Convex polar second-order Taylor approximation of AC optimal power flows

Šepetanc, Karlo

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Karlo Šepetanc

CONVEX POLAR SECOND-ORDER TAYLOR APPROXIMATION OF AC OPTIMAL POWER FLOWS

DOCTORAL THESIS

Zagreb, 2023

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Karlo Šepetanc

CONVEX POLAR SECOND-ORDER TAYLOR APPROXIMATION OF AC OPTIMAL POWER FLOWS

DOCTORAL THESIS

Supervisor: Professor Hrvoje Pandžić, PhD

Zagreb, 2023

Sveučilište u Zagrebu

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Karlo Šepetanc

KONVEKSNA POLARNA APROKSIMACIJA OPTIMALNIH IZMJENIČNIH TOKOVA SNAGA TAYLOROVIM POLINOMOM DRUGOGA REDA

DOKTORSKI RAD

Mentor: prof. dr. sc. Hrvoje Pandžić

Zagreb, 2023.

The doctoral thesis was completed at the University of Zagreb Faculty of Electrical Engineering and Computing, Department of Energy and Power Systems, Zagreb, Croatia.

Supervisor: Professor Hrvoje Pandžić, PhD

Doctoral thesis has: 80 pages

Doctoral thesis number: _____

About the Supervisor

Hrvoje Pandžić (www.fer.unizg.hr/hrvoje.pandzic) was born in 1984 in Zagreb, Croatia. He received his Masters and PhD degrees from the University of Zagreb Faculty of Electrical Engineering and Computing (UNIZG-FER) in 2007 and 2011, respectively. After being a post-doctoral researcher at the University of Washington, Seattle, 2012-2014, he became an Assistant Professor at UNIZG-FER in 2014. Currently, he is a Professor at UNIZG-FER and Head of the Department of Energy and Power Systems, as well as Head of the Demand Response Laboratory at UNIZG-FER. He has coordinated many European and national projects focused on electricity markets, energy storage, electric vehicles, microgrids and demand response. He published over 60 papers in journals categorized as Q1/Q2 according to JCR. He received numerous awards, including the Award Science by the Government of the Republic of Croatia, 2018; Award for the highest scientific and artistic achievements in Croatia by the Croatian Academy of Science and Arts for 2018; Award Vera Johanides by the Croatian Academy of Engineering for 2015. He has been an Associate Member of the Croatian Academy of Science and Arts since 2020. He is an Editor of IEEE Transactions on Power Systems journal. He is an IEEE Senior Member, as well as a member of INFORMS and CIGRE. Besides research achievements, he has led over 20 technical studies for commercial partners.

orcid.org/0000-0003-4121-4702

scholar.google.com/citations?user=cbFwhhkAAAAJ

www.researchgate.net/profile/Hrvoje-Pandzic-2

www.scopus.com/authid/detail.uri?authorId=24438098800

www.webofscience.com/wos/author/record/2706650

O mentoru

Hrvoje Pandžic (www.fer.unizg.hr/hrvoje.pandzic) rođen je 1984. godine u Zagrebu. Završio je diplomski studij 2007. i doktorski studij 2011. godine na Sveučilištu u Zagrebu Fakultetu elektrotehnike i racunarstva (UNIZG-FER). Nakon pozicije poslijedoktoranda na Sveučilištu Washington u Seattleu 2012-2014., postaje docent na UNIZG-FER 2014. godine. Trenutno je redoviti profesor i predstojnik Zavoda za visoki napon i energetiku te voditelj Laboratorija za odziv potrošnje. Vodio je mnoge europske i nacionalne istraživačke projekte usmjerene na energetska tržišta, spremnike energije, električna vozila, mikromreže i odziv potrošnje. Objavio je preko 60 znanstvenih radova u znanstvenim časopisima kvalificiranima kao Q1 ili Q2 prema JCR. Dobitnik je mnogih nagrada, uključujući nagradu Znanost Republike Hrvatske za 2018. godinu, Nagradu za najviše znanstvene i umjetničke uspjehe Hrvatske akademije znanosti i umjetnosti za 2018. godinu te nagradu Vera Johanides Hrvatske akademije tehničkih znanosti za 2015. godinu. Član-suradnik je Hrvatske akademije znanosti i umjetnosti od 2020. godine. Član je uređivačkog odbora časopisa IEEE Transactions on Power Systems. Član je profesionalnih udruženja IEEE (Senior Member), INFORMS i CIGRE. Osim znanstvenih uspjeha, vodio je preko 20 projekata za komercijalne naručitelje.

orcid.org/0000-0003-4121-4702

scholar.google.com/citations?user=cbFwhhkAAAAJ

www.researchgate.net/profile/Hrvoje-Pandzic-2

www.scopus.com/authid/detail.uri?authorId=24438098800

www.webofscience.com/wos/author/record/2706650

Acknowledgement

Special thanks to my supervisor professor Hrvoje Pandžić, associate professor Domagoj Vlah who directly contributed to the research and my parents who supported me on my endeavour.

Abstract

Modern transmission systems operation and planing problems involve computationally difficult optimizations. Transmission system operators monitor, manage power flows and plan system expansion to ensure that the system is always operating within permitted limits. Exact formulations exist, but are usually uncomputable even for small system sizes. On-the-other-hand, private power system entities optimize their investments and operation to maximize profit due to market participation. Market participation problems, due to interaction of the market participant and the market, are of even more computationally difficult bilevel optimization structure. Bilevel optimization involves two interlinked optimization problems, each optimizing its objective function and inheriting decision variables of the other optimization problem as parameters where one optimization is the master problem or the upper level, i.e. strategic player, and the other is the follower problem or the lower level problem, i.e. non-strategic player. Traditionally, solving bilevel problems involves mathematical transformation to single-level equivalent problem that of-the-shelf solvers can solve, but this procedure has severe computational performance implications. This thesis aims at improving numerical tractability of both optimization problem categories: transmission system AC optimal power flows as stand-alone optimization and bilevel optimization problems with AC transmission network constrained market clearing in the lower level.

The developed optimal power flow formulation is based on a convex approximation of Taylor expansion, as opposed to nonconvex exact formulations. Convex form is relevant for power system operation models as modern convex solvers generally handle binary variables more efficiently than nonconvex or general nonlinear solvers. The emphasis on achieving great accuracy, sufficiently close to the exact models to be used as substitution in practical cases. To achieve it, it uses the developed presolve that aims to eliminate relaxation errors. The applicability of the developed optimal power flow approach is verified on a unit commitment problem and a transmission system expansion problem. The results demonstrate faster computation times than the exact nonconvex formulation and greater accuracy than the existing convex quadratically-constrained quadratic or second-order cone formulations with similar computation times.

The presented convex optimal power flow formulations are used as a foundation for bilevel market participation modeling using single-level reduction solution techniques. This is highly relevant as the existing single-level reduction techniques do not hold for nonconvex models. The developed optimal power flow model is applied to deliver a new and accurate bilevel day-ahead AC transmission-constrained market participation model. Due to the computational difficulties of the resulting bilevel optimization formulation, second-order cone complementary condition constraints smoothening techniques is employed. The results indicate faster computation times than with other existing solution techniques and significantly higher first iteration accuracy than

with other compared existing optimal power flow models.

This thesis also presents an alternative neural network approach for solving bilevel day-ahead AC transmission constrained market participation model. The approach consists of generating neural network training dataset by computing optimal power flows for different bid values, training the neural network where the output of it is market participant's profit and optimizing the biding where the lower level is bypassed by simulating its market clearing decision with neural network in the objective function. The numerical tractability is achieved by using a smooth nonlinear activation function, as opposed to typical mixed-integer linear formulation. High accuracy is achieved due to convolutional neural network structure that mimics weak intertemporal market price dependency, i.e. its structure is physics informed. While the computation times are longer than for the single-level reduction technique, the solution process of this approach can converge for larger systems.

Keywords: AC Optimal Power Flow, Bilevel Optimization, Numerical Tractability, Smoothing Complementary Conditions, Convolutional Neural Network, Unit Commitment, Transmission System Expansion

Konveksna polarna aproksimacija optimalnih izmjeničnih tokova snaga Taylorovim polinomom drugoga reda

Postojeća europska tržišta električne energije dijele prijenosnu energetsku mrežu u zone. Svaka zona se modelira kao jedno čvorište, odnosno pretpostavlja se da unutar zone ne postoje aktivna mrežna ograničenja i zanemaruju se gubici. Čvorišta su međusobno povezana vrlo jednostavnim linearnim mrežnim modelom tokova snaga. Tim modelom zanemaruju se utjecaj kuta i magnitude napona na tokove snaga, gubici djelatne snage kao i tokovi jalovih snaga, dok nije zanemarena maksimalna snaga razmjene čvorišta i njihova topološka povezanost. Tržište se potom oslanja na naknadni postupak redispečiranja koje provodi operator prijenosnog sustava. Taj postupak uključuje provjeru izvedivosti tržišnih pozicija, eventualne korekcije proizvodnje i upravljanje naponima mijenjanjem tokova jalove snage. Ovakav način pogona sustava koji se sastoji od jednostavnog čišćenja tržišta i naknadnim redispečiranjem rezultira suboptimalnim rješenjem. Kvalitetniji model po pitanju optimalnosti je u upotrebi na, primjerice, američkim tržištima električne energije gdje vozni red elektrana određuje nezavisni operator prijenosnog sustava koristeći detaljniji istostjerni model optimalnih tokova snaga. Međutim, i taj je postupak dvokoračni jer se nakon određivanja voznog reda elektrana u zasebnom koraku upravlja jalovom snagom, što također rezultira suboptimalanim rješenjem. Težnja je modelirati prijenosni sustav računalno zahtjevnijim izmjeničnim tokovima snaga u fazi čišćenja tržišta kako bi se izbjegla potreba za naknadnim redispečiranjem i tako postigao globalni maksimum društvenog blagostanja.

Točan model optimalnih tokova snaga je nekonveksan i posljedično se može optimizirati samo s općenitim nelinearnim optimizacijskim alatima koji su manje učinkoviti kada se koriste cjelobrojne varijable. To je važno za elektroenergetski sustav jer su problemi čišćenja tržišta, određivanja voznog reda elektrana i izgradnje prijenosnog sustava numerički vrlo zahtjevni zbog utjecaja cjelobrojnih varijabli. Ova disertacija razvija novi konveksan model izmjeničnih optimalnih tokova snaga s ciljem postizanja veće preciznosti od postojećih konveksnih modela i bolje izračunljivosti u odnosu na postojeće točne nekonveksne modele. Novi model temelji se na aproksimaciji tokova snaga Taylorovim polinomom drugoga reda, a dobiveni zapis je konveksno kvadratno ograničeneog kvadratnog oblika, što omogućava upotrebu za to namijenjenih optimizacijskih alata. Model postiže preciznost tako što je Taylorov polinom razvijen oko približne radne točke, što čini model lokalnom aproksimacijom. Dodatno, koristi se razvijeni pretproračun koji uklanja pogreške relaksacije. Pogreške relaksacije nastaju zbog zamjene znaka jednakosti s nejednakosti kod kvadratnih ograničenja kako bi bila konveksna, s obzirom da ne postoji konveksno nelinearno ograničenje jednakosti. Pretproračun odabire relaksirana ograničenja za zamjenu s alternativnim ograničenjima linearih jednakosti, koja ne mogu prouzročiti pogreške relaksacije. Predmetna zamjena temelji se na predznaku ci-

jene u sijeni iz analize osjetljivosti odgovarajućih ograničenja proračuna približne radne točke. Preciznost modela može se iterativno poboljšavati tako što se novi proračun pokrene s Taylorovim polinomom razvijenim oko nove približne radne točke sustava dobivenom u prethodnoj iteraciji. Iterativni postupak, počevši s radnom točkom prazog pogona, konvergira unutar 0.00% pogreške za svaku od testirane četerdesetčetiri prijenosne mreže unutar tri iteracije od prve izvedive iteracije. Praktična primjena modela je demonstrirana na problemu voznog reda elektrana i problemu dogradnje prijenosnog sustava. Rezultati upućuju na brže vrijeme proračuna od točnog nekonveksnog modela i veću preciznost u odnosu na postojeće konveksne kvadratno ograničene kvadratne modele i modele konusa drugoga reda. Preciznost je potvrđena koristeći dodatan proračun s točnim nekonveksnim optimalnim tokovima snaga, ali s fiksiranim cjelobrojnim varijablama. Proračun se algoritamski može prikazati:

•Prora čun približne radne točke

•Pretprora čun

•Prora čun optimalnih tokova snaga

•Prora čun provjere preciznosti

Osim problematike određivanja čišćenja tržišta električne energije uvažavajući ograničenja izmjeničnih tokova snaga u prijenosnoj elektroenergetskoj mreži, vrlo je aktualan problem optimiziranja sudjelovanja tržišnog sudionika na takvim centraliziranim tržištima. Tržišni sudionik, ako raspolaže velikim snagama i kapacitetima, može utjecati na cijenu električne energije na tržištu što utječe na njegov profit i odluke o kupnji i/ili prodaji. Stoga su dvije optimizacije vezane u dvorazinsku strukturu. Donja razina je tržište električne energije koje maksimizira društveno blagostanje temeljem ponuda za kupnju i prodaju svih sudionika, dok gornju razinu čini razmatrani tržišni sudionik koji maksimizira svoj profit trgovanja uzimajući u obzir tržišnu cijenu i svoj utjecaj na nju. Postojeća tehnika rješavanja dvorazinske optimizacije temeljena je na svođenju problema na jednu razinu koristeći Karush–Kuhn–Tucker uvjete optimalnosti i linearizaciji komplementarnih ograničenja koristeći tehniku *velikog M*. Dobivena formulacija je numerički vrlo zahtjevna za računala i nije primjenjiva ukoliko je donji optimizacijski problem nelinearnog oblika. Cilj ove disertacije je izraditi precizan model i računalno efikasno riješiti problem dvorazinske optimizacije sudionika na tržištu električne energije ograničenom s izmjeničnim tokovima snaga prijenosne mreže u donjoj razini.

Konveksna formulacija optimalnih tokova snaga je osnova za dvorazinsku optimizaciju tržišnog sudjelovanja uz tehnike svođenja na jednu ekvivalentnu razinu jer u suprotnom matematičke pretpostavke tih tehnika nisu zadovoljene. Za razvoj novog preciznog dovrazinskog modela upotrebljen je također novi konveksni model optimalnih tokova snaga razvijen u prvom dijelu disertacije. Samoj dvorazinskoj optimizaciji prethodi određivanje približne radne točke sustava koja se za razmatrani problem računa kao stanje prijenosnog sustava bez kupnje ili prodaje energije razmatranog tržišnog sudionika. Zatim slijedi pretproračun za izbjegavanje poten-

cijalnih pogrešaka relaksacije optimalnih tokova snaga. Poželjno, iako opcionalno, je izračunati početne vrijednosti svih varijabli u optimizaciji kako bi se inicijalizirao glavni proračun s dobrim vrijednostima. To ne utječe na optimalno rješenje, ali olakšava optimizacijskim alatima da ga pronađu. Zadnji korak proračuna, nakon dvorazinske optimizacije, je provjera preciznosti, odnosno uspoređuju se profiti izračunati u dvorazinskoj optimizaciji s profitima postignutima uz fiksirane odluke kupnje i prodaje uz točan model optimalnih tokova snaga. Proračun se može iterativno pokretati, uz obnavljanje pretpostavljene radne točke s novom dobivenom u prethodnoj iteraciji, ali postignuta preciznost u prvoj iteraciji je vrlo visoka i rijetko ju je potrebno povećavati u naknadnim iteracijama. Opisani dvorazinski proračun može se algoritamski zapisati na sljedeći način:

•Prora čun približne radne točke

•Pretprora čun

•Prora čun primalnog zapisa optimalnih tokova snaga

•Prora čun dualnog zapisa optimalnih tokova snaga

•Dvorazinska optimizacija

•Prora čun provjere preciznosti

Naglasak disertacije je u određivanju računalno najpovoljnije metode rješavanja jednorazinskog ekvivalenta prethodno opisane dvorazinske optimizacije. Uspoređuje se sedam različitih kategorija metoda rješavanja. Usporedba je provedena temeljem profita tržišnog sudionika i vrijednosti društvenog blagostanja izračunate u dvorazinskoj optimizaciji i točnim iznosom iz provjere preciznosti i iznosom dualnog razmaka primalnog i dualnog donjeg optimizacijskog problema. Primalno-dualne tehnike su računalno najmanje zahtjevne, ali ostavljaju velik dualni razmak i stoga su niskih preciznosti. Tehnike jake dualnosti u potpunosti zatavaraju dualni razmak, ali optimizacijski alati teško konvergiraju s tim formulacijama. McCormick omotnice relaksiraju bilinearne umnoške iz jake dualnosti, što rezultira s istim rješenjima kao primalno-dualne tehnike jer je relaksacija prevelika. Nerelaksirane varijante tehnika komplementarnih uvjeta mogu u potpunosti zatvoriti dualni razmak, ali samo relaksirana agregirana varijanta je u praktičnim uvjetima izračunljiva. Pokazalo se da je teško odrediti vrijednost relaksacijskog parametra kod te podtehnike. Tehnike penalizacijskog faktora pokazuju poteškoće u konvergiranju optimizacijskih alata ukoliko je vrijednost penalizacijskog faktora visoka, dok za niže vrijednosti dualni razmak je značajan. Diskretizacijske tehnike, iako su teorijski egzaktne, računalno su gotovo neizračunljive zbog velikih binarnih stabala pretraživanja cjelobrojnih vrijednosti. Vrlo visoka preciznost i dobra izračunljivost, povoljnije u odnosu na ostale tehnike koje mogu gotovo u potpunosti smanjiti dualni razmak, postignuti su metodama uglađenja komplementarnih uvjeta. Uspoređene su sljedeće tehnike i njihove inačice:

•Primalno-dualne tehnike

- Primal-dual

  - Osnaženi primal-dual

•Tehnike jake dualnosti
  - Jaka dualnost
  - Relaksirana jaka dualnost

•McCormick omotnice

•Tehnike komplementarnih uvjeta
  - Komplementarni uvjeti
  - Relaksirani komplementarni uvjeti
  - Agregirani komplementarni uvjeti
  - Relaksirani agregirani komplementarni uvjeti

•Tehnike penalizacijskog faktora
  - Penalizacijski faktor – jaka dualnost
  - Penalizacijski faktor – komplementarna labavost

•Tehnike diskretizacije interakcije
  - Binarna ekspanzija – jaka dualnost
  - Binarna ekspanzija – penalizacijski faktor
  - Unarna ekspanzija – jaka dualnost
  - Unarna ekspanzija – penalizacijski faktor

•Tehnike ugla đenja
  - Chen–Harker–Kanzow–Smale
  - Kanzow

Disertacija također istražuje pristup rješavanja prethodno opisanog dvorazinskog optimizacijskog problema koristeći neuronske mreže. Ovim pristupom nadomješta se donji opimizacijski problem s neuronskom mrežom treniranom da replicira odluke tržišta o cijeni električne energije ovisno o ponudi gornje razine. Prvo je potrebno napraviti bazu podataka za treniranje neuronskih mreža proračunima optimalnih tokova snaga za razne vrijednosti trgovane energije gornje razine, zatim je potrebno istrenirati neuronsku mrežu za ukupan profit razmatranog sudionika na tržištu i na kraju optimizirati odluke gornje razine uz nadomještenu funkciju cilja s funkcijom neuronske mreže. Računalna izračunljivost postupka postiže se uglađenom aktivacijskom funkcijom, nasuprot uobičajenoj mješovitoj-cjelobrojnoj linearnoj formulaciji, i smanjivanjem veličine optimizacijskog problema evaluirajući neuronsku mrežu u jeziku za modeliranje umjesto u optimizacijskom alatu. Visoka preciznost postiže se konvolucijskom strukturom neuronske mreže koja oponaša slabu međuvremensku ovisnost tržišnih cijena. Drugim riječima, struktura neuronske mreže je fizikalno informirana. Vremena proračuna su duža ovim pristupom nego pristupom svođenja na jednu razinu temeljem Karush–Kuhn–Tucker uvjeta, ali postupak neuronskim mrežama može konvergirati za veće sustave. Razvijeni algoritam proračuna sažeto je prikazan na sljedeći način:

ponavljaj

- •Generirati bazu podataka za treniranje neuronske mreže (10^5 *unosa*)
- •Evaluirati odaziv donje razine za svaki podatak iz baze
- •Istrenirati 60 neuronskih mreža za aproksimaciju donje razine
- •Optimizirati gornju razinu s neuronskim mrežama umetnutim u funkciju cilja
- •Odrediti stvarne profite za optimizirane iznose trgovane energije
- •Odabrati najbolji stvarni profit od:
  - najbolji izravan rezultat;
  - rezultat dobiven usrednjavanjem odluka dobivenih sa svim neuronskim mrežama;
- •Za idu ću iteraciju smanjiti prostorno područje podataka za treniranje neuronske mreže oko najboljeg rezultata

dok: Najbolji rezultat je gori u odnosu na prethodnu iteraciju

Ključne riječi: optimalni izmjenični tokovi snaga, dvorazinska optimizacija, računalna izračunljivost, uglađenje komplementarnih uvjeta, konvolucijska neuronska mreža, vozni red elektrana, proširenje prijenosnog sustava

Contents

Chapter 1

Introduction

1.1 Background, Motivation and Objective of the Thesis

Existing European electricity markets use zonal areas to divide the operational responsibility for the pan-European transmission power system. Every zone is represented by a singe bus only, i.e. it is assumed that there are no transmission system active operating constraints and power loses are ignored. Buses are interconnected using a simple network power flow model [1]. This model does not consider the effect of physical properties of electricity such as voltage magnitude and angle on power flows and it ignores the losses and reactive power flows. The only considered properties are maximum power exchange and topological bus placements. Reliable operation after the market clearing process is achieved by the transmission system operators, who check the feasibility of the market clearing outcome and perform a redispatch of active power if required or control the voltage magnitudes by reactive power. The described two-step production optimization procedure, consisting of a simple market clearing in the first step and the transmission system operator corrections in the second one, results in suboptimal operation. Slightly different electricity market design is used in the United States of America. There, the unit commitment is determined by an independent system operator using a detailed DC optimal power flow model [2], which considers voltage magnitude and angle effect on power flows. However, this procedure also resorts to a two-step optimization that results in suboptimal solutions. Thus, the incentive is to use AC optimal power flow and to optimize all grid components in a single optimization. However, this requires abundant computational resources even for simple linear optimal power flow models. This thesis focuses on improving the numerical tractability of the AC optimal power flow while achieving very high model accuracy.

The motivation to develop a new AC optimal power flow arises from the fact that modern mixed-integer quadratic solvers handle binary variables much more efficiently than general mixed-integer nonlinear solvers. This is highly relevant since many power system problems, such as unit commitment and system expansion planing, contain plentiful of integer variables.

Furthermore, there is also a motivation to increase the accuracy of existing relaxation and approximation AC optimal power flow models since their typical inaccuracy is approximately as large as system losses, which is several percent. To adequately model the system losses, optimal power flow accuracy needs to be significantly higher than the share of losses in the total power flows. To achieve high accuracy, the proposal is to use an approximation of the transmission system operating point and a presolve to dynamically adapt the model to a transmission system state so that the potential relaxation errors are avoided. As such, the goal of the thesis is to develop an AC optimal power flow model that will be more tractable than the exact models and comparably tractable, but more accurate than the existing convex quadratically-constrained quadratic programming models.

The second research focus is on maximizing the benefit from AC transmission system constrained electricity market participation. Market participants that have available large capacities can impact the market price with their bids which, in turn, changes their optimal bid quantities. Since electricity markets maximize the overall social welfare and participant optimizes its profit due to arbitrage on a market, the two optimizations are interlinked in bilevel optimization structure. The market participant is called the leader or the upper-level problem. It decides on bid quantities anticipating the market result. The market, which is called the follower or the lower-level problem, determines market prices based on all bids. The two interlinked optimizations are much more difficult to solve than solving them individually. Existing solution techniques effectively solve bilevel problems where the lower level is linear, but severe intractabilities occur for optimizations that contain convex quadratic or more complex constrains in the lower level. The goal of this thesis is to deliver a new bilevel model that will accurately solve the AC market participation problem and address numerical tractability issues that occur due to mathematically difficult constraints. Recent mathematical advancements provide unexplored options for solving difficult optimization problems. The issue of solving the described bilevel problem is approached with two novel techniques: a) smoothing the complementary conditions and b) using neural networks to replace the lower level with its metamodel. The goal of the thesis is to develop an algorithm for each technique and demonstrate their advantages over the each other and the existing techniques.

To summarise, this thesis contribution is divided in two parts:

1. A development of an accurate and numerically tractable AC optimal power flow model and a presolve technique which further increases accuracy.
2. A development of two accurate and numerically tractable bilevel models for bilevel optimization with AC transmission system constrained electricity market in the lower level and market participant in the upper level.

1.2 Structure of the Thesis

The thesis is structured as follows:

- Chapter 2 introduces relevant mathematical concepts and presents state-of-the-art.
- Chapter 3 highlights the main contribution of the thesis and links them to the related publications;
- Chapter 4 presents the list of all relevant publications;
- Chapter 5 summarizes the author's contribution to the publications;
- Chapter 6 concludes the thesis and highlights the main findings.

Chapter 2

Concepts and State-of-the-art

This review section covers three research areas relevant to this thesis: optimal power flow and bilevel solution techniques.

2.1 Optimal power flow

2.1.1 Exact models

Optimal power flow research focuses on developing new power flow equations. Exact power flow equations are directly derived from an equivalent circuit scheme displayed in Fig. 2.1. Final expressions depend on the coordinate system for representing general complex from equations (2.1) and (2.2) and on substitution of the variables, e.g. currents can be fully substituted by voltages. Commonly used formulations are voltage-based in rectangular coordinates [3], [4] and in polar coordinates [3], [5]. In case of nonlinear impedances, rectangular current-voltage formulation without current substitutions is used [6]. An important concept in optimization is that the numerical tractability of a formulation is not a result of physical properties, but of a notation. Voltage-based rectangular and current-voltage formulations both contain nonconvex quadratic equations. However, current-voltage formulation has more variables than the former. Voltage-based formulation is of general nonlinear and nonconvex form. As a result, all exact formulations solve differently despite mathematical equivalence. Furthermore, formulation convexity is also important factor in optimization. For convex models, every local optimal solution is also global solution. Thus, for convex models it is easy to prove using first order optimality criteria if global solution is found. Thus far no finite exact convex formulation has been found, but there is infinite exact convex moment-based formulation [7].

$$S_{i,j} = (Y_{i,j} + Y_{i,j}^{sh}/|T_{i,j}|^2)^* \cdot |V_i|^2 + Y_{i,j}^* \cdot V_i \cdot V_j^*/T_{i,j} \tag{2.1}$$

$$S_{j,i} = (Y_{i,j} + Y_{i,j}^{sh})^* \cdot |V_i|^2 + Y_{i,j}^* \cdot V_i \cdot V_j^*/T_{i,j} \tag{2.2}$$

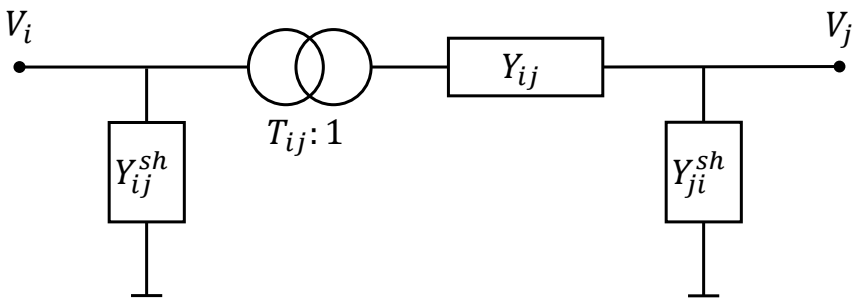

Figure 2.1: Equivalent circuit scheme for lines and transformers.

Table 2.1: Summary of exact optimal power flow formulations.

Name & reference	Optim. class	Notes
Polar, [3], [5]	NLP	Voltage expressed in terms of voltage magnitude and angle.
Rectangular, [3], [4]	NLP	Voltage expressed in terms of real and imaginary parts.
Current-voltage, [6]	NLP	Power flows expressed in terms of voltage and current real and imagniary parts.

Power system problems frequently contain discrete variables to model discrete decisions such as investments, activity states or activation levels, normally occurring in unit commitment or transmission expansion planning problems. Generally, exact formulations solve well if there are no discrete variables in the model, but show severe computational intractabilities if containing only few discrete variables. Computational difficulties are a result of the way the solvers address discrete variables. Solution procedure is based on the branch-and-bound algorithm [8] that first relaxes the discrete variables, than solves the optimization, adds integrality cuts on non-integer relaxed discrete variables and repeats the procedure until all relaxed discrete variables take integer values. The key in achieving faster computation times is in warm-starting the subsequent branch-and-bound iterations within the solver since only one constraint in the model is added in each new iteration. General nonlinear solvers do not use algorithms that can efficiently warm-start the iterations and thus optimal power flow research has been focused on developing new approximate or relaxed formulations that can use more specific solvers that handle discrete variables more efficiently. Relaxation models expand the feasibility space and provide best bound solution for exact models. If the solution of exact and relaxed model are the same, then the found solution of the exact model is proven optimal. Approximations, on the other hand, both tighten and relax the feasibility space and thus have wider set of transformations available to derive the model. Example of feasible spaces is shown in Fig. 2.2.

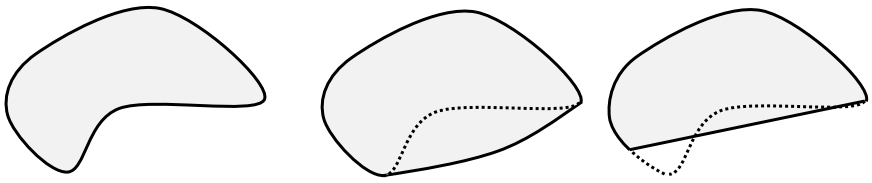

Figure 2.2: Feasible spaces: a) nonconvex; b) relaxed; c) approximated

2.1.2 Relaxation models

The least tight relaxations are the copper-plate model and the Network-flow model [1]. The copper-plate model is also know as the single-bus model. It removes all network constraints and places all devices on a single bus. It is a relaxation since its solution is always equal or more favourable than with the exact models. On the other hand, the network-flow model has network constraints, but it only accounts for network topology and maximum power exchange, ignoring dependency of the power power flows on the physical properties such as current and voltage. Both models are linear and thus solve easily with discrete variables since simplex algorithm can be used in the solution process to quickly solve branch-and-bound leaf nodes. This is a result of the simplex algorithm feature that it can be warm-started with the previous solution when solving a model with slightly changed parameters or constraints. Warm-starting it usually results in a speed-up of the two orders of magnitude which is significant. Tighter relaxations are of a more general second-order cone or an even more general semidefinite form. Jabr's second-order cone relaxation [9] is commonly used in power distribution systems. For distribution systems, i.e. radial in structure, there are proven conditions for which the relaxation is exact as elaborated in detail in paper [10]. In some cases, where the exactness conditions are not met, penalty factors can be used to retrieve the exact solution. The model is significantly less accurate for meshed transmission systems, commonly reaching inaccuracy of several percentages, and in some cases even over 10% [11]. Quadratic-convex relaxation [12] tightens the Jabr's relaxation by adding cuts due to assumed or computed maximum voltage angle differences. Unless the voltage angle differences can be determined to be very small, the model is just marginally more tight than the Jabr's model. Shor's relaxation [13] is of the most computationally demanding semidefinite form, but is also significantly tighter than the Jabr's formulation. To achieve tractability, semidefinite matrix sparsity needs to be used to separate it into more smaller ones [14]. A sparse Jabr's model achieves linear scaling with the number of system buses, but is still very demanding and currently there are very few solvers that are capable of solving mixed-integer semidefinite programming problems, namely SCIP-SDP [15], BNB [16] and CUTSDP [17]. However, these solvers are still not mature and can not solve difficult cases. Overview of relaxations, sorted in increasing tightness, is available in Table 2.2.

Table 2.2: Summary of optimal power flow relaxations, sorted in increasing tightness.

Name & reference	Optim. class	Notes
Copper-plate, [1]	LP	Single-bus model. Does not model voltages, reactive power and losses.
Network-flow, [1]	LP	Accounts for network topology, but not for physical power flow properties. The only constraint is maximum power transfer between connected nodes. Does not model voltages, reactive power and losses.
Jabr's, [9]	SOCP	SOCP relaxation of exact polar or rectangular formulation. Commonly used in distribution networks.
QC, [12]	SOCP	Further tightened Jabr's model with precomputed variable bounds.
Shor's, [13]	SDP	SDP relaxation of exact polar or rectangular formulation. Requires SDP specialised solver. Computationally demanding.

2.1.3 Approximation models

Accuracy of approximation models is typically more consistent than of the relaxation models, which are for some cases exact, but for other cases achieve large errors of over 10%. The most widely used approximation for transmission systems is the DC optimal power flow [2]. It ignores losses and reactive power flows, but models the power flow voltage angle dependence. It is accurate for transmission systems since they, due to high voltages, have low active power flow losses. The DC model is linear and thus solves well in combination with discrete variables. Further research led to generalization of the DC model to include losses and reactive power. The linear model in [18] adds a piece-wise linear expression for losses using numerically difficult and nonconvex integer variables. The model presented in [19] uses linear data-driven constraints to add losses. However, it needs to penalize the objective function to prevent negative losses. DC model extensions are generally still inaccurate for losses and reactive power flows. A convex quadratic AC optimal power flow approximation [20] has been published as a patent. It is more accurate than the linear models, but also more computationally demanding. A linear-programming AC Taylor-based optimal power flow [21] uses piece-wise linearization of quadratic terms with relaxed components to avoid discrete variables. The model considers only angle-dependent quadratic Taylor terms. After the initial publication, the author in PowerModels package [22] implemented more accurate version of the model that does not piece-wise linearize considered quadratic terms, but uses more accurate initial quadratic terms. Out of the

presented models, only the DC extension models and the linear programming AC model can be warm-started with approximate operating point to achieve greater accuracy. However, none of them utilizes this advantage efficiently, potentially resulting in even worse accuracy if warm-started. In this thesis the linear programming AC model [21] will be enhanced to better exploit the approximate operating point potential using convex quadratically constrained quadratic programming. Overview of optimal power flow approximations is given in Table 2.3.

