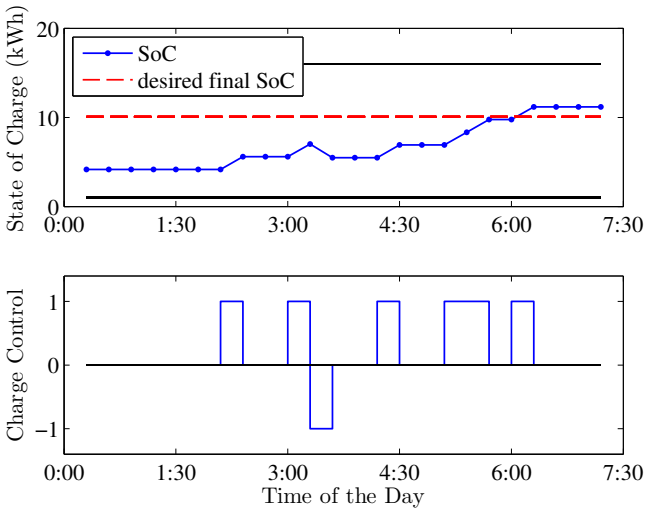


Figure 6: Local charge profile.

individual PEV. Charge and discharge control signals, as well as the evolution of the SoC are shown. The desired final state of charge is achieved by the end of the charging period.

The numeric values of these results are reported in the Appendix, see Table 2 for the charge-only experiments, and Table 3 for the results with V2G.

5. Conclusion

We have provided new results concerning the primal solutions recovered from lagrangian duals of problems structured as $\mathcal{P}$. These results are of direct practical interest, in particular if one wishes to distribute the computational burden of calculating solutions to very large instances of such mixed integer programs. The strength of our results lies in the generality of X_i, which can include very sophisticated local models and therefore accommodate a large variety of practical applications.

It appears that many solution approaches can be derived from the result given in Theorem 2.5; the one we propose in Section 3 is amenable to distributed computations and is simple to implement. It is also independent of the method used to solve the dual. Depending on the method chosen, convergence results could also be derived. One can for instance deploy the scheme exposed in [1] together with our contraction method to recover an optimizer $x^\star_{\mathrm{LP}}$ of $\overline{\mathcal{P}}_{LP}$. According to Theorem 2.5, this solution is known to satisfy integrality for at least $|I|-m$ subsystems. The non-integral components can be resolved by performing at most m local optimizations, neglecting the coupling constraints. Owing to the contraction, the resulting solution retains feasibility, and satisfies performance bounds similar to (8).

Another interesting line of research includes an investigation of the effectiveness of these methods on non-linear problems.

In Remark 2.7 we mentioned that the method presented here is directly applicable to problems with concave objective or constraint functions. Handling other non-linearities is subject of future work.

A. Appendix: Proofs

A.1. Proof of Fact 2.2

Proof. Due to the linearity of the objective function and the definition of the set X_i, it is straightforward to observe that

$$\begin{aligned}\min_{x_i \in X_i}(c_i^\top + \lambda^\top H_i)x_i &= \min_{x_i \in \operatorname{conv}(X_i)}(c_i^\top + \lambda^\top H_i)x_i \\ &= \min_{x_i \in \operatorname{vert}(X_i)}(c_i^\top + \lambda^\top H_i)x_i.\end{aligned} \tag{A.1}$$

Thus, the desired assertion readily follows from the fact that X_i are compact and non-empty. □

A.2. Proof of Theorem 2.5

Proof. Let us introduce two new LPs that are crucial for our subsequent analysis. First, we denote by x_i^j the j-th element of $\operatorname{vert}(X_i)$ for $j \in J_i$ where $J_i = \{1, \dots, |\operatorname{vert}(X_i)|\}$. In view of (A.1), one can derive an LP version of the program $\mathcal{D}$ as

$$\begin{cases} \underset{\lambda}{\text{maximize}} & -\lambda^\top b + \sum_{i \in I} \min_{j \in J_i}\left(c_i^\top x_i^j + \lambda^\top H_i x_i^j\right) \\ \text{subject to} & \lambda \geq 0, \end{cases}$$

which can then be cast as the LP

$$\begin{cases} \underset{\lambda, z}{\text{maximize}} & -\lambda^\top b + \sum_{i \in I} z_i \\ \text{subject to} & z_i \leq c_i^\top x_i^j + \lambda^\top H_i x_i^j \quad i \in I, j \in J_i \\ & \lambda \geq 0, \end{cases} \tag{A.2}$$

which, after introducing the slack variables s_i^j, is equivalent to

$$\begin{cases} \underset{\lambda, z, s}{\text{maximize}} & -\lambda^\top b + \sum_{i \in I} z_i \\ \text{subject to} & z_i = c_i^\top x_i^j + \lambda^\top H_i x_i^j - s_i^j \quad i \in I, j \in J_i \\ & s_i^j \geq 0 \qquad\qquad\qquad\quad i \in I, j \in J_i \\ & \lambda \geq 0. \end{cases} \tag{$\mathcal{D}_{lp}$}$$

