LARGE-SCALE UNIT COMMITMENT: DECENTRALIZED MIXED INTEGER PROGRAMMING APPROACHES

A Thesis
Presented to
The Academic Faculty

by

Mohammad Javad Feizollahi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2015

LARGE-SCALE UNIT COMMITMENT: DECENTRALIZED MIXED INTEGER PROGRAMMING APPROACHES

Approved by:

Professor Shabbir Ahmed, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Santiago Grijalva
Department of Electrical and Computer Engineering
Georgia Institute of Technology

Professor George L. Nemhauser
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Arkadi Nemirovski
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Andy Sun
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: July 30, 2015

To my parents,

for their tremendous blessing and inspiration,

and to my wife, Maryam,

whose exceptional support and encouragements

made this journey possible,

and to our children, Idin and Alaa.

ACKNOWLEDGEMENTS

I would like to greatly thank Prof. Shabbir Ahmed and Prof. Santiago Grijalva for their exceptional help and supervision during my Ph.D. work. This research would not have been possible without their unreserved support and remarkable vision.

I am very grateful to Prof. George Nemhauser and Prof. Arkadi Nemirovski for serving on my thesis committee, their dedication and all the invaluable things that I have learned from them.

I owe a great deal of thanks to Prof. Andy Sun for his mentorship. He gave me considerable guidance about augmented Lagrangian relaxation and duality. I am grateful for his time and collaboration.

I am grateful to the Georgia Tech professors Natashia Boland, Bill Cook, Jiangang Dai, Santanu Dey, Antonius Dieker, Magnus Egerstedt, Alan Erera, Robert Foley, David Goldberg, David Goldsman, Anton Kleywegt, Eva Lee, Leon McGinnis, Renato Monteiro, Judith Norback, R. Gary Parker, Kamran Paynabar, Sebastian Pokutta, Martin Savelsbergh, Alexander Shapiro, Alejandro Toriello, Craig Tovey and many others.

I wish to acknowledge all professors at the Department of Industrial Engineering at Sharif University of Technology, specially my former advisor Prof. Mohammad Modarres for his endless support. I also want to thank my former supervisor at University of Toronto, Prof. Igor Averback, for all great things that I learned from him and for his continuous support.

I am indebted to my peers for their friendship and support:

- Mitch Costley. Special thanks to Mitch for three years of exceptional collaboration and his tremendous help and comments on unit commitment problem,

power systems, and the software that we developed together. I also thank him for reviewing and proofreading my papers, proposal and some chapters of my dissertation.

- Masoud Nazari. Many thanks for the joint research and his great comments on my papers and dissertation.

- Burak Kocuk and Murat Yildirim. For their friendship and carefully reviewing and proofreading my thesis chapters..

- Researchers in the ACES Lab at Georgia Tech, especially Zachary Costello, Jennifer Howard, Tanguy Hubert, Arun Narasimhan, Marcelo Sandoval, Umer Tariq, Alyse Taylor, James Jamal Thomas, Leilei Xiong, and Sungkap Yeo.

- ISyE PhD students and graduates: Isil Alev, Ilke Bakir, Rodolfo Carvajal, Samaneh Ebrahimi, Erin Garcia, Akshay Gupte, Qie He, Andres Iroume, Tugce Isik, Ben Johnson, Ezgi Karabulut, Fatma Kilinc Karzan, Brian Kues, Junho Lee, Alvaro Lorca Galvez, Aly Megahed, Yasaman Mohammad Shahi, Ahmed Nazeem, Ahmet Ozlu, Alborz Parcham-Kashani, Feng Qui, Diego Ramirez, Kevin Ryan, Soheil Shayegh, Daniel F. Silva, Kael Stilp, Jan Vlachy, Fangfang Xiao, Linwei Xin, Jikai Zou, and many others.

Last but certainly not least, I would like to express my wholehearted gratitude to my wife, Maryam, for her unreserved love and support. This long journey would not have been possible without her patience, support, and persuasion. I am also very grateful to my parents and siblings for their wholehearted and sincere support and encouragements throughout my life.

Contents

List of Tables

List of Figures

SUMMARY

In this dissertation, we investigate theory and application of decentralized optimization for mixed integer programming (MIP) problems. Our focus is on loosely coupled MIPs where different blocks of the problem have mixed integer linear feasible sets and a small number of linear constraints couple these blocks together. We develop decentralized optimization approaches based on Lagrangian and augmented Lagrangian duals for such MIPs. The contributions of this dissertation are a) proof of exactness of augmented Lagrangian dual (ALD) for MIPs, b) decentralized exact and heuristic algorithms for MIPs, and c) application to decentralized unit commitment (UC).

First, we prove that ALD is able to close the duality gap for MIPs. In particular, we show that with non-negative level bounded augmenting functions, ALD is able to asymptotically achieve zero duality gap for MIPs, when the penalty coefficient is allowed to go to infinity. We further show that, under some mild conditions, using any norm as the augmenting function ALD is able to close the duality gap of a MIP with a finite penalty coefficient.

Nonlinear objective functions in ALD destroy the decomposability which exists in classical Lagrangian dual for a loosely coupled MIP. A key challenge is that, because of the non-convex nature of MIPs, classical distributed and decentralized optimization approaches such as alternating direction method of multipliers (ADMM) cannot be applied directly to find their optimal solutions. We propose three exact and one heuristic decentralized algorithms based on extensions of ADMM and dual decomposition techniques.

Finally, we apply the developed algorithms to solve distributed UC. The UC problem deals with the on/off decisions and output power levels of generating units in a power system over a given planning horizon. We present mathematical formulations for the UC problem, which are appropriate for the proposed decentralized algorithms. Privacy concerns of the participants in UC are taken into account in these formulations. We propose a solution approach for decentralized UC which exploits the structure of UC in our decentralized algorithms. We present extensive computational experiments for solving UC instances with different decentralized approaches. We illustrate the challenges arising from nonconvexity of UC problem and show how the proposed algorithms overcome these challenges. We demonstrate remarkable performance of parallel implementation of the heuristic decentralized algorithm to solve large scale UC instances on power systems of more than 3,000 buses. We also show that for small UC instances, the proposed exact algorithms are able to find global optimal solutions.

Chapter I

INTRODUCTION

1.1 Motivation and Contributions

In many real-world optimization problems, discrete integer or binary variables are used to capture system non-convexities. Generally, Mixed Integer Programming (MIP) problems include optimization models with continuous and discrete variables, linear objective function, and linear constraints. Many important applications give rise to large-scale MIP problems with specific structure of loosely coupled blocks of constraints where different blocks of the problem have mixed integer linear feasible sets and a small number of linear constraints couple these blocks together. For instance, energy scheduling problems in large-scale power networks have this structure. In this dissertation, we aim to develop decentralized optimization approaches to tackle large-scale MIP problems such as unit commitment (UC) in the domain of power systems.

The future power grid will be a very large scale dynamic and complex system including Distributed Generation (DG) [27, 124], Home Energy Management Systems (HEMS) [88], energy storage systems [8, 35], PHEV [157, 155, 93, 92, 71], increasing penetration of renewable energies [67], highly elastic and flexible loads [201], microgrids [91, 141, 112], and communications capabilities [1, 114]. These forces introduce new complexity, dimensionality, non-convexity, and uncertainty to the power system, which makes its optimization and control under the existing approaches intractable and inefficient [76, 77, 78, 181].

In the prosumer-based architecture [77, 76] for the future power grid, decentralized control and operation of the grid can play a significant role. In this architecture, a

prosumer is a power subsystem that can produce, consume, store and/or transport electricity. The prosumer concept is highly scalable, and any electric power system, from large interconnections to homes and appliances, can be modeled as a prosumer [77]. Without loss of generality, we assume each prosumer has a different owner. Thus, prosumer owners are not willing to share commercial and sensitive data such as details of cost functions, physical constraints, etc. In this thesis, we will develop a decentralized unit commitment (DUC) approach consistent with this prosumer-based architecture.

Although there are rapidly growing efforts on decentralized or parallel optimization of convex problems in the literature, few solid works with strong theoretical support are available for problems with non-convexities, specifically MIP problems. For more background on decomposition and distributed optimization methods, see [18, 148, 47, 128, 69, 127, 28]. Challenges related to the non-convexity, non-convergence, and inefficiency of naive implementation of classic decentralized approaches for UC problems were observed in [63, 171, 170, 72].

Similar to decomposition techniques, in decentralized approaches a large-scale optimization problem is divided into smaller subproblems. The ability to solve many smaller optimization problems in parallel can provide computational gains for problems of realistic operational size. Decomposition methods such as Benders and Dantzig-Wolfe use a master-slave architecture where subproblems may be solved on separate computational nodes but are coordinated by a master problem, which then requires the results of those subproblems to solve an iteration of its own algorithm. In decentralized approaches, there is no master problem or central computational node.

For many MIP formulations of UC, the difference in computational time between solving for the entire system and, for example, one utility or control area can be of several orders of magnitude. Depending on the convergence characteristics of the overall algorithm versus the complexity of the subproblems, even many hundreds of parallel

iterations may result in total computational time gains. A second motivating factor for UC problem decomposition is the potential for neighboring control regions to coordinate their operation, optimizing their interchange for global benefit. A scheme based on MIP UC formulations, as presented in this dissertation, would not require drastic changes to existing UC software to implement. Finally, the decomposition presented here can preserve much more data privacy for prosumers and participants than is possible today. Only a small amount of data regarding system state must be communicated between neighboring prosumers at each iteration. Sensitive data, such as generator identity, cost, constraints, etc. can remain private to each prosumer. It is for these reasons and because of the increased relevance of storage and demand response that we present and investigate a prosumer-based DUC.

In the case of loosely coupled optimization problems, one possible decentralized approach is to relax the coupling constraints and solve the relaxed problem in parallel by decomposing it into subpoblems. Unlike convex optimization problems, in general for MIP, a non-zero duality gap may exist when coupling constraints are relaxed by using classical Lagrangian or even augmented Lagrangian with finite penalty coefficient [193]; that is, strong duality may not hold. We first investigate the augmented Lagrangian dual (ALD) for MIPs. We provide a primal characterization for ALD and prove that ALD is able to asymptotically achieve zero duality gap for MIPs, when the penalty coefficient is allowed to go to infinity. We also show that, under some mild conditions, ALD using any norm as the augmenting function is able to close the duality gap of the MIP with a finite penalty coefficient.

In the dual decomposition algorithm which is based on the classical Lagrangian dual, each iteration of the algorithm can be performed by parallel computation of subproblems. Although this method is parallelizable, it suffers from non-zero duality gap, oscillation, and traps in local optimality when it is applied to nonconvex optimization problems. Moreover, it is not clear how to recover a primal feasible

solution which satisfies all coupling constraints. Contrarily, augmented Lagrangian methods that can close the duality gap and recover primal feasible solutions are not separable. In the context of convex optimization, a remedy for this challenge is alternating direction method of multipliers (ADMM). ADMM blends the separability of dual decomposition with the superior convergence properties of the method of multipliers (which is based on augmented Lagrangians) [28]. However, discrete variables make MIPs nonconvex, which destroys the convergence properties of ADMM. We propose different exact and heuristic decentralized algorithms for MIPs, which mitigate oscillations and traps in local optimality. The proposed algorithms are based on extensions and modifications of ADMM and dual decomposition methods.

We present scalable mathematical formulations for the UC problem, which are appropriate for the proposed decentralized algorithms. Privacy concerns of the participants in UC taken into account in these formulations. We also propose a solution approach for decentralized UC which exploits the structure of UC in our decentralized algorithms. We demonstrate remarkable performance of parallel implementation of the heuristic decentralized algorithm to solve large scale UC instances on power systems of more than 3000 buses. We also show that for small UC instances, the proposed exact algorithms are able to find the exact optimal solution.

1.2 Dissertation Overview

The structure of this dissertation is as follows. In Chapter 2, we investigate the ALD for MIPs and present some theoretical results on the zero duality gap properties of ALD with specific augmenting functions for MIPs. Considering challenges in direct application of distributed and decentralized optimization techniques for MIPs, we propose several exact and heuristic decentralized algorithms for MIPs in Chapter 3. In Chapter 4, we review the UC problem along with decentralized approaches for UC. Then, we propose a scalable method for formulating and solving a prosumer-based

decentralized UC problem. Extensive computational results for solving UC instances with different proposed decentralized approaches and partitionings are demonstrated in Chapter 5. Finally, conclusions and summary of contributions are given in Chapter 6.

Chapter II

EXACT AUGMENTED LAGRANGIAN DUAL FOR MIPS

In this chapter, we investigate the augmented Lagrangian dual (ALD) for mixed integer linear programming (MIP) problems. ALD modifies the classical Lagrangian dual by appending a nonlinear penalty function on the violation of the dualized constraints in order to reduce the duality gap. We first provide a primal characterization for ALD for MIPs and prove that ALD is able to asymptotically achieve zero duality gap when the weight on the penalty function is allowed to go to infinity. This provides an alternative characterization and proof of a recent result in Boland and Eberhard [26, Proposition 3]. We further show that, under some mild conditions, ALD using any norm as the augmenting function is able to close the duality gap of an MIP with a *finite* penalty coefficient. This generalizes the result in [26, Corollary 1] from pure integer programming problems with bounded feasible region to general MIPs. We also present an example where ALD with a quadratic augmenting function is not able to close the duality gap for any finite penalty coefficient.

2.1 Introduction

We consider the general mixed integer (linear) programming (MIP) problem

$$z^{\mathrm{IP}} := \inf\{\boldsymbol{c}^\top \boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \in X\}, \tag{1}$$

and its augmented Lagrangian dual (ALD)

$$z_\rho^{\mathrm{LD}+} := \sup_{\boldsymbol{\lambda} \in \mathbb{R}^n} \inf_{\boldsymbol{x} \in X} \{\boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})\},$$

where X is a mixed integer linear set, ρ is a given positive scalar, and $\psi(\cdot)$ is an *augmenting function* with $\psi(\boldsymbol{0}) = 0$ and $\psi(\boldsymbol{u}) > 0$ for all $\boldsymbol{u} \neq \boldsymbol{0}$. Here, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ are

the complicating constraints, and relaxing these makes the remaining problem easier. Details of the assumptions are provided in Section 2.3. ALD provides a lower bound for the problem (1), i.e. $z_\rho^{\mathrm{LD}+} \leq z^{\mathrm{IP}}$, for all $\rho > 0$.

We consider non-negative level bounded augmenting functions in ALD for solving MIPs. Because of the non-convexity in MIP (1), a non-zero duality gap may exist [193], that is $z^{\mathrm{IP}} - z_\rho^{\mathrm{LD}+} > 0$. To check the possibility of zero duality gap and exact penalization for MIPs, a possible approach is to verify the general necessary and sufficient conditions for strong duality of augmented Lagrangians based on the structure of value functions [33, 34, 148]. Boland and Eberhard [26] presented an alternative approach and showed that in ALD for MIPs, with a specific class of nonnegative convex augmenting functions, $\lim_{\rho\to\infty} z_\rho^{\mathrm{LD}+} = z^{\mathrm{IP}}$ holds. They also proved that if X is a finite set (e.g. a bounded pure IP), then there exists a finite penalty coefficient which closes the duality gap.

The main contributions of this chapter can be summarized as follows:

1. We first provide a primal characterization for the ALD of an MIP. This is an alternative characterization to the one provided in [26, Theorem 1]. Using this characterization, the ALD of an MIP can be viewed as a traditional Lagrangian dual (LD) in a lifted space.

2. We give an alternative proof for the asymptotic zero duality gap property of ALD for MIPs when the penalty coefficient is allowed to go to infinity. This was first proved in [26, Proposition 3].

3. We prove that ALD using any norm as the augmenting function with a sufficiently large but finite penalty coefficient closes the duality gap for general MIPs. This generalizes the result in [26, Corollary 1] from the case of pure integer programming with a bounded feasible region to general MIPs with unbounded feasible regions.

4. Using our primal characterization, we also present an example where ALD with a quadratic augmenting function is not able to close the duality gap for any finite penalty coefficient.

The chapter is organized as follows. Section 2.2 summarizes the related literature on ALD. Section 2.3 provides definitions and surveys existing results on Lagrangian relaxation and augmented Lagrangian relaxation of general nonlinear optimization problems and specifically of MIPs. Section 2.4 presents a primal characterization of the ALD of a general MIP and the zero duality gap property when the penalty coefficient is allowed to go to infinity. Section 2.5 proves that under mild conditions the ALD achieves zero duality gap using any norm as the augmenting function with a finite penalty coefficient.

2.2 Related Literature

Theory and application of the Lagrangian Dual (LD) for convex optimization problems have been well-studied in the last five decades [87]. Its natural separability and tractability features make LD an appealing decomposition technique for solving the large-scale convex optimization problems. It can also be used to obtain better lower bounds than linear programming (LP) relaxation for discrete optimization problems.

Contrary the convex setting, a non-zero duality gap may exist for nonconvex optimization problems when coupling constraints are relaxed by using classical LD. However, the duality gap can be closed if the dual problem corresponds to some class of functions which are not necessarily affine but they are capable of penetrating possible 'dents' in the *value function* [148] (see Definition 2.2 for value function).

To eliminate or decrease the duality gap, different forms of ALD approaches have been introduced in, e.g. [15, 30, 29, 31, 32, 87, 109, 125, 126, 147, 148, 152, 150, 149, 151, 188, 200]. Depending on the properties of the value function of the underlying optimization problem, different augmenting functions can be used to close the duality

gap. Note that under certain conditions, a *zero duality gap* can be reached asymptotically by increasing the coefficient on penalty function to infinity [188]. In some cases, duality gap can be closed with a large enough *finite* value of the penalty coefficient. In this case, we say that the corresponding ALD involves *exact penalization* or is *exact*.

In the convex optimization setting, the ALD needs weaker assumptions to converge and demonstrate more robust behavior than standard LD [148]. In contrast, nonlinear augmented term destroys the natural separability properties of the LD. A remedy for this drawback is the *alternating direction method of multipliers* (ADMM) which combines the robustness of ALD and the separability of LD [28].

Rockafellar [147] and Bertsekas [17] used convex quadratic augmenting functions. [15] proposes penalty/barrier multiplier methods which are nonquadratic augmented Lagrangians. Burke [33, 34] used norms as convex augmenting functions. For these cases, necessary and sufficient conditions for exact penalization are provided in [33, 34, 148] which we will recast in Section 2.3.

For some classes of non-convex optimization problems, the duality gap cannot be closed by using convex augmenting functions. For these problems, more general forms of ALD are needed. For example, *level-bounded* augmenting functions were used in [87] rather than non-negative convex ones. The works in [149] and [151] used a family of augmenting functions with *almost peak at zero* property, which includes the augmenting functions in [87] as special cases. Note that the class of augmenting functions in [149] and [151] are generalizations of convex augmenting functions in [148]. A weaker peak at zero property was considered in [125]. A more general form of peak at zero property was investigated in [188] to provide a unified nonlinear ALD. Using the theory of *abstract convexity*, ALD was studied in [32] and [30] in Banach and Hausdorff topological spaces, respectively. For detailed definitions of terms such as level-boundedness, peak at zero, etc. see Appendix B.

2.3 Preliminaries

Let $\mathbb{R}$, $\mathbb{Q}$, and $\mathbb{Z}$ denote the sets of real, rational, and integer numbers, respectively. Let $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ be the set of extended real numbers. For any vector $\boldsymbol{a}$ and matrix $\boldsymbol{A}$ with finite dimensions, denote their transpose by $\boldsymbol{a}^\top$ and $\boldsymbol{A}^\top$, respectively. For a given $\overline{\boldsymbol{x}} \in \mathbb{R}^n$ and $\delta > 0$, $\mathbb{B}(\overline{\boldsymbol{x}}, \delta) \subset \mathbb{R}^n$ denotes the open Euclidean ball with center $\overline{\boldsymbol{x}}$ and radius δ. For any set $S \subseteq \mathbb{R}^n$, let $\mathbf{conv}(S)$, $\mathbf{ri}(S)$ and $\mathbf{cl}(S)$ denote the convex hull, relative interior and closure of the set S, respectively. For a set $T \subset \mathbb{R}^{n+1}$ of vectors $(\boldsymbol{x}, \omega)$, $\text{Proj}_x(T) := \{\boldsymbol{x} \in \mathbb{R}^n : \exists \omega \text{ s.t. } (\boldsymbol{x}, \omega) \in T\}$ denotes the projection of T into the space of vectors $\boldsymbol{x}$ in $\mathbb{R}^n$. Moreover, let $\mathbf{diam}(S) := \sup\{\|\boldsymbol{u} - \boldsymbol{v}\|_\infty : \boldsymbol{u} \in S, \boldsymbol{v} \in S\}$ denote the diameter of set S, where $\|\cdot\|_\infty$ is the l^∞ norm.

Let $\boldsymbol{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2}$ be the vector of decision variables, where n_1 and n_2 are numbers of integer and continuous variables, respectively, and $n := n_1 + n_2$. For given $\boldsymbol{c} \in \mathbb{Q}^n$, $\boldsymbol{b} \in \mathbb{Q}^m$, and $\boldsymbol{A} \in \mathbb{Q}^{m \times n}$, consider the general MIP problem (1),

$$z^{\text{IP}} := \inf\{\boldsymbol{c}^\top \boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \in X\},$$

where m is the number of *complicating* or *coupling constraints*, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. The case with $n_2 = 0$ is called a *pure IP*, while for a MIP we have $n_2 \geq 1$ and $n_1 \geq 1$. Denote the LP relaxation of z^{IP} in problem (1) by z^{LP}. In this chapter, we consider MIP problems that satisfy the following assumption.

Assumption 2.1. *For the MIP* (1) *we have the following:*

a) X is a mixed integer linear set given by $X = \{\boldsymbol{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2} : \boldsymbol{E}\boldsymbol{x} \leq \boldsymbol{f}\}$ for some $\boldsymbol{E} \in \mathbb{Q}^{\bar{m} \times n}$ and $\boldsymbol{f} \in \mathbb{Q}^{\bar{m}}$, where $\bar{m}$ is the number of the inequality constraints in the definition of X. The problem data $\boldsymbol{A}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{E}$, and $\boldsymbol{f}$ all have rational entries, and without loss of generality, we can assume that they are integral.

b) Problem (1) *is feasible and its optimal value is bounded.*

Usually problem (1) is taken to be structured so that X includes integrality constraints, simple bounds on variables, and other simple constraints.

Remark 1. *Note that under Assumption 2.1-a,* $\mathbf{conv}(X)$ *and* $\mathbf{conv}(\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \in X\})$ *are rational polyhedra by Meyer's theorem [116]. By Assumption 2.1 (rationality of input data and boundedness of* z^{IP}*), the value of the LP relaxation of MIP* (1) *is bounded [24], i.e.* $-\infty < z^{LP} \leq z^{IP} < \infty$*. Let* $\bar{\boldsymbol{\lambda}}_{LP}$ *be a rational optimal vector of dual variables for* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *in the LP relaxation of* (1)*. Moreover,* z^{IP} *is attainable and the* inf *in the objective function of* (1) *can be replaced by* min*. That is, there exists an* optimal solution $\boldsymbol{x}^*$ *of problem* (1) *such that* $\boldsymbol{x}^* \in X$*,* $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$ *and* $z^{IP} = \boldsymbol{c}^\top \boldsymbol{x}^*$*.*

It is worth mentioning that the equality relation in $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ does not impose any restriction on the type of these constraints. Because any inequality can be replaced by an equality constraint with a new non-negative slack variable. The non-negativity of the introduced variable can be absorbed in X. Moreover, in the case of a pure IP, this slack variable will automatically be an integer variable following Assumption 2.1-a.

Definition 2.2. *(Value function). The value function (also known as the performance function, the marginal function and the perturbation function) for problem* (1) *is defined as*

$$p(\boldsymbol{u}) := \inf\{\boldsymbol{c}^\top \boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} + \boldsymbol{u}, \boldsymbol{x} \in X\}. \tag{2}$$

Note that $p(\mathbf{0}) = z^{\mathrm{IP}}$. The value function is a very important tool for the theoretical examination of constrained optimization problems [149]. The properties of the value functions for IPs and MIPs were studied in [23, 24, 25, 117, 144]. For an MIP problem with rational data, the value function is lower semicontinuous [117] and piecewise polyhedral with finitely many pieces in any bounded set [23].

Definition 2.3. *(Lagrangian relaxation and dual). For a given Lagrange multiplier vector* $\boldsymbol{\lambda} \in \mathbb{R}^m$*, the corresponding Lagrangian relaxation (LR) of* (1) *is given as*

$$z^{LR}(\boldsymbol{\lambda}) := \inf_{\boldsymbol{x} \in X} \{\boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})\}, \tag{3}$$

and the associated Lagrangian dual (LD) is

$$z^{LD} := \sup_{\boldsymbol{\lambda} \in \mathbb{R}^m} z^{LR}(\boldsymbol{\lambda}). \tag{4}$$

A well known primal characterization of LD is given by [73] as

$$z^{\mathrm{LD}} = \inf_{\boldsymbol{x}} \{\boldsymbol{c}^\top \boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \in \mathbf{conv}(X)\}. \tag{5}$$

Remark 2. *Note that by rationality of the input data in Assumption 2.1,* z^{LD} *is attainable and* inf *in the objective function of* (5) *can be replaced by* min.

Definition 2.4. *(Augmented Lagrangian relaxation and dual). The augmented Lagrangian relaxation (ALR) of* (1) *has the following form [148]:*

$$z_\rho^{LR+}(\boldsymbol{\lambda}) := \inf_{\boldsymbol{x} \in X} \{\boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})\}. \tag{6}$$

Here, $\rho > 0$ *is a fixed given parameter called* penalty coefficient *and* ψ *is an augmenting function. In this chapter, unless explicitly mentioned, we assume that* ψ *satisfies the following assumption.*

Assumption 2.5. $\psi : \mathbb{R}^m \to \mathbb{R}_+$ *is a proper, nonnegative, lower semicontinuous, and level-bounded augmenting function, that is* $\psi(\mathbf{0}) = 0$*,* $\psi(\boldsymbol{u}) > 0$ *for all* $\boldsymbol{u} \neq \mathbf{0}$*,* $\mathbf{diam}\{\boldsymbol{u} \mid \psi(\boldsymbol{u}) \leq \delta\} < +\infty$ *for all* $\delta > 0$*. Moreover* $\lim_{\delta \downarrow 0} \mathbf{diam}\{\boldsymbol{u} \mid \psi(\boldsymbol{u}) \leq \delta\} = 0$*.*

Note that non-negative convex augmenting functions satisfy Assumption 2.5. The augmented Lagrangian dual (ALD) is as follows.

$$z_\rho^{LD+} := \sup_{\boldsymbol{\lambda} \in \mathbb{R}^m} z_\rho^{LR+}(\boldsymbol{\lambda}). \tag{7}$$

Since the augmenting function $\psi(\cdot)$ is nonnegative, $z_\rho^{\mathrm{LD}+}$ is a non-decreasing function of ρ. Moreover, since $\rho\psi(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) \geq 0$ we have $z^{\mathrm{LR}}(\boldsymbol{\lambda}) \leq z_\rho^{\mathrm{LR}+}(\boldsymbol{\lambda})$ for all $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\rho > 0$, because $\rho\psi(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) \geq 0$. Therefore, $z^{\mathrm{LD}} \leq z_\rho^{\mathrm{LD}+}$ for any $\rho > 0$. Moreover

$$\begin{aligned} z_\rho^{\mathrm{LR}+}(\boldsymbol{\lambda}) &\leq \inf_{\boldsymbol{x}\in X:\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}} \{\boldsymbol{c}^\top\boldsymbol{x} + \boldsymbol{\lambda}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho\psi(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})\} \\ &= \inf_{\boldsymbol{x}\in X:\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}} \boldsymbol{c}^\top\boldsymbol{x} \\ &= z^{\mathrm{IP}}, \end{aligned}$$

where the inequality holds because $\{\boldsymbol{x} \in X : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\} \subseteq X$. The first equality follows from the facts that $\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ for all $\boldsymbol{x}$ satisfying $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, and $\psi(\boldsymbol{0}) = 0$. The second equation holds by definition of z^{IP}. Thus, $z_\rho^{\mathrm{LD}+} \leq z^{\mathrm{IP}}$. In summary,

$$-\infty < z^{\mathrm{LP}} \leq z^{\mathrm{LD}} \leq z_\rho^{\mathrm{LD}+} \leq z^{\mathrm{IP}} < +\infty, \quad \forall \rho > 0. \tag{8}$$

where the strict inequalities in the upper and lower bounds hold from Assumption 2.1.

2.3.1 Exact Penalty Representation

Definition 2.6. *(Exact penalty representation [148, Definition 11.60]). A dual vector* $\bar{\boldsymbol{\lambda}}$ *is said to support an* exact penalty representation *for problem* (1) *if, for all* ρ *sufficiently large,* $z^{IP} = z_\rho^{LR+}(\bar{\boldsymbol{\lambda}})$ *and*

$$\operatorname*{argmin}_{\boldsymbol{x}\in X:\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}} \boldsymbol{c}^\top\boldsymbol{x} = \operatorname*{argmin}_{\boldsymbol{x}\in X}\{\boldsymbol{c}^\top\boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho\psi(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})\}.$$

The smallest ρ which satisfies this property is called the *adequate penalty threshold*. A criterion for exact the penalty representation presented in [148] is as follows. Suppose that $z_\rho^{\mathrm{LR}+}(\boldsymbol{\lambda}) > -\infty$ for at least one $(\boldsymbol{\lambda}, \rho) \in \mathbb{R}^m \times (0, \infty)$. Then, a vector $\bar{\boldsymbol{\lambda}}$ supports an exact penalty representation for problem (1) if and only if there exist an open neighborhood $W \subset \mathbb{R}^m$ of $\boldsymbol{0}$ and a scalar $\hat{\rho} > 0$ such that

$$p(\boldsymbol{u}) \geq p(\boldsymbol{0}) - \bar{\boldsymbol{\lambda}}^\top\boldsymbol{u} - \hat{\rho}\psi(\boldsymbol{u}), \quad \forall \boldsymbol{u} \in W.$$

If $z_\rho^{\mathrm{LD}+} = z^{\mathrm{IP}}$ for some $\rho > 0$, then ALR (6) can recover a primal solution for the MIP problem (1).

Proposition 2.7. *Suppose Assumption 2.1 holds and $z^{IP} = z_{\hat{\rho}}^{LD+} = z_{\hat{\rho}}^{LR+}(\bar{\boldsymbol{\lambda}})$ for some finite $\hat{\rho} > 0$ and $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^m$. Then, any optimal solution of ALR (6) with $\boldsymbol{\lambda} = \bar{\boldsymbol{\lambda}}$ and $\rho = \rho^* > \hat{\rho}$ solves the MIP problem (1) and vice versa.*

Proof. Let ρ^* be any scalar such that $\rho^* > \hat{\rho}$. Let $\bar{\boldsymbol{x}}$ be an optimal solution of MIP problem (1). The existence of an optimal solution for problem (1) is guaranteed under Assumption 2.1. Then, it holds that $\bar{\boldsymbol{x}} \in X$, $\boldsymbol{A}\bar{\boldsymbol{x}} = \boldsymbol{b}$, and $\boldsymbol{c}^\top \bar{\boldsymbol{x}} = z^{\mathrm{IP}}$. Thus,

$$\boldsymbol{c}^\top \bar{\boldsymbol{x}} + \bar{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\bar{\boldsymbol{x}}) + \rho^* \psi(\boldsymbol{b} - \boldsymbol{A}\bar{\boldsymbol{x}}) = \boldsymbol{c}^\top \bar{\boldsymbol{x}} = z^{\mathrm{IP}} = z_{\rho^*}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}).$$

where the last equality follows from the facts that $z_{\hat{\rho}}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}) \leq z_{\rho^*}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}) \leq z^{\mathrm{IP}}$ and $z^{\mathrm{IP}} = z_{\hat{\rho}}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}})$. Therefore, $\bar{\boldsymbol{x}}$ solves ALR (6) with ρ^* and $\bar{\boldsymbol{\lambda}}$. Moreover, it shows that the optimality is achieved for this case of ALR (6).

To prove the other side, let $\boldsymbol{x}^* \in X$ be any optimal solution of ALR (6) with ρ^* and $\bar{\boldsymbol{\lambda}}$, i.e. $\boldsymbol{c}^\top \boldsymbol{x}^* = z_{\rho^*}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}})$. We claim that $\boldsymbol{x}^*$ solves problem (1), i.e. $\boldsymbol{x}^* \in X$, $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$ and $\boldsymbol{c}^\top \boldsymbol{x}^* = z^{\mathrm{IP}}$. Note that as a feasible solution of ALR (6), $\boldsymbol{x}^*$ belongs to X. Assume by contradiction $\boldsymbol{A}\boldsymbol{x}^* \neq \boldsymbol{b}$. Then, $\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*) > 0$ and therefore

$$\hat{\rho}\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*) < \rho^* \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*). \tag{9}$$

Moreover,

$$\begin{aligned} z^{\mathrm{IP}} = z_{\hat{\rho}}^{\mathrm{LD}+} = z_{\hat{\rho}}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}) &\leq \boldsymbol{c}^\top \boldsymbol{x}^* + \bar{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*) + \hat{\rho}\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*) \\ &< \boldsymbol{c}^\top \boldsymbol{x}^* + \bar{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*) + \rho^* \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*) \\ &= z_{\rho^*}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}), \end{aligned} \tag{10}$$

which contradicts $z_{\rho^*}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}})$ being a lower bound for z^{IP}. Therefore, $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$. Note that in (10) the equality relations hold by assumption, the first inequality holds by

definition of $z_{\hat{\rho}}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}})$, and the strict inequality follows from (9). Furthermore,

$$\begin{aligned} z^{\mathrm{IP}} &= z_{\hat{\rho}}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}) \\ &\leq z_{\rho^*}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}) = \boldsymbol{c}^\top \boldsymbol{x}^* + \bar{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*) + \rho^* \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*) = \boldsymbol{c}^\top \boldsymbol{x}^* \\ &\leq z^{\mathrm{IP}}, \end{aligned} \tag{11}$$

where the first two equalities hold by assumption and the third equality follows from $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$. Therefore, $\boldsymbol{c}^\top \boldsymbol{x}^* = z^{\mathrm{IP}}$ which completes the proof. □

Two important cases of ALR are the proximal and sharp Lagrangian. Next, we present their definitions, and necessary and sufficient conditions for supporting an exact penalty representation in these cases.

2.3.2 Proximal Lagrangian

Definition 2.8. *(Proximal Lagrangian). An ALR generated with the augmenting function $\psi(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{u}\|_2^2$ is called a proximal Lagrangian.*

Definition 2.9. *(Proximal subgradient [148, Definition 8.45]). A vector $\boldsymbol{\lambda} \in \mathbb{R}^m$ is called a proximal subgradient of a function $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ at $\bar{\boldsymbol{u}}$, a point where $f(\bar{\boldsymbol{u}})$ is finite, if there exist $\rho > 0$ and $\delta > 0$ such that*

$$f(\boldsymbol{u}) \geq f(\bar{\boldsymbol{u}}) - \boldsymbol{\lambda}^\top (\boldsymbol{u} - \bar{\boldsymbol{u}}) - \frac{1}{2}\rho \|\boldsymbol{u} - \bar{\boldsymbol{u}}\|_2^2, \ \forall \boldsymbol{u} \ s.t. \ \|\boldsymbol{u} - \bar{\boldsymbol{u}}\|_2 \leq \delta.$$

The existence of a proximal subgradient at $\bar{\boldsymbol{u}}$ corresponds to the existence of a 'local quadratic support' to f at $\bar{\boldsymbol{u}}$.

In a proximal Lagrangian, suppose that there exists $(\boldsymbol{\lambda}, \rho) \in \mathbb{R}^n \times (0, \infty)$ such that $z_{\rho}^{\mathrm{LR}+}(\boldsymbol{\lambda}) > -\infty$. Then, a necessary and sufficient condition for a vector $\bar{\boldsymbol{\lambda}}$ to support an exact penalty representation is that $\bar{\boldsymbol{\lambda}}$ is a proximal subgradient of the value function $p(\boldsymbol{u})$ at $\boldsymbol{u} = \boldsymbol{0}$ [148].

2.3.3 Sharp Lagrangian

Definition 2.10. *(Sharp Lagrangian). An ALR which uses a norm as an augmenting function, i.e. $\psi(\boldsymbol{u}) = \|\boldsymbol{u}\|$, is called a sharp Lagrangian.*

Definition 2.11. *(Calmness [148, Ch. 8.F]).* A function $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ is *calm at* $\overline{\boldsymbol{u}}$ *from below* with modulus $\kappa \in \mathbb{R}_+$ if $f(\overline{\boldsymbol{u}})$ is finite and on some open neighborhood V of $\overline{\boldsymbol{u}}$, one has

$$f(\boldsymbol{u}) \geq f(\overline{\boldsymbol{u}}) - \kappa \|\boldsymbol{u} - \overline{\boldsymbol{u}}\|, \ \forall \boldsymbol{u} \in V.$$

Note that Definition 2.11 of calmness at $\overline{\boldsymbol{u}}$ from below is equivalent to satisfying the condition

$$\liminf_{\boldsymbol{u} \to \overline{\boldsymbol{u}}} \frac{f(\boldsymbol{u}) - f(\overline{\boldsymbol{u}})}{\|\boldsymbol{u} - \overline{\boldsymbol{u}}\|} > -\infty,$$

which was given in [34]. Consider a function f which is not calm at $\overline{\boldsymbol{u}}$ from below. Then, a small shift in $\boldsymbol{u}$ can produce a proportionally unbounded downward shift in f. Calmness is a basic regularity condition under which we can study the sensitivity properties of certain variational systems [45].

In the sharp Lagrangian, suppose that $z_\rho^{\mathrm{LR}+}(\mathbf{0}) > -\infty$ for some $\rho \in (0, \infty)$. Then, a necessary and sufficient condition for the vector $\bar{\boldsymbol{\lambda}} = \mathbf{0}$ to support an exact penalty representation is that the value function $p(\boldsymbol{u})$ is calm from below at $\boldsymbol{u} = \mathbf{0}$ [33, 34, 148].

2.3.4 ALD for MIPs

For the MIP problem (1), under some technical assumptions, Boland and Eberhard [26] showed that the duality gap for ALD, $z_\rho^{\mathrm{LD}+} - z^{\mathrm{IP}}$, goes to zero as the penalty coefficient ρ goes to infinity.

Proposition 2.12. *[26, Proposition 3] Suppose ψ is of the form $\psi(\boldsymbol{u}) = \phi(\|\boldsymbol{u}\|)$ for some norm $\|\cdot\|$ in $\mathbb{R}^m$ where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a convex, monotonically increasing function for which $\phi(0) = 0$ and there exists $\delta > 0$ for which*

$$\liminf_{a \to +\infty} \frac{\phi(a)}{a} \geq \delta > 0$$

with $\mathbf{diam}\{\boldsymbol{a} | \phi(\boldsymbol{a}) \leq \delta\} \downarrow 0$ *as* $\delta \downarrow 0$. *Moreover, at least one of the following conditions holds: 1) The solution set of the LP relaxation of problem* (1) *does not contain*

a lineality space. 2) The matrices $\boldsymbol{A}$ *and* $\boldsymbol{D}$ *have rational entries and the norm* $\|.\|$ *used in the definition of* ψ *is the* l_∞ *norm. 3)* $\mathbf{conv}(X)$ *is bounded. Then*

$$z^{LD^*} := \sup_{\rho>0} z_\rho^{LD+} = \lim_{\rho\to\infty} z_\rho^{LD+} = z^{IP}.$$

Boland and Eberhard [26] also showed that if X is a finite set of discrete elements then ρ does not need to go to infinity to close duality gap.

Corollary 2.13. *[26, Corollary 1] Suppose* X *is a finite set and assumptions in Proposition 2.12 hold. Then, there exists a* ρ^* *with* $0 < \rho^* < \infty$ *such that* $z_{\rho^*}^{LD+} = z^{IP}$.

2.4 Zero Duality Gap with ALD

In this section, we first present a primal characterization of the ALD for MIPs. Then, we prove that strong duality holds for ALD of general MIPs when the penalty coefficient is allowed to go to infinity. Our primal characterization and the strong duality result hold for a general, not necessarily convex augmenting function, satisfying Assumption 2. We also discuss the relation of our results to the recent results in [26].

2.4.1 A Primal Characterization of ALD

Similar to the equivalence of (4) and (5) for the LD, we can give a primal characterization for the ALD problem (7). The key observation is that (7) can be viewed as an LD of a problem in a lifted space. Then, the primal characterization follows from strong duality in convex optimization with usual regularity conditions.

Let us first find the primal problem for the ALD problem (7).

$$\begin{aligned} z_\rho^{\mathrm{LD}+} &= \sup_{\lambda\in\mathbb{R}^m} \inf_{\boldsymbol{x}\in X} \{\boldsymbol{c}^\top\boldsymbol{x} + \boldsymbol{\lambda}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho\psi(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})\} \\ &= \sup_{\lambda\in\mathbb{R}^m} \inf_{\boldsymbol{x}\in X, \psi(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})\le\omega} \{\boldsymbol{c}^\top\boldsymbol{x} + \rho\omega + \boldsymbol{\lambda}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})\} \qquad (12) \\ &= \sup_{\lambda\in\mathbb{R}^m} \inf_{\boldsymbol{x},\omega} \{\boldsymbol{c}^\top\boldsymbol{x} + \rho\omega + \boldsymbol{\lambda}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) \ : \ (\boldsymbol{x},\omega)\in \mathbf{conv}(S_\psi)\}, \qquad (13) \end{aligned}$$

where S_ψ denotes the feasible region of the inf problem in (12), i.e.

$$S_\psi := \left\{(\boldsymbol{x}, \omega) \in \mathbb{R}^{n+1} : \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \leq \omega,\ \boldsymbol{x} \in X\right\}, \tag{14}$$

and (13) holds because the objective function in (12) is linear. Now switching the sup and inf in (13), we have the dual problem of (13) given as

$$\begin{aligned}\hat{z}_\rho^{\mathrm{LD}+} &:= \inf_{(\boldsymbol{x},\omega)\in\mathbf{conv}(S_\psi)} \sup_{\lambda\in\mathbb{R}^m} \{\boldsymbol{c}^\top \boldsymbol{x} + \rho\omega + \boldsymbol{\lambda}^\top(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})\} \\ &= \inf_{\boldsymbol{x},\omega}\{\boldsymbol{c}^\top \boldsymbol{x} + \rho\omega : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, (\boldsymbol{x}, \omega) \in \mathbf{conv}(S_\psi)\}. \end{aligned} \tag{15}$$

Theorem 2.18 below shows that, under a mild regularity condition, strong duality holds between (13) and (15), i.e. $z_\rho^{\mathrm{LD}+} = \hat{z}_\rho^{\mathrm{LD}+}$. Note that (15) only involves primal variables $\boldsymbol{x}, \omega$. Therefore, this gives a primal characterization of the ALD problem (13). To prove this result, we need a few simple propositions and a nonlinear Farkas lemma.

Proposition 2.14. $\mathrm{Proj}_x(\mathbf{conv}(S_\psi)) = \mathbf{conv}(X)$.

Proof. For any $(\boldsymbol{x}, \omega) \in \mathbf{conv}(S_\psi)$, there exist $\boldsymbol{x}^i \in X$ and $\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^i) \leq \omega^i$ for $i = 1, \ldots, n+2$ so that $\boldsymbol{x} = \sum_{i=1}^{n+2} \lambda_i \boldsymbol{x}^i$, $\omega = \sum_{i=1}^{n+2} \lambda_i \omega^i$, and $\sum_{i=1}^{n+2} \lambda_i = 1$ with $\lambda_i \geq 0$ for all $i = 1, \ldots, n+2$ (by Caratheodory's Theorem). Clearly, $\boldsymbol{x} \in \mathbf{conv}(X)$, which shows $\mathrm{Proj}_x(\mathbf{conv}(S_\psi)) \subseteq \mathbf{conv}(X)$.

For the other direction, take any $\boldsymbol{x} \in \mathbf{conv}(X)$. Then $\boldsymbol{x}$ can be written as $\boldsymbol{x} = \sum_{i=1}^{n+1} \lambda_i \boldsymbol{x}^i$ for each $\boldsymbol{x}^i \in X$ and λ_i's form a convex combination. Let $\omega_i := \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^i)$ and $\omega := \sum_i \lambda_i \omega_i$. Then, for each i, $(\boldsymbol{x}^i, \omega^i) \in S_\psi$, and $(\boldsymbol{x}, \omega) = \sum_i \lambda_i (\boldsymbol{x}^i, \omega^i)$. Therefore, $(\boldsymbol{x}, \omega) \in \mathbf{conv}(S_\psi)$, i.e. $\boldsymbol{x} \in \mathrm{Proj}_x(\mathbf{conv}(S_\psi))$. This completes the proof. □

Proposition 2.15. *Let S be a nonempty convex set in $\mathbb{R}^{n+1}$. Then* $\mathbf{ri}(\mathrm{Proj}_x(S)) = \mathrm{Proj}_x(\mathbf{ri}(S))$.

This follows from the well-known fact $\mathbf{ri}(A(S)) = A(\mathbf{ri}(S))$, where A is a linear transformation and S is a convex set. See e.g., [19].

Proposition 2.16. *There exists* $\boldsymbol{x} \in \mathbf{ri}(\mathbf{conv}(X))$ *and* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *if and only if Problem* (15) *has a feasible point in* $\mathbf{ri}(\mathbf{conv}(S_\psi))$.

Proof. If (15) has a feasible point $(\bar{\boldsymbol{x}}, \bar{\omega})$ in $\mathbf{ri}(\mathbf{conv}(S_\psi))$, then $\boldsymbol{A}\bar{\boldsymbol{x}} = \boldsymbol{b}$ and $\bar{\boldsymbol{x}} \in \mathrm{Proj}_x(\mathbf{ri}(\mathbf{conv}(S_\psi)))$. by Proposition 2.15, $\bar{\boldsymbol{x}} \in \mathbf{ri}(\mathrm{Proj}_x(\mathbf{conv}(S_\psi)))$. By Proposition 2.14, we have $\bar{\boldsymbol{x}} \in \mathbf{ri}(\mathbf{conv}(X))$.

For the other direction, take any $\bar{\boldsymbol{x}} \in \mathbf{ri}(\mathbf{conv}(X))$ and $\boldsymbol{A}\bar{\boldsymbol{x}} = \boldsymbol{b}$. By Proposition 2.14, we have $\bar{\boldsymbol{x}} \in \mathbf{ri}(\mathrm{Proj}_x(\mathbf{conv}(S_\psi)))$. By Proposition 2.15, then we know that $\bar{\boldsymbol{x}} \in \mathrm{Proj}_x(\mathbf{ri}(\mathbf{conv}(S_\psi)))$, i.e. there exists $(\bar{\boldsymbol{x}}, \bar{\omega}) \in \mathbf{ri}(\mathbf{conv}(S_\psi))$ and $\boldsymbol{A}\bar{\boldsymbol{x}} = \boldsymbol{b}$. □

Lemma 2.17 (Nonlinear Farkas' Lemma (Prop. 3.5.4, [19]))**.** *Let* C *be a nonempty convex subset of* $\mathbb{R}^n$*, and let* $f : C \to \mathbb{R}$ *and* $g_j : C \to \mathbb{R}$*, for* $j = 1, \ldots, r$ *be convex functions. Consider the set* F *given by* $F = \{\boldsymbol{x} \in C : g(\boldsymbol{x}) \leq 0\}$*, where* $g(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_r(\boldsymbol{x}))$*, and assume that* $f(\boldsymbol{x}) \geq 0$ *for all* $\boldsymbol{x} \in F$*. Consider the subset* Q^* *of* $\mathbb{R}^r$ *given by*

$$Q^* = \left\{\boldsymbol{\lambda} \in \mathbb{R}^r : \boldsymbol{\lambda} \geq \mathbf{0}, f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top g(\boldsymbol{x}) \geq 0, \forall \boldsymbol{x} \in C\right\}.$$

Then, Q^* *is nonempty if the functions* g_j *for* $j = 1, \ldots, r$ *are affine, and* F *contains a relative interior point of* C.

Next, we present the primal characterization of the ALD problem (7) as following theorem.