Table 2.3: Summary of optimal power flow approximations.

Name & reference	Optim. class	Notes
DC, [2]	LP	No reactive power.
DC extensions, [18], [19]	MILP/LP	Based on linear Taylor expansion and added linear or mixed-integer linear losses.
LPAC, [21]	LP/QCQP	Based on first-order and select second-order Taylor expansion terms.
QPAC, [20]	QCQP	Published patent. Based on series of convex quadratic approximations of initial exact nonconvex quadratic equations.

2.2 Bilevel solution techniques

Bilevel modeling is mainly used for operation and planning optimization. Literature covers a wide range of topics, with the more common ones being: pricing [23] and arbitrage [24] schemes, maintenance scheduling [25] and transmission system expansion planning [26]. Bilevel optimization is an optimization structure where one optimization is embedded, i.e. nested, into the other. Its formal mathematical notation is provided in equation (2.3), where F, G and H are general functions. The two optimizations exchange decision variables x and y. Variables of one problem may be parameters in the other and vice versa. The embedded problem is called the lower-level problem and the outer problem is called the upper-level problem. Such optimization problems cannot be solved directly using off-the-shelf solvers and have to be converted into a more suitable form. Research of bilevel solution techniques focuses on improving numerical tractability of solution algorithms. Bilevel problems are difficult to solve even if problems are linear.

$$\begin{aligned} \min_{x} \quad & F_1(x,y) \\ \text{subject to} \quad & G_1(x,y) \leq 0 \\ & H_1(x,y) = 0 \\ y \in \arg\min \quad & \{F_2(x,y) : G_2(x,y) \leq 0, \quad H_2(x,y) = 0\} \end{aligned} \tag{2.3}$$

There are two main bilevel groups of solution approaches: the classical approaches and the evolutionary approaches. Classical approaches are based on the duality theory and are applicable to cases where the strong duality holds. In general, the strong duality holds for all convex problems that are strongly feasible, i.e. there exists a solution to a convex optimization problem that is not at the boundary of the feasible space. Conditions that encompass broader set of problems also exist, but they are difficult to prove and validity does not only depend on the problem structure, but also on the solution point. Thus, in practice they only apply to optimization problems with strong regularity assumptions such as linearity or convexity and continuity of the lower-level problem. Most widely used classical approach is the KKT-based single-level reduction. This way the lower-level optimization problem is replaced with a set of constraints that have the same solution as the original optimization problem. The set of equations is then added to the upper level as constraints. The resulting formulation is of the standard optimization form that can be solved using off-the-shelf solvers. However, computational costs are moderately high due to combinatorial characteristics of complementary conditions, i.e. equations that require that bilinear terms are equal to zero. The single-level reduction approach has been used to solve either bilevel problems with linear lower levels [27] or convex quadratic lower levels [28]. In a special case when interaction between the two levels is discrete, it has also been used to solve bilevel problems with convex quadratically constrained quadratic lower levels [29]. Other classical approaches are the descent method, the penalty function method, the trust-region method and the parametric programming method. The descent method uses a gradient to determine optimal variable change. Numerical difficulties occur due to the lower level being feasible only when it is optimal, so it is difficult to determine a variable change that respects feasibility. This technique was used in paper [30]. The penalty method transfers constraints into an objective function penalizing the constraint violations. In such way, the lower-level constraints are transferred to the objective function in [31], while in [32] constraints of the both levels are transferred to the objective function. The trust region algorithms are based on an iterative approximation of the lower level feasible space around the operating point by a linear problem or a quadratic problem [33]. This thesis researches which solution method for KKT-based single-level reduction works the best in terms of tractability and accuracy with the developed convex AC optimal power flow approximation model for bilevel transmission-system-constrained market participation.

On the other hand, the evolutionary approaches for solving bilevel problems mimic biological evolution by allowing generational propagation of the most favourable numbers by some metric, usually the combination of the objective function value and the infeasibility penalty. These approaches are efficient in finding good solutions of mathematically very irregular, nonlinear or discrete problems, but usually cannot prove or find global optimal solutions. Evolutionary approaches normally involve solving the lower-level problem separately using simplex or interior point algorithm for any upper-level variable values, in a nested way, that have been determined using evolution such as in [34] and [35]. However, the lower level can also be solved using an evolution, as in [36]. The nested evolutionary method cannot solve large problems due to an exponential growth of the lower-level optimizations with the number of upper-level variables. Single-level reduction can also be used for evolutionary approaches where the evolution is applied to a reduced form, as in [37]. The approach inherits mathematical regularity requirements of the classical approaches for the lower level, but can solve for more irregular upper levels. Finally, bilevel problems can be solved by adding a metamodel to the upper level as a substitution for the lower level. This can be done using a reaction set mapping to approximate the lower-level variable values as a response to the upper-level decisions, as demonstrated in [38]. Another way of substituting the lower level with a metamodel is by replacing its objective statement, i.e. objective function and minimize and maximize goal, with a constraint requiring that its objective function value is at least equal or better than function approximating optimal value of it in terms of upper level variables. The described approach is called the optimal lower-level function value approach and was used in [39]. More generally, but widely unexplored, the lower levels can be bypassed completely by metamodels. The latest advancements in machine learning provide new tools to train statistics-based metamodels. With a goal of increasing the numerical tractability and solution accuracy, this thesis focuses on applying machine learning techniques on solving bilevel models. The considered bilevel transmission-system-constrained market participation model does not conform to the required regularity constraints for classical approaches in its original exact formulation since it is nonconvex, which makes it suitable for evolutionary approaches. The most recent trend to achieve modeling accuracy and efficiency is to make metamodels, in this case neural networks, physics informed about the original models that they replace. This thesis thereby researches how to make good physics-informed neural networks as a metamodel for bypassing the lower level completely.

Chapter 3

Main Scientific Contribution

This thesis' contribution is divided into three parts. The first one provides a new convex optimal power flow formulation that uses approximate operating point to warm-start the model and achieve greater accuracy. The second part investigates using presolve to dynamically determine between quadratic inequality and linear equality constraints to further increase the accuracy of the developed optimal power flow model. The third part of contribution proposes a bilevel model for AC transmission system constrained day-ahead electricity market participation with contributions of achieved accuracy and numerical tractability. The contribution is listed here and then explained in the follow-up sections:

- a convex polar second-order Taylor approximation model of optimal AC transmission system power flow;
- a presolve method for selecting between quadratic inequality and linear equality constraints to improve accuracy of the optimal power flow models based on second-order Taylor approximation;
- a bilevel optimization model of an AC power-transmission-system-constrained day-ahead electricity market participation.

3.1 Convex polar second-order Taylor approximation model of optimal AC transmission system power flow

The developed convex AC optimal power flow model from [P_1] enables the use of modern, efficient solvers for mixed-integer programming. The achieved accuracy is greater than for the existing relaxation and approximation models. Computational tractability is comparable to other existing convex quadratically-constrained quadratic or second-order cone models. Accuracy and tractability are demonstrated on a unit commitment case study in [P_1] and on a transmission system expansion case study in [P_5].

3.2 Presolve method for selecting between quadratic inequality and linear equality constraints to improve accuracy of the optimal power flow models based on second-order Taylor approximation

Convex optimal power flows are susceptible to relaxation errors. These errors have gross impact to the model accuracy. The developed presolve dynamically adapts the model based on computation of an approximate operating point. It selects the relaxed constraints likely to cause relaxation errors and replaces them with linear equality alternatives that cannot result in relaxation errors. Its effectiveness and implementation are presented in paper [P_1]. The results indicate that on stable test cases, i.e. without relaxation errors, the presolve keeps more accurate quadratic constraints. However, in less stable test cases, where relaxation errors would otherwise occur, it successfully selects the correct constraints for replacement with the nonrelaxed constraints.

3.3 Bilevel optimization model of an AC power-transmission-system-constrained day-ahead electricity market participation

The last part of the contribution focuses on accuracy and numerical tractability of bilevel modeling of AC power-transmission-system constrained day-ahead electricity market participation. The exact AC optimal power flow is a nonconvex formulation and thus difficult to solve when in the lower level since it does not satisfy the mathematical regularity requirements. This thesis presents two approaches to solve this problem: using the developed convex AC optimal power flow model and a single-level reduction technique with smoothing of complementary conditions, as presented in two-part papers [P_2] and [P_3] and using a convolutional neural network metamodel to bypass the lower level, as presented in [P_4]. The first part of the two part paper [P_2] presents the model and the algorithm of the bilevel model, while the second part [P_3] presents case studies for accuracy demonstration, economic profit due to increased model accuracy and comparison of the solution techniques. The results indicate that the smoothing technique has the best tractability out of all single-level reduction-based techniques that are able to almost fully close the duality gap. Comparing the two approaches, i.e. the single-level reduction and the metamodel, the results indicate that the single-level reduction achieves greater accuracy and better computation times for smaller networks. However, for larger networks, the single-level reduction approach diverges, which is not the case with the metamodel. Thus, the

metamodel approach is more suitable for more difficult large cases.

Chapter 4

List of Publications

The publications published within this thesis and considered as the main contribution are divided into two sections: journal papers and conference papers. Other published papers can be found under biography Chapter 6.2.

4.1 Journal Papers

Published

[P_1]K. Šepetanc and H. Pandži ć, "Convex Polar Second-Order Taylor Approximation of AC Power Flows: A Unit Commitment Study," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3585-3594, July 2021,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2020.304697

[P_2]K. Šepetanc, H. Pandži ć and T. Capuder, "Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part I: Model Description and the Algorithm," *IEEE Transactions on Power Systems*, 2022,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2022.3207088

[P_3]K. Šepetanc, H. Pandži ć and T. Capuder, "Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part II: Solution Techniques and Case Study," *IEEE Transactions on Power Systems*, 2022,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2022.3207097

[P_4]D. Vlah, K. Šepetanc and H. Pandži ć, "Solving Bilevel Optimal Bidding Problems Using Deep Convolutional Neural Networks," *IEEE System Journal*, 2023,
ISSN: 1937-9234, DOI: 10.1109/JSYST.2022.3232942

4.2 Conference Papers

Published and Presented

[P_5]O. Heide, K. Šepetanc and H. Pandži ć, "Transmission Expansion Planning using a Highly Accurate AC Optimal Power Flow Approximation," in *2021 International Conference on Smart Energy Systems and Technologies (SEST)*, 2021, pp. 1-6, 10.1109/SEST50973.2021.9543431

Chapter 5

Author's Contribution to the Publications

The contribution of this thesis was achieved during the period of 2019–2022 at the University of Zagreb Faculty of Electrical Engineering and Computing, Unska 3, HR-10000 Zagreb, Croatia. The research was conducted under under the projects listed below:

- Project *Advanced Tools Towards cost-efficient decarbonisation of future reliable Energy SysTems* (ATTEST), founded by European Union's Horizon 2020 research and innovation programme,
- Project *A novel smart grid architecture that facilitates high RES penetration through innovative markets towards efficient interaction between advanced electricity grid management and intelligent stakeholders* (FLEXGRID), founded by European Union's Horizon 2020 research and innovation programme

The employment of Karlo Šepetanc at University of Zagreb Faculty of Electrical Engineering and Computing was fully funded by the Croatian Science Foundation under programme DOK-2018-09.

The author's main contribution to each paper is listed below:

[P_1] In the journal paper *"Convex Polar Second-Order Taylor Approximation of AC Power Flows: A Unit Commitment Study"* : literature review, development and implementation of the optimal power flow model, development and implementation of the presolve technique, creation of the case studies, paper writing and results elaboration.

[P_2] In the journal paper *"Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part I: Model Description and the Algorithm"* : proposal and implementation of a bilevel market participation model, proposal of the solution algorithm and paper writing.

[P_3] In the journal paper *"Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part II: Solution Techniques and Case Study"* : listing of the solution techniques, creation of the case studies, paper writing and results elaboration.

[P_4] In the journal paper *"Solving Bilevel Optimal Bidding Problems Using Deep Convolu-*

tional Neural Networks" : literature review, proposal of the smooth neural network activation function, training dataset creation, creation of the case studies, paper writing and results elaboration.

[P_5]In the conference paper *"Transmission Expansion Planning using a Highly Accurate AC Optimal Power Flow Approximation"* : proposal of suitable optimal power flow model, paper writing and live presentation.

Chapter 6

Conclusions and Future Work

6.1 Main Conclusions of the Thesis

The developed convex AC optimal power flow model demonstrated higher accuracy than the existing convex models. This is a result of the new power flow equations developed using extensive numerical experiments to determine the importance of various equation terms. The developed optimal power flow model works in tandem with a presolve to avoid relaxation errors and retain high accuracy even in otherwise unfavourable cases. With the new optimal power flow and the presolve, discrete power system problems, such as unit commitment and transmission expansion planning, can now be solved with better computation times than with the exact models and with better accuracy and similar computation times than using the existing approximations and relaxations.

The thesis proposes two new bilevel AC transmission system constrained market participation models. One is based on the KKT single-level reduction, smoothing of the complementary conditions and the developed convex AC optimal power flow model. This approach solves the problem accurately, but solution process diverges for large networks. The second model replaces the lower level with convolutional neural network metamodel. The accuracy of this approach is high, but lower than the accuracy of the first model. Computation times are as well longer for the metamodel approach. However, unlike the KKT approach, the metamodel solution process converges for larger networks. As a result, using the developed two bilevel models, both large and small cases can be solved to high accuracy.

6.2 Future Work

The future research may diverge in multiple directions. One direction is the simplification the presolve to compute the required constraint marginals directly from the approximate operating point. The second one is to further improve numerical tractability of bilevel models based on

KKT single-level reduction approach. Furthermore, future work may also focus on applying the developed models to real-life applications.

Bibliography

[1]Coffrin, C., Hijazi, H., Hentenryck, P. V., “Network flow and copper plate relaxations for ac transmission systems”. IEEE, 6 2016, pp. 1-8.

[2]Stott, B., Jardim, J., Alsac, O., “Dc power flow revisited”, IEEE Transactions on Power Systems, Vol. 24, 8 2009, pp. 1290-1300.

[3]Frank, S., Rebennack, S., “An introduction to optimal power flow: Theory, formulation, and examples”, IIE Transactions, Vol. 48, 12 2016, pp. 1172-1197.

[4]Torres, G., Quintana, V., “An interior-point method for nonlinear optimal power flow using voltage rectangular coordinates”, IEEE Transactions on Power Systems, Vol. 13, 1998, pp. 1211-1218.

[5]Sun, D., Ashley, B., Brewer, B., Hughes, A., Tinney, W., “Optimal power flow by newton approach”, IEEE Transactions on Power Apparatus and Systems, Vol. PAS-103, 10 1984, pp. 2864-2880.

[6]O’neill, R. P., Castillo, A., Cain, M. B., Cain, M., “The iv formulation and linear approximations of the ac optimal power flow problem”, 2013.

[7]Molzahn, D. K., Hiskens, I. A., “Moment-based relaxation of the optimal power flow problem”. IEEE, 8 2014, pp. 1-7.

[8]Dillon, T. S., Edwin, K. W., Kochs, H.-D., Taud, R. J., “Integer programming approach to the problem of optimal unit commitment with probabilistic reserve determination”, IEEE Transactions on Power Apparatus and Systems, Vol. PAS-97, 11 1978, pp. 2154-2166.

[9]Jabr, R., “Radial distribution load flow using conic programming”, IEEE Transactions on Power Systems, Vol. 21, 8 2006, pp. 1458-1459.

[10]Gan, L., Li, N., Topcu, U., Low, S. H., “Exact convex relaxation of optimal power flow in radial networks”, IEEE Transactions on Automatic Control, Vol. 60, No. 1, 2015, pp. 72-87.

[11]“Pglib opf”, Available: https://github.com/power-grid-lib/pglib-opf/blob/master/BASELINE.md

[12]Hijazi, H., Coffrin, C., Hentenryck, P. V., “Convex quadratic relaxations for mixed-integer nonlinear programs in power systems”, Mathematical Programming Computation, Vol. 9, 9 2017, pp. 321-367.

[13]Bai, X., Wei, H., Fujisawa, K., Wang, Y., “Semidefinite programming for optimal power flow problems”, International Journal of Electrical Power & Energy Systems, Vol. 30, 7 2008, pp. 383-392.

[14]Jabr, R. A., “Exploiting sparsity in sdp relaxations of the opf problem”, IEEE Transactions on Power Systems, Vol. 27, 5 2012, pp. 1138-1139.

[15], Available: http://www.opt.tu-darmstadt.de/scipsdp/

[16]“Bnb”, Available: https://yalmip.github.io/solver/bnb/ Sep 2016.

[17]“Cutsdp”, Available: https://yalmip.github.io/solver/cutsdp/ Sep 2016.

[18]Zhang, H., Heydt, G. T., Vittal, V., Quintero, J., “An improved network model for transmission expansion planning considering reactive power and network losses”, IEEE Transactions on Power Systems, Vol. 28, 8 2013, pp. 3471-3479.

[19]Yang, Z., Zhong, H., Bose, A., Zheng, T., Xia, Q., Kang, C., “A linearized opf model with reactive power and voltage magnitude: A pathway to improve the mw-only dc opf”, IEEE Transactions on Power Systems, Vol. 33, 3 2018, pp. 1734-1745.

[20]“Alternating current (ac) power flow analysis in an electrical power network - wo2013173879a1”, Available: https://patents.google.com/patent/WO2013173879A1/zh

[21]Coffrin, C., Hentenryck, P. V., “A linear-programming approximation of ac power flows”, INFORMS Journal on Computing, Vol. 26, 11 2014, pp. 718-734.

[22]“Powermodels”, Available: https://lanl-ansi.github.io/PowerModels.jl/stable/

[23]Momber, I., Wogrin, S., Gómez San Román, T., “Retail pricing: A bilevel program for pev aggregator decisions using indirect load control”, IEEE Transactions on Power Systems, Vol. 31, No. 1, 2016, pp. 464-473.

[24]Pandži ć, H., Kuzle, I., “Energy storage operation in the day-ahead electricity market”, in 2015 12th International Conference on the European Energy Market (EEM), 2015, pp. 1-6.

[25]Pandzic, H., Conejo, A. J., Kuzle, I., Caro, E., “Yearly maintenance scheduling of transmission lines within a market environment”, IEEE Transactions on Power Systems, Vol. 27, No. 1, 2012, pp. 407-415.

[26]Garces, L. P., Conejo, A. J., Garcia-Bertrand, R., Romero, R., “A bilevel approach to transmission expansion planning within a market environment”, IEEE Transactions on Power Systems, Vol. 24, No. 3, 2009, pp. 1513-1522.

[27]Fortuny-Amat, J., McCarl, B., “A representation and economic interpretation of a two-level programming problem”, Journal of the Operational Research Society, Vol. 32, 9 1981, pp. 783-792.

[28]Edmunds, T., Bard, J., “Algorithms for nonlinear bilevel mathematical programs”, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 21, 1991, pp. 83-89.

[29]Zolfaghari, S., Akbari, T., “Bilevel transmission expansion planning using second-order cone programming considering wind investment”, Energy, Vol. 154, 7 2018, pp. 455-465.

[30]Kolstad, C. D., Lasdon, L. S., “Derivative evaluation and computational experience with large bilevel mathematical programs”, Journal of Optimization Theory and Applications, Vol. 65, 6 1990, pp. 485-499.

[31]Aiyoshi, E., Shimizu, K., “A solution method for the static constrained stackelberg problem via penalty method”, IEEE Transactions on Automatic Control, Vol. 29, 12 1984, pp. 1111-1114.

[32]Ishizuka, Y., Aiyoshi, E., “Double penalty method for bilevel optimization problems”, Annals of Operations Research, Vol. 34, 12 1992, pp. 73-88.

[33]Colson, B., Marcotte, P., Savard, G., “A trust-region method for nonlinear bilevel programming: Algorithm and computational experience”, Computational Optimization and Applications, Vol. 30, 3 2005, pp. 211-227.

[34]Li, X., Tian, P., Min, X., “A hierarchical particle swarm optimization for solving bilevel programming problems”, pp. 1169-1178, 2006.

[35]Zhu, X., Yu, Q., Wang, X., “A hybrid differential evolution algorithm for solving nonlinear bilevel programming with linear constraints”. IEEE, 7 2006, pp. 126-131.

[36]Angelo, J. S., Krempser, E., Barbosa, H. J., “Differential evolution for bilevel programming”. IEEE, 6 2013, pp. 470-477.

[37]Hejazi, S., Memariani, A., Jahanshahloo, G., Sepehri, M., “Linear bilevel programming solution by genetic algorithm”, Computers & Operations Research, Vol. 29, 11 2002, pp. 1913-1925.

[38]Sinha, A., Malo, P., Deb, K., “Evolutionary algorithm for bilevel optimization using approximations of the lower level optimal solution mapping”, European Journal of Operational Research, Vol. 257, 3 2017, pp. 395-411.

[39]Sinha, A., Malo, P., Deb, K., “Solving optimistic bilevel programs by iteratively approximating lower level optimal value function”. IEEE, 7 2016, pp. 1877-1884.

Nomenclature

$S_{i,j}$	Apparent power between nodes i and j
$T_{i,j}$	Transformer between node i and j tap ratio
V_i	Voltage at node i
$Y_{i,j}^{sh}$	Line or transformer between nodes i and j shunt admittance
$Y_{i,j}$	Line or transformer between nodes i and j series admittance

List of Figures

List of Tables

Abbreviation

ACAlternating current

DCDirect current

KKTKarush-Kuhn-Tucker

NLPNonlinear programming

LPLinear programming

LPACLinear programming AC

MILPMixed-integer linear programming

QCQPQuadratically constrained quadratic programming

QPACQuadratic programming AC

SOCPSecond-order cone programming

QCQuadratically constrained

SDPSemidefinite programming

Publications

There are in total 4 journal papers and 1 presented conference paper under this thesis. All journal and conference papers are attached bellow.

Journal Papers

[P_1]K. Šepetanc and H. Pandži ć, "Convex Polar Second-Order Taylor Approximation of AC Power Flows: A Unit Commitment Study," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3585-3594, July 2021,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2020.304697

[P_2]K. Šepetanc, H. Pandži ć and T. Capuder, "Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part I: Model Description and the Algorithm," *IEEE Transactions on Power Systems*, 2022,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2022.3207088

[P_3]K. Šepetanc, H. Pandži ć and T. Capuder, "Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part II: Solution Techniques and Case Study," *IEEE Transactions on Power Systems*, 2022,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2022.3207097

[P_4]D. Vlah, K. Šepetanc and H. Pandži ć, "Solving Bilevel Optimal Bidding Problems Using Deep Convolutional Neural Networks," *IEEE System Journal*, 2023,
ISSN: 1937-9234, DOI: 10.1109/JSYST.2022.3232942

Published and Presented

[P_5]O. Heide, K. Šepetanc and H. Pandži ć, "Transmission Expansion Planning using a Highly Accurate AC Optimal Power Flow Approximation," in *2021 International Conference on Smart Energy Systems and Technologies (SEST)*, 2021, pp. 1-6,
10.1109/SEST50973.2021.9543431

Convex Polar Second-Order Taylor Approximation of AC Power Flows: A Unit Commitment Study

Karlo Šepetanc, *Student Member, IEEE*, and Hrvoje Pandžić, *Senior Member, IEEE*

***Abstract*—Modern mixed-integer quadratic solvers generally handle binary variables more efficiently than nonlinear mixed-integer solvers. This is relevant to the power system operation models as the unit commitment formulations typically contain a large number of binary variables. This paper investigates how to achieve the accuracy level close to the one of the exact nonlinear models, but by utilising convex models and solvers. The presented unit commitment model is based on a Taylor-series expansion where both the voltage magnitude and angle are quadratically constrained. To achieve high accuracy, the model takes advantage of the meshed transmission network structure that enables replacement of the quadratic inequality constraints that cause constraint relaxation errors with the linear equality constraints. Quadratic constraints to be replaced as well as the operating point parameters are determined based on the presolve. The first presented case study validates the model's accuracy and the convergence of the iterative algorithm, while the second is a non-iterative full unit commitment problem. Unit commitment results show superior accuracy and similar computation times to the existing quadratic formulations on one hand and faster computation times than the exact nonlinear polar formulation on the other.**

***Index Terms*—Mixed-integer quadratically constrained quadratic program, network-constrained unit commitment, optimal power flow approximation.**

NOMENCLATURE

Sets and Indices

N	Set of buses, indexed by i and j.
N^{P}	Tuple set of paired buses aligned with branch E orientation, indexed by (i,j).
R	Set of reference buses, indexed by i.
E, E^{R}	Tuple set of branches, forward and reverse orientation, indexed by (e,i,j).
E_i, E_i^{R}	Array of tuple sets of branches at bus i, forward and reverse orientation, indexed by (e,i,j).
G, G_i	Set of all generators and array of sets of generators at bus i, indexed by k.
L_i	Array of sets of loads at bus i, indexed by l.
S_i	Array of sets of shunts at bus i, indexed by s.
τ	Set of time steps, indexed by t and h.
Ξ	Set of decision variables.

Parameters

$\ddot{\boldsymbol{c}}_k, \dot{\boldsymbol{c}}_k, \boldsymbol{c}_k$	Generator cost coefficients.
$\boldsymbol{P}^{\mathrm{d}}_{t,l}, \boldsymbol{Q}^{\mathrm{d}}_{t,l}$	Active and reactive power load.
$\boldsymbol{g}^{\mathrm{sh}}_s, \boldsymbol{b}^{\mathrm{sh}}_s$	Bus shunt conductance and susceptance.
$\boldsymbol{g}_e, \boldsymbol{g}^{\mathrm{fr}}_e, \boldsymbol{g}^{\mathrm{to}}_e$	Branch π-section conductances.
$\boldsymbol{b}_e, \boldsymbol{b}^{\mathrm{fr}}_e, \boldsymbol{b}^{\mathrm{to}}_e$	Branch π-section susceptances.
$\boldsymbol{\tau}_e, \boldsymbol{\sigma}_e$	Branch tap magnitude and shift angle.
$\underline{\boldsymbol{P}}^{\mathrm{g}}_k, \overline{\boldsymbol{P}}^{\mathrm{g}}_k$	Generator minimum and maximum active power production.
$\underline{\boldsymbol{Q}}^{\mathrm{g}}_k, \overline{\boldsymbol{Q}}^{\mathrm{g}}_k$	Generator minimum and maximum reactive power production.
$\overline{\boldsymbol{S}}_e$	Branch maximum apparent power.
$\underline{\boldsymbol{\theta}}_{i,j}, \overline{\boldsymbol{\theta}}_{i,j}$	Bus-pair minimum and maximum voltage angle difference.
$\underline{\boldsymbol{V}}_i, \overline{\boldsymbol{V}}_i$	Bus minimum and maximum voltage magnitude.
$\boldsymbol{V}^{\mathrm{op}}_{t,i}, \boldsymbol{\theta}^{\mathrm{op}}_{t,i}$	Assumed bus voltage magnitude and angle operating points.
$\boldsymbol{c}^{\mathrm{su}}_k$	Generator start-up cost.
$\boldsymbol{RU}_k, \boldsymbol{RD}_k$	Generator ramp-up and -down limits.
$\boldsymbol{MU}_k, \boldsymbol{MD}_k$	Generator minimum up and down time.
$\Lambda_{t,e}, \Gamma_{t,i,j}$	Boolean parameters which indicate whether to use quadratic form of voltage and cosine representations respectively.

Variables

Continuous variables

$P^{\mathrm{g}}_{t,k}, Q^{\mathrm{g}}_{t,k}$	Generator active and reactive power production.
$P_{t,e,i,j}, Q_{t,e,i,j}$	Branch active and reactive power flow.
$V^{\Delta}_{t,i}, \theta^{\Delta}_{t,i}$	Bus voltage magnitude and angle change.
$V_{t,i}, \theta_{t,i}$	Bus voltage magnitude and angle.
$\widehat{cos}_{t,i,j}$	Cosine approximation.

Manuscript received July 9, 2020; revised November 3, 2020; accepted December 19, 2020. Date of publication December 23, 2020; date of current version June 18, 2021. The work of Karlo Šepetanc was supported by the Croatian Science Foundation under programme DOK-2018-09. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 864 298 (project ATTEST). Paper no. TPWRS-01 146-2020. *(Corresponding author: Hrvoje Pandzic.)*

The authors are with the Innovation Centre Nikola Tesla, 10000 Zagreb, Croatia, and the University of Zagreb Faculty of Electrical Engineering and Computing, Zagreb 10000, Croatia (e-mail: karlo.sepetanc@fer.hr; hrvoje.pandzic@ieee.org).

Digital Object Identifier 10.1109/TPWRS.2020.3046970

$\breve{V}_{t,e}$ Second-order Taylor series voltage magnitude term approximation.

Binary variables

$x_{t,k}$ Generator activity state indicator.

$y_{t,k}, z_{t,k}$ Generator start-up and shut-down indicators.

I. Introduction

A. Motivation

UNIT commitment is an optimization problem that determines the least-cost production of the generators to satisfy the demand while considering the generators' physical limitations. Besides generator capacity limits, it typically encompasses output ramp limits, minimum up and down times and start-up costs, whose modeling requires computationally expensive binary variables. Computational burden of the unit commitment problem can be assessed from two perspectives: the binary formulation tractability and the grid formulation tractability.

Novel binary formulations are typically studied without the network constraints to better demonstrate and isolate the source of numerical difficulties. To this end, there is even a unit commitment benchmark generally accepted by the scientific community [1], but without the network constraints. However, the ISO-type electricity markets consider transmission constraints already at the market-clearing phase. Hence, we focus on the network-constrained unit commitment problem and deliver a new transmission system power flow formulation that retains an accuracy level close to the exact nonlinear AC models, but allows for the use of more specific and performant mixed-integer quadratically constrained quadratic program (MIQCQP) solvers as opposed to the mixed-integer nonlinear (MINLP) ones. It is a common practice in transmission system modeling to reduce a MINLP to a mixed-integer linear (MILP), e.g. by applying DC network approximation. However, linear models are inaccurate when it comes to modeling reactive power flows, voltage magnitudes and losses. On the other hand, the existing convex quadratic approximations or relaxations of AC power flows may achieve good or even perfect accuracy when there are no quadratic constraint relaxation errors. These are a consequence of the convexification process that requires the quadratic equality constraints to be relaxed into quadratic inequalities. However, when such errors do occur, they are very large. Since unit commitment is a multi-period optimization problem thus simulating the grid under various conditions, including generating units nonconvexities, the chances of having relaxation errors in at least one of the simulated periods are relatively high, rendering the relaxation model inaccurate. A motivation to develop a model based on the Taylor expansion comes from the thought that constraint relaxation errors, which in this case can occur due to convexification of the second-order parts of the expansion, can be avoided without otherwise gross changes to the power flow equations by upfront neglecting these terms on per-branch basis as determined by the presolve. Constraint relaxation errors due to convexification are avoided since when the second-order terms are neglected, the resulting constraints are linear equalities instead of quadratic inequalities. Additionally, meshed transmission network structure acts favourably as it provides a sufficient number of quadratic constraints to preserve the accuracy and iterative convergence of the algorithm that reruns the model around an updated Taylor operating point despite some neglected second-order terms. A parallel can be drawn to the Newton's power flow calculation method that converges faster when using higher-order modifications [2]. Fast convergence is important as it allows us to achieve high warm-start single iteration accuracy using the approximate warm-start operating point parameters obtained by first solving the problem in its continuous version where binary variables are relaxed into continuous in the range from 0 to 1. Utilising the described features and the proposed transmission system power flow formulation, we solve the network-constrained unit commitment problem while cooptimizing the real and reactive powers to exploit the unused monetary value in the traditional separate optimization, as demonstrated in [3], but by using an implementation without any loop statements.

B. Literature Review

Unit commitment research started with the development of the branch-and-bound algorithm [4], which is the basis of modern mixed-integer solvers. The MILP approach is still considered as the state-of-the-art due to its computational tractability. The early works grasped the unit commitment problem in its simple form without the network constraints, thus focusing on tightening the binary formulation, which can be expressed using three such variables per generator, as in [5], [6] and [7], or using a single binary variable as in [8]. Because of reliance on the branch-and-bound algorithm to solve mixed-integer problems, less variables does not necessarily imply better computational tractability. Size of the problem can be decreased by clustering similar generators [9], but this requires simplifications that reduce accuracy and applicability.

The subsequent unit commitment research branches out in multiple directions, mainly focusing on uncertainties [10] security constraints [11], and network constraints [12]. Security constraints add an additional contingency scenario dimension to the unit commitment problem drastically increasing the problem size. Our work and this literature review are focused on network constraints, whose inclusion in the unit commitment model also has a detrimental effect on the computational time. The inclusion of both security and network constraints forms an even more demanding problem, however in this work we focus on the network constraints to better isolate their difficulties and features. To reduce the problem complexity, DC optimal power flow is widely used. This transmission grid approximation results in good accuracy of the active power flows [13]. There are various attempts to generalize the DC model to include losses and reactive power, e.g., by expanding the first-order Taylor series around the operating point and adding only the voltage-angle-dependent nonconvex piece-wise linear losses [14], which require integer variables, or by linear loss estimation [15], which needs a penalty factor to prevent negative losses. Quadratic approximations [16] are much more accurate, but also more

TABLE I
OVERVIEW OF OPTIMAL POWER FLOW TECHNIQUES (OP – OPERATING POINT)

Name	Type	Form	Convex	Voltage	Reactive power	Losses
DC [13]	Approx.	LP	Yes	1 p.u.	No	No
DC extensions [14], [15]	Approx.	MILP/LP	No/Yes	OP+Var.*	Yes	Yes*
LPAC [17]	Approx.	LP/QCQP	Yes	OP+Var.	Yes	Yes
QPAC [16]	Approx.	QCQP	Yes	Variable	Yes	Yes
Jabr's [19]	Relax.	SOCP	Yes	Variable	Yes	Yes
QC [20]	Relax.	SOCP	Yes	Variable	Yes	Yes
Shor's [21]	Relax.	SDP	Yes	Variable	Yes	Yes
IV [23]	Exact	QCQP	No	Variable	Yes	Yes
Rectangular [24], [25]	Exact	QCQP	No	Variable	Yes	Yes
Polar [24], [26]	Exact	NLP	No	Variable	Yes	Yes

*some simple formulations do not include voltage variables nor losses

computationally demanding than the linear models. Work [17] proposes a Taylor-based piece-wise linearization with no integer variables, i.e. linear programming approximation of AC power flows (LPAC) of the initially quadratic approximation, that considers angle-only dependent losses. In a post-publication in PowerModels package [18], the author implemented an enhanced variant of the model with better accuracy by using the quadratic losses constraint instead of the piece-wise linear one.