The second LP is described as

$$\begin{cases} \underset{p}{\text{minimize}} & \sum_{i \in I}\sum_{j \in J_i} p_i^j c_i^\top x_i^j \\ \text{subject to} & \sum_{i \in I}\sum_{j \in J_i} p_i^j H_i x_i^j \leq b \\ & \sum_{j \in J_i} p_i^j = 1 \qquad\quad i \in I \\ & p_i^j \geq 0 \qquad\quad i \in I, j \in J_i, \end{cases} \tag{$\mathcal{P}_{lp}$}$$

where $p_i^j \in [0, 1]$ is the scalar optimization variable associated to the vertex x_i^j. In fact, (A.2) is the dual program of $\mathcal{P}_{lp}$, and in our subsequent analysis we will use its equivalent problem $\mathcal{D}_{lp}$, since we will need to work with the slack variables s. Let us denote by $p^\star$ an optimizer of $\mathcal{P}_{lp}$. Note that $\mathcal{P}_{lp}$ corresponds to an extended LP version of $\mathcal{P}_{\text{LP}}$, yet they are not entirely equivalent problems. In particular, each $p^\star$ leads to a unique $x_{\text{LP}}^\star$, but the reverse does not hold, i.e., uniqueness of $x_{\text{LP}}^\star$ does not imply uniqueness of $p^\star$.[4] We split the proof of the theorem by proving the following steps:

(a) Let $I_1 \subset I$ be a subset of indices where $(x_{\text{LP}}^\star)_i \in \operatorname{vert}(X_i)$ for all $i \in I_1$. Then, $(x_{\text{LP}}^\star)_i$ is an optimizer of the inner problem, i.e., $(x_{\text{LP}}^\star)_i \in \underset{x_i \in X_i}{\operatorname{argmin}}(c_i^\top x + \lambda^{\star\top} H_i x)$ where $\lambda^\star$ is an optimizer of $\mathcal{D}$.

(b) Let $(\lambda^\star, z^\star, s^\star)$ be an optimal solution of $\mathcal{D}_{lp}$ and $p^\star$ be an optimal solution of $\mathcal{P}_{lp}$ with the corresponding optimizer $x_{\text{LP}}^\star$ for $\mathcal{P}$. If the optimal pair $(p^\star, s^\star)$ is strictly complementary, then $(x_{\text{LP}}^\star)_i = x_i(\lambda^\star)$ for all i in the subset I_1 as defined in (a).

(c) If $x_{\text{LP}}^\star$ is a vertex for the program $\mathcal{P}$, then the subset I_1 in (a) can be selected such that $|I_1| \geq |I| - m$.

Before proceeding with the proofs of the above results, let us highlight how the desired assertion, under the unique primal and dual optimizers, follows from these three steps. First, note that if the optimal solution of $\mathcal{D}$ is unique, then $(\lambda^\star, z^\star, s^\star)$ is the unique solution to $\mathcal{D}_{lp}$: $\lambda^\star$ coincides for $\mathcal{D}$ and $\mathcal{D}_{lp}$ according to [19, p. 89]; $z^\star$ is the optimal objective of the i-th inner problem, and is thus uniquely determined for fixed λ; and finally $(s^\star)_i^j$ is also uniquely determined by the equality constraints in $\mathcal{D}_{lp}$, in which it is the only variable left undetermined. Therefore, $s^\star$ always belongs to the pair $(p^\star, s^\star)$ of primal-dual optimizers for which strict complementarity holds; the existence of such pair is guaranteed in the LP setting [20, Thm. 2.1]. Moreover, if $x_{\text{LP}}^\star$ is unique, then it is always a vertex. Hence, the requirements of the above results are fulfilled and the theorem assertion is concluded.

Proof of (a)*:*. Let $(x_{\text{LP}}^\star)_i \in \operatorname{vert}(X_i)$. Then, owing to the uniqueness of $x_{\text{LP}}^\star$, for any solution $p^\star$ of $\mathcal{P}_{lp}$ we have $(p^\star)_i^{\hat{\jmath}} = 1$ for the corresponding $\hat{\jmath} \in J_i$. Therefore, by complementary slackness, the dual optimizer has $(s^\star)_i^{\hat{\jmath}} = 0$, and the step (a) follows by

$$z_i^\star = c_i^\top (x_{\text{LP}}^\star)_i + \lambda^{\star\top} H_i (x_{\text{LP}}^\star)_i \leq c_i^\top x_i^j + \lambda^{\star\top} H_i x_i^j, \qquad \forall j \in J_i. \tag{A.3}$$

Proof of (b)*:*. Let $i \in I_1$ and, as explained in the proof of (a), $(p^\star)_i^{\hat{\jmath}} = 1$ for the corresponding $\hat{\jmath} \in J_i$. In light of the equality constraint $\sum_{j \in J_i} p_i^j = 1$, we have $(p^\star)_i^j = 0$ for all $j \neq \hat{\jmath}$. The assumed strict complementarity now implies $(s^\star)_i^j \neq 0$ for all $j \neq \hat{\jmath}$, which leads to a strict inequality in (A.3). Hence, the inner problem $\min_{x_i \in X_i}(c_i^\top x + \lambda^{\star\top} H_i x)$ has the unique solution $x_i(\lambda^\star)$. Now the desired assertion follows from the step (a).