Theorem 2.18. *If there exists* $\boldsymbol{x} \in \mathbf{ri}(\mathbf{conv}(X))$ *such that* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$*, and* $z_\rho^{LD+} > -\infty$*, then for all* $\rho > 0$*,*

$$z_\rho^{LD+} = \inf_{\boldsymbol{x},\omega} \quad \boldsymbol{c}^\top \boldsymbol{x} + \rho\omega \tag{16a}$$

$$s.t. \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{16b}$$

$$(\boldsymbol{x}, \omega) \in \mathbf{conv}(S_\psi). \tag{16c}$$

Proof. Essentially, we want to show that strong duality holds between the primal and dual pair of convex programs (13) and (15), i.e. $\hat{z}_\rho^{LD+} = z_\rho^{LD+}$ (recall that $\hat{z}_\rho^{LD+}$ is defined in (15)). By Proposition 2.16, if there exists $\boldsymbol{x} \in \mathbf{ri}(\mathbf{conv}(X))$ and $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, then (16) has a feasible point in $\mathbf{ri}(\mathbf{conv}(S_\psi))$. To apply the nonlinear Farkas' lemma, we first rewrite the linear equality constraints in (16b) as linear inequalities $\tilde{\boldsymbol{A}}\boldsymbol{x} \leq \tilde{\boldsymbol{b}}$ with $\tilde{\boldsymbol{A}} = [\boldsymbol{A}^\top, -\boldsymbol{A}^\top]^\top$ and $\tilde{\boldsymbol{b}} = [\boldsymbol{b}^\top, -\boldsymbol{b}^\top]^\top$; we can also subtract $\hat{z}_\rho^{LD+}$ from the objective function of (16a) so the new optimal value is zero. Furthermore, denote the feasible region of (16) as

$$F := \left\{ (\boldsymbol{x}, \omega) \in \mathbf{conv}(S_\psi) : \tilde{\boldsymbol{A}}\boldsymbol{x} \leq \tilde{\boldsymbol{b}} \right\}.$$

Since F contains a point in the relative interior of $\mathbf{conv}(S_\psi)$, by Lemma 2.17, we know that there exists a multiplier vector $\boldsymbol{\lambda}^* \leq \mathbf{0}$ such that

$$\boldsymbol{c}^\top \boldsymbol{x} + \rho\omega - \hat{z}_\rho^{\mathrm{LD}+} + (\boldsymbol{\lambda}^*)^\top (\tilde{\boldsymbol{b}} - \tilde{\boldsymbol{A}}\boldsymbol{x}) \geq 0, \qquad \forall (\boldsymbol{x}, \omega) \in \mathbf{conv}(S_\psi).$$

From this, we obtain

$$\inf_{(\boldsymbol{x},\omega)\in\mathbf{conv}(S_\psi)} \boldsymbol{c}^\top \boldsymbol{x} + \rho\omega + (\boldsymbol{\lambda}^*)^\top (\tilde{\boldsymbol{b}} - \tilde{\boldsymbol{A}}\boldsymbol{x}) \geq \hat{z}_\rho^{LD+}$$

$$\Rightarrow z_\rho^{LD+} = \sup_{\boldsymbol{\lambda}\leq \mathbf{0}} \inf_{(\boldsymbol{x},\omega)\in\mathbf{conv}(S_\psi)} \boldsymbol{c}^\top \boldsymbol{x} + \rho\omega + \boldsymbol{\lambda}^\top (\tilde{\boldsymbol{b}} - \tilde{\boldsymbol{A}}\boldsymbol{x}) \geq \hat{z}_\rho^{LD+}.$$

By the weak duality between (13) and (15), we already have $z_\rho^{LD+} \leq \hat{z}_\rho^{LD+}$, therefore, this shows that $z_\rho^{\mathrm{LD}+} = \hat{z}_\rho^{\mathrm{LD}+}$ for all $\rho > 0$. □

Remark 3. *A similar primal characterization of* (7) *is given in [26, Theorem 1]. In particular, the primal characterization in [26] has the following form*

$$z_\rho^{LD+} = \min_{\hat{\omega}>0} \left\{ \rho\hat{\omega} + \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \in X_\psi(\hat{\omega}) \right\} \right\}, \tag{17}$$

where $X_\psi(\hat{\omega}) := \mathbf{conv}(\{\boldsymbol{x} \in \mathbb{R}^n : \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \leq \hat{\omega},\ \boldsymbol{x} \in X\})$. *Note that* (17) *first minimizes over* $\hat{\omega}$ *then over* $\boldsymbol{x}$*, whereas the primal characterization obtained in* (16) *minimizes* $\boldsymbol{x}, \omega$ *jointly. Of course,* (16) *can also be written in this order as*

$$z_\rho^{LD+} = \min_{\hat{\omega}>0} \left\{ \rho\hat{\omega} + \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \in X'_\psi(\hat{\omega}) \right\} \right\}, \tag{18}$$

where $X'_\psi(\hat{\omega}) := \{\boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x}, \hat{\omega}) \in \mathbf{conv}(S_\psi)\}$. *The difference between* (17) *and* (16) *is more clear if we rewrite the sets* $X_\psi(\hat{\omega})$ *and* $X'_\psi(\hat{\omega})$ *as follows,*

$$\begin{aligned} X_\psi(\hat{\omega}) &= Proj_x \left(\mathbf{conv} \left(S_\psi \cap \{(\boldsymbol{x}, \omega) : \omega = \hat{\omega}\} \right) \right) \\ X'_\psi(\hat{\omega}) &= Proj_x \left(\mathbf{conv} \left(S_\psi \right) \cap \{(\boldsymbol{x}, \omega) : \omega = \hat{\omega}\} \right). \end{aligned} \tag{19}$$

From this, we can see $X_\psi(\hat{\omega}) \subseteq X'_\psi(\hat{\omega})$. *In this sense,* (17) *provides a stronger characterization than* (18)*, when the joint minimization over* $(\boldsymbol{x}, \omega)$ *is split out in the order of* ω *and* $\boldsymbol{x}$*. In fact, the proof in [26] that established* (17) *is quite involved. The difficulty exactly lies in characterizing the properties of the optimal objective value of the inner minimization in* (17) *as a single variable function in* ω*. In comparison, our primal characterization* (16) *bypasses this difficulty by only looking at the joint minimization problem. It seems that this insight to view the ALD as a traditional LD problem in a lifted space is new, which makes the derivation of* (16) *quite simpler. Our primal characterization also requires less assumptions than* (17)*. In particular,* (17) *requires that the augmenting function is convex in a particular form and at least one of the three assumptions stated in Proposition 2.12 hold, whereas* our primal characterization works for both convex and non-convex augmenting functions*, and the relative interior condition in Theorem 2.18 is a rather mild regularity condition. In addition, as we will show now, Assumptions 2.1 and 2.5 are enough to prove the zero duality gap result for ALD of general MIPs. A similar result is also proved in [26, Proposition 3] through their characterization* (17)*, again under more restricted conditions.*

2.4.2 Zero Duality Gap for MIPs

From the primal characterization (16) we can see the $z_\rho^{\mathrm{LD}+}$ is a non-decreasing function of ρ. Since $z_\rho^{\mathrm{LD}+}$ is upper bounded by z^{IP}, therefore we have

$$-\infty < z^{\mathrm{LD}^*} := \sup_{\rho>0} z_\rho^{\mathrm{LD}+} = \lim_{\rho\to+\infty} z_\rho^{\mathrm{LD}+} \le z^{\mathrm{IP}}.$$

We want to show that in fact $z^{\mathrm{LD}^*} = z^{\mathrm{IP}}$. Recall that $\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}$ is defined as a rational optimal vector of dual variables for $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ in the LP relaxation of problem (1).

Proposition 2.19. *Suppose Assumptions 2.1 and 2.5 hold. For given $\rho > 0$ and $\epsilon > 0$, define $\omega^*_{\rho,\epsilon}$ as*

$$\begin{aligned} \omega^*_{\rho,\epsilon} := \inf_{\boldsymbol{x},\omega}\ & \omega \\ s.t.\ & \boldsymbol{x} \in X, \\ & \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \le \omega, \\ & \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{LP}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho\omega - z_\rho^{LR+}(\bar{\boldsymbol{\lambda}}_{LP}) \le \epsilon. \end{aligned} \tag{20}$$

Then, the limit $\omega^*_\rho := \lim_{\epsilon\downarrow 0} \omega^*_{\rho,\epsilon}$ exists and $\lim_{\rho\to+\infty} \omega^*_\rho = 0$.

Proof. From definition and finiteness of $z_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}})$, and feasibility of problem (1), we know that problem (20) is feasible. For all $\rho > 0$ and $\epsilon > 0$, nonnegativity of ψ implies $\omega^*_{\rho,\epsilon} \ge 0$. Moreover, the first and third constraints in (20) imply

$$\begin{aligned} \omega^*_{\rho,\epsilon} &\le \frac{1}{\rho}(z_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) + \epsilon - \boldsymbol{c}^\top \boldsymbol{x} - \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})), \text{ for some } \boldsymbol{x} \in X \\ &\le \frac{1}{\rho}(z^{\mathrm{IP}} + \epsilon - z^{\mathrm{LP}}), \end{aligned} \tag{21}$$

where the second inequality follows from the facts that $z_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) \le z^{\mathrm{IP}}$ and $z^{\mathrm{LP}} \le \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$, for all $\boldsymbol{x} \in X$. By taking limits $\epsilon \downarrow 0$ on both sides of (21) we have

$$0 \le \omega^*_\rho = \lim_{\epsilon\downarrow 0} \omega^*_{\rho,\epsilon} \le \lim_{\epsilon\downarrow 0} \frac{1}{\rho}(z^{\mathrm{IP}} + \epsilon - z^{\mathrm{LP}}) = \frac{1}{\rho}(z^{\mathrm{IP}} - z^{\mathrm{LP}}) \tag{22}$$

Note that $\omega^*_{\rho,\epsilon}$ is non-decreasing as $\epsilon \downarrow 0$. Moreover, $\omega^*_{\rho,\epsilon}$ is upper bounded. Then, $\lim_{\epsilon\downarrow 0} \omega^*_{\rho,\epsilon}$ exists. By taking limits $\rho \to +\infty$ on both sides of (22) we have $\lim_{\rho\to+\infty} \omega^*_\rho = 0$.

□

Lemma 2.20. *Consider ω_ρ^* as described in Proposition 2.19. Let us define $\tilde{z}_\rho^{LR+}(\bar{\boldsymbol{\lambda}}_{LP})$ as follows:*

$$\begin{aligned} \tilde{z}_\rho^{LR+}(\bar{\boldsymbol{\lambda}}_{LP}) := \inf_{\boldsymbol{x},\omega} \ & \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{LP}^\top(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho\omega\} \\ s.t. \ & \boldsymbol{x} \in X, \\ & \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \le \omega, \\ & (1-\delta)\omega_\rho^* \le \omega \le (1+\delta)\omega_\rho^*. \end{aligned} \tag{23}$$

Then,

$$\begin{aligned} z_\rho^{LR+}(\bar{\boldsymbol{\lambda}}_{LP}) =& \tilde{z}_\rho^{LR+}(\bar{\boldsymbol{\lambda}}_{LP}) \\ \ge & \inf_{\boldsymbol{x}} \ \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{LP}^\top(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho(1-\delta)\omega_\rho^*\} \\ & s.t. \ \boldsymbol{x} \in X, \\ & \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \le (1+\delta)\omega_\rho^*, \\ \ge & \inf_{\boldsymbol{x}} \ \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{LP}^\top(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})\} \\ & s.t. \ \boldsymbol{x} \in X, \\ & \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \le (1+\delta)\omega_\rho^*, \end{aligned} \tag{24}$$

for any $0 < \delta < 1$.

Proof. Clearly, $z_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) \le \tilde{z}_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}})$, due to the last constraint in (23). Let $\alpha_\rho := \tilde{z}_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) - z_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}})$. Assume by contradiction, $z_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) < \tilde{z}_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}})$ or equivalently $\alpha_\rho > 0$. Then, for all $(\boldsymbol{x}, \omega)$ satisfying constraints of (23) it holds

$$\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho\omega \ge \tilde{z}_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) = z_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) + \alpha_\rho,$$

which implies $(\boldsymbol{x}, \omega)$ is infeasible for problem (20) if $0 < \epsilon < \alpha_\rho$. Therefore, $\omega_{\rho,\epsilon}^* \notin ((1-\delta)\omega_\rho^*, (1+\delta)\omega_\rho^*)$ for $0 < \epsilon < \alpha_\rho$, which contradicts with $\omega_\rho^* = \lim_{\epsilon \downarrow 0} \omega_{\rho,\epsilon}^*$. Therefore, $z_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) = \tilde{z}_\rho^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}})$. Inequalities in (24) hold, because $\rho\omega \ge \rho(1-\delta)\omega_\rho^* \ge 0$ and $\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \le (1+\delta)\omega_\rho^*$, for all $(\boldsymbol{x}, \omega)$ satisfying constraints of (23). □

Theorem 2.21. *Suppose Assumptions 2.1 and 2.5 hold. Then,* $\sup_{\rho>0} z_\rho^{LD+} = z^{IP}$.

Proof. Following (8), it is enough to show that $\sup_{\rho>0} z_\rho^{\mathrm{LD}+} \geq z^{\mathrm{IP}}$. Let δ be a given positive scalar in $(0,1)$. By definition of ALD, we have

$$\begin{aligned}
z_\rho^{\mathrm{LD}+} &= \sup_{\boldsymbol{\lambda}\in\mathbb{R}^m} \inf_{\boldsymbol{x},\omega} \{\boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho\omega : \boldsymbol{x}\in X,\ \psi(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) \leq \omega\} \\
&\geq \inf_{\boldsymbol{x},\omega} \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho\omega : \boldsymbol{x}\in X,\ \psi(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) \leq \omega\} \\
&\geq \inf_{\boldsymbol{x}} \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) : \boldsymbol{x}\in X,\ \psi(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) \leq (1+\delta)\omega_\rho^*\} \qquad (25a) \\
&\geq \inf_{\boldsymbol{x}} \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) : \boldsymbol{x}\in X,\ \|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_\infty \leq \kappa_\rho\} \qquad (25b) \\
&= \min_{\boldsymbol{x}} \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) : \boldsymbol{x}\in X,\ \|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_\infty \leq \kappa_\rho\} \qquad (25c)
\end{aligned}$$

where

$$\kappa_\rho := \mathbf{diam}\{\boldsymbol{u} \mid \psi(\boldsymbol{u}) \leq (1+\delta)\omega_\rho^*\} = \sup\{\|\boldsymbol{u}\|_\infty \mid \psi(\boldsymbol{u}) \leq (1+\delta)\omega_\rho^*\} < \infty.$$

Inequality (25a) holds by Lemma 2.20, and (25b) follows from level boundedness of ψ. Equality (25c) is valid by Assumption 2.1. By taking limits on both sides of (25b) we have

$$\begin{aligned}
\lim_{\rho\to+\infty} z_\rho^{\mathrm{LD}+} &\geq \lim_{\rho\to+\infty} \min_{\boldsymbol{x}} \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) : \boldsymbol{x}\in X, \|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_\infty \leq \kappa_\rho\} \\
&\geq \min_{\boldsymbol{x}} \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) : \boldsymbol{x}\in X, \|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_\infty \leq \lim_{\rho\to+\infty} \kappa_\rho\} \qquad (26a) \\
&= \min_{\boldsymbol{x}} \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) : \boldsymbol{x}\in X, \|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_\infty \leq 0\} \qquad (26b) \\
&= \min_{\boldsymbol{x}} \{\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) : \boldsymbol{x}\in X, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\} \qquad (26c) \\
&= \min_{\boldsymbol{x}} \{\boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x}\in X, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\} \qquad (26d) \\
&= z^{\mathrm{IP}}. \qquad (26e)
\end{aligned}$$

where (26a) follows from lower semicontinuity of value functions for MIPs with rational data [117]. Equality (26b) holds by Assumption 2.5, i.e. $\lim_{\rho\to+\infty} \kappa_\rho = 0$. This completes the proof. □

2.5 Exact ALD for MIPs

2.5.1 Pure IP case

A special case of problem (1) is the *pure IP case*, where all variables are integral, i.e. $n_2 = 0$. Zero duality gap and exact penalty representation using proximal Lagrangian for pure IPs were established in [24, Theorem 1.5]. Boland and Eberhard [26, Corollary 1] proved exact penalty representation for ALD of pure IPs with a bounded feasible region, i.e. X is finite, and the augmenting functions satisfying assumptions in Proposition 2.12. In this section, we extend this recent result to show exact penalty representation for pure IPs under weaker assumptions on the augmenting functions (e.g., the augmenting function does not have to be convex) and X may not be necessarily finite.

Theorem 2.22. *Suppose problem* (1) *is a pure IP with potentially infinitely many feasible solutions, and Assumptions 2.1 holds. If*

$$\inf\{\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) : \boldsymbol{x} \in X, \boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{b}\} \geq \delta > 0 \tag{27}$$

for some strictly positive value of δ, then there exists a finite $\rho^* \in (0, +\infty)$ such that $z^{LD+}_{\rho^*} = z^{IP}$.

Proof. Following (8), it suffices to find a finite ρ^* such that $z^{\mathrm{LD}+}_{\rho^*} \geq z^{\mathrm{IP}}$. Let $\bar{\rho} > 0$ be any positive penalty coefficient. By assumption, there exists a $\delta > 0$ which satisfies (27). Furthermore, let $\boldsymbol{x}^0$ be any arbitrary feasible solution of (1), i.e. $\boldsymbol{x}^0 \in X$ and $\boldsymbol{A}\boldsymbol{x}^0 = \boldsymbol{b}$. Set $\rho^* = \frac{\boldsymbol{c}^\top \boldsymbol{x}^0 - z^{\mathrm{LP}}}{\delta}$. Note that $0 < \rho^* < +\infty$, because

$\delta > 0$ and $-\infty < z^{\mathrm{LP}} \leq \boldsymbol{c}^\top \boldsymbol{x}^0 < +\infty$. We claim that $z^{\mathrm{LD}+}_{\rho^*} \geq z^{\mathrm{IP}}$. Observe that we have

$$z^{\mathrm{LD}+}_{\rho^*} = \sup_{\boldsymbol{\lambda}} z^{\mathrm{LD}+}_{\rho^*}(\boldsymbol{\lambda}) \geq z^{\mathrm{LD}+}_{\rho^*}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) = \inf_{\boldsymbol{x} \in X} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top_{\mathrm{LP}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho^* \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \right\}. \tag{28}$$

There are two cases.

1. For all $\boldsymbol{x} \in X$ with $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$,

$$\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho^* \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{c}^\top \boldsymbol{x} \geq z^{\mathrm{IP}}. \tag{29}$$

2. For all $\boldsymbol{x} \in X$ with $\boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{b}$,

$$\begin{aligned}
&\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho^* \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \\
&\quad = \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \left(\frac{\boldsymbol{c}^\top \boldsymbol{x}^0 - z^{\mathrm{LP}}}{\delta} \right) \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \\
&\quad \geq \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \left(\boldsymbol{c}^\top \boldsymbol{x}^0 - z^{\mathrm{LP}} \right) && \text{(30a)} \\
&\quad \geq z^{\mathrm{LP}} + \left(\boldsymbol{c}^\top \boldsymbol{x}^0 - z^{\mathrm{LP}} \right) && \text{(30b)} \\
&\quad = \boldsymbol{c}^\top \boldsymbol{x}^0 \\
&\quad \geq z^{\mathrm{IP}},
\end{aligned}$$

where (30a) holds because $-\infty < z^{\mathrm{LP}} \leq z^{\mathrm{IP}} \leq \boldsymbol{c}^\top \boldsymbol{x}^0$ and $\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \geq \delta > 0$ for all $\boldsymbol{x} \in X$ with $\boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{b}$ by (27). Inequality (30b) follows by definition of $\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}$.

Inequalities (29) and (30) imply

$$\inf_{\boldsymbol{x} \in X} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho^* \psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \right\} \geq z^{\mathrm{IP}}.$$

Together with (28), we have

$$z_{\rho^*}^{\mathrm{LD}+} \geq z^{\mathrm{IP}}.$$

This completes the proof. □

Note that for the pure IP case of problem (1) under Assumption 2.1, any augmenting function defined in Proposition 2.12 satisfies (27). Even the index function $\mathbb{I} : \mathbb{R}^m \to \{0, 1\}$ where

$$\mathbb{I}(\boldsymbol{u}) = \begin{cases} 0 & \text{if } \boldsymbol{u} = \boldsymbol{0}, \\ 1 & \text{otherwise} \end{cases}$$

can be used as an augmenting function $\psi(\cdot)$ to satisfy (27).

2.5.2 MIP case

For a general MIP with both continuous and integer variables, we need more conditions on the augmenting function to have an exact penalty representation. For example, if $\psi(\cdot) = \|\cdot\|_2^2$, i.e. the proximal Lagrangian case, this augmenting function satisfies the assumptions in Proposition 2.12 as well as (27) when X is a pure integer set. However, for a general MIP, there may not exist a finite $0 < \rho^* < \infty$ such that $z_{\rho^*}^{\mathrm{LD}+} = z^{\mathrm{IP}}$ under this augmenting function. In this section, we first give an example to show that proximal Lagrangian fails to have an exact penalty representation for a simple MIP in three variables. Then we prove that, when the augmenting function is any norm (but not the squared norm) i.e., for the sharp Lagrangians, the ALD always has an exact penalty representation for general MIPs. Finally, we extend this result to some classes of augmenting functions that are not convex.

2.5.2.1 Counterexample MIP for Proximal Lagrangian

Proposition 2.23. *There exists an MIP problem of the form* (1) *and an augmenting function satisfying assumptions in Proposition 2.12 such that* $z_\rho^{LD+} < z^{IP}$ *for all finite* $\rho > 0$.

Next, we verify this proposition with a simple example.

Example 1. *Consider the following MIP problem, with one binary and two continuous variables.*

$$\begin{aligned} z^{IP} = \min_{x_1,x_2,x_3} \quad & -x_1 - x_2 \\ s.t. \quad & -x_1 + x_2 = 0 \\ & 0 \le x_1 \le x_3 \\ & 0 \le x_2 \le 1 - x_3 \\ & x_3 \in \{0,1\} \end{aligned} \tag{31}$$

The only feasible points for (31) are $(x_1, x_2, x_3) = (0, 0, 0)$ and $(x_1, x_2, x_3) = (0, 0, 1)$ with objective value 0. Then, $z^{\text{IP}} = 0$. Projection of the feasible region of (31) without the constraint $-x_1 + x_2 = 0$ into the space of x_1 and x_2 contains the blue lines in Figure 1. The points satisfying $-x_1 + x_2 = 0$ are depicted by a red line in this space.

We show that in Example 1, for ALD with $\psi(\cdot) = \|\cdot\|_2^2$, $z_\rho^{\text{LD}+} < 0$ for all $\rho > 0$. From Theorem 2.18, ALD (15) with $\psi(\cdot) = \|\cdot\|_2^2$ becomes

$$\begin{aligned} z_\rho^{\text{LD}+} = \inf \quad & -x_1 - x_2 + \rho\omega \\ \text{s.t.} \quad & (x_1, x_2, x_3, \omega) \in \mathbf{conv}(S_2) \\ & -x_1 + x_2 = 0. \end{aligned} \tag{32}$$

where,

$$S_2 := \left\{ (x_1, x_2, x_3, \omega) \in \mathbb{R}^4 : \begin{array}{c} \omega \geq (-x_1 + x_2)^2 \\ 0 \leq x_1 \leq x_3 \\ 0 \leq x_2 \leq 1 - x_3 \\ x_3 \in \{0, 1\} \end{array} \right\}.$$

Consider $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\omega}) = (0, r, 0, r^2)$ and $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{\omega}) = (r, 0, 1, r^2)$ where $r(\rho) = \min\{1, \frac{1}{2\rho}\}$. Obviously, both $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\omega})$ and $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{\omega})$ belong to S_2. Then, $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\omega}) := \frac{1}{2}(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\omega}) + \frac{1}{2}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{\omega}) = (\frac{r}{2}, \frac{r}{2}, \frac{1}{2}, r^2)$ belongs to $\text{Conv}(S_2)$.

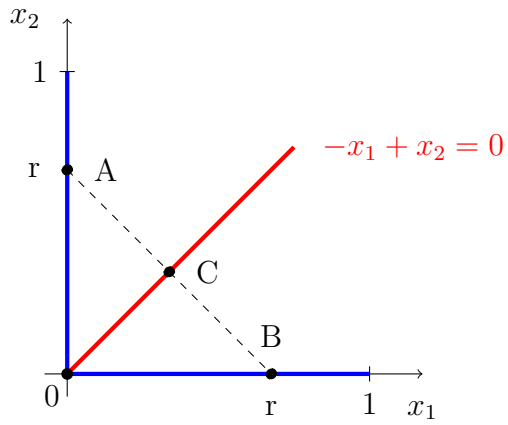

Figure 1: Projection of the feasible region of Example 1 in the space of x_1 and x_2.

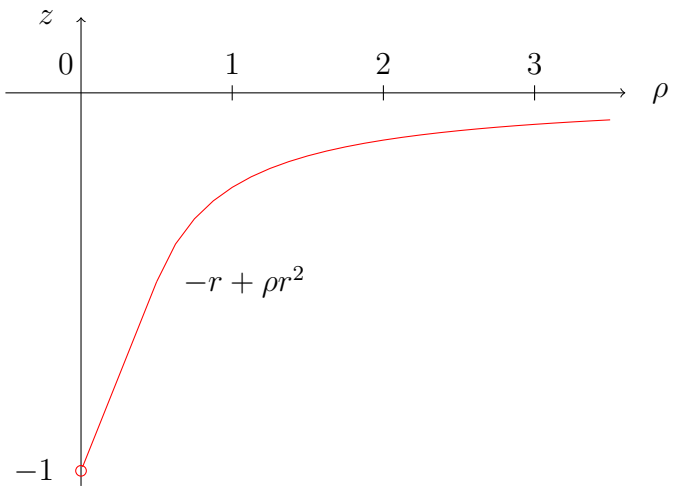

Figure 2: Value of $-\bar{x}_1 - \bar{x}_2 + \rho\bar{\omega} = -r + \rho r^2$ versus ρ for $\rho > 0$

Projection of the points $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\omega})$, $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{\omega})$ and $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\omega})$ in the space of x_1 and x_2 can be depicted as points A, B and C, respectively, in Figure 1. Because $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\omega}) \in \mathbf{conv}(S_2)$ and $-\bar{x}_1 + \bar{x}_2 = 0$, the point $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\omega})$ is a feasible solution of (32). Therefore,

$$z_\rho^{\mathrm{LD}+} \leq -\bar{x}_1 - \bar{x}_2 + \rho\bar{\omega} = -r + \rho r^2 \leq \max\left\{-\frac{1}{2}, -\frac{1}{4\rho}\right\} < 0 = z^{\mathrm{IP}}, \quad \forall \rho > 0 \qquad (33)$$

which shows $z_\rho^{\mathrm{LD}+} < z^{\mathrm{IP}}$ for all $\rho > 0$, i.e. there is no finite ρ^* such that $z_{\rho^*}^{\mathrm{LD}+} = z^{\mathrm{IP}}$. Note that the second inequality in (33) follows from the fact that

$$-r + \rho r^2 = \begin{cases} -1 + \rho \times 1^2 = -1 + \rho \leq -\frac{1}{2}, & \text{if } 0 < \rho < \frac{1}{2} \\ -\frac{1}{2\rho} + \rho \times \left(\frac{1}{2\rho}\right)^2 = -\frac{1}{4\rho}, & \text{if } \frac{1}{2} \leq \rho. \end{cases}$$

An alternative way to show this result is as follows. From relations (6) and (7), for any $\lambda \in \mathbb{R}$,

$$\begin{aligned} z_\rho^{\mathrm{LR}+}(\lambda) = \inf_{x_1,x_2,x_3} & \{-x_1 - x_2 + \lambda(-x_1 + x_2) + \rho(-x_1 + x_2)^2\} \\ \text{s.t.} \quad & 0 \leq x_1 \leq x_3 \\ & 0 \leq x_2 \leq 1 - x_3 \\ & x_3 \in \{0, 1\} \\ = \min & \left\{ \min_{0 \leq x_1 \leq 1} \{-(\lambda + 1)x_1 + \rho x_1^2\}, \min_{0 \leq x_2 \leq 1} \{(\lambda - 1)x_2 + \rho x_2^2\}, \right\} \end{aligned}$$

where the second equality follows from enumeration of the values of x_3. Note that

$$\min_{0\le x_1\le 1}\{-(\lambda+1)x_1+\rho x_1^2\}=\begin{cases}0, & \lambda\le -1\\ \rho-\lambda-1, & \lambda\ge -1+2\rho\\ -\frac{(\lambda+1)^2}{4\rho}, & -1<\lambda<-1+2\rho,\end{cases}$$

and

$$\min_{0\le x_2\le 1}\{(\lambda-1)x_2+\rho x_2^2\}=\begin{cases}0, & \lambda\ge 1\\ \rho+\lambda-1, & \lambda\le 1-2\rho\\ -\frac{(\lambda-1)^2}{4\rho}, & 1-2\rho<\lambda<1.\end{cases}$$

Then, by considering all possible values of λ and $\rho>0$, we have:

$$\begin{aligned} z_\rho^{\mathrm{LR}+}(\lambda)=&\begin{cases}\rho+\lambda-1\le-\rho, & \text{if } \lambda\le -1,\lambda\le 1-2\rho\\ -\frac{(\lambda-1)^2}{4\rho}\le-\frac{1}{\rho}, & \text{else if } \lambda\le -1,\lambda>1-2\rho\\ \le\rho-\lambda-1<-\rho, & \text{else if } \lambda\ge -1+2\rho\\ -\frac{(\lambda+1)^2}{4\rho}\le-\frac{1}{\rho}, & \text{else if } -1<\lambda<-1+2\rho,\ \lambda\ge 1\\ \le\rho+\lambda-1<-\rho, & \text{else if } -1<\lambda<-1+2\rho,\ \lambda\le 1-2\rho\\ \min\{-\frac{(\lambda+1)^2}{4\rho},-\frac{(\lambda-1)^2}{4\rho}\}\le-\frac{1}{4\rho}, & \text{Otherwise}\end{cases}\\ \le&\max\{-\rho,-\frac{1}{4\rho}\}\end{aligned}$$

Therefore,

$$z_\rho^{\mathrm{LD}+}=\sup_{\lambda\in\mathbb{R}^m}\ z_\rho^{\mathrm{LR}+}(\lambda)\le\max\{-\rho,-\frac{1}{4\rho}\}<0,$$

which implies $z_\rho^{\mathrm{LD}+}<0$ for all $\rho>0$.

Another point of view for this example is exploring its value function with respect

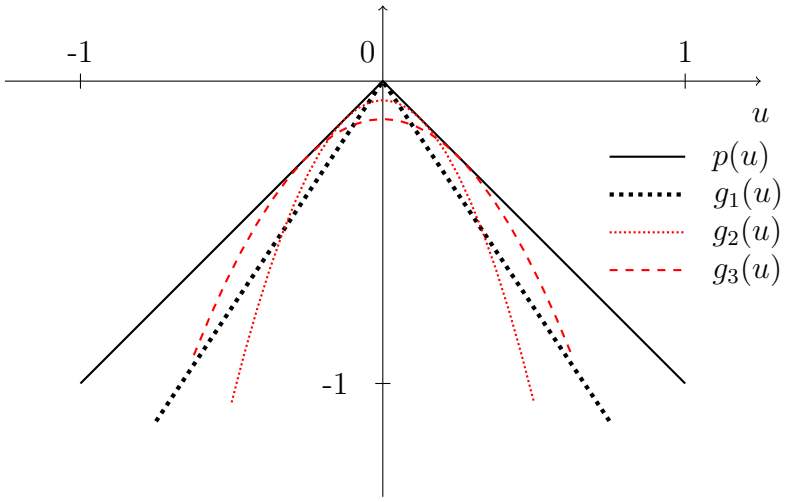

Figure 3: Value function and some augmenting functions for Example 1

to the first constraint, which is

$$\begin{aligned} p(u) = \inf_{x_1,x_2,x_3} \quad & -x_1 - x_2 \\ \text{s.t.} \quad & -x_1 + x_2 = u \\ & 0 \le x_1 \le x_3 \\ & 0 \le x_2 \le 1 - x_3 \\ & x_3 \in \{0,1\}. \end{aligned} \tag{34}$$

It is easy to check that $p(u) = -|u|$, for any $u \in [-1,1]$. Note that the problem (34) is infeasible if $u \notin [-1,1]$. Consider the supporting functions $g_1(u)$, $g_2(u)$ and $g_3(u)$ which are based on augmenting functions $\psi_1(u) = 1.5|u|$, $\psi_2(u) = u^2$ and $\psi_3(u) = 0.5u^2$, respectively. The value function and these supporting functions are depicted in Figure 3. From this figure, it is clear that $\psi_2(u) = u^2$ can result in a strictly smaller duality gap comparing to $\psi_3(u) = 0.5u^2$. But, no quadratic supporting function to $p(\cdot)$ can reach $p(0) = 0$. However, the sharp augmenting function $\psi_1(u) = 1.5|u|$ closes the duality gap in this example.

Example 1 showed that, for $\psi(\cdot) = \|\cdot\|_2^2$, there may exist MIP problems such that $z_\rho^{\text{LD}+} < z^{\text{IP}}$, for any finite value of ρ.

2.5.2.2 Exact ALD with the sharp Lagrangian for MIPs

Next, we show that using any norm as an augmenting function with a sufficiently large penalty coefficient closes the duality gap for the general MIPs. First, we provide a self contained proof for this result. Another approach is to verify the calmness condition from Subsection 2.3.3 for the value function, which we will provide as an alternative proof.

Theorem 2.24. *Consider problem* (1) *with both integer and continuous variables. Suppose Assumption 2.1 holds, and* $\psi(\cdot) = \|\cdot\|$*, where* $\|\cdot\|$ *is any norm. Then there exists a finite* $0 < \rho^* < +\infty$ *such that* $z^{LD+}_{\rho^*} = z^{IP}$*.*

Proof. First, let us show the result for $\psi(\cdot) = \|\cdot\|_\infty$. Then, we extend it to any norm by the equivalence of norms in a Euclidean space. Let $\psi(\cdot) = \|\cdot\|_\infty$ and $\mathbf{1}_m$ be an m dimensional vector with all entries equal to 1. Then S_ψ is a polyhedron,

$$\begin{aligned} S_\psi &= \left\{(\boldsymbol{x}, \omega) \in \mathbb{R}^{n+1} : \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_\infty \leq \omega,\ \boldsymbol{x} \in X\right\} \\ &= \left\{(\boldsymbol{x}, \omega) \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2+1} : -\mathbf{1}_m\omega \leq \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \leq \mathbf{1}_m\omega,\ \boldsymbol{E}\boldsymbol{x} \leq \boldsymbol{f}\right\}. \end{aligned} \tag{35}$$

Then, by Assumption 2.1 and due to Meyer's theorem [116], there is a rational polyhedral representation for the set $\mathbf{conv}(S_{\|.\|_\infty}) \cap \{(\boldsymbol{x}, \omega) \in \mathbb{R}^{n+1} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$. Denote this representation by $\boldsymbol{H}\left[\begin{smallmatrix}\boldsymbol{x}\\ \omega\end{smallmatrix}\right] \geq \boldsymbol{h}$, where $\boldsymbol{H} \in \mathbb{Q}^{\hat{m}\times(n+1)}$ and $\boldsymbol{h} \in \mathbb{Q}^{\hat{m}}$, for some finite integer $\hat{m}$. Then, by the primal characterization of ALD in Theorem 2.18, the ALD problem (7) for a given $\rho > 0$ can be written as follows,

$$\begin{aligned} z^{\mathrm{LD+}}_\rho = \inf_{\boldsymbol{x},\omega}\ & \boldsymbol{c}^\top\boldsymbol{x} + \rho\omega \\ \text{s.t. } & \boldsymbol{H}\begin{bmatrix}\boldsymbol{x}\\ \omega\end{bmatrix} \geq \boldsymbol{h}. \end{aligned} \tag{36}$$

Note that, for a given $\rho > 0$, problem (36) is an LP and its dual can be written as

follows.

$$
\begin{aligned}
z_\rho^{\mathrm{DLD}+} := \sup_{\boldsymbol{y}} \ & \boldsymbol{h}^\top \boldsymbol{y} \\
\text{s.t. } & \boldsymbol{H}^\top \boldsymbol{y} = \begin{bmatrix} \boldsymbol{c} \\ \rho \end{bmatrix} \\
& \boldsymbol{y} \geq \boldsymbol{0}.
\end{aligned} \tag{37}
$$

Note that $z_\rho^{\mathrm{LD}+} = z_\rho^{\mathrm{DLD}+}$, since $z_\rho^{\mathrm{LD}+} > -\infty$ and by strong duality for LPs. We are interested in a finite positive ρ^* such that $z_{\rho^*}^{\mathrm{LD}+} \geq \boldsymbol{c}^\top \boldsymbol{x}^*$, where $\boldsymbol{x}^*$ is an optimal solution of (1). The existence of such a ρ^* is equivalent to the existence of $(\boldsymbol{y}^*, \xi^*, \rho^*)$ with $\xi^* = 0$ for the following feasibility problem in $(\boldsymbol{y}, \xi, \rho)$,

$$
\begin{aligned}
\boldsymbol{h}^\top \boldsymbol{y} + \xi &\geq \boldsymbol{c}^\top \boldsymbol{x}^* \\
\boldsymbol{H}^\top \boldsymbol{y} &= \begin{bmatrix} \boldsymbol{c} \\ \rho \end{bmatrix} \\
\boldsymbol{y} &\geq \boldsymbol{0} \\
\rho &\geq 1 \\
\xi &\geq 0.
\end{aligned} \tag{38}
$$

Let Ξ be the projection of the feasible set of (38) into the ξ space. Note that by Fourier-Motzkin Elimination, Ξ is itself a polyhedron. Then, Ξ is a closed set.

Consider a sequence $\xi_k \downarrow 0$ as $k \to +\infty$. Since $z_\rho^{\mathrm{DLD}+} = z_\rho^{\mathrm{LD}+} \leq z^{\mathrm{IP}}$ for any $\rho \geq 0$, and $\psi(\cdot) = \|\cdot\|_\infty$ satisfies Assumption 2.5, by Theorem 2.12, $z_\rho^{\mathrm{LD}+} \uparrow z^{\mathrm{IP}} = \boldsymbol{c}^\top \boldsymbol{x}^*$ as $\rho \to \infty$. By closedness of Ξ, $0 \in \Xi$ because $\xi^* = 0$ is a cluster point of Ξ. That is, there exists some $\boldsymbol{y}^*$ and ρ^* such that $(\boldsymbol{y}^*, 0, \rho^*)$ is a feasible solution of (38). Therefore, $z_{\rho^*}^{\mathrm{LD}+} \geq z^{\mathrm{IP}}$, which along with $z_{\rho^*}^{\mathrm{LD}+}$ being a lower bound for z^{IP}, we

conclude $z^{\mathrm{LD}+}_{\rho^*} = z^{\mathrm{IP}} = \boldsymbol{c}^\top \boldsymbol{x}^*$. Note that

$$\begin{aligned} z^{\mathrm{IP}} &= \sup_{\boldsymbol{\lambda}\in\mathbb{R}^m} \inf_{\boldsymbol{x}\in X} \quad \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho^*\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_\infty \\ &= \sup_{\boldsymbol{\lambda}\in\mathbb{R}^m} \inf_{(\boldsymbol{x},\omega)\in S_{\|\cdot\|_\infty}} \quad \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho^*\omega \\ &= \sup_{\boldsymbol{\lambda}\in\mathbb{R}^m} \inf_{(\boldsymbol{x},\omega)\in \mathbf{conv}(S_{\|\cdot\|_\infty})} \quad \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho^*\omega \\ &= \inf_{(\boldsymbol{x},\omega)\in \mathbf{conv}(S_{\|\cdot\|_\infty})} \quad \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho^*\omega \\ &= \inf_{(\boldsymbol{x},\omega)\in S_{\|\cdot\|_\infty}} \quad \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho^*\omega \\ &= \inf_{\boldsymbol{x}\in X} \quad \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho^*\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_\infty \end{aligned} \tag{39}$$

for some $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^m$, where the second equality follows from definition of $S_{\|\cdot\|_\infty}$ in (35). The third and fifth equations hold because minimizing a linear objective function on a set is equivalent to minimizing it on the convex hull of that set. The fourth equality is valid by strong duality for LPs, because under Assumption 2.1 and due to Meyer's theorem [116], $\mathbf{conv}(S_{\|\cdot\|_\infty})$ is a rational polyhedron. The last equality follow from definition of $S_{\|\cdot\|_\infty}$ in (35). Then,

$$\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho^*\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_\infty \geq z^{\mathrm{IP}},\ \forall \boldsymbol{x}\in X$$

Recall that for any norm $\|\cdot\|$ in finite dimensions there exists $0<\gamma<1$ such that $\frac{1}{\gamma}\|\boldsymbol{u}\| \geq \|\boldsymbol{u}\|_\infty \geq \gamma\|\boldsymbol{u}\|$, by the equivalence of norms. Take $\hat{\rho} = \frac{\rho^*}{\gamma}$. Then,

$$\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \hat{\rho}\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\| \geq \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \rho^*\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_\infty,\ \forall \boldsymbol{x}\in X$$

which implies

$$z^{\mathrm{LD}+}_{\hat{\rho}} \geq z^{\mathrm{LR}+}_{\hat{\rho}}(\bar{\boldsymbol{\lambda}}) = \inf_{\boldsymbol{x}\in X} \quad \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) + \hat{\rho}\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\| \geq z^{\mathrm{IP}} \tag{40}$$

On the other hand, $z^{\mathrm{LD}+}_{\hat{\rho}} \leq z^{\mathrm{IP}}$ by (8). Therefore, $z^{\mathrm{LD}+}_{\hat{\rho}} = z^{\mathrm{IP}}$. □

Remark 4. *Note that $\hat{\rho}$ and $\bar{\boldsymbol{\lambda}}$ in the proof of Theorem 2.24 satisfy the assumptions in Proposition 2.7. Therefore, any optimal solution of ALR* (6) *with $\boldsymbol{\lambda} = \bar{\boldsymbol{\lambda}}$ and $\rho > \hat{\rho}$ solves the MIP problem* (1)*.*

Another approach to prove Theorem 2.24 is to show that for sharp Lagrangian there exists some $\rho \in (0, \infty)$ such that $z_\rho^{\mathrm{LR}+}(\mathbf{0}) > -\infty$, and the value function of MIP problem (1) is calm from below at $\mathbf{0}$ (see Subsection 2.3.3). In Lemmas 2.25 and 2.26 we verify these conditions.

Lemma 2.25. *For a MIP problem* (1) *satisfying Assumption 2.1, consider ALR with* $\psi(\cdot) = \|\cdot\|$ *where* $\|\cdot\|$ *is any norm. Then, there exists some* $\rho \in (0, \infty)$ *such that* $z_\rho^{LR+}(\mathbf{0}) > -\infty$.

Proof. Let $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^m$ be the vector of optimal dual values for $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ in the LP relaxation of MIP (1). Then,

$$-\infty < z^{\mathrm{LP}} \leq \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \leq \boldsymbol{c}^\top \boldsymbol{x} + \|\bar{\boldsymbol{\lambda}}\|_2 \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2, \quad \forall \boldsymbol{x} \in X \tag{41}$$

Therefore, for $\psi(\cdot) = \|\cdot\|_2$ and $\rho = \|\bar{\boldsymbol{\lambda}}\|_2$, $z_\rho^{\mathrm{LR}+}(\mathbf{0}) \geq z^{\mathrm{LP}} > -\infty$ holds. This result can be generalized to other norms as augmenting functions by the equivalence of norms. □

Lemma 2.26. *The value function of MIP problem* (1) *satisfying Assumption 2.1 is calm from below at* $\mathbf{0}$.

Proof. By Assumption 2.1, we know that $-\infty < p(\mathbf{0}) = z^{\mathrm{IP}} < \infty$. Consider a small enough open bounded neighborhood $W \subset \mathbb{R}^n$ of $\mathbf{0}$. For an MIP problem with rational data, the value function is lower semicontinuous [117] and piecewise polyhedral with finitely many regions in any bounded set [23]. Then, there are finitely many, disjoint regions $W_1, \cdots, W_r \subset W$ with some $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_r \in \mathbb{R}^m$ and $\beta_1, \cdots, \beta_r \in \mathbb{R}$ such that $W = W_1 \cup \cdots \cup W_r$ and $p(\boldsymbol{u}) = \boldsymbol{\alpha}_k^\top \boldsymbol{u} + \beta_k$ if $\boldsymbol{u} \in W_k$ and $\boldsymbol{u}$ is feasible. Without loss of generality, assume $\mathbf{0} \in \mathbf{cl}(W_k)$ for all $k = 1, \cdots, r$; otherwise W can be chosen small enough to exclude W_k.

By lower semicontinuity of the value function $p(\boldsymbol{u})$ at $\boldsymbol{u} = \mathbf{0}$ and the fact that $\mathbf{0} \in \mathbf{cl}(W_k)$ for all $k = 1, \cdots, r$, we have $\boldsymbol{\alpha}_k^\top \mathbf{0} + \beta_k = \beta_k \geq p(\mathbf{0})$ for all $k = 1, \cdots, r$.

Let $\kappa = \max\{\|\boldsymbol{\alpha}_1\|_2, \cdots, \|\boldsymbol{\alpha}_r\|_2\}$. For each $k = 1, \cdots, r$, if $\boldsymbol{u} \in W_k$, then

$$p(\boldsymbol{u}) - p(\boldsymbol{0}) = \boldsymbol{\alpha}_k^\top \boldsymbol{u} + \beta_k - p(\boldsymbol{0}) \geq \boldsymbol{\alpha}_k^\top \boldsymbol{u} \geq -\|\boldsymbol{\alpha}_k\|_2 \times \|\boldsymbol{u}\|_2 \geq -\kappa \|\boldsymbol{u}\|_2 = -\kappa \|\boldsymbol{u} - \boldsymbol{0}\|_2.$$

Finally, we have

$$p(\boldsymbol{u}) - p(\boldsymbol{0}) \geq -\kappa \|\boldsymbol{u} - \boldsymbol{0}\|_2, \ \forall \boldsymbol{u} \in W,$$

which, from Definition 2.11, concludes that $p(\boldsymbol{u})$ is calm from below at $\boldsymbol{u} = \boldsymbol{0}$. □

Next, we show that the value of $\bar{\boldsymbol{\lambda}}$ in the proof of Theorem 2.24 really does not matter.

Proposition 2.27. *Consider problem* (1) *under Assumption 2.1. Suppose* $\psi(\cdot) = \|\cdot\|$, *where* $\|\cdot\|$ *is any norm. For any* $\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^m$, *there exists a finite* $\rho^*(\tilde{\boldsymbol{\lambda}})$ *such that* $z^{LR+}_{\rho^*}(\tilde{\boldsymbol{\lambda}}) = z^{IP}$.

Proof. Let $\hat{\rho}$ and $\bar{\boldsymbol{\lambda}}$ be as considered in (40). By the equivalence of norms, there exists $0 < \gamma < 1$ such that $\frac{1}{\gamma}\|\boldsymbol{u}\| \geq \|\boldsymbol{u}\|_2 \geq \gamma \|\boldsymbol{u}\|$. From Cauchy-Schwarz inequality, for all $\boldsymbol{x} \in X$, it holds

$$\tilde{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \geq -\|\tilde{\boldsymbol{\lambda}}\|_2 \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2 \geq -\frac{1}{\gamma}\|\tilde{\boldsymbol{\lambda}}\|_2 \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|,$$

$$\bar{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \leq \|\bar{\boldsymbol{\lambda}}\|_2 \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \gamma \|\bar{\boldsymbol{\lambda}}\|_2 \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|,$$

and consequently,

$$\tilde{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \geq \bar{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) - \left(\frac{1}{\gamma}\|\tilde{\boldsymbol{\lambda}}\|_2 + \gamma \|\bar{\boldsymbol{\lambda}}\|_2\right) \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|.$$

Take $\rho^* = \hat{\rho} + \left(\frac{1}{\gamma}\|\tilde{\boldsymbol{\lambda}}\|_2 + \gamma \|\bar{\boldsymbol{\lambda}}\|_2\right)$. Then,

$$\boldsymbol{c}^\top \boldsymbol{x} + \tilde{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho^* \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\| \geq \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \hat{\rho} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|. \tag{42}$$

By taking $\inf_{\boldsymbol{x} \in X}$ from both sides of (42) and considering (40) it is implied that $z^{\mathrm{LR}+}_{\rho^*}(\tilde{\boldsymbol{\lambda}}) \geq z^{\mathrm{IP}}$. This result along with $z^{\mathrm{LR}+}_{\rho^*}(\tilde{\boldsymbol{\lambda}})$ being a lower bound for z^{IP}, concludes $z^{\mathrm{LR}+}_{\rho^*}(\tilde{\boldsymbol{\lambda}}) = z^{\mathrm{IP}}$.

□

Next, we extend Theorem 2.24 to a more general class of augmenting functions than norms.

Theorem 2.28. *Consider an MIP problem* (1) *satisfying Assumption 2.1. Then, there exists a finite* $\hat{\rho}$ *such that* $z^{LD+}_{\hat{\rho}} = z^{LR+}_{\hat{\rho}}(\bar{\boldsymbol{\lambda}}_{LP}) = z^{IP}$ *if* ψ *is an augmenting function such that*

- $\psi(\mathbf{0}) = 0$,
- $\psi(\boldsymbol{u}) \geq \delta > 0$, *for all* $\boldsymbol{u} \notin V$,
- $\psi(\boldsymbol{u}) \geq \gamma \|\boldsymbol{u}\|_\infty$, *for all* $\boldsymbol{u} \in V$,

for some open neighborhood V *of* $\mathbf{0}$, *and positive scalars* $\delta, \gamma > 0$.

Proof. From Proposition 2.27, there exists a finite ρ^* such that $z^{\mathrm{LR}+}_{\rho^*}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) = z^{\mathrm{IP}}$ for $\psi(\cdot) = \|\cdot\|_\infty$. Now, consider the cases where ψ is not a norm but it satisfies the conditions stated above. Take $\hat{\rho} = \max\left\{\frac{z^{\mathrm{IP}} - z^{\mathrm{LP}}}{\delta}, \frac{\rho^*}{\gamma}\right\}$. There are two cases.

1. For all $\boldsymbol{x} \in X$ such that $(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \notin V$, it holds

$$\begin{aligned}
\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \hat{\rho}\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) &\geq z^{\mathrm{LP}} + \hat{\rho}\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \\
&\geq z^{\mathrm{LP}} + \frac{z^{\mathrm{IP}} - z^{\mathrm{LP}}}{\delta}\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \\
&\geq z^{\mathrm{LP}} + z^{\mathrm{IP}} - z^{\mathrm{LP}} \\
&\geq z^{\mathrm{IP}}
\end{aligned}$$

2. For all $\boldsymbol{x} \in X$ such that $(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \in V$, it holds

$$\begin{aligned}
\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \hat{\rho}\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) &\geq \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \frac{\rho^*}{\gamma}\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \\
&\geq \boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \rho^* \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_\infty \\
&\geq z^{\mathrm{IP}}.
\end{aligned}$$

Then,

$$\boldsymbol{c}^\top \boldsymbol{x} + \bar{\boldsymbol{\lambda}}_{\mathrm{LP}}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + \hat{\rho}\psi(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \geq z^{\mathrm{IP}},\ \forall \boldsymbol{x} \in X, \tag{43}$$

which implies $z_{\hat{\rho}}^{\mathrm{LD}+} \geq z_{\hat{\rho}}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}}) \geq z^{\mathrm{IP}}$. This result along with $z_{\hat{\rho}}^{\mathrm{LR}+}(\bar{\boldsymbol{\lambda}}_{\mathrm{LP}})$ being a lower bound for z^{IP}, concludes $z_{\hat{\rho}}^{\mathrm{LR}+}(\tilde{\boldsymbol{\lambda}}) = z^{\mathrm{IP}}$. □

2.6 Conclusions and future work

In this chapter we studied ALD for general linear MIP problems. We presented a primal characterization of ALD for MIPs and showed the asymptotic zero duality gap property with non-negative level bounded and not necessarily convex augmenting functions. Moreover, we showed that under some mild assumptions, ALD achieves zero duality gap for general MIPs with a finite penalty coefficient and a general class of augmenting functions. We also show that some augmenting functions such as the squared Euclidean norm are exact in the pure IP cases, but there exists MIP counterexamples for which these augmenting functions may result in a non-zero duality gap for any value of the penalty coefficient.