Relaxations, on the other hand, have less persistent accuracy that is highly dependent on the test case. Performance of Jabr's (Second-order Cone Programming – SOCP) [19], quadratic-convex (QC) [20] and Shor's (semi-definite programming – SDP) [21] relaxations were analysed in [22], which showed that Jabr's relaxation is dominated in terms of tightness by the both remaining formulations. Finally, there are nonconvex exact rectangular current-voltage [23] and voltage-based rectangular [24], [25] and polar [24], [26] formulations that can be directly utilised for unit commitment, but with the highest computational burden. An overview of the described optimal power flow (OPF) techniques is provided in Table I. Our work builds upon the quadratic implementation of the LPAC model by introducing the voltage magnitude-dependent losses and by expanding the Taylor series around a general operating point.

C. Paper Contribution and Structure

Contribution of the paper consists of the following:

- We develop new transmission network AC equations based on the Taylor's expansion that approximates the second-order voltage terms. The approximation consists of distributing power losses to both branch ends based on the forward- and reverse-orientation power flows.
- We develop a presolve technique for deciding whether to use the quadratic or the linear form of power flow constraints to avoid constraint relaxation errors due to convexification.
- The resulting MIQCQP solution is obtained much quicker than the MINLP solution without sacrificing accuracy.

Rest of the paper is structured as follows. Section II mathematically derives and states the proposed model. It is divided in three subsections: SubSection II-A presents the Taylor expansion analysis, SubSection II-B introduces the presolve technique and subSection II-C presents the model components. Case study Section III consists of the three main parts: the description and set-up III-A; and two case studies. In the first case study III-B we solve optimal power flow problem on a number of networks to demonstrate convergence and accuracy of the model. The second case study III-C solves unit commitment problems to demonstrate computational tractability and accuracy of the model. Section IV provides relevant conclusions and guidelines for future work.

II. MATHEMATICAL MODEL

A. Taylor Expansion Analysis

Our analysis of the optimal power flow starts from the polar formulation for the branch power flow in equations (1.1) and (1.2). For clarity of the analysis, a general branch, which encompasses lines and transformers according to the PowerModels 0.13 standard [18], is simplified to a line without the shunt sections, i.e. tap is $1\angle 0^\circ$ and $\boldsymbol{g}_e^{\mathbf{fr}}, \boldsymbol{g}_e^{\mathbf{to}}, \boldsymbol{b}_e^{\mathbf{fr}}, \boldsymbol{b}_e^{\mathbf{to}} = 0$. Also, to shorten the expressions, two substitutions are made: $\theta_{t,i,j} = \theta_{t,i} - \theta_{t,j}$ and $\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j} = \boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j}$. Otherwise, the final presented model (2.1)–(2.20) and all the benchmarked models include a general branch without any simplifications or substitutions.

$$P_{t,e,i,j} = V_{t,i}^2 \cdot \boldsymbol{g}_e - V_{t,i} \cdot V_{t,j}(\boldsymbol{g}_e \cdot \cos(\theta_{t,i,j}) + \boldsymbol{b}_e \cdot \sin(\theta_{t,i,j})) \tag{1.1}$$

$$Q_{t,e,i,j} = -V_{t,i}^2 \cdot \boldsymbol{b}_e - V_{t,i} \cdot V_{t,j}(\boldsymbol{g}_e \cdot \sin(\theta_{t,i,j}) - \boldsymbol{b}_e \cdot \cos(\theta_{t,i,j})) \tag{1.2}$$

Full second-order Taylor series for the branch active and reactive power flow is structurally written in expressions (1.3) and (1.4). The first line contains the zeroth-order part (in blue), the next three lines the first-order part (in green) and the last six lines the second-order part (in red) of the Taylor series expanded over variables $V_{t,i}$, $V_{t,j}$ and $\theta_{t,i,j}$ around the operating point parameters $\boldsymbol{V}^{\mathbf{op}}_{t,i}$, $\boldsymbol{V}^{\mathbf{op}}_{t,j}$ and $\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}$. Variables representing a change from the operating point (delta variables) are at the beginning of the row, while the corresponding coefficients are within the square brackets.

$$\begin{aligned}
&P_{t,e,i,j} = \\
&(\boldsymbol{V}^{\mathbf{op}}_{t,i})^2 \cdot \boldsymbol{g}_e - \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j}(\boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})) \\
&+V^{\Delta}_{t,i} \cdot [2 \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{g}_e - \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot (\boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))] \\
&-V^{\Delta}_{t,j} \cdot [\boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot (\boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))] \\
&-\theta^{\Delta}_{t,i,j} \cdot [\boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot (\boldsymbol{b}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}) - \boldsymbol{g}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))] \\
&+(V^{\Delta}_{t,i})^2 \cdot [\boldsymbol{g}_e] \\
&-V^{\Delta}_{t,i} \cdot V^{\Delta}_{t,j} \cdot [\boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})]
\end{aligned}$$

$$
\begin{aligned}
&+(V^{\Delta}_{t,j})^2\cdot 0\\
&+(\theta^{\Delta}_{t,i,j})^2\cdot[\frac{\boldsymbol{V}^{\mathbf{op}}_{t,i}\cdot\boldsymbol{V}^{\mathbf{op}}_{t,j}}{2}\cdot(\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})+\boldsymbol{b}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))]\\
&-V^{\Delta}_{t,i}\cdot\theta^{\Delta}_{t,i,j}\cdot[\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))]\\
&-V^{\Delta}_{t,j}\cdot\theta^{\Delta}_{t,i,j}\cdot[\boldsymbol{V}^{\mathbf{op}}_{t,i}\cdot(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))] \qquad (1.3)
\end{aligned}
$$

$$
\begin{aligned}
&Q_{t,e,i,j}=\\
&-(\boldsymbol{V}^{\mathbf{op}}_{t,i})^2\cdot\boldsymbol{b}_e+\boldsymbol{V}^{\mathbf{op}}_{t,i}\cdot\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))\\
&+V^{\Delta}_{t,i}\cdot[-2\cdot\boldsymbol{V}^{\mathbf{op}}_{t,i}\cdot\boldsymbol{b}_e+\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))]\\
&+V^{\Delta}_{t,j}\cdot[\boldsymbol{V}^{\mathbf{op}}_{t,i}\cdot(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))]\\
&-\theta^{\Delta}_{t,i,j}\cdot[\boldsymbol{V}^{\mathbf{op}}_{t,i}\cdot\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot(\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})+\boldsymbol{b}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))]\\
&-(V^{\Delta}_{t,i})^2\cdot[\boldsymbol{b}_e]\\
&+V^{\Delta}_{t,i}\cdot V^{\Delta}_{t,j}\cdot[\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})]\\
&+(V^{\Delta}_{t,j})^2\cdot 0\\
&-(\theta^{\Delta}_{t,i,j})^2\cdot[\frac{\boldsymbol{V}^{\mathbf{op}}_{t,i}\cdot\boldsymbol{V}^{\mathbf{op}}_{t,j}}{2}\cdot(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))]\\
&-V^{\Delta}_{t,i}\cdot\theta^{\Delta}_{t,i,j}\cdot[\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot(\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})+\boldsymbol{b}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))]\\
&-V^{\Delta}_{t,j}\cdot\theta^{\Delta}_{t,i,j}\cdot[\boldsymbol{V}^{\mathbf{op}}_{t,i}\cdot(\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})+\boldsymbol{b}_e\cdot\sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}))] \qquad (1.4)
\end{aligned}
$$

Expressions (1.3) and (1.4) are nonconvex because i) they are equalities, and no quadratic equality is convex; ii) $(V^{\Delta}_{t,j})^2$ term is multiplied by zero. Convexification of the polar AC OPF by Taylor series was studied in [17], where the model was obtained by ignoring all but one of the second-order terms, $(\theta^{\Delta}_{t,i,j})^2$, and by taking $\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j}=0$, which leaves the model voltage-wise loosely constrained, i.e. without the voltage second-order terms.

We performed extensive tests to determine the importance of different second-order terms by removing them one-by-one from the series in expressions (1.3) and (1.4) and analysing accuracy and iterative algorithm convergence which reruns the model around an operating point obtained from a previous solve. Since the Taylor series is nonconvex (quadratic equality), nonlinear IPOPT solver was used. As a result, all combinations without the $(V^{\Delta}_{t,i})^2$ term failed to algorithmically fully converge and the version with only $(V^{\Delta}_{t,i})^2$ and $(\theta^{\Delta}_{t,i,j})^2$ terms converged extremely slowly, requiring over 100 iterations. However, the versions with $(V^{\Delta}_{t,i})^2$, $V^{\Delta}_{t,i}\cdot V^{\Delta}_{t,j}$ and $(\theta^{\Delta}_{t,i,j})^2$ exhibited fast convergence and good first iteration accuracy, while $V^{\Delta}_{t,i}\cdot\theta^{\Delta}_{t,i,j}$ and $V^{\Delta}_{t,j}\cdot\theta^{\Delta}_{t,i,j}$ terms had little effect. Also, reactive power voltage second-order terms, which in the tests were shown to be better ignored than approximated, have small impact on convergence and overall accuracy due to an absence of reactive power in the objective function. Based on these results, we include the quadratic angle term $(\theta^{\Delta}_{t,i,j})^2$ convexified by a relaxation in constraint (1.5), as well as the voltage $(V^{\Delta}_{t,i})^2$ and $V^{\Delta}_{t,i}\cdot V^{\Delta}_{t,j}$ terms for active power flow by convex approximation in (1.6). The relaxation in constraints (1.5) and (1.6) refers to the swap of the equality sign with the inequality. However, since the constraints are obtained by the Taylor expansion, which is an approximation, even with inequality sign they are still an approximation, albeit convex.

$$
\widehat{cos}_{t,i,j}\le 1-\frac{(\theta^{\Delta}_{t,i,j})^2}{2},\quad \forall t,(i,j)\in N^{\mathrm{P}} \qquad (1.5)
$$

$$
\begin{aligned}
\breve{V}_{t,e}\ge\ &\boldsymbol{g}_e\cdot(V^{\Delta}_{t,i})^2-2\cdot\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i,j})\cdot V^{\Delta}_{t,i}\cdot V^{\Delta}_{t,j}\\
&+\boldsymbol{g}_e\cdot(V^{\Delta}_{t,j})^2,\quad \forall t,(e,i,j)\in E
\end{aligned} \qquad (1.6)
$$

Constraints (1.5) and (1.6) are embedded in the main model as constraints (2.9.1) and (2.8.1), respectively, without simplifications and substitutions introduced at the beginning of this section. To retain similarity with the existing formulations in the literature, the right-hand side of (1.5) is a second-order Taylor series of a cosine function and, once multiplied with a parameter in the power flow constraints (2.4)–(2.7), it forms a part of the power flow zero-order Taylor term and $(\theta^{\Delta}_{t,i,j})^2$ term. The approximation in expression (1.6) is obtained by summing the voltage $(V^{\Delta}_{t,i})^2$ and $V^{\Delta}_{t,i}\cdot V^{\Delta}_{t,j}$ Taylor terms from both the forward and the reverse orientations of the branch active power flow, effectively representing losses evenly distributed between both branch ends in active power constraints (2.4) and (2.5). Second-order voltage approximation is always at least marginally convex, assuming $\boldsymbol{g}_e>0$.

While the second-order voltage approximation from (1.6) could mathematically be applied to reactive power as well (by simply extracting and transferring factor $\boldsymbol{g}_e$ from the approximation into the active power flow constraints and $-\boldsymbol{b}_e$ for the reactive power flow constraints), it was determined that this approximation is in some cases inadequate for reactive power, as it leads more commonly to an infeasible model using the flat start operating point assumptions, i.e. $\boldsymbol{V}^{\mathbf{op}}_{t,i}=1$ p.u. and $\boldsymbol{\theta}^{\mathbf{op}}_{t,i}=0$ rad. Reactive power flows are generally about an order of magnitude lower than active power flows, while the dominant factor for the approximation in reactive power flows, susceptance $\boldsymbol{b}_e$, is an order of magnitude higher than the dominant factor, conductance $\boldsymbol{g}_e$, in active power flows. Applying approximation only to active power flows does not limit its purpose to better constrain voltages since active power has much stronger effect on the objective function and thus on constraining the voltages. Also, no accuracy drawbacks were observed when applying the approximation only on active power flows in conditions where $\boldsymbol{b}_e/\boldsymbol{g}_e$ was close to 1 and in lightly loaded networks. Constraint (2.8.1), the subsequent of constraint (1.6), is applied in the model as a constraint for every branch in the forward orientation E rather than bus-pairs N^{P}, so it can account for parallel conductances ($\boldsymbol{g}^{\mathbf{fr}}_e$ and $\boldsymbol{g}^{\mathbf{to}}_e$) and differing taps in parallel branches.

B. Presolve Technique

An important problem feature is a meshed structure of the transmission networks. It makes all the delta variables $V^{\Delta}_{t,i}$ and $\theta^{\Delta}_{t,i}$ quadratically constrained despite removing some quadratic inequality constraints and swapping them with the equality constraints since the delta variables appear in multiple instances of (2.8.1) and (2.9.1) constraints. Swapping the quadratic inequality constraints with linear equality constraints avoids constraint relaxation errors due to convexifications, i.e. errors due to a

deviance from the inequality boundary, without a significant loss in a single iteration accuracy or multiple iterations convergence. Constraint (2.8.1) is a second-order voltage approximation without the simplifications introduced in the analysis, while (2.8.2) is its linear alternative. Similarly, (2.9.1) and (2.9.2) are quadratic and linear representations of the Taylor series cosine. Quadratic forms of the constraints are used only if the respective Boolean parameter $\Lambda_{t,e}$ or $\Gamma_{t,i,j}$ is true and if conductance $\boldsymbol{g}_e$ is positive in the case of voltage approximation constraint (2.8.1).

Our proposition is to determine the value of the Boolean parameters, i.e. decide whether to use linear or quadratic forms of the voltage and cosine approximation constraints, in the presolve computation step. The AC unit commitment is a difficult problem mostly due to binary variables needed for generators in combination with network constraints, which slow down the solution process and limit the selection of solvers that can be used. However, if binary variables are fixed, i.e. replaced with a parameter, the problem is relatively easy to solve even if the network constraints are nonconvex or nonlinear. Furthermore, since the constraint marginal represents the sensitivity of the objective function on adding a small positive constant to the right-hand side of the constraint, sign of the equality constraint marginal, which is computed by default by, e.g., IPOPT and Knitro solvers, indicates if this constraint would be binding if it was relaxed into an inequality constraint. For constraint (2.8.1) to be binding, due to its greater-or-equal sign, $\check{V}_{t,e}$ should have a tendency to be as small as possible. Adding a positive constant to the right-hand side of (2.8.1) would in this binding scenario increase $\check{V}_{t,e}$ and thus worsen the objective function, i.e. the marginal would be positive. Oppositely, for constraint (2.9.1) to be binding, since its less-or-equal sign, its marginal needs to be negative. As a result, a nonconvex presolve with only quadratic approximation constraints (2.8.1) and (2.9.1) in the equality form and fixed binary variables is proposed. Values of the fixed variables are determined simultaneously with an approximate operating point as described in SubSection III-C. This way, the main unit commitment solve will have only the convex quadratic approximation constraints that the presolve flagged as binding. All other quadratic constraints are replaced with linear ones to avoid constraint relaxation errors.

C. Optimization Model

This section presents the whole network-constrained unit commitment model. The objective function (2.1) is a variable generation and start-up cost minimization. While the ISO markets typically consider piecewise linear cost curves, we tend to avoid further approximations and thus use the quadratic cost curves. Constraints (2.2) and (2.3) are the bus balance constraints, (2.4)–(2.7) are power flow equations which also contain second-order term approximation variables $\check{V}_{t,e}$ and $\widehat{cos}_{t,i,j}$ from (2.8.1)–(2.9.2). Constraints (2.10) and (2.11) are generator production constraints that disable production when a generator is inactive ($x_{t,k} = 0$). (2.12) is the branch apparent power constraint for both orientations, (2.13) is the reference bus angle constraint, (2.14) and (2.15) are voltage magnitude and bus-pair angle constraints, and (2.16) is the generator ramp-up and -down constraint. Constraints (2.17) and (2.18) model the interaction between the generator activity binary variable and the start-up and shut-down binary variables. Constraints (2.19) and (2.20) ensure generator minimum up and down time requirements. Such start-up and shut-down formulation using three binary variables was first presented in [6] and was proven efficient in [27]. The presented problem is of MIQCQP [28] class with convex objective function and constraints, except for binary variables.

$$\underset{\Xi}{\text{Min}} \sum_{t,k} (\ddot{\boldsymbol{c}}_k \cdot (P^{\text{g}}_{t,k})^2 + \dot{\boldsymbol{c}}_k \cdot P^{\text{g}}_{t,k} + \boldsymbol{c}_k \cdot x_{t,k} + \boldsymbol{c}^{\mathbf{su}}_k \cdot y_{t,k}) \tag{2.1}$$

$$\sum_{k \in G_i} P^{\text{g}}_{t,k} - \sum_{l \in L_i} \boldsymbol{P}^{\mathbf{d}}_{t,l} - \sum_{(e,i,j) \in E_i \cup E^{\text{R}}_i} P_{t,e,i,j} - ((\boldsymbol{V}^{\mathbf{op}}_{t,i})^2 + 2 \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot V^{\Delta}_{t,i}) \cdot \sum_{s \in S_i} \boldsymbol{g}^{\mathbf{sh}}_s = 0, \quad \forall t, i \tag{2.2}$$

$$\sum_{k \in G_i} Q^{\text{g}}_{t,k} - \sum_{l \in L_i} \boldsymbol{Q}^{\mathbf{d}}_{t,l} - \sum_{(e,i,j) \in E_i \cup E^{\text{R}}_i} Q_{t,e,i,j} + ((\boldsymbol{V}^{\mathbf{op}}_{t,i})^2 + 2 \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot V^{\Delta}_{t,i}) \cdot \sum_{s \in S_i} \boldsymbol{b}^{\mathbf{sh}}_s = 0, \quad \forall t, i \tag{2.3}$$

$$P_{t,e,i,j} = \frac{((\boldsymbol{V}^{\mathbf{op}}_{t,i})^2 + 2 \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot V^{\Delta}_{t,i}) \cdot (\boldsymbol{g}_e + \boldsymbol{g}^{\mathbf{fr}}_e)}{\boldsymbol{\tau}^2_e} + \frac{\check{V}_{t,e}}{2} - (\boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e)) \cdot (\boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot \widehat{cos}_{t,i,j} + V^{\Delta}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} + V^{\Delta}_{t,j} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i})/\boldsymbol{\tau}_e - (\boldsymbol{b}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e) - \boldsymbol{g}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e)) \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot (\theta^{\Delta}_{t,i} - \theta^{\Delta}_{t,j})/\boldsymbol{\tau}_e, \quad \forall t, (e,i,j) \in E \tag{2.4}$$

$$P_{t,e,i,j} = ((\boldsymbol{V}^{\mathbf{op}}_{t,i})^2 + 2 \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot V^{\Delta}_{t,i}) \cdot (\boldsymbol{g}_e + \boldsymbol{g}^{\mathbf{to}}_e) + \frac{\check{V}_{t,e}}{2} - (\boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e)) \cdot (\boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot \widehat{cos}_{t,j,i} + V^{\Delta}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} + V^{\Delta}_{t,j} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i})/\boldsymbol{\tau}_e - (\boldsymbol{b}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e) - \boldsymbol{g}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e)) \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot (\theta^{\Delta}_{t,i} - \theta^{\Delta}_{t,j})/\boldsymbol{\tau}_e, \quad \forall t, (e,i,j) \in E^{\text{R}} \tag{2.5}$$

$$Q_{t,e,i,j} = -\frac{((\boldsymbol{V}^{\mathbf{op}}_{t,i})^2 + 2 \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot V^{\Delta}_{t,i}) \cdot (\boldsymbol{b}_e + \boldsymbol{b}^{\mathbf{fr}}_e)}{\boldsymbol{\tau}^2_e} + (\boldsymbol{b}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e) - \boldsymbol{g}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e)) \cdot (\boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot \widehat{cos}_{t,i,j} + V^{\Delta}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} + V^{\Delta}_{t,j} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i})/\boldsymbol{\tau}_e - (\boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e)) \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot (\theta^{\Delta}_{t,i} - \theta^{\Delta}_{t,j})/\boldsymbol{\tau}_e, \quad \forall t, (e,i,j) \in E \tag{2.6}$$

$$Q_{t,e,i,j} = -((\boldsymbol{V}^{\mathbf{op}}_{t,i})^2 + 2 \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot V^{\Delta}_{t,i}) \cdot (\boldsymbol{b}_e + \boldsymbol{b}^{\mathbf{to}}_e)$$
$$+ (\boldsymbol{b}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e) - \boldsymbol{g}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e)) \cdot$$
$$(\boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot \widehat{cos}_{t,j,i} + V^{\Delta}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} + V^{\Delta}_{t,j} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,i}) / \boldsymbol{\tau}_e$$
$$- (\boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e)) \cdot$$
$$\boldsymbol{V}^{\mathbf{op}}_{t,i} \cdot \boldsymbol{V}^{\mathbf{op}}_{t,j} \cdot (\theta^{\Delta}_{t,i} - \theta^{\Delta}_{t,j}) / \boldsymbol{\tau}_e, \quad \forall t, (e,i,j) \in E^{\mathrm{R}} \tag{2.7}$$

$$\breve{V}_{t,e} \geq \frac{\boldsymbol{g}_e + \boldsymbol{g}^{\mathbf{fr}}_e}{\boldsymbol{\tau}^2_e} \cdot (V^{\Delta}_{t,i})^2 - \frac{2 \cdot \boldsymbol{g}_e}{\boldsymbol{\tau}_e} \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e) \cdot V^{\Delta}_{t,i} \cdot V^{\Delta}_{t,j}$$
$$+ (\boldsymbol{g}_e + \boldsymbol{g}^{\mathbf{to}}_e) \cdot (V^{\Delta}_{t,j})^2, \quad \forall t, (e,i,j) \in E : \boldsymbol{g}_e > 0 \wedge \Lambda_{t,e} \tag{2.8.1}$$

$$\breve{V}_{t,e} = 0, \quad \forall t, (e,i,j) \in E : \boldsymbol{g}_e \leq 0 \vee \neg \Lambda_{t,e} \tag{2.8.2}$$

$$\widehat{cos}_{t,i,j} \leq 1 - \frac{(\theta^{\Delta}_{t,i} - \theta^{\Delta}_{t,j})^2}{2}, \quad \forall t, (i,j) \in N^{\mathrm{P}} : \Gamma_{t,i,j} \tag{2.9.1}$$

$$\widehat{cos}_{t,i,j} = 1, \quad \forall t, (i,j) \in N^{\mathrm{P}} : \neg \Gamma_{t,i,j} \tag{2.9.2}$$

$$\underline{\boldsymbol{P}}^{\mathbf{g}}_k \cdot x_{t,k} \leq P^{\mathrm{g}}_{t,k} \leq \overline{\boldsymbol{P}}^{\mathbf{g}}_k \cdot x_{t,k}, \quad \forall t, k \tag{2.10}$$

$$\underline{\boldsymbol{Q}}^{\mathbf{g}}_k \cdot x_{t,k} \leq Q^{\mathrm{g}}_{t,k} \leq \overline{\boldsymbol{Q}}^{\mathbf{g}}_k \cdot x_{t,k}, \quad \forall t, k \tag{2.11}$$

$$P^2_{t,e,i,j} + Q^2_{t,e,i,j} \leq \overline{\boldsymbol{S}}^2_e, \quad \forall t, (e,i,j) \in E \cup E^{\mathrm{R}} : \exists \overline{\boldsymbol{S}}_e \tag{2.12}$$

$$\boldsymbol{\theta}^{\mathbf{op}}_{t,i} + \theta^{\Delta}_{t,i} = 0, \quad \forall t, i \in R \tag{2.13}$$

$$\underline{\boldsymbol{V}}_i \leq \boldsymbol{V}^{\mathbf{op}}_{t,i} + V^{\Delta}_{t,i} \leq \overline{\boldsymbol{V}}_i, \quad \forall t, i \tag{2.14}$$

$$\underline{\boldsymbol{\theta}}_{i,j} \leq (\boldsymbol{\theta}^{\mathbf{op}}_{t,i} + \theta^{\Delta}_{t,i}) - (\boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \theta^{\Delta}_{t,j}) \leq \overline{\boldsymbol{\theta}}_{i,j}, \ \forall t, (i,j) \in N^{\mathrm{P}} \tag{2.15}$$

$$\boldsymbol{RD}_k \leq P^{\mathrm{g}}_{t,k} - P^{\mathrm{g}}_{t-1,k} \leq \boldsymbol{RU}_k, \quad \forall t, k \tag{2.16}$$

$$y_{t,k} - z_{t,k} = x_{t,k} - x_{t-1,k}, \quad \forall t, k \tag{2.17}$$

$$y_{t,k} + z_{t,k} \leq 1, \quad \forall t, k \tag{2.18}$$

$$\sum_{h=t-\boldsymbol{MU}_k+1}^{t} y_{h,k} \leq x_{t,k}, \quad \forall t, k \tag{2.19}$$

$$\sum_{h=t-\boldsymbol{MD}_k+1}^{t} z_{h,k} \leq 1 - x_{t,k}, \quad \forall t, k \tag{2.20}$$

III. Case Study

A. Description and Set-Up

Although the main case study is elaborated in SubSection III-C, where we demonstrate the effectiveness of the proposed model, we start by applying the model on a number of OPF problems to show convergence speed and accuracy of the model. In all cases we compare the obtained results, both in terms of accuracy and computational effort, to the exact nonlinear polar model [24], i.e. based on constraints (1.1) and (1.2), which serves as a reference. Additionally, in the unit commitment case study, our results are compared with the LPAC approximation in the warm-start and the quadratic implementation [17], [18] and with the QC relaxation [20]. Our implementations of the existing models were verified in a side-by-side comparison to match the PowerModels [18] implementation.

Algorithm 1: Convergence and Accuracy Demonstration.

1: Run exact polar model ▷ results in Table II
2: $\boldsymbol{V}^{\mathbf{op}}_{t,i} \leftarrow 1$ p.u.; $\boldsymbol{\theta}^{\mathbf{op}}_{t,i} \leftarrow 0$ rad
3: **repeat**
4: Run nonconvex model
5: Select constraints by evaluating marginals
6: Run convex model ▷ results in Table II
7: Update operating point
8: **until** $| * |gap < 0.005\%$ ▷ gap to exact polar

In SubSection III-B all models were solved using IPOPT solver that has proven to be numerically highly robust and was set up to run with the HSL linear MA27 and scaling MC19 modules, option to always apply scaling and at most 500 solver iterations. In SubSection III-C the convex MIQCQP/MISOCP models were solved using Gurobi 9.0.2, while all the other models using Knitro 11.1. Both Gurobi and Knitro were run under the default settings on Intel i5 7600 CPU on 4 threads on a machine with 16 GB of RAM, while IPOPT was run on only one thread since it showed no performance scaling using additional threads. The solvers were instructed to compute to full optimality, unless the time limit of 1 h is reached. GAMS 31.1.1 was used as a modelling language and all continuous variables were declared as free variables. In the SubSection III-B, variable start values used to generate the model's Jacobian and Hessian matrices for the nonlinear solvers were reset to flat start assumptions after every computation. The specifics on bounds and variable starting values are important since they affect nonlinear solvers. All variable and parameter units are in p.u. or dimensionless.

B. Convergence and Accuracy Demonstration

This section demonstrates the accuracy of the model using OPF under assumption that the constraint type selection is known, i.e. determined by steps 4 and 5 in Algorithm 1. In an iterative procedure, the model is first run in its nonconvex form in which constraints (2.8.1) and (2.9.1) are equalities. Signs of their marginal values determine which constraints would deviate from the inequality boundary if they were relaxed. Subsequently, the model is rerun but in its convex form with selected quadratic or linear constraints to avoid constraint relaxation errors due to convexifications. The convex model updates the operating point voltage and angle parameters initially set to 1 p.u. and 0 rd. The procedure continues until fully converged in comparison to the exact nonlinear polar model [24], as described in Algorithm 1. The purpose of this case study is purely to explore the model's behavior on a large number of cases and not to solve an OPF as there are better existing options for problems with no integer variables, e.g. using the exact nonlinear models.

A well-established power grid model benchmark PGLib-OPF v19.05 [29] is used to perform a single time step optimization. Thus, the unit commitment and ramping constraints (2.16)–(2.20) as well as the unit commitment binary variables are

TABLE II
MODEL CONVERGENCE AND ACCURACY DEMONSTRATION BASED ON PGLIB-OPF TYPICAL CASES BENCHMARK [29]

				Exact Polar NLP	Iteration 1 (flat start)						Iteration 2						Iteration 3						Iteration 4					
Name pglib_opf_case	$\|E\|$	$\|N^{\mathrm{P}}\|$	$\|E\|$: $\boldsymbol{g}_e \leqslant 0$	Objective	Time [s]	Gap [%]	Time [s]	#(2.8.2) : $\boldsymbol{g}_e > 0$	#(2.9.2)	#(2.8.1) errors	#(2.9.1) errors	Gap [%]	Time [s]	#(2.8.2) : $\boldsymbol{g}_e > 0$	#(2.9.2)	#(2.8.1) errors	#(2.9.1) errors	Gap [%]	Time [s]	#(2.8.2) : $\boldsymbol{g}_e > 0$	#(2.9.2)	#(2.8.1) errors	#(2.9.1) errors	Gap [%]	Time [s]	#(2.8.2) : $\boldsymbol{g}_e > 0$	#(2.9.2)	#(2.8.1) errors	#(2.9.1) errors
3_lmbd	3	3	0	5.8126E+3	0	3.77	0	0	1			0.35	0	0	1			0.00	0	0	1			0.00	0	0	1		
5_pjm	6	6	0	1.7552E+4	0	0.56	0	0	0			-0.09	0	0	0			0.00	0	0	0			0.00	0	0	0		
14_ieee	20	20	5	2.1781E+3	0	0.36	0	0	3			0.03	0	0	0			0.00	0	0	0			0.00	0	0	0		
24_ieee_rts	38	34	0	6.3352E+4	0	0.24	0	0	0			0.00	0	0	0			0.00	0	0	0			0.00	0	0	0		
30_as	41	41	7	8.0313E+2	0	0.20	0	0	3			0.00	0	0	1			0.00	0	0	1			0.00	0	0	1		
30_fsr	41	41	7	5.7577E+2	0	-0.01	0	0	2			0.00	0	0	2			0.00	0	0	2			0.00	0	0	2		
30_ieee	41	41	7	8.2085E+3	0	0.23	0	0	1			0.04	0	0	0			0.00	0	0	0			0.00	0	0	0		
39_epri	46	46	4	1.3842E+5	0	0.11	0	0	2			0.01	0	0	2			0.00	0	0	1			0.00	0	0	1		
57_ieee	80	78	18	3.7589E+4	0	0.01	0	0	0			0.00	0	0	1			0.00	0	0	1			0.00	0	0	1		
73_ieee_rts	120	108	1	1.8976E+5	0	0.25	0	0	0			0.01	0	0	1			0.00	0	0	1			0.00	0	0	1		
89_pegase	210	206	10	1.0729E+5	0	it/inf	48	4	64	167	85	0.12	0	2	7			0.00	0	2	7			0.00	0	2	7		
118_ieee	186	179	9	9.7214E+4	0	0.39	0	0	6			0.02	0	0	2			0.00	0	0	2			0.00	0	0	2		
162_ieee_dtc	284	280	33	1.0808E+5	0	it/inf	2	22	157	5	6	1.08	0	0	88		3	0.04	0	0	90			0.00	0	0	88		
179_goc	263	222	72	7.5427E+5	0	0.03	0	0	69		6	0.00	0	0	60			0.00	0	0	60			0.00	0	0	60		
200_tamu	245	245	0	2.7558E+4	0	0.04	0	0	0			0.02	0	0	0			0.00	0	0	0			0.00	0	0	0		
240_pserc	448	348	93	3.3297E+6	1	2.45	1	0	61			0.06	1	0	63		1	0.00	1	0	59			0.00	1	0	59		
300_ieee	411	409	64	5.6522E+5	1	-1.20	1	0	6			0.00	1	0	5			0.00	1	0	5			0.00	1	0	5		
500_tamu	597	584	0	7.2578E+4	1	0.46	1	0	0			0.01	1	0	2			0.00	1	0	2			0.00	1	0	2		
588_sdet	686	677	54	3.1314E+5	1	0.14	1	0	24			0.01	1	0	32			0.00	1	0	31			0.00	1	0	31		
1354_pegase	1991	1710	1	1.2588E+6	2	0.44	5	0	2			0.01	5	0	7			0.00	9	0	7			0.00	9	0	8		
1888_rte	2531	2308	248	1.4025E+6	10	it/inf	39	174	854	4	5	2.47	16	6	637	2	18	0.13	13	6	601		2	0.00	16	7	601		
1951_rte	2596	2375	339	2.0856E+6	6	8.24	9	2	311	2	10	0.01	23	0	243			0.00	7	0	243			0.00	7	0	243		
2000_tamu	3206	2667	0	1.2285E+6	4	0.26	4	0	0			0.08	4	0	21			0.00	4	0	31			0.00	4	0	31		
2316_sdet	3017	2834	58	1.7753E+6	2	0.15	9	0	43		1	0.02	9	0	63			0.00	13	0	64			0.00	17	0	64		
2383wp_k	2896	2886	195	1.8682E+6	5	1.13	6	0	58		2	0.02	8	0	76			0.00	10	0	76			0.00	7	0	76		
2736sp_k	3269	3263	10	1.3080E+6	4	0.37	9	0	1			0.01	9	0	8			0.00	6	0	9			0.00	7	0	9		
2737sop_k	3269	3263	10	7.7773E+5	3	0.24	6	0	4			0.02	8	0	8			0.00	7	0	11			0.00	6	0	11		
2746wop_k	3307	3299	10	1.2083E+6	3	0.34	7	0	2			0.01	10	0	13			0.00	12	0	13			0.00	10	0	13		
2746wp_k	3279	3273	10	1.6317E+6	4	0.46	8	0	0			0.00	8	0	3			0.00	8	0	3			0.00	8	0	3		
2848_rte	3776	3442	209	1.2866E+6	21	0.34	17	0	392		42	0.01	27	1	237		1	0.00	26	1	235			0.00	23	1	235		
2853_sdet	3921	3635	386	2.0524E+6	5	0.19	9	0	218		1	0.02	10	0	193			0.00	11	0	202			0.00	13	0	199		
2868_rte	3808	3471	257	2.0096E+6	19	0.35	17	0	228		1	0.01	47	0	247			0.00	14	0	248			0.00	45	0	248		
2869_pegase	4582	3968	136	2.4628E+6	5	0.38	35	0	18			0.03	33	0	46			0.00	31	0	48			0.00	31	0	47		
3012wp_k	3572	3566	20	2.6008E+6	4	it/inf	43	519	309	1	3	-0.01	9	1	17	1		0.00	9	2	16			0.00	9	2	16		
3120sp_k	3693	3684	21	2.1480E+6	5	it/inf	48	1263	1475			0.05	11	1	33			0.00	10	0	34			0.00	11	0	34		
3375wp_k	4161	4068	34	7.4382E+6	8	it/inf	52	882	624	43	20	0.03	18	2	29		1	0.00	14	2	23			0.00	37	3	24		
4661_sdet	5997	5751	115	2.2513E+6	8	0.17	37	3	89			0.01	28	1	145			0.00	18	3	142			0.00	17	3	142		
6468_rte	9000	8065	188	2.0697E+6	43	it/inf	150	540	1384	3	2	it/inf	171	1726	3209	11	3	it/inf	151	2330	4596	2	6	it/inf	158	4080	4760	13	126
6470_rte	9005	8066	217	2.2376E+6	41	it/inf	163	276	1064	3	4	it/inf	121	1873	2023		156	0.03	132	2	318			0.00	153	2	318		
6495_rte	9019	8084	253	3.0678E+6	49	it/inf	166	2902	4281	1	23	it/inf	169	1073	2592	47	80	it/inf	186	1158	2908		1	it/inf	157	2794	4462	420	503
6515_rte	9037	8104	263	2.8255E+6	45	it/inf	136	1711	2350	19	25	-1.04	138	11	310		6	0.02	131	24	434		1	0.00	137	25	431		
9241_pegase	16049	14207	921	6.2431E+6	49	0.43	103	0	241		1	0.03	155	0	384		1	0.00	175	0	383			0.00	185	0	384		
10000_tamu	12706	12217	179	2.4859E+6	63	-0.07	62	2	1322		48	0.05	77	0	199		1	0.00	60	0	216			0.00	61	0	216		
13659_pegase	20467	18625	1503	8.9480E+6	62	0.27	111	0	666		22	0.03	152	1	689		7	0.00	151	1	705			0.00	152	1	705		

removed. The resulting convex and nonconvex models minimize (2.1) subject to (2.2)–(2.15) over variables $\Xi = \{P^{\mathrm{g}}_{t,k}, Q^{\mathrm{g}}_{t,k}, P_{t,e,i,j}, Q_{t,e,i,j}, V^{\Delta}_{t,i}, \theta^{\Delta}_{t,i}, \widehat{cos}_{t,i,j}, \breve{V}_{t,e}\}$. The nonconvex model has only constraints (2.8.1) and (2.9.1) in the equality form, while the convex one has both linear and relaxed quadratic forms of constraints determined by the nonconvex presolve.