[4] Since $(p^\star)_i$ represents a way to combine the vertices of $\operatorname{conv}(X_i)$ to produce a certain $(x_{\text{LP}}^\star)_i$, many such combinations can lead to the same $(x_{\text{LP}}^\star)_i$.

Proof of (c)*:*. Problem $\mathcal{P}_{lp}$ has m inequality constraints ($b \in \mathbb{R}^m$) and $|I|$ equality constraints, plus the positivity constraints on p_i^j. We can add slack variables to the complicating constraints thus obtaining a problem with $|I|+m$ equality constraints and positivity constraints on all the optimization variables, which are now the slacks $q \in \mathbb{R}_+^m$ and the variables p_i^j. The constraints of $\mathcal{P}_{lp}$ can therefore be rewritten as $\mathbb{H}(p^\top, q^\top)^\top = (b^\top, 1 \dots, 1)^\top$, $p, q \geq 0$, where the matrix $\mathbb{H}$, is defined as

$$\mathbb{H} = \left[\begin{array}{ccc|c|ccc|c} H_1 x_1^1 & \cdots & H_1 x_1^{J_1} & \cdots & H_{|I|} x_{|I|}^1 & \cdots & H_{|I|} x_{|I|}^{J_{|I|}} & I_{m\times m} \\ \hline 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 & 0 \\ \hline & \vdots & & \ddots & & \vdots & & \vdots \\ \hline 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 & 0 \end{array}\right] \tag{A.4}$$
$$\underbrace{\qquad\qquad}_{\mathbb{H}_1} \qquad\qquad \underbrace{\qquad\qquad}_{\mathbb{H}_{|I|}}$$

in which we have also defined the submatrices $\mathbb{H}_i$, $i \in I$. It is well known (see [9, Prop. 2.1.4 (b)]) that for a problem in this form any feasible point is a vertex if and only if the columns of $\mathbb{H}$ corresponding to the non-zero coordinates of the point are linearly independent. This is then true for any optimal vertex. Thus, $\text{supp}(p^\star) \leq |I| + m$, as the number of rows of $\mathbb{H}$ is $|I| + m$. On the other hand, the constraint $\sum_{j\in J_i} p_i^j = 1$, $i \in I$ in $\mathcal{P}_{lp}$ forces any feasible solution to have at least one variable p_i^j larger than zero for each $i \in I$, i.e. $\text{supp}(p^\star) \geq |I|$. It thus follows that at least $|I| - m$ entries must be set to 1 at any feasible vertex solution, including an optimal one.

□

A.3. Proof of Theorem 3.1

Proof. Note that by construction $x(\bar\lambda^\star) \in \mathcal{X}(\bar\lambda^\star)$ for all $i \in I$. Then, it only suffices to show $\sum_{i\in I} H_i x_i(\bar\lambda^\star) \leq b$. By virtue of Theorem 2.5, we know that there exists a subset $I_1 \subset I$ such that $|I_1| \geq |I| - m$ and $x_i(\bar\lambda^\star) = (\bar x_{\text{LP}}^\star)_i$. Setting $I_2 = I \setminus I_1$, we have

$$\begin{aligned}\sum_{i\in I} H_i x_i(\bar\lambda^\star) &= \sum_{i\in I_1} H_i x_i(\bar\lambda^\star) + \sum_{i\in I_2} H_i x_i(\bar\lambda^\star) \\ &= \sum_{i\in I_1} H_i (\bar x_{\text{LP}}^\star)_i + \sum_{i\in I_2} H_i x_i(\bar\lambda^\star) \\ &= \underbrace{\sum_{i\in I} H_i (\bar x_{\text{LP}}^\star)_i}_{\leq \bar b} + \underbrace{\sum_{i\in I_2} \left(H_i x_i(\bar\lambda^\star) - H_i (\bar x_{\text{LP}}^\star)_i\right)}_{\leq \rho} \leq b.\end{aligned}$$