Solving IP and MIP problems by ALD may have computational advantages over the classical Lagrangian relaxation approaches, since ALD may produce better dual bounds and provide primal solutions. The main drawback of ALR and ALD methods is that the resulting subproblems are not separable because of the nonlinear augmenting terms. In the convex setting, the alternating direction method of multipliers (ADMM) [28] and related scheme have been developed to overcome this issue. It is not at all clear how to decompose ALD for MIP problems and utilize parallel computation for solving smaller subproblems. In the next chapter, we will develop different exact and heuristic algorithms to solve loosely coupled MIPs in a decentralized fashion.

Chapter III

DECENTRALIZED MIP ALGORITHMS

In this chapter, we propose different exact and heuristic decentralized algorithms for mixed integer programming (MIP) problems. A key challenge is that, because of the non-convex nature of MIPs, classical distributed and decentralized optimization approaches cannot be applied directly to find their optimal solutions. The heuristic method extends the ADMM to mitigate oscillations and traps in local optimality that result from the nonconvexity of MIPs. The proposed exact algorithms are based on adding primal cuts and restricting the Lagrangian relaxation of the original MIP problem.

3.1 Introduction

Consider the MIP problem

$$\begin{aligned} z^{\mathrm{IP}} := \min_{\boldsymbol{x}_1,\cdots\boldsymbol{x}_N} & \sum_{\nu\in\mathcal{P}} \boldsymbol{c}_\nu^\top \boldsymbol{x}_\nu \\ \text{s.t. } & \boldsymbol{x}_\nu \in X_\nu,\ \forall \nu \in \mathcal{P}, \\ & \sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu = \boldsymbol{b}, \end{aligned} \tag{44}$$

where $\mathcal{P} = \{1, \cdots, N\}$ is the set of blocks. Each block ν has its own n_ν dimensional vector $\boldsymbol{x}_\nu$ of decision variables , and local linear constraints

$$\boldsymbol{x}_\nu \in X_\nu, \tag{45}$$

where X_ν is a linear mixed integer set. Different blocks of the problem (44) are linked to each other via the following linear *coupling constraints*:

$$\sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu = \boldsymbol{b}. \tag{46}$$

Each $\boldsymbol{A}_\nu$ is a $m \times n_\nu$ matrix, for all $\nu \in \mathcal{P}$, $\boldsymbol{b}$ is a m dimensional vector, where m is the number of coupling constraints (46).

Details of assumptions and notations are provided in Section 3.2. If $\boldsymbol{A}_\nu$s are sparse matrices and the number of coupling constraints (46) is relatively small comparing to the total number of local constraints of type (45), then we call the problem (44) a *loosely coupled* MIP.

Note that any MIP problem with bounded integer variables can be recast as formulation (44). For example, two-stage 0-1 stochastic MIP programs can be represented as (44), where each scenario is assumed to be a block and the nonanticipativity constraints couple different scenarios. In these problems, the sets U_is are identical and the binary vectors $\boldsymbol{u}_i$s should be the same for different scenarios, due to nonanticipativity constraints.

Another example is the Unit Commitment (UC) problem, which is a challenging MIP problem in power systems. In UC, each electric power generating unit has its own variables and constraints, and only a small percentage of the constraints couple the generators to each other. In UC, binary variables are used to represent the on/off status of each generator at each time. In Chapter 4, we decompose the UC problem on a region based where data privacy is an important issue for different regions.

In this chapter, we propose different exact and heuristic algorithms to solve loosely coupled MIPs in a decentralized fashion. A key challenge is that, because of the non-convex nature of MIPs, classical distributed and decentralized optimization approaches for convex optimization cannot be applied directly to find optimal solutions for MIPs. The proposed algorithms are based on extensions of dual decomposition and ADMM methods. ADMM is an algorithm that is intended to blend the decomposability of dual ascent with the superior convergence properties of the method of multipliers [28]. If the problem (44) is solvable and the sets X_ν are convex, closed, and non-empty for all $\nu \in \mathcal{P}$, then ADMM can solve (44) in a decentralized framework

(see [28] for convergence properties of ADMM). However, binary variables make the sets X_ν nonconvex, which destroys the convergence properties of ADMM.

The chapter is organized as follows. Section 3.2 provides the notations and problem statement. Section 3.3 summarizes the related literature on dual decomposition and ADMM methods. Our exact decentralized algorithms for MIPs are presented in Section 3.5. Our heuristic method is presented in Section 3.4. Experimental results of implementing these algorithms for unit commitment instances will be presented in Chapter 5.

3.2 Preliminaries

Let $\mathbb{R}$ and $\mathbb{Q}$ denote the sets of real and rational numbers. For a finite dimensional vector $\boldsymbol{a}$, denote its transpose by $\boldsymbol{a}^\top$. For a set $\mathcal{S}$, denote its cardinality by $|\mathcal{S}|$. In this chapter, we consider MIP problem (44) which satisfies the following assumptions.

Assumption 3.1. *For the MIP* (44) *we have the following:*

a) For each block $\nu \in \mathcal{P}$, X_ν is a linear mixed integer set defined by

$$X_\nu := \{(\boldsymbol{u}_\nu^\top, \boldsymbol{y}_\nu^\top)^\top : \boldsymbol{u}_\nu \in U_\nu,\ \boldsymbol{y}_\nu \in Y_\nu(\boldsymbol{u}_\nu)\}, \tag{47}$$

where $\boldsymbol{u}_\nu \in \{0,1\}^{n_\nu^1}$ and $\boldsymbol{y}_\nu \in \mathbb{R}^{n_\nu^2}$ are the subvectors of n_ν^1 binary and n_ν^2 continuous decision variables, respectively, with $n_\nu = n_\nu^1 + n_\nu^2$.

b) In description (47) *of X_ν, U_ν and $Y_\nu(\boldsymbol{u}_\nu)$ are subsets of $\{0,1\}^{n_\nu^1}$ and $\mathbb{R}^{n_\nu^2}$, respectively. Because U_ν is a finite set, it can be represented by a set of linear inequalities and integrality constraints. For a given $\boldsymbol{u}_\nu \in U_\nu$, we assume $Y_\nu(\boldsymbol{u}_\nu)$ is a (possibly empty) polyhedron. In particular, let $Y_\nu(\boldsymbol{u}_\nu) = \{\boldsymbol{y}_\nu : \mathbb{R}^{n_\nu^2} : \boldsymbol{E}_\nu \boldsymbol{u}_\nu + \boldsymbol{F}_\nu \boldsymbol{y}_\nu \leq \boldsymbol{g}_\nu\}$, where $\boldsymbol{E}_\nu$ and $\boldsymbol{F}_\nu$ are matrices and $\boldsymbol{g}_\nu$ is a vector of appropriate finite dimensions, independent of the value of $\boldsymbol{u}_\nu$.*

c) $\boldsymbol{c}_\nu$, $\boldsymbol{A}_\nu$, $\boldsymbol{E}_\nu$, $\boldsymbol{F}_\nu$ and $\boldsymbol{g}_\nu$, for all $\nu \in \mathcal{P}$, and $\boldsymbol{b}$ have rational entries.

d) The optimal value of the problem (44) *is bounded and feasible.*

Let $n^1 := \sum_{\nu\in\mathcal{P}} n^1_\nu$ and $n^2 := \sum_{\nu\in\mathcal{P}} n^2_\nu$ denote total number of binary and continuous variables, respectively, and $n = n^1 + n^2$. For convenience, let

$$\boldsymbol{c} := \begin{bmatrix} \boldsymbol{c}_1 \\ \vdots \\ \boldsymbol{c}_N \end{bmatrix},\ \boldsymbol{x} := \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_N \end{bmatrix},\ \boldsymbol{u} := \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_N \end{bmatrix},\ \boldsymbol{y} := \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix},$$

$$\boldsymbol{A} := [\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N],\ X := X_1 \times \cdots \times X_N,\ U := U_1 \times \cdots \times U_N,$$

$$Y(\boldsymbol{u}) := Y_1(\boldsymbol{u}_1) \times \cdots Y_N(\boldsymbol{u}_N).$$

Then, problem (44) can be recast as $z^{\mathrm{IP}} = \min_{\boldsymbol{x}}\{\boldsymbol{c}^\top\boldsymbol{x} : \boldsymbol{x} \in X, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$, which is exactly the MIP problem (1) in Chapter 2.

By Assumption 3.1-d), there exists a solution $\boldsymbol{x}^*$ which satisfies constraints (45) and (46), and $\boldsymbol{c}^\top\boldsymbol{x}^* = z^{\mathrm{IP}}$. Therefore, by data rationality assumption in part (c), the value of the LP relaxation of (44) is bounded [24], i.e. $-\infty < z^{\mathrm{LP}} \leq z^{\mathrm{IP}} < \infty$.

Example 2. *Following is an example for problem* (44) *with two blocks.*

$$
\begin{aligned}
\min \quad & 70u_{11} + 70u_{12} + 110u_{13} + 2y_{11} + 2y_{12} + 48u_{21} + 48u_{22} + 52u_{23} + 3y_{21} + 3y_{22} \\
& \left.\begin{aligned}
& u_{12} - u_{11} - u_{13} \le 0, \\
& 30u_{11} \le y_{11} \le 100u_{11}, \\
s.t. \quad & 30u_{12} \le y_{12} \le 100u_{12}, \\
& -35 \le y_{12} - y_{11} \le 35, \\
& u_{11}, u_{12}, u_{13} \in \{0,1\},
\end{aligned}\right\} \textit{Local constraints for block 1} \\
& \left.\begin{aligned}
& u_{22} - u_{21} - u_{23} \le 0, \\
& 20u_{11} \le y_{21} \le 80u_{21}, \\
& 20u_{12} \le y_{22} \le 80u_{22}, \\
& -30 \le y_{22} - y_{21} \le 30, \\
& u_{21}, u_{22}, u_{23} \in \{0,1\},
\end{aligned}\right\} \textit{Local constraints for block 2} \\
& \left.\begin{aligned}
& y_{11} + y_{21} = 90, \\
& y_{12} + y_{22} = 120.
\end{aligned}\right\} \textit{Coupling constraints}
\end{aligned}
\tag{48}
$$

Recalling the notations described in Sections 3.1 and 3.2, $\boldsymbol{u}_1 = (u_{11}, u_{12}, u_{13})^\top$ *and* $\boldsymbol{u}_2 = (u_{21}, u_{22}, u_{23})^\top$ *are the vectors of binary variables for blocks 1 and 2, respectively. Similarly,* $\boldsymbol{y}_1 = (y_{11}, y_{12})^\top$ *and* $\boldsymbol{y}_2 = (y_{21}, y_{22})^\top$ *are the vectors of continuous variables for blocks 1 and 2, respectively. Then,* $\boldsymbol{x}_1 = (\boldsymbol{u}_1^\top, \boldsymbol{y}_1^\top)^\top$ *and* $\boldsymbol{x}_2 = (\boldsymbol{u}_2^\top, \boldsymbol{y}_2^\top)^\top$ *are the vectors of decision variables for blocks 1 and 2, respectively. Moreover,* $\boldsymbol{u} = (\boldsymbol{u}_1^\top, \boldsymbol{u}_2^\top) = (u_{11}, u_{12}, u_{13}, u_{21}, u_{22}, u_{23})^\top$ *and* $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top) = (y_{11}, y_{12}, y_{21}, y_{22})^\top$ *are*

the overall vectors of binary and continuous variables. In this example, we have

$$\boldsymbol{c}_1 = \begin{bmatrix} 70 \\ 70 \\ 110 \\ 2 \\ 2 \end{bmatrix}, \ \boldsymbol{c}_2 = \begin{bmatrix} 48 \\ 48 \\ 52 \\ 3 \\ 3 \end{bmatrix}, \ \textit{and } \boldsymbol{A}_1 = \boldsymbol{A}_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Moreover,

$$U_1 = \{\boldsymbol{u}_1 \in \{0,1\}^3 : u_{12} - u_{11} - u_{13} \leq 0\},$$

$$U_2 = \{\boldsymbol{u}_2 \in \{0,1\}^3 : u_{22} - u_{21} - u_{23} \leq 0\},$$

$$U = U_1 \times U_2 = \left\{ \boldsymbol{u} \in \{0,1\}^6 : \begin{array}{r} u_{12} - u_{11} - u_{13} \leq 0, \\ u_{22} - u_{21} - u_{23} \leq 0 \end{array} \right\},$$

$$Y_1(\boldsymbol{u}_1) = \left\{ \boldsymbol{y}_1 \in \mathbb{R}^2 : \begin{array}{r} 30u_{11} \leq y_{11} \leq 100u_{11}, \\ 30u_{12} \leq y_{12} \leq 100u_{12}, \\ -35 \leq y_{12} - y_{11} \leq 35 \end{array} \right\},$$

$$Y_2(\boldsymbol{u}_2) = \left\{ \boldsymbol{y}_2 \in \mathbb{R}^2 : \begin{array}{r} 20u_{11} \leq y_{21} \leq 80u_{21} \\ 20u_{12} \leq y_{22} \leq 80u_{22} \\ -30 \leq y_{22} - y_{21} \leq 30 \end{array} \right\},$$

$$X_1 = \left\{ \boldsymbol{x}_1 = (\boldsymbol{u}_1^\top, \boldsymbol{y}_1^\top) \in \{0,1\}^3 \times \mathbb{R}^2 : \begin{array}{r} u_{12} - u_{11} - u_{13} \leq 0, \\ 30u_{11} \leq y_{11} \leq 100u_{11}, \\ 30u_{12} \leq y_{12} \leq 100u_{12}, \\ -35 \leq y_{12} - y_{11} \leq 35, \end{array} \right\},$$

$$X_2 = \left\{ \boldsymbol{x}_2 = (\boldsymbol{u}_2^\top, \boldsymbol{y}_2^\top)^\top \in \{0,1\}^3 \times \mathbb{R}^2 : \begin{array}{r} u_{22} - u_{21} - u_{23} \leq 0, \\ 20u_{11} \leq y_{21} \leq 80u_{21}, \\ 20u_{12} \leq y_{22} \leq 80u_{22}, \\ -30 \leq y_{22} - y_{21} \leq 30, \end{array} \right\}.$$

Recalling from Section 2.3, for a given vector of the dual variables, $\boldsymbol{\mu} \in \mathbb{R}^m$, the standard Lagrangian relaxation (LR) for problem (44) is

$$\begin{aligned} z^{\mathrm{LR}}(\boldsymbol{\mu}) := \boldsymbol{\mu}^\top \boldsymbol{b} + \min_{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N} & \sum_{\nu \in \mathcal{P}} \mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}) \\ \text{s.t. } & \boldsymbol{x}_\nu \in X_\nu, \, \forall \nu \in \mathcal{P}, \end{aligned} \tag{49}$$

where

$$\mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}) := (\boldsymbol{c}_\nu^\top - \boldsymbol{\mu}^\top \boldsymbol{A}_\nu)\boldsymbol{x}_\nu, \, \forall \nu \in \mathcal{P},$$

and the corresponding Lagrangian dual (LD) value is

$$z^{\mathrm{LD}} := \sup_{\boldsymbol{\mu} \in \mathbb{R}^m} z^{\mathrm{LR}}(\boldsymbol{\mu}). \tag{50}$$

Since (49) is a relaxation of (44), $z^{\mathrm{LR}}(\boldsymbol{\mu}) \leq z^{\mathrm{LD}} \leq z^{\mathrm{IP}}$ holds, for any $\boldsymbol{\mu} \in \mathbb{R}^m$. Due to the presence of binary variables, a nonzero duality gap may exists [193], i.e. $z^{\mathrm{LD}} < z^{\mathrm{IP}}$ is possible. Let $\boldsymbol{\mu}^*$ be a maximizer in (50), if there exists one. Obtaining $\boldsymbol{\mu}^*$ and z^{LD} are not straight forward in practice. A popular and easy approach to solve (50) is the subgradient decent method, where the problem (49) is solved iteratively and the dual multipliers are updated at each iteration. Note that problem (49) is separable and it can be solved by computing $\min_{\boldsymbol{x}_\nu} \{\mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}) : \boldsymbol{x}_\nu \in X_\nu\}$ for each block ν.

Even with $\boldsymbol{\mu}^*$ at hand, a *primal feasible solution*, one that satisfies all constraints in model (44), is not readily available. For a given $\hat{\boldsymbol{u}} \in U$, the best corresponding primal feasible solution, if there exists one, and its objective value, $z(\hat{\boldsymbol{u}})$, can be computed by solving the following LP:

$$
\begin{aligned}
z(\hat{\boldsymbol{u}}) := \min_{\boldsymbol{y}_1,\cdots,\boldsymbol{y}_N} & \sum_{\nu\in\mathcal{P}} \boldsymbol{c}_\nu^\top \begin{bmatrix} \hat{\boldsymbol{u}}_\nu \\ \boldsymbol{y}_\nu \end{bmatrix} \\
\text{s.t. } & \boldsymbol{y}_\nu \in Y_\nu(\hat{\boldsymbol{u}}_\nu),\ \forall \nu \in \mathcal{P}, \\
& \sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \begin{bmatrix} \hat{\boldsymbol{u}}_\nu \\ \boldsymbol{y}_\nu \end{bmatrix} = \boldsymbol{b}.
\end{aligned} \tag{51}
$$

Problem (51) is an LP and can be solved with a distributed algorithm. Denote the upper and lower bounds on z^{IP} by ub and lb, respectively. Then, $z(\boldsymbol{u})$ and $z^{\mathrm{LR}}(\boldsymbol{\mu})$ are valid ub and lb, respectively, for all $\boldsymbol{u} \in U$ and $\boldsymbol{\mu} \in \mathbb{R}^m$. In fact, $z^{\mathrm{IP}} = \min_{\boldsymbol{u}\in U} z(\boldsymbol{u})$.

3.3 Related Literature

Dual decomposition and ADMM are two well known distributed optimization technique in the context of convex optimization. Our decentralized MIP algorithms in this chapter are based on extensions of these two techniques. Next, we present these schemes.

3.3.1 Dual Decomposition

Dual decomposition is a well known technique to solve large scale optimization problems. Early works on application of dual decomposition for large scale linear programming can be found in [16, 52, 53, 58].

Let $\rho_\mu^k > 0$ be the step size for updating the dual vector $\boldsymbol{\mu}$ at iteration k. Algorithm 1 represents an overall scheme of a dual decomposition method to solve (44). Each iteration of this method requires a "broadcast" and a "gather" operation. Dual update step (line 11) requires $\boldsymbol{A}_\nu \boldsymbol{x}_\nu^k$ values from all blocks. Once $\boldsymbol{\mu}^k$ is computed, it must be broadcast to all blocks.

A lower bound for z^{IP} can be obtained from Algorithm 1. If $\sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu^k = \boldsymbol{b}$ in some iteration k of this algorithm, $\boldsymbol{x}^k$ is a feasible and optimal solution of (44). But, this case is not likely in practice and there is no hope to find a feasible solution for

(44) by running only Algorithm 1. Therefore, in general we cannot expect to get an upper bound for z^{IP} from this algorithm. A modified version of dual decomposition technique is presented in Algorithm 5 which is able to provide upper bounds for z^{IP}.

Algorithm 1 Basic Dual Decomposition

1: $lb \leftarrow -\infty$, $\boldsymbol{\mu}^0 \leftarrow \mathbf{0}$, and $k \leftarrow 0$.
2: **while** some termination criteria is not met **do**
3: $k \leftarrow k + 1$
4: **for** $\nu := 1$ to N **do**
5: solve $\min_{\boldsymbol{x}_\nu}\{\mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}^{k-1}) : \boldsymbol{x}_\nu \in X_\nu\}$
6: let v_ν^k be the optimal value and $\boldsymbol{x}_\nu^k$ be an optimal solution
7: **end for**
8: **if** $lb < \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu\in\mathcal{P}} v_\nu^k$ **then**
9: $lb \leftarrow \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu\in\mathcal{P}} v_\nu^k$
10: **end if**
11: $\boldsymbol{\mu}^k \leftarrow \boldsymbol{\mu}^{k-1} + \rho_\mu^k \left(\boldsymbol{b} - \sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu^k\right)$
12: **end while**

3.3.2 Alternating Direction Method of Multipliers (ADMM)

ADMM is an algorithm that is intended to blend the separability of dual decomposition with the superior convergence properties of the method of multipliers [28]. Recalling from Section 2.3, for $\rho > 0$ and $\boldsymbol{\mu} \in \mathbb{R}^m$, the augmented Lagrangian with squared Euclidean norm has the following form.

$$\mathcal{L}_\rho^+(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N, \boldsymbol{\mu}) = \sum_{\nu\in\mathcal{P}} \boldsymbol{c}_\nu^\top \boldsymbol{x}_\nu + \boldsymbol{\mu}^\top \left(\boldsymbol{b} - \sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu\right) + \frac{\rho}{2} \left\| \boldsymbol{b} - \sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu \right\|_2^2. \quad (52)$$

It is obvious that $\mathcal{L}_\rho^+(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N, \boldsymbol{\mu})$ in (52) is not separable between different blocks, because the nonlinear (quadratic) terms are coupling different block to each other. For convex optimization problems, a decomposable algorithm to minimize $\mathcal{L}_\rho^+(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N, \boldsymbol{\mu})$ over $\boldsymbol{x} \in X$ is ADMM [28].

3.3.2.1 ADMM with two blocks

Algorithm 2 presents an ADMM approach for an optimization problem with two blocks. In this algorithm, $\mathcal{L}_\rho^+(\boldsymbol{x}_1, \boldsymbol{x}_2^{k-1}, \boldsymbol{\mu}^{k-1})$ is first minimized with respect to $\boldsymbol{x}_1$, assuming that $\boldsymbol{x}_2$ is fixed at its previous value $\boldsymbol{x}_2^{k-1}$. Then, $\mathcal{L}_\rho^+(\boldsymbol{x}_1^k, \boldsymbol{x}_2, \boldsymbol{\mu}^{k-1})$ is minimized with respected to $\boldsymbol{x}_2$, assuming that $\boldsymbol{x}_1$ is fixed at its previous value $\boldsymbol{x}_1^k$. Finally, the vector of dual variables $\boldsymbol{\mu}^k$ is updated. Note that $\rho > 0$ is a given and fixed penalty factor.

Algorithm 2 ADMM procedure for two blocks

1: $\boldsymbol{x}_2^0 \leftarrow \mathbf{0}$, $\boldsymbol{\mu}^0 \leftarrow \mathbf{0}$, and $k \leftarrow 0$
2: **while** some termination criteria is not met **do**
3: $\quad k \leftarrow k + 1$
4: $\quad \boldsymbol{x}_1^k \leftarrow \arg\min_{\boldsymbol{x}_1 \in X_1} \mathcal{L}_\rho^+(\boldsymbol{x}_1, \boldsymbol{x}_2^{k-1}, \boldsymbol{\mu}^{k-1})$
5: $\quad \boldsymbol{x}_2^k \leftarrow \arg\min_{\boldsymbol{x}_2 \in X_2} \mathcal{L}_\rho^+(\boldsymbol{x}_1^k, \boldsymbol{x}_2, \boldsymbol{\mu}^{k-1})$
6: $\quad$ Update $\boldsymbol{\mu}^k \leftarrow \boldsymbol{\mu}^{k-1} + \rho \times [\boldsymbol{b} - (\boldsymbol{A}_1\boldsymbol{x}_1^k + \boldsymbol{A}_2\boldsymbol{x}_2^k)]$
7: **end while**

Let $\boldsymbol{\alpha}^k$ and $\boldsymbol{\beta}^k$ denote vector of primal and dual residuals at iteration k. Then,

$$\boldsymbol{\alpha}^k = \boldsymbol{b} - (A_1\boldsymbol{x}_1^k + A_2\boldsymbol{x}_2^k) \text{ and } \boldsymbol{\beta}^k = \rho\boldsymbol{A}_1^\top\boldsymbol{A}_2(\boldsymbol{x}_2^k - \boldsymbol{x}_2^{k-1}).$$

If problem (44) is solvable and the sets X_1 and X_2 are convex, closed, and nonempty, Algorithm 2 can solve (44) in a decentralized framework. In this case, primal residuals ($\boldsymbol{\alpha}^k$) converge to zero. Moreover, dual variables ($\boldsymbol{\mu}^k$) and objective value converge to their optimal values [28]. In practice, ADMM converges to modest accuracy –sufficient for many applications– within a few tens of iterations [28]. However, direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent [42].

3.3.2.2 Global Variable Consensus Problem with ADMM

To extend ADMM for multi-block minimization problems, a global variable consensus problem can be constructed. An equivalent optimization problem for (44) is as follows.

$$
\begin{aligned}
z^{\mathrm{IP}} := \min_{\boldsymbol{x}_1,\cdots \boldsymbol{x}_N, \bar{\boldsymbol{x}}_1,\cdots,\bar{\boldsymbol{x}}_N} \quad & \sum_{\nu\in\mathcal{P}} \boldsymbol{c}_\nu^\top \boldsymbol{x}_\nu \\
\text{s.t.} \quad & \boldsymbol{x}_\nu \in X_\nu,\ \forall \nu \in \mathcal{P}, \\
& \sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \bar{\boldsymbol{x}}_\nu = \boldsymbol{b}, && \text{(53a)} \\
& \bar{\boldsymbol{x}}_\nu = \boldsymbol{x}_\nu,\ \forall \nu \in \mathcal{P}. && \text{(53b)}
\end{aligned}
$$

Formulation (53) can be decomposed into two parts, where one part includes variables $\boldsymbol{x}_1, \cdots \boldsymbol{x}_N$ and constraints $\boldsymbol{x}_\nu \in X_\nu$, for all $\nu \in \mathcal{P}$, and the other part contains variables $\bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_N$ and constraints (53a). In this case, constraints (53b) are coupling these two parts and Algorithm 2, ADMM with two blocks, can be adjusted to solve problem (53) in a decentralized manner. Algorithm 3, consensus ADMM, represents this process.

Algorithm 3 Consensus ADMM

$\bar{\boldsymbol{x}}^0 \leftarrow \mathbf{0}$, $\boldsymbol{\mu}^0 \leftarrow \mathbf{0}$, and $k \leftarrow 0$
while some termination criteria is not met **do**
  $k \leftarrow k + 1$
  for $\nu := 1$ to N **do**
    $\boldsymbol{x}_\nu^k \leftarrow \arg\min_{\boldsymbol{x}_\nu \in X_\nu} \mathcal{L}_{\rho,\nu}^+(\boldsymbol{x}_\nu, \bar{\boldsymbol{x}}_\nu^{k-1}, \boldsymbol{\mu}_\nu^{k-1})$
  end for
  $\bar{\boldsymbol{x}}^k \leftarrow \arg\min_{\bar{\boldsymbol{x}}} \left\{\mathcal{L}_\rho^+(\boldsymbol{x}^k, \bar{\boldsymbol{x}}, \boldsymbol{\mu}^{k-1}) : \sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \bar{\boldsymbol{x}}_\nu = \boldsymbol{b}\right\}$ by using (54)
  for $\nu := 1$ to N **do**
    $\boldsymbol{\mu}_\nu^k \leftarrow \boldsymbol{\mu}_\nu^{k-1} + \rho \times (\boldsymbol{x}_\nu^k - \bar{\boldsymbol{x}}_\nu^k)$
  end for
end while

Let

$$\mathcal{L}_\rho^+(\boldsymbol{x}, \bar{\boldsymbol{x}}, \boldsymbol{\mu}) := \sum_{\nu\in\mathcal{P}} \mathcal{L}_{\rho,\nu}^+(\boldsymbol{x}_\nu, \bar{\boldsymbol{x}}_\nu, \boldsymbol{\mu}_\nu),$$

where $\mathcal{L}_{\rho,\nu}^+(\boldsymbol{x}_\nu, \bar{\boldsymbol{x}}_\nu, \boldsymbol{\mu}_\nu) := \boldsymbol{c}_\nu^\top \boldsymbol{x}_\nu + \boldsymbol{\mu}_\nu^\top (\boldsymbol{x}_\nu - \bar{\boldsymbol{x}}_\nu) + \frac{\rho}{2}\|\boldsymbol{x}_\nu - \bar{\boldsymbol{x}}_\nu\|_2^2$. Then, the subproblem for part one is $\min_{\boldsymbol{x}}\{\mathcal{L}_\rho^+(\boldsymbol{x}, \bar{\boldsymbol{x}}, \boldsymbol{\mu}) : \boldsymbol{x}_\nu \in X_\nu, \forall \nu \in \mathcal{P}\}$, which is separable between blocks and can be solved in parallel. Moreover, the subproblem for part two is

$$\min_{\bar{\boldsymbol{x}}}\{\mathcal{L}_\rho^+(\boldsymbol{x}, \bar{\boldsymbol{x}}, \boldsymbol{\mu}) : \sum_{\nu\in\mathcal{P}} \boldsymbol{A}_\nu \bar{\boldsymbol{x}}_\nu = \boldsymbol{b}\}$$

which has a closed form solution as follows (assuming $\boldsymbol{A}$ has full row rank):

$$\begin{aligned}\arg\min_{\bar{\boldsymbol{x}}} \left\{ \mathcal{L}_\rho^+(\boldsymbol{x}, \bar{\boldsymbol{x}}, \boldsymbol{\mu}) : \sum_{\nu \in \mathcal{P}} \boldsymbol{A}_\nu \bar{\boldsymbol{x}}_\nu = \boldsymbol{b} \right\} &= \arg\min_{\bar{\boldsymbol{x}}} \left\{ \|\boldsymbol{x} + \frac{\boldsymbol{\mu}}{\rho} - \bar{\boldsymbol{x}}\|_2^2 : \boldsymbol{A}\bar{\boldsymbol{x}} = \boldsymbol{b} \right\} \\ &= [I - \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)^{-1} \boldsymbol{A}](\boldsymbol{x} + \frac{\boldsymbol{\mu}}{\rho}) + \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)^{-1} \boldsymbol{b}\end{aligned} \tag{54}$$

where the second equality is well know in linear algebra for finding the orthogonal projection of a point onto an affine subspace (see e.g. [115, 140]). In general, to compute inverse matrices is not easy [84], but it can be done efficiently for sparse matrices with specific structures.

In decentralized consensus optimization, ADMM has a linear convergence rate [165]. Consensus ADMM can be interpreted as a method for solving problems in which the objective and constraints are distributed across multiple processors. Each processor only has to handle its own objective and constraint term, plus a quadratic term which is updated each iteration. The linear parts of the quadratic terms are updated in such a way that the variables converge to a common value, which is the solution of the full problem [28].

In our context of MIP (44), consensus ADMM (Algorithm 3) can be used for upper bounding z^{IP}. For a set $\hat{\mathcal{S}} \subset U$ an upper bounding method is as Algorithm 4.

Algorithm 4 Upper Bounding Algoritm

1: **for** $\hat{\boldsymbol{u}} \in \hat{\mathcal{S}}$ **do**
2: compute $z(\hat{\boldsymbol{u}})$ by solving LP (51) with consensus ADMM, Algorithm 3
3: **if** $z(\hat{\boldsymbol{u}}) < ub$ **then**
4: $ub \leftarrow z(\hat{\boldsymbol{u}})$
5: $\boldsymbol{u}^* \leftarrow \hat{\boldsymbol{u}}$
6: **end if**
7: **end for**

3.3.3 Combination of Dual Decomposition and Consensus ADMM

A combination of Algorithm 1 (dual decomposition) and Algorithm 3 (consensus ADMM) can be used to generate lower and upper bounds for z^{IP}. Algorithm 5

presents a modified version of Algorithm 1. In this algorithm, for a given binary vector $\hat{\boldsymbol{u}}$, Algorithm 3 (consensus ADMM) is used to refine continuous variables $\boldsymbol{y}$, and obtain an upper bound for z^{IP}. Besides the issues related to the non zero duality gap and the challenges in finding the the best dual vector $\boldsymbol{\mu}^*$, which is a maximizer in (50), it is possible for Algorithms 1 and 5 to cycle between non-optimal solutions forever.

Algorithm 5 Modified Dual Decomposition for MIPs

$ub \leftarrow +\infty$, $\mathcal{S} \leftarrow \emptyset$, $\boldsymbol{u}^* \leftarrow \emptyset$, and $k \leftarrow 0$.
Solve LP relaxation of (44) with ADMM, Algorithm 3. Let z^{LP} be its optimal value, and $\boldsymbol{\mu}^0$ be the dual values for the coupling constraints (46).
$lb \leftarrow z^{\mathrm{LP}}$
while some termination criteria is not met **do**
    $k \leftarrow k + 1$
    for $\nu := 1$ to N **do**
        solve $\min_{\boldsymbol{x}_\nu}\{\mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}^{k-1}) : \boldsymbol{x}_\nu \in X_\nu\}$
        let v_ν^k be the optimal value and $\boldsymbol{x}_\nu^k = (\boldsymbol{u}_\nu^k, \boldsymbol{y}_\nu^k)$ be an optimal solution
    end for
    if $lb < \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} v_\nu^k$ **then**
        $lb \leftarrow \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} v_\nu^k$
    end if
    $\boldsymbol{\mu}^k \leftarrow \boldsymbol{\mu}^{k-1} + \rho_\mu^k \left(\boldsymbol{b} - \sum_{\nu \in \mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu^k \right)$
    if $\boldsymbol{u}_\nu^k \notin \mathcal{S}$ **then**
        $\mathcal{S} \leftarrow \mathcal{S} \cup \{\boldsymbol{u}_\nu^k\}$
        compute $z(\boldsymbol{u}_\nu^k)$ by solving (51) with ADMM, Algorithm 3
        if $z(\boldsymbol{u}_\nu^{k+1}) < ub$ **then**
            $ub \leftarrow z(\boldsymbol{u}_\nu^k)$
            $\boldsymbol{u}^* \leftarrow \boldsymbol{u}_\nu^k$
        end if
    end if
end while

3.4 Heuristic Release-and-Fix Method

Recalling from Section 3.1, binary variables make the sets X_ν nonconvex, which destroys the convergence properties of ADMM. In this section, we propose modifications

to the direct application of ADMM to mitigate this issue. Note that the subproblems in ADMM always have quadratic objective functions because the augmenting function is squared Euclidean norm.

3.4.1 ADMM-Relax

ADMM-Relax denotes the application of ADMM to the continuous relaxation $\mathrm{CR}(X_\nu)$ of X_ν. That is, each of the variables in $\boldsymbol{u}_\nu$ is allowed to take any value between 0 and 1. The subproblems then become convex quadratic programs (QPs), for which ADMM converges [28]. ADMM-Relax can provide a good lower bound when the MIP formulation is tight. Moreover, the primal and dual solutions from this stage can be used for warm-starting ADMM-Release, which we will describe next.

3.4.2 ADMM-Release

ADMM-Release refers to the application of ADMM where the binary variables are required to take 0 or 1 values. In this case, the subproblems then become mixed integer quadratic programs (MIQPs) which are much harder than the corresponding QPs in ADMM-Relax. The computation time due to the MIQP subproblems can be reduced by observing the transitions in binary variables. If, when solving the MIQP subproblems, some regions do not change the values of their binary variables for some number of consecutive iterations (i.e. some elements of $\boldsymbol{u}_\nu$ are remaining constant), then we may fix those binary variables temporarily. This is consistent with the empirical observation that only a subset of binary variables are actively being searched at any given stage of the solution process.

Furthermore, the penalty factor ρ can be decreased or increased at different points depending on the history of solutions. If new binary solutions are needed – as in the case that we have found a feasible global solution and wish to explore for others with superior objective function value – ρ can be decreased to encourage exploration of new binary values. If the ADMM-Release stage has been running for many iterations,

ρ may be increased to force settlement on a binary solution.

3.4.3 ADMM-Fix

When a binary-feasible solution is found, the binary variables can be fixed while the solution of continuous variables is refined through further ADMM iterations. This differs from ADMM-Release in that the whole vector $\boldsymbol{u}_\nu$ is fixed for all ν. Therefore, the only active decision variables are elements of $\boldsymbol{y}_\nu$. With sufficiently low residual tolerances, the solution resulting from this algorithm is implementable, unlike ADMM-Relax. Similar to ADMM-Relax, ADMM-Fix converges because the subproblems are convex QPs.

3.4.4 Release-and-Fix process

The flowchart in Figure 4 illustrates the basic process of the Release-and-Fix (R&F) approach. The ADMM-Relax stage is used to provide good starting points for many decision variables and the dual variables. The following stage is composed of cycles between the ADMM-Release and ADMM-Fix processes where binary solutions are explored, discovered, refined, and recorded before searching for more binary solutions.

Although, R&F algorithm provides an upper bound for the optimality gap of its solutions, it is a heuristic method and may not provide an optimality certificate for a given solution of a MIP problem.

3.4.5 Improvements

The performance of the R&F algorithm can be significantly improved by some of the following:

3.4.5.1 Strengthening the CR

It is desirable to obtain as tight a formulation as possible for each of the subproblems. Specifically, the ideal formulation for the region ν subproblem would describe the convex hull of S_ν, denoted by $\mathbf{conv}(X_\nu)$. By definition, $X_\nu \subseteq \mathbf{conv}(X_\nu) \subseteq \mathrm{CR}(X_\nu)$.

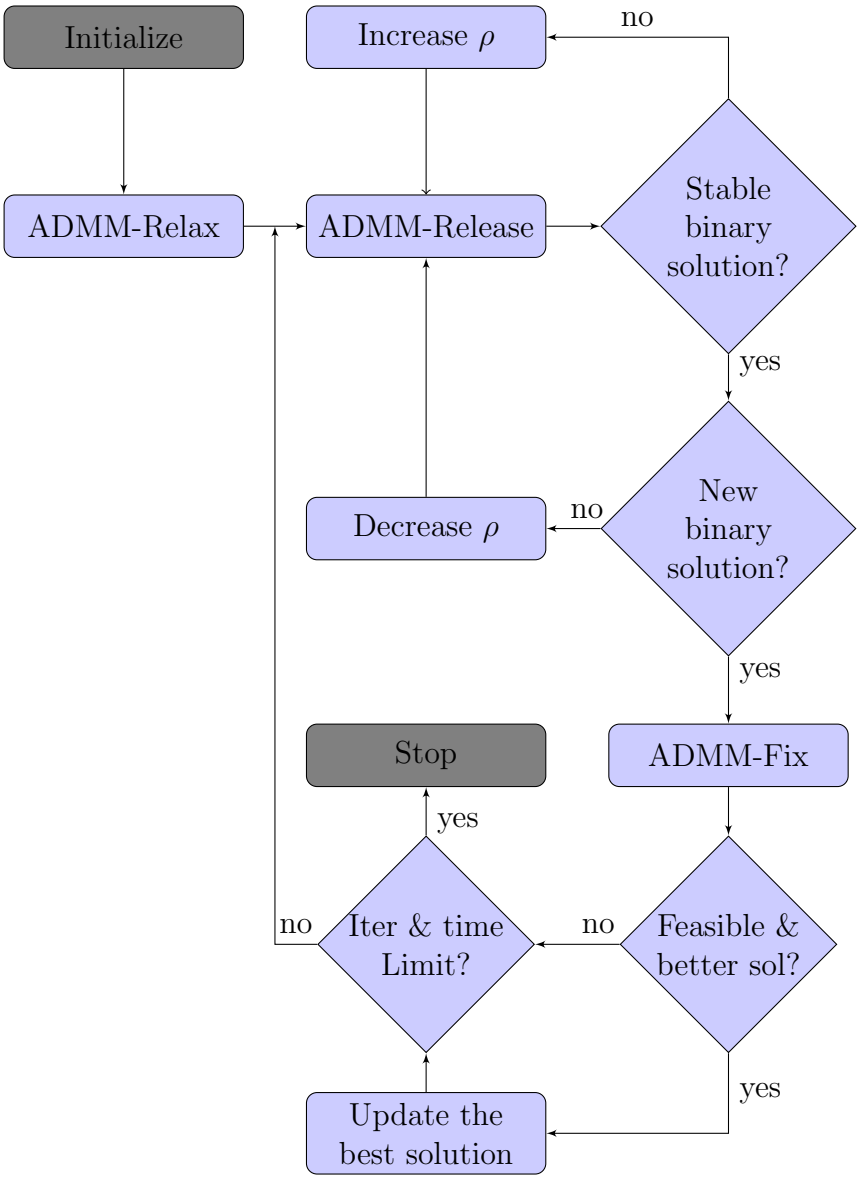

Figure 4: Composition of solution processes for R&F.

Adding valid inequalities to $\mathrm{CR}(X_\nu)$ helps to better approximate $\mathbf{conv}(X_\nu)$. With a quadratic objective, standard valid inequalities for MIPs may not be useful. Linearizing around an intermediate solution point may provide useful cuts as the solution progresses.

One such valid inequality can be determined by solving an auxiliary MIP periodically. Given a solution $\boldsymbol{x}_\nu^* \in \mathrm{CR}(X_\nu)$, find the gradient $\boldsymbol{g} := \nabla_{\boldsymbol{x}_\nu} \mathcal{L}^+_{\rho,\nu}(\boldsymbol{x}_\nu, \boldsymbol{\mu})|_{\boldsymbol{x}_\nu^*}$ at $\boldsymbol{x}_\nu^*$. Then, solve the following auxiliary MIP problem:

$$\alpha_\nu(\boldsymbol{g}) := \min\{\boldsymbol{g}^\top \boldsymbol{x}_\nu : \boldsymbol{x}_\nu \in X_\nu\}. \tag{55}$$

Then, the following valid inequality may be added to the subproblem of region ν:

$$\alpha_\nu(\boldsymbol{g}) \leq \boldsymbol{g}^\top \boldsymbol{x}_\nu. \tag{56}$$

Validity of inequality (56) follows from the fact that $\alpha_\nu(\boldsymbol{g}) \leq \boldsymbol{g}^\top \boldsymbol{x}_\nu$, for all $\boldsymbol{x}_\nu \in X_\nu$, by definition of $\alpha_\nu(\boldsymbol{g})$.

3.4.5.2 Subproblem acceleration

Subproblems in R&F can be accelerated by several tweaks. First, in CPLEX or other MIP solvers, selecting the appropriate root node algorithm for the subproblems can greatly reduce subproblem solution time. When subproblems are smaller, a dual simplex approach is more beneficial than an interior point method, and vice-versa. Furthermore, if a simplex method is used, inheriting the basis from the solution of a previous iteration provides even more speed. In CPLEX 12.6, it was found that the full MIP preprocessing was often run for problems in the ADMM-Fix stage (i.e. all binary variables fixed), even though the problem to solve was effectively a QP. Manually changing the problem type was necessary to leverage this knowledge. Furthermore, if memory limit is not restrictive, different IloCPLEX and IloModel objects can be assigned for each subproblem in ADMM-Release and ADMM-Fix phases.

3.4.5.3 Warm start

As will be specifically seen later from experimental results, initializing the subproblems with primal and dual variables from a hypothetical previous day's solution can sometimes significantly reduce solution time.

3.5 Exact Methods

We propose a decentralized MIP approach where each block solves its own Lagrangian relaxation (LR) subproblem iteratively. The approach evaluates the cost of binary solutions as candidate partial solutions and refines them to get a primal feasible solutions to the overall problem. To improve the lower bound and prevent cycling in Algorithm 5, the explored binary solutions are then cut-off from future consideration in all subproblems.

This idea is similar to the scenario decomposition algorithm for two-stage 0-1 stochastic MIP problems proposed in [3]. In the two-stage 0-1 stochastic MIP model at [3], each scenario is assumed to be a block and nonanticipativity constraints are coupling different scenarios. In that model, binary variables are only present in the first stage and they are the same for different scenarios. Therefore, it is straightforward to cutoff explored binary solutions from the feasible regions of all subproblems. On the contrary, in our loosely coupled MIP model (44), binary variables are not the same for different blocks. Then, it is not clear how to cutoff a global binary solution from the feasible regions of subproblems. For instance, in Example 2, $\boldsymbol{u}_1 = (u_{11}, u_{12}, u_{13})^\top$ and $\boldsymbol{u}_2 = (u_{21}, u_{22}, u_{23})^\top$ are completely different binary vectors for blocks 1 and 2, respectively. In Example 2, consider $\hat{\boldsymbol{u}} = (\hat{\boldsymbol{u}}_1^\top, \hat{\boldsymbol{u}}_2^\top)^\top$ where $\hat{\boldsymbol{u}}_1 = (1, 1, 0)^\top \in U_1$ and $\hat{\boldsymbol{u}}_2 = (0, 0, 0)^\top \in U_2$. Then, it is a challenge to cutoff $\hat{\boldsymbol{u}} = (1, 1, 0, 0, 0, 0)^\top \in U_1 \times U_2$ from the local feasible regions of blocks 1 and 2 in a decentralized and parallel fashion. In this section, we propose three exact algorithms to handle this process in a distributed framework.

For given $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\mathcal{S} \subset U$, we define the *restricted Lagrangian relaxation* (RLR)

$$\begin{aligned} z^{\mathrm{RLR}}(\boldsymbol{\mu}, \mathcal{S}) := \boldsymbol{\mu}^\top \boldsymbol{b} + \min_{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N} & \sum_{\nu \in \mathcal{P}} \mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}) \\ & \text{s.t. } \boldsymbol{x}_\nu \in X_\nu, \forall \nu \in \mathcal{P}, \\ & \boldsymbol{u} \notin \mathcal{S}. \end{aligned} \tag{57}$$

Recall from Assumption 3.1, $\boldsymbol{x}_\nu$ consists of the binary variables' subvector $\boldsymbol{u}$ and the continuous variables' subvector $\boldsymbol{u}$. Note that, $ub = \min_{\hat{\boldsymbol{u}} \in \mathcal{S}} \{z(\hat{\boldsymbol{u}})\}$ and $lb = \min\{z^{\mathrm{RLR}}(\boldsymbol{\mu}, \mathcal{S}), ub\}$ are valid bounds for z^{IP}. Moreover, $z^{\mathrm{RLR}}(\boldsymbol{\mu}, \mathcal{S}) \leq z^{\mathrm{RLR}}(\boldsymbol{\mu}, \mathcal{T})$ for any pair of sets $\mathcal{S}$ and $\mathcal{T}$ such that $\mathcal{S} \subset \mathcal{T} \subset U$. Furthermore, $z^{\mathrm{RLR}}(\boldsymbol{\mu}, U) = +\infty$. Therefore, for any $\boldsymbol{\mu} \in \mathbb{R}^m$, there exists a set $\mathcal{S}(\boldsymbol{\mu}) \subset U$ such that $z^{\mathrm{RLR}}(\boldsymbol{\mu}, \mathcal{S}(\boldsymbol{\mu})) \geq z^{\mathrm{IP}}$.

For a given binary vector $\hat{\boldsymbol{u}} \in \{0,1\}^r$ let us define the simple binary cut (SBC) of $\hat{\boldsymbol{u}}$ in terms of binary decision vector $\boldsymbol{u} \in \{0,1\}^{n^1}$ as follows:

$$\mathrm{SBC}(\boldsymbol{u}, \hat{\boldsymbol{u}}): \quad \sum_{k:\hat{u}_k=0} u_k + \sum_{k:\hat{u}_k=1} (1 - u_k) \geq 1. \tag{58}$$

Then, $\mathrm{SBC}(\boldsymbol{u}, \hat{\boldsymbol{u}})$ for $\hat{\boldsymbol{u}} = (1,1,0,0,0,0)^\top$ in Example 2 is the following inequality:

$$-u_{11} - u_{12} + u_{13} + u_{21} + u_{22} + u_{23} \geq 1. \tag{59}$$

To cutoff multiple solutions, stronger cuts can be used as described in [9]. Then, the constraint $\boldsymbol{u} \notin \mathcal{S}$ in (57) can be represented as $\mathrm{SBC}(\boldsymbol{u}, \hat{\boldsymbol{u}})$, for all $\hat{\boldsymbol{u}} \in \mathcal{S}$. However this constraint couples different blocks to each other and defeats the goal of problem decomposition. For example, in constraint (59), all binary variables from blocks 1 and 2 are present. Next, we propose different techniques to overcome this issue by introducing equivalent formulations of (57) which are decomposable.

3.5.1 Binary Variables Duplication

In our first approach of decoupling the constraint $\boldsymbol{u} \notin \mathcal{S}$ in (57), we propose to duplicate the whole vector of binary variables and give a copy of it to each block.

For each pair of $\nu, \nu' \in \mathcal{P}$, let $\tilde{\boldsymbol{u}}_{\nu,\nu'} \in \tilde{U}_{\nu,\nu'} \subset \{0,1\}^{n^1_{\nu'}}$ be block ν's perception of $\boldsymbol{u}_{\nu'}$, where $U_{\nu,\nu'}$ is the set of all possible values for $\tilde{\boldsymbol{u}}_{\nu,\nu'}$. For convenience, let $\tilde{\boldsymbol{u}}_\nu$ and $\tilde{U}_\nu$ be block ν's perception of $\boldsymbol{u}$ and U. Note that $\tilde{\boldsymbol{u}}_\nu \in \{0,1\}^{n^1}$ and $\tilde{U}_\nu \subset \{0,1\}^{n^1}$, for all $\nu \in \mathcal{P}$.