Results of the described iterative procedure for the convex step 6 in Algorithm 1 are shown in Table II. The iteration gap is defined as a percentage deviation from the exact nonlinear polar solution. Other metrics used are the computation time, the number of linear voltage constraints (2.8.2) with positive $\boldsymbol{g}_e$, the number of linear cosine constraints (2.9.2) and the number of quadratic constraints (2.8.1) and (2.9.1) that deviated from the inequality boundary. Total number of voltage linear constraints is equal to the number of branches with nonpositive $\boldsymbol{g}_e$ plus those decided to be replaced by the iteration's presolve. For a better overview of the grids' sizes, there are also columns with number of branches and bus-pairs, while the grid names contain the number of buses.

The model converges fully on all grids within three iterations from the first feasible iteration step. Full convergence results in 0.00% approximation error since the Taylor expansion is exact at the expansion point and there are no relaxation errors related to constraints (2.8.1) and (2.9.1), as displayed in Table II. This is because the presolve removes those relaxed quadratic constraints that would deviate from the inequality boundary with their linear equality alternatives. However, the flat start operating point assumption at the first iteration is not good enough for the model to provide a feasible solution in all cases. The solver usually does not recognise infeasible models, but reaches iteration limit marked by "it/inf" in the gap column. Despite being infeasible, a solution is returned by the solver and used to update the operating point for the next iteration. The only two grids that did not converge within four iterations are 6468_rte and 6495_rte. Very similar grids, 6470_rte and 6515_rte, did converge because they got updated with more favorable operating points quicker than the former two grids. It took seven iterations for 6468_rte and 6495_rte grids to fully converge as their first feasible solution was reached in the fifth iteration. Fast convergence indicates that very high accuracy can be achieved if the model is warm-started with a reasonably good operating point and constraints to be linearized selected by the presolve, which is pivotal for SubSection III-C.

The model's working principle can be well described on the 3_lmbd grid, which is small but congested. The second-order voltage approximation variable seldom deviates from the inequality boundary in (2.8.1) since such deviation would only increase active power losses at both branch ends. Cosine variable deviation from its bound in quadratic constraint (2.9.1) causes much higher reactive power losses than the active power losses due to branch susceptance being much higher than conductance. In 3_lmbd grid the ratio $\boldsymbol{b}_e/\boldsymbol{g}_e$ is -30 for the branch with (2.9.2) applied. That branch in the optimal solution is congested and also producing reactive power. Thus, reactive power losses are favorable as they enable greater transfer of active power. To prevent false losses by the cosine approximation constraint relaxation, the presolve determined that the linear equality constraint should be applied. However, congestion is not a necessary condition for potential cosine

Algorithm 2: Unit Commitment (UC).

1: Run exact polar model using relaxed binary variables to determine operating point (NLP) ▷ comp. time t^{A}
2: Run nonconvex quadratic presolve using fixed binary variables and operating point from the previous solve to select constrains for the main UC computation by evaluating constraint marginals (nonconvex QCQP) ▷ comp. time t^{B}
3: Run the UC (MIQCQP) ▷ comp. time t^{C}
4: Run exact polar model with fixed binary variables to UC solution to determine approximation error (NLP) ▷ comp. time t^{D}

variable relaxation errors. Some branches simply operate at such state that a deviation is favorable even without congestion. That occurs more commonly with high negative $\boldsymbol{b}_e/\boldsymbol{g}_e$ ratios.

Iterations that were infeasible or reached the iteration limit have an unrealistic number of linear constraints due to equivocal information from the proposed nonconvex presolve that is also infeasible. It is important that the network data contains entries for branch conductances as otherwise the model would contain only linear voltage constraints (2.8.2), making it insufficiently quadratically constrained for convergence and warm-start accuracy. This case study demonstrates, however, that even with partial conductance data, the model operates well as grid 179_goc has 27% of all branches with null conductances.

C. Unit Commitment

We use a 24-hour network-constrained unit commitment to evaluate accuracy and computational performance of the model. Due to availability of a limited amount of unit commitment network data, grids from the OPF benchmark [29] are adapted for unit commitment purposes using generic ramp limits, start-up costs, minimum up- and down-times, load curve and generator costs. Ramp limits are set so the mean of the allowed production range can be reached within a single time period, i.e. $(|\overline{\boldsymbol{P}}^{\mathbf{g}}_k| + |\underline{\boldsymbol{P}}^{\mathbf{g}}_k|)/2$, start-up costs are set to 1500 cost units for all generators, minimum up- and down-times are set to 2 hours for generators with $\leq$1 p.u. (100 MW) and 4 hours for larger ones, generator fixed costs $\boldsymbol{c}_k$ are modified to 10% of the linear costs $\dot{\boldsymbol{c}}_k$ if their original value is zero. Loads use scaling factors for the winter weekday periods from IEEE RTS-96 [30].

The proposed quadratic approximation model is designed to operate well in the vicinity of the operating point parameters $\boldsymbol{V}^{\mathbf{op}}_{t,i}$ and $\boldsymbol{\theta}^{\mathbf{op}}_{t,i}$. Thus, the first step is to provide a good operating point by solving the continuous exact nonlinear model with relaxed binary variables. The next step is to run the nonconvex continuous quadratic presolve, i.e. with (2.8.1) and (2.9.1) as equalities, using fixed (replaced with a parameter) binary variables and around the operating point computed in the previous solve. The fixed binary variables inherit relaxed, i.e. continuous, values. Its solution is the same as in the first step, but it returns the constraints' marginal values whose sign determines the constraint type (linear or quadratic) for every branch and bus pair of the main unit commitment solve. The third step is to run the mixed-integer unit commitment. The last, fourth step is to determine the approximation error by running the exact continuous polar formulation [24], but with binary variables fixed to the solution of the mixed-integer unit commitment run from the previous step. Inexact models may return binary variable values for which no solution is possible. In that case the approximation error is marked with "inf" in Table III. Compared to the existing LPAC, the proposed model has an additional, but easy to solve, second step to select constraints to be linearized. The described procedure is itemized in Algorithm 2. Optimization source code is provided on GitHub [31].

The simulation results are provided in Table III. Convex nature of the presented model allows for the use of traditional solvers typically used to solve MILP problems as they handle binary variables more efficiently than MINLP solvers. Thus, the presented model outperforms the exact polar MINLP in all but one test case, with computation times up to 77 times faster for 39_epri grid. Test cases 24_ieee_rts and 73_ieee_rts are difficult due to a large number of generators resulting in a large number of binary variables, which severely impacts computational tractability. It was found that in these two test cases the other solvers outperformed Gurobi. For the proposed MIQCQP model applied to 24_ieee_rts, CPLEX 12.10 reached 0.09% MIP gap as compared to 1.26% achieved by Gurobi. The best performance on 73_ieee_rts was achieved with Xpress 8.8, which reached 2.78% MIP gap, while Gurobi did not find a feasible solution. Furthermore, this MIP gap is much lower than the 10.29% gap obtained using Knitro 11.1 with the exact polar MINLP model. Both grids at the solution point had 0.00% objective function approximation errors. Test cases 162_ieee_dtc and 179_goc were infeasible even for the QC relaxation and were eliminated from the study, as well as networks larger than 200_tamu as they are too difficult to solve.

The most significant result of the study is very low, almost nonexistent, approximation error. For 13 test cases, the highest objective function error is -0.02%. The presented model is very accurate around a broad vicinity of the operating point, which can be sufficiently well estimated by solving the model with relaxed binary variables. Furthermore, at the operating point the quadratic constraints that can cause large errors by deviating from the inequality bounds are replaced with the linear ones. Quadratic constraints that finally do deviate from the inequality bound in the unit commitment solve cause a very small relaxation error. If this error was larger, quadratic constraints would deviate at the operating point and thus be replaced with linear ones. Furthermore, we evaluate the approximation errors of the individual variables using average and maximum normalized distance of solution variables [32]. Normalization is carried out by dividing the absolute variable error value by the variable feasible range. For $P_{t,e,i,j}$ and $Q_{t,e,i,j}$, the feasible range is $2\overline{\boldsymbol{S}}_e$, for $V_{t,i}$ it is $\overline{\boldsymbol{V}}_i - \underline{\boldsymbol{V}}_i$ and for the branch angle difference $\theta_{t,i,j}$ it is $\overline{\boldsymbol{\theta}}_{i,j} - \underline{\boldsymbol{\theta}}_{i,j}$. The results are displayed in Table IV. The highest average active power flow variable distance is 0.06%, reactive power flow distance 0.14%, voltage distance 1.42% and branch

TABLE III
UNIT COMMITMENT BENCHMARK RESULTS

					Exact Polar MINLP			Proposed MIQCQP approximation								LPAC approximation warm-start, MIQCQP variant				QC relaxation MISOCP			
Name pglib_opf_case	$\lvert E\rvert$	$\lvert N^{\mathrm{P}}\rvert$	$\lvert E\rvert$: $\boldsymbol{g}_e \leqslant 0$	$\lvert G\rvert$: $\underline{\boldsymbol{P}}_k^{\mathrm{g}} \neq \overline{\boldsymbol{P}}_k^{\mathrm{g}}$	MIP gap [%]	Objective	Time [s] t^{C}	MIP gap [%]	Objective	Approx. error [%]	Time [s] $t^{\mathrm{A}}+t^{\mathrm{B}}+t^{\mathrm{C}}+t^{\mathrm{D}}$	#(2.8.2) : $\boldsymbol{g}_e > 0$	#(2.9.2)	#(2.8.1) errors	#(2.9.1) errors	MIP gap [%]	Objective	Approx. error [%]	Time [s] $t^{\mathrm{A}}+t^{\mathrm{C}}+t^{\mathrm{D}}$	MIP gap [%]	Objective	Relax. error [%]	Time [s] $t^{\mathrm{C}}+t^{\mathrm{D}}$
3_lmbd	3	3	0	2	0	1.0224E+5	4	0	1.0224E+5	0.00	0+0+1+0	0	14			0	1.0150E+5	-0.72	0+1+0	0	1.0167E+5	-0.56	1+0
5_pjm	6	6	0	5	0	3.1233E+5	12	0	3.1233E+5	0.00	0+0+2+0	0	0			0	3.1240E+5	0.02	0+2+0	0	2.7507E+5	inf	3+1
14_ieee	20	20	5	2	0	4.7154E+4	6	0	4.7156E+4	0.00	0+0+5+0	0	4			0	4.7126E+4	-0.06	0+4+0	0	4.7117E+4	-0.08	7+0
24_ieee_rts	38	34	0	32	6.35	9.1310E+5	3600	1.26	8.7739E+5	0.00	1+0+3600+0	0	8			0.95	8.7385E+5	inf	1+3600+6	2.20	8.8185E+5	-0.02	3600+1
30_as	41	41	7	6	0	1.7281E+4	3559	0	1.7280E+4	0.00	1+0+54+0	0	13	3		0	1.7213E+4	-0.39	1+34+0	0	1.7230E+4	inf	500+6
30_fsr	41	41	7	6	0	1.3553E+4	839	0	1.3552E+4	-0.01	1+0+87+0	0	4	13		0	1.3531E+4	-0.16	1+35+0	0	1.3107E+4	inf	403+1
30_ieee	41	41	7	2	0	1.5810E+5	32	0	1.5810E+5	0.00	1+0+9+0	0	18			0	1.5771E+5	-0.25	1+7+0	0	1.3166E+5	inf	73+2
39_epri	46	46	4	10	0	2.5967E+6	1547	0	2.5967E+6	0.00	1+0+20+1	0	57	16		0	2.5947E+6	-0.08	1+11+1	0	2.5879E+6	inf	65+40
57_ieee	80	78	18	4	0	7.4452E+5	160	0	7.4450E+5	0.00	4+1+26+0	0	37	8		0	7.4309E+5	-0.19	4+21+0	0	7.4323E+5	-0.17	90+0
73_ieee_rts	120	108	1	96	10.29	2.8347E+6	3600	-	-	-	6+1+3600+ -	0	35	-	-	-	-	-	6+3600+ -	-	-	-	3600+ -
89_pegase	210	206	10	12	4.41	1.7638E+6	3600	0	1.6957E+6	-0.02	3+2+479+3	4	77	5		0	1.6826E+6	inf	4+91+21	0.06	1.6766E+6	inf	3600+11
118_ieee	186	179	9	19	0.83	1.9406E+6	3600	0	1.9310E+6	0.01	3+3+2300+2	0	106			0	1.9250E+6	-0.33	3+1318+2	0	1.9026E+6	inf	2252+122
200_tamu	245	245	0	32	25.15	2.9787E+5	3600	0	2.2587E+5	0.00	5+2+2008+3	0	10			0	2.2530E+5	inf	5+1699+19	-	num. dif.	-	395+ -

TABLE IV
NORMALIZED DISTANCE OF THE SOLUTION VARIABLES (IN [%])

Name	$P_{t,e,i,j}$		$Q_{t,e,i,j}$		$V_{t,i}$		$\theta_{t,i,j}$	
pglib_opf_case	avg	max	avg	max	avg	max	avg	max
3_lmbd	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00
5_pjm	0.00	0.06	0.02	0.06	0.02	0.11	0.00	0.01
14_ieee	0.00	0.01	0.00	0.11	0.03	0.56	0.00	0.13
24_ieee_rts	0.00	0.11	0.03	1.27	0.17	3.43	0.01	0.16
30_as	0.02	0.43	0.12	3.07	0.54	3.62	0.01	0.07
30_fsr	0.02	0.56	0.14	10.44	0.28	5.92	0.01	0.38
30_ieee	0.00	0.00	0.00	0.01	0.00	0.04	0.00	0.00
39_epri	0.00	0.11	0.04	3.89	0.24	9.99	0.00	0.29
57_ieee	0.01	0.15	0.02	0.80	0.14	3.65	0.00	0.11
73_ieee_rts	-	-	-	-	-	-	-	-
89_pegase	0.03	1.95	0.13	12.03	1.42	15.16	0.02	0.81
118_ieee	0.06	2.20	0.08	3.64	0.41	15.47	0.02	0.58
200_tamu	0.00	0.00	0.00	0.15	0.02	0.37	0.00	0.15

angle distance 0.02%, which is significantly lower than what can be expected from the competing QC and even Shor's model that normally achieve overall average distances in the range from 5 to 10% (see Fig. 3 in [32]). The approximation errors can be even further reduced by iteratively running the unit commitment solve by updating the operating point and retesting the constraints for that new operating point. However, as an alternative, due to a low objective function approximation error, we recommend using the verified solution, i.e. Algorithm's 2 step 4, as the final solution to the unit commitment problem.

Achieving great accuracy in unit commitment is important for two reasons: a) active power losses are only about 2% of the total production, thus the accuracy needs to be far greater than 2% to account for them properly; b) inaccurate solutions can result in infeasible decisions for generator activity status, i.e. binary variables. The LPAC model in its more accurate quadratic variant, despite being warm-started similarly like the proposed model, has the highest approximation errors -0.72%, -0.39% and -0.33% and three out of 13 cases are infeasible. QC relaxation, despite proven to be tighter than the Jabr's relaxation, is still overly optimistic and thus results in 7 infeasible cases, while the greatest error of feasible models is -0.56%. These results highlight the importance of quadratically constraining the voltage in combination with the constraint type selection procedure. Computation-time-wise, the proposed MIQCQP model is in between the faster LPAC approximation and the slower QC approximation.

IV. CONCLUSION

The presented case studies demonstrate that the proposed MIQCQP model is moderately accurate assuming flat start operating point parameters, but almost perfectly accurate if the assumed operating point is in a broader vicinity of the optimal solution. Accuracy is achieved by having both the voltage magnitude and the angle quadratically constrained and by replacing some of the quadratic inequality constraints with linear equality constraints to avoid constraint relaxation errors due to convexification as determined by the presolve. Meshed transmission network structure acts favourably to further constrain the model so it can withstand the replacement of some quadratic constraints with linear ones and stay accurate and convergent.

Since the model's power-flow-related constraints are convex quadratic, the model supports solvers that handle binary variables more efficiently than the MINLP solvers. The proposed model displays average computational tractability for the MIQCQP/MISOCP problem class, which means it is slightly slower than the quadratic implementation of the LPAC approximation [17], [18] and quicker than the QC relaxation [20]. Accuracy of the existing quadratic models is insufficient for network-constrained unit commitment leading to frequent infeasible decisions for binary variables or significant approximation or relaxation errors.

ACKNOWLEDGMENT

The sole responsibility for the content of this document lies with the authors. It does not necessarily reflect the opinion of the Innovation and Networks Executive Agency (INEA) or the European Commission (EC). INEA or the EC are not responsible for any use that may be made of the information contained therein.

REFERENCES

[1] C. Coffrin and B. Knueven, "PGLib-UC," *GitHub*, Aug. 7, 2019. Accessed: Dec. 30, 2020. [Online]. Available: github.com/power-grid-lib/pglib-uc

[2] Y. Ham, C. Chun and S. Lee, "Some higher-order modifications of newton's method for solving nonlinear equations," *J. Comput. Appl. Math.*, vol. 222, no. 2, pp. 477–486, Dec. 2008.

[3] A. Castillo, C. Laird, C. A. Silva-Monroy, J. Watson and R. P. O'Neill, "The unit commitment problem with AC optimal power flow constraints," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4853–4866, Nov. 2016.

[4] T. S. Dillon, K. W. Edwin, H. D. Kochs and R. J. Taud, "Integer programming approach to the problem of optimal unit commitment with probabilistic reserve determination," *IEEE Trans. Power App. Syst.*, vol. PAS-97, no. 6, pp. 2154–2166, Nov. 1978.

[5] J. M. Arroyo and A. J. Conejo, "Optimal response of a thermal unit to an electricity spot market," *IEEE Trans. Power Syst.*, vol. 15, no. 3, pp. 1098–1104, Aug. 2000.
[6] D. Rajan and S. Takriti, "Minimum up/down polytopes of the unit commitment problem with start-up costs," IBM, New York, NY, USA, Tech. Rep. RC23628, pp. 1–18, Jun. 2005, Accessed: Dec. 30, 2020.[Online]. Available: preview.tinyurl.com/yc5c3ey9
[7] C. K. Simoglou, P. N. Biskas, and A. G. Bakirtzis, "Optimal self-scheduling of a thermal producer in short-term electricity markets by MILP," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1965–1977, Nov. 2010.
[8] M. Carrion and J. M. Arroyo, "A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem," *IEEE Trans. Power Syst*. vol. 21, no. 3, pp. 1371–1378, Aug. 2006.
[9] J. Meus, K. Poncelet and E. Delarue, "Applicability of a clustered unit commitment model in power system modeling," *IEEE Trans. Power Syst*. vol. 33, no. 2, pp. 2195–2204, Mar. 2018.
[10] K. Jurković, H. Pandžić, and I. Kuzle, "Review on unit commitment under uncertainty approaches," in *Proc. 38th Int. Conv. MIPRO*, Opatija, Croatia, 2015, pp. 1093–1097.
[11] Y. Wu, Y. Li, W. Hsu, and B. Lan, "Review of security-constrained unit commitment in a large power system," in *Proc. IEEE Int. Conf. ICASI*, Chiba, Japan, 2018, pp. 1310–1313.
[12] D. K. Molzahn and I. A. Hiskens, "A survey of relaxations and approximations of the power flow equations," *Found. Trends Electr. Energy Syst.*, vol. 4, no. 1-2, pp. 1–221, Feb. 2019.
[13] B. Stott, J. Jardim, and O. Alsac, "DC power flow revisited," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1290–1300, Aug. 2009.
[14] H. Zhang, G. T. Heydt, V. Vittal and J. Quintero, "An improved network model for transmission expansion planning considering reactive power and network losses," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3471–3479, Aug. 2013.
[15] Z. Yang, H. Zhong, A. Bose, T. Zheng, Q. Xia and C. Kang, "A linearized OPF model with reactive power and voltage magnitude: A. pathway to improve the MW-only DC OPF," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1734–1745, Mar. 2018.
[16] C. Coffrin, H. Hijazi, and P. Van Hentenryck, "Alternating current (AC) power flow analysis in an electrical power network," WO Patent WO2013173879A1, Nov. 28, 2013.
[17] C. Coffrin and P. Van Hentenryck, "A linear-programming approximation of AC power flows," *INFORMS J. Comp.*, vol. 26, no. 4, pp. 718–734, 2014.
[18] C. Coffrin, R. Bent, K. Sundar, Y. Ng and M. Lubin, "PowerModels.jl: An open-source framework for exploring power flow formulations," in *Proc. Power Syst. Comput. Conf.*, pp. 1–8, Jun. 2018.
[19] R. A. Jabr, "Radial distribution load flow using conic programming," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1458–1459, Aug. 2006.
[20] H. Hijazi, C. Coffrin and P. Van Hentenryck, "Convex quadratic relaxations of mixed-integer nonlinear programs in power systems," *Math. Prog. Comp.*, Sep. 2016.
[21] X. Bai, H. Wei, K. Fujisawa and Y. Wang, "Semidefinite programming for optimal power flow problems," *Int. J. Elec. Power Energy Syst.*, vol. 30, no. 6, pp. 383–392, Jul. 2008.
[22] C. Coffrin, H. L. Hijazi, and P. Van Hentenryck, "The QC relaxation: A. theoretical and computational study on optimal power flow," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3008–3018, Jul. 2016.
[23] R. P. O'Neill, A. Castillo, and M. B. Cain, "The IV formulation and linear approximations of the ac optimal power flow problem," FERC, Washington, DC, USA, Tech. Rep., Dec. 2012. Accessed: Dec. 30, 2020. [Online]. Available: http://cms.ferc.gov/sites/default/files/2020-05/acopf-2-iv-linearization.pdf
[24] S. Frank and S. Rebennack, "An introduction to optimal power flow: Theory, formulation, and examples," *IIE Trans.*, vol. 48, no. 12, pp. 1172–1197, 2016.
[25] G. L. Torres and V. H. Quintana, "An interior-point method for nonlinear optimal power flow using voltage rectangular coordinates," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1211–1218, Nov. 1998.
[26] D. I. Sun, B. Ashley, B. Brewer, A. Hughes and W. F. Tinney, "Optimal power flow by Newton approach," *IEEE Trans. Power App. Syst.*, vol. PAS-103, no. 10, pp. 2864–2880, Oct. 1984.
[27] H. Pandžić, Ting Qiu, and D. S. Kirschen, "Comparison of state-of-the-art transmission constrained unit commitment formulations," in *Proc. IEEE PES Gen. Meeting*, Vancouver, BC, 2013, pp. 1–5.
[28] J. Currie, "Mixed integer quadratically constrained quadratic program (MIQCQP)," OPTI Toolbox. Dec. 21, 2016. Accessed: Dec. 30, 2020. [Online]. Available: http://inverseproblem.co.nz/OPTI/index.php/Probs/MIQCQP
[29] C. Coffrin, PGLib-OPF v19.05, GitHub, May 13, 2019. Accessed: Dec. 30, 2020. [Online]. Available: Github.com/power-Grid-Lib/pglib-Opf/tree/v19.05
[30] C. Grigg *et al.*, "The IEEE reliability test System-1996," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1010–1020, Aug. 1999.
[31] K. Šepetanc and H. Pandžić, *CPSOTA-UC*, GitHub, Dec. 21, 2020. Accessed: Dec. 30, 2020. [Online]. Available: github.com/KSepetanc/CPSOTA-UC
[32] A. Venzke, S. Chatzivasileiadis and D. K. Molzahn, "Inexact convex relaxations for AC optimal power flow: Towards AC feasibility," *Electr. Power Syst. Res.*, vol. 187, Oct. 2020.

Karlo Šepetanc (Student Member, IEEE) received the M.E.E. degree in 2018 from the Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, where he is currently working toward the Ph.D. degree. He is also a Researcher with Innovation Centre Nikola Tesla. His research interests include planing, operation, and economics of power and energy systems.

Hrvoje Pandžić (Senior Member, IEEE) received the M.E.E. and Ph.D. degrees from the Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, in 2007 and 2011, respectively. From 2012 to 2014, he was a Postdoctoral Researcher with the University of Washington, Seattle, WA, USA.

He is currently an Associate Professor and the Head of the Department of Energy and Power Systems, Faculty of Electrical Engineering and Computing, University of Zagreb. He is also a Leading Researcher with Innovation Centre Nikola Tesla. His research interests include planning, operation, control, and economics of power and energy systems.

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207088

Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part I: Model Description and the Algorithm

K. Šepetanc, *Student Member*, *IEEE*, H. Pandžić, *Senior Member*, *IEEE* and T. Capuder, *Member*, *IEEE*

***Abstract*—The existing research on market price-affecting agents, i.e. price makers, neglects or simplifies the nature of AC power flows in the power system as it predominantly relies on DC power flows. This paper proposes a novel bilevel formulation based on the smoothing technique, where any price-affecting strategic player can be modelled in the upper level, while the market clearing problem in the lower level uses convex quadratic transmission AC optimal power flow (AC OPF), with the goal of achieving accuracy close to the one of the exact nonlinear formulations. Achieving convexity in the lower level is the foundation for bilevel modeling since traditional single-level reduction techniques do not hold for nonconvex models. The bilevel market participation problem with the AC OPF formulation in the lower level is transformed into a single-level problem and solved using multiple techniques such as the primal-dual counterpart, the strong duality theorem, the McCormick envelopes, the complementary slackness, the penalty factor, the interaction discretization as well as the proposed smoothing techniques.**

Due to an extensive amount of information and descriptions, the overall work is presented as a two-part paper. This first part provides a literature overview, positions the work and presents the model and the solution algorithm, while the solution techniques and case studies are provided in the accompanying paper.

***Index Terms*—Bilevel models, AC OPF, complementary condition smoothing functions.**

NOMENCLATURE

A. Abbreviations

OPFOptimal power flow.
SOCSecond-order cone.
SOCPSecond-order cone programming.
QCQuadraticaly constrained.
SDPSemidefinite programming.
IVCurrent-voltage.
LPACLinear programming AC.
QPACQuadratic programming AC.
ESEnergy storage.
CPSOTAConvex polar second-order Taylor approximation.
KKTKarush–Kuhn–Tucker.
NLPNonlinear programming.
OPOperating point.
NCNonconvex.
LLLower level.

The authors are with the Innovation Centre Nikola Tesla (ICENT) and the University of Zagreb Faculty of Electrical Engineering and Computing (e-mails: karlo.sepetanc@fer.hr; hrvoje.pandzic@fer.hr; tomislav.capuder@fer.hr). Employment of Karlo Šepetanc is funded by the Croatian Science Foundation under programme DOK-2018-09. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864298 (project ATTEST). The sole responsibility for the content of this document lies with the authors. It does not necessarily reflect the opinion of the Innovation and Networks Executive Agency (INEA) or the European Commission (EC). INEA or the EC are not responsible for any use that may be made of the information contained therein.

B. Sets and Indices

N	Set of buses, indexed by i and j.
β	ES's bus location singleton, indexed by i.
R	Singleton set containing reference bus, indexed by i.
E, E^{R}	Tuple set of branches, forward and reverse orientation, indexed by (e, i, j).
$N^{\mathrm{P}}, N^{\mathrm{PR}}$	Tuple set of paired buses aligned with branch E and E^{R} orientations, indexed by (i, j).
G, G_i	Set of all generators and array of sets of generators at bus i, indexed by k.
L_i	Array of sets of loads at bus i, indexed by l.
S_i	Array of sets of shunts at bus i, indexed by s.
τ	Set of time steps, indexed by t and h.
$\Xi^{[\cdot]}$	Set of decision variables.

C. Parameters

$\ddot{\boldsymbol{c}}_k, \dot{\boldsymbol{c}}_k, \boldsymbol{c}_k$	Generator cost coefficients.
$\boldsymbol{P}^{\mathbf{d}}_{t,l}, \boldsymbol{Q}^{\mathbf{d}}_{t,l}$	Active and reactive power load.
$\boldsymbol{g}^{\mathbf{sh}}_s, \boldsymbol{b}^{\mathbf{sh}}_s$	Bus shunt conductance and susceptance.
$\boldsymbol{g}_e, \boldsymbol{g}^{\mathbf{fr}}_e, \boldsymbol{g}^{\mathbf{to}}_e$	Branch π-section conductances.
$\boldsymbol{b}_e, \boldsymbol{b}^{\mathbf{fr}}_e, \boldsymbol{b}^{\mathbf{to}}_e$	Branch π-section susceptances.
$\boldsymbol{\tau}_e, \boldsymbol{\sigma}_e$	Branch tap magnitude and shift angle.
$\underline{\boldsymbol{P}}^{\mathbf{g}}_k, \overline{\boldsymbol{P}}^{\mathbf{g}}_k$	Generator minimum and maximum active power production.
$\underline{\boldsymbol{Q}}^{\mathbf{g}}_k, \overline{\boldsymbol{Q}}^{\mathbf{g}}_k$	Generator minimum and maximum reactive power production.
$\overline{\boldsymbol{S}}_e$	Branch maximum apparent power.
$\underline{\boldsymbol{V}}_i, \overline{\boldsymbol{V}}_i$	Bus minimum and maximum voltage magnitude.
$\boldsymbol{V}^{\mathbf{op}}_{t,i}, \boldsymbol{\theta}^{\mathbf{op}}_{t,i}$	Assumed bus voltage magnitude and angle operating points.
$\boldsymbol{\Lambda}_{t,e}, \boldsymbol{\Gamma}_{t,i,j}$	Boolean parameters which indicate whether to use the quadratic form of the voltage and the cosine representations, respectively.
$\boldsymbol{\Phi}_{t,e,i,j}$	Boolean parameter indicating if the branch power limit is imposed.
$\overline{\boldsymbol{SoE}}$	Energy storage capacity.
$\overline{\boldsymbol{s}}^{\mathbf{es}}$	Energy storage maximum power.
$\boldsymbol{\eta}^{\mathbf{ch}}, \boldsymbol{\eta}^{\mathbf{dis}}$	Energy storage (dis)charging efficiency.

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207088

D. Variables

Continuous variables

$P^{\mathrm{g}}_{t,k}, Q^{\mathrm{g}}_{t,k}$	Generator active and reactive power production.
$P_{t,e,i,j}, Q_{t,e,i,j}$	Branch active and reactive power flow.
$V^{\Delta}_{t,i}, \theta^{\Delta}_{t,i}$	Bus voltage magnitude and angle change.
$\widehat{cos}_{t,i,j}$	Cosine approximation.
$\breve{V}_{t,e}$	Second-order Taylor series voltage magnitude term approximation.
SoE_t	Energy storage state-of-energy.
$p^{\mathrm{es}}_t, q^{\mathrm{es}}_t$	Energy storage active and reactive power.
$p^{\mathrm{ch}}_t, p^{\mathrm{dis}}_t$	Energy storage (dis)charging active power.