□

A.4. Proof of Theorem 3.3

Proof. Note that

$$J_{\mathcal{P}}(x(\bar\lambda^\star)) - J_{\mathcal{P}}^\star = \underbrace{\left[J_{\mathcal{P}}(x(\bar\lambda^\star)) - J_{\overline{\mathcal{P}}_{\text{LP}}}^\star\right]}_{\text{(i)}} + \underbrace{\left[J_{\overline{\mathcal{P}}_{\text{LP}}}^\star - J_{\mathcal{P}_{\text{LP}}}^\star\right]}_{\text{(ii)}} + \underbrace{\left[J_{\mathcal{P}_{\text{LP}}}^\star - J_{\mathcal{P}}^\star\right]}_{\text{(iii)}},$$

where each term can be bounded as follows:

(i) According to Theorem 2.5, there exists an index set I_1 with $|I_1| \geq |I| - m$ such that, for all $i \in I_1$, $(\bar x_{\text{LP}}^\star)_i = x_i(\bar\lambda^\star)$. Defining $I_2 \doteq I \setminus I_1$, we have

$$\begin{aligned}J_{\mathcal{P}}(x(\bar\lambda^\star)) - J_{\mathcal{P}}(\bar x_{\text{LP}}^\star) &= \sum_{i\in I_2} \left(c_i^\top x_i(\bar\lambda^\star) - c_i^\top (\bar x_{\text{LP}}^\star)_i\right) \\ &\leq m \cdot \max_{i\in I} \left(\max_{x_i\in X_i} c_i^\top x_i - \min_{x_i\in X_i} c_i^\top x_i\right) \\ &= m \max_{i\in I} \gamma_i.\end{aligned}$$

(ii) By virtue of [28, Lemma 1], given the Slater's point $\hat x$ we can bound $\|\bar\lambda^\star\|_1$ by

$$\|\lambda^\star\|_1 \leq \frac{1}{\zeta|I|}\Big(\sum_{i\in I} c_i^\top \hat x_i - \Big(\sum_{i\in I} \min_{x_i\in X_i} \left(c_i^\top + \lambda H_i\right) x_i\Big) - \lambda^\top b\Big), \qquad \forall \lambda \geq 0.$$

Setting $\lambda = 0$ in the above, we arrive at

$$\|\lambda^\star\|_1 \leq \frac{1}{\zeta} \max_{i\in I} \gamma_i, \qquad \gamma_i \doteq \max_{x_i\in X_i} c_i^\top x_i - \min_{x_i\in X_i} c_i^\top x_i.$$

In light of perturbation theory [12, Sec. 5.6.2], one can bound the term (ii) from above by $(\bar\lambda^\star)^\top \rho$, where $\bar\lambda^\star$ is the optimizer of the program $\overline{\mathcal{D}}$ and ρ is the contraction vector as defined in (6). Thus,

$$J_{\overline{\mathcal{P}}_{LP}}^\star - J_{\mathcal{P}_{\text{LP}}}^\star \leq (\bar\lambda^\star)^\top \rho \leq \|\bar\lambda^\star\|_1 \|\rho\|_\infty \leq \frac{\|\rho\|_\infty}{\zeta} \max_{i\in I} \gamma_i.$$

(iii) By definition, $\overline{\mathcal{P}}_{\text{LP}}$ is a relaxed version of $\overline{\mathcal{P}}$. Hence $J_{\mathcal{P}_{\text{LP}}}^\star - J_{\mathcal{P}}^\star \leq 0$.

□

A.5. Proof of Theorem 3.4

For a given $x(\lambda^\star) \in \mathcal{X}(\lambda^\star)$, let us introduce $\widetilde I = \left\{i \in I \mid (x_{\text{LP}}^\star)_i \neq x_i(\lambda^\star)\right\}$. For the k-th complicating constraint we then have

$$\begin{aligned}\sum_{i\in I} H_i^k x_i(\lambda^\star) &= \sum_{i\in I\setminus\widetilde I} H_i^k (x_{\text{LP}}^\star)_i + \sum_{i\in\widetilde I} H_i^k x_i(\lambda^\star) \\ &\leq b + \sum_{i\in\widetilde I} H_i^k (x_i(\lambda^\star) - (x_{\text{LP}}^\star)_i) \\ &= b + \sum_{i\in\widetilde I\cap I_k} H_i^k (x_i(\lambda^\star) - (x_{\text{LP}}^\star)_i) \\ &\leq b + |\widetilde I \cap I_k| \cdot \max_{i\in I_k} \left(\max_{x_i\in X_i} H_i^k x_i - \min_{x_i\in X_i} H_i^k x_i\right)\end{aligned}$$

In order to get a bound on $|\widetilde I \cap I_k|$, we resort again to the program $\mathcal{P}_{lp}$. We know that, under Assumption 2.4, $x_i(\lambda^\star) \neq (x_{\text{LP}}^\star)_i$ if

and only if $(x^\star_{\mathrm{LP}})_i \notin \mathrm{vert}(X_i)$, as shown in Appendix A.2. Thus, if $i \in \widetilde{I}$ there are at least two $j \in J_i$ such that $(p^\star)_i^j > 0$ in the corresponding program $\mathcal{P}_{lp}$. And for every $i \in I$, there is always at least one $j \in J_i$ such that $(p^\star)_i^j > 0$. Thus

$$|\mathrm{supp}([p^\star]_{i\in I_k})| \geq |I_k \setminus \widetilde{I}| + 2|\widetilde{I} \cap I_k| = |I_k| + |\widetilde{I} \cap I_k|.$$