It can be assumed $U_{\nu'} \subset \tilde{U}_{\nu,\nu'}$ for all $\nu \neq \nu'$ where it is possible that $U_{\nu'} \neq \tilde{U}_{\nu,\nu'}$. For example one may assume $\tilde{U}_{\nu,\nu'} = \{0,1\}^{n^1_{\nu'}}$. Therefore, it may happen $\tilde{U}_{\nu,\nu'} \backslash U_{\nu'} \neq \emptyset$; i.e. block ν may not know any explicit or implicit descriptions of $U_{\nu'}$ and consequently its perception of $\boldsymbol{u}_\nu$ can be infeasible. But, block ν should receive an infeasibility alert from block ν', if $\hat{\boldsymbol{u}}_{\nu,\nu'} \notin U_{\nu'}$. Then, $\hat{\boldsymbol{u}}_{\nu,\nu'}$ can be cut off from $\tilde{U}_{\nu,\nu'}$ using SBC$(\boldsymbol{u}_{\nu,\nu'}, \hat{\boldsymbol{u}}_{\nu,\nu'})$ as defined in (58). In this algorithm, we assume $\tilde{U}_\nu = U$, for the sake of simplicity. Later, we will present other algorithms where the blocks do not need to know anything about the feasibility regions of the other blocks.

For Example 2, blocks 1 and 2 perceptions of the overall binary vector $\boldsymbol{u}$ are $\tilde{\boldsymbol{u}}_1 = (\tilde{u}_{111}, \tilde{u}_{112}, \tilde{u}_{113}, \tilde{u}_{121}, \tilde{u}_{122}, \tilde{u}_{123})^\top$ and $\tilde{\boldsymbol{u}}_2 = (\tilde{u}_{211}, \tilde{u}_{212}, \tilde{u}_{213}, \tilde{u}_{221}, \tilde{u}_{222}, \tilde{u}_{223})^\top$, respectively. In this case, $\tilde{\boldsymbol{u}}_{11} = (\tilde{u}_{111}, \tilde{u}_{112}, \tilde{u}_{113})^\top \in U_1$, $\tilde{\boldsymbol{u}}_{12} = (\tilde{u}_{121}, \tilde{u}_{122}, \tilde{u}_{123})^\top \in U_1$, $\tilde{\boldsymbol{u}}_{21} = (\tilde{u}_{211}, \tilde{u}_{212}, \tilde{u}_{213})^\top \in U_1$, and $\tilde{\boldsymbol{u}}_{22} = (\tilde{u}_{221}, \tilde{u}_{222}, \tilde{u}_{223})^\top \in U_2$. Then, SBC$(\boldsymbol{u}, \hat{\boldsymbol{u}})$ cut (59) for $\hat{\boldsymbol{u}} = (1,1,0,0,0,0)^\top$ can be reformulated as

$$-\tilde{\boldsymbol{u}}_{111} - \tilde{\boldsymbol{u}}_{112} + \tilde{\boldsymbol{u}}_{113} + \tilde{\boldsymbol{u}}_{121} + \tilde{\boldsymbol{u}}_{122} + \tilde{\boldsymbol{u}}_{123} \geq 1, \tag{60}$$

and

$$-\tilde{\boldsymbol{u}}_{211} - \tilde{\boldsymbol{u}}_{212} + \tilde{\boldsymbol{u}}_{213} + \tilde{\boldsymbol{u}}_{221} + \tilde{\boldsymbol{u}}_{222} + \tilde{\boldsymbol{u}}_{223} \geq 1. \tag{61}$$

for blocks 1 and 2, respectively. Note that in inequality (60), only (perception) binary variables from block 1 are present. in Similarly, in inequality (61), only (perception) binary variables from block 2 are present.

An equivalent formulation for (57) can be constructed by using the binary vectors $\tilde{\boldsymbol{u}}_1, \cdots, \tilde{\boldsymbol{u}}_N$, where all the blocks have the same perceptions of $\boldsymbol{u}$, i.e.

$$\tilde{\boldsymbol{u}}_1 = \cdots = \tilde{\boldsymbol{u}}_N, \tag{62}$$

and the $\boldsymbol{u} \notin \mathcal{S}$ is replaced by

$$\tilde{\boldsymbol{u}}_\nu \in U \backslash \mathcal{S}. \tag{63}$$

In Example 2, constraint (62) has the following form

$$\begin{aligned}
\tilde{u}_{111} &= \tilde{u}_{211}, \\
\tilde{u}_{112} &= \tilde{u}_{212}, \\
\tilde{u}_{113} &= \tilde{u}_{213}, \\
\tilde{u}_{121} &= \tilde{u}_{221}, \\
\tilde{u}_{122} &= \tilde{u}_{222}, \\
\tilde{u}_{123} &= \tilde{u}_{223}.
\end{aligned}$$

Note that for all $\nu' \neq \nu$, binary vectors $\tilde{\boldsymbol{u}}_{\nu,\nu'}$ are redundant. But, they make it possible to cut a global binary solution $\hat{\boldsymbol{u}}$ from the feasible region of all blocks. In other words, we use $\tilde{\boldsymbol{u}}_{\nu,\nu'}$ for $\nu' \neq \nu$ to handle constraint (63). Let $\tilde{\boldsymbol{x}}_\nu := (\tilde{\boldsymbol{u}}_\nu, \boldsymbol{y}_\nu) \in \{0,1\}^{n^1} \times \mathbb{R}^{n^2_\nu}$. Note that for all $\nu \in \mathcal{P}$, $\boldsymbol{u}_\nu$ is a subvector of $\tilde{\boldsymbol{u}}_\nu$ and consequently, $\boldsymbol{x}_\nu = (\boldsymbol{u}_\nu, \boldsymbol{y}_\nu)$ is a subvector of $\tilde{\boldsymbol{x}}_\nu$. Then, problem (57) can be reformulated as follows:

$$\begin{aligned}
z^{\mathrm{RLR}}(\boldsymbol{\mu}, \mathcal{S}) = \boldsymbol{\mu}^\top \boldsymbol{b} + \min_{\tilde{\boldsymbol{x}}_1, \cdots, \tilde{\boldsymbol{x}}_N} & \sum_{\nu \in \mathcal{P}} \mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}) \\
\text{s.t. } & \boldsymbol{x}_\nu \in X_\nu \text{ and } \tilde{\boldsymbol{u}}_\nu \in U \backslash \mathcal{S}, \ \forall \nu \in \mathcal{P} \\
& \tilde{\boldsymbol{u}}_1 = \cdots = \tilde{\boldsymbol{u}}_N.
\end{aligned} \tag{64}$$

In the model (64), the consensus constraints (62) are joint between different blocks. To decouple these constraints, we use vectors of dual variables $\boldsymbol{\lambda}_\nu \in \mathbb{R}^{n^1}$, for all ν such that $\sum_{\nu \in \mathcal{P}} \boldsymbol{\lambda}_\nu = \boldsymbol{0}$. Then, the new restricted Lagrangian relaxation for the model (44) is

$$\begin{aligned}
z^{\mathrm{RLR}'}(\boldsymbol{\mu}, \boldsymbol{\lambda}, \mathcal{S}) := \boldsymbol{\mu}^\top \boldsymbol{b} + \min_{\tilde{\boldsymbol{x}}_1, \cdots, \tilde{\boldsymbol{x}}_N} & \sum_{\nu \in \mathcal{P}} \mathcal{L}'_\nu(\tilde{\boldsymbol{x}}_\nu, \boldsymbol{\mu}, \boldsymbol{\lambda}_\nu) \\
\text{s.t. } & \boldsymbol{x}_\nu \in X_\nu \text{ and } \tilde{\boldsymbol{u}}_\nu \in U \backslash \mathcal{S}, \ \forall \nu \in \mathcal{P},
\end{aligned} \tag{65}$$

where $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_N)$ and $\mathcal{L}'_\nu(\tilde{\boldsymbol{x}}_\nu, \boldsymbol{\mu}, \boldsymbol{\lambda}_\nu) := (\boldsymbol{c}_\nu^\top - \boldsymbol{\mu}^\top \boldsymbol{A}_\nu)\boldsymbol{x}_\nu + \boldsymbol{\lambda}_\nu^\top \tilde{\boldsymbol{u}}_\nu$.

To solve problem (65), it is sufficient for each block ν to solve its subproblem of $\min_{\tilde{\boldsymbol{x}}_\nu}\{\mathcal{L}'_\nu(\tilde{\boldsymbol{x}}_\nu, \boldsymbol{\mu}, \boldsymbol{\lambda}_\nu) : \boldsymbol{x}_\nu \in X_\nu \text{ and } \tilde{\boldsymbol{u}}_\nu \in U \backslash \mathcal{S}\}$

Note that $z^{\mathrm{RLR}'}(\boldsymbol{\mu}, \boldsymbol{\lambda}, \mathcal{S}) \leq z^{\mathrm{RLR}}(\boldsymbol{\mu}, \mathcal{S})$, for all $\mathcal{S} \subset \{0,1\}^{n^1}$, $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\boldsymbol{\lambda}_\nu \in \mathbb{R}^{n^1}$, $\forall \nu \in \mathcal{P}$ such that $\sum_{\nu \in \mathcal{P}} \boldsymbol{\lambda}_\nu = \mathbf{0}$.

Algorithm 6 Distributed MIP with Binary Variables Duplication

1: Run Algorithm 5 to initialize ub, lb, $\boldsymbol{u}^*$, $\boldsymbol{\mu}^0$ and $\mathcal{S}$.
2: $\boldsymbol{\lambda}^0 \leftarrow \mathbf{0}$ and $k \leftarrow 0$.
3: **while** $ub > lb$ **do**
4: Lower bounding:
5: **while** some termination criteria is not met **do**
6: $k \leftarrow k + 1$
7: **for** $\nu := 1$ to N **do**
8: solve $\min_{\tilde{\boldsymbol{x}}_\nu}\{\mathcal{L}'_\nu(\tilde{\boldsymbol{x}}_\nu, \boldsymbol{\mu}^{k-1}, \boldsymbol{\lambda}_\nu^{k-1}) : \boldsymbol{x}_\nu \in X_\nu \text{ and } \tilde{\boldsymbol{u}}_\nu \in U \backslash \mathcal{S}\}$.
9: let v_ν^k be the optimal value and $\tilde{\boldsymbol{x}}_\nu^k = (\tilde{\boldsymbol{u}}_\nu^k, \boldsymbol{y}_\nu^k)$ be an optimal solution
10: **end for**
11: **if** $lb < \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} v_\nu^k$ **then**
12: $lb \leftarrow \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} v_\nu^k$
13: **end if**
14: $\bar{\boldsymbol{u}}^k \leftarrow \frac{1}{|\mathcal{P}|} \sum_{\nu \in \mathcal{P}} \tilde{\boldsymbol{u}}_\nu^k$
15: $\boldsymbol{\mu}^k \leftarrow \boldsymbol{\mu}^{k-1} + \rho_\mu^k \left(\boldsymbol{b} - \sum_{\nu \in \mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu^k \right)$ and $\boldsymbol{\lambda}_\nu^k \leftarrow \boldsymbol{\lambda}_\nu^{k-1} + \rho_\lambda^k \left(\tilde{\boldsymbol{u}}_\nu^k - \bar{\boldsymbol{u}}^k \right)$
16: **end while**
17: Let $\hat{\mathcal{S}}^k = \cup_{\nu \in \mathcal{P}} \{\tilde{\boldsymbol{u}}_\nu^k\}$.
18: Upper bounding: run Algorithm 4 for set $\hat{\mathcal{S}}$ to update ub and $\boldsymbol{u}^*$.
19: $\mathcal{S} \leftarrow \mathcal{S} \cup \hat{\mathcal{S}}^k$
20: **end while**

Let $\rho_\mu^k, \rho_\lambda^k > 0$ be the step size for updating the dual vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ at iteration k. Then, our first exact decentralized MIP method is as Algorithm 6. This algorithm is initialized by running ADMM to solve the LP relaxation and then switches to dual decomposition. In fact, this step initializes upper and lower bounds as well as dual vectors. In the lower bounding loop (lines 5-16) of Algorithm 6, problem (65) is solved in parallel by each block and the dual vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are updated as well as the lower bound and candidate binary subvectors. Then, each candidate binary subvector is evaluated by solving an LP with ADMM method. In this step, the upper

bound is updated. Finally, the candidate binary subvectors are added to set $\mathcal{S}$ and consequently are cutoff from feasible regions of all blocks. The algorithm continues until the lower bound hits the upper bound.

3.5.2 Auxiliary Binary Variables

In Algorithm 6, each block has as many binary variables as n^1, the number of overall binaries in the original MIP problem (44). Moreover, each block ν needs to know the constraints defining the set $U_{\nu'}$, for all $\nu' \neq \nu$ or to be able to check the feasibility of $\tilde{\boldsymbol{u}}_{\nu,\nu'}$. Next, we propose another algorithm by introducing some auxiliary binary variables, in which different blocks do not need to know about other blocks' binary variables or feasible regions.

For a given $\mathcal{S} \subset U$, let $\mathcal{S}_\nu$, for all $\nu \in \mathcal{P}$, be the minimal sets such that $\mathcal{S}_\nu \subset U_\nu$ and $\mathcal{S} \subset \mathcal{S}_1 \times \cdots \times \mathcal{S}_N$. That is for all $\hat{\boldsymbol{u}}_\nu \in \mathcal{S}_\nu$ and $\nu \in \mathcal{P}$, there exists a $\hat{\boldsymbol{u}} \in \mathcal{S}$ such that the νth block of $\hat{\boldsymbol{u}}$ is $\hat{\boldsymbol{u}}_\nu$. Let $S_\nu := \{1, \cdots, |\mathcal{S}_\nu|\}$ and denote the lth solution of $\mathcal{S}_\nu$ by $\hat{\boldsymbol{u}}_\nu(l)$.

Example 3. *Consider Example 2 with* $\mathcal{S} = \{(1,1,0,0,0,0), (1,1,0,0,1,1)\}$. *Then, it holds* $\mathcal{S}_1 = \{(1,1,0)\}$ *and* $\mathcal{S}_2 = \{(0,0,0), (0,1,1)\}$.

For $\nu, \nu' \in \mathcal{P}$ and $l \in S_{\nu'}$, let $w_{\nu,\nu',l}$ be a binary variable which is 1, if block ν's perception of $\boldsymbol{u}_{\nu'}$ is $\hat{\boldsymbol{u}}_{\nu'}(l)$, and 0 otherwise. For convenience, let $w_{\nu,\nu',0}$ be a binary variable which is 1, if block ν's perception of $\boldsymbol{u}_{\nu'}$ is not in $\mathcal{S}_{\nu'}$, and 0 otherwise. Then,

$$w_{\nu,\nu',l} \in \{0,1\},\ \forall \nu' \in N,\ l \in S_{\nu'} \cup \{0\}. \tag{66}$$

Then, for Example 3, block 1 has auxiliary binary variables $w_{1,1,0}$, $w_{1,1,1}$, $w_{1,2,0}$, $w_{1,2,1}$, $w_{1,2,2}$. Binary variable w_{111} is 1 if and only if block 1 perception of $\boldsymbol{u}_1$ are $(1,1,0)$. Binary variables w_{121} and w_{122} are 1 if and only if blocks 1 perceptions of $\boldsymbol{u}_2$ are $(0,0,0)$ and $(0,1,1)$, respectively. Similarly, w_{110} and w_{120} are 1 if and only if blocks 1 perceptions of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ do not exist in $\mathcal{S}_1$ and $\mathcal{S}_1$, respectively. Likewise, block 2 has auxiliary binary variables $w_{2,1,0}$, $w_{2,1,1}$, $w_{2,2,0}$, $w_{2,2,1}$, $w_{2,2,2}$.

Note that block ν does not know the length of $\boldsymbol{u}_{\nu'}$ or the values in the $\hat{\boldsymbol{u}}_\nu(l)$, unless $\nu = \nu'$. Therefore, $\boldsymbol{u}_\nu = \hat{\boldsymbol{u}}_\nu(l)$ if and only if $w_{\nu,\nu,l} = 1$. This relation between the binary vector $\boldsymbol{u}_\nu$ and the binary variable $w_{\nu,\nu,l}$ can be imposed by constraints (67) and (68).

$$\begin{cases} u_{\nu k} \geq w_{\nu,\nu,l}, & \text{if } \hat{u}_{\nu k}(l) = 1 \\ u_{\nu k} \leq 1 - w_{\nu,\nu,l}, & \text{Otherwise} \end{cases} \quad \forall l \in S_\nu, k = 1, \cdots, n_\nu^1, \tag{67}$$

$$\sum_{k:\hat{u}_{\nu k}(l)=0} u_{\nu k} + \sum_{k:\hat{u}_{\nu k}(l)=1} (1 - u_{\nu k}) \geq w_{\nu,\nu,0}, \; \forall l \in S_\nu. \tag{68}$$

Each block ν should consider exactly one of the binary solutions $\hat{\boldsymbol{u}}_{\nu'}$ in $\mathcal{S}_{\nu'}$, for all $\nu' \in \mathcal{P}$, i.e.

$$\sum_{l \in S_\nu \cup \{0\}} w_{\nu,\nu',l} = 1, \; \forall \nu' \in \mathcal{P}. \tag{69}$$

Inequality (70) cuts the explored binary solutions to prevent cycling.

$$\sum_{\nu' \in \mathcal{P}} \left[\sum_{l:\hat{\boldsymbol{u}}_{\nu'}(l) \neq \hat{\boldsymbol{u}}_{\nu'}(s)} w_{\nu,\nu',l} + \sum_{l:\hat{\boldsymbol{u}}_{\nu'}(l) = \hat{\boldsymbol{u}}_{\nu'}(s)} (1 - w_{\nu,\nu',l}) \right] \geq 1, \; \forall s \in \mathcal{S}, \tag{70}$$

Because of the constraints (66) and (69), constraint (70) can be strengthened as follows:

$$\sum_{\substack{\nu' \in \mathcal{P} \\ l \in S_{\nu'}: \hat{\boldsymbol{u}}_{\nu'}(l) \neq \hat{\boldsymbol{u}}_{\nu'}(s)}} w_{\nu,\nu',l} \leq N - 1, \; \forall s \in \mathcal{S}. \tag{71}$$

Constraints (67)-(69), and (71) for block 2 in Example 3 have the following form:

$$\left.\begin{array}{l} u_{21} \leq 1 - w_{221}, \; u_{22} \leq 1 - w_{221}, \; u_{23} \leq 1 - w_{221}, \\ u_{21} \leq 1 - w_{222}, \; u_{22} \geq w_{222}, \; u_{23} \geq w_{222}, \end{array}\right\} \text{Constraint (67)}$$

$$\left.\begin{array}{l} u_{21} + u_{22} + u_{22} \geq w_{220}, \\ u_{21} + 1 - u_{22} + 1 - u_{22} \geq w_{220}, \end{array}\right\} \text{Constraint (68)}$$

$$\left.\begin{array}{l} w_{210} + w_{211} = 1, \\ w_{220} + w_{221} + w_{223} = 1, \end{array}\right\} \text{Constraint (69)}$$

$$\left.\begin{array}{l} w_{211} + w_{221} \leq 1, \\ w_{211} + w_{222} \leq 1. \end{array}\right\} \text{Constraint (71)}$$

Let $\boldsymbol{w}_\nu$ be the vector of all binary variables $w_{\nu,\nu',l}$, for all $\nu' \in \mathcal{P}$ and all $l \in S_{\nu'}$ In the second distributed MIP approach, we use the auxiliary binary vector $\boldsymbol{w}_\nu \in \{0,1\}^{|\mathcal{P}|+\sum_{\nu'\in\mathcal{P}}|\mathcal{S}_{\nu'}|}$, for all $\nu \in \mathcal{P}$, to develop another equivalent model for (57). Considering the consensus constraints

$$\boldsymbol{w}_1 = \cdots = \boldsymbol{w}_N, \tag{72}$$

problem (57) can be reformulated as follows.

$$\begin{aligned} z^{\text{RLR}}(\boldsymbol{\mu}, \mathcal{S}) = \boldsymbol{\mu}^\top \boldsymbol{b} + \min_{\boldsymbol{x},\boldsymbol{w}_1,\cdots,\boldsymbol{w}_N} & \sum_{\nu\in\mathcal{P}} \mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}) \\ \text{s.t. } & \boldsymbol{x}_\nu \in X_\nu \text{ and } (66)-(69),(71),\ \forall \nu \in \mathcal{P}, \\ & \boldsymbol{w}_1 = \cdots = \boldsymbol{w}_N. \end{aligned} \tag{73}$$

Consensus constraints (72) are coupling different block in the problem (73). To decouple these constraints, we use the dual variable vectors $\boldsymbol{\gamma}_\nu \in \mathbb{R}^{|\mathcal{P}|+\sum_{\nu'\in\mathcal{P}}|\mathcal{S}_{\nu'}|}$, for all $\nu \in \mathcal{P}$ such that $\sum_{\nu\in\mathcal{P}} \boldsymbol{\gamma}_\nu = \boldsymbol{0}$. Then, the new restricted Lagrangian relaxation for the model (44) is

$$\begin{aligned} z_{\mathcal{S}}^{\text{RLR}''}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathcal{S}) := \boldsymbol{\mu}^\top \boldsymbol{b} + \min_{\boldsymbol{x},\boldsymbol{w}_1,\cdots,\boldsymbol{w}_N} & \sum_{\nu\in\mathcal{P}} \mathcal{L}''_\nu(\boldsymbol{x}_\nu, \boldsymbol{w}_\nu, \boldsymbol{\mu}, \boldsymbol{\gamma}_\nu) \\ \text{s.t.} \quad & \boldsymbol{x}_\nu \in X_\nu, \text{ and } (66)-(69),(71),\ \forall i \in \mathcal{N}, \end{aligned} \tag{74}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_N)$ and $\mathcal{L}''_\nu(\boldsymbol{x}_\nu, \boldsymbol{w}_\nu, \boldsymbol{\mu}, \boldsymbol{\gamma}_\nu) := (\boldsymbol{c}_\nu^\top - \boldsymbol{\mu}^\top \boldsymbol{A}_\nu)\boldsymbol{x}_\nu + \boldsymbol{\gamma}_\nu \boldsymbol{w}_\nu$.

Let $\rho_\gamma^k > 0$ be the step size for updating the dual vector $\boldsymbol{\gamma}$ at iteration k. Then, our second exact decentralized MIP approach is as Algorithm 7. The overall scheme of Algorithm 7 is similar to Algorithm 6. The main difference is that instead of problem (65), problem (74) is solved in parallel in the lower bounding loop (lines 6-17) of Algorithm 7. Different blocks do not need to know about other blocks' vector $\boldsymbol{u}_\nu$ of binary variables or feasible regions U_ν to solve problem (74) in parallel. Moreover, in

line 20 of Algorithm 7, a new binary solution is added to $\mathcal{S}_\nu$ which results in adding a new corresponding binary variable w and a new dual variable γ to all blocks.

Algorithm 7 Distributed MIP with Auxiliary Binary Variables

1: Run Algorithm 5 to initialize ub, lb, $\boldsymbol{u}^*$, $\boldsymbol{\mu}^0$ and $\mathcal{S}$.
2: Based on $\mathcal{S}$, set up the sets $\mathcal{S}_\nu$, for all $\nu \in \mathcal{P}$.
3: $\boldsymbol{\gamma}^0 \leftarrow \mathbf{0}$ and $k \leftarrow 0$.
4: **while** $ub > lb$ **do**
5: Lower bounding:
6: **while** some termination criteria is not met **do**
7: $k \leftarrow k + 1$.
8: **for** $\nu := 1$ to N **do**
9: solve $\min_{\boldsymbol{x}_\nu, \boldsymbol{w}_\nu} \{\mathcal{L}''_\nu(\boldsymbol{x}_\nu, \boldsymbol{w}_\nu, \boldsymbol{\mu}^{k-1}, \boldsymbol{\gamma}_\nu^{k-1}) : \boldsymbol{x}_\nu \in X_\nu, (66) - (69), (71)\}$
10: let v_ν^k be the optimal value and $(\boldsymbol{x}_\nu^k, \boldsymbol{w}_\nu^k)$ be an optimal solution
11: **end for**
12: **if** $lb < \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} v_\nu^k$ **then**
13: $lb \leftarrow \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} v_\nu^k$
14: **end if**
15: $\bar{\boldsymbol{w}}^k \leftarrow \frac{1}{|\mathcal{P}|} \sum_{\nu \in \mathcal{P}} \boldsymbol{w}_\nu^k$
16: $\boldsymbol{\mu}^k \leftarrow \boldsymbol{\mu}^{k-1} + \rho_\mu^k \left(\boldsymbol{b} - \sum_{\nu \in \mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu^k \right)$ and $\boldsymbol{\gamma}_\nu^k \leftarrow \boldsymbol{\gamma}_\nu^{k-1} + \rho_\gamma^k \left(\boldsymbol{w}_\nu^k - \bar{\boldsymbol{w}}^k \right)$
17: **end while**
18: **for** $\nu := 1$ to N **do**
19: **if** $\sum_{\nu' \in \mathcal{P}} w_{\nu,\nu',0} \geq 1$ **then**
20: $\mathcal{S}_\nu \leftarrow \mathcal{S}_\nu \cup \{\boldsymbol{u}_\nu(0)\}$
21: **end if**
22: **end for**
23: Let $\tilde{\boldsymbol{u}}_\nu^k$ be the corresponding $\tilde{\boldsymbol{u}}_\nu \in U$ to $\boldsymbol{w}_\nu^k$
24: $\hat{\mathcal{S}} \leftarrow \cup_{\nu \in \mathcal{P}} \{\tilde{\boldsymbol{u}}_\nu^k\}$.
25: Upper bounding: run Algorithm 4 for set $\hat{\mathcal{S}}$ to update ub and $\boldsymbol{u}^*$.
26: $\mathcal{S} \leftarrow \mathcal{S} \cup \hat{\mathcal{S}}^k$
27: **end while**

3.5.3 Auxiliary IP Approach

Next, we propose and extension of Algorithm 7 where dual vectors $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are updated in different loops by solving pure 0-1 and simple MIP subproblems, respectively.

For a given $\hat{\boldsymbol{u}}_\nu \in U_\nu$ and $\boldsymbol{\mu} \in \mathbb{R}^m$, let us define

$$\begin{aligned} z'_\nu(\hat{\boldsymbol{u}}_\nu, \boldsymbol{\mu}) :=& \min_{\boldsymbol{x}_\nu} \{\mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}) : \boldsymbol{x}_\nu \in X_\nu, \boldsymbol{u}_\nu = \hat{\boldsymbol{u}}_\nu\} \\ =& \min_{\boldsymbol{y}_\nu} \left\{ (\boldsymbol{c}_i^\top - \boldsymbol{\mu}^\top \boldsymbol{A}_i) \begin{bmatrix} \hat{\boldsymbol{u}}_\nu \\ \boldsymbol{y}_\nu \end{bmatrix} : \boldsymbol{y}_\nu \in Y_\nu(\hat{\boldsymbol{u}}_\nu) \right\}, \end{aligned} \tag{75}$$

and

$$\boldsymbol{u}'_\nu(\mathcal{S}_\nu, \boldsymbol{\mu}) = \arg \min_{\boldsymbol{u}_\nu \in U_\nu \setminus \mathcal{S}_\nu} \{z'_\nu(\boldsymbol{u}_\nu, \boldsymbol{\mu})\}. \tag{76}$$

Note that calculating $z'_\nu(\hat{\boldsymbol{u}}_\nu, \boldsymbol{\mu})$ in (75) requires solving a LP while $\boldsymbol{u}'_\nu(\mathcal{S}_\nu, \boldsymbol{\mu})$ in (76) can be obtained by solving a MIP problem. It is easy to check that

$$\begin{aligned} z^{\mathrm{LR}}(\boldsymbol{\mu}) &= \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} \min_{\boldsymbol{x}_\nu} \{\mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}) : \boldsymbol{x}_\nu \in X_\nu\} \\ &= \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} \min_{\boldsymbol{u}_\nu \in U_\nu} \min_{\boldsymbol{y}_\nu} \{\mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}) : \boldsymbol{y}_\nu \in Y_\nu(\boldsymbol{u}_\nu)\} \\ &= \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} \min_{\boldsymbol{u}_\nu \in U_\nu} z'_\nu(\boldsymbol{u}_\nu, \boldsymbol{\mu}). \end{aligned}$$

For the sake of simplicity, for fixed $\mathcal{S}_\nu$, $\boldsymbol{\mu}$ and for all $l \in S_\mu \cup \{0\}$, we denote $\boldsymbol{u}'_\nu(\mathcal{S}_\nu, \boldsymbol{\mu})$ and $z'_\nu(\hat{\boldsymbol{u}}_\nu(l), \boldsymbol{\mu})$ by $\hat{\boldsymbol{u}}_\nu(0)$ and $\hat{z}_\nu(l)$, respectively. Then,

$$\begin{aligned} z^{\mathrm{RLR}}(\boldsymbol{\mu}, \mathcal{S}) = \boldsymbol{\mu}^\top \boldsymbol{b} + & \min_{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_N} \sum_{\nu \in \mathcal{P}} \sum_{l \in S_\nu \cup \{0\}} \hat{z}_\nu(l) w_{\nu,\nu,l} \\ & \text{s.t. (66), (69) and (71), } \forall \nu \in \mathcal{P}, \\ & \boldsymbol{w}_1 = \cdots = \boldsymbol{w}_N \end{aligned} \tag{77}$$

Note that (77) which is a pure IP problem, has the same optimal value as (73), but $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{y})$ is not the vector of decision variables in (77). Moreover, model (77) does not have the constraints $\boldsymbol{x}_\nu \in X_\nu$, (67) and (68). The Lagrangian relaxation of (77) can be represented as follows:

$$\begin{aligned} z^{\mathrm{RLR}'''}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathcal{S}) := \boldsymbol{\mu}^\top \boldsymbol{b} + & \min_{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_N} \sum_{\nu \in \mathcal{P}} \mathcal{L}'''_\nu(\boldsymbol{w}_\nu, \boldsymbol{\mu}, \boldsymbol{\gamma}_\nu) \\ & \text{s.t.} \quad \text{(66), (69) and (71), } \forall i \in \mathcal{N}, \end{aligned} \tag{78}$$

where,

$$\mathcal{L}'''_\nu(\boldsymbol{w}_\nu, \boldsymbol{\mu}, \boldsymbol{\gamma}_\nu) := \sum_{l \in S_\nu \cup \{0\}} \hat{z}_\nu(l) w_{\nu,\nu,l} + \boldsymbol{\gamma}_\nu \boldsymbol{w}_\nu.$$

Algorithm 8 Distributed MIP with an Auxiliary IP

Run Algorithm 5 to initialize ub, lb, $\boldsymbol{u}^*$, $\boldsymbol{\mu}^0$, $\boldsymbol{x}_\nu^0$ and $\mathcal{S}$.
Based on $\mathcal{S}$, set up the sets $\mathcal{S}_\nu$, for all $\nu \in \mathcal{P}$.
$\boldsymbol{\gamma}^0 \leftarrow \mathbf{0}$, $k \leftarrow 0$ and $r \leftarrow 0$.
while $ub > lb$ **do**
    Lower bounding:
    while some termination criteria is not met **do**
        $r \leftarrow r + 1$
        $\boldsymbol{\mu}^r \leftarrow \boldsymbol{\mu}^{r-1} + \rho_\mu^r \left(\boldsymbol{b} - \sum_{\nu \in \mathcal{P}} \boldsymbol{A}_\nu \boldsymbol{x}_\nu^{r-1} \right)$
        for $\nu := 1$ to N **do**
            solve $\min_{\boldsymbol{x}_\nu} \{\mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}^r) : \boldsymbol{x}_\nu \in X_\nu, \boldsymbol{u}_\nu \in U_\nu \backslash \mathcal{S}_\nu\}$
            Let $z_\nu(0)$ be the optimal value and $\hat{\boldsymbol{x}}_\nu^r = (\boldsymbol{u}_\nu(0), \boldsymbol{y}_\nu(0))$ be an optimal solution
            for $l := 1$ to S_ν **do**
                Update $\hat{z}_\nu(l) \leftarrow z'_\nu(\hat{\boldsymbol{x}}_\nu(l), \boldsymbol{\mu})$ by solving the LP in (75).
            end for
        end for
    end while
    while some termination criteria is not met **do**
        $k \leftarrow k + 1$
        for $\nu := 1$ to N **do**
            solve $\min_{\boldsymbol{w}_\nu} \{\mathcal{L}'''_\nu(\boldsymbol{w}_\nu, \boldsymbol{\mu}^r, \boldsymbol{\gamma}_\nu^k) : (66), (69) \text{ and } (71)\}$
            let v_ν^k be the optimal value and $\boldsymbol{w}_\nu^k$ be an optimal solution
        end for
        if $lb < \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} v_\nu^k$ **then**
            $lb \leftarrow \boldsymbol{\mu}^\top \boldsymbol{b} + \sum_{\nu \in \mathcal{P}} v_\nu^k$
        end if
        $\bar{\boldsymbol{w}}^k \leftarrow \frac{1}{|\mathcal{P}|} \sum_{\nu \in \mathcal{P}} \boldsymbol{w}_\nu^k$
        $\boldsymbol{\gamma}_\nu^k \leftarrow \boldsymbol{\gamma}_\nu^{k-1} + \rho_\gamma^k \left(\boldsymbol{w}_\nu^k - \bar{\boldsymbol{w}}^k \right)$
    end while
    for $\nu := 1$ to N **do**
        if $\sum_{\nu' \in \mathcal{P}} w_{\nu,\nu',0} \geq 1$ **then**
            $\mathcal{S}_\nu \leftarrow \mathcal{S}_\nu \cup \{\boldsymbol{u}_\nu(0)\}$
        end if
    end for
    Let $\tilde{\boldsymbol{u}}_\nu^k$ be the corresponding $\tilde{\boldsymbol{u}}_\nu \in U$ to $\boldsymbol{w}_\nu^k$
    $\hat{\mathcal{S}} \leftarrow \cup_{\nu \in \mathcal{P}} \{\tilde{\boldsymbol{u}}_\nu^k\}$.
    Upper bounding: run Algorithm 4 for set $\hat{\mathcal{S}}$ to update ub and $\boldsymbol{u}^*$.
    $\mathcal{S} \leftarrow \mathcal{S} \cup \hat{\mathcal{S}}^k$
end while

Then, we propose an extension of Algorithm 7 as Algorithm 8. These algorithms have two main differences. In the first inner while loop (lines 6-16) of Algorithm 8, values of $\boldsymbol{\mu}$ and $\hat{z}_\nu(l)$ are updated by solving the MIP problem $\min_{\boldsymbol{x}_\nu} \{\mathcal{L}_\nu(\boldsymbol{x}_\nu, \boldsymbol{\mu}^r) : \boldsymbol{x}_\nu \in X_\nu, \boldsymbol{u}_\nu \in U_\nu \backslash \mathcal{S}_\nu\}$, and LP (75), respectively. Note that as long as there is no change in value of $\boldsymbol{\mu}$ and set $\mathcal{S}_\nu$ these updates are not required. In the second inner while loop (lines 17-28), values of $\boldsymbol{\gamma}$ and lb are updated by solving the auxiliary pure 0-1 problem (78) in parallel. In this way, binary variables $\boldsymbol{w}_\nu$ and mixed variables $\boldsymbol{x}_\nu$ are decoupled for each block ν. Moreover, constraints (67) and (68) are no longer needed to couple $\boldsymbol{w}_\nu$ and $\boldsymbol{x}_\nu$.

3.6 Conclusions and future work

In this chapter, we proposed different exact and heuristic decentralized algorithms for MIPs. These algorithms were extensions of dual decomposition and ADMM. To mitigate oscillations and traps in local optimality, a modified version of ADMM (R&F) was developed as a heuristic decentralized method for MIPs. In the exact approaches, primal cuts were added to restrict the Lagrangian relaxation and improve the lower bound on the objective function of the original MIP problem.

A possible direction for future research is to blend the speed of R&F and precision of the exact methods. Another topic for future work is investigating stronger primal cuts to speed up the proposed exact methods. Moreover, the proposed methods can be improved for specific applications by exploiting the problem structures.

In Chapter 4, we decompose the unit commitment problem on a region based where data privacy is an important issue for different regions. Then, in Chapter 5, we present extensive computational experiments for solving UC instances with different decentralized approaches.

Chapter IV

DECENTRALIZED UNIT COMMITMENT

In this chapter, we present the notation and the problem statement for the unit commitment (UC) problems. Moreover, a literature review for the decentralized UC (DUC) is provided. We also present formulations for DUC which can be solved by decentralized MIP algorithms proposed in Chapter 3. Finally, we propose a solution approach for DUC which exploits the structure of UC in decentralized algorithms.

This chapter and some parts of the next chapter are based on work performed jointly with Mitch Costley. This work is described in reference [61], which Mr. Costley and Mr. Feizollahi co-authored, and it is largely reproduced here. Mr. Costley's contributions included the approach to handle system reserve constraints (96), which is described in Section 4.2.2.1, and the large-scale software systems to conduct the centralized and decentralized experiments in Section 5.2.

4.1 Introduction

Power generating companies (GENCOs) and Independent System Operators (ISOs) use some forms of the Unit Commitment (UC) problem to determine status of power plants in the day-ahead to week-ahead timeframe. The UC problem deals with the on/off decisions and output power levels of generating units in a power system over a given planning horizon. Its objective is minimizing the total system cost or maximizing system profit depending on the market design. This problem typically considers technological, economic and regulatory factors and constraints such as physical limitation of generators or reserve requirements. For most ISOs, UC problem is solved as part of the day-ahead market clearing process [164]. Other applications of UC

may involve an intra-day commitment, such as [183, 184, 190]. Further, some processes, such as hydroelectric generation scheduling, require time frames in the weeks to months [40, 64, 7].

As power grids grow significantly in size, complexity, and user requirements, new operational frameworks will be needed. One option is to integrate system operations such as economic dispatch, contingency analysis, and UC at the interconnection level, potentially leading to the formation of a system operator for the entire Eastern interconnection [57, 82]. Ostensibly, the goal of such an effort would be to ensure that coupled systems (operated by PJM, MISO, etc.) with many interfaces would be coordinated in such a way that the most reliable and economic global system state is achieved. Solving operational problems of this size in the required time scales has proven quite difficult.

A compelling alternative is to decentralize, rather than centralize, these operations more [75, 106]. System operators and software systems can manage smaller problems more effectively, allowing for more detailed modeling and more modest computational requirements. The central question then becomes how the individual neighboring areas can coordinate to optimize economy while ensuring reliability. In this chapter, we examine a new decentralized approach to the UC problem.

At the height of the deregulation movement in the United States, a literature around decentralized UC (DUC) sought to investigate the market dynamics of UC conducted by generating companies only (so-called *self-commitment*). In this framework, the ISO posts hourly energy prices calculated based on the load forecast. Generating companies then conduct UC for their assets as price-takers [198, 169]. Additional iterations may occur to search for an equilibrium where all the load is served with the minimum prices. The existence of such an equilibrium under certain conditions was established in [121, 122] and investigated empirically in [50]. In simulation results reported in [170], the cost of anarchy under self-commitment was found to

be roughly 4%. Challenges related to the nonconvexity, oscillation, high cost, and inefficiency of self-commitment were explored in [63, 171, 170, 72].

In this chapter, we do not study self-commitment as described above. The UC process presented here differs from self-commitment DUC in that it is a direct translation of the traditional centralized UC problem to subproblems corresponding to partitions of the system. Works studying optimal self-scheduling and bidding strategies under some forecast of prices, such as [169, 199, 167, 101, 104, 83], are largely concerned with determining generating company behavior in a framework where some other coordinator (e.g. an ISO) determines prices to induce desirable system-level behavior.

As a decomposition of the UC, our work is more analogous to, for example, the reliability UC (RUC) described in [89] conducted following the day-ahead market clearing process. The analyses of [198, 121, 122, 50, 170, 63, 171, 72] study the revenue adequacy of market participants and describe pricing approaches to overcome inefficiencies. In short, these works describe how pricing should occur and market participants should respond in a system without centralized UC, whereas the process described here retains the centralized UC functionality but decentralizes its solution.

Our work retains some of the advantages of self-commitment by ensuring the privacy of commercial data. Further, it differs from most of the models provided in the works mentioned above by including not only all the traditional generator temporal constraints, but also network flow constraints, which self-commitment cannot easily address. Subproblems can truly be solved in parallel with minimal requirements for information exchange in each iteration. Although it is certainly possible to implement this in a centralized scheme, it allows for entirely separate entities to coordinate their operations even without a strong centralized computational node.

To the extent that DUC is deployed in a single computational environment in a centralized framework, it can be compared to other problem decompositions that

leverage parallelizable subproblems. Some of the most important decomposition techniques for UC are Benders [110, 129, 142, 59, 194, 5], dual or Lagrangian [207, 187, 176, 177, 202, 79, 133, 62, 111], and Dantzig-Wolfe [65]. Decomposition methods such as Benders and Dantzig-Wolfe use a master-slave architecture where subproblems may be solved on separate computational nodes but are coordinated by a master problem, which then requires the results of those subproblems to solve an iteration of its own algorithm. In our approach, there is no master problem or central computational node.

Other problem decompositions are more similar to that presented here in that they have a decentralized structure without being a form of self-commitment DUC discussed above. Batut and Renaud first applied a regional decomposition approach with duplication of variables to power system problems in [13]. Kim and Baldick [95] similarly used a linearized augmented Lagrangian approach along with the auxiliary problem principle to solve optimal power flow (OPF) problems in parallel. They then showed in [94] how to extend this formulation to use several different solution algorithms, including ADMM (see Section 3.3.2 for details of ADMM algorithm). In Section III.B of [94], it is highlighted that ADMM is limited in parallel applications because of the interdependency of the two minimization problems in ADMM.

The ADMM method of solving OPF problems in a distributed manner was implemented and tested in [44], but the UC master problem after the Benders decomposition remained centralized. ADMM is also applied to solve security-constrained OPF with AC constraints in [139]. In [130], a Lagrangian method was used to solve the multi-area OPF problem with AC constraints. The solution computations were largely distributed with a central coordinator needed for some simple calculations in each iteration. An extension of [130] was presented in [75] to solve the decentralized AC power flow using neural networks to solve the nonlinear programming subproblems. Bakirtzis and Biskas [11] proposed a solution method to the DC-OPF problem

using a similar formulation to [95] around phase angle variable duplication but with a nearly fully decentralized solution approach.

Recently, [204] demonstrated a decentralized solution to the security-constrained DC-OPF using a marginal equivalent decomposition that requires exchanging shift factors and binding constraint data in each iteration. This method was proven to converge in finite iterations to the global optimum under some mild assumptions.

Finally, Li and Luh [103] provided a DUC framework using a two-level decomposition where regions solve their own optimization subproblems and communicate to coordinate marginal prices on borders. The formulation uses shift factors, requiring a centralized computation of line-injection sensitivities, and proposes that heuristic methods be used to find feasible solutions to the binary variables.

Our formulation can be implemented in a peer-to-peer framework that limits information exchange between subproblems, enabling a decentralized structure while preserving the confidentiality of data internal to the regions. Experimental results of parallel implementation of DUC is presented in Section 5.2. In this way, we improve upon the form of ADMM given in [94]. We also address the UC problem as opposed to the OPF problem addressed in [95, 11, 94, 139, 130, 204]. Our formulation differs from [13] in that we solve mixed-integer quadratic programming (MIQP) subproblems instead of finding generation schedules through dynamic programming. It differs from [103] in that we use an augmented Lagrangian formulation and a new heuristic for finding feasible binary solutions. Further, we demonstrate the performance of our algorithm on systems of over 3,000 buses.

The contributions of this work fall into two categories: decentralization of the UC solution method and computational speed gain of UC solution searches. Decentralization of UC has structural benefits in that it provides data privacy for the confidential information of generating companies. However, as will be shown in Chapter 5, our method can also find near optimal solutions to large-scale UC problems faster than

conventional approaches. The boost in computational speed has impact in at least two different ways: First, as the size of some electricity markets grow, UC solution engines that previously performed well are now experiencing new performance challenges [43]. The improved scalability to large-scale problems may help alleviate these emerging issues. Second, if good solutions to UC problems can be found more quickly than in conventional approaches, then it is possible to conduct more "rolling-horizon" UC studies throughout an operating day, adjusting commitment decisions to react to new information about forecast of wind generation or load. For example, wind forecasts are generally much more accurate an hour or two ahead of time than they are 24 hours ahead. The less time that UC takes to calculate, the closer to the dispatch time the calculation can begin, meaning the more accurate near-term forecasts can be used. This capability has implications for both reliability and cost of system operation.

The remainder of this chapter is organized as follows. In Section 4.2, we review the notation, problem statement and solution methods for classical UC problems. In Section 4.3, we present formulations for DUC which can be solved by decentralized MIP algorithms proposed in Chapter 3. In Section 4.4, we propose a solution approach for DUC. Extensive computational experiments for DUC will be presented in Chapter 5.

4.2 Classical Unit Commitment Formulation

In the context of this research, we consider day-ahead, 24-hour UC problems with the objective of minimizing total system cost over the decision variables of generator active power outputs and on/off status. Constraints mainly include technological aspects such as minimum and maximum operating levels of generators or thermal limit of transmission lines. Regulatory constraints include various forms of operating reserve and contingency constraints to protect the system against the loss of some

generation or transmission elements.

Let the power system network be a connected graph $G = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{E}$ is the set of edges. In this formulation, nodes correspond to buses and edges to branches (that is, transmission lines or transformers). Buses may have any number of generators and loads connected to them. Detailed formulations of the constraints and cost function for each generator and load can be found in [38, 162, 119, 134].

Here we recast the tight MIP formulations for the UC problem presented in [119, 134] with some slight adjustments. In our formulation:

- The power system is partitioned into regions.
- DC power flow approximation is used.
- System reserve requirements are included.
- Both cold and hot startup are considered.
- Quadratic costs for power production are approximated by piecewise linear functions.
- All demand is served.

4.2.1 Generator Cost Function and Constraints

For unit g at time t, let p_{gt}, r_{gt}^{SR}, r_{gt}^{NSR}, and r_{gt}^{OR} be continuous variables representing power output above the minimum output, spinning reserve, non-spinning reserve, and operating reserve, respectively, all in MW. Moreover, suppose u_{gt}, v_{gt}, v_{gt}^{HS}, w_{gt} are binary variables for commitment, startup, hot startup and shutdown status, respectively, of unit g at time t. Denote the vector of all decision variables related to unit g by $\boldsymbol{x}_g$. Let $\mathcal{G}$ be the set of all generators and $\mathcal{G}_i$ be the set of generators connected bus i.

Let $\overline{P}_g$, $\underline{P}_g$, RD_g, RU_g, SD_g, SU_g, T_g^{CS}, TD_g, and TU_g be maximum and minimum power outputs (MW), ramp down and up rate limits (MW/h), shutdown and startup capabilities (MW), cold startup time (h), and minimum downtime (h) and uptime (h) of unit g. Let C_g^{CS}, C_g^{HS}, C_g^{SD}, C_g^{LV}, C_g^{NL}, and C_g^{Q} be cold startup, hot startup, shutdown, linear, no-load, and quadratic costs, respectively, of unit g.

4.2.1.1 Cost function

Operational cost of thermal generating units includes:

- Fuel cost: This cost is depending on the generated power, and most often, the incremental fuel cost is used in UC. Although in reality these functions are possibly non-continuous and non-convex, approximate polynomial or piece-wise linear functions are used in optimization models to prevent trapping in local solutions [160, 161].

- Startup costs: Bringing an "off" unit into operation leads to an extra cost due to fuel used in the controlled heating of the unit and pressurization and decompression of the boilers. This reduces the effective life of the unit. Some generator start up costs can be decomposed further into a hot and cold start up costs. This is due to the fact that as these generators cool down, it becomes more expensive to start them up. Therefore, to accurately model the practical unit commitment problem, one would need to include these different start up costs and their dependency to the state of the generating unit [123].

- Shutdown costs: These costs are often much lower than the startup costs, and sometimes they are not considered in the UC, since they can be included in the startup costs.

Then, the total cost of unit g in all times in the study horizon, $C_g(\boldsymbol{x}_g)$, can be

formulated as follows.

$$C_g(\boldsymbol{x}_g) = \sum_{t=1}^{T} \big[C_g^{\mathrm{CS}} v_{gt} + (C_g^{\mathrm{HS}} - C_g^{\mathrm{CS}}) v_{gt}^{\mathrm{HS}} + C_g^{\mathrm{SD}} w_{gt} + C_g^{\mathrm{Q}} p_{gt}^2 + (C_g^{\mathrm{LV}} + C_g^{\mathrm{Q}} \underline{P}_g) p_{gt} + (C_g^{\mathrm{NL}} + C_g^{\mathrm{LV}} \underline{P}_g + C_g^{\mathrm{Q}} \underline{P}_g^2) u_{gt}\big]. \tag{79}$$

4.2.1.2 Constraints for piece-wise linear cost functions

Sometimes instead of quadratic energy cost function, convex piece-wise linear functions are used to reduce the computational burden of UC problem. In this case, the quadratic term $C_g^{\mathrm{Q}} p_{gt}^2$ in (79) is replaced by a non-negative continuous variable $\tilde{c}_{gt}$, along with the following constraints.

$$\tilde{c}_{gt} \geq a_{gk} + b_{gk} p_{gt}, \quad \forall t = 1, \cdots, T, k = 1, \cdots, K_g, \tag{80}$$

In this research, quadratic costs for power production are approximated by piece-wise linear functions with $K_g = 5$ line segments with equal length. Specifically, we assume $b_{gk} = 2C_g^{\mathrm{Q}} \times \frac{(k-0.5)(\overline{P}_g - \underline{P}_g)}{K_g}$ and $a_{gk} = -C_g^{\mathrm{Q}} \times \left(\frac{(k-0.5)(\overline{P}_g - \underline{P}_g)}{K_g}\right)^2$. Note that Similar to [119], we defined p_{gt} as power output above the minimum output. Therefore, $0 \leq p_{gt} \leq \underline{P}_g - \overline{P}_g$, for all $g \in \mathcal{G}$ and $t \in \mathcal{T}$.