Binary variables

x^{p}_t	Disables simultaneous charging and discharging of energy storage.

I. Introduction

Increasing the integration of renewable energy resources, as well as the electrification trends in the context of the zero-carbon energy future, pushes the power system operation to its technical limits. Operational challenges and increased needs for system flexibility require advances towards new, close to real-time, market products. However, the existing market bidding models do not adequately model the relevant technical aspects, such as voltage magnitudes, node angles and loses, as they rely on simplifications or approximations in the optimal power flow models. Although market designs both in Europe and the US still employ the DC optimal power flow (DC OPF) to perform market clearing, there is a growing interest for using AC optimal power flow (AC OPF), primarily to gain a more accurate insight into the network state and enable a more complete market design that accounts for ancillary services, see e.g. [1] and [2].

Due to the interest of the power industry, there has been a growing interest for development of tractable and accurate AC OPF models. Unlike the linear DC OPF problem, where linearity of the power flow approximation results in good numerical computability even with many binary variables, the nonconvexity of the AC OPF problem makes it computationally challenging. The AC OPF models can be grouped into three categories: exact models, relaxations and approximations.

Exact OPF models are based on an exact power flow function, most commonly expressed either using rectangular coordinates [3], as a nonconvex quadratic optimization problem, or polar coordinates [4], as a nonlinear function. There is also the IV (current-voltage) rectangular formulation [5], typically used for modeling loads with irregular power-voltage curves, also known as ZIP loads. On the other hand, relaxation models provide an upper bound to the objective function value of the exact models. Thus, if an exact optimization achieves the same objective function value a relaxation model, then this solution is necessarily the global one. Otherwise, global optimality can not be guaranteed.

One of the first relaxation models, which was developed by Jabr [6], is based on the second-order cone relaxation (SOCP), which can be used to relax any nonconvex quadratic formulation. It achieves good results in radial distribution networks due to the absence of the closed loop angle consistency criterion. However, when applied to the transmission network, it results in objective function errors of app. 2% [7], which is comparable to the total transmission network losses. An extension of the Jabr's model, developed by Coffrin, and called quadratically constrained relaxation (QC) [8], utilizes the McCormick envelopes to tighten the feasibility area based on maximum bus voltage angle differences. The QC model is at least as tight relaxation as the SOCP relaxation [9]. Shor's semidefinite relaxation (SDP) [10] is numerically the most demanding out of the commonly used models. It is also at least as tight as the SOCP relaxation [9] and usually more accurate than the QC relaxation. The QC model is not considered in this paper since it normally results in similar accuracy as the Jabr's model, but at an increased computation time [9]. Shor's SDP model is also not considered, as there are no solvers that can solve a combination of nonlinear and SDP optimizations incurred by the solution techniques presented in the Part II paper.

In approximations, unlike relaxations, the objective function can deviate both positively and negatively from the global optimum. The commonly used DC model, a linear model, also belongs to this category. Linear models in general have difficulties with modeling active power losses since they are in quadratic dependence on the voltage magnitude difference. To prevent negative active power losses, the linear model from [11] uses penalty factors in the objective function. The model from [12] approaches the problem of quadratic losses by introducing nonconvex piecewise linear losses, but at an expense of binary and integer variables. Similar approaches are also published in [13] and [14]. On the other hand, the linear AC model (LPAC) from [15] approximates the quadratic function with a series of linear inequalities that form a convex space and, thus, do not require binary variables. Except the approximation errors, this model also exhibits relaxations errors due to inequality constraints that replace the intended equality bound. The convex quadratic model (QPAC) was published in 2013 as a still active patent [16]. If there are no errors due to deviation from the inequality bounds, this model approximates the power flows well. However, occurrence of these errors depends on the system state and network configuration.

From the reviewed models, only LPAC can potentially increase its accuracy by using a presolve to approximate the operating point data. Unfortunately, it does not fully use this potential and sometimes it is even more accurate when using the cold start assumptions. This is in detailed discussed in our previous work [17], demonstrating the benefits of warm start information. In this paper we extend that model by applying it to applicability to the bilevel problems with AC OPF in the lower level. Convex polar second-order Taylor approximation of AC power flows (CPSOTA) from [17] achieves high accuracy by incorporating both the voltage magnitude and the angle second-order Taylor components and by replacing some of the quadratic inequality constraints with linear equality constraints to avoid constraint relaxation errors due to convexification. The replacement is determined by the presolve also developed in [17]. CPSOTA acts as a local AC OPF approximation at a reference operating point which can be well estimated in a

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207088

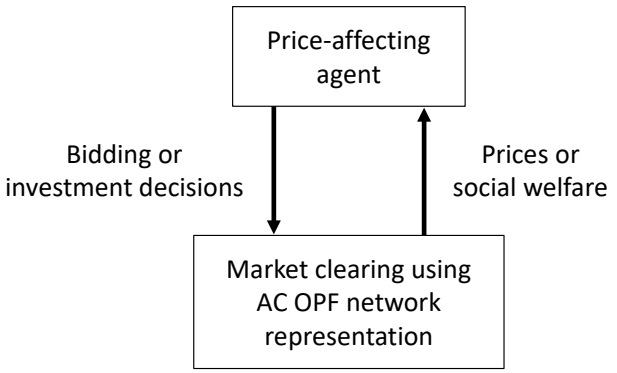

Fig. 1: Visualization of the upper-level and lower-level interaction in bilevel problems with market clearing based on AC OPF in the lower level.

bilevel optimization by simply assuming no market influence of the strategic market player. Good numerical tractability of the proposed method is, however, a result of the smoothing techniques presented in Part II, Section II.G.

Convex optimal power flow formulations are a foundation for strategic bidding models relying on bilevel optimization. Generally, in such models the upper-level problem determines the optimal bidding strategy of a strategic player, i.e. the player whose bidding decisions can affect market prices. This player's bidding prices or quantities affect the power flows and market prices. The lower level is used to simulate the market outcome pertaining to the strategic player's bidding actions, as visualized in Figure 1. Examples of such bilevel models are bidding of large consumers [18], generators [19], energy storage [20] and aggregators [21]. The considered markets may range from the day-ahead market, such as in [18]– [21], to a number of markets, such as in [22], where the strategic agent takes part in the day-ahead as well as in the reserve/balancing market. Bilevel models can also be used to determine an optimal investment by considering a number of representative market clearing days within a year. As an example, in [23] the authors seek an optimal generation investment considering the expected market-clearing conditions.

Strategic agents are not the only ones that use bilevel problems with market clearing in the lower level. The system operators could be interested in increasing the social welfare by investing in new transmission lines [24]. As an example, in [25] and [26] the authors consider transmission expansion planning where the SOCP relaxation of the AC OPF is formulated in the lower level. Both papers compare the results only to the option where the lower-level model is replaced with the DC OPF approximation and do not discuss the aspects of achieving accuracy of the exact formulations. More complex models include three levels, one considering the independent agent investments, another considering the system operator investments, and the final one to simulate market outcomes. For example, [27] formulates a trilevel model where the upper-level problem optimizes the system operator's transmission line and energy storage investments, the middle-level problem determines the merchant energy storage investment decisions, and the lower-level problem simulates a market clearing process for representative days using DC OPF. There are multiple ways of solving trilevel problems, but the first step is always to merge the middle- and lower-level problem into a single-level equivalent.

For conciseness, in this paper we consider a bilevel structure, however, the described procedure can be employed to trilevel problems as well. Also for brevity reasons, we choose a simple energy storage bidding model in the day-ahead market as the upper-level problem and focus on the AC OPF in the lower-level problem and on effectively converting and solving the initial problem.

For a strategic player (or a system operator) to solve a bilevel problem, it needs to first convert it into a single-level equivalent. However, the conversion techniques do not hold for nonconvex models. An excellent review of bilevel optimization approaches can be found in [28] and [29], where different techniques, later applied for solving the problem described in this paper, are explained in detail. Traditionally, the reduction techniques of transforming a bilevel problem into a single-level one are based on the primal-dual theory and the Karush–Kuhn–Tucker optimality conditions [30]. Both techniques follow the idea of replacing the lower-level problem with a set of equations and inequalities that have the same solution as the original lower-level problem. This set of equations is then added as a set of constraints to the upper-level problem, finally forming a single-level problem equivalent to the initial bilevel problem.

The main obstacle in the AC OPF formulations for strategic bidding are complementarity condition constraints, which are difficult for any interior point solver since they do not satisfy the Mangasarian-Fromovitz constraint qualification, meaning there is no strictly feasible point, making the nonlinear programming numerically unstable [31]. This aspect is one of the reasons why, to the best of the authors' knowledge, all existing bilevel market models rely either on the DC approximation [32] or linearized AC models [33]. This is because in these cases the complementary conditions can be reshaped into a mixed-integer linear form and solved using the branch-and-bound method. We found only two publications that either explicitly address or can be generalized to bilevel formulations and include exact, relaxed or approximate AC OPF formulations in the lower-level problem. The authors of [34] propose a bilevel problem of the worst contingency under attack, where they address the nonlinearities of the AC model as convex SOCP and SDP relaxations. However, this single-level reduction approach is applicable only because the interaction variable between the upper level and the lower level is binary, thus allowing for an exact reformulation of the resulting bilinear terms using the McCormick envelopes. In the second one [1], the authors propose a successive linear algorithm based on the dual form of a linear Taylor expansion of the IV-ACOPF model to solve the AC OPF. Their approach can be expanded to bilevel modeling with AC OPF-constrained market clearing in the lower level, however, their algorithm on the presented 14_ieee network does not converge. After six iterations and a DC OPF-based starting point, the LMP oscillations are at 0.20%. If the upper level is added to the model, its interaction with the lower level could potentially further destabilise convergency. It also uses generally non-trivial constraint violation penalties and constrains maximum

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207088

deviations of the Taylor delta variables.

Besides the techniques listed above, there are some interesting mathematical tools whose applicability has not yet been explored in the power systems community. In paper [35], the authors examine the Lipschitzian and differential properties of smoothing functions and their application to optimization with SOC complementarity constraints. Application of the smoothing techniques still results in nonlinear expressions, however, much easier to solve since they have all derivations in every point of the function. An example of smoothing a linear complementary condition is shown in Figure 2. The perpendicular lines represent the following three constraints $x \geqslant 0$, $y \geqslant 0$ and $x \cdot y = 0$, while the corresponding smooth curve is represented by only a single smooth constraint $x + y - \sqrt{x^2 + y^2 + 2\boldsymbol{\epsilon}^2} = 0$. Despite the promising theoretical foundations of the reformulated smooth constraints to achieve good numerical tractability with interior point based solvers, the technique has not yet been practically used or demonstrated. In this paper we employ the novel smoothing techniques, first proposed and developed in [35], to tractably solve bilevel problems with AC OPF in the lower level. This allows us to avoid model linearization and use any AC OPF formulation of continuous SOCP class or simpler. Smoothing technique implementation details are explained in Section II.G of Part II of this work.

Based on the above, the paper brings the following original contribution:

- For the first time we develop and present a mathematical formulation of a bilevel problem based on the smoothing techniques, where a strategic player's profit maximization is in the upper level, while the AC OPF problem is the lower level. For demonstration purposes, we choose energy storage (ES) as the strategic player making the bidding decisions in the upper-level problem. However any other strategic player, e.g. generator, demand response, aggregator of different flexibility sources, or a system operator can be plugged into the upper level. Furthermore, the model can be easily expanded into an investment model that considers multiple representative days.
- The bilevel problem is reduced to a single-level problem using the known techniques, i.e. primal-dual counterpart, strong duality theorem, McCormick envelopes, complementary slackness, penalty factor and interaction discretization. For all these techniques the computational tractability, effort and accuracy are analyzed and issues are detected and elaborated.
- All the developed models and codes are made available as open access to the scientific community at [36].

Rest of the paper is structured as follows. Section II mathematically states the proposed model. It is divided into two parts: Subsection II-A presents the initial model, which is reformulated into its SOC form in Subsection II-B. Section III introduces the algorithm and presolve to obtain the model's prerequisites and verifies the accuracy of the obtained solution. Solution techniques and the case studies are presented in the companion paper.

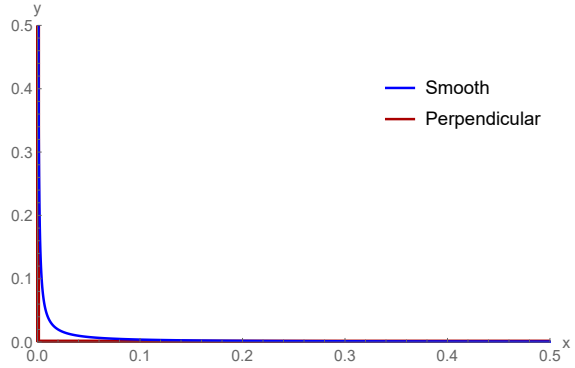

Fig. 2: Smoothing example of the perpendicular relation.

II. Mathematical Model

The presented bilevel optimization model consists of a simple ES active and reactive power bidding model in the upper level and the AC OPF in the lower level. Mathematical focus is on demonstrating an accurate and computationally tractable bilevel AC OPF solution.

A. Initial Model

1) Upper Level:

The upper level considers a large ES unit located at bus $i \in \beta$. The objective function (1.1) maximizes its profit by performing the day-ahead energy and reactive power market arbitrage while its impact on the prices is acknowledged by the dual variables $\lambda_{1,t,i}$ for active power and $\lambda_{2,t,i}$ for reactive. Constraint (1.2) models the storage (dis)charging process considering its efficiency, while (1.3)–(1.5) limit the maximum ES capacity, charging and discharging rates. Constraint (1.6) combines the charged and discharged energy into a cumulative quantity p_t^{es}. Binary variable x_t^{p} disables simultaneous charging and discharging which could otherwise occur in periods with negative prices. However, in many cases in the case study x_t^{p} is dropped-out to evaluate the performance of the solution techniques on a fully continuous model. When x_t^{p} is dropped out, constraints (1.4) and (1.5) are conceptually not needed due to constraint (1.7) which limits ES the apparent power.

$$\underset{\Xi^{\text{ul}}}{\text{Max}} \sum_{t,i\in\beta} (p_t^{\text{es}} \cdot \lambda_{1,t,i} + q_t^{\text{es}} \cdot \lambda_{2,t,i}) \tag{1.1}$$

$$SoE_t = SoE_{t-1} + p_t^{\text{ch}} \cdot \boldsymbol{\eta}^{\mathbf{ch}} - p_t^{\text{dis}} / \boldsymbol{\eta}^{\mathbf{dis}}, \quad \forall t \tag{1.2}$$

$$0 \leqslant SoE_t \leqslant \overline{\boldsymbol{SoE}}, \quad \forall t \tag{1.3}$$

$$0 \leqslant p_t^{\text{ch}} \leqslant \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot x_t^{\text{p}}, \quad \forall t \tag{1.4}$$

$$0 \leqslant p_t^{\text{dis}} \leqslant \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot (1 - x_t^{\text{p}}), \quad \forall t \tag{1.5}$$

$$p_t^{\text{es}} = p_t^{\text{ch}} - p_t^{\text{dis}}, \quad \forall t \tag{1.6}$$

$$(p_t^{\text{es}})^2 + (q_t^{\text{es}})^2 \leqslant (\overline{\boldsymbol{s}}^{\mathbf{es}})^2, \quad \forall t \tag{1.7}$$

The upper-level set of variables is $\Xi^{\text{ul}} = \{p_t^{\text{es}}, p_t^{\text{ch}}, p_t^{\text{dis}}, q_t^{\text{es}}, SoE_t, x_t^{\text{p}}\}$. x_t^{p} is an optional binary variable that can be neglected in case of nonnegative market prices.

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207088

2) Lower Level:

Transmission-constrained electricity markets perform market clearing by running an OPF optimization. Here we model the network using the CPSOTA [17]. The CPSOTA model is chosen due to convexity which is needed to satisfy the single-level reduction regularity conditions and due to superb warm-start transmission system OPF accuracy as demonstrated in [17] as well as in the case study of the second part of this paper. The CPSOTA model is designed to exploit a meshed network structure to achieve accuracy and should not be used with radial networks. Both the loads' and ES bids are considered inelastic, meaning they will always be cleared. This assumption holds if the loads (including the ES when buying) bid very high prices and the ES bids very low price when selling energy. Thus, both the loads and the ES (dis)charged energy p_t^{es} are modeled as parameters in the lower level. This simplification is introduced to shorten the upper-level objective function convexification (4.1) and its follow-up versions appearing in the Part II paper.

The lower level minimizes the quadratic production costs in its objective function (2.1). It is assumed that generators bid reactive power at zero prices. Bus power balance is imposed in (2.2) and (2.3). The upper-level variables p_t^{es} and q_t^{es} appear in constraints (2.2) and (2.3) only for the bus at which the ES is located, i.e. when condition $:i\in\beta$ under the variable is true. Our constraint writing style also assumes that summation indices are first fixed by the outer *for all* statement and then the summation rolls over to the remaining unfixed indices. This way $\sum_{(e,i,j)\in E\cup E^{\mathrm{R}}} P_{t,e,i,j}$, $\forall t,i$ sums the active powers from all branches originating from bus i. Constraints (2.4)–(2.7) are active and reactive power flow equations. Since these are based on the Taylor expansion, the computed voltage magnitude is equal to the assumed operating point value $\boldsymbol{V}_{t,i}^{\mathbf{op}}$ plus deviation $V_{t,i}^{\Delta}$. Similarly, the computed voltage angle is $\boldsymbol{\theta}_{t,i}^{\mathbf{op}}+\theta_{t,i}^{\Delta}$. To shorten the expressions, parameters $\boldsymbol{cps}_{t,e,i,j}$ and $\boldsymbol{cms}_{t,e,i,j}$ are defined over the other basic parameters, i.e. the ones present in the nomenclature, as introduced in the Appendix (A.1.1)–(A.2.2). Constraints (2.8.1) and (2.9.1) are second-order Taylor voltage magnitude and angle approximations, respectively, whose variables $\breve{V}_{t,e}$ and $\widehat{cos}_{t,i,j}$ also appear in the power flow equations. To achieve convexity, these constraints are inequalities as no quadratic equality is convex. At the solution point, it is intended that these constraints are binding so that no errors occur due to a swap of the equality sign with inequality. To achieve this, there is a presolve step in the optimization, as described in Section III, which decides on swapping the quadratic inequality constraints that are likely to diverge from the inequality boundary with simple linear equality constraints (2.8.2) and (2.9.2), thus avoiding gross errors. Applying the linear equality constraint variants disregards the second-order part of the Taylor. Even the zero-order Taylor expansion is exact at the expansion point, so disregarding a few second-order terms on per-branch or per-bus pair basis does not worsen the solution significantly in the vicinity of the expansion point. Constraints (2.10) and (2.11) limit the generators' production capabilities. Constraint (2.12) limits the maximum branch apparent power. Since (2.12) is a major source of computationally demanding quadratic equations, the presolve reduces their number for the main solve using a preset threshold at the initial operating point. More on the multi-step presolve procedure which finds an approximate operating point $\boldsymbol{V}_{t,i}^{\mathbf{op}}$, $\boldsymbol{\theta}_{t,i}^{\mathbf{op}}$ and decides on the use of quadratic constraints, i.e. determines the values of the Boolean parameters $\boldsymbol{\Lambda}_{t,e}$, $\boldsymbol{\Gamma}_{t,i,j}$ and $\boldsymbol{\Phi}_{t,e,i,j}$, can be found in Section III. Finally, constraint (2.13) sets the reference bus angle to zero and constraint (2.14) sets the voltage magnitude bounds.

$$\underset{\Xi^{\mathrm{ll}}}{\mathrm{Min}}\ \Omega^{\mathrm{p}} := \sum_{t,k}(\ddot{\boldsymbol{c}}_k\cdot(P_{t,k}^{\mathrm{g}})^2+\dot{\boldsymbol{c}}_k\cdot P_{t,k}^{\mathrm{g}}+\boldsymbol{c}_k) \tag{2.1}$$

$$\begin{aligned}&\sum_{k\in G_i}P_{t,k}^{\mathrm{g}}-\sum_{l\in L_i}\boldsymbol{P}_{t,l}^{\mathbf{d}}-\sum_{(e,i,j)\in E\cup E^{\mathrm{R}}}P_{t,e,i,j}-\underset{:i\in\beta}{p_t^{\mathrm{es}}}\\&-((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot\sum_{s\in S_i}\boldsymbol{g}_s^{\mathbf{sh}}=0,\quad\forall t,i\quad:\lambda_{1,t,i}\end{aligned} \tag{2.2}$$

$$\begin{aligned}&\sum_{k\in G_i}Q_{t,k}^{\mathrm{g}}-\sum_{l\in L_i}\boldsymbol{Q}_{t,l}^{\mathbf{d}}-\sum_{(e,i,j)\in E\cup E^{\mathrm{R}}}Q_{t,e,i,j}-\underset{:i\in\beta}{q_t^{\mathrm{es}}}\\&+((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot\sum_{s\in S_i}\boldsymbol{b}_s^{\mathbf{sh}}=0,\quad\forall t,i\quad:\lambda_{2,t,i}\end{aligned} \tag{2.3}$$

$$\begin{aligned}P_{t,e,i,j}=&((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot(\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{fr}})/\boldsymbol{\tau}_e^2+\breve{V}_{t,e}/2\\&-\boldsymbol{cps}_{t,e,i,j}\cdot(\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot\widehat{cos}_{t,i,j}+V_{t,i}^{\Delta}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}+V_{t,j}^{\Delta}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}})/\boldsymbol{\tau}_e\\&-\boldsymbol{cms}_{t,e,i,j}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})/\boldsymbol{\tau}_e,\\&\qquad\forall t,(e,i,j)\in E\quad:\lambda_{3,t,e,i,j}\end{aligned} \tag{2.4}$$

$$\begin{aligned}P_{t,e,i,j}=&((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot(\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{to}})+\breve{V}_{t,e}/2\\&-\boldsymbol{cps}_{t,e,i,j}\cdot(\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot\widehat{cos}_{t,j,i}+V_{t,i}^{\Delta}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}+V_{t,j}^{\Delta}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}})/\boldsymbol{\tau}_e\\&-\boldsymbol{cms}_{t,e,i,j}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})/\boldsymbol{\tau}_e,\\&\qquad\forall t,(e,i,j)\in E^{\mathrm{R}}\quad:\lambda_{4,t,e,i,j}\end{aligned} \tag{2.5}$$

$$\begin{aligned}Q_{t,e,i,j}=&-((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot(\boldsymbol{b}_e+\boldsymbol{b}_e^{\mathbf{fr}})/\boldsymbol{\tau}_e^2\\&+\boldsymbol{cms}_{t,e,i,j}\cdot(\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot\widehat{cos}_{t,i,j}+V_{t,i}^{\Delta}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}+V_{t,j}^{\Delta}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}})/\boldsymbol{\tau}_e\\&-\boldsymbol{cps}_{t,e,i,j}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})/\boldsymbol{\tau}_e,\\&\qquad\forall t,(e,i,j)\in E\quad:\lambda_{5,t,e,i,j}\end{aligned} \tag{2.6}$$

$$\begin{aligned}Q_{t,e,i,j}=&-((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot(\boldsymbol{b}_e+\boldsymbol{b}_e^{\mathbf{to}})\\&+\boldsymbol{cms}_{t,e,i,j}\cdot(\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot\widehat{cos}_{t,j,i}+V_{t,i}^{\Delta}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}+V_{t,j}^{\Delta}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}})/\boldsymbol{\tau}_e\\&-\boldsymbol{cps}_{t,e,i,j}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})/\boldsymbol{\tau}_e,\\&\qquad\forall t,(e,i,j)\in E^{\mathrm{R}}\quad:\lambda_{6,t,e,i,j}\end{aligned} \tag{2.7}$$

$$\breve{V}_{t,e}\geqslant(\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{fr}})\cdot(V_{t,i}^{\Delta})^2/\boldsymbol{\tau}_e^2-2\cdot\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e)\cdot V_{t,i}^{\Delta}V_{t,j}^{\Delta}/\boldsymbol{\tau}_e+(\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{to}})\cdot(V_{t,j}^{\Delta})^2,\quad\forall t,(e,i,j)\in E:\boldsymbol{\Lambda}_{t,e} \tag{2.8.1}$$

$$\breve{V}_{t,e}=0,\quad\forall t,(e,i,j)\in E:\neg\boldsymbol{\Lambda}_{t,e}\quad:\lambda_{10,t,e,i,j} \tag{2.8.2}$$

$$\widehat{cos}_{t,i,j}\leqslant 1-(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})^2/2,\quad\forall t,(i,j)\in N^{\mathrm{P}}:\boldsymbol{\Gamma}_{t,i,j} \tag{2.9.1}$$

$$\widehat{cos}_{t,i,j}=1,\quad\forall t,(i,j)\in N^{\mathrm{P}}:\neg\boldsymbol{\Gamma}_{t,i,j}\quad:\lambda_{13,t,i,j} \tag{2.9.2}$$

$$\underline{\boldsymbol{P}}_k^{\mathbf{g}}\leqslant P_{t,k}^{\mathrm{g}}\leqslant\overline{\boldsymbol{P}}_k^{\mathbf{g}},\quad\forall t,k\quad:\underline{\mu}_{3,t,k},\overline{\mu}_{3,t,k} \tag{2.10}$$

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207088

$$\underline{\boldsymbol{Q}}^{\mathbf{g}}_{k} \leqslant Q^{\mathrm{g}}_{t,k} \leqslant \overline{\boldsymbol{Q}}^{\mathbf{g}}_{k}, \quad \forall t,k \quad : \underline{\mu}_{4,t,k}, \overline{\mu}_{4,t,k} \tag{2.11}$$

$$P^2_{t,e,i,j}+Q^2_{t,e,i,j} \leqslant \overline{\boldsymbol{S}}^2_{e}, \quad \forall t,(e,i,j)\in E\cup E^{\mathrm{R}} : \boldsymbol{\Phi}_{t,e,i,j}, \\ : \lambda_{14,t,e,i,j}, \lambda_{15,t,e,i,j}, \mu_{5,t,e,i,j} \tag{2.12}$$

$$\boldsymbol{\theta}^{\mathbf{op}}_{t,i} + \theta^{\Delta}_{t,i} = 0, \quad \forall t, i \in R \quad : \lambda_{16,t,i} \tag{2.13}$$

$$\underline{\boldsymbol{V}}_{i} \leqslant \boldsymbol{V}^{\mathbf{op}}_{t,i} + V^{\Delta}_{t,i} \leqslant \overline{\boldsymbol{V}}_{i}, \quad \forall t, i \quad : \underline{\mu}_{6,t,i}, \overline{\mu}_{6,t,i} \tag{2.14}$$

Lower level set of primal variables: $\Xi^{\mathrm{ll}} = \{\theta^{\Delta}_{t,i}, V^{\Delta}_{t,i}, P_{t,e,i,j}, Q_{t,e,i,j}, P^{\mathrm{g}}_{t,k}, Q^{\mathrm{g}}_{t,k}\, \widehat{cos}_{t,i,j}, \breve{V}_{t,e}\}$.

B. SOC Constraint Reformulation

To derive the dual problem needed for a single-level reduction, quadratic primal constraints are first converted into their equivalent SOC form. Otherwise, the direct conversion would result in a general nonlinear formulation provided in [37], as opposed to our SOCP dual. Thus, this conversion is essential for solution techniques reliant on the SOCP form, i.e. primal-dual counterpart, McCormick envelopes, interaction discretization and smoothing techniques from the Part II paper. A SOC constraint conversion is nontrivial because there exist infinitely many SOC formulations of the same convex quadratic constraint. On the other hand, there also exists a formula with a unique positive semidefinite solution for conversion of the convex quadratic constraints to the SOC form [38]. However, we decided to manually choose our own SOC form since the general formula can result in a form with more than minimum possible number of cone variables and extensively complicated coefficients.

The resulting basic cone constraints are (3.1.1) and (3.2.1). Together with the substitution defining (3.1.2)–(3.1.5) and (3.2.2)–(3.2.4), they are equivalent to the initial quadratic voltage magnitude (2.8.1) and angle (2.9.1) constraints, respectively. To shorten the expressions, coefficients, i.e. parameters, $\boldsymbol{p}_{1,t,e,i,j}$, $\boldsymbol{p}_{2,t,e}$ and $\boldsymbol{p}_{3,t,e,i,j}$ are defined in the Appendix (A.3)–(A.5). Also, it is useful to note that the basic cone constraints share the same dual variable with their right-hand-side substitution constraint.

The remaining quadratic parts of the model are the branch apparent power limit constraint (2.12), which is already in the SOC form, and the objective function (2.1). To derive the dual, we recognize three ways of dealing with a quadratic objective function:

- apply the QP duality theory;
- transform it into a single large (multidimensional) SOC constraint;
- transform it into multiple three dimensional SOC constraints.

In this work we apply the QP duality theory to the objective function, while the rest of the model is converted into the dual using the SOCP procedure. By the QP duality theory, a dual is derived by writing the Karush–Kuhn–Tucker (KKT) conditions and then eliminating the remaining primal variables, which remain after derivations due to the squares in the primal objective function, by substituting them from the KKTs into the Lagrange function as described in lecture [39]. The SOC constraints are converted into the dual by using the mathematical theorem that states that the basic cones are self-dual [40], i.e. for every primal basic SOC constraint there is a dual basic SOC constraint (see primal-dual constraint pair (4.10) and (4.11) from Part II paper). Other constraints are linear and their KKT conditions can be used to derive the dual. This way the dual model is of the same optimization class as the primal, i.e. the objective function is convex quadratic and the constraints are SOC. The write-out of the dual is available in the Appendix. The two other objective function transformation approaches are not preferred over the QP procedure since they enlarge both the primal and dual models and, since the SOC constraints are inequalities, introduce additional complementary slackness conditions.

$$w^2_{1,t,e,i,j}+w^2_{2,t,e,i,j}+w^2_{3,t,e,i,j} \leqslant w^2_{0,t,e,i,j}, \quad \forall t,(e,i,j)\in E : \boldsymbol{\Lambda}_{t,e} \tag{3.1.1}$$

$$w_{1,t,e,i,j} = (1-\breve{V}_{t,e})/2, \quad \forall t,(e,i,j)\in E : \boldsymbol{\Lambda}_{t,e} \quad : \lambda_{7,t,e,i,j} \tag{3.1.2}$$

$$w_{2,t,e,i,j} = \boldsymbol{p}_{1,t,e,i,j} \cdot V^{\Delta}_{t,i} - \boldsymbol{p}_{2,t,e} \cdot V^{\Delta}_{t,j}, \\ \forall t,(e,i,j)\in E : \boldsymbol{\Lambda}_{t,e} \quad : \lambda_{8,t,e,i,j} \tag{3.1.3}$$

$$w_{3,t,e,i,j} = \boldsymbol{p}_{3,t,e,i,j} \cdot V^{\Delta}_{t,i}, \quad \forall t,(e,i,j)\in E : \boldsymbol{\Lambda}_{t,e} \quad : \lambda_{9,t,e,i,j} \tag{3.1.4}$$

$$w_{0,t,e,i,j} = (1+\breve{V}_{t,e})/2, \quad \forall t,(e,i,j)\in E : \boldsymbol{\Lambda}_{t,e} \quad : \mu_{1,t,e,i,j} \tag{3.1.5}$$

$$f^2_{1,t,i,j}+f^2_{2,t,i,j} \leqslant f^2_{0,t,i,j}, \quad \forall t,(i,j)\in N^{\mathrm{P}} : \boldsymbol{\Gamma}_{t,i,j} \tag{3.2.1}$$

$$f_{1,t,i,j} = (\theta^{\Delta}_{t,i} - \theta^{\Delta}_{t,j})/\sqrt{2}, \ \forall t,(i,j)\in N^{\mathrm{P}} : \boldsymbol{\Gamma}_{t,i,j} \ : \lambda_{11,t,i,j} \tag{3.2.2}$$

$$f_{2,t,i,j} = \widehat{cos}_{t,i,j} - 3/4, \ \forall t,(i,j)\in N^{\mathrm{P}} : \boldsymbol{\Gamma}_{t,i,j} \ : \lambda_{12,t,i,j} \tag{3.2.3}$$

$$f_{0,t,i,j} = -\widehat{cos}_{t,i,j} + 5/4, \ \forall t,(i,j)\in N^{\mathrm{P}} : \boldsymbol{\Gamma}_{t,i,j} \ : \mu_{2,t,i,j} \tag{3.2.4}$$

The reformulated lower-level set of variables is $\Xi^{\mathrm{r}} = \Xi^{\mathrm{ll}} \cup \{w_{0,t,e,i,j}, w_{1,t,e,i,j}, w_{2,t,e,i,j}, w_{3,t,e,i,j}, f_{0,t,i,j}, f_{1,t,i,j}, f_{2,t,i,j}\}$.

III. Algorithm and Presolve

The presented bilevel model requires as an input both the numerical and the Boolean parameters that need to be determined beforehand. This section introduces the algorithm to obtain prerequisites and verify the accuracy of the obtained solution.

The lower-level problem consists of an AC OPF model based on a Taylor expansion and thus requires an approximate operating point for voltage magnitude $\boldsymbol{V}^{\mathbf{op}}_{t,i}$ and angle $\boldsymbol{\theta}^{\mathbf{op}}_{t,i}$ as inputs. The first step in the Algorithm is to compute it using the exact polar model. The computation is based on the assumption that the strategic player in the upper-level problem is passive, i.e. the energy storage is neither charging nor discharging. Thus, the optimization is a typical, single-level AC OPF. To reduce the computational burden of the main bilevel optimization, the first step determines on which lines in the forthcoming steps the power limits will not be imposed, controlled with Boolean parameter $\boldsymbol{\Phi}_{t,e,i,j}$, using a preset threshold. In case the final solution overloads some of the lines, the threshold can be changed.