On the other hand, in view of [9, Prop. 2.1.4 (b)], and as discussed in Appendix A.2, the columns within the matrix $\mathbb{H}$ (defined in Equation (A.4)) corresponding to non-zero $(p^\star)_i^j$ coordinates must be linearly independent. Hence $|\mathrm{supp}(p^\star)| \leq \mathrm{rank}(\mathbb{H})$ and in particular

$$|\mathrm{supp}([p^\star]_{i\in I_k})| \leq \mathrm{rank}([\mathbb{H}_i]_{i\in I_k}).$$

Finally, from the structure of $\mathbb{H}$ defined in Equation (A.4), it is clear that

$$\mathrm{rank}([\mathbb{H}_i]_{i\in I_k}) \leq \mathrm{rank}([H_i]_{i\in I_k}) + |I_k|.$$

Combining the above inequalities immediately leads to

$$\mathrm{rank}([H_i]_{i\in I_k}) \geq |\widetilde{I} \cap I_k|,$$

as desired.

A.6. *Proof of Proposition 3.7*

The objective is to establish a connection from the sensitivity of the large scale, but structured, optimization program $\mathcal{P}_{\mathrm{LP}}$ to a reduced version in which only m subsystems appear. To this end, we first start with some preparatory lemmas.

Lemma A.1. *Let $J : \mathbb{R}_+ \to \mathbb{R}$ be a convex function. Suppose there exist a constant L and a sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that $\varepsilon_n \to 0$ as n goes to infinity and $J(0) - J(\varepsilon_n) \leq L\varepsilon_n$ for all $n \in \mathbb{N}$. Then, $J(0) - J(\varepsilon) \leq L\varepsilon$ for all $\varepsilon \in \mathbb{R}_+$.*

Proof. For the sake of contradiction, suppose there exists an $\bar{\varepsilon}$ such that $J(0) - J(\bar{\varepsilon}) > L\bar{\varepsilon}$. Let n be large enough so that $\varepsilon_n \in (0, \bar{\varepsilon})$ and $\alpha \doteq \frac{\varepsilon_n}{\bar{\varepsilon}}$. In light of convexity of J, we have

$$\begin{aligned} J(\varepsilon_n) &\leq (1-\alpha)J(0) + \alpha J(\bar{\varepsilon}) \\ &< (1-\alpha)J(0) + \alpha(J(0) - L\bar{\varepsilon}) = J(0) - L\varepsilon_n, \end{aligned}$$

which is obviously in contradiction with our assumption. □

Lemma A.2. *Consider the parametrized LP*

$$\begin{cases} \underset{x}{\text{minimize}} & cx \\ \text{subject to} & Ax \leq b + \varepsilon\mathbb{1}, \end{cases} \tag{A.5}$$

where $\varepsilon \in \mathbb{R}_+$ is the parameter and $\mathbb{1} \doteq [1, \ldots, 1]^\top \in \mathbb{R}^m$. Suppose the program admits a vertex optimizer whose objective value is denoted by $J(\varepsilon)$. Then, there exists a constant independent of the resource vector b, denoted by $L(A, c)$, such that

$$0 \leq J(0) - J(\varepsilon) \leq L(A, c)\varepsilon, \qquad \forall \varepsilon \in \mathbb{R}_+.$$

Proof. We only need to prove the right-hand side of the inequality as the left-hand side trivially holds since the parameter ε is non-negative and only relaxes the constraint. Let $x^\star(\varepsilon)$ be a vertex optimizer for (A.5). By virtue of [9, Prop. 2.1.4 (a)], given a fixed ε, we know that there exists a collection of m linearly independent rows of the matrix A, denoted by the invertible submatrix $[A](\varepsilon)$, such that $[A](\varepsilon)x^\star(\varepsilon) = b + \varepsilon\mathbb{1}$. Note that the number of submatrices of matrix A is, of course, finite. Therefore, one can always pick a sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that $\varepsilon_n \to 0$ as n goes to infinity and the corresponding submatrix $[A](\varepsilon_n)$ is constant; let us denote this submatrix by $[A]$. We thus have

$$J(0) - J(\varepsilon_n) = cx^\star(0) - cx^\star(\varepsilon_n) = -c[A]^{-1}\mathbb{1}\varepsilon_n \leq L(c, A)\varepsilon_n,$$

where the constant can be, for example, $L(c, A) \doteq m\|c\|_2\|[A]^{-1}\|_2$. Note that, by construction, the submatrix $[A]$ is invertible and the norm $\|[A]^{-1}\|$ is bounded. The desired assertion now follows from the convexity of the perturbation mapping $\varepsilon \mapsto J(\varepsilon)$ [12, Sec. 5.6.2] and Lemma A.1. □