4.2.1.3 Unit commitment relation with startup and shutdown variables

To facilitate startup and shutdown costs, and minimum up and down times we use startup (v_{gt}), hot startup (v_{gt}^{HS}), and shutdown (w_{gt}) variables in addition to the on/off state variables (u_{gt}). Startup and shutdown variables are constrained by

$$v_{gt}^{\mathrm{HS}} \leq \sum_{\tau=t-T_g^{\mathrm{CS}}-\mathrm{TD}_g}^{t-1} w_{gt}, \quad \forall t = T_g^{\mathrm{CS}} + \mathrm{TD}_g + 1, \cdots, T, \tag{81}$$

$$v_{gt}^{\mathrm{HS}} \leq v_{gt}, \quad \forall t = 1, \cdots, T, \tag{82}$$

$$u_{gt} - u_{g,t-1} = v_{gt} - w_{gt}, \quad \forall t = 2, \cdots, T. \tag{83}$$

Let T_g^{Init} be the number of hours unit g has been online (+) or offline (-) prior to the first period of the commitment study. For $t = 1$, constraint (83) can be

adjusted as follows. If $T_g^{\text{Init}} > 0$ then $v_{g1} = 0$ and $w_{g1} = 1 - u_{g1}$. Otherwise, $w_{g1} = 0$ and $u_{g1} = v_{g1}$. Note that $v_{gt} = 0$ implies $v_{gt}^{\text{HS}} = 0$ by (82), and there will be no associated startup cost in (79). For the case that $v_{gt} = 1$, if no shutdown happened in $[t - T_g^{\text{CS}} - \text{TD}_g, t-1]$, then $v_{gt}^{\text{HS}} = 0$ and the startup cost will be C^{CS}. Otherwise $v_{gt}^{\text{HS}} = 1$ and the startup cost will be C^{HS} by (79).

4.2.1.4 Generator output power limits

Generation limit constraint formulations depend on the generator's minimum up time requirement. If $\text{TU}_g = 1$,

$$p_{gt} \leq (\overline{P}_g - \underline{P}_g)u_{gt} - (\overline{P}_g - \text{SU}_g)v_{gt}, \quad \forall t = 1, \cdots, T \tag{84}$$

$$p_{gt} \leq (\overline{P}_g - \underline{P}_g)u_{gt} - (\overline{P}_g - \text{SD}_g)w_{g,t+1}, \ \forall t = 1, \cdots, T-1 \tag{85}$$

If $\text{TU}_g \geq 2$,

$$p_{gt} \leq (\overline{P}_g - \underline{P}_g)u_{gt} - (\overline{P}_g - \text{SU}_g)v_{gt} - (\overline{P}_g - \text{SD}_g)w_{g,t+1}, \quad \forall t = 1, \cdots, T-1. \tag{86}$$

For $t = T$, constraint (86) can be adjusted as

$$p_{gT} \leq (\overline{P}_g - \underline{P}_g)u_{gT} - (\overline{P}_g - \text{SU}_g)v_{gT}.$$

4.2.1.5 Minimum up and down time constraints

If a unit must be on for a certain number of hours before it can be shut off, then a minimum up-time is set. By contrast, minimum down-time is the number of hours a unit must be off before it can be brought on again. Minimum up and down time constraints are represented by

$$\sum_{\tau=t-\text{TU}_g+1}^{t} v_{g\tau} \leq u_{gt}, \quad \forall t = \text{TU}_g, \cdots, T, \tag{87}$$

$$\sum_{\tau=t-\text{TD}_g+1}^{t} w_{g\tau} \leq 1 - u_{gt}, \quad \forall t = \text{TD}_g, \cdots, T, \tag{88}$$

Initial minimum up and down time constraints can be enforced as follows. If $0 < T_g^{\text{Init}} < \text{TU}_g$ then $u_{gt} = 1, \quad \forall t = 1, \cdots, \text{TU}_g - T_g^{\text{Init}}$. If $-\text{TD}_g < T_g^{\text{Init}} < 0$ then $u_{gt} = 0, \quad \forall t = 1, \cdots, T_g^{\text{Init}} - \text{TD}_g$. One may also include must-run, outaged, or fixed output power constraints for certain units in certain times [161].

4.2.1.6 Ramp up and down limits

Ramping limits can be expressed as

$$-\text{RD}_g \leq p_{gt} - p_{g,t-1} \leq \text{RU}_g, \quad \forall t = 2, \cdots, T \tag{89}$$

Constraint (89) should be modified for $t = 1$ as $-\text{RD}_g \leq p_{g1} - p_g^{\text{Init}} \leq \text{RU}_g$, where p_g^{Init} is the initial power output (above minimum output) for unit g.

4.2.1.7 Reserve Constraints

To overcome unexpected events such as generator failures, some extra capacity is required in UC [56, 164, 4]. The generator-level constraints for spinning, non-spinning, and operating reserve for unit g can be modeled as follows.

$$r_{gt}^{\text{SR}} \leq \frac{\text{RU}_g}{6} u_{gt}, \quad \forall t = 1, \cdots, T, \tag{90}$$

$$r_{gt}^{\text{SR}} \leq \overline{P}_g - \left(p_{gt} + \underline{P}_g\right), \quad \forall t = 1, \cdots, T, \tag{91}$$

$$r_{gt}^{\text{NSR}} \leq \frac{\text{SU}_g}{6}(1 - u_{gt}) + \frac{\text{RU}_g}{6} u_{gt}, \quad \forall t = 1, \cdots, T, \tag{92}$$

$$r_{gt}^{\text{NSR}} \leq \overline{P}_g - \left(p_{gt} + \underline{P}_g\right) - r_{gt}^{\text{SR}}, \quad \forall t = 1, \cdots, T, \tag{93}$$

$$r_{gt}^{\text{OR}} \leq \overline{P}_g - \left(p_{gt} + \underline{P}_g\right) - r_{gt}^{\text{SR}} - r_{gt}^{\text{NSR}}, \quad \forall t = 1, \cdots, T, \tag{94}$$

4.2.1.8 Variable type

The various decision variables included above are constrained as follows:

$$p_{gt}, r_{gt}^{\text{SR}}, r_{gt}^{\text{NSR}}, r_{gt}^{\text{OR}}, \tilde{c}_{gt} \geq 0, \ u_{gt}, v_{gt}, w_{gt}, v_{gt}^{HS} \in \{0, 1\}, \ \forall t = 1, \cdots, T. \tag{95}$$

4.2.2 Power and Reserve Requirements

For time t, let $d_{i,t}$ and $\hat{r}_t^q$ be the expected load (MW) at bus i and system reserve requirement for product q (MW), respectively. These requirements should be satisfied by the power output and reserve products of generators

4.2.2.1 System requirements for reserve products

The total of each reserve product q in the system must meet the minimal system requirement. For this study, the system requirement was equal to the size of the largest contingency for the 10-minute contingency reserve and the size of the second largest contingency for the 30-minute operating reserve. Spinning reserve was required to be at least half of the 10-minute reserve. The system-level constraints for reserve products are

$$\sum_{g\in\mathcal{G}} r_{g,t}^q \geq \hat{r}_t^q, \quad \forall q \in \{\text{SR, NSR, OR}\}, \forall t = 1, \cdots, T. \tag{96}$$

4.2.2.2 Power demand and supply balance

Generation must equal demand at each time period, expressed as

$$\sum_{g\in\mathcal{G}} [\underline{P}_g u_{gt} + p_{gt}] = \sum_{i\in\mathcal{N}} d_{it}, \forall t = 1, \cdots, T. \tag{97}$$

4.2.3 Mathematical Formulations for UC

Next we present three mathematical formulations for UC which are UC without network constraints, network-constrained UC with line sensitivities and network-constrained UC with voltage phase angles.

4.2.3.1 UC without network constraints

In this formulation, the total system cost should be minimized subject to generator constraints (81)-(95) for all units, reserve requirement (96), and demand and supply

balance (97) as follows:

$$\begin{aligned} \min \quad & \sum_{g\in\mathcal{G}} C_g(\boldsymbol{x}_g) \\ \text{s.t.} \quad & (80)-(95), \quad \forall g \in \mathcal{G} \\ & (96),\ (97) \end{aligned} \tag{98}$$

In above formulation, constraints (96) and (97) are coupling different generators to each other.

4.2.3.2 Network-constrained UC with line sensitivities

For branch ij, let $\overline{F}_{ij}$ be the active power flow limit. In network-constrained UC, thermal limits are considered for branch elements. These are represented as

$$-\overline{F}_{ij} \le F_{ij,t} \le \overline{F}_{ij},\ \forall t = 1, \cdots, T. \tag{99}$$

Many UC formulations express line power flows in terms of power injections at buses by using line sensitivities as in [110]. Under the DC power flow approximation, the power flow in line ij at time t, which is $F_{ij,t}$, is assumed to be a linear function of net power injections $P^{\text{net}}_{k,t}$ in all buses $k \in \mathcal{N}$; that is,

$$F_{ij,t} = \sum_{k\in\mathcal{N}} \gamma_{ij,k} P^{\text{net}}_{k,t} \tag{100}$$

and

$$P^{\text{net}}_{k,t} := \sum_{g\in\mathcal{G}_k} [\underline{P}_g u_{gt} + p_{gt}] - d_{k,t}. \tag{101}$$

The sensitivities $\gamma_{ij,k}$ are the power transfer distribution factors (PTDFs), also referred to as generation shift factors (GSFs), which may be calculated according to [96].

In network-constrained UC using GSFs, constraints (99) for the monitored lines $ij \in \mathcal{E}' \subset \mathcal{E}$ are added to the model (98). In practice, the set of monitored lines, $\mathcal{E}'$, contains about 10% of all lines in $\mathcal{E}$. Note that there is no need to actually add

the additional variables $F_{ij,t}$ and $P^{\text{net}}_{i,t}$ and constraints (100)-(101) to the optimization model since they can be substituted by the right hand sides of (100) and (101). Then, constraints (99) can be recast as:

$$-\overline{F}_{ij} \leq \sum_{k\in\mathcal{N}} \gamma_{ij,k} \left(\sum_{g\in\mathcal{G}_k} [\underline{P}_g u_{gt} + p_{gt}] - d_{k,t} \right) \leq \overline{F}_{ij}, \quad \forall ij \in \mathcal{E}' \subset \mathcal{E}, t, \tag{102}$$

Therefore, the network-constrained UC with line sensitivities can be formulated as follows.

$$\begin{aligned} \min \quad & \sum_{g\in\mathcal{G}} C_g(\boldsymbol{x}_g) \\ \text{s.t.} \quad & (80)-(95), \quad \forall g \in \mathcal{G} \\ & (96),\ (97), (102) \end{aligned} \tag{103}$$

In above formulation, constraints (96), (97) and (102) are coupling different generators to each other.

4.2.3.3 Network-constrained UC with voltage phase angles

In this work, we forgo the use of sensitivity factors in order to attain better decomposability. Instead, bus phase angles are treated as decision variables in order to better formulate the mathematical line flow and bus power balance constraints. The output of the UC process remains the generator power output set points; bus phase angles need not be regulated set points. We seek a formulation that is as bus-centered as possible.

Let B_{ij} denotes element ij of the DC power flow Jacobian. Let $\theta_{i,t}$ be the voltage phase angle at bus i and time t. Without loss of generality, bus 1 has been designated the reference bus, giving $\theta_{1,t} = 0$, $\forall t$. Then, the DC power flow in line ij and time t can be formulated as

$$F_{ij,t} := B_{ij}(\theta_{i,t} - \theta_{j,t}). \tag{104}$$

In this case, constraints (99) can be represented as

$$-\overline{F}_{ij} \leq B_{ij}(\theta_{i,t} - \theta_{j,t}) \leq \overline{F}_{ij}, \quad \forall\ ij \in \mathcal{E}, t = 1, \cdots, T. \tag{105}$$

Bus voltage phase angles should satisfy the active power balance constraints

$$\sum_{g\in\mathcal{G}_k}[\underline{P}_g u_{gt}+p_{gt}]-d_{k,t}=\sum_{j\in\delta_i\bigcup\{i\}}B_{ij}\theta_{j,t},\quad \forall k\in\mathcal{N},t=1,\cdots,T, \tag{106}$$

where, δ_i is the set of all buses connected to bus i. Note that $B_{ii}=\sum_{j\in\delta_i}B_{ij}$, $\forall i\in\mathcal{N}$, and bus balance equations (106) imply the global demand and supply equation (97). Then, the network-constrained UC with voltage phase angles can be formulated as

$$\begin{aligned}\min\quad & \sum_{g\in\mathcal{G}}C_g(\boldsymbol{x}_g)\\ \text{s.t.}\quad & (80)-(95),\quad \forall g\in\mathcal{G},\\ & (96),(105),(106).\end{aligned} \tag{107}$$

In above formulation, constraints (96), (105) and (106) are coupling different generators to each other.

4.2.4 Other variants of UC

In addition to the UC formulations discussed in the preceding section, other variants of UC are also used in literature. These variants differ in terms of the additional constraints or modified objective functions they consider.

4.2.4.1 Emission constrained UC

Power generation from fossil fuels can pollute the air with different rates at startup, steady-state, and shut down phases. The main pollutants are SO_2, CO_2, and NO_2, and their emission depends on the consumed fuel and generated power amount. To manage and reduce air pollution, energy planners and regulators provide environmental cost estimates as a function of unit emission [74]. In some works, such as [74] and [192], emission limits are enforce by adding constraints to the UC problem. Emission constraints may limit power generation of a single unit, a group of units or the whole system. They may be given to a specific period of time, or for a certain number of periods [132]. In other works like [97], emissions are modeled as a part of objective function to be minimized.

4.2.4.2 Fuel constrained UC

In thermal units with the dominant fuel cost, managing fuel consumption becomes an important daily task. It may have some limitations such as fuel contracts, congestion in the fuel delivery system, and limited storage. Considering fuel management makes the problem rather complex, modelers make some assumptions to dramatically reduce the level of complexity. Researchers in [10, 99, 100, 186, 179, 180, 182] provide some assumptions and formulation for fuel constrained unit commitment.

4.2.4.3 UC with combined cycle plants

Because of the advantages like high efficiency, fast response, environmental friendliness, flexibility, and shorter installation time, combined cycle units have become popular in recent decades [107]. To model combined cycle units, instead of two "on" and "off" states for the generator, several states can be considered. In [107], authors modeled the startup cost as an exponential function of the time during which unit was "off". In [22, 46, 108], the authors used multiple configurations for combined cycle units. Their common point is that they all try to approximate the non-convex cost functions.

4.2.4.4 UC with other devices

In recent literature, scheduling methods have been introduced for thermal and energy storage system (ESS) unit commitment [162]. Pumped-hydro storage [102, 5] and various types of flexible loads [138] are available in the literature.

4.2.4.5 Integrating renewables in UC

Recently, new UC models were developed to include renewable [12, 189] or vehicle-to-grid idea [156]. UC models with wind power are considered in [137, 184].

4.2.4.6 UC under uncertainty

In the security constrained unit commitment (SCUC) (see e.g. [80, 168, 120, 70, 65, 51, 189, 21, 59, 136, 6, 163, 158, 191, 20]) different failure scenarios are considered. To address uncertainties arising from volatile demand, unexpected failures, or variable generations from renewable resources, robust unit commitment [118, 203, 90, 205, 191, 20] and stochastic unit commitment [37, 178, 36, 176, 179, 135, 166, 131, 12, 196, 39, 48, 154, 153, 49, 197] were developed and studied.

In other variants of UC problem, more complexities such as AC power flow relations [164, 64, 110, 195, 66, 105] and frequency regulation constraints [145] are included. Moreover, in price-based unit commitment (PBUC), generating companies try to maximize their profit based on forecasted prices [37, 146] or prices issued from a market [104].

4.2.5 Solution Methods

There is a rich literature of exact and heuristic solution methods for different variants of UC problem. Some of exact methods are MIP [38, 41, 56, 79, 81, 143, 119, 134, 162, 174], Branch and Bound [98, 163], dynamic programming [173], Benders decomposition [110, 194] and primal and dual method [55]. Besides, some heuristics and approximate methods used Lagrangian relaxation [207, 187, 176, 177, 202, 79, 133, 111], genetic algorithm [54], harmony search algorithm [2], tabu search [113], simulated annealing [206], fuzzy dynamic programming [175], particle swarm [68], memetic algorithm [185], artificial neural networks [159], ant colony search algorithm [172], and combination of above methods [86].

4.3 Decentralized Unit Commitment Formulation

Recalling from Section 4.2, the power system network is considered a connected graph with $\mathcal{N}$ as the set of nodes (buses) and $\mathcal{E}$ as the set of edges (branches). Suppose

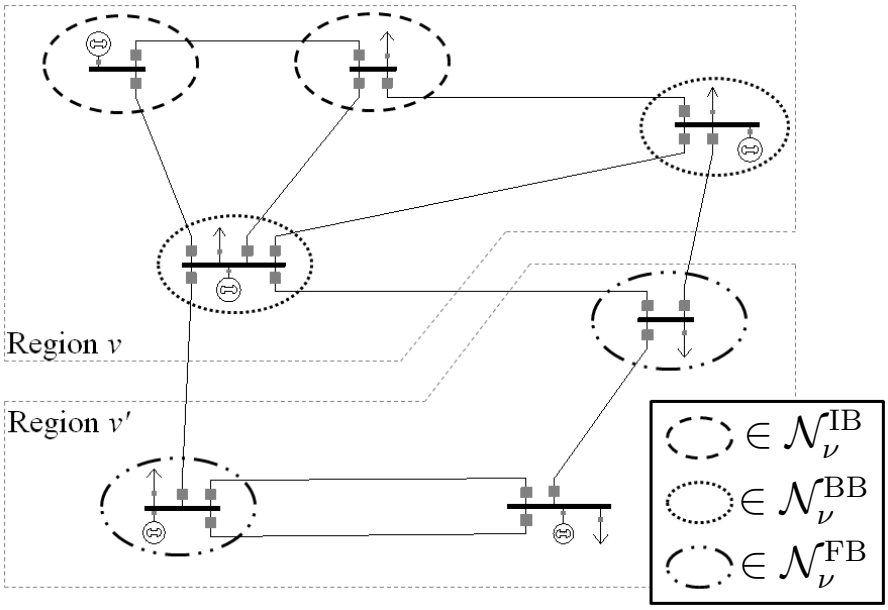

Figure 5: Regions ν and ν' outlined with bus sets of region ν.

the power system is partitioned into n exclusive regions with respect to buses, that is $\mathcal{N} = \bigcup_{\nu \in \mathcal{P}} \mathcal{N}_\nu$ and $\mathcal{N}_\nu \cap \mathcal{N}_{\nu'} = \emptyset$ for all $\nu, \nu' \in \mathcal{P}$, $\nu \neq \nu'$, with $\mathcal{P}$ being the set of regions. For a region ν denote the sets of its all, internal, and boundary buses by $\mathcal{N}_\nu$, $\mathcal{N}_\nu^{\mathrm{IB}}$ and $\mathcal{N}_\nu^{\mathrm{BB}}$, respectively. Moreover, let $\mathcal{N}_\nu^{\mathrm{FB}}$ be the set of boundary buses of the other regions connected to ν. Note that $j \in \mathcal{N}_\nu^{\mathrm{FB}}$ implies $j \in \mathcal{N}_{\nu'}^{\mathrm{BB}}$ for some region ν', which is a neighbor of ν. See Figure 5 for an illustration of how these three sets relate to the buses of region ν. Let δ_i and Δ_i be the sets of all buses and regions, respectively, connected to bus i.

In GSF network-constrained UC (103) or UC without network constraints (98), the constraints (96), (97), and (102) are globally coupled between different regions, i.e., all of the regions participate in each of these constraints. We intend to decompose the centralized UC problem (107) into sub-areas, which we will call *regions* or *prosumers* in this research. Because each generator belongs to exactly one region, all of the constraints and variables are directly decomposable to regions, and in fact buses, except for the following:

- Capacity constraints (105) on branches bridging regional boundaries (denoted as *boundary lines*).

- Bus power balance constraints (106) on buses connected by boundary lines (denoted as *boundary buses*).

Specifically, boundary lines and buses are coupling between only the neighboring regions. The reserve constraint (96) is globally coupled between all regions. Next, we propose a reformulation of model (107) that is appropriate to use in our decentralized approach.

In model (107), $\theta_{i,t}$ is shared between different regions if i is a boundary bus. To facilitate distributing the UC model among regions, we will assume each region ν connected to boundary bus i has the perception $\tilde{\theta}_{\nu,i,t}$ of the voltage phase angle at bus i and time t. This formulation is similar to that used by Kim and Baldick [95, 94] and the one used by Bakirtzis and Biskas [11], both of which duplicated some variables associated with buses in adjacent regions in order to seek convergence between the regions. However, the formulation presented here differs from the those of [95, 94] by not requiring any dummy buses – all buses belong to some region in the system. Also, this formulation differs from the one in [11] since we use an augmented Lagrangian with different objective terms. Reference [85] similarly uses a variable perception and duplication strategy to formulate a frequency control problem and applies ADMM to solve it in a distributed way.

There are three possibilities for constraints (105) of line ij. If $i, j \in \mathcal{N}_\nu^{\text{IB}}$, then $i, j \notin \mathcal{N}_\nu^{\text{BB}}$ and constraints (105) can be used as-is internal to region ν. For the cases with one internal and one boundary bus as end points of line ij, without loss of generality, let us assume $i \in \mathcal{N}_\nu^{\text{IB}}$ and $j \in \mathcal{N}_\nu^{\text{BB}}$. Then,

$$-\overline{F}_{ij} \leq B_{ij}(\theta_{i,t} - \tilde{\theta}_{\nu,j,t}) \leq \overline{F}_{ij} \quad \forall t. \tag{108}$$

For the cases with one boundary and one foreign bus as end points of line ij, without

loss of generality, let us assume $i \in \mathcal{N}_\nu^{\mathrm{BB}}$ and $j \in \mathcal{N}_\nu^{\mathrm{FB}}$. Then,

$$-\overline{F}_{ij} \leq B_{ij}(\tilde{\theta}_{\nu,i,t} - \tilde{\theta}_{\nu,j,t}) \leq \overline{F}_{ij} \quad \forall t. \tag{109}$$

The power balance constraint (106) for bus $i \in \mathcal{N}_\nu$ can be rewritten as

$$\sum_{g \in \mathcal{G}_k} [\underline{P}_g u_{gt} + p_{gt}] - d_{k,t} = \sum_{\substack{j \in \delta_i \bigcup \{i\} \\ \text{s.t. } j \in \mathcal{N}_\nu^{\mathrm{IB}}}} B_{ij}\theta_{j,t} + \sum_{\substack{j \in \delta_i \bigcup \{i\} \\ \text{s.t. } j \in \mathcal{N}_\nu^{\mathrm{BB}} \bigcup \mathcal{N}_\nu^{\mathrm{FB}}}} B_{ij}\tilde{\theta}_{\nu,j,t}, \ \forall t. \tag{110}$$

To link the actual phase angles with the perceptions of those phase angles, we need the additional constraints

$$\begin{aligned} \tilde{\theta}_{\nu,i,t} = \tilde{\theta}_{\nu',i,t}, \quad & \forall \nu, \nu' \in \mathcal{P}, \ \forall t \\ \text{s.t.} \quad & i \in \mathcal{N}_\nu^{\mathrm{BB}} \cap \mathcal{N}_{\nu'}^{\mathrm{FB}} \text{ or } i \in \mathcal{N}_\nu^{\mathrm{FB}} \cap \mathcal{N}_{\nu'}^{\mathrm{BB}}. \end{aligned} \tag{111}$$

With these, an equivalent expression of the problem (107) is

$$\begin{aligned} \min \quad & \sum_{\nu \in \mathcal{P}} C_\nu(\boldsymbol{x}_\nu) \\ \text{s.t.} \quad & \boldsymbol{x}_\nu \in X_\nu, \ \forall \nu \in \mathcal{P}, \\ & (96), \ (111), \end{aligned} \tag{112}$$

where $\boldsymbol{x}_\nu$ is the vector of all decision variables in region ν, including all variables of generators in region ν, variables $\theta_{i,t}$ for internal buses $i \in \mathcal{N}_\nu^{\mathrm{IB}}$, and $\tilde{\theta}_{\nu,i,t}$ of all buses $i \in \mathcal{N}_\nu^{\mathrm{BB}} \cup \mathcal{N}_\nu^{\mathrm{FB}}$, and

$$C_\nu(\boldsymbol{x}_\nu) = \sum_{i \in \mathcal{N}_\nu} \sum_{g \in \mathcal{G}_i} C_g(\boldsymbol{x}_g). \tag{113}$$

The set of feasible solutions, X_ν, which is a linear mixed integer set is defined as follows.

$$X_\nu := \left\{ \boldsymbol{x}_\nu : \left\{ \begin{array}{l} (80) - (95), \quad \forall g \in \mathcal{G}_i, \ i \in \mathcal{N}_\nu, \\ (110), \ \forall i \in \mathcal{N}_\nu \\ (105), \ \forall ij \in \left\{ kl \in \mathcal{E} : k, l \in \mathcal{N}_\nu^{\mathrm{IB}} \right\} \\ (108), \forall ij \in \left\{ kl \in \mathcal{E} : k \in \mathcal{N}_\nu^{\mathrm{IB}}, l \in \mathcal{N}_\nu^{\mathrm{BB}} \right\} \\ (109), \forall ij \in \left\{ kl \in \mathcal{E} : k \in \mathcal{N}_\nu^{\mathrm{BB}}, l \in \mathcal{N}_\nu^{\mathrm{FB}} \right\} \end{array} \right\}. \right. \tag{114}$$

Note that in model (112) all of the variable and constraints are local except constraints (96) and (111). By relaxing these constraints, augmenting them in the objective function through use of ADMM [28], we propose an iterative method to solve (112) in a decentralized framework.

4.4 Solution Approach for DUC

In this section, we propose an ADMM based approach to solve UC model (112) in a decentralized fashion.

4.4.1 Application of ADMM

Recall consensus ADMM, Algorithm 53, from Section 3.3.2. At iteration m, define $\bar{\tilde{\theta}}^m_{\nu,i,t}$, $\bar{\tilde{r}}^{q,m}_{\nu,t}$ and $\bar{\tilde{F}}^m_{\nu,ij,t}$ as follows:

$$\bar{\tilde{\theta}}^m_{\nu,i,t} := \frac{\sum\limits_{\nu' \in \Delta_i} \tilde{\theta}^m_{\nu',i,t}}{|\Delta_i|} \quad \forall i \in \mathcal{N}^{\text{BB}}_\nu \cup \mathcal{N}^{\text{FB}}_\nu, \forall t, \tag{115}$$

$$\bar{\tilde{r}}^{q,m}_{\nu,t} := r^{q,m}_{\nu,t} + \frac{\hat{r}^q_t - \sum\limits_{\nu' \in \mathcal{P}} r^{q,m}_{\nu',t}}{|\mathcal{P}|} \quad \forall \nu \in \mathcal{P}, \forall q, \forall t, \tag{116}$$

$$\bar{\tilde{F}}^m_{\nu,ij,t} := \frac{\tilde{F}^m_{\nu,ij,t} - \tilde{F}^m_{\nu',ji,t}}{2} = \frac{B_{ij}\left(\tilde{\theta}^m_{\nu,i,t} - \tilde{\theta}^m_{\nu,j,t}\right) - B_{ij}\left(\tilde{\theta}^m_{\nu',j,t} - \tilde{\theta}^m_{\nu',i,t}\right)}{2}, \tag{117}$$
$$\forall \nu \in \mathcal{P}, \forall t, \forall ij : (i \in \mathcal{N}^{\text{BB}}_\nu, j \in \mathcal{N}^{\text{FB}}_\nu) \text{ or } (j \in \mathcal{N}^{\text{BB}}_\nu, i \in \mathcal{N}^{\text{FB}}_\nu).$$

Here, $|.|$ is the cardinality of a set, and Δ_i is the set of regions (including region ν) connected to bus $i \in \mathcal{N}^{\text{BB}}_\nu$. Moreover, let $\lambda^m_{\nu,i,t}$, $\omega^{q,m}_{\nu,t}$ and $\mu^m_{\nu,ij,t}$ be the dual variables corresponding to deviations of $\tilde{\theta}_{\nu,i,t}$ from $\bar{\tilde{\theta}}^m_{\nu,i,t}$, $r^q_{\nu,t}$ from $\bar{\tilde{r}}^{q,m}_{\nu,t}$, and $\tilde{F}_{\nu,ij,t}$ from $\bar{\tilde{F}}^m_{\nu,ij,t}$, respectively. For the sake of simplicity, let $\tilde{\boldsymbol{\theta}}_\nu$, $\bar{\tilde{\boldsymbol{\theta}}}^m_\nu$, $\boldsymbol{r}_\nu$, $\bar{\tilde{\boldsymbol{r}}}^m_\nu$, $\tilde{\boldsymbol{F}}_\nu$, $\bar{\tilde{\boldsymbol{F}}}^m_\nu$, $\boldsymbol{\lambda}^m_\nu$, $\boldsymbol{\omega}^m_\nu$ and $\boldsymbol{\mu}^m_\nu$ be the vectors of $\tilde{\theta}_{\nu,i,t}$, $\bar{\tilde{\theta}}^m_{\nu,i,t}$, $r^{q,m}_{\nu,t}$, $\bar{\tilde{r}}^{q,m}_{\nu,t}$, $\tilde{F}_{\nu,ij,t}$, $\bar{\tilde{F}}^m_{\nu,ij,t}$, $\lambda^m_{\nu,i,t}$, $\omega^{q,m}_{\nu,t}$ and $\mu^m_{\nu,ij,t}$ respectively. Now, let the augmented Lagrangian function of region ν at iteration $m+1$ be

$$\begin{aligned}\mathcal{L}^+_{\rho,\nu}\left(\boldsymbol{x}_\nu, \bar{\tilde{\boldsymbol{\theta}}}^m_\nu, \bar{\tilde{\boldsymbol{F}}}^m_\nu, \bar{\tilde{\boldsymbol{r}}}^m_\nu, \boldsymbol{\lambda}^m_\nu, \boldsymbol{\mu}^m_\nu, \boldsymbol{\omega}^m_\nu\right) &= C_\nu(\boldsymbol{x}_\nu) + \boldsymbol{\lambda}^m_\nu\left(\tilde{\boldsymbol{\theta}}_\nu - \bar{\tilde{\boldsymbol{\theta}}}^m_\nu\right) + \frac{\rho}{2}\left\|\tilde{\boldsymbol{\theta}}_\nu - \bar{\tilde{\boldsymbol{\theta}}}^m_\nu\right\|_2^2 \\ &+ \boldsymbol{\mu}^m_\nu\left(\tilde{\boldsymbol{F}}_\nu - \bar{\tilde{\boldsymbol{F}}}^m_\nu\right) + \frac{\rho}{2}\left\|\tilde{\boldsymbol{F}}_\nu - \bar{\tilde{\boldsymbol{F}}}^m_\nu\right\|_2^2 + \boldsymbol{\omega}^m_\nu\left(\boldsymbol{r}_\nu - \bar{\tilde{\boldsymbol{r}}}^m_\nu\right) + \frac{\rho}{2}\|\boldsymbol{r}_\nu - \bar{\tilde{\boldsymbol{r}}}^m_\nu\|_2^2,\end{aligned} \tag{118}$$

where $\rho > 0$ is a given penalty factor. The terms related to $\tilde{\boldsymbol{F}}_\nu - \bar{\bar{\boldsymbol{F}}}^m_\nu$ in (118) are not required for the augmented Lagrangian relaxation of problem (112). But, having these terms speeds up convergence of the proposed ADMM based methods specially for large scale DUC instances.

The dual variables are updated as

$$\begin{aligned}
\boldsymbol{\lambda}^m_\nu &= \boldsymbol{\lambda}^{m-1}_\nu + \rho\left(\tilde{\boldsymbol{\theta}}^m_\nu - \bar{\bar{\boldsymbol{\theta}}}^m_\nu\right), \\
\boldsymbol{\mu}^m_\nu &= \boldsymbol{\mu}^{m-1}_\nu + \rho\left(\tilde{\boldsymbol{F}}^m_\nu - \bar{\bar{\boldsymbol{F}}}^m_\nu\right), \\
\boldsymbol{\omega}^m_\nu &= \boldsymbol{\omega}^{m-1}_\nu + \rho\left(\boldsymbol{r}^m_\nu - \bar{\bar{\boldsymbol{r}}}^m_\nu\right).
\end{aligned} \tag{119}$$

The primal residuals corresponding to $\tilde{\theta}^m_{\nu,i,t}$, $r^{q,m}_{\nu,t}$ and $\tilde{F}^m_{\nu,ij,t}$ at iteration m are $(\tilde{\theta}^m_{\nu,i,t} - \bar{\bar{\theta}}^m_{\nu,i,t})$, $(r^{q,m}_{\nu,t} - \bar{\bar{r}}^{q,m}_{\nu,t})$ and $(\tilde{F}^m_{\nu,ij,t} - \bar{\bar{F}}^m_{\nu,ij,t})$, respectively. The dual residuals corresponding to $\bar{\bar{\theta}}^m_{\nu,i,t}$, $\bar{\bar{r}}^{q,m}_{\nu,t}$ and $\bar{\bar{F}}^m_{\nu,ij,t}$ at iteration m are $\rho(\bar{\bar{\theta}}^m_{\nu,i,t} - \bar{\bar{\theta}}^{m-1}_{\nu,i,t})$, $\rho(\bar{\bar{r}}^{q,m}_{\nu,t} - \bar{\bar{r}}^{q,m-1}_{\nu,t})$ and $\rho(\bar{\bar{F}}^m_{\nu,ij,t} - \bar{\bar{F}}^{m-1}_{\nu,ij,t})$, respectively. Let $\boldsymbol{\alpha}^m$ and $\boldsymbol{\beta}^m$ be the vectors of all primal and dual residuals, respectively, at iteration m. We propose the basic DUC method as Algorithm 9.

4.4.2 Challenges and Remedies

For feasible convex optimization problems with finite optimal values, and convex closed feasible regions for subproblems, ADMM has nice convergence properties. In this case, primal and dual residuals converge to zero. Moreover, dual variables and objective value converge to their optimal values [28]. Convexity may be an appropriate assumption for the traditional economic dispatch, where all variables are continuous and cost curves are generally convex (if not in reality, then due to market rules). Similarly, LP relaxation of UC is a convex optimization problem.

If there are binary variables present, as there are in the UC problem, a convexity assumption is inappropriate. Therefore, although economic dispatch or LP relaxation of UC would have nice ADMM convergence properties, such properties are not expected for UC. In direct application of ADMM for non-convex optimization problems, there is no guarantee to converge to the global optimal objective value. In fact, there

Algorithm 9 Basic DUC algorithm based on ADMM

1: Initialize $\boldsymbol{x}_\nu^0$, $\boldsymbol{\lambda}_\nu^0$, $\boldsymbol{\omega}_\nu^0$ and $\boldsymbol{\mu}_\nu^0$ for each region $\nu \in \mathcal{P}$; $m \leftarrow 0$.
2: **for** $\nu \in \mathcal{P}$ **do**
3: **for** $i \in \mathcal{N}_\nu^{\mathrm{BB}}$ **do**
4: **for** $\nu' \in \Delta_i \backslash \{\nu\}$ **do**
5: **for** $j \in \delta_i \cap \mathcal{N}_{\nu'}^{\mathrm{BB}}$ **do**
6: Region ν sends $\tilde{\boldsymbol{\theta}}_{\nu,i}^m$ and $\tilde{\boldsymbol{\theta}}_{\nu,j}^m$ to its neighbor ν'.
7: **end for**
8: **end for**
9: **end for**
10: Region ν sends $\boldsymbol{r}_\nu^m$ to a designated region ν^*.
11: **end for**
12: Region ν^* calculates $\bar{\bar{\boldsymbol{r}}}_\nu^m$, the new reserve targets for each region, by (116).
13: **for** $\nu \in \mathcal{P}$ **do**
14: Region ν^* sends $\bar{\bar{\boldsymbol{r}}}_\nu^m$ to region ν.
15: **for** $i \in \mathcal{N}_\nu^{\mathrm{BB}}$ **do**
16: Region ν computes $\bar{\bar{\boldsymbol{\theta}}}_{\nu,i}^m$, using (115) .
17: **for** $\nu' \in \Delta_i \backslash \{\nu\}$ **do**
18: Region ν sends $\bar{\bar{\boldsymbol{\theta}}}_{\nu,i,t}^m$ to regions ν'.
19: **end for**
20: **for** $j \in \delta_i \cap \mathcal{N}_\nu^{\mathrm{FB}}$ **do**
21: Region ν updates $\bar{\bar{\boldsymbol{F}}}_{\nu,ij}^m$ from (117).
22: **end for**
23: **end for**
24: **end for**
25: **for** $\nu \in \mathcal{P}$ **do**
26: Region ν updates its primal and dual residuals, and dual variables.
27: **end for**
28: **if** $m > 0$, $\|\boldsymbol{\alpha}^m\| \leq \epsilon^{\mathrm{Pri}}$ and $\|\boldsymbol{\beta}^m\| \leq \epsilon^{\mathrm{Dual}}$ **then**
29: Stop and output $\boldsymbol{x}_\nu^m$ as optimal decision for each region ν.
30: **end if**
31: **for** $\nu \in \mathcal{P}$ **do**
32: Region ν solves $\min_{\boldsymbol{x}_\nu \in S_\nu} \ \mathcal{L}_{\rho,\nu}^+ \left(\boldsymbol{x}_\nu, \bar{\bar{\boldsymbol{\theta}}}_\nu^m, \bar{\bar{\boldsymbol{F}}}_\nu^m, \bar{\bar{\boldsymbol{r}}}_\nu^m, \boldsymbol{\lambda}_\nu^m, \boldsymbol{\mu}_\nu^m, \boldsymbol{\omega}_\nu^m \right)$ to updates $\boldsymbol{x}_\nu^{m+1}$.
33: **end for**
34: Update $m \leftarrow m + 1$ and go to Line 2.

is no guarantee to converge and the method may oscillate for ever. Experimental results in Section 5.1.1 confirm these issues.

Decentralized MIP algorithms proposed in Chapter 3 provide some remedies for these challenges. Exact decentralized MIP procedures can be implemented to obtain the optimal solution of a small UC instance. For large scale UC instances, exact decentralized methods are inefficient while the heuristic Release-and-Fix (R&F) approach demonstrates remarkable performance. For details of these algorithms we refer the reader to Chapter 3. Extensive computational experiments will be presented in Chapter 5.

4.4.3 Improvements of R&F for DUC

The performance of R&F algorithm can be significantly improved by implementing some of the following:

4.4.3.1 Phase angle perception simplifications

Suppose $i, j \in \mathcal{N}_\nu^{\text{BB}}$, $k \in \mathcal{N}_\nu^{\text{FB}} \cap \mathcal{N}_{\nu'}^{\text{BB}}$, $ik \in \mathcal{E}$, and $jk \in \mathcal{E}$. That is, the boundary buses i and j in region ν connected to k in region ν'. One approach to reformulate problem (107) is to consider separate perceptions $\tilde{\theta}_{i,k}$ and $\tilde{\theta}_{j,k}$ of the phase angle θ_k for each of the buses i and j, respectively. Through a simple exchange of self-identified bus names, it is possible for each region to determine whether it is multiply connected to the same boundary bus in a neighboring region. We can therefore add the equality $\tilde{\theta}_{i,k} = \tilde{\theta}_{j,k}$ to the formulation to reduce the search space of the region ν subproblem and enhance convergence. Or simply consider perception $\tilde{\theta}_{\nu,k}$ of phase angle θ_k for region ν , instead of $\tilde{\theta}_{i,k}$ and $\tilde{\theta}_{j,k}$ as in formulation (112).

4.4.3.2 Problem reformulation

As mentioned earlier, the terms related to $\tilde{\boldsymbol{F}}_\nu - \bar{\bar{\boldsymbol{F}}}_\nu^m$ are not required for the augmented Lagrangian (118) of problem (112). Without including these terms, we observed that

after thousands of iterations, primal residuals ($\tilde{\boldsymbol{\theta}}_\nu - \bar{\bar{\boldsymbol{\theta}}}^m_\nu$) are very small, but a considerable gap between power demand and supply remains. In preliminary experiments, we observed that large susceptance values may cause this issue. For example, in the IEEE case with 2736 buses, for connected buses i and j, B_{ij} in susceptance matrix varies from 2.2 to 15873. In fact, preliminary evaluation of residual convergence behavior revealed that large susceptance values B_{ij} resulted in poor solution quality. Because the power flowing across a line ij is $F_{ij,t} = B_{ij}(\theta_{i,t} - \theta_{j,t})$, small absolute differences in phase angles may result in nontrivial disagreements between regions on boundary line power flows. To counter this problem, we include additional penalty terms related to $\tilde{\boldsymbol{F}}_\nu - \bar{\bar{\boldsymbol{F}}}^m_\nu$ in augmented Lagrangian (118) for the actual power flows on boundary lines.

4.4.3.3 Smart network partitioning

In some cases, partitioning the network into regions or prosumers depends on decision maker. In these cases, smart partitioning can result in dramatic speed up in DUC. For this purpose, we designed a heuristic partitioning algorithm to pursue two objectives. The first is minimizing the total number of boundary lines. Although this problem is NP-hard in general, the heuristic method performed adequately. While constructing each region, the buses with the most connections were annexed first to limit the number of edges on the partition boundary. This objective helps to reduce the number of quadratic penalty terms in the augmented Lagrangian of each subproblem. The second objective is to equalize the number of generators in each region, since generators are the main contributor to problem complexity. Subproblems are therefore expected to be of roughly similar sizes. Cases are easy to imagine in which one region is significantly larger than another in terms of generators or problem size, especially if regions are strictly considered to be utilities, generating companies, or transmission operators. However, in these cases, a decomposition could be performed

internal to those large regions to achieve better overall congruity.

4.4.3.4 Stopping criteria

Adjusting stopping criteria depending on the problem structure can facilitate a balance of solution speed and feasibility. For problems with many boundary lines, the magnitudes of the primal and dual residual vectors will naturally be larger than problems with fewer boundary lines. By scaling the total residuals against the totals of the variables they are associated with, more effective stopping criteria can be defined.

Additionally, distinct values of ϵ can be used for the various solution processes of R&F in 3.4. ADMM-Release benefits from higher residual tolerances so that new binary solutions can be explored, while ADMM-Fix requires lower tolerances so that the binary-feasible solutions are well-refined.

Chapter V

EXPERIMENTAL RESULTS

In this chapter, we present extensive computational experiments for solving UC instances with different decentralized approaches. In Section 5.1, we present preliminary results of direct application of ADMM and R&F algorithm to solve UC instances where the subproblems were solved sequentially. We also discuss the challenges arising from nonconvexity of UC problem. In Section 5.2, we demonstrate remarkable performance of parallel implementation of R&F heuristic algorithm to solve large scale UC instances. Finally, we present numerical results for exact decentralized MIP algorithms to solve small UC instances in Section 5.3.

5.1 Preliminary Results

In this section, we present preliminary results of applying ADMM based decentralized methods to solve different UC instances. First, we discuss the results from direct application of ADMM to solve UC and its LP relaxation. Then, we demonstrate the effect of power network configuration on convergence of the proposed methods. Moreover, we present numerical results of the primitive and improved versions of R&F algorithms. We also compare our method with self commitment approach. Finally, we discuss the effect of partitioning approach and communication losses on the performance of R&F.

In the experiments for this section, the subproblems were solved sequentially, but the hypothetical parallel solution time based on the maximum computation time of subproblems in each iteration is reported. The algorithms were coded in C++. IBM ILOG CPLEX version 12.4 was used for solving QP and MIQP subproblems. Experiments were conducted on a UNIX machine with four cores rated at 2.27 GHz

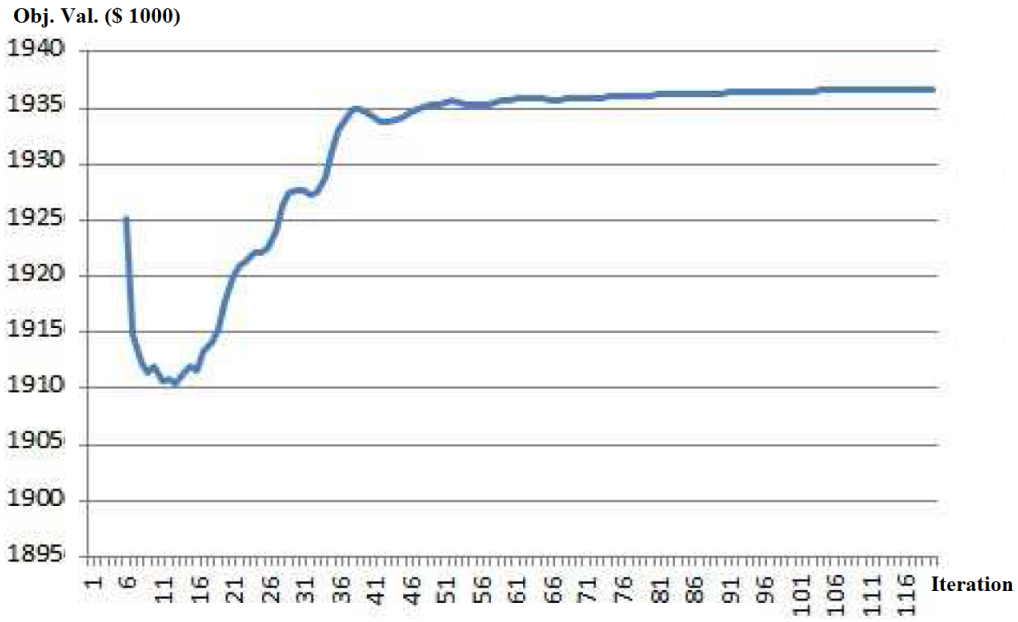

Figure 6: Objective value in ADMM for LP relaxation of a UC instance.

with 24 GB RAM total.

5.1.1 Direct Application of ADMM

For convex optimization problems, ADMM has nice convergence properties. However, directly applying classic ADMM to the DUC problem results in vulnerability to binary variable oscillations and being trapped in local optima. UC is a nonconvex MIP problem, while its LP relaxation is a convex optimization problem. Figures 6 and 7 show the convergence of objective value and total primal residuals, respectively, in ADMM for LP relaxation of a small UC instance with 5 generators.

The objective value and total primal residual in direct application of ADMM to solve the previous UC instance are depicted in Figures 8 and 9. Oscillation and nonconverging behavior of this approach are clear in Figures 8 and 9. In Figure 10, number of swaps in binary variables is presented. It is obvious that binary variables, on/off state of generators, are changing in a continuing base. Therefore, objective value and primal residuals are oscillating.

Figure 11 demonstrates the behavior of two cases of the 3,012-bus system (the

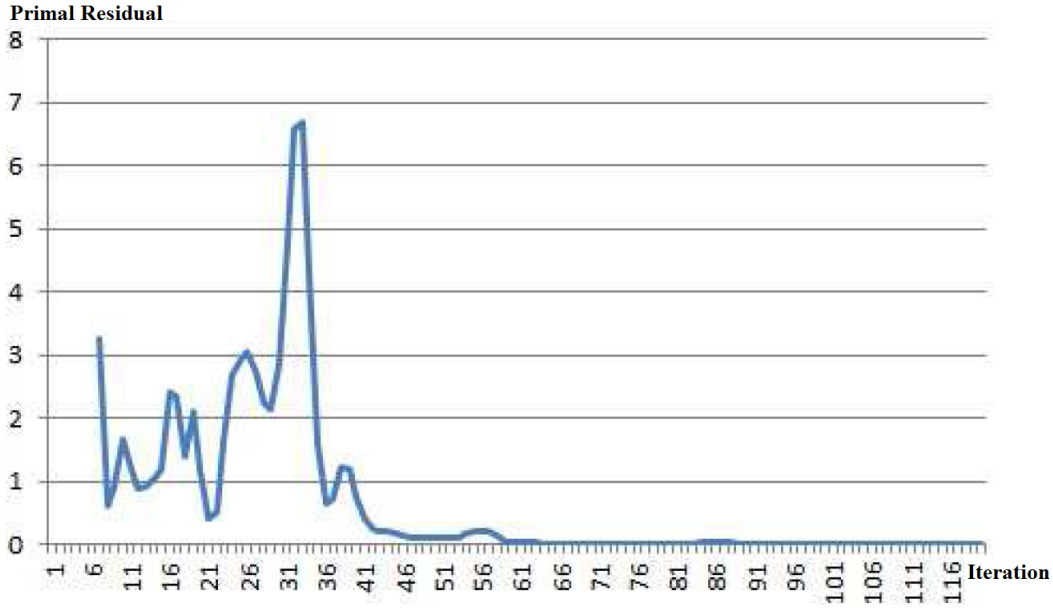

Figure 7: Primal residual in ADMM for LP relaxation of a UC instance.

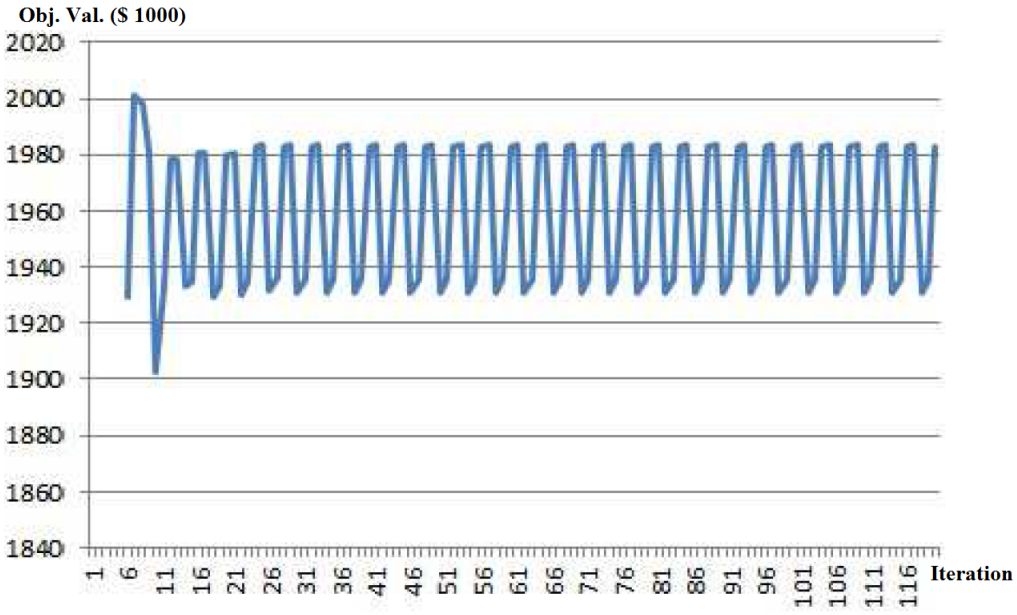

Figure 8: Objective value in direct application of ADMM for a UC instance.