The second step is the presolve, aimed to drastically improve the AC OPF approximation model accuracy. The main

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207088

potential inaccuracy source is the relaxation of the equality sign in constraints (2.8.1) and (2.9.1) into the inequality sign to achieve convexity. To prevent relaxation errors in the forthcoming steps, this step marks all the constraints that would at the operating point deviate from the inequality boundary and replaces them with their linear equality variants (2.8.2) and (2.9.2) controlled with Boolean parameters $\mathbf{\Lambda}_{t,e}$ and $\mathbf{\Gamma}_{t,i,j}$. This step itself is nonconvex. It reruns the operating point using CPSOTA, but uses exclusively (2.8.1) and (2.9.1) quadratic constraints as equalities, thus not susceptible to relaxation errors and not requiring any information about the $\mathbf{\Lambda}_{t,e}$ and $\mathbf{\Gamma}_{t,i,j}$ parameters which it determines for the forthcoming steps. The selection of constraints for the forthcoming steps is based on their marginal value, computed by default by many solvers, e.g. IPOPT and Knitro. A constraint marginal is a sensitivity of the primal objective function on adding a small positive constant to the right-hand side of the constraint and thus its sign indicates whether an equality constraint would be binding if it were relaxed into inequality. In case of the voltage constraint (2.8.1), it would be binding if the marginal is positive and thus $\mathbf{\Lambda}_{t,e}$ is true. The cosine constraint (2.9.1) it would be binding if the marginal is negative and thus $\mathbf{\Gamma}_{t,i,j}$ is true. This step has the same solution as the first one and all delta variables $V^{\Delta}_{t,i}$ and $\theta^{\Delta}_{t,i}$ are zero. The described presolve working principle is further explained in [17]. It minimizes (2.1) subject to (2.2)–(2.14) with (2.8.1) and (2.9.1) as equalities, without (2.8.2) and (2.9.2), with respect to the variables set Ξ^{ll}.

The third and fourth steps are convex and simply resolve the primal and dual problems at the determined operating point ($\boldsymbol{V}^{\mathbf{op}}_{t,i}$ and $\boldsymbol{\theta}^{\mathbf{op}}_{t,i}$) for the selected constraints ($\mathbf{\Lambda}_{t,e}$, $\mathbf{\Gamma}_{t,i,j}$ and $\mathbf{\Phi}_{t,e,i,j}$) to supply the variables with their warm start values for the bilevel solve. Warm start values initialize the interior-point-based solvers, e.g. initializing Jacobian and Hessian matrices, enhancing the numerical tractability. The strategic player is still considered passive and, at the solution, the objective function values are the same as in the previous steps. The third step solves the SOCP version of the lower-level problem and the fourth step solves the SOCP dual. Specifically, third step minimizes (2.1) subject to (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4) with respect to the variables set Ξ^{r} and the forth step maximizes (B.1) subject to (B.2)–(B.12) with respect to the variables set Ξ^{du}.

The sixth step, which comes after the bilevel solve step five, verifies the solution accuracy. It solves the exact polar AC OPF with fixed (dis)charging decisions to the bilevel solve. It determines the actual system expenses and the upper-level profit for decisions from the bilevel problem. For computing the upper-level profit, nodal prices are obtained from the nodal power balance constraint marginal. The described procedure is itemized in Algorithm 1.

The proposed Algorithm is conceptually iterable to improve accuracy. Steps 1–6 can be run in a loop where the first step computes a new operating point assuming fixed (dis)charging decisions from the last bilevel solve. When looping, Steps 1 and 6 solve the same problem and can be performed in a single optimization. The described Algorithm is visually presented in Figure 3.

Algorithm 1 Bilevel optimization

1: Run exact polar model ▷ NLP; determine OP, $\mathbf{\Phi}_{t,e,i,j}$
2: Run presolve ▷ NC-QCQP; determine $\mathbf{\Lambda}_{t,e}$, $\mathbf{\Gamma}_{t,i,j}$
3: Run LL-primal ▷ SOCP; warm start solver
4: Run LL-dual ▷ SOCP; warm start solver
5: Run bilevel optimization ▷ Opt. class varies; deter. p^{es}_t, q^{es}_t
6: Run exact polar model ▷ NLP; verify solution

IV. Conclusion

This paper presented the mathematical model of a strategic energy storage acting in the day-ahead market with the market clearing based on the AC OPF in the lower level. It presents the SOC constraint reformulation and proposes an algorithm to accurately solve such complex structure in a time-efficient manner. The solution techniques as well as the case studies are presented in the accompanying paper.

Appendix

A. Parameters

The following parameters are defined over parameters from the nomenclature and are used to shorten the model formulation. Parameters defined in (A.1.1)–(A.2.2) are used in the lower-level primal problem for power flow constraints (2.4)–(2.7) and subsequently in the dual problem. The naming scheme is inspired by the parameters definition, i.e. *cps* stands for *cosine-plus-sine* and *cms* stands for *cosine-minus-sine*. The other three parameters defined in (A.3)–(A.5) are used to shorten the SOC constraint reformulation from Subsection II-B and, subsequently, in the dual problem.

$$\boldsymbol{cps}_{t,e,i,j} := \boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e), \quad \forall t, (e,i,j) \in E \quad \text{(A.1.1)}$$

$$\boldsymbol{cps}_{t,e,i,j} := \boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e) + \boldsymbol{b}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e), \quad \forall t, (e,i,j) \in E^{\mathrm{R}} \quad \text{(A.1.2)}$$

$$\boldsymbol{cms}_{t,e,i,j} := \boldsymbol{b}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e) - \boldsymbol{g}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e), \quad \forall t, (e,i,j) \in E \quad \text{(A.2.1)}$$

$$\boldsymbol{cms}_{t,e,i,j} := \boldsymbol{b}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e) - \boldsymbol{g}_e \cdot \sin(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} + \boldsymbol{\sigma}_e), \quad \forall t, (e,i,j) \in E^{\mathrm{R}} \quad \text{(A.2.2)}$$

$$\boldsymbol{p}_{1,t,e,i,j} := \frac{-\boldsymbol{g}_e \cdot \cos(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e)}{\sqrt{\boldsymbol{g}_e + \boldsymbol{g}^{\mathbf{to}}_e \cdot \boldsymbol{\tau}_e}}, \quad \forall t, (e,i,j) \in E \quad \text{(A.3)}$$

$$\boldsymbol{p}_{2,t,e} := \sqrt{\boldsymbol{g}_e + \boldsymbol{g}^{\mathbf{fr}}_e}, \quad \forall t, e \quad \text{(A.4)}$$

$$\boldsymbol{p}_{3,t,e,i,j} := \sqrt{\frac{\boldsymbol{g}^2_e \cdot \sin^2(\boldsymbol{\theta}^{\mathbf{op}}_{t,i} - \boldsymbol{\theta}^{\mathbf{op}}_{t,j} - \boldsymbol{\sigma}_e) + \boldsymbol{g}_e \cdot (\boldsymbol{g}^{\mathbf{fr}}_e + \boldsymbol{g}^{\mathbf{to}}_e) + \boldsymbol{g}^{\mathbf{fr}}_e \cdot \boldsymbol{g}^{\mathbf{to}}_e}{\boldsymbol{g}_e + \boldsymbol{g}^{\mathbf{fr}}_e}}, \quad \forall t, (e,i,j) \in E \quad \text{(A.5)}$$

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207088

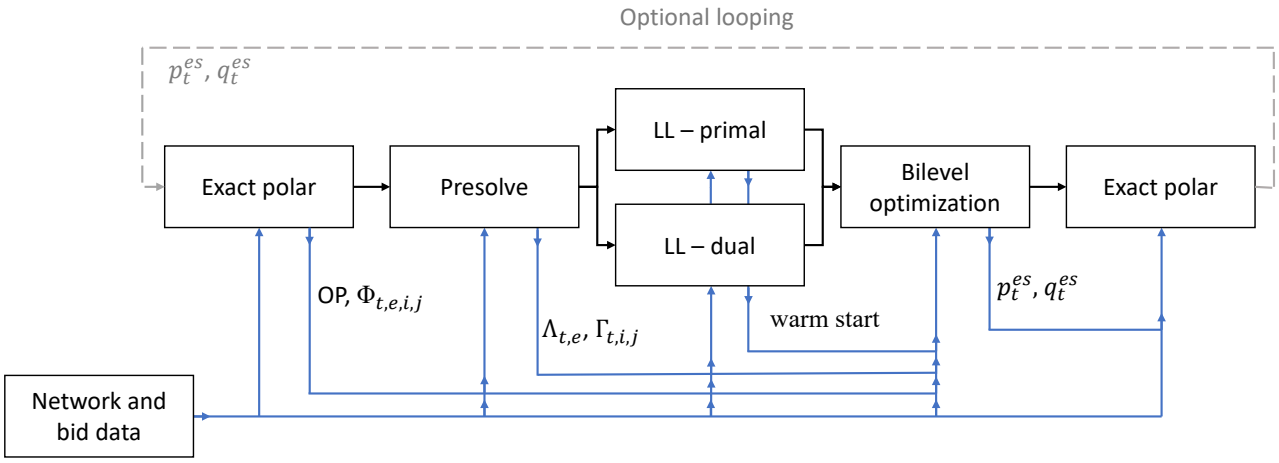

Fig. 3: Visualization of the Algorithm 1 and input data.

B. Dual

The dual objective function (B.1) is followed by constraints (B.2)–(B.9), obtained by derivation of the primal problem in variable order as appearing in the Ξ^{ll} set from Subsection II-A. The next three constraints (B.10)–(B.12) are the dual SOC constraints of (3.1.1), (3.2.1) and (2.12), respectively (due to the self-duality principle of the basic cones). The last constraint (B.13) is a nonnegativity condition for the dual variables associated to the primal inequality constraints.

$$-\lambda_{1,t,i}+\underset{:(e,i,j)\in E}{\lambda_{3,t,e,i,j}}+\underset{:(e,i,j)\in E^{\mathrm{R}}}{\lambda_{4,t,e,i,j}}-\underset{:\boldsymbol{\Phi}_{t,e,i,j}}{\lambda_{14,t,e,i,j}}=0,\ \forall t,(e,i,j)\in E\cup E^{\mathrm{R}} \tag{B.4}$$

$$-\lambda_{2,t,i}+\underset{:(e,i,j)\in E}{\lambda_{5,t,e,i,j}}+\underset{:(e,i,j)\in E^{\mathrm{R}}}{\lambda_{6,t,e,i,j}}-\underset{:\boldsymbol{\Phi}_{t,e,i,j}}{\lambda_{15,t,e,i,j}}=0,\ \forall t,(e,i,j)\in E\cup E^{\mathrm{R}} \tag{B.5}$$

$$\dot{\boldsymbol{c}}_k+\overline{\mu}_{3,t,k}-\underline{\mu}_{3,t,k}+\underset{:k\in G_i}{\sum_{i}}\lambda_{1,t,i}=0,\quad \forall t,k:\ddot{\boldsymbol{c}}_k=0 \tag{B.6}$$

$$\overline{\mu}_{4,t,k}-\underline{\mu}_{4,t,k}+\underset{:k\in G_i}{\sum_{i}}\lambda_{2,t,i}=0,\quad \forall t,k \tag{B.7}$$

$$\underset{:\boldsymbol{\Gamma}_{t,i,j}}{\mu_{2,t,i,j}}-\underset{:\boldsymbol{\Gamma}_{t,i,j}}{\lambda_{12,t,i,j}}+\underset{:\neg\boldsymbol{\Gamma}_{t,i,j}}{\lambda_{13,t,i,j}}+\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot\underset{(e,i,j)\in E}{\sum}(\boldsymbol{cps}_{t,e,i,j}\cdot\lambda_{3,t,e,i,j}+\boldsymbol{cps}_{t,e,j,i}\cdot\lambda_{4,t,e,j,i}-\boldsymbol{cms}_{t,e,i,j}\cdot\lambda_{5,t,e,i,j}-\boldsymbol{cms}_{t,e,j,i}\cdot\lambda_{6,t,e,j,i})/\boldsymbol{\tau}_e=0,\quad \forall t,(i,j)\in N^{\mathrm{P}} \tag{B.8}$$

$$(-\lambda_{3,t,e,i,j}-\lambda_{4,t,e,j,i}+\underset{:\boldsymbol{\Lambda}_{t,e}}{\lambda_{7,t,e,i,j}}-\underset{:\boldsymbol{\Lambda}_{t,e}}{\mu_{1,t,e,i,j}})/2+\underset{:\neg\boldsymbol{\Lambda}_{t,e}}{\lambda_{10,t,e,i,j}}=0,\quad \forall t,(e,i,j)\in E \tag{B.9}$$

$$\lambda_{7,t,e,i,j}^2+\lambda_{8,t,e,i,j}^2+\lambda_{9,t,e,i,j}^2\leqslant\mu_{1,t,e,i,j}^2,\quad \forall t,(e,i,j)\in E:\boldsymbol{\Lambda}_{t,e} \tag{B.10}$$

$$\lambda_{11,t,i,j}^2+\lambda_{12,t,i,j}^2\leqslant\mu_{2,t,i,j}^2,\quad \forall t,(i,j)\in N^{\mathrm{P}}:\boldsymbol{\Gamma}_{t,i,j} \tag{B.11}$$

$$\lambda_{14,t,e,i,j}^2+\lambda_{15,t,e,i,j}^2\leqslant\mu_{5,t,e,i,j}^2,\quad \forall t,(e,i,j)\in E:\boldsymbol{\Phi}_{t,e,i,j} \tag{B.12}$$

$$\mu\geqslant 0 \tag{B.13}$$

$$\begin{aligned}
\underset{\Xi^{\mathrm{du}}}{\mathrm{Max}}\ \Omega^{\mathrm{d}}:=&\sum_{t,k}(\boldsymbol{c}_k-\overline{\boldsymbol{P}}_k^{\mathbf{g}}\cdot\overline{\mu}_{3,t,k}+\underline{\boldsymbol{P}}_k^{\mathbf{g}}\cdot\underline{\mu}_{3,t,k}-\overline{\boldsymbol{Q}}_k^{\mathbf{g}}\cdot\overline{\mu}_{4,t,k}+\underline{\boldsymbol{Q}}_k^{\mathbf{g}}\cdot\underline{\mu}_{4,t,k})-\underset{:\ddot{\boldsymbol{c}}_k>0}{\sum_{t,k}}(\dot{\boldsymbol{c}}_k+\overline{\mu}_{3,t,k}-\underline{\mu}_{3,t,k}+\underset{:k\in G_i}{\sum_{i}}\lambda_{1,t,i})^2/(4\cdot\ddot{\boldsymbol{c}}_k)\\
&+\sum_{t,i}\Big((\boldsymbol{V}_{t,i}^{\mathbf{op}}-\overline{\boldsymbol{V}}_i)\cdot\overline{\mu}_{6,t,i}+(\underline{\boldsymbol{V}}_i-\boldsymbol{V}_{t,i}^{\mathbf{op}})\cdot\underline{\mu}_{6,t,i}\Big)-\sum_{t,i}\sum_{l\in L_i}(\boldsymbol{P}_{t,l}^{\mathbf{d}}\cdot\lambda_{1,t,i}+\boldsymbol{Q}_{t,l}^{\mathbf{d}}\cdot\lambda_{2,t,i})-\sum_{t,i}\sum_{s\in S_i}(\boldsymbol{g}_s^{\mathbf{sh}}\cdot\lambda_{1,t,i}-\boldsymbol{b}_s^{\mathbf{sh}}\cdot\lambda_{2,t,i})-\underset{:\neg\boldsymbol{\Gamma}_{t,i,j}}{\sum_{t,(i,j)\in N^{\mathrm{P}}}}\lambda_{13,t,i,j}\\
&+\underset{:\boldsymbol{\Gamma}_{t,i,j}}{\sum_{t,(i,j)\in N^{\mathrm{P}}}}(3/4\cdot\lambda_{12,t,i,j}-5/4\cdot\mu_{2,t,i,j})-\underset{:\boldsymbol{\Lambda}_{t,e}}{\sum_{t,(e,i,j)\in E}}(\lambda_{7,t,e,i,j}+\mu_{1,t,e,i,j})/2+\sum_{t,i\in R}\boldsymbol{\theta}_{t,i}^{\mathbf{op}}\cdot\lambda_{16,t,i}-\sum_{t,i\in\beta}p_t^{\mathrm{es}}\cdot\lambda_{1,t,i}-\sum_{t,i\in\beta}q_t^{\mathrm{es}}\cdot\lambda_{2,t,i}-\underset{:\boldsymbol{\Phi}_{t,e,i,j}}{\sum_{t,(e,i,j)\in E\cup E^{\mathrm{R}}}}\overline{\boldsymbol{S}}_e\cdot\mu_{5,t,e,i,j}\\
&-\sum_{t,(e,i,j)\in E}\big((\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{fr}})\cdot\lambda_{3,t,e,i,j}-(\boldsymbol{b}_e+\boldsymbol{b}_e^{\mathbf{fr}})\cdot\lambda_{5,t,e,i,j}\big)\cdot(\boldsymbol{V}_{t,i}^{\mathbf{op}})^2/\boldsymbol{\tau}_e^2-\sum_{t,(e,i,j)\in E^{\mathrm{R}}}\big((\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{to}})\cdot\lambda_{4,t,e,i,j}-(\boldsymbol{b}_e+\boldsymbol{b}_e^{\mathbf{to}})\cdot\lambda_{6,t,e,i,j}\big)\cdot(\boldsymbol{V}_{t,i}^{\mathbf{op}})^2
\end{aligned} \tag{B.1}$$

$$\begin{aligned}
&\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\sum_{(e,i,j)\in E}\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\boldsymbol{cms}_{t,e,i,j}\cdot\lambda_{3,t,e,i,j}-\boldsymbol{cms}_{t,e,j,i}\cdot\lambda_{4,t,e,j,i}+\boldsymbol{cps}_{t,e,i,j}\cdot\lambda_{5,t,e,i,j}-\boldsymbol{cps}_{t,e,j,i}\cdot\lambda_{6,t,e,j,i})/\boldsymbol{\tau}_e\\
&-\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\sum_{(e,i,j)\in E^{\mathrm{R}}}\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\boldsymbol{cms}_{t,e,j,i}\cdot\lambda_{3,t,e,j,i}-\boldsymbol{cms}_{t,e,i,j}\cdot\lambda_{4,t,e,i,j}+\boldsymbol{cps}_{t,e,j,i}\cdot\lambda_{5,t,e,j,i}-\boldsymbol{cps}_{t,e,i,j}\cdot\lambda_{6,t,e,i,j})/\boldsymbol{\tau}_e\\
&+\underset{:i\in R}{\lambda_{16,t,i}}-\underset{:\boldsymbol{\Gamma}_{t,i,j}}{\sum_{(i,j)\in N^{\mathrm{P}}}}\lambda_{11,t,i,j}/\sqrt{2}+\underset{:\boldsymbol{\Gamma}_{t,j,i}}{\sum_{(i,j)\in N^{\mathrm{PR}}}}\lambda_{11,t,j,i}/\sqrt{2}=0,\quad \forall t,i
\end{aligned} \tag{B.2}$$

$$\overline{\mu}_{6,t,i}-\underline{\mu}_{6,t,i}+2\cdot\boldsymbol{V}^{\mathbf{op}}_{t,i}\cdot\Big(\lambda_{2,t,i}\cdot\sum_{s\in S_i}\boldsymbol{b}^{\mathbf{sh}}_{s}-\lambda_{1,t,i}\cdot\sum_{s\in S_i}\boldsymbol{g}^{\mathbf{sh}}_{s}\Big)+\sum_{\substack{(e,i,j)\in E^{\mathrm{R}}\\ :\boldsymbol{\Lambda}_{t,e}}}\boldsymbol{p}_{2,t,e}\cdot\lambda_{8,t,e,j,i}-\sum_{\substack{(e,i,j)\in E\\ :\boldsymbol{\Lambda}_{t,e}}}\left(\boldsymbol{p}_{1,t,e,i,j}\cdot\lambda_{8,t,e,i,j}+\boldsymbol{p}_{3,t,e,i,j}\cdot\lambda_{9,t,e,i,j}\right)$$
$$+\sum_{(e,i,j)\in E}\Big[\left(-2\cdot(\boldsymbol{g}_e+\boldsymbol{g}^{\mathbf{fr}}_e)\cdot\boldsymbol{V}^{\mathbf{op}}_{t,i}/\boldsymbol{\tau}^2_e+\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot\boldsymbol{cps}_{t,e,i,j}/\boldsymbol{\tau}_e\right)\cdot\lambda_{3,t,e,i,j}+\left(2\cdot(\boldsymbol{b}_e+\boldsymbol{b}^{\mathbf{fr}}_e)\cdot\boldsymbol{V}^{\mathbf{op}}_{t,i}/\boldsymbol{\tau}^2_e-\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot\boldsymbol{cms}_{t,e,i,j}/\boldsymbol{\tau}_e\right)\cdot\lambda_{5,t,e,i,j}$$
$$+\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot(\boldsymbol{cps}_{t,e,j,i}\cdot\lambda_{4,t,e,j,i}-\boldsymbol{cms}_{t,e,j,i}\cdot\lambda_{6,t,e,j,i})/\boldsymbol{\tau}_e\Big]$$
$$+\sum_{(e,i,j)\in E^{\mathrm{R}}}\Big[\left(-2\cdot(\boldsymbol{g}_e+\boldsymbol{g}^{\mathbf{to}}_e)\cdot\boldsymbol{V}^{\mathbf{op}}_{t,i}+\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot\boldsymbol{cps}_{t,e,i,j}/\boldsymbol{\tau}_e\right)\cdot\lambda_{4,t,e,i,j}+\left(2\cdot(\boldsymbol{b}_e+\boldsymbol{b}^{\mathbf{to}}_e)\cdot\boldsymbol{V}^{\mathbf{op}}_{t,i}-\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot\boldsymbol{cms}_{t,e,i,j}/\boldsymbol{\tau}_e\right)\cdot\lambda_{6,t,e,i,j}$$
$$+\boldsymbol{V}^{\mathbf{op}}_{t,j}\cdot(\boldsymbol{cps}_{t,e,j,i}\cdot\lambda_{3,t,e,j,i}-\boldsymbol{cms}_{t,e,j,i}\cdot\lambda_{5,t,e,j,i})/\boldsymbol{\tau}_e\Big]=0,\quad\forall t,i \tag{B.3}$$

Set of dual variables Ξ^{du} contains all λ and μ variables.

REFERENCES

[1]P. Lipka, S. S. Oren, R. P. O'Neill and A. Castillo, "Running a More Complete Market With the SLP-IV-ACOPF," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1139–1148, March 2017.

[2]T. Wu, M. Rothleder, Z. Alaywan and A. D. Papalexopoulos, "Pricing energy and ancillary services in integrated market systems by an optimal power flow," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 339–347, Feb. 2004.

[3]G. L. Torres and V. H. Quintana, "An interior-point method for nonlinear optimal power flow using voltage rectangular coordinates," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1211–1218, Nov. 1998.

[4]D. I. Sun, B. Ashley, B. Brewer, A. Hughes and W. F. Tinney, "Optimal power flow by Newton approach," *IEEE Trans. Power App. Syst.*, vol. PAS-103, no. 10, pp. 2864–2880, Oct. 1984.

[5]R. P. O'Neill, A. Castillo and M. B. Cain, "The IV formulation and linear approximations of the ac optimal power flow problem," FERC, Washington, DC, USA, Tech. Rep., Dec. 2012. Accessed on: Jun. 9, 2021. [Online] Available at: cms.ferc.gov/sites/default/files/2020-05/acopf-2-iv-linearization.pdf

[6]R. A. Jabr, "Radial distribution load flow using conic programming," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1458–1459, Aug. 2006.

[7]C. Coffrin *et al. PGLib-OPF v20.07*, GitHub, Jul 29, 2020. Accessed on: March 23, 2021. [Online] Available at: github.com/power-grid-lib/pglib-opf/tree/v20.07

[8]H. Hijazi, C. Coffrin and P. Van Hentenryck, "Convex Quadratic Relaxations of Mixed-Integer Nonlinear Programs in Power Systems," *Math. Prog. Comp.*, Sept. 2016.

[9]C. Coffrin, H. L. Hijazi, and P. Van Hentenryck, "The QC Relaxation: A Theoretical and Computational Study on Optimal Power Flow," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3008–3018, July 2016.

[10]X. Bai, H. Wei, K. Fujisawa and Y. Wang, "Semidefinite programming for optimal power flow problems," *Int. J. Elec. Power Energy Syst.*, vol. 30, no. 6, pp. 383–392, July 2008.

[11]Z. Yang, H. Zhong, A. Bose, T. Zheng, Q. Xia and C. Kang, "A Linearized OPF Model With Reactive Power and Voltage Magnitude: A Pathway to Improve the MW-Only DC OPF," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1734–1745, March 2018.

[12]H. Zhang, G. T. Heydt, V. Vittal and J. Quintero, "An Improved Network Model for Transmission Expansion Planning Considering Reactive Power and Network Losses," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3471–3479, Aug. 2013.

[13]T. Akbari and M. Tavakoli Bina, "A linearized formulation of AC multi-year transmission expansion planning: A mixed-integer linear programming approach," *Electric Power Systems Research*, vol. 114, pp. 93–100, Sept. 2014.

[14]T. Akbari and M. Tavakoli Bina, "Linear approximated formulation of AC optimal power flow using binary discretisation," *Electric Power Systems Research*, vol. 10, no. 6, pp. 1117–1123, Apr. 2016.

[15]C. Coffrin and P. Van Hentenryck, "A Linear-Programming Approximation of AC Power Flows," *INFORMS J. Comp.*, vol. 26, no. 4, pp. 718–734, May 2014.

[16]C. Coffrin, H. Hijazi, and P. Van Hentenryck, "Alternating current (AC) power flow analysis in an electrical power network," WO Patent WO2013173879A1, Nov. 28, 2013.

[17]K. Šepetanc and H. Pandžić, "Convex Polar Second-Order Taylor Approximation of AC Power Flows: A Unit Commitment Study," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3585–3594, July 2021.

[18]S. J. Kazempour, A. J. Conejo and C. Ruiz, "Strategic Bidding for a Large Consumer," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 848–856, March 2015.

[19]A. Badri, S. Jadid and M. P. Moghaddam, "Impact of generators' different bidding strategies on system nash equilibrium point," *2007 International Power Engineering Conference (IPEC 2007)*, 2007, pp. 1–5.

[20]H. Pand žić and I. Kuzle, "Energy storage operation in the day-ahead electricity market," *2015 12th International Conference on the European Energy Market (EEM)*, 2015, pp. 1–6.

[21]M. Gonz ález Vayá and G. Andersson, "Optimal Bidding Strategy of a Plug-In Electric Vehicle Aggregator in Day-Ahead Electricity Markets Under Uncertainty," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2375–2385, Sept. 2015.

[22]K. Pandzic, K. Bruninx and H. Pandzic, "Managing Risks Faced by Strategic Battery Storage in Joint Energy-Reserve Markets," *IEEE Trans. Power Syst.*, Early Access.

[23]S. J. Kazempour and A. J. Conejo, "Strategic Generation Investment Under Uncertainty Via Benders Decomposition," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 424–432, Feb. 2012.

[24]L. P. Garces, A. J. Conejo, R. Garcia-Bertrand and R. Romero, "A Bilevel Approach to Transmission Expansion Planning Within a Market Environment," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1513–1522, Aug. 2009.

[25]Z. Saeed, and T. Akbari, "Bilevel transmission expansion planning using second-order cone programming considering wind investment," *Energy*, vol. 154, pp. 455–465, Apr. 2018.

[26]H. Hossein, and B. Zeng, "Bilevel conic transmission expansion planning," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4640–4642, May 2018.

[27]K. Pandzic, H. Pandzic and I. Kuzle, "Coordination of Regulated and Merchant Energy Storage Investments," *IEEE Trans. Sust. Energy*, vol. 9, no. 3, pp. 1244–1254, July 2018.

[28]A. Sinha, P. Malo and K. Deb, "A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications," *IEEE Trans. Evolutionary Computation*, vol. 22, no. 2, pp. 276–295, Apr. 2018.

[29]S. Pineda, H. Bylling and J. M. Morales, "Efficiently solving linear bilevel programming problems using off-the-shelf optimization software," *Optimization and Engineering*, vol. 19, no. 1, pp. 187—211, March 2018.

[30]A. Conejo, E. Castillo, R. Minguez and R. Garcia-Bertrand, "Decomposition Techniques in Mathematical Programming," Springer: Berlin/Heidelberg, Germany, 2006.

[31]H. Scheel and S. Scholtes, "Mathematical programs with complementarity constraints: Regularity, optimality conditions, and sensitivity," *Mathematics of Operations Research*, vol. 25, no. 1, pp. 1–22, Feb. 2000.

[32]C. Zhao and R. Jiang, "Distributionally Robust Contingency-Constrained Unit Commitment," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 94–102, Jan. 2018.

[33]Z. Yang, H. Zhong, A. Bose, T. Zheng, Q. Xia and C. Kang, "A Linearized OPF Model With Reactive Power and Voltage Magnitude: A Pathway to Improve the MW-Only DC OPF," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1734–1745, March 2018.

[34]B. Dandurand, K. Kim and S. Leyffer, "A Bilevel Approach for Identifying the Worst Contingencies for Nonconvex Alternating Current Power Systems," *SIAM Journal on Optimization*, vol. 31, no. 1, pp. 702-–726, Feb. 2021.

[35]M. Fukushima, Z. Luo and P. Tseng, "Smoothing Functions for Second-

Order-Cone Complementarity Problems," *SIAM Journal on Optimization*, vol. 12, no. 2, pp. 436–460, Jan. 2002.

[36]K. Šepetanc and H. Pandžić, *Bilevel-AC-OPF*, GitHub, June. 11, 2021. Accessed on: June 11, 2021. [Online] Available at: github.com/KSepetanc/Bilevel-AC-OPF

[37]M. Ç. Pınar, "A simple duality proof in convex quadratic programming with a quadratic constraint, and some applications," *European Journal of Operational Research*, vol. 124, no. 1, pp. 151–158, Jul. 2000.

[38]W. Wright, "Three Excursions around Conic Duality," April 2017. Accessed on: Jun. 9, 2021. [Online] Available at: galois.math.ucdavis.edu/lib/exe/fetch.php?media=conic_duality_talk.pdf

[39]A. Iouditski, "Lecture 4: Convex Programming and Lagrange Duality", Nov, 2015. [Online] Available at: www-ljk.imag.fr/membres/Anatoli.Iouditski/cours/convex/chapitre_4.pdf

[40]D. P. Williamson, "Mathematical Programming I: Lecture 26," Nov. 2014. [Online] Available at: people.orie.cornell.edu/dpw/orie6300/Lectures/lec26.pdf

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part II: Solution Techniques and Case Study

K. Šepetanc, *Student Member*, *IEEE*, H. Pandžić, *Senior Member*, *IEEE* and T. Capuder, *Member*, *IEEE*

***Abstract*—This is a second part of the research on AC optimal power flow being used in the lower level of the bilevel strategic bidding or investment models. As an example of a suitable upper-level problem, we observe a strategic bidding of energy storage and propose a novel formulation based on the smoothing technique.**

After presenting the idea and scope of our work, as well as the model itself and the solution algorithm in the companion paper (Part I), this paper presents a number of existing solution techniques and the proposed one based on smoothing the complementary conditions. The superiority of the proposed algorithm and smoothing techniques is demonstrated in terms of accuracy and computational tractability over multiple transmission networks of different sizes and different OPF models. The results indicate that the proposed approach outperforms all other options in both metrics by a significant margin. This is especially noticeable in the metric of accuracy where out of total 422 optimizations over 9 meshed networks the greatest AC OPF error is 0.023% that is further reduced to 3.3e-4% in the second iteration of our algorithm.

***Index Terms*—Bilevel models, AC OPF, complementary condition smoothing functions.**

I. Introduction

In the Part I of this work, we develop and present a mathematical formulation of a bilevel problem based on the smoothing techniques, where a strategic player's profit maximization is in the upper level and the AC OPF approximation in the lower level. Building upon Algorithm I presented in the Part I paper, in this paper we present a number of solution techniques that can be used to solve the bilevel strategic bidding problem at hand, as well as any other bilevel problem with the AC optimal power flow (AC OPF)-based market clearing algorithm in the lower level. These solution techniques, described in Section II of this paper, act as a baseline to which we compare the proposed smoothing techniques which differ in the smoothing function. The case study Section III consists of three main parts: the description and set-up Subsection III-A; and the two case studies. The first case study described in III-B demonstrates the model's accuracy, while the second one in III-D presents an in-depth solution techniques analysis. The final Section IV provides conclusive remarks.

The authors are with the Innovation Centre Nikola Tesla (ICENT) and the University of Zagreb Faculty of Electrical Engineering and Computing (e-mails: karlo.sepetanc@fer.hr; hrvoje.pandzic@fer.hr; tomislav.capuder@fer.hr). Employment of Karlo Šepetanc is funded by the Croatian Science Foundation under programme DOK-2018-09. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864298 (project ATTEST). The sole responsibility for the content of this document lies with the authors. It does not necessarily reflect the opinion of the Innovation and Networks Executive Agency (INEA) or the European Commission (EC). INEA or the EC are not responsible for any use that may be made of the information contained therein.

II. Solution Techniques

In this work we consider all classical techniques to solve a single-level reduced bilevel optimization problem, i.e. Step 5 of the Algorithm 1 from the Part I paper. The techniques differ in how well they close the duality gap, numerical tractability and ease of finding or converging to the global optimality as opposed to a local one. In the following subsections we first present the classical techniques, i.e. the primal-dual counterpart, the strong duality, the McCormic envelopes, the complementarity slackness, the penalty factor, the interaction discretization, followed by the proposed smoothing techniques, Chen-Harker-Kanzow-Smale and Kanzow.

A. Primal-dual counterpart

1) Primal-dual (PD):

This technique relies on the convexified objective function (4.1) to act as a penalty factor and does not enforce closure of the duality gap in any other way. One of the first such convexifications is presented in [1]. The resulting model simply consists of the lower-level primal and dual constraints as well as the upper-level constraints and the convexified quadratic objective function. Using this technique, the problem belongs to the second-order cone programming (SOCP) optimization class (with convex quadratic objective function). Dual constraints are available in the Appendix of the Part I paper.