Proof (Theorem 3.7). Given the partition $I = I_1 \cup I_2$, we introduce a reduced version of $\mathcal{P}_{\mathrm{LP}}(\varepsilon)$ associated with the index set I_2 as follows:

$$\begin{cases} \underset{(x_i)_{i\in I_2}}{\text{minimize}} & \sum_{i\in I_2} c_i^\top x_i \\ \text{subject to} & \sum_{i\in I_2} H_i x_i \leq b - \sum_{i\in I_1} H_i (x^\star_{\mathrm{LP}})_i + \varepsilon\mathbb{1} \\ & x_i \in \mathrm{conv}(X_i) \qquad\qquad i \in I_2, \end{cases} \tag{$\mathcal{R}_{I_2}(\varepsilon)$}$$

where $x^\star_{\mathrm{LP}}$ is an optimizer of the program $\mathcal{P}_{\mathrm{LP}}$. We denote the optimal value of $\mathcal{R}_{I_2}(\varepsilon)$ by $J^\star_{\mathcal{R}_{I_2}}(\varepsilon)$. Let us highlight that for any partition of the index set $I = I_1 \cup I_2$ the program $\mathcal{R}_{I_2}(\varepsilon)$ is always feasible as $\left(x^\star_{\mathrm{LP}}\right)_{i\in I_2}$ trivially satisfies the constraints for any $\varepsilon \in \mathbb{R}_+$. As a first step in the proof, we show that there exist an index subset I_2 and a sequence of $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that $|I_2| \leq m$ and the optimal values $J^\star_{\mathcal{P}_{LP}}(\varepsilon_n)$ and $J^\star_{\mathcal{R}_{I_2}}(\varepsilon_n)$ have the same sensitivity in terms of the parameter ε.

Let $x^\star_{\mathrm{LP}}(\varepsilon)$ be a vertex optimizer of the program $\mathcal{P}_{LP}(\varepsilon)$; the existence of such a vertex is always ensured since the feasible set of $\mathcal{P}_{LP}(\varepsilon)$ is a compact polytope. In light of part (c) in the proof of Theorem 2.5, we know that for each $x^\star_{\mathrm{LP}}(\varepsilon)$ there exists a partition $I = I_1(\varepsilon) \cup I_2(\varepsilon)$ where $|I_2(\varepsilon)| \leq m$ and $(x^\star_{LP}(\varepsilon))_i \in \mathrm{vert}(X_i)$ for all $i \in I_1(\varepsilon)$. Due to the fact that the number of the subsets of I as well as the set $\mathrm{vert}(X_i)$ is finite, then there exists a partition $I = I_1 \cup I_2$ and a subsequence of $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that $|I_2| \leq m$ and $(x^\star_{LP}(\varepsilon_n))_i$ are constants for $i \in I_1$. By compactness we can, without loss of generality, assume that this sequence is convergent. It is a well-known result in the context of perturbation theory that the mapping $\varepsilon \mapsto J^\star_{\mathcal{P}_{\mathrm{LP}}}(\varepsilon)$ is convex and closed on $[0, \infty)$, and in particular continuous [31, Sec. 28]. Hence, one can infer that $(x^\star_{LP}(\varepsilon_n))_i$ converges to an optimizer of $\mathcal{P}_{\mathrm{LP}}$, which consequently implies $(x^\star_{LP}(\varepsilon_n))_i = (x^\star_{\mathrm{LP}})_i$ for all $i \in I_1$. Therefore, by construction of the auxiliary program $\mathcal{R}_{I_2}(\varepsilon)$ we can deduce

$$J^\star_{\mathcal{P}_{LP}}(0) - J^\star_{\mathcal{P}_{LP}}(\varepsilon_n) = J^\star_{\mathcal{R}_{I_2}}(0) - J^\star_{\mathcal{R}_{I_2}}(\varepsilon_n), \qquad \forall n \in \mathbb{N}.$$

Now, in view of Lemma A.2, we know that the right-hand side of the above equality is non-negative and can be upper bounded by a constant only depending on the data of the subsystems indexed in I_2, i.e., $(\mathfrak{D}i)_{i\in I_2}$. Let us denote this constant by $L(I_2)$. Then, we have

$$0 \leq J^{\star}_{\mathcal{P}_{LP}}(0) - J^{\star}_{\mathcal{P}_{LP}}(\varepsilon_n) \leq L(I_2)\varepsilon_n, \qquad \forall n \in \mathbb{N},$$

that by virtue of Lemma A.1 leads to the desired assertion. □

B. Simulation Tables

Table 1 contains the parameters used in the simulation. Values in brackets are sampled from a uniform distribution over the given interval. Tables 2 and 3 report the numeric values of the performance results derived from the simulations discussed in Section 4.