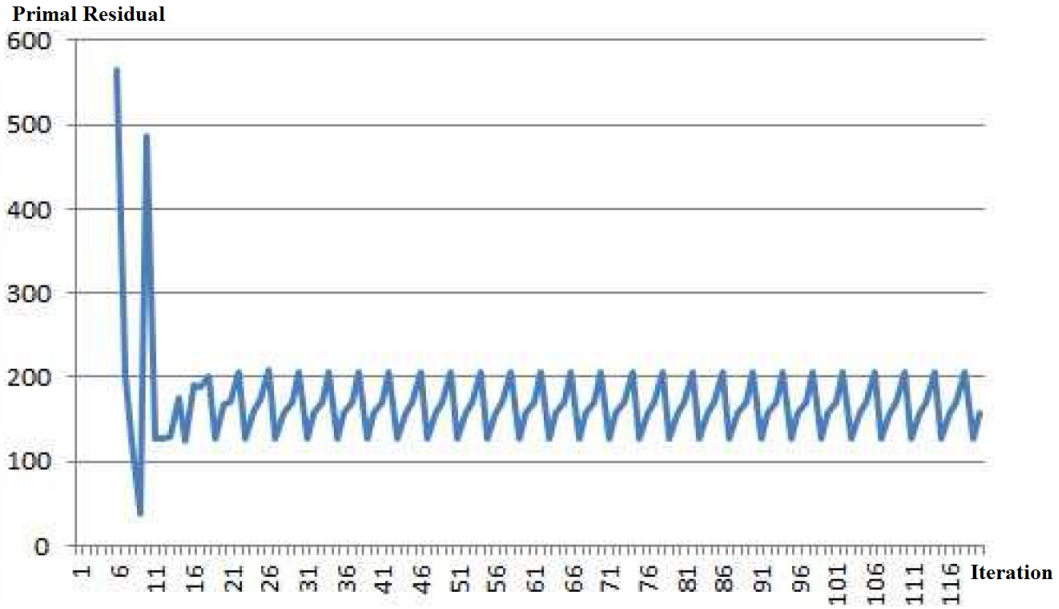

Figure 9: Primal residual in direct application of ADMM for a UC instance.

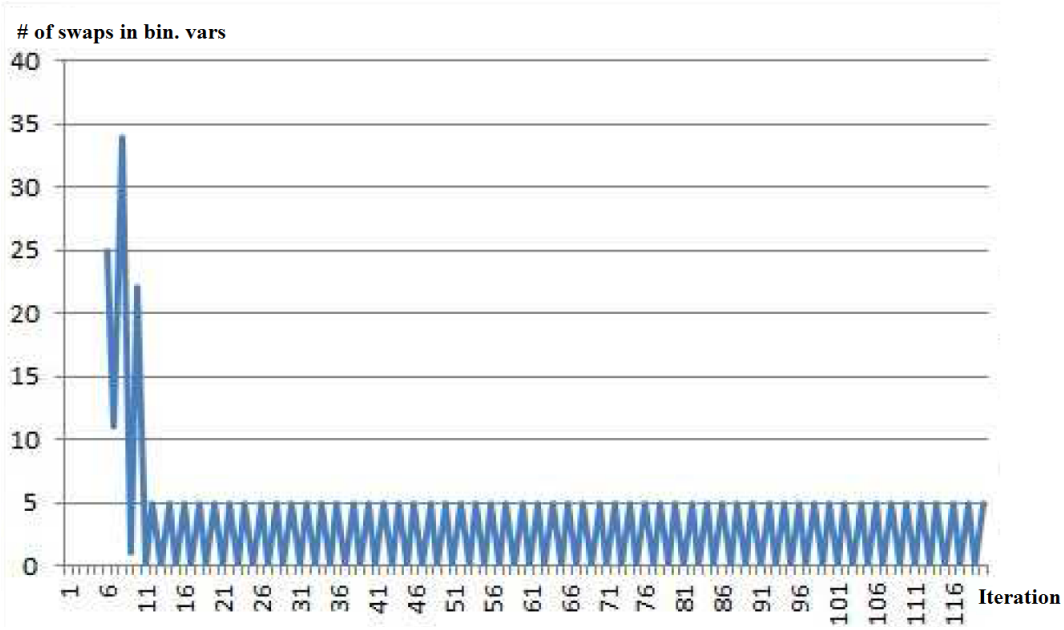

Figure 10: Number of swaps in binary variables in direct application of ADMM for a UC instance.

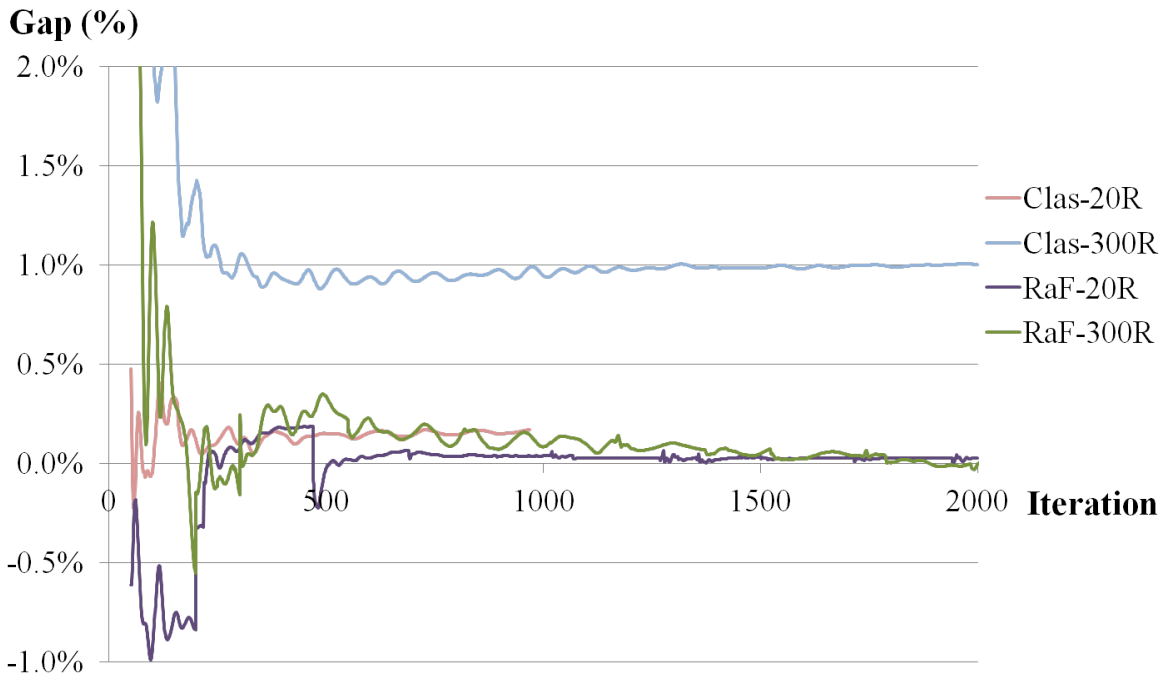

Figure 11: Optimality gaps for 3,012-bus case with classic ADMM and R&F with 20-region and 300-region decompositions.

20-region and 300-region decompositions) under classic ADMM and the improved R&F algorithm. With the 300-region decomposition, classic ADMM converges to a solution with 1% optimality while R&F approaches the global optimum much more closely. The 20-region decomposition under R&F provided similar results in terms of optimality, but the same case with classic ADMM terminated before 1,000 iterations because of the computational time limit of four hours. Note that in the classic ADMM for UC, subproblems are always MIQP and computationally hard to solve. This result highlights the advantages of both more granular decompositions and iteratively fixing and releasing binary variables in R&F.

5.1.2 Network Topology Effect

Convergence rate of ADMM for DUC depends on the topology of underlying power network. We tested ADMM for UC instances with five differently configured networks as follows:

- No line (NL): In this case, UC formulation (98) without network constraints is considered. All regions send their total power output to a designated region to compute the discrepancy in power demand and supply.
- Path
- Circle (Cr)
- Tree (Tr)
- A meshed example from PowerWorld (PW)

Each test case has 10 prosumers, and each prosumer has 5 generators. The load is owned by prosumer 1. Load and generators data are from the le_50_0_3 w.mod available at http://people.brunel.ac.uk/m̃astjjb/jeb/orlib/les/unitnew.zip. The demand profile is as shown in Figure 12. For this experiment, we assumed that the capacities of lines are not restrictive.

To better capture convergence speed of ADMM for different configurations, all figures are in 2 different scales, one for iterations up to 150 and the other for the remaining iterations. Figure 13 depicts the optimality gap (%) versus solution time for LP relaxation of UC instances with different network configurations. In Figure 14, this gap is shown for direct application of ADMM. We also used the solution of LP relaxation as warm start for direct application of ADMM for UC. The optimality gap for this case is shown in Figure 15.

Based on these numerical results we can observe and make the following remarks.

- LP relaxation of DUC converges fast to its optimal solution.
- The realistic, meshed topology case (PW) converges faster than most of the other configurations.
- The path topology, having the least connectivity, shows the worst behavior. Note that for communication network, we considered the same topology as

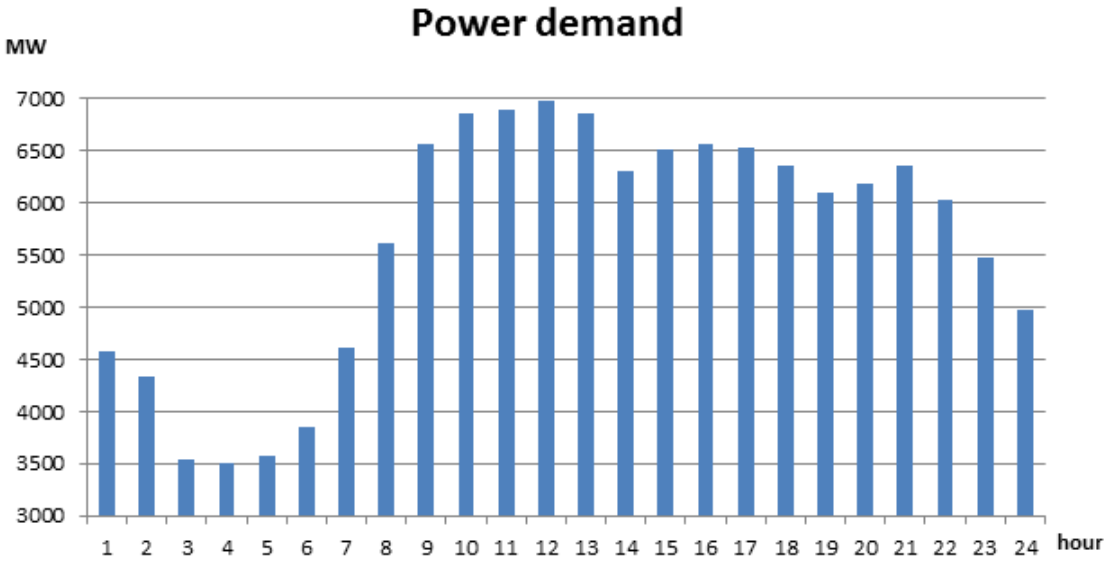

Figure 12: Demand profile for test cases with different network configurations.

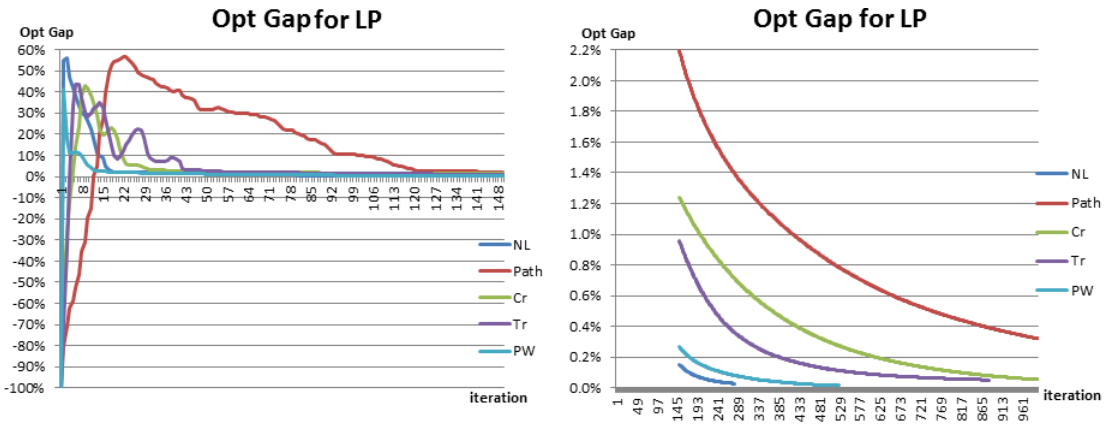

Figure 13: Optimality gap (%) for LP relaxation of UC instances with different network configurations.

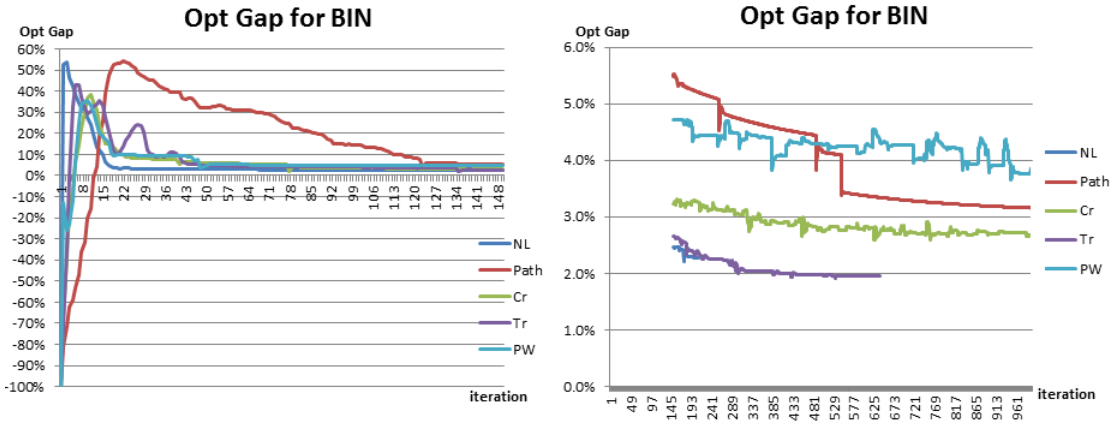

Figure 14: Optimality gap (%) for UC instances with different network configurations.

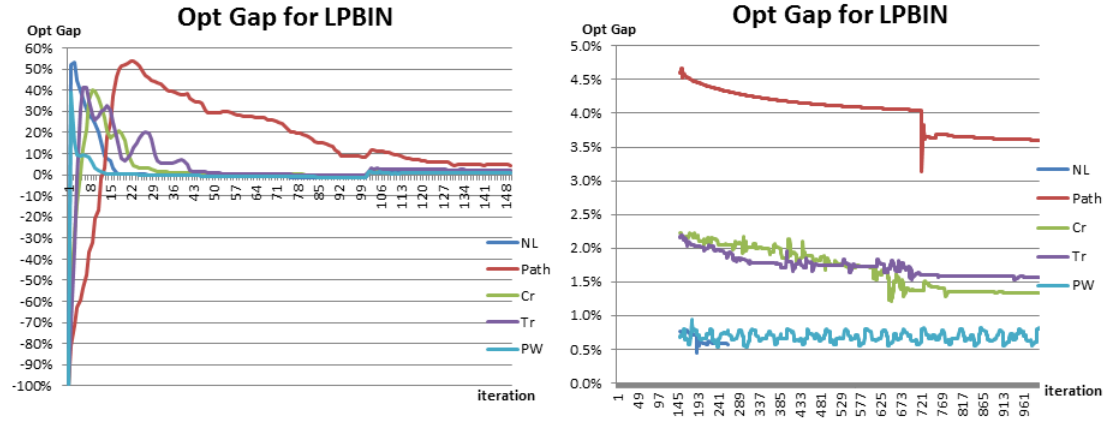

Figure 15: Optimality gap (%) for UC instances with different network configurations, using LP relaxation as warm start.

power transmission network. Then, more connectivity in power transmission network of regions means more information exchange between regions.

- Using LP relaxation in the first iterations improves the quality of solutions from direct application of ADMM for UC.

- If binary variables are not fixed after a while, there is no guarantee to prevent abnormal behaviors and large oscillations.

5.1.3 Primitive R&F Algorithm

In this subsection, we present results of testing the primitive version of R&F algorithm on small standard UC instance. In this version of algorithm, reserve constraints were not included. Moreover, instead of $\mathcal{L}^{+}_{\rho,\nu}\left(\boldsymbol{x}_\nu, \bar{\bar{\boldsymbol{\theta}}}^m_\nu, \bar{\bar{\boldsymbol{F}}}^m_\nu, \bar{\bar{\boldsymbol{r}}}^m_\nu, \boldsymbol{\lambda}^m_\nu, \boldsymbol{\mu}^m_\nu, \boldsymbol{\omega}^m_\nu\right)$ in (118), we used $\mathcal{L}^{+'}_{\rho,\nu}\left(\boldsymbol{x}_\nu, \bar{\bar{\boldsymbol{\theta}}}^m_\nu, \boldsymbol{\lambda}^m_\nu\right)$ as defined below:

$$\mathcal{L}^{+'}_{\rho,\nu}\left(\boldsymbol{x}_\nu, \bar{\bar{\boldsymbol{\theta}}}^m_\nu, \boldsymbol{\lambda}^m_\nu\right) := C_\nu(\boldsymbol{x}_\nu) + \boldsymbol{\lambda}^m_\nu\left(\tilde{\boldsymbol{\theta}}_\nu - \bar{\bar{\boldsymbol{\theta}}}^m_\nu\right) + \frac{\rho}{2}\left\|\tilde{\boldsymbol{\theta}}_\nu - \bar{\bar{\boldsymbol{\theta}}}^m_\nu\right\|_2^2 \tag{120}$$

5.1.3.1 Test cases

We used test cases available in MATPOWER package from [208] to generate our problem instances. These test cases are mainly for optimal power flow problems and do not have inter-temporal data such as minimum up and down times, ramp rate limits, start-up and shut down costs, initial state of generators and demand forecast of loads at each time period. We generated these missing data semi-randomly, as follows:

- Minimum up and down time of a generator: Minimum up time is set to be 1, 2, 3, 5, 6 and 8 for generators with maximum capacity in $(0, 50]$, $(50, 100]$, $(100, 200]$, $(200, 300]$, $(300, 600]$, and $(600, \infty]$, respectively. Note that maximum capacities for generators are available in MATPOWER package [208]. Minimum down time is randomly chosen from integers within ± 1 of minimum up time.

Table 1: Test case details

Buses	Generators	Lines	Cont. Var.	Bin. Var.	Constraints	t_c (s)
6	3	11	337	66	679	0.86
9	3	9	340	67	768	0.83
14	5	20	472	114	815	0.21
30	6	41	699	135	1533	203.06
39	10	46	1097	195	2389	0.25
57	7	80	817	152	1260	8.48

- Start-up and shut down costs: These costs were set to be 10% of the production cost of maximum power output.

- Initial state of the generator: We assumed that with probability 0.2 each generator was off initially for a random number of periods between 0 and twice minimum down time. Also, with probability 0.8, it was on for a random number between 0 and twice minimum up time. If the generator was on, the initial power generation level of generator is randomly chosen between its minimum and maximum operating level.

- Demand forecast of a load at time period: In the test cases of MATPOWER [208], there is a single number as demand of the load. We considered this demand as peak time demand and multiplied it by a load factor (see Fig. 16) at each hour to calculate the base demand of the load. Then, demand forecast was chosen randomly within $\pm 15\%$ of its base demand value.

In all of the cases, 24 hours of planning with 1 hour time granularity are considered. In the first three columns of Table 1, the number of buses, generators, transmission lines are shown for each case. The next three columns indicate the numbers of continuous and binary variables and linear constraints of each case after the presolve stage in CPLEX. CPU time (in seconds) needed to solve each case is presented in the last column.

Each test case is split into different numbers of prosumers. In the first case with

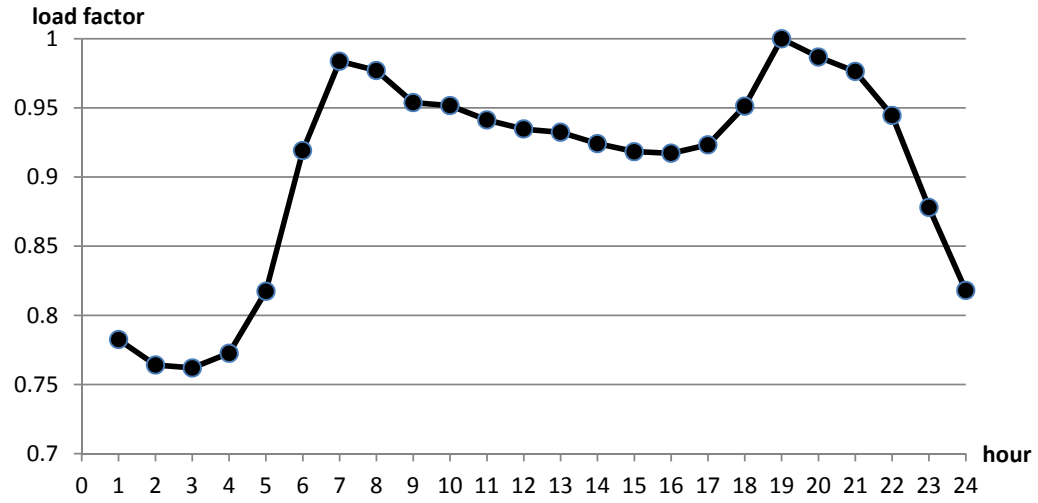

Figure 16: Load factor at each hour

39 buses and 3 prosumers (3^a), each control area is assumed to be a prosumer. In the other cases, buses and devices connected to them are partitioned into prosumers with roughly the same number of buses.

5.1.3.2 Implementation details

We tested the primitive R&F with the following details. First, the algorithm had 10,000 iterations limit to run ADMM-Relax (see Section 3.4.1) for solving continuous relaxation of the UC problem. In our experiments, we used the origin $\mathbf{0}$ as initial primal and dual values. In this phase, subproblems were QPs and relatively easy to solve. The acceptable tolerance was reaching less than 10^{-3} maximum and 10^{-1} total primal infeasibility (measured by residuals) and less than 0.5% gap between total augmented objective values $\mathcal{L}'^{+}_{\rho,\nu}(\boldsymbol{x}_\nu^{k+1}, \bar{\bar{\boldsymbol{\theta}}}^k, \boldsymbol{\lambda}_\nu^k)$ and total base objective values $C_\nu(\boldsymbol{x}_\nu^{k+1})$.

After terminating ADMM-Relax, R&F algorithm switched to ADMM-Release (see Section 3.4.2). In this phase, it had 200 iterations to solve ADMM with MIQP subproblems. After every five iterations, if there was a binary variable change, the penalty term ρ was increased by 5%. If there was no binary variable change within

last consecutive 30 iterations or acceptable tolerance was reached, this phase was terminated.

Next, if the current binary solution was not previously explored, R&F algorithm switched to ADMM-Fix (see Section 3.4.3). In ADMM-Fix, binary variables were fixed until reaching the acceptable tolerance or maximum iterations of this step which was 800.

In addition to the iterations spent on ADMM-Relax, the R&F algorithm had 5000 iterations to spend on ADMM-Release and ADMM-Fix. Total accumulative time limit of 3600 seconds for the R&F was assumed.

5.1.3.3 Numerical results

Numerical results of running the primitive R&F Algorithm are presented in Table 2. In this table, the first three columns correspond to the numbers of buses, prosumers and boundary lines of the cases. $\text{iter}_{\text{Relax}}$ and t_{Relax} stand for the number of iterations and CPU time in seconds spent on ADMM-Relax. The rest of the columns consist of the iterations and CPU times spent to find the first and best solutions by R&F algorithm. $\text{iter}_{\text{Release}}$ and iter_{Fix} are the number of iterations spent on Algorithms ADMM-Release and ADMM-Fix, respectively. $t_{\text{R\&F}}$ is time spent to find the solution. "Gap (%)" is the gap between optimal objective value of the case (which was solved in a centralized framework beforehand for comparison) and the average of total augmented and base objective value of the solution found by R&F. "Cycle" indicates the number of cycles of releasing and fixing binary variables (switches between Algorithms ADMM-Release and ADMM-Fix) until finding the best solution within the iteration and time limits.

Table 2: Numerical results for primitive R&F

					First feasible solution				Best feasible solution				
bus	pro	bnd	$\text{iter}_{\text{Relax}}$	t_{Relax}	$\text{iter}_{\text{Release}}$	iter_{Fix}	$t_{\text{R\&F}}$	gap (%)	$\text{iter}_{\text{Release}}$	iter_{Fix}	cycle	$t_{\text{R\&F}}$	gap (%)
6	3	8	45	0.27	1	60	0.9	5.18	1175	310	5	24.1	0.58
	6	11	74	0.33	45	27	1.2	6.64	1634	518	7	46.7	0.52
9	3	6	72	0.30	12	27	0.8	< 0.01	12	27	1	0.8	< 0.01
	9	9	359	0.88	40	387	2.9	< 0.01	40	387	1	2.9	< 0.01
14	3	9	195	1.62	1	159	3.5	17.39	1148	1600	6	45.5	< 0.01
	7	14	243	1.29	1	212	2.9	3.28	570	744	3	16.4	0.17
	14	20	545	2.18	1	209	3.5	14.93	694	1837	6	22.3	0.29
30	3	9	107	2.18	11	23	4.5	5.87	1743	331	10	464.1	< 0.01
	5	19	259	2.89	1	75	4.6	3.67	2639	1133	15	397.3	0.04
	10	30	346	2.83	1	122	4.3	4.53	2157	2158	9	173.4	1.13
39	3^a	6	262	8.87	1	43	11.8	< 0.01	1	43	1	11.8	< 0.01
	3^b	18	858	27.11	1	172	35.0	< 0.01	1	172	1	35.0	< 0.01
	5	21	845	15.52	38	279	23.0	< 0.01	38	279	1	23.0	< 0.01
	10	23	900	8.37	1	570	15.8	< 0.01	1	570	1	15.8	< 0.01
57	3	15	2866	209.33	1	799	291.7	0.71	198	1598	2	491.3	0.02
	10	37	5751	55.55	1158	2462	136.9	0.09	1158	2462	5	136.9	0.09

For the cases with 9 and 39 buses, the best solutions of R&F were found in the first cycle of releasing and fixing binary variables. For the 30-bus and 57-bus cases with 3 prosumers, because the generators are not evenly distributed between prosumers, one of the prosumers had much harder subproblems than others. The best solution was found within 1 minute for cases with 6, 9, 14 and 39 bus cases and between 2 to 8 minutes for 30 and 57 bus cases.

In most cases, many iterations and much time are are necessary for solving CR with good accuracy. This phase can be eliminated or accelerated by using some reasonable initial solutions such as primal and dual values from running R&F for the day before.

For the cases with the same number of buses, larger numbers of prosumers cause a decrease in t in some cases (e.g. 30 and 57 bus cases) and an increase in the rest. Therefore, number of prosumers and the strategy for partitioning buses into prosumers can be investigated to determine the best way of establishing prosumers.

By comparing the CPU times $t_{\text{R\&F}}$ of primitive R&F in Table 2 and t_c of the centralized method in Table 1, we observe that in most cases (except 30 bus case), the centralized method is faster than R&F. For the cases with 30 buses, the first feasible solutions are found in less than 5 seconds which have less than 6 % optimality gap, while it took 203 seconds to solve the corresponding UC problem centrally. For these cases, the CPU times of the best solutions (which have very small optimality gap) are also comparable to the centralized method. Although more experiments are needed to solve larger and harder cases to observe the real effect of decentralizing the UC problem, intuitively one expects faster parallel solution times for the cases with thousands of buses and tens to hundreds of prosumers.

The main conclusions are:

- Although there is no theoretical proof of convergence for ADMM-type algorithms applied to discrete optimization problems, in practice, optimal or near

optimal solutions with less than 1% optimality gap and acceptable primal infeasibility of 10^{-3} were obtained by the proposed primitive R&F for the DUC cases.

- Because of the scalability of the proposed framework, control regions of a large power system can be modelled as prosumers where neighbors can coordinate their operation, optimizing their interchange for global benefit.
- The proposed scheme would not require drastic changes to existing UC solution software to implement.
- In the proposed framework, neighboring prosumers need to locally communicate with each other. Only a small amount of data (related to boundary bus angles) must be communicated between neighboring prosumers at each iteration. Sensitive data, such as generator costs and ramp rates, can remain private to each prosumer.
- For most of the test cases that we solved, solving a centralized UC problem needs less CPU time than solving PDUC with Hybrid ADMM.

Next, we present the results for an improved version of R&F.

5.1.4 Improved R&F Algorithm

To speed up the R&F algorithm for solving DUC, we improved it by techniques presented in Sections 3.4.5 and 4.4.3. In this section, we test the improved version of R&F on medium to large size standard power systems. In these instances, reserve constraints were not included. Moreover, instead of $\mathcal{L}^{+}_{\rho,\nu}\left(\boldsymbol{x}_\nu, \bar{\bar{\boldsymbol{\theta}}}^m_\nu, \bar{\bar{\boldsymbol{F}}}^m_\nu, \bar{\bar{\boldsymbol{r}}}^m_\nu, \boldsymbol{\lambda}^m_\nu, \boldsymbol{\mu}^m_\nu, \boldsymbol{\omega}^m_\nu\right)$ in (118), we used $\mathcal{L}^{+''}_{\rho,\nu}\left(\boldsymbol{x}_\nu, \bar{\bar{\boldsymbol{\theta}}}^m_\nu, \bar{\bar{\boldsymbol{F}}}^m_\nu, \boldsymbol{\lambda}^m_\nu, \boldsymbol{\mu}^m_\nu\right)$ as defined below:

$$\begin{aligned}\mathcal{L}^{+''}_{\rho,\nu}\left(\boldsymbol{x}_\nu, \bar{\bar{\boldsymbol{\theta}}}^m_\nu, \bar{\bar{\boldsymbol{F}}}^m_\nu, \boldsymbol{\lambda}^m_\nu, \boldsymbol{\mu}^m_\nu\right) =& C_\nu(\boldsymbol{x}_\nu) + \boldsymbol{\lambda}^m_\nu\left(\tilde{\boldsymbol{\theta}}_\nu - \bar{\bar{\boldsymbol{\theta}}}^m_\nu\right) + \frac{\rho}{2}\left\|\tilde{\theta}_\nu - \bar{\bar{\boldsymbol{\theta}}}^m_\nu\right\|_2^2 \\ &+ \boldsymbol{\mu}^m_\nu\left(\tilde{\boldsymbol{F}}_\nu - \bar{\bar{\boldsymbol{F}}}^m_\nu\right) + \frac{\rho}{2}\left\|\tilde{\boldsymbol{F}}_\nu - \bar{\bar{\boldsymbol{F}}}^m_\nu\right\|_2^2\end{aligned} \tag{121}$$

Table 3: Centralized Test Cases

Buses	Generators	Lines	Cont. Var.	Bin. Var.	Constraints	t_c (s)
57	7	80	817	152	1260	8.48
118	54	186	6448	1221	10553	5194
300	69	411	9130	1487	16013	7200
3,012	496	3572	43977	8361	124793	1898

Having new terms related to $\tilde{\boldsymbol{F}}_\nu - \bar{\bar{\boldsymbol{F}}}_\nu^m$ in $\mathcal{L}_{\rho,\nu}^{+''}\left(\boldsymbol{x}_\nu, \bar{\bar{\boldsymbol{\theta}}}_\nu^m, \bar{\bar{\boldsymbol{F}}}_\nu^m, \boldsymbol{\lambda}_\nu^m, \boldsymbol{\mu}_\nu^m\right)$ speeds up convergence of the proposed ADMM based DUC methods.

5.1.4.1 Test cases

Experiments were conducted using the IEEE 57, 118, 300, and 3,012-bus cases. These cases are available from the MATPOWER software package [208]. As mentioned in Section 5.1.3.1, these test cases are mainly for OPF problems and do not have inter-temporal data for UC. With the same logic described in Section 5.1.3.1, we generated these missing data. Divisions of each case into a specific number of regions were determined by the heuristic partitioning method described in Section 4.4.3.3.

Table 3 reports several metrics regarding each case's size and centralized solution time. The load dataset for all four cases was taken from the summer peak scenario of a large utility system and scaled to match the peak load observed in each case's input data. For the 3,012-bus case, a second day in the same scenario was used. Fig. 17 shows the load factors for "Day 1" and "Day 2". A load factor of 1.0 corresponds to a total of 1,250 MW for the 57-bus case, 4,242 MW for the 118-bus case, 23,525 MW for the 300-bus case, and 27,170 MW for the 3,012-bus case. The cases with 57, 118, and 300 buses had quadratic cost functions specified for generators, while the 3,012-bus case had linear generator costs. This fact may explain the much longer solution times necessary for the 118-bus and 300-bus cases seen in Table 3. Solution times to reach 0% optimality are reported for the centralized cases. For the 300-bus case, an optimality gap of 0.21% remained after two hours.

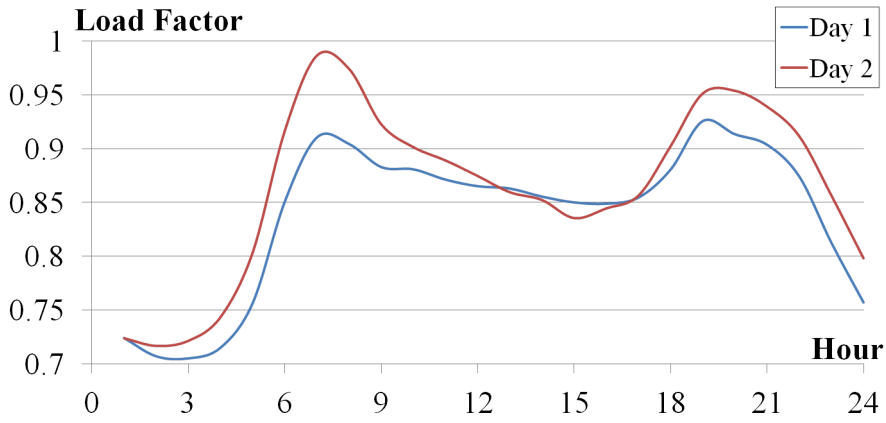

Figure 17: Load factors at each hour for days 1 and 2 of 3,012-bus case.

5.1.4.2 Numerical results

Table 4 summarizes the results of the DUC experiments conducted with improved R&F. The column headings "Buses" and "Prosumers" are the numbers of buses and prosumers in each experiment. The adjacent columns $\text{iter}_{\text{relax}}$ and t_{relax} are the number of iterations and total CPU time in seconds spent in ADMM-Relax. The next columns $\text{iter}_{\text{Release}}$, iter_{Fix}, $t_{\text{R\&F}}$, and "Gap %" are, respectively, the number of iterations in ADMM-Release, the number of iterations in ADMM-Fix, the total CPU time spent in R&F, and the percentage gap between the best solutions found through R&F and through centralized branch-and-cut (i.e., the cases in Table 3).

In Table 4, the solution times for all cases strictly decrease with respect to increasing decomposition granularity (i.e. more and increasingly smaller regions). Although the numbers of iterations in the various R&F stages generally increased with more regions, total computational time decreased dramatically. Certainly the individual MIQP subproblems required much less time with diminished size as the combinatorics of the binary solution search would suggest, but QP subproblems were also solved with much greater speed, as can be seen from the results of the ADMM-Relax.

The 3,012-bus case was easily the largest case tested, providing some of the most interesting results. Most importantly, parallel solution time showed a monotonic

Table 4: Decentralized Test Cases and Results

Case Info		CR Stage		Best feasible solution			
Buses	Prosumers	$\text{iter}_{\text{relax}}$	t_{Relax} (s)	$\text{iter}_{\text{Release}}$	iter_{Fix}	$t_{\text{R\&F}}$ (s)	Gap %
57	3	86	4.9	486	182	410.6	< 0.01
	10	220	3.7	687	899	43.8	< 0.01
118	4	85	24.0	240	40	313.1	0.14
	10	89	4.2	326	50	74.9	< 0.01
	20	98	1.5	519	59	33.1	< 0.01
300	4	105	24.0	532	92	668.4	0.29
	10	223	12.0	368	57	413.8	< 0.01
	30	283	3.5	427	168	43.2	< 0.01
3,012	10	227	278	152	1204	3449	0.01
	20	218	107	179	1545	1989	0.03
	50	248	67	150	1783	537	0.02
	100	270	44	99	1576	215	0.11
	200	241	17	147	1402	143	0.03
	300	301	17	157	1527	94	< 0.01

decrease with an increasing number of regions. Although the numbers of iterations in ADMM-Release and ADMM-Fix do not show any clear trend, the MIQP subproblems can be solved much faster when the system is divided into more regions. For example, while the centralized 3,012-bus MIQP took 1898 s to solve, each subproblem in the 300-region case with 3,012 buses could be solved in about 40 ms.

Figure 18 shows the solution times of the 3,012-bus scenarios for the day 1 and day 2 load profiles shown in Figure 17. The problem for the second day was initialized with the best solution from the first day. Even though the forecast demand is different enough to result in a different commitment, providing a high quality binary solution and dual variables are enough to greatly speed up the search process for the day 2 solution.

The times to discovery and the optimality gaps for each binary-feasible solution to the 3,012-bus case are shown in Figures 19 and 20 for load profiles of day 1 and day 2, respectively. Intersection with the horizontal axis indicates that a globally optimal solution was found (corresponding to optimality gap 0.0%). In the 300-region case, the global optimum objective was reached in 94 s, so the algorithm

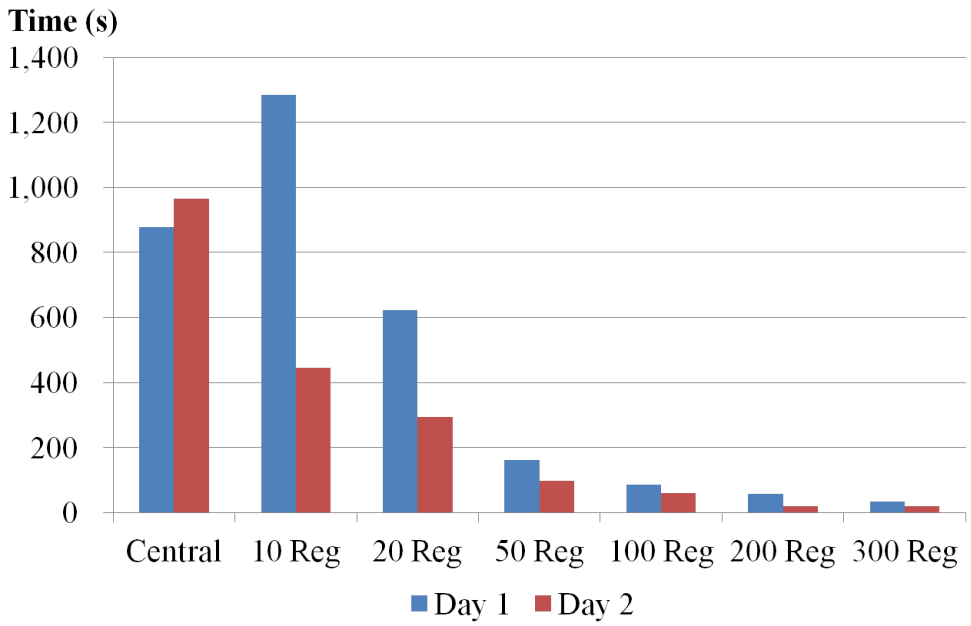

Figure 18: Times to find first feasible solution for each regional decomposition of the 3,012-bus case for days 1 and 2. The day 2 case was warm-started with the best solution from day 1.

halted. The other cases reached one of two stopping criteria: solution time limit (1 hour) or maximum number of iterations (2000). The 300, 200, 100, and 50-region decompositions all found many binary solutions before the first feasible solution was found by the centralized method.

5.1.5 Self Commitment

As discussed in Section 4.1, the DUC formulation presented in this research is fundamentally different from a self-commitment framework. A comparison is provided here to illustrate some differences in approach and results. The model of [171, 170] was implemented as a representative self-commitment process. The solution process of the self-commitment method was as follows: an auctioneer announces a set of hourly energy prices, price-taking generators individually determine hourly commitments and dispatch to maximize profit, generators submit hourly energy offers to the auctioneer, the auctioneer adjusts the hourly energy prices, and the process repeats

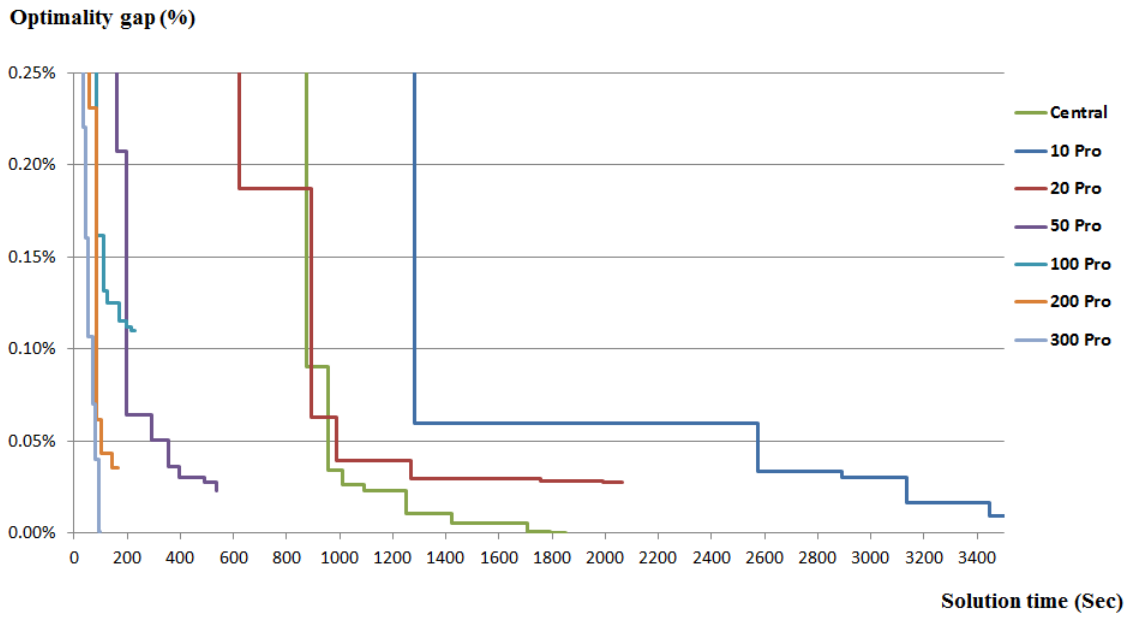

Figure 19: Feasible solution optimality gaps for 3,012-bus case with day 1 load profile and different regional decompositions.

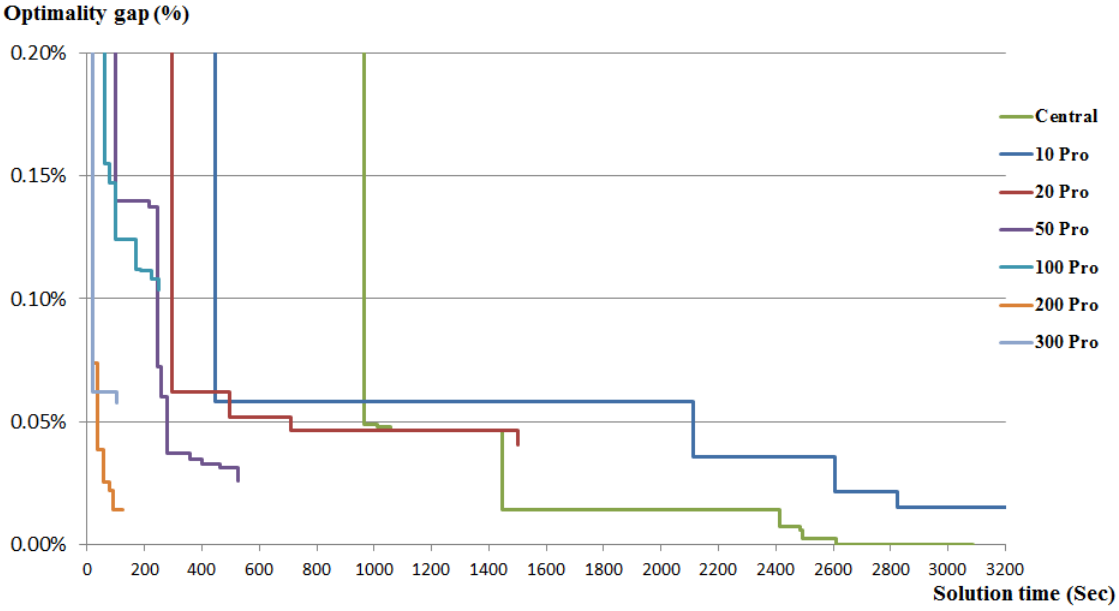

Figure 20: Feasible solution optimality gaps for 3,012-bus case with day 2 load profile and different regional decompositions.

Table 5: Comparison of R&F and Self-Commitment Optimality Gaps for 3,012-Bus Case

Number of Regions	10	20	50	100	200	300
Self-Commitment (%)	11.7	12.5	10.8	10.1	10.5	10.5
Release-and-Fix (%)	0.03	0.02	0.07	0.15	0.04	0.24

Table 6: Details of non-uniform prosumer decomposition

Prosumer type	No. of prosumers	No. of buses	No. of Generators
Big	10	100	15
Medium	20	50	8
Small	100	10	0–3

until sufficient generation is acquired to serve the load with minimal prices.

Similar to the findings of [171, 170], self-commitment solutions for the 3,012-bus case with varying regional decompositions were inefficient compared to DUC with R&F in terms of total system cost. Because the self-commitment method cannot find minimal prices that clear the market, excess generation is always present. Heuristics to match the generation to load following this stage are subject to being trapped in local optima. From the results in Table 5, the cost of anarchy can be seen to be around 10% with a self-commitment approach.

5.1.6 Effect of Partitioning Approach

In previous configurations, prosumers had roughly the same number of buses and generators. To investigate the effect of non-uniform prosumer decompositions, we considered a special configuration. In the IEEE 3,012 bus case, we assumed three sizes of prosumers: big, medium and small. For each prosumer type, the numbers of prosumers, buses and generators are reported in Table 6. There were 130 prosumers in total. Thus, we denote it by "130 Pro".

Figure 21 shows optimality gap versus solution time for different configurations of IEEE 3,012. In this figure, results from non-uniform "130 Pro" are located between "20 Pro" and "50 Pro" because the size of the big prosumers in "130 Pro" is between

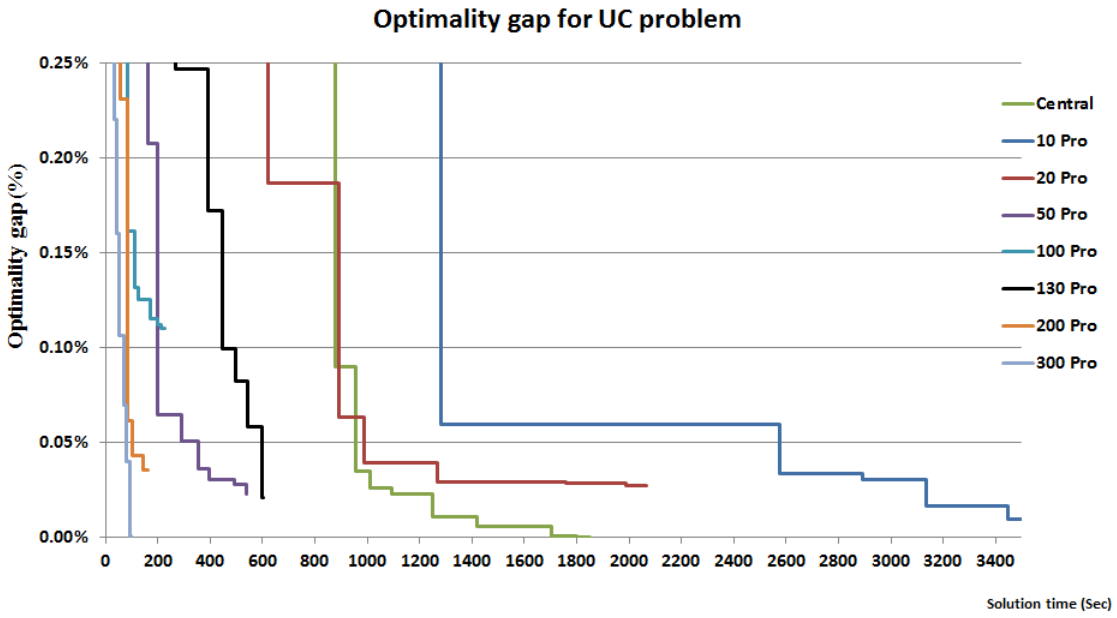

Figure 21: Optimality gap for R&F with different partitioning.

the size of big (uniform) prosumers in "20 Pro" and "50 Pro".

5.1.7 Communication Loss

Up to now, we have assumed perfect communication; that is, every packet is transmitted successfully between prosumers. In practical environments, some packets will be lost. To investigate communication losses on R&F and quality of solutions, we assumed that with some probability p a message is not delivered. When a prosumer does not get a message with new angles or residuals, it uses old values from the last successful iteration. The results for scenarios with $p = 0.0\%$, 0.1%, 1.0%, and 5.0% for "130 Pro" are depicted in Figure 22.

To check the robustness of the proposed algorithm to communication losses with small probabilities, Figures 23 and 24 shows the results from 10 random runs of R&F with $p = 0.1\%$, 1.0%. Figure 25 presents average of 10 runs for p= 0.1 %, 1.0 %.