The upper-level objective function is convexified by adding the term $\Omega^{\mathrm{d}} - \Omega^{\mathrm{p}}$. This way, the bilinear terms that cause nonconvexity, $p_t^{\mathrm{es}} \cdot \lambda_{1,t,i}$ and $q_t^{\mathrm{es}} \cdot \lambda_{2,t,i}$, are canceled since they also appear in the dual objective function Ω^{d}. The convexified objective function is equivalent to the original one if zero duality gap is ensured by the solution technique, i.e. $\Omega^{\mathrm{d}} = \Omega^{\mathrm{p}}$.

$$\mathrm{Max} \sum_{t,i\in\beta} \left(p_t^{\mathrm{es}} \cdot \lambda_{1,t,i} + q_t^{\mathrm{es}} \cdot \lambda_{2,t,i}\right) + \Omega^{\mathrm{d}} - \Omega^{\mathrm{p}} \tag{4.1}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4) and (B.2)–(B.13), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$. Note that equations (1)–(3) and (B) are from the Part I paper, where (1) denote the upper-level

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

constraints, constraints (2) lower-level primal constraints (primal feasibility KKT conditions), (3) are reformulated lower-level SOC constraints, i.e. reformulated constraints (2), while (B) are lower-level dual constraints, i.e. stationarity and dual feasibility KKT conditions.

2) Strengthened primal-dual (PD-S):

This technique applies an additional linear constraint (4.2) on top of the regular primal-dual technique. This constraint is obtained by writing the Karush–Kuhn–Tucker (KKT) stationarity conditions for $P^{\mathrm{g}}_{t,k}$. Otherwise, it is used to derive the dual as a part of dealing with the objective function convex quadraticity by substituting $P^{\mathrm{g}}_{t,k}$ from the Langrange function. Therefore, this constraint is not directly a part of the dual model, but it closes the duality gap since it connects both the primal and the dual variables. This technique is applicable only if there are generators with square cost bids ($\ddot{\boldsymbol{c}}_k > 0$). The optimization class is SOCP.

$$\dot{\boldsymbol{c}}_k + \overline{\mu}_{3,t,k} - \underline{\mu}_{3,t,k} + \sum_{\substack{i \\ :k \in G_i}} \lambda_{1,t,i} + 2 \cdot \ddot{\boldsymbol{c}}_k \cdot P^{\mathrm{g}}_{t,k} = 0, \quad \forall t, k : \ddot{\boldsymbol{c}}_k > 0 \tag{4.2}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13) and (4.2), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

B. Strong Duality

1) Strong Duality (SD):

Strong duality technique directly enforces zero duality gap by enforcing constraint (4.3), as explained in [2]. The formulation is nonconvex quadratic due to the equality sign and bilinear $p^{\mathrm{es}}_t \cdot \lambda_{1,t,i}$ and $q^{\mathrm{es}}_t \cdot \lambda_{2,t,i}$ terms in eq. (4.3).

$$\Omega^{\mathrm{p}} = \Omega^{\mathrm{d}} \tag{4.3}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13) and (4.3), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

2) Relaxed Strong Duality (SD-R):

To potentially improve numerical stability, the strong duality constraint can be relaxed so that small gaps are allowed. For any primal-dual optimization problem pair that has a finite solution, assuming that the goal of the primal is minimization, $\Omega^{\mathrm{p}} \geqslant \Omega^{\mathrm{d}}$ weak duality holds even without such constraint in the model. Thus, it is sufficient to add a constraint with an opposite inequality sign to close the duality gap. Constraint (4.4) allows an absolute duality gap $\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is a small positive constant. The formulation is nonconvex quadratic due to bilinear $p^{\mathrm{es}}_t \cdot \lambda_{1,t,i}$ and $q^{\mathrm{es}}_t \cdot \lambda_{2,t,i}$ terms in constraint (4.4).

$$\Omega^{\mathrm{p}} \leqslant \Omega^{\mathrm{d}} + \boldsymbol{\epsilon} \tag{4.4}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13) and (4.4), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

C. McCormick Envelopes (MC)

McCormick envelopes [3] relax a bilinear term into a plane-bounded region. The technique requires an assumption on the bounds of the electricity price, i.e. $\underline{\boldsymbol{\lambda}}_{1,t,i}$, $\underline{\boldsymbol{\lambda}}_{2,t,i}$ for the lower bound and $\overline{\boldsymbol{\lambda}}_{1,t,i}$, $\overline{\boldsymbol{\lambda}}_{2,t,i}$ for the upper bound. Values of these parameters can be estimated from the obtained operating point in Step 1 of Algorithm 1, e.g. using fixed intervals around the computed prices. Variables w^{p}_t and w^{q}_t represent relaxations of $p^{\mathrm{es}}_t \cdot \lambda_{1,t,i}$ and $q^{\mathrm{es}}_t \cdot \lambda_{2,t,i}$ respectively. Constraint (4.5) is the strong duality constraint in the inequality form. $+p^{\mathrm{es}}_t \cdot \lambda_{1,t,i}$ and $+q^{\mathrm{es}}_t \cdot \lambda_{2,t,i}$ terms from (4.5) cancel the original negative terms from Ω^{d}, which are replaced with $-w^{\mathrm{p}}_t$ and $-w^{\mathrm{q}}_t$. Constraints (4.6) and (4.7) are McCormick underestimator planes, while constraints (4.8) and (4.9) are McCormick overestimator planes. Together, they form a relaxed convex feasible space. To shorten the formulation, they are written in the matrix form. The formulation belongs to the SOCP optimization class.

$$\Omega^{\mathrm{p}} \leqslant \Omega^{\mathrm{d}} + \sum_{t,i \in \beta} p^{\mathrm{es}}_t \cdot \lambda_{1,t,i} + \sum_{t,i \in \beta} q^{\mathrm{es}}_t \cdot \lambda_{2,t,i} - \sum_t w^{\mathrm{p}}_t - \sum_t w^{\mathrm{q}}_t \tag{4.5}$$

$$\begin{aligned} \begin{pmatrix} w^{\mathrm{p}}_t & w^{\mathrm{q}}_t \end{pmatrix}^{\intercal} \geqslant & -\overline{\boldsymbol{s}}^{\mathbf{es}} \cdot \begin{pmatrix} \lambda_{1,t,i} & \lambda_{2,t,i} \end{pmatrix}^{\intercal} + \begin{pmatrix} p^{\mathrm{es}}_t & q^{\mathrm{es}}_t \end{pmatrix}^{\intercal} \cdot \begin{pmatrix} \underline{\boldsymbol{\lambda}}_{1,t,i} & \underline{\boldsymbol{\lambda}}_{2,t,i} \end{pmatrix} \\ & + \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot \begin{pmatrix} \underline{\boldsymbol{\lambda}}_{1,t,i} & \underline{\boldsymbol{\lambda}}_{2,t,i} \end{pmatrix}^{\intercal}, \quad \forall t, i \in \beta \end{aligned} \tag{4.6}$$

$$\begin{aligned} \begin{pmatrix} w^{\mathrm{p}}_t & w^{\mathrm{q}}_t \end{pmatrix}^{\intercal} \geqslant & \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot \begin{pmatrix} \lambda_{1,t,i} & \lambda_{2,t,i} \end{pmatrix}^{\intercal} + \begin{pmatrix} p^{\mathrm{es}}_t & q^{\mathrm{es}}_t \end{pmatrix}^{\intercal} \cdot \begin{pmatrix} \overline{\boldsymbol{\lambda}}_{1,t,i} & \overline{\boldsymbol{\lambda}}_{2,t,i} \end{pmatrix} \\ & - \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot \begin{pmatrix} \overline{\boldsymbol{\lambda}}_{1,t,i} & \overline{\boldsymbol{\lambda}}_{2,t,i} \end{pmatrix}^{\intercal}, \quad \forall t, i \in \beta \end{aligned} \tag{4.7}$$

$$\begin{aligned} \begin{pmatrix} w^{\mathrm{p}}_t & w^{\mathrm{q}}_t \end{pmatrix}^{\intercal} \leqslant & \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot \begin{pmatrix} \lambda_{1,t,i} & \lambda_{2,t,i} \end{pmatrix}^{\intercal} + \begin{pmatrix} p^{\mathrm{es}}_t & q^{\mathrm{es}}_t \end{pmatrix}^{\intercal} \cdot \begin{pmatrix} \underline{\boldsymbol{\lambda}}_{1,t,i} & \underline{\boldsymbol{\lambda}}_{2,t,i} \end{pmatrix} \\ & - \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot \begin{pmatrix} \underline{\boldsymbol{\lambda}}_{1,t,i} & \underline{\boldsymbol{\lambda}}_{2,t,i} \end{pmatrix}^{\intercal}, \quad \forall t, i \in \beta \end{aligned} \tag{4.8}$$

$$\begin{aligned} \begin{pmatrix} w^{\mathrm{p}}_t & w^{\mathrm{q}}_t \end{pmatrix}^{\intercal} \leqslant & -\overline{\boldsymbol{s}}^{\mathbf{es}} \cdot \begin{pmatrix} \lambda_{1,t,i} & \lambda_{2,t,i} \end{pmatrix}^{\intercal} + \begin{pmatrix} p^{\mathrm{es}}_t & q^{\mathrm{es}}_t \end{pmatrix}^{\intercal} \cdot \begin{pmatrix} \overline{\boldsymbol{\lambda}}_{1,t,i} & \overline{\boldsymbol{\lambda}}_{2,t,i} \end{pmatrix} \\ & + \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot \begin{pmatrix} \overline{\boldsymbol{\lambda}}_{1,t,i} & \overline{\boldsymbol{\lambda}}_{2,t,i} \end{pmatrix}^{\intercal}, \quad \forall t, i \in \beta \end{aligned} \tag{4.9}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13) and (4.5)–(4.9), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}} \cup \{w^{\mathrm{p}}_t, w^{\mathrm{q}}_t\}$.

D. Complementary Slackness

1) Complementary Slackness (CS):

For any given basic SOC primal-dual constraint inequality pair (4.10) and (4.11), assuming primal vector variable $x = \begin{pmatrix} x_0 & \overline{x} \end{pmatrix}$, where $\overline{x} = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots \end{pmatrix}$, and analogous dual vector variable y, there is a complementary slackness condition (4.12). In case of linear inequalities, $\overline{x}$ is an empty vector. Thus, for linear inequalities, constraints (4.10)–(4.12) take the following forms respectively: $0 \leqslant x_0$, $0 \leqslant y_0$ and $x_0 \cdot y_0 = 0$. Normally, complementary slackness conditions fully close the duality gap. However, due to applying the QP duality theory to deal with a quadratic objective function of the lower-level (2.1) to derive the dual, as explained in the Part I paper, Section 2.B, constraint (4.2) is also required to obtain zero duality gap. The resulting formulation is nonconvex quadratic.

$$x_1^2 + x_2^2 + x_3^2 + \ldots \leqslant x_0^2 \tag{4.10}$$

$$y_1^2 + y_2^2 + y_3^2 + \ldots \leqslant y_0^2 \tag{4.11}$$

$$x_0 \cdot y_0 + x_1 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3 + \ldots = 0 \tag{4.12}$$

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13), (4.2) and constraints based on (4.12) (one for every primal-dual inequality pair), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

2) Relaxed Complementary Slackness (CS-R):

The relaxed complementary slackness technique enhances numerical tractability by allowing a small deviance of the complementary slackness conditions, as in constraint (4.13) and shown in Section 12.3.1.1 in [5]. The constraint is only bounded from the upper side since the left-hand side is always nonnegative due to SOC constraints from the primal and the dual. The resulting formulation is nonconvex quadratic.

$$x_0 \cdot y_0 + x_1 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3 + ... \leqslant \boldsymbol{\epsilon} \tag{4.13}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13), (4.2) and constraints based on (4.13) (one for every primal-dual inequality pair), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

3) Aggregated Complementary Slackness (CS-A):

Since $x^{\intercal}y$ is always nonnegative due to the primal and dual SOC constraints, complementary conditions can also be aggregated into a single large constraint as in (4.14) which sums over all primal and dual vector variable pairs x and y from set ξ. The resulting formulation is also nonconvex quadratic.

$$\sum_{(x,y)\in\xi} x^{\intercal}y = 0 \tag{4.14}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13), (4.2) and the constraint based on (4.14), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

4) Relaxed Aggregated Complementary Slackness (CS-AR):

The same way the individual complementary slackness conditions can be relaxed to potentially improve numerical tractability, the aggregated constraint can be relaxed as well. To make an easier comparison to the nonaggregated version, $\boldsymbol{\epsilon}$ parameter is enlarged for every primal-dual inequality pair. The resulting formulation is nonconvex quadratic.

$$\sum_{(x,y)\in\xi} x^{\intercal}y \leqslant \sum_{(x,y)\in\xi} \boldsymbol{\epsilon} \tag{4.15}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13), (4.2) and constraint based on (4.15), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

E. Penalty Factor

1) Penalty Factor – Strong Duality (PF-SD):

The penalty factor technique closes the duality gap by penalizing it in the main objective function (4.16). Conceptually, as the penalty factor $\boldsymbol{\pi}$ goes to infinity, the duality gap closes to zero. However, the formulation is nonconvex due to a bilinear term in the objective function and thus the global optimality can not be guaranteed and numerical issues may occur for high penalty factors. Properties of the penalty factor techniques are discused in [4].

$$\mathrm{Max} \sum_{t,i\in\beta} p_t^{\mathrm{es}} \cdot \lambda_{1,t,i} + \sum_{t,i\in\beta} q_t^{\mathrm{es}} \cdot \lambda_{2,t,i} + (1+\boldsymbol{\pi}) \cdot (\Omega^{\mathrm{d}} - \Omega^{\mathrm{p}}) \tag{4.16}$$

The problem maximizes (4.16) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4) and (B.2)–(B.13), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

2) Penalty Factor – Complementary Slackness (PF-CS):

Duality gap can also be closed by penalizing the deviance of complementary slackness from zero. The same as the strong duality version, this version is also nonconvex.

$$\mathrm{Max} \sum_{t,i\in\beta} p_t^{\mathrm{es}} \cdot \lambda_{1,t,i} + \sum_{t,i\in\beta} q_t^{\mathrm{es}} \cdot \lambda_{2,t,i} + (\Omega^{\mathrm{d}} - \Omega^{\mathrm{p}}) - \boldsymbol{\pi} \cdot \sum_{(x,y)\in\xi} x^{\intercal}y \tag{4.17}$$

The problem maximizes (4.17) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13) and (4.2), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

F. Interaction Discretization

For this group of techniques we define additional sets. B is a set of binary variables for binary expansion variables indexed with letter b (i.e. $\{1,2,...,\lceil \log_2 D \rceil\}$), and U is a set of binary variables for unary expansion indexed by u (i.e. $\{1,2,...,D\}$), where D is a number of discretization steps. This technique discretizeses the allowed charging p_t^{ch} and discharging p_t^{dis} quantity values so that the bilinear term $p_t^{\mathrm{es}} \cdot \lambda_{1,t,i}$ from the strong duality can be reformulated into a mixed-integer linear one. Analogous discretizations also need to be done for reactive power. Thus, we introduce variables for consumed and produced reactive powers q_t^{ch} and q_t^{dis}, respectively, and the binary variable x_t^{q} that prevents simultaneous power exchange in both directions. Their values are defined in constraints (4.18)–(4.20) in analogous way as in constraints (1.4)–(1.6). The technique can use either the strong duality constraint (in the inequality form) or the complementary-slackness-penalized objective function. Since interaction discretization techniques linearize the bilinear terms, which are the only source of nonconvexity except for the introduced discrete variables, the formulations are of mixed-integer SOCP (MISOCP) class. Techniques belonging to this group can theoretically find proven optimal solution with zero duality gap (assuming high penalty factor for penalty version). General formulation of the expansions can be found in [6]. The following constraints split the positive and negative ES reactive power in two variables since they are used separately in the following subsections.

$$0 \leqslant q_t^{\mathrm{ch}} \leqslant \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot x_t^{\mathrm{q}}, \quad \forall t \tag{4.18}$$

$$0 \leqslant q_t^{\mathrm{dis}} \leqslant \overline{\boldsymbol{s}}^{\mathbf{es}} \cdot (1 - x_t^{\mathrm{q}}), \quad \forall t \tag{4.19}$$

$$q_t^{\mathrm{es}} = q_t^{\mathrm{ch}} - q_t^{\mathrm{dis}}, \quad \forall t \tag{4.20}$$

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

1) Binary Expansion – Strong Duality (BE-SD):

Discretization by binary expansion uses binary variables with exponentially increasing (base 2) assigned weights to denote an integer number. The obtained integer number divided by a maximum achievable integer number represents the value the continuous variable will take. If the ratio is 1, the continuous variable takes the upper bound value, and for ratio 0 it takes the lower bound value. The other discrete states are spread evenly. The number of binary variables grows logarithmically with the number of discrete states. The introduced variables are $y_t^{\mathrm{p}}, y_t^{\mathrm{q}}$, integer variables, $x_{t,b}^{\mathrm{p,be}}, x_{t,b}^{\mathrm{q,be}}$ binary variables with assigned weights, and auxiliary variables $w_{t,b}^{\mathrm{p,be}}, w_{t,b}^{\mathrm{q,be}}\, w_t^{\mathrm{p,ch}}, w_t^{\mathrm{p,dis}}, w_t^{\mathrm{q,ch}}, w_t^{\mathrm{q,dis}}$. Using this technique, the strong duality constraint (4.21) takes a convex form since bilinear terms cancel out each other and are effectively replaced by terms $w_t^{\mathrm{p,ch}}, w_t^{\mathrm{p,dis}}, w_t^{\mathrm{q,ch}}, w_t^{\mathrm{q,dis}}$.

$$\Omega^{\mathrm{p}} \leqslant \Omega^{\mathrm{d}} + \sum_{t,i\in\beta}(p_t^{\mathrm{es}}\cdot\lambda_{1,t,i} + q_t^{\mathrm{es}}\cdot\lambda_{2,t,i}) - \sum_{t}(w_t^{\mathrm{p,ch}} - w_t^{\mathrm{p,dis}} + w_t^{\mathrm{q,ch}} - w_t^{\mathrm{q,dis}}) \tag{4.21}$$

$$0 \leqslant \left(y_t^{\mathrm{p}}\ y_t^{\mathrm{q}}\right) \leqslant 2^{|B|} - 1, \quad \forall t \tag{4.22}$$

$$\left(y_t^{\mathrm{p}}\ y_t^{\mathrm{q}}\right) = \sum_{b\in B} 2^{b-1}\cdot\left(x_{t,b}^{\mathrm{p,be}}\ x_{t,b}^{\mathrm{q,be}}\right), \quad \forall t \tag{4.23}$$

$$\left(p_t^{\mathrm{ch}}\ q_t^{\mathrm{ch}}\right)/\overline{\boldsymbol{s}}^{\mathbf{es}} + \left(p_t^{\mathrm{dis}}\ q_t^{\mathrm{dis}}\right)/\overline{\boldsymbol{s}}^{\mathbf{es}} = \left(y_t^{\mathrm{p}}\ y_t^{\mathrm{q}}\right)/(2^{|B|}-1), \ \forall t \tag{4.24}$$

$$\left(w_t^{\mathrm{p,ch}}\ w_t^{\mathrm{q,ch}}\right)/\overline{\boldsymbol{s}}^{\mathbf{es}} + \left(w_t^{\mathrm{p,dis}}\ w_t^{\mathrm{q,dis}}\right)/\overline{\boldsymbol{s}}^{\mathbf{es}} = \sum_{b\in B} 2^{b-1}\cdot\left(w_{t,b}^{\mathrm{p,be}}\ w_{t,b}^{\mathrm{q,be}}\right)/(2^{|B|}-1), \quad \forall t \tag{4.25}$$

$$\left(\underline{\boldsymbol{\lambda}}_{1,t,i}\ \underline{\boldsymbol{\lambda}}_{2,t,i}\right)^{\intercal}\cdot\left(x_t^{\mathrm{p}}\ x_t^{\mathrm{q}}\right)\cdot\overline{\boldsymbol{s}}^{\mathbf{es}} \leqslant \left(w_t^{\mathrm{p,ch}}\ w_t^{\mathrm{q,ch}}\right) \leqslant \left(\overline{\boldsymbol{\lambda}}_{1,t,i}\ \overline{\boldsymbol{\lambda}}_{2,t,i}\right)^{\intercal}\cdot\left(x_t^{\mathrm{p}}\ x_t^{\mathrm{q}}\right)\cdot\overline{\boldsymbol{s}}^{\mathbf{es}}, \quad \forall t \tag{4.26}$$

$$\left(\underline{\boldsymbol{\lambda}}_{1,t,i}\ \underline{\boldsymbol{\lambda}}_{2,t,i}\right)^{\intercal}\cdot\left(1-x_t^{\mathrm{p}}\ 1-x_t^{\mathrm{q}}\right)\cdot\overline{\boldsymbol{s}}^{\mathbf{es}} \leqslant \left(w_t^{\mathrm{p,dis}}\ w_t^{\mathrm{q,dis}}\right)^{\intercal} \leqslant \left(\overline{\boldsymbol{\lambda}}_{1,t,i}\ \overline{\boldsymbol{\lambda}}_{2,t,i}\right)^{\intercal}\cdot\left(1-x_t^{\mathrm{p}}\ 1-x_t^{\mathrm{q}}\right)\cdot\overline{\boldsymbol{s}}^{\mathbf{es}}, \quad \forall t \tag{4.27}$$

$$\left(x_{t,b}^{\mathrm{p,be}}\ x_{t,b}^{\mathrm{q,be}}\right)^{\intercal}\cdot\left(\underline{\boldsymbol{\lambda}}_{1,t,i}\ \underline{\boldsymbol{\lambda}}_{2,t,i}\right) \leqslant \left(w_{t,b}^{\mathrm{p,be}}\ w_{t,b}^{\mathrm{q,be}}\right)^{\intercal} \leqslant \left(x_{t,b}^{\mathrm{p,be}}\ x_{t,b}^{\mathrm{q,be}}\right)^{\intercal}\cdot\left(\overline{\boldsymbol{\lambda}}_{1,t,i}\ \overline{\boldsymbol{\lambda}}_{2,t,i}\right), \quad \forall t,i\in\beta,b\in B \tag{4.28}$$

$$\left(\lambda_{1,t,i}\ \lambda_{2,t,i}\right)^{\intercal} + \left(x_{t,b}^{\mathrm{p,be}}\ x_{t,b}^{\mathrm{q,be}}\right)^{\intercal}\cdot\left(\overline{\boldsymbol{\lambda}}_{1,t,i}\ \overline{\boldsymbol{\lambda}}_{2,t,i}\right) - \left(\overline{\boldsymbol{\lambda}}_{1,t,i}\ \overline{\boldsymbol{\lambda}}_{2,t,i}\right)^{\intercal} \leqslant \left(w_{t,b}^{\mathrm{p,be}} w_{t,b}^{\mathrm{q,be}}\right)^{\intercal} \leqslant \left(\lambda_{1,t,i}\ \lambda_{2,t,i}\right)^{\intercal} + \left(x_{t,b}^{\mathrm{p,be}} x_{t,b}^{\mathrm{q,be}}\right)^{\intercal}\cdot\left(\underline{\boldsymbol{\lambda}}_{1,t,i}\ \underline{\boldsymbol{\lambda}}_{2,t,i}\right) - \left(\underline{\boldsymbol{\lambda}}_{1,t,i}\ \underline{\boldsymbol{\lambda}}_{2,t,i}\right)^{\intercal}, \quad \forall t,i\in\beta,b\in B \tag{4.29}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13), (4.18)–(4.29), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}} \cup \{w_t^{\mathrm{p,ch}}, w_t^{\mathrm{q,ch}}, w_t^{\mathrm{p,dis}}, w_t^{\mathrm{q,dis}}, w_{t,b}^{\mathrm{p,be}}, w_{t,b}^{\mathrm{q,be}}, y_t^{\mathrm{p}}, y_t^{\mathrm{q}}, x_{t,b}^{\mathrm{p,be}}, x_{t,b}^{\mathrm{q,be}}\}$.

2) Binary Expansion – Penalty Factor (BE-PF):

The binary expansion technique can also be applied to the bilinear term $p_t^{\mathrm{es}}\cdot\lambda_{1,t,i}$ and its reactive power counterpart appearing in the objective function which penalizes the duality gap (4.16). In the resulting objective function (4.30), the bilinear terms from $\Omega^{\mathrm{d}} - \Omega^{\mathrm{p}}$ are first canceled out with an explicit addition of itself and then replaced with their equivalent linear expression $w_t^{\mathrm{p,ch}} - w_t^{\mathrm{p,dis}} + w_t^{\mathrm{q,ch}} - w_t^{\mathrm{p,dis}}$. Except for the lack of the strong duality constraint, the other constraints are the same as for the strong duality version of this technique.

$$\mathrm{Max} \sum_{t,i\in\beta}(1+\boldsymbol{\pi})\cdot(p_t^{\mathrm{es}}\cdot\lambda_{1,t,i} + q_t^{\mathrm{es}}\cdot\lambda_{2,t,i}) + (1+\boldsymbol{\pi})\cdot(\Omega^{\mathrm{d}} - \Omega^{\mathrm{p}}) - \boldsymbol{\pi}\cdot\sum_{t}(w_t^{\mathrm{p,ch}} - w_t^{\mathrm{p,dis}} + w_t^{\mathrm{q,ch}} - w_t^{\mathrm{p,dis}}) \tag{4.30}$$

The problem maximizes (4.30) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13), (4.18)–(4.20) and (4.22)–(4.29), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}} \cup \{w_t^{\mathrm{p,ch}}, w_t^{\mathrm{q,ch}}, w_t^{\mathrm{p,dis}}, w_t^{\mathrm{q,dis}}, w_{t,b}^{\mathrm{p,be}}, w_{t,b}^{\mathrm{q,be}}, y_t^{\mathrm{p}}, y_t^{\mathrm{q}}, x_{t,b}^{\mathrm{p,be}}, x_{t,b}^{\mathrm{q,be}}\}$.

3) Unary Expansion – Strong Duality (UE-SD):

Similarly to the binary expansion, the unary expansion uses binary variables to represent an integer number. However, the assigned weights in this technique increase linearly and only up to one binary variable is allowed to take value 1. The number of binary variables grows linearly with the number of discrete states. In addition to the variables from the binary expansion, the introduced variables are binary variables $x_{t,u}^{\mathrm{p,ue}}, x_{t,u}^{\mathrm{q,ue}}$ and auxiliary variables $w_{t,u}^{\mathrm{p,ue}}, w_{t,u}^{\mathrm{q,ue}}$.

$$0 \leqslant \left(y_t^{\mathrm{p}}\ y_t^{\mathrm{q}}\right) \leqslant |U|, \quad \forall t \tag{4.31}$$

$$\left(y_t^{\mathrm{p}}\ y_t^{\mathrm{q}}\right) = \sum_{u\in U} u\cdot\left(x_{t,u}^{\mathrm{p,ue}}\ x_{t,u}^{\mathrm{q,ue}}\right), \quad \forall t \tag{4.32}$$

$$\sum_{u\in U}\left(x_{t,u}^{\mathrm{p,ue}}\ x_{t,u}^{\mathrm{q,ue}}\right) \leqslant 1, \quad \forall t \tag{4.33}$$

$$\left(p_t^{\mathrm{ch}}\ q_t^{\mathrm{ch}}\right)/\overline{\boldsymbol{s}}^{\mathbf{es}} + \left(p_t^{\mathrm{dis}}\ q_t^{\mathrm{dis}}\right)/\overline{\boldsymbol{s}}^{\mathbf{es}} = \left(y_t^{\mathrm{p}}\ y_t^{\mathrm{q}}\right)/|U|, \quad \forall t \tag{4.34}$$

$$\left(w_t^{\mathrm{p,ch}}\ w_t^{\mathrm{q,ch}}\right)/\overline{\boldsymbol{s}}^{\mathbf{es}} + \left(w_t^{\mathrm{p,dis}}\ w_t^{\mathrm{q,dis}}\right)/\overline{\boldsymbol{s}}^{\mathbf{es}} = \sum_{u\in U} u\cdot w_{t,u}^{\mathrm{p,ue}}/|U|, \ \forall t \tag{4.35}$$

$$\left(x_{t,u}^{\mathrm{p,ue}}\ x_{t,u}^{\mathrm{q,ue}}\right)^{\intercal}\cdot\left(\underline{\boldsymbol{\lambda}}_{1,t,i}\ \underline{\boldsymbol{\lambda}}_{2,t,i}\right) \leqslant \left(w_{t,u}^{\mathrm{p,ue}}\ w_{t,u}^{\mathrm{q,ue}}\right)^{\intercal} \leqslant \left(x_{t,u}^{\mathrm{p,ue}}\ x_{t,u}^{\mathrm{q,ue}}\right)^{\intercal}\cdot\left(\overline{\boldsymbol{\lambda}}_{1,t,i}\ \overline{\boldsymbol{\lambda}}_{2,t,i}\right), \quad \forall t,i\in\beta,b\in B \tag{4.36}$$

$$\left(\lambda_{1,t,i}\ \lambda_{2,t,i}\right)^{\intercal} + \left(x_{t,u}^{\mathrm{p,ue}}\ x_{t,u}^{\mathrm{q,ue}}\right)^{\intercal}\cdot\left(\overline{\boldsymbol{\lambda}}_{1,t,i}\ \overline{\boldsymbol{\lambda}}_{2,t,i}\right) - \left(\overline{\boldsymbol{\lambda}}_{1,t,i}\ \overline{\boldsymbol{\lambda}}_{2,t,i}\right)^{\intercal} \leqslant \left(w_{t,u}^{\mathrm{p,ue}} w_{t,u}^{\mathrm{q,ue}}\right)^{\intercal} \leqslant \left(\lambda_{1,t,i}\ \lambda_{2,t,i}\right)^{\intercal} + \left(x_{t,u}^{\mathrm{p,ue}} x_{t,u}^{\mathrm{q,ue}}\right)^{\intercal}\cdot\left(\underline{\boldsymbol{\lambda}}_{1,t,i}\ \underline{\boldsymbol{\lambda}}_{2,t,i}\right) - \left(\underline{\boldsymbol{\lambda}}_{1,t,i}\ \underline{\boldsymbol{\lambda}}_{2,t,i}\right)^{\intercal}, \quad \forall t,i\in\beta,b\in B \tag{4.37}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.14) excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13), (4.18)–(4.21), (4.26), (4.27) and (4.31)–(4.37) with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}} \cup \{w_t^{\mathrm{p,ch}}, w_t^{\mathrm{q,ch}} w_t^{\mathrm{p,dis}}, w_t^{\mathrm{q,dis}}, w_{t,u}^{\mathrm{p,ue}}, w_{t,u}^{\mathrm{q,ue}}, y_t^{\mathrm{p}}, y_t^{\mathrm{q}}, x_{t,u}^{\mathrm{p,ue}}, x_{t,u}^{\mathrm{q,ue}}\}$.

4) Unary Expansion – Penalty Factor (UE-PF):

The same as the binary expansion, the unary expansion can also be applied to the penalized objective function (4.30). Other constraints, except for the lack of the strong duality constraint, are the same as for the strong duality version of the unary expansion.

The problem maximizes (4.30) subject to constraints (1.2)–(1.7), (2.2)–(2.14), excluding (2.8.1) and (2.9.1) in favor of (3.1.1)–(3.2.4), (B.2)–(B.13), (4.18)–(4.20), (4.26), (4.27) and (4.31)–(4.37), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}} \cup \{w_t^{\mathrm{p,ch}}, w_t^{\mathrm{q,ch}}, w_t^{\mathrm{p,dis}}, w_t^{\mathrm{q,dis}}, w_{t,u}^{\mathrm{p,ue}}, w_{t,u}^{\mathrm{q,ue}}, y_t^{\mathrm{p}}, y_t^{\mathrm{q}}, x_{t,u}^{\mathrm{p,ue}}, x_{t,u}^{\mathrm{q,ue}}\}$.

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

G. Smoothing techniques

In this section we consider smoothing of the complementary slackness conditions. Smoothing techniques are presented in [7], where they are generalized from smoothing the linear complementary conditions to smoothing the SOCP complementary conditions. The resulting equality constraints replace both the primal-dual inequality pair and the complementary slackness conditions (4.10)–(4.12) for SOC and linear inequalities. The only inequality constraints left in the model are those from the upper-level problem. Smoothing techniques become exact as $\boldsymbol{\epsilon}$ parameter approaches zero from the positive side. Formulations using the smoothing techniques belong to the nonconvex nonlinear class.

1) Chen–Harker–Kanzow–Smale (SM1):

This complementary slackness conditions smoothing technique is a special case of the Chen-Mangasarian smoothing functions. Only the vector constraint (4.38) is directly part of the optimization problem. Formulas (4.39)–(4.41) only define parts for substitution of constraint (4.38). Specifically, (4.39) defines a function, (4.40) defines a numerical expression and (4.41) defines a vector.

$$x - \boldsymbol{\epsilon} \cdot (F(\psi_1/\boldsymbol{\epsilon}) \cdot u_1 + F(\psi_2/\boldsymbol{\epsilon}) \cdot u_2) = 0 \tag{4.38}$$

$$F(\alpha) = (\sqrt{\alpha^2 + 4} + \alpha)/2 \tag{4.39}$$

$$\psi_n = x_0 - y_0 + (-1)^n \cdot \|\overline{x} - \overline{y}\| \tag{4.40}$$

$$u_n = \begin{pmatrix} 1/2 & 1/2 \cdot (-1)^n \cdot \frac{\overline{x}-\overline{y}}{\|\overline{x}-\overline{y}\|} \end{pmatrix} \tag{4.41}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.7), (2.8.2), (2.9.2), (2.13), (3.1.2)–(3.1.5), (3.2.2)–(3.2.4), (B.2)–(B.9), (4.2) and constraints based on (4.38) (one vector constraint for every primal-dual inequality and complementary condition pair), whose parts are defined in (4.39)–(4.41), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

2) Kanzow (SM2):

This smoothing function is a variation of the Fischer-Burmeister function. Analogous to the previous smoothing technique, (4.42) is a vector constraint, while (4.43) and (4.44) are only used to define parts of (4.42).

$$x + y - (\sqrt{\psi_1} \cdot u_1 + \sqrt{\psi_2} \cdot u_2) = 0 \tag{4.42}$$

$$\psi_n = \|x\|^2 + \|y\|^2 + 2 \cdot \boldsymbol{\epsilon}^2 + 2 \cdot (-1)^n \cdot \|x_0 \cdot \overline{x} + y_0 \cdot \overline{y}\| \tag{4.43}$$

$$u_n = \begin{pmatrix} 1/2 & 1/2 \cdot (-1)^n \cdot \frac{x_0\overline{x}+y_0\overline{y}}{\|x_0\overline{x}+y_0\overline{y}\|} \end{pmatrix} \tag{4.44}$$

The problem maximizes (4.1) subject to constraints (1.2)–(1.7), (2.2)–(2.7), (2.8.2), (2.9.2), (2.13), (3.1.2)–(3.1.5), (3.2.2)–(3.2.4), (B.2)–(B.9), (4.2) and constraints based on (4.42) (one vector constraint for every primal-dual inequality and complementary condition pair), whose parts are defined in (4.43) and (4.44), with respect to the variables set $\Xi^{\mathrm{ul}} \cup \Xi^{\mathrm{r}} \cup \Xi^{\mathrm{du}}$.