References

[1] K. M. Anstreicher and L. A. Wolsey. Two "well-known" properties of subgradient optimization. *Math. Program.*, 120(1):213–220, Apr. 2009.

[2] J. P. Aubin and I. Ekeland. Estimates of the duality gap in nonconvex optimization. *Mathematics of Operations Research*, 1(3):225–245, Jan. 1976.

[3] C. Barnhart, E. L. Johnson, G. L. Nemhauser, M. W. Savelsbergh, and P. H. Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations research*, 46(3):316–329, 1998.

[4] P. Baumann and N. Trautmann. Portfolio-optimization models for small investors. *Mathematical Methods of Operations Research*, 77(3):345–356, June 2013.

[5] M. Bergner, A. Caprara, F. Furini, M. Luebbecke, E. Malaguti, and E. Traversi. Partial convexification of general mips by dantzig-wolfe reformulation. In *Proceedings of the 15th international conference on Integer Programming and combinatorial optimization*, pages 39–51, 2011.

[6] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, 1 edition, Jan. 1996.

[7] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition, Sept. 1999.

[8] D. P. Bertsekas. *Dynamic Programming & Optimal Control, Vol. I*. Athena Scientific, 3rd edition, May 2005.

[9] D. P. Bertsekas. *Convex Optimization Theory*. Athena Scientific, 1st edition, June 2009.

[10] D. P. Bertsekas, G. Lauer, N. Sandell, and T. Posbergh. Optimal short-term scheduling of large-scale power systems. *IEEE Transactions on Automatic Control*, 28(1):1– 11, Jan. 1983.

[11] J. R. Birge and M. a. H. Dempstert. Stochastic programming approaches to stochastic scheduling. *Journal of Global Optimization*, 9(3-4):417–451, Dec. 1996.

[12] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Mar. 2004.

[13] D. Callaway and I. Hiskens. Achieving controllability of electric loads. *Proceedings of the IEEE*, 99(1):184–199, 2011.

[14] C. C. Caroe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1):37–45, 1999.

[15] M. Dawande, S. Gavirneni, and S. Tayur. Effective heuristics for multi-product partial shipment models. *Operations Research*, 54(2):337–352, Jan. 2006.

[16] J. Desrosiers and M. E. Lübbecke. *A primer in column generation*. Springer, 2005.

[17] A. Di Giorgio, F. Liberati, and S. Canale. Optimal electric vehicles to grid power control for active demand services in distribution grids. In *2012 20th Mediterranean Conference on Control Automation (MED)*, pages 1309–1315, 2012.

[18] L. Gan, U. Topcu, and S. Low. Stochastic decentralized protocols for electric vehicle charging with discrete charging rates. In *IEEE Power and Energy Society General Meeting*, 2012.

[19] A. M. Geoffrion. *Lagrangean relaxation for integer programming*, volume 2 of *Mathematical Programming Studies*. Springer Berlin Heidelberg, 1974.

[20] H. J. Greenberg. The use of the optimal partition in a linear programming solution for postoptimal analysis. *Operations Research Letters*, 15:179–185, 1994.

[21] M. Held and R. M. Karp. The Traveling-Salesman problem and minimum spanning trees. *Operations Research*, 18(6):1138–1162, Nov. 1970.

[22] IBM. ILOG CPLEX 12.5, available on http://www.ilog.com/products/cplex/.

[23] S. Kundu and I. A. Hiskens. Hysteresis-based charging control of plug-in electric vehicles. In *2012 IEEE 51st Annual Conference on Decision and Control*, pages 5598–5604, 2012.

[24] C. Lemarechal. Lagrangian relaxation. In M. Juenger and D. Naddef, editors, *Computational Combinatorial Optimization*, volume 2241, pages 112–156. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[25] J. Lopes, F. Soares, and P. Almeida. Integration of electric vehicles in the electric power system. *Proceedings of the IEEE*, 99(1):168–183, Jan. 2011.

[26] Z. Ma, D. Callaway, and I. Hiskens. Decentralized charging control for large populations of plug-in electric vehicles. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 206–212. IEEE, 2010.

[27] O. Mangasarian. Uniqueness of solution in linear programming. *Linear Algebra and its Applications*, 25:151–162, June 1979.

[28] A. Nedic and A. Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM J. on Optimization*, 19(4):1757–1780, Feb. 2009.

[29] N. Redondo and A. Conejo. Short-term hydro-thermal coordination by lagrangian relaxation: solution of the dual problem. *Power Systems, IEEE Transactions on*, 14(1):89–95, 1999.

[30] M. Riezenman. The search for better batteries. *IEEE Spectrum*, 32(5), May 1995.

[31] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[32] A. P. Ruszczyński. *Nonlinear optimization*, volume 13. Princeton university press, 2006.