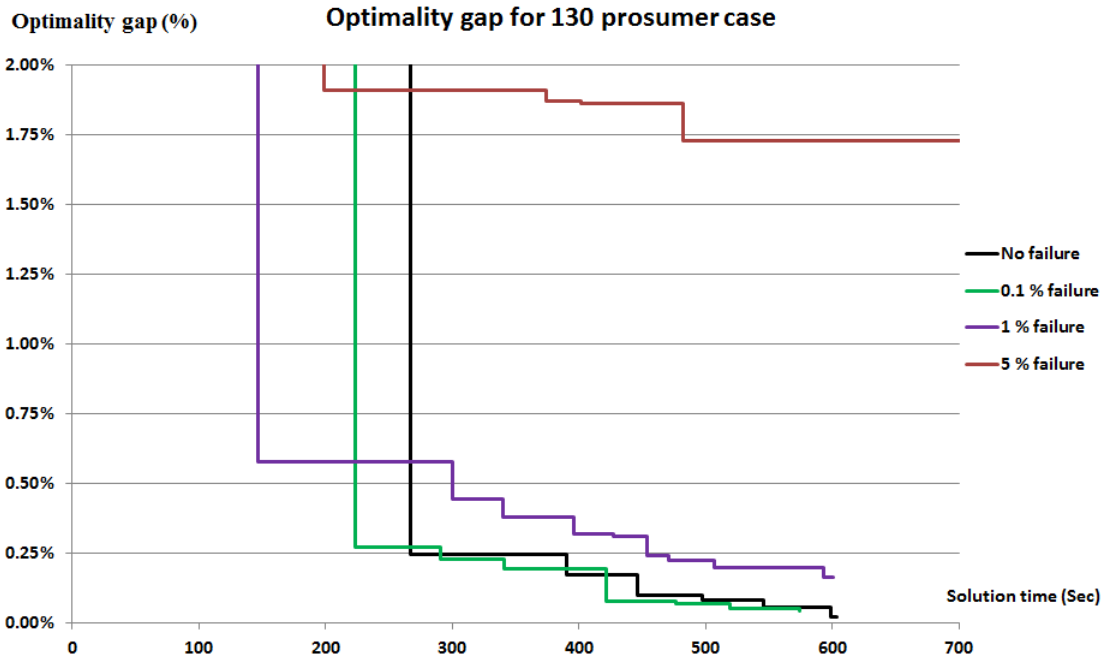

Figure 22: Optimality gap under different scenarios for communication loss.

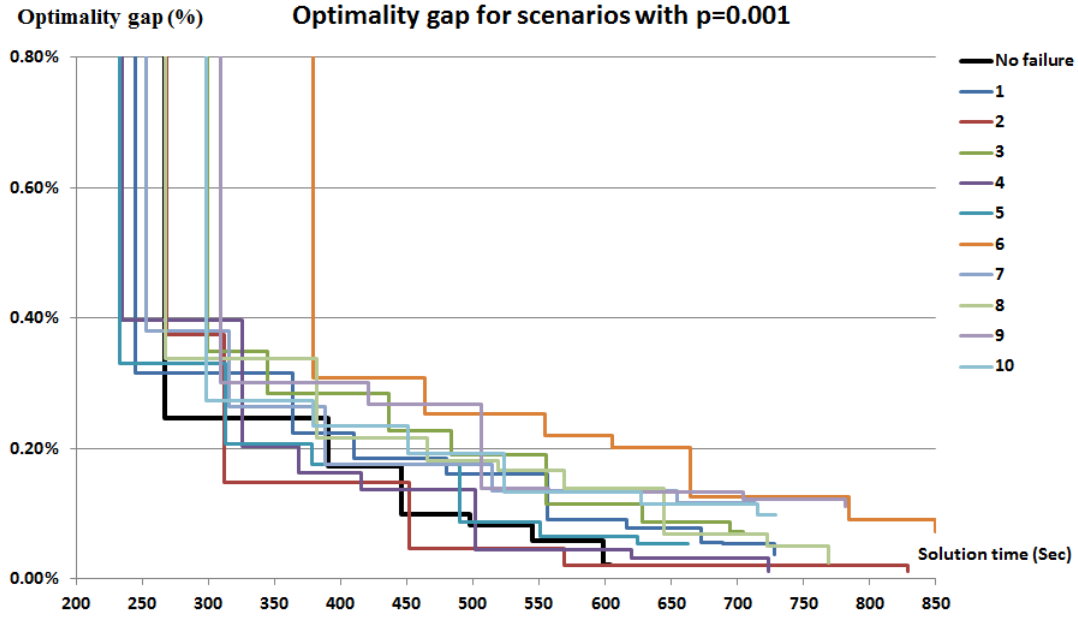

Figure 23: Optimality gap for 10 experiments with communication loss probability $p = 0.001$.

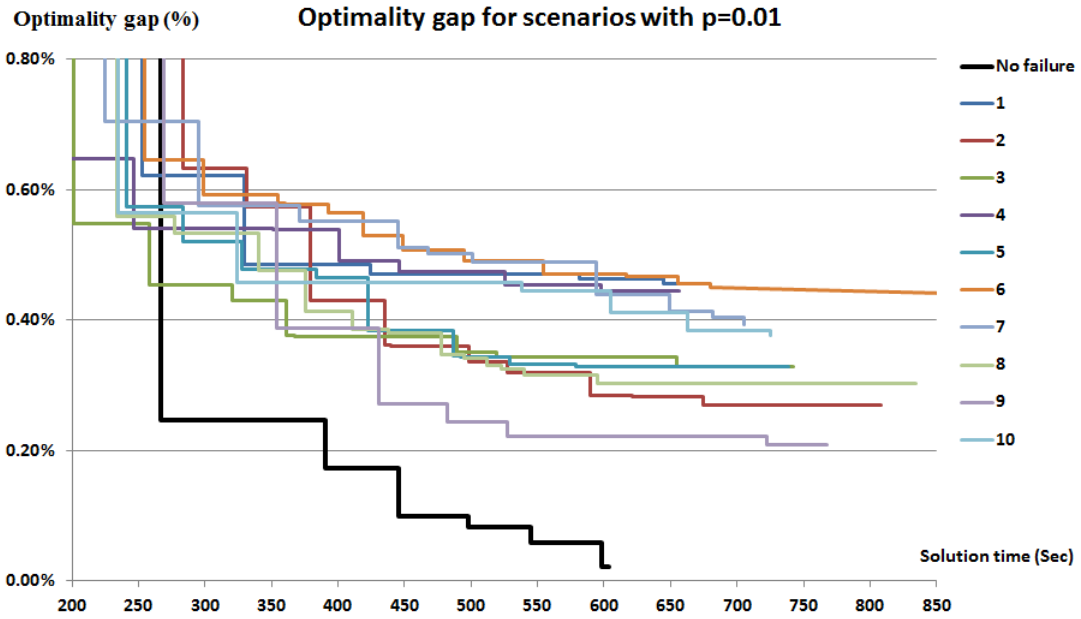

Figure 24: Optimality gap for 10 experiments with communication loss probability $p = 0.01$.

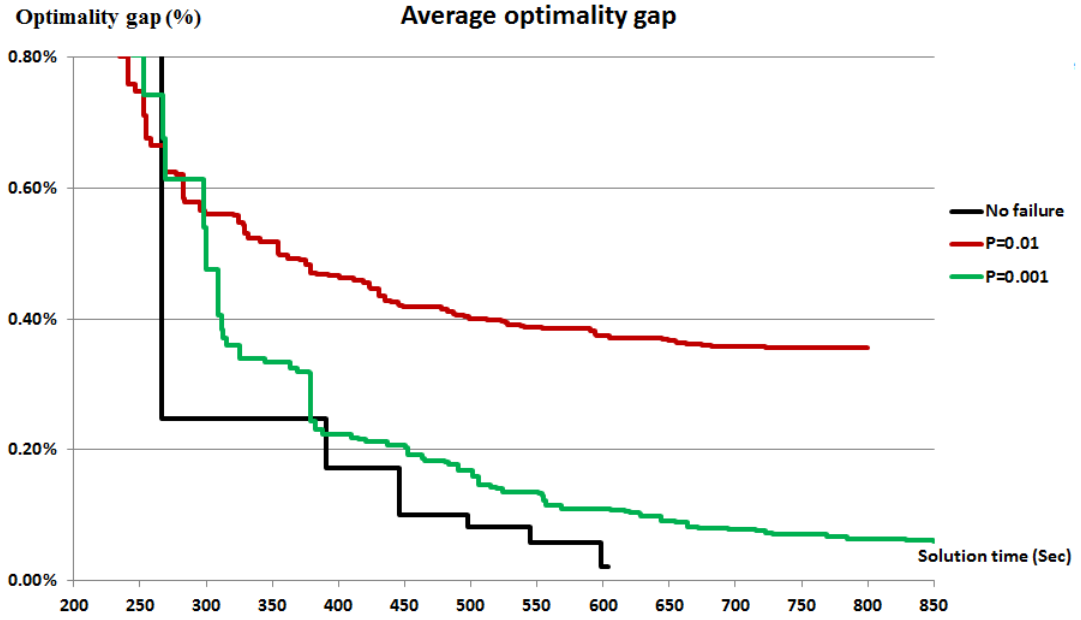

Figure 25: Average optimality gap for experiments with communication loss probability $p = 0.001, 0.01$.

Table 7: Power System Test Cases

Power System	Number of each generator type								Total Gens	Capacity (MW)	Buses	Lines
	1	2	3	4	5	6	7	8				
A	19	19	23	25	25	18	8	13	150	30415	3012	3566
B	42	36	12	10	21	14	10	7	152	44107	3374	4068

5.2 Parallel Implementation of R&F with MPI

In this section, we present the results of implementing the complete version of R&F in a truly parallel computational environment with Message Passing Interface (MPI). Here, we consider reserve constraints as well as full version of the Lagrangian function $\mathcal{L}^{+}_{\rho,\nu}\left(\boldsymbol{x}_\nu, \bar{\bar{\boldsymbol{\theta}}}^m_\nu, \bar{\bar{\boldsymbol{F}}}^m_\nu, \bar{\bar{\boldsymbol{r}}}^m_\nu, \boldsymbol{\lambda}^m_\nu, \boldsymbol{\mu}^m_\nu, \boldsymbol{\omega}^m_\nu\right)$ as noted in (118).

5.2.1 Test Cases

Two power systems, A and B, were used to conduct experiments. Some basic information regarding their structures is shown in Table 7. Network topologies for the systems A and B are adapted from the IEEE 3,012- and 3,375-bus cases, respectively, available in the MATPOWER software package [208]. Because one of the buses in the 3,375-bus case was not connected to the rest of the network, that bus was removed, leaving 3,374 buses. Moreover, parallel lines between buses were replaced by their equivalents.

In the MATPOWER cases, which were originally intended for OPF, most of the data needed for UC such as minimum up and down times, ramp up and down rates, and startup costs are not available. Thus, we replaced those generators with the eight classes of generators used in [38, 134, 119] for UC problems with no network constraints (see Table 8). In our experiments, quadratic generation costs have been approximated by piecewise linear costs with five line segments of equal length.

The total of each reserve product q in the system must meet the minimal system requirement. For this study, the system requirement was equal to the size of the

Table 8: Generator Data [38]

Gen	Technical Information						Cost Coefficients				
	$\overline{P}$ (MW)	$\underline{P}$ (MW)	TU/TD (h)	RU/RD (MW/h)	T^{Init} (h)	T^{cold} (h)	C^{NL} (\$/h)	C^{LV} (\$/MWh)	C^{Q} (\$/MW^{2}h)	C^{HS} (\$)	C^{CS} (\$)
1	455	150	8	225	+8	5	1000	16.19	0.00048	4500	9000
2	455	150	8	225	+8	5	970	17.26	0.00031	5000	10000
3	130	20	5	50	-5	4	700	16.60	0.00200	550	1100
4	130	20	5	50	-5	4	680	16.50	0.00211	560	1120
5	162	25	6	60	-5	4	450	19.70	0.00398	900	1800
6	80	20	3	60	-3	2	370	22.26	0.00712	170	340
7	85	25	3	60	-3	2	480	27.74	0.00079	260	520
8	55	10	1	135	-1	0	660	25.92	0.00413	30	60

Table 9: Total Demand (% of Total Capacity)

Time	1	2	3	4	5	6	7	8	9	10	11	12
Demand	71%	65%	62%	60%	58%	58%	60%	64%	73%	80%	82%	83%
Time	13	14	15	16	17	18	19	20	21	22	23	24
Demand	82%	80%	79%	79%	83%	91%	90%	88%	85%	84%	79%	74%

largest contingency for the 10-minute contingency reserve and the size of the second largest contingency for the 30-minute operating reserve. Spinning reserve was required to be at least half of the 10-minute reserve.

For each power system A and B, there were three test cases. In test cases A_1 and B_1, we considered 24 hourly periods, where the total system demand at each hour is determined as given in Table 9. Test cases A_2 and B_2 have also 24 hourly periods, but the total system demand is obtained by shifting the demand values in Table 9 earlier by one hour with demand from hour 1 wrapping around to hour 24 (i.e. hour 1 demand is 65%, hour 2 demand is 62%, and so on). Test cases A_3 and B_3 have 72 hourly periods (three days), where the demand in each of the three days is the same as A_1 and B_1, respectively. Distribution of demand among buses followed the proportions of the original load data in the MATPOWER files. To evaluate the

decentralized approach, the systems A and B were partitioned into n regions where $n \in \{20, 30, 40, 50, 60, 70, 75, 80, 90, 100, 120, 150, 200\}$.

All test case input files and regional assignments of buses in each partition are online, available at [60].

5.2.2 Implementation Details

All algorithms were coded in C++ using CPLEX 12.6 through the Concert API. Experiments were conducted on a UNIX cluster with cores rated between 2.0 and 3.0 GHz and addressable memory limited to 4 GB. The cluster machines are primarily a variety of Xeon E5 and X5 models. Central UC instances were solved using internal CPLEX multi-threading with four cores. Test cases A_1, and B_1 were solved using a computational time limit of two hours, while test cases A_3 and B_3 had a time limit of ten hours. All cases were solved with a 1% relative optimality gap tolerance. In both centralized and decentralized methods, the barrier method was used to solve root node problems.

The penalty factor ρ was initialized to a value of 2. Whenever the ADMM-Release stage did not find a new binary solution, ρ was multiplied by 0.95 to encourage more exploration in the space of binary variables. Relative primal and dual residual tolerances ϵ^{Pri} and ϵ^{Dual} and the relative tolerance ϵ^{Obj} between the base objective value and augmented objective value were set to 0.5% for all cases. R&F switched from ADMM-Relax to ADMM-Release after reaching all of the ϵ^{Pri}, ϵ^{Dual}, and ϵ^{Obj} tolerances or the iteration limit of 400.

In ADMM-Release, if there were no changes in binary variables in the last 15 iterations, or an iteration limit of 50 was reached, or the solution satisfied the ϵ^{Pri}, ϵ^{Dual}, and ϵ^{Obj} tolerances and the binary solution was not previously explored, R&F switched to ADMM-Fix. In the test cases A_1, A_2, B_1 and B_2, the time limit to solve each subproblem was set to 70, 50, 40, and 30 seconds for the configurations with the

number of regions set to 20, 30, 40 and 50, respectively. For the other configurations, the time limit was 20 seconds. For A_3 and B_3, the above time limits were multiplied by 2. For each MIQP subproblem, the optimal solution of each iteration was used as a warm start for the next iteration. When the MIQP subproblem was being solved for the first time, since there was no warm start, the time limit was multiplied by 3 to find the first feasible integer solution. The optimality gap tolerance was set to 1% for all MIQP subproblems.

In ADMM-Fix, if there was no decrease in primal residuals for 50 consecutive iterations, that solution was considered an infeasible solution and discarded. Otherwise, it continued until it satisfied the ϵ^{Pri}, ϵ^{Dual}, and ϵ^{Obj} tolerances or reached 100 iterations. At the end of this phase, if the solution satisfied the tolerances and provided a better objective value than the best one recorded yet, the new best solution was recorded. A total iteration limit of 2,000 was used for the whole algorithm.

The message passing interface (MPI) standard was used to develop the distributed software. Decentralized test cases used $n + 1$ computational nodes where each node used a single core. Each region was assigned to one node with the final node being used as a simple coordinator. The coordinator node kept track of the R&F stage, checking stopping criteria, and recording the binary solutions. Note that most of the information internal to each region does not need to pass through the coordinator.

5.2.3 Numerical results

To provide a benchmark for the DUC approach, three centralized UC formulations were solved for each test case: one without any network constraints, one with network constraints represented through line sensitivities (GSFs), and one with voltage phase angles (107). The GSF network model included only line constraints which were binding or near binding at the optimal solution of the no-network model (98). This corresponded to 44 lines for the A cases and 39 lines for the B cases.

Table 10: Case Sizes in Centralized Models

	No Network Model			GSF Network Model			Phase Angle Model		
Case	# Bin. Vars.	# Cont. Vars.	# Constr.	# Bin. Vars.	# Cont. Vars.	# Constr.	# Bin. Vars.	# Cont. Vars.	# Constr.
A_1 A_2	3,600	28,800	67,375	3,600	29,856	68,431	3,600	101,088	310,807
A_3	10,800	86,400	207,199	10,800	89,568	210,367	10,800	303264	937495
B_1 B_2	3,648	29,184	68,323	3,648	30,120	69,259	3,648	110,160	344,539
B_3	10,944	87,552	211,123	10,944	90,360	213,931	10,944	330,480	1,039,771

Table 11: Centralized Solution Information

	No Network Model			GSF Network Model			Phase Angle Model		
Case	UB ($ 10^6)	LB ($ 10^6)	Gap (%)	UB ($ 10^6)	LB ($ 10^6)	Gap (%)	UB ($ 10^6)	LB ($ 10^6)	Gap (%)
A_1	11.298	11.261	0.33	11.314	11.275	0.34	11.316	11.276	0.35
A_2	11.216	11.155	0.54	11.201	11.168	0.30	11.650	11.169	4.13
A_3	33.577	33.408	0.50	33.605	33.440	0.49	35.771	33.446	6.50
B_1	15.490	15.454	0.24	15.536	15.511	0.16	16.179	15.469	4.39
B_2	15.496	15.461	0.23	15.544	15.516	0.19	16.020	15.489	3.32
B_3	46.524	46.346	0.38	46.627	46.501	0.27	–	46.394	–

Table 10 shows the problem size of each test case under each of the models. Table 11 shows the best upper and lower bounds discovered with the three models, and Table 12 presents the computational time required to obtain those results. Some cases could not be solved to 1% optimality gap under all models. For example, only the root node relaxation was solved for the three-day case B_3. Some of the cases could be solved under the GSF network model within the time limit, but they did not reach 1% optimality gap. In such cases, the time $t_{1\%}$ is blank in Table 12.

Note that the lower bounds provide some validation for the models. In the model without network constraints, the lower bounds are the smallest, whereas the lower bounds under the phase angle formulation are the highest. This is exactly as expected since binding line constraints drive up the total system cost. Therefore the GSF model has a higher lower bound than the model without network constraints, and the phase angle model has an even higher lower bound since all line constraints are included.

Table 12: Centralized UC Solution Times

Case	No Net. Model		GSF Model		P. Ang. Model	
	t_{RR} (s)	$t_{1\%}$ (s)	t_{RR} (s)	$t_{1\%}$ (s)	t_{RR} (s)	$t_{1\%}$ (s)
A_1	3	625	26	648	32	3,975
A_2	3	870	34	1,566	26	–
A_3	12	6,466	170	26,520	952	–
B_1	2	315	31	822	133	–
B_2	2	76	29	598	108	–
B_3	10	4,161	131	9,068	2,164	–

Table 13: Cases A_1 and B_1 Decentralized Solutions

	Case A_1					Case B_1				
# Region	# Iter	t_{best} (s)	# Cycle	Cost ($\$ 10^6$)	Gap (%)	# Iter	t_{best} (s)	# Cycle	Cost ($\$ 10^6$)	Gap (%)
20	949	7,179	16.2	11.304	0.25	983	2,677	20.0	15.502	0.15
30	753	4,256	21.8	11.302	0.23	860	1,437	16.2	15.511	0.21
40	796	3,365	24.0	11.300	0.21	874	986	22.8	15.514	0.23
50	583	556	12.2	11.299	0.20	811	665	20.8	15.505	0.17
60	850	905	22.4	11.298	0.20	832	701	23.4	15.510	0.20
70	749	520	17.6	11.301	0.22	950	577	21.6	15.510	0.20
75	704	449	21.0	11.297	0.18	702	323	9.6	15.495	0.10
80	662	389	10.0	11.300	0.21	903	415	25.0	15.500	0.14
90	831	511	13.0	11.297	0.19	927	507	16.8	15.514	0.23
100	646	364	6.4	11.298	0.20	912	521	24.0	15.513	0.22
120	809	483	14.6	11.300	0.21	812	402	16.4	15.511	0.21
150	822	414	20.4	11.294	0.16	790	382	12.6	15.508	0.18
200	775	441	24.0	11.300	0.21	770	378	16.6	15.524	0.29

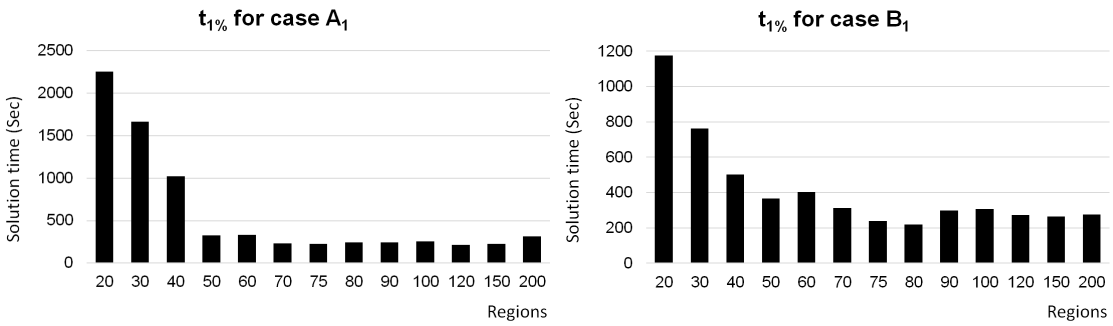

Figure 26: Solution time of cases A_1 and B_1.

We now describe the results of the DUC experiments. Five runs were conducted of each instance to normalize for occasional performance differences between machines recruited by the cluster job scheduler, which can alter observed solution times. The results in Table 13 and Figs. 26-28 are based on averages among five runs for each instance. Table 13 reflects the results of a parallel implementation of R&F to solve cases A_1 and B_1 with the number of regions varying from 20 to 200. In this table, "# Region" indicates the number of regions (partitions) for each instance. The columns labeled "# Iter", "t_{best}" and "# Cycle" represent the number of ADMM iterations, clock time in seconds spent to get the best solution, and the number of cycles between ADMM-Release and ADMM-Fix, respectively. All times reported are averages of actual wall clock times. This is an important consideration since the cluster is not solely dedicated to our problems. The "Cost" and "Gap" columns denote the augmented Lagrangian value and relative optimality gap of the best solution found in R&F, respectively. The optimality gap is based on the best lower bound (LB) obtained from the centralized UC solutions in Table 11 as $\text{Gap} = (\text{Cost} - \text{LB})/\text{Cost} \times 100$. Note that, if the UC problem cannot be solved centrally and there is no lower bound available from the central approach, we can use the optimal value at the end of ADMM-Relax as a weaker lower bound.

As shown in Table 13 for cases A_1 and B_1 with the number of regions $n \geq 50$, an optimality gap of 0.3% was reached in less that 10 minutes. Between 650 and

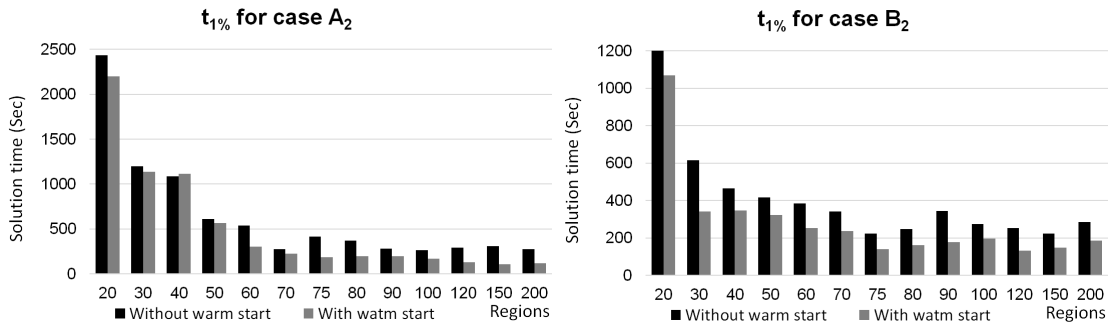

Figure 27: Solution time of cases A_2 and B_2.

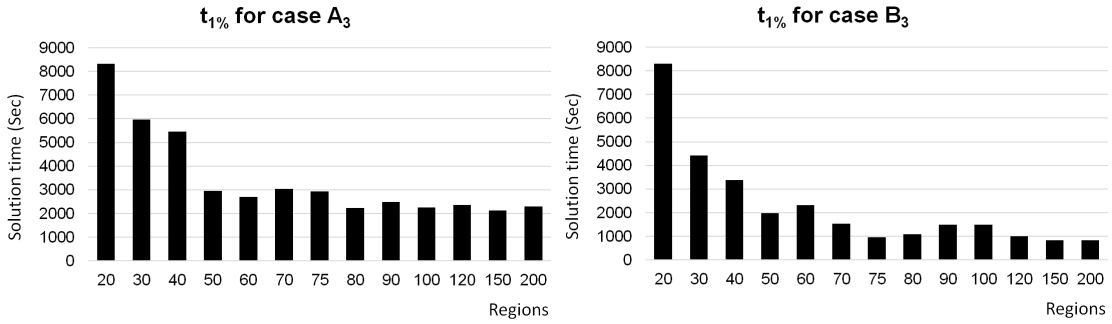

Figure 28: Solution time of cases A_3 and B_3.

1,000 iterations and 5 to 25 cycles of ADMM-Release and ADMM-Fix were needed in R&F. Decompositions into small numbers of regions generally did not perform well. Specifically, the 20 and 30 region instances of case A_1 took more than an hour to reach optimality gaps less than 0.3%, which the 150 region instance achieved in under 7 minutes. In most of the cases, in the first two cycles of ADMM-Release and ADMM-Fix, R&F was able to achieve 1% optimality gap. Figs. 26-28 depict wall clock time up to the point of attaining 1% optimality gap for all cases.

5.2.4 Concluding Remarks

In this work, the UC problem was formulated in a way suitable for the application of ADMM. The mathematical formulation, addition of new heuristics, and adaptation of parameters based on empirical observations were all studied, tested, and described. The contributions of this chapter are as follows:

1. Specification of R&F algorithm parameters and penalty terms providing good

performance.

2. Practical strategies for implementing R&F.

3. Experimental results demonstrating greatly superior solution time for DUC compared to centralized UC for large-scale power systems.

4. Demonstration of scalability in terms of system size and number of regions.

The experimental results show that, under a decentralized operations framework, a realistically-sized UC problem can be solved in reasonable time with each region sharing *only phase angle information at boundary buses*. Even in centralized operations, inter-ISO transactions and even near-real-time operations can benefit from a fast, scalable DUC methodology. In fact, the data-sharing requirements of such an application align well with the R&F procedure since no market participant cost data is directly exchanged. Two large ISOs would best leverage the large-scale test results by dividing their interior problems into many regions and then exchanging data on the bordering regions. Because of the poor behavior observed when a small number of large regions are used (see the 20-region test cases), formulating the problem as two ISO-sized subproblems would likely be an inefficient strategy.

Besides the enhancements to data privacy and multi-area coordination offered by R&F to DUC, the computational speed benefit can be exploited by a single ISO with no change in market or system control architecture. The potential speed increase is enough to open many possibilities, such as:

- Conducting multiple studies for various scenarios under uncertainty.
- Exploring stochastic or robust UC approaches.
- Extending the time horizon or granularity of UC studies.

Regarding the use of release-and-fix by ISOs, it should be noted that the procedure can be used in control centers without any change in the way that generating companies determine their optimal bids for production. Therefore, as a technique for ISOs to solve UC problems faster, our method does not pose any concerns for the general structure of electricity markets. However, as a method for conducting decentralized UC among the generating companies, load-serving entities, and network operators without central coordination by an ISO, it would imply an operational structure very different from today. Although our procedure is effectively a regional decomposition of the centralized problem, wherein generators with high cost would tend to be committed after generators with low cost, the role of the independent market monitor would have to be significantly expanded to mitigate the exercise of market power by some participants since the generator "bid" information would not leave the region.

To achieve a full security-constrained UC in a decentralized setting, some additional steps are needed. A method for conducting contingency analysis without global visibility does not yet exist but is needed even for today's operations. It remains to be determined whether a global contingency analysis can be conducted among decentralized regions that captures full $N-1$ system security. Joint security assessments between neighboring systems are an important topic of ongoing research. Further, the performance of R&F considering AC system constraints should be evaluated. It is hoped that the small subproblems attainable with the DUC formulation might enable reformulation with nonlinear voltage constraints.

It is noted that the applications of a decentralized MIP or MIQP solution methodology extend beyond UC and even power systems. Such an approach might be used to coordinate energy scheduling among a campus of buildings where the number of integer variables, representing states of load operation, may be larger than can practically be solved by other methods. It might also be used to coordinate energy scheduling among a neighborhood of homes aggregated as a demand response unit without

concentrating or communicating any data about actual appliance status within the homes.

The future electricity grid will require decentralized operations and control architectures for many reasons outlined in Section 4.1, and the R&F algorithm is designed with such requirements in mind. Given the experimental results obtained so far, it appears that R&F has the potential to solve large-scale energy scheduling problems with a highly distributed structure.

5.3 Exact DUC

In this section, we present numerical results testing the exact decentralized MIP Algorithms 6 and 7 on small UC instances. We used 6 small UC instances with 3, 4 and 5 generators for T=12 and 24 hours of planning. Table 14 presents details of these instances. In Table 14, "# Gen" and "Gen. types" denote the number and types of generator in each instance (see Table 8 for details of each generator type). The total system demand at each hour is determined as given in Table 9. The labels "# Bin. Vars.", "# Cont. Vars.", and "# Constr." denote the number of binary variables, continuous variables, and constraints, respectively, for each test case. Moreover, the columns z^{LP}, z^{IP}, "Duality Gap", and t_{C} represent optimal objective value of LP relaxation and MIP formulation for UC, relative duality gap in percentage (between z^{LP} and z^{IP}), and the solution time (in seconds) in central approach, respectively. An estimation for Lagrangian dual, which is obtained as the best lower bound in 100 iterations of the dual decomposition method, is denoted by $\tilde{z}^{\mathrm{LD}}$. Note that finding an optimal vector of dual variables in the dual decomposition algorithm is not guaranteed. Then, $\tilde{z}^{\mathrm{LD}}$ is not necessarily equal or close to the value of Lagrangian dual.

All algorithms were coded in C++ using CPLEX 12.6 through the Concert API. Central UC instances were solved using internal CPLEX multi-threading with four

Table 14: Test case details for exact algorithms

# Gen	Gen. types	T	# Bin. Vars.	# Cont. Vars.	# Constr.	z^{LP}	z^{IP}	Duality Gap (%)	t_{C} (Sec)	z^{LD}
3	6,7,8	24	216	144	891	139896	146403	4.44	0.03	139933
		12	108	72	435	68212	70945	3.85	0.09	68226
4	3,5,6,8	24	288	192	1244	207068	212771	2.68	0.14	207100
		12	144	96	596	101676	104381	2.59	0.09	101686
5	1,5,6,7,8	24	360	240	1514	354684	359197	1.26	0.22	354705
		12	180	120	722	171099	172994	1.10	0.11	171110

Table 15: Summary of the results for the exact Algorithm 7

# Gen	T	t_0	t_1	t^*	t_{all}	iter_1	iter^*	iter_{all}	# Feas.	# Cut
3	24	3.85	4.02	4.67	4.69	1	5	5	12	16
	12	2.16	2.23	2.59	3	1	4	7	12	19
4	24	4.54	4.7	7.4	193.1	1	11	118	90	530
	12	2.23	2.31	5.36	34.9	1	16	61	38	252
5	24	5.07	5.29	97.08	1621.72	1	42	190*	303	1004
	12	2.18	2.29	3.01	715.88	1	5	190*	290	962

cores. The step sizes ρ_μ, ρ_λ and ρ_γ were set to be 0.01, 10 and 50, respectively. The algorithms start with running ADMM to solve the LP relaxations of the UC instances to initialize the vector of dual variables $\boldsymbol{\mu}$ and the lower bound lb. Then, they do 100 iterations of the dual decomposition algorithm to improve the lower bound. Then, the main body of Algorithms 6 and 7 starts with 200 iterations limit where the first 10 iterations are spent on updating dual vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ without adding cuts. In each iteration, the lower bounding phase does 10 sub-iterations. Then, new candidate binary vectors are explored by the upper bounding procedure and cutoff from the feasible regions of all blocks.

Summary of the results for exact Algorithms 6 and 7 are presented in Tables 15 and 16, respectively. In Tables 15 and 16, t_0, t_1, t^*, and t_{all} are the estimated parallel times spent to initialize the algorithm, to find the first and best feasible solution, and to terminate the algorithm, respectively. The exact algorithms were initialized by

Table 16: Summary of the results for the exact Algorithm 7

# Gen	T	t_0	t_1	t^*	t_{all}	iter_1	iter^*	iter_{all}	# Feas.	# Cut
3	24	4.07	4.39	4.39	13.47	2	2	45	4	9
	12	2.35	2.49	2.49	3.97	1	1	11	4	7
4	24	4.56	4.81	4.81	6.81	2	2	10	12	23
	12	2.07	2.17	2.17	2.87	1	1	6	6	11
5	24	4.45	5.05	5.05	601.12	4	4	190*	84	373
	12	1.98	2.19	2.19	24.06	3	3	38	54	122

running ADMM for the LP relaxation and 100 iterations of the dual decomposition. "iter_1", "iter^*", "iter_{all}" are the corresponding number of iteration to t_1, t^*, and t_{all}, respectively. "# Feas.", "# Cut" are the number of feasible explored solutions and cuts (all explored binary solution), respectively.

For the 5 generator cases with T=24 and 12, Algorithm 6 terminated with % 1.078 and %0.911 optimality gaps after 190 iterations. For the 5 generator case with T=24, Algorithm 7 terminated with %0.671 optimality gap after 190 iterations. All other cases were solved to optimality. Based on the results in Tables 15 and 16, for most cases, Algorithm 7 outperforms Algorithm 6, in the sense that it requires less solution time (t_{all}), total number of iterations (iter_{all}) and cuts. It is worth mentioning that these exact algorithms are proof-of-concept implementations to verify possibility of obtaining the global optimal solutions of MIPs in a decentralized manner. Hence, the focus is not on computational times or number of iterations. With the current implementation and numerical results, the main advantage of Algorithms 6 and 7 is that they preserve data privacy for different blocks. Based on the results in Tables 14-16, these exact decentralized algorithms take much more time than the central approach. In particular, the solution times for Algorithms 6 and 7 are 3 seconds to 30 minutes while the central problems are solved in less than a second.

Chapter VI

CONCLUSIONS

In this dissertation, we investigated theory and application of decentralized optimization for loosely coupled mixed integer programming (MIP) problems. We developed decentralized optimization approaches based on Lagrangian and augmented Lagrangian duals for MIPs. The contributions of this dissertation are as follows:

- Proof of exactness of augmented Lagrangian dual (ALD) for MIPs;
- Decentralized exact and heuristic algorithms for MIPs;
- Application to decentralized unit commitment (UC).

6.1 Proof of Exactness of ALD for MIPs

For loosely coupled optimization problems, one possible decentralized approach is to relax the coupling constraints and solve the relaxed problem in parallel by decomposing it into subproblems. Unlike convex optimization problems, in general for MIP, a non-zero duality gap may exist when coupling constraints are relaxed by using classical Lagrangian or even augmented Lagrangian. We provided a primal characterization for ALD and proved that ALD is able to asymptotically achieve zero duality gap for MIPs, when the penalty coefficient is allowed to go to infinity. We also showed that, under some mild conditions, ALD using any norm as the augmenting function is able to close the duality gap of the MIP with a finite penalty coefficient.

One possible direction for future research is the application of ALD (without decomposition) for solving MIP problems. Iterative algorithms can be developed to optimally solve ALDs (with specific augmenting functions) for MIPs in a central framework. These algorithms may have computational advantages for some certain

classes of MIPs. For example, for a linear 0-1 problem with loosely coupled blocks of constraints, if the augmenting function is submodular, the augmented Lagrangian will be submodular, which is known to be easy to minimize. Developing theories and algorithms to solve ALD for MIPs in a distributed way is another future research direction with utmost practical importance. Even approximate decomposable methods in this context may produce better bounds than classical Lagrangian dual. Finally, investigating ALD for (convex) mixed integer nonlinear programs (MINLPs) can be pursued in future. Note that discreteness and nonlinearity in feasible MINLPs cause challenges such as lack of closedness and optimal solutions.

6.2 Decentralized Exact and Heuristic Algorithms for MIPs

Although ALD is able to close the duality gap, nonlinear objective functions in ALD destroy the decomposability which exists in classical Lagrangian dual for a MIP. A key challenge is that, because of the non-convex nature of MIPs, classical distributed and decentralized optimization approaches such as alternating direction method of multipliers (ADMM) cannot be applied directly to find their optimal solutions. We proposed three exact and one heuristic decentralized algorithms which are based on extensions of ADMM and dual decomposition techniques. The proposed heuristic method extends ADMM by periodically fixing and releasing binary variables to mitigate oscillations and traps in local optimality that result from the nonconvexity of MIPs. Our exact algorithms are based on adding primal cuts and restricting the Lagrangian relaxation of the original MIP problem. The exact algorithms evaluate the cost of the binary solutions as candidate partial solutions and refine them to get a primal feasible solution to the overall problem. To improve the lower bound and prevent cycling, the explored binary solutions are then cut-off from future consideration in all subproblems.

A possible direction for future research is to blend the speed of the proposed heuristic and precision of the exact decentralized algorithms. Another topic for future work is investigating stronger primal cuts to speed up the proposed exact methods. Moreover, the proposed methods can be improved for specific applications by exploiting the problem structures.

6.3 Application to Decentralized Unit Commitment

In the prosumer-based architecture for the future power grid, decentralized control and operation of the grid can play a significant role. Computational speed gain, data privacy, scalability, distributed operational databases, and multi-area coordination are the key motivations to use decentralized UC in power systems. First, we presented mathematical formulations for the UC problem, which are appropriate for the proposed decentralized algorithms. Privacy concerns of the participants in UC are taken into account in these formulations. Next, we proposed a solution approach for decentralized UC, which exploits the structure of UC in our decentralized algorithms. Finally, we presented extensive computational experiments for solving UC instances with different decentralized approaches. We demonstrated remarkable performance of parallel implementation of the heuristic decentralized algorithm to solve large scale UC instances on power systems of more than 3,000 buses. We also showed that for small UC instances, the proposed exact algorithms are able to find global optimal solutions.

In the context of UC, a possible future topic is application of the proposed decentralized algorithms for other variants of UC which have side constraints. For example, these algorithms can be applied and customized for emission-constrained, fuel-constrained, security-constrained, or stochastic UC problems to solve them in a faster and decentralized fashion.

Appendix A

NOTATION

A.1 Abbreviations

AC	Alternating Current
ALD	Augmented Lagrangian Dual
ALR	Augmented Lagrangian Relaxation
ADMM	Alternating Direction Method of Multipliers
CUC	Central Unit Commitment
DC	Direct Current
DUC	Decentralized Unit Commitment
GENCo	(Power) Generating Company
GSF	Generation Shift Factor
ISO	Independent System Operator
LD	Lagrangian Dual
LP	Linear Programming
LR	Lagrangian Relaxation
MIP	Mixed Integer (Linear) Programming
MIQP	Mixed Integer Quadratic Programming
MISO	Midcontinent Independent System Operator
NSR	Nonspinning Reserve
OPF	Optimal Power Flow
OR	Operating Reserve
PM	Pennsylvania-New Jersey-Maryland Interconnection
PTDF	Power Transfer Distribution Factor

QP	Quadratic Programming
R&F	Release and Fix
RHS	Right hand side
SR	Spinning Reserve
SCUC	Security Constrained Unit Commitment
UC	Unit Commitment

A.2 Nomenclature for Decentralized MIP Algorithms in Chapter 3

A.2.1 Input Data and Other Parameters

$\boldsymbol{\alpha}$	Vector of primal residuals
$\boldsymbol{\beta}$	Vector of dual residuals
$\boldsymbol{A}_\nu$	Matrix of coefficients for $\boldsymbol{x}_\nu$ in the coupling constraints
$\boldsymbol{A}$	Matrix of all $\boldsymbol{A}_\nu$
$\boldsymbol{b}$	RSH vector of coupling constraints
CR(.)	Continuous relaxation of a discrete set
$\boldsymbol{c}_\nu$	Vector of objective coefficients for $\boldsymbol{x}_\nu$
$\boldsymbol{E}_\nu$, $\boldsymbol{F}_\nu$, $\boldsymbol{g}_\nu$	Matrices of coefficients for $\boldsymbol{u}_\nu$, $\boldsymbol{y}_\nu$, and RHS vector, respectively, in description of $Y_\nu(\boldsymbol{u}_\nu)$
$\mathcal{L}_\nu(\cdot)$	Lagrangian function for block ν
$\mathcal{L}_\rho^+(\cdot)$	Augmented Lagrangian function
lb	Lower bound for z^{IP}
m	Number of coupling constraints
n	Number of decision variables for all blocks
n^1	Number of binary decision variables for all blocks
n^2	Number of continuous decision variables for all blocks
n_ν	Number of decision variables for block ν

n_ν^1	Number of binary decision variables for block ν
n_ν^2	Number of continuous decision variables for block ν
$\mathcal{P}$	Set of blocks, running from 1 to N
U_ν	Feasible set of $\boldsymbol{u}_\nu$
U	Cartesian product of $U_1, \cdots, U_N$
ub	Upper bound for z^{IP}
X_ν	Feasible set of $\boldsymbol{x}_\nu$
X	Cartesian product of $X_1, \cdots, X_N$
$\boldsymbol{x}^*$	Optimal solution of MIP (44)
$Y_\nu(\boldsymbol{u}_\nu)$	Feasible set of $\boldsymbol{y}_\nu$, for a given $\boldsymbol{u}_\nu$
$Y(\boldsymbol{u})$	Cartesian product of $Y_1(\boldsymbol{u}_1), \cdots, Y_N(\boldsymbol{u}_N)$
z^{LD}	Optimal value of Lagrangian dual for MIP (44)
z^{IP}	Optimal value of MIP (44)
z^{LP}	Optimal value of LP relaxation MIP (44)
$z^{\mathrm{LR}}(\boldsymbol{\mu})$	Optimal value of Lagrangian relaxation for MIP (44), for a given $\boldsymbol{\mu}$
$z(\hat{\boldsymbol{u}})$	Optimal value of MIP (44) when binary variables are fixed to be $\hat{\boldsymbol{u}}$
$\boldsymbol{\mu}$	Vector of dual multipliers for the coupling constraints
ρ	penalty coefficient in ADMM
ρ_μ^k	Step size for updating the dual vector $\boldsymbol{\mu}$ in iteration k
ρ_λ^k	Step size for updating the dual vector $\boldsymbol{\lambda}$ in iteration k
ρ_γ^k	Step size for updating the dual vector $\boldsymbol{\gamma}$ in iteration k

A.2.2 Decision variables

$\boldsymbol{u}_\nu$	Vector of binary decision variables for block ν
$\boldsymbol{u}$	Vector of all $\boldsymbol{u}_\nu$
$\boldsymbol{x}_\nu$	Vector of decision variables for block ν
$\boldsymbol{x}$	Vector of $\boldsymbol{x}_\nu$

$\boldsymbol{y}_\nu$	Vector of continuous decision variables for block ν
$\boldsymbol{y}$	Vector of all $\boldsymbol{y}_\nu$

A.3 Nomenclature for UC Problems

A.3.1 Sets and Indices

$\nu \in \mathcal{P}$	Regions, running from 1 to n
$\nu' \in \Delta_i$	Set of regions (including region ν) connected to bus $i \in \mathcal{N}_\nu^{\text{BB}}$
$g \in \mathcal{G}$	Generating units, running from 1 to G
$g \in \mathcal{G}_i$	Generating units located in bus i
$i \in \mathcal{N}$	Set of buses, running from 1 to N
$i \in \mathcal{N}_\nu$	Set of all buses in region ν
$i \in \mathcal{N}_\nu^{\text{BB}}$	Set of boundary buses in region ν
$i \in \mathcal{N}_\nu^{\text{FB}}$	Set of foreign buses not in region ν connected to some bus in region ν
$i \in \mathcal{N}_\nu^{\text{IB}}$	Set of internal buses in region ν
$ij \in \mathcal{E}$	Transmission lines, running from 1 to E
$ij \in \mathcal{E}'$	Set of monitored transmission lines
$j \in \delta_i$	Set of all buses connected to bus i
$t \in \mathcal{T}$	Hourly periods, running from 1 to T hours
$\boldsymbol{x}_\nu \in X_\nu$	Set of all feasible solutions for region ν

A.3.2 Input Data and Other Parameters

α^m	Vector of primal residuals at iteration m
β^m	Vector of dual residuals at iteration m
$\gamma_{ij,k}$	Power flow sensitivity of line ij with respect to injection transfer from bus k to slack bus
ϵ^{Dual}	Dual tolerance

ϵ^{Pri}	Primal tolerance
ϵ^{Obj}	Relative objective function tolerance
$\lambda^{m}_{\nu,i,t}$	Value of dual variable corresponding to deviation from $\bar{\tilde{\theta}}^{m}_{\nu,i,t}$, iteration m
$\mu^{m}_{\nu,i,t}$	Value of dual variable corresponding to deviation from $\bar{\tilde{F}}^{m}_{\nu,ij,t}$, iteration m
ρ	Penalty factor
$\tilde{\theta}^{m}_{\nu,i,t}$	Optimal value of $\tilde{\theta}_{\nu,i,t}$ at iteration m
$\bar{\tilde{\theta}}^{m}_{\nu,i,t}$	Average value for $\tilde{\theta}_{\nu,i,t}$ at iteration m
$\omega^{q,m}_{\nu t}$	Value of dual variable corresponding to deviation from $\tilde{r}^{q,m}_{\nu t}$
a_{gk}	Intercept of kth line segment for energy cost of unit g (\$)
b_{gk}	Slope of kth line segment for energy cost of unit g (\$/MWh)
B_{ij}	Element ij of DC power flow Jacobian
$C_{\nu}(.)$	Total cost function of region ν
C^{CS}_{g}	Cold startup cost of unit g (\$)
C^{HS}_{g}	Hot startup cost of unit g (\$)
C^{LV}_{g}	Linear energy cost of unit g (\$/MWh)
C^{NL}_{g}	No-load cost of unit g (\$/h)
C^{Q}_{g}	Quadratic energy cost of unit g (\$/MW^{2}h)
C^{SD}_{g}	Shutdown cost of unit g (\$)
CR(.)	Continuous relaxation of a discrete set
$d_{i,t}$	Expected load at bus i, time t
D_t	Total expected load at time t
$\overline{F}_{ij}$	Active power flow limit of branch ij
$\tilde{F}^{m}_{\nu,ij,t}$	Optimal value of $\tilde{F}_{\nu,ij,t}$ at iteration m
$\bar{\tilde{F}}^{m}_{\nu,ij,t}$	Average value for $\tilde{F}_{\nu,ij,t}$ at iteration m
$\mathcal{L}_{\rho,\nu}(.)$	Augmented Lagrangian of region ν

$\overline{P}_g$ Maximum power output of unit g (MW)

$\underline{P}_g$ Minimum power output of unit g (MW)

p_g^{Init} Initial power output of unit g above the minimum output $\underline{P}_g$ (MW)

$\hat{r}_\nu^q$ Reserve requirement for region ν, product q (MW)

$\hat{r}_{\text{sys}}^q$ System reserve requirement for product q (MW)

RD_g Ramp-down rate of unit g (MW/h)

RU_g Ramp-up rate of unit g (MW/h)

SD_g Shutdown capability of unit g (MW)

SU_g Startup capability of unit g (MW)

T_g^{CS} Cold startup time of unit g (h)

T_g^{Init} Number of hours unit g has been online (+) or offline (-) prior to the first period of the commitment study

TD_g Minimum downtime of unit g (h)

TD_g^0 Initial minimum downtime of unit g (h)

TU_g Minimum uptime of unit g (h)

TU_g^0 Initial minimum uptime of unit g (h)

t_{RR} Computational time spent solving the root relaxation node for centralized cases (s)

$t_{1\%}$ Computational time spent finding a solution with 1% relative optimality gap (s)

t_{best} Computational time spent finding the best binary-feasible solution (s)

A.3.3 Decision variables

$\theta_{i,t}$ Voltage phase angle at bus i, time t

$\tilde{\theta}_{\nu,i,t}$ Perception by region ν of voltage phase angle at bus $i \in \mathcal{N}_\nu^{\text{BB}} \bigcup \mathcal{N}_\nu^{\text{FB}}$, time t

$\tilde{c}_{gt}$ Approximated quadratic cost of unit g at hour t (\$)

$F_{ij,t}$ Power flow from bus i to bus j at time t

$\tilde{F}_{\nu,ij,t}$ Perception by region ν of power flow from bus $i \in \mathcal{N}_{\nu}^{\mathrm{BB}}$ to bus $j \in \mathcal{N}_{\nu}^{\mathrm{FB}}$ at time t

$P_{i,t}^{\mathrm{net}}$ Total real power injection from generators and loads at bus i, time t

p_{gt} Power output of unit g at hour t above the minimum output $\underline{P}_g$ (MW)

r_{gt}^{q} Amount of reserve product q allocated to unit g at time t (MW)

u_{gt} Commitment status of unit g at hour t, equal to 1 if the unit is online and 0 if offline

v_{gt} Startup status of unit g, which takes the value of 1 if the unit starts up in hour t and 0 otherwise

v_{gt}^{HS} Hot startup status of unit g, which takes the value of 1 in the hour t if the unit starts up and shuts down in the interval $[t - T_g^{\mathrm{CS}} - TD_g, t-1]$

w_{gt} Shutdown status of unit g, which takes the value of 1 if the unit shuts down in hour t and 0 otherwise

$\boldsymbol{x}_{\nu}$ Vector of all decision variables in region ν, including variables u_{gt}, v_{gt}, v_{gt}^{HS}, w_{gt}, and p_{gt} of generators $g \in \mathcal{G}_{\nu}$, variables $\theta_{i,t}(.)$ for internal buses $i \in \mathcal{N}_{\nu}^{\mathrm{IB}}$, and $\tilde{\theta}_{\nu,i,t}$ of all buses $i \in \mathcal{N}_{\nu}^{\mathrm{BB}} \bigcup \mathcal{N}_{\nu}^{\mathrm{FB}}$, for all $t \in \mathcal{T}$

Appendix B

SOME DEFINITIONS

Definition B.1. *(Lower limits and lower semicontinuity, [148]). The lower limit of a function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *at* $\overline{\boldsymbol{x}}$ *is the value in* $\overline{\mathbb{R}}$ *defined by*

$$\liminf_{\boldsymbol{x}\to\overline{\boldsymbol{x}}} f(\boldsymbol{x}) := \lim_{\delta\downarrow 0}\left[\inf_{\boldsymbol{x}\in\mathbb{B}(\overline{\boldsymbol{x}},\delta)} f(\boldsymbol{x})\right] = \sup_{\delta\downarrow 0}\left[\inf_{\boldsymbol{x}\in\mathbb{B}(\overline{\boldsymbol{x}},\delta)} f(\boldsymbol{x})\right]. \tag{122}$$

The function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is lower semicontinuous at* $\overline{\boldsymbol{x}}$ *if* $\liminf_{\boldsymbol{x}\to\overline{\boldsymbol{x}}} f(\boldsymbol{x}) \geq f(\overline{\boldsymbol{x}})$ *or equivalently* $\liminf_{\boldsymbol{x}\to\overline{\boldsymbol{x}}} f(\boldsymbol{x}) = f(\overline{\boldsymbol{x}})$.