[33] N. Shor, K. Kiwiel, and A. Ruszczynski. *Minimization methods for non-differentiable functions*, volume 121. Springer Verlag, 1985.

[34] R. Sioshansi. Modeling the impacts of electricity tariffs on plug-in hybrid electric vehicle charging, costs, and emissions. *Operations Research*, 60(3):506–516, Jan. 2012.

[35] F. Vanderbeck. Implementing mixed integer column generation. In *Column Generation*, pages 331–358. Springer, 2005.

[36] R. Vujanic, P. M. Esfahani, P. Goulart, and M. Morari. Electric Vehicles Aggregator Optimization: a Fast and Solver-Free Solution Method. *To appear in the Proceedings of the IEEE Conference on Decision and Control*, Dec. 2014.

[37] R. Vujanic, P. M. Esfahani, P. Goulart, and M. Morari. Large Scale Mixed-Integer Optimization: a Solution Method with Supply Chain Applications. *To appear in the Proceedings of the Mediterranean Conference on Control and Automation*, Feb. 2014.

[38] C. Wilbaut, S. Hanafi, and S. Salhi. A survey of effective heuristics and their application to a variety of knapsack problems. *IMA Journal of Management Mathematics*, 19(3):227–244, July 2008.

[39] H. Yamin. Review on methods of generation scheduling in electric power systems. *Electric Power Systems Research*, 69:227–248, May 2004.

[40] W. Yu and R. Lui. Dual methods for nonconvex spectrum optimization of multicarrier systems. *IEEE Transactions on Communications*, 54(7):1310–1322, 2006.

Parameter	$\|I\|$	P_i	$E_i^{\min}$	$E_i^{\max}$	E_i^{init}	E_i^{ref}
Unit	PEVs	kW	kWh	kWh	kWh	kWh
Value	200 – 10000	$[3; 5]$	1	$[8; 16]$	$[0.2; 0.5] \cdot E_i^{\max}$	$[0.55; 0.8] \cdot E_i^{\max}$

Parameter	ζ_i	ΔT	N	$P^{\max}$	$P^{\min}$	$C^u[k]$	$C^v[k]$	δ_i^u, δ_i^v
Unit	–	min	–	kW	kW	€/MWh	€/MWh	€/MWh
Value	$[0.015; 0.075]$	20	24	$3 \cdot \|I\|$	$-P^{\max}$	$[19; 35]$	$1.1 \cdot C^u[k]$	$[-0.3; 0.3]$

Table 1: Parameters used in the simulations. Values in the brackets are sampled from a uniform distribution.

	Proposed Method						CPLEX		
	Opt. Gap (%)			Solve time† (sec)			Solve time (sec)		
# PEVs	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
200	3.24	3.32	3.41	*	*	*	1.97	2.16	3.74
350	2.21	2.44	2.58	*	*	*	1.13	1.79	2.30
500	1.40	1.46	1.54	*	*	*	1.02	1.24	1.52
700	1.01	1.05	1.10	0.31	0.31	0.31	1.27	1.29	1.31
1000	0.68	0.72	0.76	0.44	0.44	0.44	1.68	1.70	1.73
1500	0.46	0.47	0.49	0.67	0.70	0.70	2.39	2.42	2.45
2000	0.33	0.35	0.36	0.88	0.88	0.89	3.22	3.30	3.41
5000	0.13	0.14	0.14	2.17	2.21	2.23	8.00	8.18	8.43
7000	0.05	0.05	0.06	3.14	3.15	3.16	11.40	11.74	13.25
10000	0.03	0.03	0.04	4.45	4.51	4.52	17.41	17.77	18.43

(*) ≤ 0.3 sec (imprecise measurements).

Table 2: Charging only.

	Proposed Method						CPLEX		
	Opt. Gap (%)			Solve time (min)			Solve time (min)		
# PEVs	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
200	8.82	10.51	12.37	1.05	1.06	1.08	0.06	0.07	0.07
350	2.93	3.24	3.51	1.48	1.49	1.52	1.56	6.89	15.81
500	2.05	2.15	2.24	1.85	1.93	2.62	15.21*	65.10*	262.81*
700	1.48	1.54	1.61	2.43	2.44	2.48	–	–	–
1000	1.01	1.05	1.10	3.24	3.26	3.28	–	–	–
1500	0.65	0.68	0.72	4.72	4.74	4.81	–	–	–
2000	0.45	0.50	0.53	6.19	6.21	6.23	–	–	–
5000	0.12	0.15	0.20	14.88	14.90	14.95	–	–	–
7000	0.09	0.10	0.12	20.59	20.77	21.94	–	–	–
10000	0.06	0.07	0.07	29.34	29.39	29.58	–	–	–

(*) failed to solve two instances (out of memory)

(–) out of memory before attaining the desired optimality gap

Table 3: Charging and V2G.