Lower semicontinuity of f on $\mathbb{R}^n$ is equal to the closedness of epigraph of f in $\mathbb{R}^n \times \mathbb{R}$.

Definition B.2. *(Dualizing parameterization function, [87]). Consider the primal optimization problem* $\inf_{\boldsymbol{x}\in\mathbb{R}^n} \phi(\boldsymbol{x})$, *where* $\phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is an extended real-valued function. A function* $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ *is said to be a dualizing parameterization function for* ϕ *if* $\phi(\boldsymbol{x}) = f(\boldsymbol{x}, \mathbf{0})$, $\forall \boldsymbol{x} \in \mathbb{R}^n$.

Definition B.3. *(Level boundedness, [148]). Let* $X \subset \mathbb{R}^n$ *be a closed subset and* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is an extended real-valued function. The function* f *is said to be level bounded on* X *if, for any* $\alpha \in \mathbb{R}$, *the set* $\{\boldsymbol{x} \in X : f(\boldsymbol{x}) \leq \alpha\}$ *is bounded.*

Definition B.4. *(Generalized augmenting function, [87]). A function* $\sigma : \mathbb{R}^m \to \mathbb{R}_+ \cup \{+\infty\}$ *is said to be a generalized augmenting function if it is proper, lower semicontinuous, level-bounded on* $\mathbb{R}^m$, $\arg\min_{\boldsymbol{y}} \sigma(\boldsymbol{y}) = \{\mathbf{0}\}$, *and* $\sigma(\mathbf{0}) = 0$.

Definition B.5. *(Almost peak at zero, Definition 2.2 in [149]). Let* U *be a certain set. A family* $(\mu_u)_{u\in U}$ *of continuous functions* μ *defined on a normed space* Z *is called an almost peak at zero one if* $\mu_u(\mathbf{0}) = 0$ *for all* $u \in U$, *and for each* $\delta > 0$ *and* $\epsilon > 0$ *there exists* $u \in U$, $\epsilon' \in (0, \epsilon)$ *and* $r > 0$ *such that*

i) $\mu_u(\boldsymbol{z}) \leq \epsilon'/r$ *if* $\|\boldsymbol{z}\| < \delta$;

ii) $\mu_u(\boldsymbol{z}) \leq (\epsilon' - 1)/r$ *if* $\|\boldsymbol{z}\| \geq \delta$;

Definition B.6. *(Peak at zero, Definition 2.3 in [149]).*

1. *A family* $(\mu_u)_{u \in U}$ *of continuous functions* μ *defined on a normed space* Z *is called a peak at zero if*

   i) $\mu_u(\boldsymbol{z}) \leq 0 = \mu_u(\mathbf{0})$ *for all* $u \in U$ *and* $\boldsymbol{z} \in Z$;

   ii) for each $\delta > 0$ *there exists* $u \in U$ *such that* $\sup_{\|\boldsymbol{z}\| \geq \delta} \mu_u(\boldsymbol{z}) < 0$.

2. *A continuous functions* μ *defined on* Z *is called a peak at zero if*

   i) $\mu(\boldsymbol{z}) < 0 = \mu(\mathbf{0})$ *for all* $\boldsymbol{z} \neq 0$;

   ii) $\sup_{\|\boldsymbol{z}\| \geq \delta} \mu(\boldsymbol{z}) < 0$ *for all* $\delta > 0$.

Definition B.7. *(Weak peak at zero, Assumption 2 in [125]). A function* σ *satisfies a weak peak at zero property if:*

a) $\sigma(\boldsymbol{u}) \geq 0$, *for all* $\boldsymbol{u}$;

b) For any given sequence $\{\boldsymbol{u}_k\} \subset \mathbb{R}^m$, *if* $\sigma(\boldsymbol{u}_k) \to 0$, *then* $\boldsymbol{u}_k^+ \to 0$.

Note that the above property (b) is equivalent to the following condition:

$$\inf_{\{\boldsymbol{u} | \operatorname{dist}(\boldsymbol{u}, \mathbb{R}^m_-) \geq \delta\}} \sigma(\boldsymbol{u}) > 0, \quad \forall \delta > 0.$$

Definition B.8. *(Generalized peak at zero, [188]). Consider a continuous, not identically equal to 0 and nondecreasing function* $e : \mathbb{R} \to \mathbb{R}$ *with* $e(0) = 0$. *The variable substitution* $\alpha(\boldsymbol{u})$ *for* $\boldsymbol{u}$ *is defined as*

$$\alpha(\boldsymbol{u}) := (e(u_1), \cdots, e(u_m)).$$

$U(r)$ *Denotes the effective domain of the nonlinear augmenting penalty function with respect to the dual variable* $\boldsymbol{u}$. *Consider the nonlinear augmenting penalty function*

$\rho : U(r) \times V \times \mathbb{R}_{++} \to \mathbb{R}$. *$\rho$ satisfies a general peak at zero condition if* $\forall \epsilon > 0$, *and* $(\boldsymbol{v}, \bar{r}) \in V \times \mathbb{R}_{++}$,

$$\inf\{\rho(\boldsymbol{u}, \boldsymbol{v}, r) - \rho(\boldsymbol{u}, \boldsymbol{v}, \bar{r}) | u \in U(r), \|\alpha(\boldsymbol{u})\| \geq \epsilon\} > 0, \forall r > \bar{r}.$$

Definition B.9. *(Coercivity properties, Definition 3.25 in [148]). A function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is level-coercive if it is bounded below on bounded sets and satisfies*

$$\liminf_{\|\boldsymbol{x}\| \to \infty} \frac{f(\boldsymbol{x})}{\|\boldsymbol{x}\|} > 0 \tag{123}$$

whereas it is coercive if it is bounded below on bounded sets and

$$\liminf_{\|\boldsymbol{x}\| \to \infty} \frac{f(\boldsymbol{x})}{\|\boldsymbol{x}\|} = \infty \tag{124}$$

For any proper, lower semicontinuous function f on $\mathbb{R}^n$, level coercivity implies level boundedness. When f is convex the two properties are equivalent [148].

Definition B.10. *(Abstract convexity, Definition 2.1 in [149]). Let Z be a set and H be a set of finite functions defined on Z. For two functions f and h defined on Z, the notation $h \leq f$ means that $h(\boldsymbol{z}) \leq f(\boldsymbol{z})$ for all $\boldsymbol{z} \in Z$. Let $f : Z \to \overline{\mathbb{R}}$.*

1. *The set $supp(f, H) = \{h \in H : h \leq f\}$ is called the support set of f.*
2. *The function $co_H f : Z \to \overline{\mathbb{R}}$, defined by $co_H f(\boldsymbol{z}) = \sup\{h(\boldsymbol{z}) : h \in supp(f, H)\}$, $\boldsymbol{z} \in Z$, is called the H-convex hull of f.*
3. *f is called abstract convex with respect to H, or H-convex if $f(\boldsymbol{z}) = co_H f(\boldsymbol{z})$ for all $\boldsymbol{z} \in Z$.*

Classical convexity is equivalent to abstract convexity with respect to the set of continuous affine functions [32].

Definition B.11. *(Banach space, [14]). A Banach space is a vector space X over the field real or complex numbers, which is equipped with a norm and which is complete with respect to that norm.*

REFERENCES

[1] Abe, R., Taoka, H., and McQuilkin, D., "Digital grid: Communicative electrical grids of the future," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 399–410, 2011.

[2] Afkousi-Paqaleh, M., Rashidinejad, M., and Pourakbari-Kasmaei, M., "An implementation of harmony search algorithm to unit commitment problem," *Electrical Engineering*, vol. 92, no. 6, pp. 215–225, 2010.

[3] Ahmed, S., "A scenario decomposition algorithm for 0–1 stochastic programs," *Operations Research Letters*, vol. 41, no. 6, pp. 565–569, 2013.

[4] Aminifar, F., Fotuhi-Firuzabad, M., and Shahidehpour, M., "Unit commitment with probabilistic spinning reserve and interruptible load considerations," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 388–397, 2009.

[5] Amjady, N. and Ansari, M. R., "Hydrothermal unit commitment with AC constraints by a new solution method based on benders decomposition," *Energy Conversion and Management*, vol. 65, no. 0, pp. 57 – 65, 2013.

[6] Amjady, N. and Nasiri-Rad, H., "Security constrained unit commitment by a new adaptive hybrid stochastic search technique," *Energy Conversion and Management*, vol. 52, no. 2, pp. 1097–1106, 2011.

[7] Andrianesis, P., Liberopoulos, G., Biskas, P., and Bakirtzis, A., "Medium-term unit commitment in a pool market," in *Energy Market (EEM), 2011 8th International Conference on the European*, pp. 461–466, May 2011.

[8] Angheliţă, P., Chefneux, M., Balaban, R., and Trocan, L., "Energy storage systems for buildings equipped with photovoltaic cells," in *Electrical and Electronics Engineering (ISEEE), 2010 3rd International Symposium on*, pp. 332–335, September 2010.

[9] Angulo, G., Ahmed, S., Dey, S. S., and Kaibel, V., "Forbidden vertices," *Mathematics of Operations Research*, vol. 40, no. 2, pp. 350–360, 2015.

[10] Aoki, K., Itoh, M., Satoh, T., Nara, K., and Kanezashi, M., "Optimal long-term unit commitment in large scale systems including fuel constrained thermal and pumped-storage hydro," *IEEE Transactions on Power Systems*, vol. 4, no. 3, pp. 1065–1073, 1989.

[11] Bakirtzis, A. G. and Biskas, P. N., "A decentralized solution to the DC-OPF of interconnected power systems," *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1007–1013, 2003.

[12] Barth, R., Brand, H., Meibom, P., and Weber, C., "A stochastic unit-commitment model for the evaluation of the impacts of integration of large amounts of intermittent wind power," in *Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on*, pp. 1–8, IEEE, 2006.

[13] Batut, J. and Renaud, A., "Daily generation scheduling optimization with transmission constraints: a new class of algorithms," *IEEE Transactions on Power Systems*, vol. 7, no. 3, pp. 982–989, 1992.

[14] Beauzamy, B., *Introduction to Banach spaces and their geometry*, vol. 68. Elsevier, 2011.

[15] Ben-Tal, A. and Zibulevsky, M., "Penalty/barrier multiplier methods for convex programming problems," *SIAM Journal on Optimization*, vol. 7, no. 2, pp. 347–366, 1997.

[16] Benders, J. F., "Partitioning procedures for solving mixed-variables programming problems," *Numerische mathematik*, vol. 4, no. 1, pp. 238–252, 1962.

[17] Bertsekas, D. P., *Constrained optimization and Lagrange multiplier methods*. Academic press, 1982.

[18] Bertsekas, D. P., *Nonlinear programming*. Athena Scientific, 1999.

[19] Bertsekas, D. P., Nedić, A., and Ozdaglar, A. E., *Convex analysis and optimization*. Athena Scientific Belmont, 2003.

[20] Bertsimas, D., Litvinov, E., Sun, X., Zhao, J., and Zheng, T., "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 52–63, 2013.

[21] Bienstock, D. and Verma, A., "The n-k problem in power grids: New models, formulations, and numerical experiments," *SIAM Journal on Optimization*, vol. 20, no. 5, pp. 2352–2380, 2010.

[22] Bjelogrlic, M. R., "Inclusion of combined cycle plants into optimal resource scheduling," in *IEEE Power Engineering Society Summer Meeting*, vol. 1, pp. 189–194, IEEE, 2000.

[23] Blair, C. and Jeroslow, R., "The value function of a mixed integer program: I," *Discrete Mathematics*, vol. 19, no. 2, pp. 121–138, 1977.

[24] Blair, C. and Jeroslow, R., "The value function of a mixed integer program: II," *Discrete Mathematics*, vol. 25, no. 1, pp. 7–19, 1979.

[25] Blair, C. and Jeroslow, R., "The value function of an integer program," *Mathematical Programming*, vol. 23, no. 1, pp. 237–273, 1982.

[26] Boland, N. L. and Eberhard, A. C., "On the augmented Lagrangian dual for integer programming," *Mathematical Programming*, vol. 150, no. 2, pp. 491–509, 2015.

[27] Borbely, A. M. and Kreider, J. F., *Distributed Generation: The Power Paradigm for the New Millennium*. Mechanical Engineering Series, Taylor & Francis, 2010.

[28] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[29] Burachik, R. S., Iusem, A. N., and Melo, J. G., "The exact penalty map for nonsmooth and nonconvex optimization," *Optimization*, vol. 64, no. 4, pp. 717–738, 2015.

[30] Burachik, R. S., Iusem, A. N., and Melo, J. G., "Duality and exact penalization for general augmented lagrangians," *Journal of optimization theory and applications*, vol. 147, no. 1, pp. 125–140, 2010.

[31] Burachik, R. S. and Rubinov, A., "On the absence of duality gap for lagrange-type functions," *Journal of Industrial and Management Optimization*, vol. 1, no. 1, p. 33, 2005.

[32] Burachik, R. S. and Rubinov, A., "Abstract convexity and augmented Lagrangians," *SIAM Journal on Optimization*, vol. 18, no. 2, pp. 413–436, 2007.

[33] Burke, J. V., "Calmness and exact penalization," *SIAM Journal on Control and Optimization*, vol. 29, no. 2, pp. 493–497, 1991.

[34] Burke, J. V., "An exact penalization viewpoint of constrained optimization," *SIAM Journal on Control and Optimization*, vol. 29, no. 4, pp. 968–998, 1991.

[35] Campoccia, A., Dusonchet, L., Telaretti, E., and Zizzo, G., "Economic impact of ice thermal energy storage systems in residential buildings in presence of double-tariffs contracts for electricity," in *Energy Market, 2009. EEM 2009. 6th International Conference on the European*, pp. 1–5, May 2009.

[36] Carøe, C. C. and Schultz, R., *A two-stage stochastic program for unit commitment under uncertainty in a hydro-thermal power system*. ZIB, 1998.

[37] Carpentier, P., Gohen, G., Culioli, J.-C., and Renaud, A., "Stochastic optimization of unit commitment: a new decomposition framework," *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 1067–1073, 1996.

[38] Carrión, M. and Arroyo, J. M., "A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1371–1378, 2006.

[39] Cerisola, S., Baíllo, Á., Fernández-López, J. M., Ramos, A., and Gollmer, R., "Stochastic power generation unit commitment in electricity markets: A novel formulation and a comparison of solution methods," *Operations Research*, vol. 57, no. 1, pp. 32–46, 2009.

[40] Chang, C. W. and Waight, J. G., "A mixed integer linear programming based hydro unit commitment," in *Power Engineering Society Summer Meeting, 1999. IEEE*, vol. 2, pp. 924–928, 1999.

[41] Chang, G., Tsai, Y., Lai, C., and Chung, J., "A practical mixed integer linear programming based approach for unit commitment," in *Power Engineering Society General Meeting, 2004. IEEE*, pp. 221–225, IEEE, 2004.

[42] Chen, C., He, B., Ye, Y., and Yuan, X., "The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent," *Mathematical Programming*, pp. 1–23, 2015.

[43] Chen, Y., Wang, X., and Wang, Q., "Overcoming computational challenges on large-scale security constrained unit commitment (SCUC) problems," in *FERC Technical Conference on increasing Real-Time and Day-Ahead Market Efficiency through Improved Software*, June 2014.

[44] Chung, K., Kim, B., and Hur, D., "Multi-area generation scheduling algorithm with regionally distributed optimal power flow using alternating direction method," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 9, pp. 1527–1535, 2011.

[45] Clarke, F. H., "A new approach to Lagrange multipliers," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 165–174, 1976.

[46] Cohen, A. I. and Ostrowski, G., "Scheduling units with multiple operating modes in unit commitment," in *IEEE Power Industry Computer Application Conference*, pp. 494–500, IEEE, 1995.

[47] Conejo, A. J., Castillo, E., García-Bertrand, R., and Mínguez, R., *Decomposition techniques in mathematical programming: engineering and science applications*. Springer Berlin, 2006.

[48] Constantinescu, E., Zavala, V., Rocklin, M., Lee, S., and Anitescu, M., "Unit commitment with wind power generation: integrating wind forecast uncertainty and stochastic programming," tech. rep., Argonne National Laboratory (ANL), 2009.

[49] Constantinescu, E. M., Zavala, V. M., Rocklin, M., Lee, S., and Anitescu, M., "A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 431–441, 2011.

[50] Correia, P., "Decentralised unit commitment in a market structure: problem formulation and solution advancement," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 153, no. 1, pp. 121–126, 2006.

[51] Daneshi, H., Choobbari, A. L., Shahidehpour, M., and Li, Z., "Mixed integer programming method to solve security constrained unit commitment with restricted operating zone limits," in *Electro/Information Technology, 2008. EIT 2008. IEEE International Conference on*, pp. 187–192, IEEE, 2008.

[52] Dantzig, G. B. and Wolfe, P., "Decomposition principle for linear programs," *Operations research*, vol. 8, no. 1, pp. 101–111, 1960.

[53] Dantzig, G. B., *Linear programming and extensions*. RAND Corporation, 1963.

[54] Dasgupta, D. and McGregor, D. R., "Thermal unit commitment using genetic algorithms," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 141, no. 5, pp. 459–465, 1994.

[55] Dentcheva, D., Gollmer, R., Möller, A., Römisch, W., and Schultz, R., "Solving the unit commitment problem in power generation by primal and dual methods," in *IN: PROGRESS IN INDUSTRIAL MATHEMATICS AT ECMI 96 (M. BR NS, MP BENDS E AND MP S RENSEN EDS.), TEUBNER*, Citeseer, 1997.

[56] Dillon, T. S., Edwin, K. W., Kochs, H.-D., and Taud, R., "Integer programming approach to the problem of optimal unit commitment with probabilistic reserve determination," *IEEE Transactions on Power Apparatus and Systems*, no. 6, pp. 2154–2166, 1978.

[57] Donnelly, M., Ingram, M., and Carroll, J. R., "Eastern interconnection phasor project," in *Power Systems Conference and Exposition*, pp. 336–342, 2006.

[58] Everett, H., "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations research*, vol. 11, no. 3, pp. 399–417, 1963.

[59] Ezzati, S., Yousefi, G., Pedram, M., and Baghdadi, M., "Security-constrained unit commitment based on hybrid benders decomposition and mixed integer non-linear programming," in *Energy Conference and Exhibition (EnergyCon), 2010 IEEE International*, pp. 233–237, IEEE, 2010.

[60] Feizollahi, M. J. and Costley, M., "DUC test cases," 2014. [Online]. Available: http://www.ece.gatech.edu/research/labs/aces/pages/ductestdata.zip.

[61] Feizollahi, M. J., Costley, M., Ahmed, S., and Grijalva, S., "Large-scale decentralized unit commitment," *International Journal of Electric Power & energy Systems*, vol. 73, no. 1, pp. 97–106, 2015.

[62] Finardi, E. C. and da Silva, E. L., "Solving the hydro unit commitment problem via dual decomposition and sequential quadratic programming," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 835–844, 2006.

[63] Fischer, R. and Serra, P., "Energy prices in the presence of plant indivisibilities," *Energy Economics*, vol. 25, no. 4, pp. 303–314, 2003.

[64] Fu, Y. and Shahidehpour, M., "Fast SCUC for large-scale power systems," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2144–2151, 2007.

[65] Fu, Y., Shahidehpour, M., and Li, Z., "Long-term security-constrained unit commitment: hybrid Dantzig-Wolfe decomposition and subgradient approach," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 2093–2106, 2005.

[66] Fu, Y., Shahidehpour, M., and Li, Z., "Security-constrained unit commitment with ac constraints," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1001–1013, 2005.

[67] Gadonneix, P., de Castro, F. B., de Medeiros, N. F., Drouin, R., Jain, C., Kim, Y. D., Ferioli, J., Marie-JoséNadeau, Sambo, A., Teyssen, J., Naqi, A. A., Graham Ward, C., Guobao, Z., and Frei, C., "2010 survey of energy resources." World Energy Council, 2010.

[68] Gaing, Z.-L., "Discrete particle swarm optimization algorithm for unit commitment," in *Power Engineering Society General Meeting, 2003, IEEE*, vol. 1, IEEE, 2003.

[69] Galati, M., *Decomposition methods for integer linear programming.* PhD thesis, Lehigh University, 2010.

[70] Galiana, F., "Bound estimates of the severity of line outages in power system contingency analysis and ranking," *IEEE Transactions on Power Apparatus and Systems*, no. 9, pp. 2612–2624, 1984.

[71] Galus, M. D. and Andersson, G., "Demand management of grid connected plug-in hybrid electric vehicles (PHEV)," in *Energy 2030 Conference, 2008. ENERGY 2008. IEEE*, pp. 1 –8, November 2008.

[72] Gatterbauer, W., "Economic efficiency of decentralized unit commitment from a generators perspective," in *Engineering Electricity Services of the Future*, Kluwer Academic Publishers (Springer Verlag), 2010.

[73] Geoffrion, A. M., "Lagrangean relaxation for integer programming," *Mathematical Programming Study*, vol. 2, pp. 82–114, 1974.

[74] Gjengedal, T., "Emission constrained unit-commitment (ecuc)," *IEEE Transactions on Energy Conversion*, vol. 11, no. 1, pp. 132–138, 1996.

[75] Granada E, M., Rider, M. J., Mantovani, J. R. S., and Shahidehpour, M., "Decentralized AC power flow for real-time multi-TSO power system operation," in *IEEE Power & Energy Society General Meeting*, pp. 1–7, 2010.

[76] Grijalva, S., Costley, M., and Ainsworth, N., "Prosumer-based control architecture for the future electricity grid," in *Control Applications (CCA), 2011 IEEE International Conference on*, pp. 43–48, 2011.

[77] Grijalva, S. and Tariq, M. U., "Prosumer-based smart grid architecture enables a flat, sustainable electricity industry," in *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, pp. 1–6, January 2011.

[78] Grijalva, S., "Research needs in multi-dimentional, multi-sacle modeling and algorithms for next generation electricity grids," in *Computational Needs for Next Generation Electric Power Grid Proceedings, 2011 DOE Workshop on*, pp. 5–1 to 5–32, April 2011.

[79] Guan, X., Zhai, Q., and Papalexopoulos, A., "Optimization based methods for unit commitment: Lagrangian relaxation versus general mixed integer programming," in *Power Engineering Society General Meeting, 2003, IEEE*, vol. 2, IEEE, 2003.

[80] Guy, J., "Security constrained unit commitment," *IEEE Transactions on Power Apparatus and Systems*, no. 3, pp. 1385–1390, 1971.

[81] Hedman, K. W., O'Neill, R. P., and Oren, S. S., "Analyzing valid inequalities of the generation unit commitment problem," in *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, pp. 1–6, IEEE, 2009.

[82] Helman, U. and Hobbs, B. F., "Large-scale market power modeling: Analysis of the U.S. eastern interconnection and regulatory applications," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1434–1448, 2010.

[83] Heredia, F. J., Rider, M. J., and Corchero, C., "Optimal bidding strategies for thermal and combined cycle units in the day-ahead electricity market with bilateral contracts," in *IEEE Power & Energy Society General Meeting*, pp. 1–6, 2009.

[84] Higham, N. J., *Accuracy and stability of numerical algorithms*. Philadelphia, PA: SIAM, 2nd ed., 2002.

[85] Honarvar Nazari, M., Costello, Z., Feizollahi, M. J., Grijalva, S., and Egerstedt, M., "Distributed frequency control of prosumer-based electric energy systems," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2934 – 2942, 2014.

[86] Huang, C.-L. and others, "Application of genetic-based neural networks to thermal unit commitment," *IEEE Transactions on Power Systems*, vol. 12, no. 2, pp. 654–660, 1997.

[87] Huang, X. and Yang, X., "A unified augmented lagrangian approach to duality and exact penalization," *Mathematics of Operations Research*, vol. 28, no. 3, pp. 533–552, 2003.

[88] Hubert, T. and Grijalva, S., "Modeling for residential electricity optimization in dynamic pricing environments," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 2224–2231, 2012.

[89] Hui, H., Yu, C.-N., and Moorty, S., "Reliability unit commitment in the new ERCOT nodal electricity market," in *IEEE Power & Energy Society General Meeting*, pp. 1–8, 2009.

[90] Jiang, R., Wang, J., and Guan, Y., "Robust unit commitment with wind power and pumped storage hydro," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 800–810, 2012.

[91] Katiraei, F. and Iravani, M. R., "Power management strategies for a microgrid with multiple distributed generation units," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1821–1831, 2006.

[92] Kempton, W. and Tomić, J., "Vehicle-to-grid power fundamentals: Calculating capacity and net revenue," *Journal of Power Sources*, vol. 144, no. 1, pp. 268–279, 2005.

[93] Kempton, W. and Tomić, J., "Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy," *Journal of Power Sources*, vol. 144, no. 1, pp. 280–294, 2005.

[94] Kim, B. H. and Baldick, R., "A comparison of distributed optimal power flow algorithms," *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 599–604, 2000.

[95] Kim, B. and Baldick, R., "Coarse-grained distributed optimal power flow," *IEEE Transactions on Power Systems*, vol. 12, no. 2, pp. 932–939, 1997.

[96] Krall, E., Higgins, M., and O'Neill, R. P., "RTO unit commitment test system," tech. rep., Federal Energy Regulatory Commission FERC, 2012. [Online]. Available: http://www.ferc.gov/legal/staff-reports/rto-COMMITMENT-TEST.pdf.

[97] Kuloor, S., Hope, G., and Malik, O., "Environmentally constrained unit commitment," in *Generation, Transmission and Distribution, IEE Proceedings C*, vol. 139, pp. 122–128, IET, 1992.

[98] Lauer, G., Sandell, N., Bertsekas, D., and Posbergh, T., "Solution of large-scale optimal unit commitment problems," *IEEE Transactions on Power Apparatus and Systems*, no. 1, pp. 79–86, 1982.

[99] Lee, F. N., "A fuel-constrained unit commitment method," *IEEE Transactions on Power Systems*, vol. 4, no. 3, pp. 1208–1218, 1989.

[100] Lee, F. N., "The coordination of multiple constrained fuels," *IEEE transactions on power systems*, vol. 6, no. 2, pp. 699–707, 1991.

[101] Li, C., Svoboda, A. J., Xiaohong, G., and Singh, H., "Revenue adequate bidding strategies in competitive electricity markets," *IEEE Transactions Power Systems*, vol. 14, no. 2, pp. 492–497, 1999.

[102] Li, C.-A., Svoboda, A. J., Tseng, C.-L., Johnson, R. B., and Hsu, E., "Hydro unit commitment in hydro-thermal optimization," *IEEE Transactions on Power Systems*, vol. 12, pp. 764–769, May 1997.

[103] Li, M. and Luh, P. B., "A decentralized framework of unit commitment for future power markets," in *IEEE Power & Energy Society General Meeting*, pp. 1–5, 2013.

[104] Li, T. and Shahidehpour, M., "Price-based unit commitment: a case of lagrangian relaxation versus mixed integer programming," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 2015–2025, 2005.

[105] Li, Z. and Shahidehpour, M., "Security-constrained unit commitment for simultaneous clearing of energy and ancillary services markets," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1079–1088, 2005.

[106] Lo, C.-H. and Ansari, N., "Decentralized controls and communications for autonomous distribution networks in smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 66–77, 2013.

[107] Lu, B. and Shahidehpour, M., "Short-term scheduling of combined cycle units," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1616–1625, 2004.

[108] Lu, B. and Shahidehpour, M., "Unit commitment with flexible generating units," *Power Systems, IEEE Transactions on*, vol. 20, no. 2, pp. 1022–1034, 2005.

[109] Luo, Z.-Q., Pang, J.-S., and Ralph, D., *Mathematical programs with equilibrium constraints*. Cambridge University Press, 1996.

[110] Ma, H. and Shahidehpour, S., "Transmission-constrained unit commitment based on benders decomposition," *International Journal of Electrical Power & Energy Systems*, vol. 20, no. 4, pp. 287 – 294, 1998.

[111] Maheswari, S. and Vijayalakshmi, C., "A lagrangian decomposition model for unit commitment problem," *International Journal of Computer Applications*, vol. 43, no. 12, pp. 21–25, 2012.

[112] Majumder, R., Ghosh, A., Ledwich, G., and Zare, F., "Load sharing and power quality enhanced operation of a distributed microgrid," *Renewable Power Generation, IET*, vol. 3, no. 2, pp. 109–119, 2009.

[113] Mantawy, A., Abdel-Magid, Y. L., and Selim, S. Z., "Unit commitment by tabu search," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 145, no. 1, pp. 56–64, 1998.

[114] McNaughton, G. A. and Saint, R., "Enterprise integration implications for home-area network technologies," in *Innovative Smart Grid Technologies (ISGT), 2010*, pp. 1–5, January 2010.

[115] Meyer, C. D., *Matrix analysis and applied linear algebra*. Philadelphia, PA: SIAM, 2000.

[116] Meyer, R. R., "On the existence of optimal solutions to integer and mixed-integer programming problems," *Mathematical Programming*, vol. 7, no. 1, pp. 223–235, 1974.

[117] Meyer, R. R., "Integer and mixed-integer programming models: general properties," *Journal of Optimization Theory and Applications*, vol. 16, no. 3-4, pp. 191–206, 1975.

[118] Momoh, J. A. and Zhang, Y., "Unit commitment using adaptive dynamic programming," in *Intelligent Systems Application to Power Systems, 2005. Proceedings of the 13th International Conference on*, pp. 4–pp, IEEE, 2005.

[119] Morales-Espana, G., Latorre, J. M., and Ramos, A., "Tight and compact MILP formulation for the thermal unit commitment problem," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4897–4908, 2013.

[120] Morante, Q., Ranaldo, N., Vaccaro, A., and Zimeo, E., "Pervasive grid for large-scale power systems contingency analysis," *IEEE Transactions on Industrial Informatics*, vol. 2, no. 3, pp. 165–175, 2006.

[121] Motto, A. L. and Galiana, F. D., "Equilibrium of auction markets with unit commitment: The need for augmented pricing," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 798–805, 2002.

[122] Motto, A. L. and Galiana, F. D., "Coordination in markets with nonconvexities as a mathematical program with equilibrium constraints-Part I: a solution procedure," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 308–316, 2004.

[123] Naidoo, R. D., "A mixed integer programming formulation of generator startup costs," in *Power Engineering Conference, 2007. IPEC 2007. International*, pp. 173–176, IEEE, 2007.

[124] Nazari, M. H. and Ilić, M., "Potential for efficiency improvement of future electric energy systems with distributed generation units," in *IEEE Power & Energy Society General Meeting*, pp. 1–9, July 2010.

[125] Nedić, A. and Ozdaglar, A., "A geometric framework for nonconvex optimization duality using augmented Lagrangian functions," *Journal of Global Optimization*, vol. 40, no. 4, pp. 545–573, 2008.

[126] Nedic, A. and Ozdaglar, A., "Separation of nonconvex sets with general augmenting functions," *Mathematics of Operations Research*, vol. 33, no. 3, pp. 587–605, 2008.

[127] Nedic, A. and Ozdaglar, A., "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[128] Nemhauser, G. L. and Wolsey, L. A., *Integer and combinatorial optimization*, vol. 18. Wiley New York, 1988.

[129] Niknam, T., Khodaei, A., and Fallahi, F., "A new decomposition approach for the thermal unit commitment problem," *Applied Energy*, vol. 86, no. 9, pp. 1667–1674, 2009.

[130] Nogales, F. J., Prieto, F. J., and Conejo, A. J., "A decomposition methodology applied to the multi-area optimal power flow problem," *Annals of Operations Research*, vol. 120, no. 1-4, pp. 99–116, 2003.

[131] Nowak, M. P., Schultz, R., and Westphalen, M., "A stochastic integer programming model for incorporating day-ahead trading of electricity into hydro-thermal unit commitment," *Optimization and Engineering*, vol. 6, no. 2, pp. 163–176, 2005.

[132] O'Neill, R. P., Dautel, T., and Krall, E., "Recent iso software enhancements and future software and modeling plans," tech. rep., Federal Energy Regulatory Commission FERC, 2011. [Online]. Available: www.ferc.gov/industries/electric/indus-act/rto/rto-iso-soft-2011.pdf.

[133] Ongsakul, W. and Petcharaks, N., "Unit commitment by enhanced adaptive lagrangian relaxation," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 620–628, 2004.

[134] Ostrowski, J., Anjos, M. F., and Vannelli, A., "Tight mixed integer linear programming formulations for the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 39–46, 2012.

[135] Ozturk, U. A., Mazumdar, M., and Norman, B. A., "A solution to the stochastic unit commitment problem using chance constrained programming," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1589–1598, 2004.

[136] ONeill, R. P., Hedman, K. W., Krall, E. A., Papavasiliou, A., and Oren, S. S., "Economic analysis of the n-1 reliable unit commitment and transmission switching problem using duality concepts," *Energy Systems*, vol. 1, no. 2, pp. 165–195, 2010.

[137] Papavasiliou, A., Oren, S. S., and O'Neill, R. P., "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Transactions on Power Systems*, vol. PP, pp. 1–1, 2011.

[138] Parvania, M. and Fotuhi-Firuzabad, M., "Demand response scheduling by stochastic scuc," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 89–98, 2010.

[139] Phan, D. and Kalagnanam, J., "Distributed methods for solving the security-constrained optimal power flow problem," in *IEEE PES Innovative Smart Grid Technologies (ISGT)*, pp. 1–7, 2012.

[140] Plesník, J., "Finding the orthogonal projection of a point onto an affine subspace," *Linear algebra and its applications*, vol. 422, no. 2, pp. 455–470, 2007.

[141] Pogaku, N., Prodanovic, M., and Green, T. C., "Modeling, analysis and testing of autonomous operation of an inverter-based microgrid," *IEEE Transactions on Power Electronics*, vol. 22, no. 2, pp. 613–625, 2007.

[142] Rahimi, S., Niknam, T., and Fallahi, F., "A new approach based on benders decomposition for unit commitment problem," *World Applied Sciences Journal*, vol. 6, no. 12, pp. 1665–1672, 2009.

[143] Rajan, D. and Takriti, S., "Minimum up/down polytopes of the unit commitment problem with start-up costs," *IBM Res. Rep*, 2005.

[144] Ralphs, T. K. and Hassanzadeh, A., "On the value function of a mixed integer linear optimization problem and an algorithm for its construction," tech. rep., Laboratory for Computational Optimization at Lehigh (CORL), Department of Industrial and Systems Engineering, Lehigh University, Technical Report 14T-004, 2014.

[145] Restrepo, J. F. and Galiana, F. D., "Unit commitment with primary frequency regulation constraints," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1836–1842, 2005.

[146] Richter Jr, C. W. and Sheble, G. B., "A profit-based unit commitment ga for the competitive environment," *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 715–721, 2000.

[147] Rockafellar, R. T., "Augmented Lagrange multiplier functions and duality in nonconvex programming," *SIAM Journal on Control*, vol. 12, no. 2, pp. 268–285, 1974.

[148] Rockafellar, R. T. and Wets, R. J.-B., *Variational Analysis*, vol. 317. Springer, 1998.

[149] Rubinov, A. M., Huang, X., and Yang, X., "The zero duality gap property and lower semicontinuity of the perturbation function," *Mathematics of Operations Research*, vol. 27, no. 4, pp. 775–791, 2002.

[150] Rubinov, A. M., Glover, B. M., and Yang, X., "Decreasing functions with applications to penalization," *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 289–313, 1999.

[151] Rubinov, A. M. and Yang, X., *Lagrange-type functions in constrained non-convex optimization*, vol. 85. Springer Science & Business Media, 2003.

[152] Rubinov, A., Yang, X., and Bagirov, A., "Penalty functions with a small penalty parameter," *Optimization Methods and Software*, vol. 17, no. 5, pp. 931–964, 2002.

[153] Ruiz, P. A., Philbrick, C. R., and Sauer, P. W., "Wind power day-ahead uncertainty management through stochastic unit commitment policies," in *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, pp. 1–9, IEEE, 2009.

[154] Ruiz, P. A., Philbrick, C. R., Zak, E., Cheung, K. W., and Sauer, P. W., "Uncertainty management in the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 642–651, 2009.

[155] Saber, A. Y. and Venayagamoorthy, G. K., "Efficient utilization of renewable energy sources by gridable vehicles in cyber-physical energy systems," *Systems Journal, IEEE*, vol. 4, no. 3, pp. 285–294, 2010.

[156] Saber, A. Y. and Venayagamoorthy, G. K., "Unit commitment with vehicle-to-grid using particle swarm optimization," in *PowerTech, 2009 IEEE Bucharest*, pp. 1–8, IEEE, 2009.

[157] Saber, A. Y. and Venayagamoorthy, G. K., "Intelligent unit commitment with vehicle-to-grida cost-emission optimization," *Journal of Power Sources*, vol. 195, no. 3, pp. 898–911, 2010.

[158] Samiee, M., Amjady, N., and Sharifzadeh, H., "Security constrained unit commitment of power systems by a new combinatorial solution strategy composed of enhanced harmony search algorithm and numerical optimization," *International Journal of Electrical Power & Energy Systems*, vol. 44, no. 1, pp. 471–481, 2013.

[159] Sasaki, H., Watanabe, M., Kubokawa, J., Yorino, N., and Yokoyama, R., "A solution method of unit commitment by artificial neural networks," *IEEE Transactions on Power Systems*, vol. 7, no. 3, pp. 974–981, 1992.

[160] Sen, S. and Kothari, D., "Optimal thermal generating unit commitment: a review," *International Journal of Electrical Power & Energy Systems*, vol. 20, no. 7, pp. 443–451, 1998.

[161] Sen, S. and Kothari, D., "Optimal thermal generating unit commitment of large power system: a novel approach," in *TENCON'98. 1998 IEEE Region 10 International Conference on Global Connectivity in Energy, Computer, Communication and Control*, vol. 2, pp. 474–478, IEEE, 1998.

[162] Senjyu, T., Miyagi, T., Yousuf, S. A., Urasaki, N., and Funabashi, T., "A technique for unit commitment with energy storage system," *International Journal of Electrical Power & Energy Systems*, vol. 29, no. 1, pp. 91 – 98, 2007.

[163] Shafie-Khah, M., Parsa Moghaddam, M., and Sheikh-El-Eslami, M., "Unified solution of a non-convex scuc problem using combination of modified branch-and-bound method with quadratic programming," *Energy Conversion and Management*, vol. 52, no. 12, pp. 3425–3432, 2011.

[164] Shahidehpour, M., Yamin, H., and Li, Z., *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*. Wiley, 2002.

[165] Shi, W., Ling, Q., Yuan, K., Wu, G., and Yin, W., "On the linear convergence of the ADMM in decentralized consensus optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1750–1761, 2014.

[166] Shiina, T. and Birge, J. R., "Stochastic unit commitment problem," *International Transactions in Operational Research*, vol. 11, no. 1, pp. 19–32, 2004.

[167] Shrestha, G., Pokharel, B., Lie, T., and Fleten, S.-E., "Price-based unit commitment for bidding under price uncertainty," *IET Generation, Transmission & Distribution*, vol. 1, no. 4, pp. 663–669, 2007.

[168] Silva, I. d. J., Rider, M., Romero, R., Garcia, A., and Murari, C., "Transmission network expansion planning with security constraints," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 152, no. 6, pp. 828–836, 2005.

[169] Simoglou, C. K., Biskas, P. N., and Bakirtzis, A. G., "Optimal self-scheduling of a thermal producer in short-term electricity markets by MILP," *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1965–1977, 2010.

[170] Sioshansi, R., O'Neill, R., and Oren, S. S., "Economic consequences of alternative solution methods for centralized unit commitment in day-ahead electricity markets," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 344–352, 2008.

[171] Sioshansi, R., Oren, S. S., and ONeill, R., "The cost of anarchy in self-commitment based electricity markets," in *Competitive Electricity Markets: Design, Implementation and Performance*, pp. 245–266, Elsevier, 2008.

[172] Sisworahardjo, N. and El-Keib, A., "Unit commitment using the ant colony search algorithm," in *Power Engineering 2002 Large Engineering Systems Conference on, LESCOPE 02*, pp. 2–6, IEEE, 2002.

[173] Snyder, W. L., Powell, H. D., and Rayburn, J. C., "Dynamic programming approach to unit commitment," *IEEE Transactions on Power Systems*, vol. 2, no. 2, pp. 339–348, 1987.

[174] Streiffert, D., Philbrick, R., and Ott, A., "A mixed integer programming solution for market clearing and reliability analysis," in *Power Engineering Society General Meeting, 2005. IEEE*, pp. 2724–2731, IEEE, 2005.

[175] Su, C.-C. and Hsu, Y.-Y., "Fuzzy dynamic programming: an application to unit commitment," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 1231–1237, 1991.

[176] Takriti, S. and Birge, J. R., "Lagrangian solution techniques and bounds for loosely coupled mixed-integer stochastic programs," *Operations Research*, vol. 48, no. 1, pp. 91–98, 2000.

[177] Takriti, S. and Birge, J. R., "Using integer programming to refine lagrangian-based unit commitment solutions," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 151–156, 2000.

[178] Takriti, S., Birge, J. R., and Long, E., "A stochastic model for the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 11, no. 3, pp. 1497–1508, 1996.

[179] Takriti, S., Krasenbrink, B., and Wu, L. S.-Y., "Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem," *Operations Research*, vol. 48, no. 2, pp. 268–280, 2000.

[180] Takriti, S., Supatgiat, C., and Wu, L.-Y., "Coordinating fuel inventory and electric power generation under uncertainty," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 603–608, 2001.

[181] Tariq, M. U., Grijalva, S., and Wolf, M., "Towards a distributed, service-oriented control infrastructure for smart grid," in *Cyber-Physical Systems (ICCPS), 2011 IEEE/ACM International Conference on*, pp. 35–44, April 2011.

[182] Tong, S. and Shahidehpour, S., "An innovative approach to generations scheduling in large-scale hydro-thermal power systems with fuel constrained units," *IEEE Transactions on Power Systems*, vol. 5, no. 2, pp. 665–673, 1990.

[183] Tuohy, A., Denny, E., and O'Malley, M., "Rolling unit commitment for systems with significant installed wind capacity," in *Power Tech, 2007 IEEE Lausanne*, pp. 1380–1385, July 2007.

[184] Tuohy, A., Meibom, P., Denny, E., and O'Malley, M., "Unit commitment for systems with significant wind penetration," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 592–601, 2009.

[185] Valenzuela, J. and Smith, A. E., "A seeded memetic algorithm for large unit commitment problems," *Journal of Heuristics*, vol. 8, no. 2, pp. 173–195, 2002.

[186] Vemuri, S. and Lemonidis, L., "Fuel constrained unit commitment," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 410–415, 1992.

[187] Virmani, S., Adrian, E. C., Imhof, K., and Mukherjee, S., "Implementation of a lagrangian relaxation based unit commitment problem," *IEEE Transactions on Power Systems*, vol. 4, no. 4, pp. 1373–1380, 1989.

[188] Wang, C. Y., Yang, X. Q., and Yang, X. M., "Nonlinear augmented Lagrangian and duality theory," *Mathematics of Operations Research*, vol. 38, no. 4, pp. 740–760, 2013.

[189] Wang, J., Shahidehpour, M., and Li, Z., "Security-constrained unit commitment with volatile wind power generation," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1319–1327, 2008.

[190] Wang, K., Zhang, B., Wu, X., Zhai, J., Shao, W., and Duan, Y., "Multi-time scales coordination scheduling of wind power integrated system," in *Innovative Smart Grid Technologies - Asia (ISGT Asia), 2012 IEEE*, pp. 1–4, May 2012.

[191] Wang, Q., Watson, J.-P., and Guan, Y., "Two-stage robust optimization for n-k contingency-constrained unit commitment," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2366–2375, 2013.

[192] Wang, S., Shahidehpour, S., Kirschen, D., Mokhtari, S., and Irisarri, G., "Short-term generation scheduling with transmission and environmental constraints using an augmented lagrangian relaxation," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1294–1301, 1995.

[193] Wolsey, L. A. and Nemhauser, G. L., *Integer and Combinatorial Optimization*. Wiley Series in Discrete Mathematics and Optimization, Wiley, 1999.

[194] Wu, L. and Shahidehpour, M., "Accelerating the Benders decomposition for network-constrained unit commitment problems," *Energy Systems*, vol. 1, no. 3, pp. 339–376, 2010.

[195] Wu, L., Shahidehpour, M., and Li, T., "Stochastic security-constrained unit commitment," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 800–811, 2007.

[196] Wu, L., Shahidehpour, M., and Li, T., "Cost of reliability analysis based on stochastic unit commitment," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1364–1374, 2008.

[197] Xiong, P. and Jirutitijaroen, P., "Stochastic unit commitment using multi-cut decomposition algorithm with partial aggregation," in *IEEE Power & Energy Society General Meeting*, pp. 1–8, IEEE, 2011.

[198] Xu, J. and Christie, R. D., "Decentralized unit commitment in competitive energy markets," in *The Next Generation of Electric Power Unit Commitment Models*, pp. 293–313, Springer, 2001.

[199] Yamin, H. and Shahidehpour, S., "Self-scheduling and energy bidding in competitive electricity markets," *Electric Power Systems Research*, vol. 71, no. 3, pp. 203–209, 2004.

[200] Yang, X. and Huang, X., "A nonlinear Lagrangian approach to constrained optimization problems," *SIAM Journal on Optimization*, vol. 11, no. 4, pp. 1119–1144, 2001.

[201] Zavala, V. M., Botterud, A., Constantinescu, E., and Wang, J., "Computational and economic limitations of dispatch operations in the next-generation power grid," in *Innovative Technologies for an Efficient and Reliable Electricity Supply (CITRES), 2010 IEEE Conference on*, pp. 401–406, September 2010.

[202] Zhai, Q., Guan, X., and Cui, J., "Unit commitment with identical units successive subproblem solving method based on lagrangian relaxation," *IEEE Transactions on Power Systems*, vol. 17, no. 4, pp. 1250–1257, 2002.

[203] Zhang, M. and Guan, Y., "Two-stage robust unit commitment problem," *University of Florida, USA*, 2009.

[204] Zhao, F., Litvinov, E., and Zheng, T., "A marginal equivalent decomposition method and its application to multi-area optimal power flow problems," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 53–61, 2014.

[205] Zhao, L. and Zeng, B., "Robust unit commitment problem with demand response and wind energy," in *IEEE Power & Energy Society General Meeting*, pp. 1–8, IEEE, 2012.

[206] Zhuang, F. and Galiana, F., "Unit commitment by simulated annealing," *IEEE Transactions on Power Systems*, vol. 5, no. 1, pp. 311–318, 1990.

[207] Zhuang, F. and Galiana, F. D., "Towards a more rigorous and practical unit commitment by lagrangian relaxation," *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 763–773, 1988.

[208] Zimmerman, R. D., Murillo-Sánchez, C. E., and Gan, D., "A matlab power system simulation package, available online at: http://www. pserc. conrnell. edu/matpower," 2005.

VITA

Mohammad Javad Feizollahi is a PhD candidate in Operations Research in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. He received the BSc and MSc degrees in Industrial Engineering in 2005 and 2007, respectively, from Sharif University of Technology, Tehran, Iran. His main research interest is optimization, particularly theory and applications of decentralized mixed integer programming as well as optimization under uncertainty.

In his PhD work, he worked on decentralized control and operations for future power grids as part of the ARPA-E GENI program. In particular, he developed mathematical formulations and algorithms for the decentralized unit commitment and frequency control problems at the Advanced Computational Electricity Systems (ACES) Lab at Georgia Tech.

He developed robust counterparts and algorithms for some important optimization problems such as *redundancy allocation* and *quadratic assignment* while he was an M.Sc. student at Sharif University of Technology and a visiting scholar at Georgia Tech.

Mr. Feizollahi has several years of experience as an operations research analyst in various simulation and optimization projects in different companies. His responsibilities in these experiences include managing or conducting discrete event simulation, optimal inventory replenishment, optimal resource allocation, and multi-period asset allocation.