III. Case Study

A. Description and Set-Up

The case study consists of three parts. The first one demonstrates the accuracy of our convex polar second-order Taylor approximation (CPSOTA) model [8] in comparison to implementations that use the existing convex OPF formulations in the lower level, i.e. Jabr and DC. The accuracy comparison is based on the smoothing solution technique that effectively ensures strong duality ($\Omega^{\mathrm{p}} = \Omega^{\mathrm{d}}$). The second case study shows economic benefits for the ES and the system by comparing profits with only active power bids from the ES and both reactive and active power bids. The third case study evaluates different solution techniques with focus on accuracy, i.e. duality gap and objective function value, and numerical tractability.

The case study considers a large energy storage unit with 100 MWh capacity (1 p.u.), 60 MW (0.6 p.u.) maximum (dis)charging rate and 0.9 (dis)charging efficiency. The initial state-of-energy is set to 50% and the storage is allowed to end a day at any state-of-energy. All variable and parameter units are in p.u. or dimensionless.

Transmission system meshed networks were taken from the PGLib-OPF v19.05 [9] database. A 24-hour time horizon was added by scaling the loads with winter weekday profile factors from IEEE RTS-96 [10].

All problems were solved on a desktop PC (i7 9700; 32 GB, 2.67 GHz RAM) in AMPL. Convex and mixed-integer convex problems were solved in Xpress 8.10.1, while all other problems in KNITRO 12.3. The default solver settings were used, except for the Algorithm 1 step 5 (see Part I) for which the settings are stated individually in the case studies. Algorithms 1 steps 1–4 and 6 are single-level continuous optimizations and thus easy to solve. Their run times are less than a second even for the largest considered network with 57 buses. The threshold for imposing the line thermal power limits, controlled with Boolean parameter $\boldsymbol{\Phi}_{t,e,i,j}$, was set to 85% and no final solution violated the thermal limits.

B. Case Study I: Model Accuracy

Accuracy of the considered bilevel model mostly depends on the accuracy of the OPF in its lower level and the technique applied to ensure that the strong duality holds at the solution point. Since convexity of OPF is a requirement for strong duality, this case study compares our proposed model with the models based on commonly used convex OPFs: Jabr's [11] and DC. In addition, the bilevel approach is compared to two single-level simplifications. One assumes constant prices in the lower level regardless of the energy storage bidding strategy, and the other, centralized, models the ES as if it were owned by the system operator who minimizes the system costs. Prices for the first simplification are obtained in a pre-run by considering an idle energy storage, thus preventing its impact on market prices. For bilevel models, the strong duality is ensured with Chen–Harker–Kanzow–Smale (SM1) smoothing technique, described in Subsection II-G1. This technique's defining parameter $\boldsymbol{\epsilon}$ was set to 1e-3, which in practice ensures a duality gap of 1e-5% or better. To make

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

TABLE I
MODEL ACCURACY TESTED ON 3_LMBD NETWORK (IP – INTERIOR POINT, S – SIMPLEX)

	ES at bus 1						ES at bus 2						ES at bus 3					
	ES profit			System expenses			ES profit			System expenses			ES profit			System expenses		
ES active-power-only bids																		
Model	Actual	Computed	Diff [%]	Actual	Computed	Diff [%]	Actual	Computed	Diff [%]	Actual	Computed	Diff [%]	Actual	Computed	Diff [%]	Actual	Computed	Diff [%]
CPSOTA	1818.65	1818.56	-4.9e-3	98827.27	98827.34	7.1e-5	1359.88	1359.80	-5.9e-3	99342.16	99342.20	4.0e-5	2016.85	2017.79	0.047	98496.77	98496.94	1.7e-4
Jabr's [11]	1817.61	1641.17	-9.7	98821.80	98443.00	-0.38	1327.67	1582.75	19.2	99327.24	98506.39	-0.83	2007.06	1754.30	-13	98486.16	98304.91	-0.18
DC	1818.35	1757.64	-3.3	98824.47	97319.61	-1.5	1359.04	1364.71	0.42	99341.82	97774.13	-1.6	1986.50	1729.57	-13	98531.12	97236.05	-1.3
Centralized	1795.61			98811.38			1347.65			99317.95			1988.96			98481.34		
Fixed prices – IP	1329.03	2224.99	67	99013.29	98529.70	-0.49	719.18	1560.60	117	99614.43	99194.09	-0.42	1138.13	2915.78	156	98966.00	97838.91	-1.1
Fixed prices – S	1122.16	2224.99	98	99122.00	98529.70	-0.60	490.31	1560.60	218	99729.13	99194.09	-0.54	-150.98	2915.78	-2031	99345.15	97838.91	-1.5

the comparison easier, but without the loss of accuracy due to positive prices, i.e. simultaneous charging and discharging does not occur, models were run without the binary variable x_t^{P} for energy storage (ES) (dis)charging, otherwise present in constraints (1.4) and (1.5) in the Part I paper. Since the final model (Algorithm 1 step 5) is nonconvex-nonlinear, KNITRO multistart feature (16 starting points, ± 0.6 variable perturbations) was used to the increase chances of finding a global optimal solution. Numerical stability was increased by tightening the solver default convergence and infeasibility tolerances by a factor of 100 and by enabling the solver's warm start option. To better differentiate between active and reactive power accuracy, the case study contains separate analysis for the cases when ES can bid only active power and when it can bid both active and reactive power. The exception are the buses with were zero reactive power prices observed at the assumed operating point (Algorithm 1 Step 1) as in that case the two solutions are the same.

Accuracy comparison for three- and five-bus networks are presented in Tables I and II. The analysis is performed for an ES placed at each bus individually. Table data columns include actual ES profits and system expenses, i.e. verified quantities using Algorithm 1 Step 6, computed profits or expenses, using Algorithm 1 Step 5, and the percentage difference between the actual and the computed profits. The rows include considered models: CPSOTA, Jabr's [11], DC, centralized and fixed-price models solved using two different solver methods (interior point – IP and simplex – S). The tree-bus network table does not include active and reactive power bids data since the reactive power prices at all buses are zero. Similarly, the five-bus network table does not include data for active and reactive power bids at buses 4 and 5 due to zero reactive power prices. In the case of zero reactive power prices, the results are the same as in the case of active-power-only bids.

The results in Table I indicate that, despite the significant inaccuracy (actual vs. computed columns), all bilevel approaches still make generally good decisions for an ES bidding active power. For example, for ES at bus 1, Jabr's model underestimates the ES profit at only 1641.17 (10% underestimated), while the actual profit 1817.61 is very close to the best achieved actual profit of 1818.65. On the other hand, for the ES located at bus 2, the Jabr's model greatly overestimates the computed profit at 1582.75 (16% overestimated), while the actual profit is only 2.4% away from the best actual value, achieve by CPSOTA. For the ES at bus 3, the Jabr's model again underestimates the profit. For the DC model, when the ES is located at buses 1 or 2, the computed and actual ES profits are quite close, indicating that the model provides good estimates of the ES profit regardless on the lossless network representation. However, when the ES is at bus 3, which is the bus without generators capable of producing active power and is under the effect of congestion during multiple hours, the DC model underperforms. The centralized model achieves the lowest system expenses in all cases. However, the ES profits are worse then with the DC model at first two buses and better at the third. The fixed price approaches are inadequate for the considered systems. The actual profits are unfavourable and much lower than the computed ones. These severe differences between the computed and the actual profits when neglecting the impact of being a strategic player is in line with findings in [12]. Results of the fixed price approaches are also very susceptible to the solver method. The interior point method, as opposed to the simplex method, has a tendency of finding solutions with intermediate variable values, balancing the ES (dis)charging during low- or high-price periods since there are multiple hours with the same prices. A more even charging across multiple hours results in more favourable actual profits, despite the identical computed profits for both fixed-prices methods. In contrast, the proposed bilevel model based on CPSOTA AC OPF almost perfectly matches the computed and the verified values regardless of the considered ES bus placement. The CPSOTA's accuracy, being a Taylor expansion-based model, can be iteratively even further enhanced by reevaluating the operating point parameters $\boldsymbol{V}_{t,i}^{\mathbf{OP}}$ and $\boldsymbol{\theta}_{t,i}^{\mathbf{OP}}$. Regarding the system expenses, i.e. the lower-level objective function value, the Jabr's model consistently underestimates it, as the model is a relaxation. The same goes for the DC model since it does not consider network losses. The fixed-prices system expenses are computed assuming that they increase at marginal prices from a base point, i.e. systems expenses without the ES performing arbitrage. These as well underestimate the verified system expenses. Finally, the CPSOTA model results in negligible system expenses inaccuracy, indicating that the model almost perfectly computes the AC OPF.

Model accuracy is further examined on the 5_pjm network, whose results are displayed in Table II. Comparing the active power only bids, the Jabr's model highly overestimates the ES profit when connected to buses 1, 2 and 5. On the other hand, it underestimates the ES profit when connected to bus 4, while for the ES at bus 3 it accurately computes the ES profit. The reason for such diverse ES profit accuracy is

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

TABLE II
MODELS' ACCURACY TESTED ON 5_PJM NETWORK (IP – INTERIOR POINT, S – SIMPLEX)

Model		ES at bus 1			ES at bus 2			ES at bus 3			ES at bus 4			ES at bus 5		
		Actual	Computed	Diff [%]	Actual	Computed	Diff [%]	Actual	Computed	Diff [%]	Actual	Computed	Diff [%]	Actual	Computed	Diff [%]
ES active-power-only bids																
CPSOTA	ES profit	804.94	804.94	1.5e-4	1648.09	1648.07	-7.7e-4	1958.23	1958.22	-5.4e-4	2833.07	2833.10	9.5e-4	696.45	696.38	-0.010
	System expenses	295944.24	295944.23	-1.0e-6	295100.82	295100.82	1.4e-6	294790.81	294790.81	-2.8e-7	293915.06	293915.04	-6.5e-6	296050.47	296050.49	-5.9e-6
Jabr's [11]	ES profit	792.26	1923.73	143	1648.07	1956.98	19	1958.23	1958.22	5.1e-4	2833.06	1950.47	-31	171.65	1916.34	1016
	System expenses	295956.89	257518.89	-13	295100.80	257485.69	-13	294790.81	257484.66	-13	293915.01	257491.99	-12	296577.55	257526.16	-13
DC	ES profit	755.19	839.01	11	1590.18	1676.82	5.4	1901.61	2002.22	5.3	2761.95	2897.07	4.9	655.15	695.08	6.1
	System expenses	295947.22	294248.75	-0.57	295106.07	293410.94	-0.57	294795.54	293085.54	-0.58	293930.42	292190.70	-0.59	296066.12	294392.69	-0.57
Centralized	ES profit	761.87			1601.78			1911.97			2786.38			627.35		
	System expenses	295944.21			295100.67			294790.66			293914.75			296049.57		
Fixed prices – IP	ES profit	486.53	1083.59	123	1336.49	1945.05	46	1647.12	2255.19	37	2468.82	3131.36	27	-149.81	864.25	-677
	System expenses	296089.10	295665.66	-0.14	295236.83	294804.20	-0.15	294926.79	294494.06	-0.15	294051.53	293617.89	-0.15	296622.61	295885.00	-0.25
Fixed prices – S	ES profit	486.43	1083.59	123	1336.29	1945.05	46	1647.11	2255.19	37	2191.22	3131.36	43	-183.87	864.25	-570
	System expenses	296089.15	295665.66	-0.14	295236.93	294804.20	-0.15	294926.79	294494.06	-0.15	294055.85	293617.89	-0.15	296610.75	295885.00	-0.24
ES active and reactive power bids																
CPSOTA	ES profit	1170.06	1169.34	-0.061	1999.66	1989.00	-0.53	2016.55	2008.37	-0.41						
	System expenses	295575.18	295575.52	1.2e-4	294728.23	294733.50	1.8e-3	294711.51	294715.58	1.4e-3						
Jabr's [11]	ES profit	951.65	1923.73	102	1982.75	1980.23	-0.13	1958.22	1966.43	0.42						
	System expenses	295796.91	257518.89	-13	294747.12	257462.46	-13	294782.46	257484.66	-13						
Centralized	ES profit	1127.21			1952.71			1970.94								
	System expenses	295575.17			294727.43			294708.75								

that due to large AC OPF relaxation errors observed in the computed system expenses, wrong generators can be claimed as marginal ones, thus largely impacting the prices. This effect is very pronounced since the 5_pjm network only has linear generator cost curves. The DC model is again relatively accurate, but overestimating the ES profit at all buses by 4-11%. The centralized model again achieves similarly favourable ES profits as the DC model, i.e. slightly better at four busses and slightly worse at bus 5, but always achieves the lowest system expenses. Both fixed-price models also highly overestimate the ES profits, regardless on the ES position. This is because these models do not consider the effect the ES bidding strategy has on market prices. Namely, the ES tends to increase market prices when purchasing energy and reduce them when selling energy. The verified profits when the ES is located at bus 5 are actually negative for both fixed-price models, which is a result of ignoring the price changes by the ES's bidding actions. The proposed CPSOTA model results in almost perfect ES profit accuracy at all buses. The proposed model also results in almost identical computed and verified system expenses. Again, both the Jabr's and the DC models, as well as the fixed-price models, underestimate the actual system expenses.

Table II also includes the results when the ES bids both active and reactive power. These results generally show lower accuracy than in the case when only active power is bid. Reactive power is generally more difficult to accurately model with typically a few times greater power flow inaccuracy than for active power [13]. The Jabr's model again highly overestimates the ES profit at bus 1. At buses 2 and 3 the computed values are close to the actual ones, however, the actual profits are farther from the highest achieved with CPSOTA as compared to the active-power-only bidding. CPSOTA still achieves the highest actual ES profits. On the other hand, the centralized model is more consistent in terms of accuracy than the Jabr's model due to the use of an exact AC OPF, but has 2-4% lower ES profits than CPSOTA due to lack of bilevel optimization structure. The actual profit increase due to reactive power bidding highly depends on the ES placement. At bus 1 the increase is 365.66 (45.4%), at bus 2 it is 351.57 (21.3%), at bus 3 it is 58.32 (3.0%) and no increase at buses 4 and 5 due to zero reactive power prices.

Since the proposed CPSOTA model resulted in higher than normal inaccuracy on the 3_lmbd network for active-power-only bids for ES placed at bus 3 (however, this inaccuracy is still extremely low, less than 0.05%) and at buses 2 and 3 for active and reactive power bids on the 5_pjm network (0.54% and 0.41% error, respectively), we ran the second iteration of the Algorithm 1 (presented in the Part I paper), whose results are presented in Table III. The second iteration basically eliminates the remaining errors, displaying fast iterative convergency of the proposed algorithm. Thus, if an extremely high accuracy is required, this can be achieved by running the second iteration, which brings the error virtually to zero since the greatest remaining ES profit error is less than 0.02%.

Statistical data for larger networks obtained by running the CPSOTA-based model for ES at each bus is presented in Table IV. Buses with zero reactive power prices are excluded from the statistics for bidding both the active and reactive power. This makes for a total of 422 optimizations, 232 for active-power-only bids and 190 for active and reactive power bids. The shown errors are computed as a percentage relative difference between the actual and the computed values. As the optimization is run for the ES placement at each bus, the median, the mean and the maximum (Max) errors are listed in Table IV. The median errors better represent the most common error values than the mean errors since mean are significantly influenced by the outliers. The active power bidding median, as well as the mean ES profit errors, are

TABLE III
CPSOTA MODEL ACCURACY IN THE SECOND ITERATION

Network	ES at bus	ES profit			System expenses		
		Actual	Computed	Diff [%]	Actual	Computed	Diff [%]
ES active-power-only bids							
3_lmbd	3	2016.876	2016.884	4.0e-4	98497.667	98497.666	-1.0e-6
ES active and reactive power bids							
5_pjm	2	1999.662	1999.651	-5.5e-4	294728.228	294728.227	-3.4e-7
	3	2017.189	2016.792	-0.020	294708.966	294708.974	2.7e-6

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

TABLE IV
STATISTICS ON CPSOTA BILEVEL ACCURACY IN THE FIRST ITERATION

	ES profit errors (%)			System expenses errors (%)		
Network	Median	Mean	Max	Median	Mean	Max
ES active-power-only bids						
3_lmbd	5.8e-3	0.019	0.047	7.1e-5	9.3e-5	1.7e-4
5_pjm	9.5e-4	2.6e-3	0.010	1.3e-6	3.0e-6	6.5e-6
14_ieee	5.7e-4	8.5e-4	3.3e-3	2.4e-6	4.4e-6	1.7e-5
24_ieee_rts	1.9e-3	2.3e-3	8.3e-3	2.7e-6	3.2e-6	1.1e-5
30_as	2.6e-3	4.9e-3	0.033	8.8e-6	1.2e-5	6.4e-5
30_fsr	2.7e-3	4.6e-3	0.026	7.1e-6	1.2e-5	5.1e-5
30_ieee	0.014	0.025	0.19	1.4e-4	2.5e-4	1.7e-3
39_epri	1.3e-3	3.7e-3	0.024	4.1e-7	9.5e-7	6.5e-6
57_ieee	3.2e-3	5.9e-3	0.027	2.9e-6	8.7e-6	6.0e-5
ES active and reactive power bids						
5_pjm	0.41	0.33	0.53	1.4e-3	1.1e-3	1.8e-3
14_ieee	4.4e-3	0.10	0.88	4.2e-5	2.4e-4	1.8e-3
24_ieee_rts	3.7e-3	8.5e-3	0.072	1.0e-5	4.5e-5	3.3e-4
30_as	2.6e-3	7.2e-3	0.043	2.3e-5	3.2e-5	1.1e-4
30_fsr	6.1e-3	0.22	3.97	4.9e-5	1.3e-3	0.023
30_ieee	0.034	0.13	1.68	8.8e-4	1.5e-3	1.6e-3
39_epri	0.017	0.064	1.12	4.3e-6	2.9e-5	5.1e-4
57_ieee	0.087	0.18	1.28	1.4e-4	2.5e-4	1.3e-3

mainly in the range 0.01%–1e-3% and the system expenses median and mean errors, i.e. AC OPF errors, are in the range 1e-4%–1e-6%. The maximum errors, which occurred at the same bus and network, when the ES is bidding only active power are 0.19% for the ES profit and 1.7e-3% for the system expenses, which reduces to 4.0e-3% and 1.5e-7% in the second iteration of Algorithm 1. Errors when the ES bids both active and reactive powers are higher. The median and mean ES profit errors mainly range from 0.10% to 1e-3% and system expenses mean and median errors are in the range 1e-3–1e-5%. However, the maximum ES profit errors are significant ($\geqslant 1\%$) on four networks, reaching 3.97% at bus 9 of the 30_fsr network. We run the second iteration of Algorithm 1 for that case and the error was reduced to 0.035%. The maximum system expense error, which also occurred at the same bus and network as the maximum ES profit error, is 0.023%. In the second iteration it is reduced to 3.3e-4%.

C. Case Study II: Economical Benefits of ES Reactive Power Bids

Reactive power bids provide both the financial opportunity for the ES and benefit for the system. Reactive power prices are commonly 10 to 100 times lower than active power prices since generators can produce them without costs, leaving only indirect active power savings to influence the price. However, since the reactive power does not consume the ES state-of-energy, but only its power capacity, it can bid it in large quantities. 3_lmbd network is dropped from the following analysis since it has zero reactive power prices at all buses and time periods so the ES profit increase and the system savings are 0. Figure 1 shows the percentage profit increase for the ES due to reactive power bids sorted in a descending order. The profit increase significantly depends on the network and ES placement. At 5_pjm, 39_epri and 57_ieee networks the highest ES profit increases are in the range 26%–47% with one outlining profit increase of 731% at bus 30 of the 39_epri network. This large relative increase is a result of a low active power profit due to the constant active power marginal prices at this bus caused by a large generator, i.e. the ES just discharges all the stored energy and performs no other arbitrage. On the other hand, high locational reactive power prices are a result of a large generator with high minimum reactive power output connected to bus 30. At most of the other buses and considered networks, the profit increases are much lower and range from 0% to 5% as shown in the Figure 1.

TABLE V
AVERAGE ABSOLUTE SYSTEM EXPENSE SAVINGS TO AVERAGE ES PROFIT INCREASE RATIO DUE TO REACTIVE POWER BIDS

Network	5_pjm	14_ieee	24_ieee_rts	30_as	30_fsr	30_ieee	39_epri	57_ieee
Ratio	1.06	1.59	1.48	1.53	1.57	1.44	1.36	1.50

As the dual lower-level objective function Ω^{d}, i.e. system expenses, contains the upper-level profit term with a negative sign, $-\sum_{t,i\in\beta}(p_t^{\mathrm{es}}\cdot\lambda_{1,t,i}+q_t^{\mathrm{es}}\cdot\lambda_{2,t,i})$, the upper-level profit increase normally results in system savings, thus the interests of the ES and the system generally align. In Figure 2 we see that the savings are of similar distribution as the profit increases from Figure 1. However, ES savings on average result in even grater system savings, as shown in Table V. The ratio of absolute savings and profit increase also depends on the network, however, it mostly ranges from 1.4 to 1.6. On rare cases, the ES reactive power bids can be counter productive for the system. Figure 1 shows that at three buses at 30_fsr network, the system expenses have increased. The magnitude of the increase is, however, too low to be of significance (note that Figure 2 contains absolute values and not percentages).

D. Case Study III: Solution Techniques Study

This case study evaluates effectiveness of all solution techniques from Section II on two networks. A broad analysis allowing for both the ES active and reative power bids is performed on a small 3_lmbd network consisting of only three buses to identify viable techniques. The results of this analysis are presented in Table VI. Only the select techniques are applied to the 24_ieee_rts network for ES at bus 3. We selected the 24 bus network as it is the first larger network from the benchmark library [9] that has quadratic generators bid curves

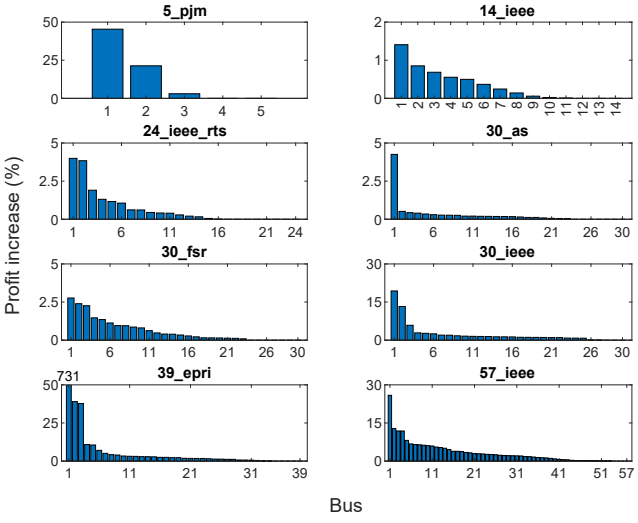

Fig. 1: Relative profit increase due to reactive power bids, in descending order.

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

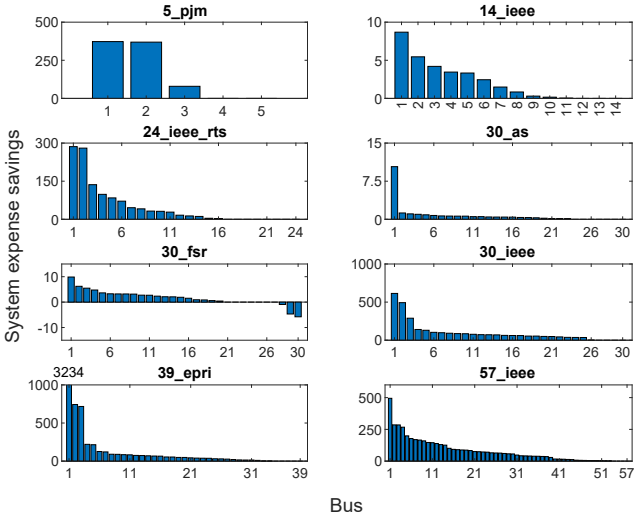

Fig. 2: Absolute system expense savings due to ES reactive power bids, in descending order.

and bus 3 since it is the first bus with non zero reactive power prices. The results of this analysis are displayed in Table VII and show the techniques performance when the ES bids only active power and both the active and reactive powers. All the simulations, except for those referring to the interaction discretization techniques, do not have binary variables to forbid simultaneous charging and discharging. Leaving out binary variables from optimization does not impact the solutions due to nonnegative energy prices. Finally, Table VIII presents an analysis for the most reduced and the best-performing set of solution techniques to solve the problem considering also the ES (dis)charging binary variables.

As displayed in Table VI, the primal-dual counterpart (PD) and its straightened variant (PD-S) are the easiest to compute since they belong to the convex SOCP optimization class. However, they leave large duality gaps and the straightened variant is only applicable when using the generators' quadratic bid curves. The best performing techniques (observing both the accuracy and numerical tractability) are SM1 with ϵ=1e-4, SM2 with ϵ=1e-4, CS-R with ϵ=0.01, PF-SD with π=100 and PF-CS with π=10. They converge in 0.1–0.2 seconds and require only 21–42 iterations, achieving small duality gaps of 8.6e-9%, 5.3e-9%, 1.2e-3%, 2.3e-5% and 6.3e-4%, respectively. Other defacto exact techniques are strong duality (SD), discretization techniques (BE and UE) and complementary slackness (CS and CS-A). However, a significant numerical tractability issues are observed. They manifest either as a large number of solver iterations to achieve convergence or an inability to adequately close the mixed-integer programming (MIP) gap (in the case of discretization techniques). Solvers using discretization techniques also fail to adequately set the best bound, with the strong duality version setting it to the solution of the PD technique and with the penalty factor version setting it to the effective infinity, which leaves large MIP gaps. With these techniques we used 32 discretization segments. The PF-SD technique with high penalty factor (π=300 and 1000) converges to a suboptimal solution as can occur with nonconvex formulations, while the complementary slackness

TABLE VI
TECHNIQUES COMPARISON ON 3_LMBD NETWORK

	ES at bus 3					
	ES profit			System expenses	Numerical tractability	
Technique	Actual	Computed	Diff [%]	Duality gap [%]	Time [s]	Iterations / MIP gap
PD	1988.07	2273.58	14	0.33	0.08	33
PD-S	2013.94	2094.55	4.0	8.6e-2	0.06	20
MC	1988.06	2273.58	14	0.33	0.09	35
SD	2016.84	2019.07	0.11	5.2E-13	2.0	440
SD-R ϵ=0.1	Converged to infeas. point				11	1734
SD-R ϵ=1	2016.75	2046.98	1.5	1.0E-03	2.3	396
SD-R ϵ=10	2016.18	2103.68	4.3	0.010	0.34	76
BE-SD	1484.62	1519.93	2.4	2.1E-04	1800	51.93%
UE-SD	No solution found				1800	-
BE-PF π=100	1250.53	1291.77	3.3	2.9E-05	1800	(2.8e+5)%
UE-PF π=100	-379.88	-361.00	-5.0	4.7E-05	1800	(1.2e+6)%
PF-SD π=10	2016.69	2038.54	1.1	1.9E-03	0.17	40
PF-SD π=30	2016.81	2025.13	0.41	2.4E-04	0.21	50
PF-SD π=100	2016.82	2020.00	0.16	2.3E-05	0.10	35
PF-SD π=300	1958.59	1958.50	-4.6e-3	7.8E-06	1.5	238
PF-SD π=1000	2015.15	2015.93	0.039	5.5E-07	2.4	334
PF-CS π=3	2016.72	2036.52	0.98	4.8E-03	0.21	40
PF-CS π=10	2016.81	2024.56	0.38	6.3E-04	0.16	42
PF-CS π=30	2016.83	2020.18	0.17	7.9E-05	0.65	134
PF-CS π=100	2016.84	2018.52	0.083	7.4E-06	0.36	77
PF-CS π=300	1915.59	1916.60	0.053	3.9E-06	0.55	109
CS	Converged to infeas. point				20	1152
CS-R ϵ=1e-4	2016.84	2018.97	0.11	1.4E-05	8.6	864
CS-R ϵ=1e-3	2016.84	2021.72	0.24	1.2E-04	0.26	55
CS-R ϵ=0.01	2016.83	2029.56	0.63	1.2E-03	0.12	28
CS-R ϵ=0.1	2016.73	2049.16	1.6	9.6E-03	0.12	28
CS-A	2016.84	2017.86	0.051	1.6E-12	5.2	974
CS-AR ϵ=1e-4	1798.02	1802.96	0.27	6.2E-05	0.79	172
CS-AR ϵ=1e-3	2016.81	2030.67	0.69	6.2E-04	0.82	186
CS-AR ϵ=0.01	2016.69	2054.46	1.9	6.2E-03	0.56	128
CS-AR ϵ=0.1	2015.52	2093.16	3.9	6.2E-02	0.07	17
SM1 ϵ=1e-4	2016.84	2017.78	0.047	8.6E-09	0.11	22
SM1 ϵ=1e-3	2016.84	2017.78	0.047	5.7E-07	0.43	69
SM1 ϵ=0.01	2016.84	2017.75	0.045	3.4E-05	0.14	32
SM1 ϵ=0.1	2016.85	2012.52	-0.21	5.7E-03	0.22	45
SM1 ϵ=1	2016.84	1487.22	-26	0.57	0.09	22
SM2 ϵ=1e-4	2016.84	2017.78	0.047	5.3E-09	0.12	21
SM2 ϵ=1e-3	2016.84	2017.78	0.047	4.9E-07	0.55	74
SM2 ϵ=0.01	2016.84	2017.74	0.045	4.6E-05	0.10	21
SM2 ϵ=0.1	2016.85	2012.52	-0.21	5.7E-03	0.13	26
SM2 ϵ=1	2016.84	1487.22	-26	0.57	0.12	27

technique fails to find any solution. Relaxing the techniques or reducing the penalty factor enhances the numerical tractability (for SD-R, PF-SD, PF-CS, CS-R and CS-AR), but also reduces accuracy. The McCormick technique (MC), which relaxes the bilinear terms in the strong duality, achieves the same solution as if there was no relaxed strong duality constraint which means that the relaxation is too strong to be useful. For MC we used fixed envelope bounds around the operating point of $\pm$ 1000 [1/p.u.] (i.e. 10 per MW) for active power and $\pm$ 300 [1/p.u.] (i.e. 3 per MW) for reactive power price.

The select versions of the primal-dual, strong duality, penalty factor, complementary slackness and the two smoothing techniques are tested on a larger network with results displayed in Table VII. The previously well performing techniques PF-SD and PF-CS (see Table VI) did not perform well. In the case of bidding both the active and reactive powers it took 62 and 85 second to compute them, which is about 6 times longer than for the two smoothing techniques and about 3 times longer when only active power bidding is allowed. CS-R also takes 6 times longer to compute than the smoothing techniques in case of bidding both the active and reactive powers, but it finishes slightly faster then the smoothing techniques when only active power bidding is allowed. It closes the duality gap only moderately well with 5.6e-4% in

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3207097

TABLE VII
COMPARISON OF SELECT TECHNIQUES ON 24_IEEE_RTS NETWORK

	ES at bus 3					
	ES profit			System expenses	Numerical tractability	
Technique	Actual	Computed	Diff [%]	Duality gap [%]	Time [s]	Iterations
ES active-power-only bids						
PD	4841.37	4921.32	1.7	4.3e-3	0.86	45
PD-S	4841.11	4902.15	1.3	3.3e-3	0.83	42
SD	-10452.73	-10444.28	-0.081	3.5e-11	199	2563
SD-R ϵ=10	4843.38	4901.84	1.2	8.1e-4	948	8102
PF-SD π=10	4848.66	4860.68	0.25	7.8e-5	36	139
PF-CS π=10	4848.81	4854.19	0.11	8.4e-5	14	134
CS	Converged to infeas. point				475	852
CS-R ϵ=0.1	4848.79	4880.99	0.66	9.3e-4	4.0	67
CS-A	Converged to infeas. point				400	3296
CS-AR ϵ=0.01	4841.10	4902.14	1.3	3.3e-3	2.8	46
SM1 ϵ=1e-4	4848.98	4848.88	-2.1e-3	1.2e-6	10	93
SM2 ϵ=1e-4	4848.89	4848.97	1.6e-3	4.2e-9	5.5	55
ES active and reactive power bids						
PD	4989.28	5271.89	5.7	2.2e-2	0.91	49
PD-S	4988.72	5251.51	5.3	2.1e-2	0.83	43
SD	Converged to infeas. point				367	4870
SD-R ϵ=10	5031.49	5118.21	1.7	8.1e-4	104	276
PF-SD π=10	5034.90	5056.21	0.42	1.5e-4	62	105
PF-CS π=10	5035.12	5051.37	0.32	1.2e-4	85	138
CS	Converged to infeas. point				682	138
CS-R ϵ=0.1	5034.79	5070.47	0.71	5.6e-4	72	743
CS-A	5035.14	5053.21	0.36	3.9e-10	599	5329
CS-AR ϵ=0.1	4988.77	5251.44	5.3	2.1e-2	40	180
SM1 ϵ=1e-4	5035.02	5035.54	0.010	5.1e-9	10	95
SM2 ϵ=1e-4	5035.02	5035.54	0.010	4.4e-9	13	112

TABLE VIII
PERFORMANCE IN DISCRETE OPTIMIZATION

		ES profit			System expenses	Numerical tractability	
Network	Technique	Actual	Computed	Diff [%]	Duality gap [%]	Time [s]	Nodes
ES active-power-only bids							
24_ieee_rts (bus 3)	SM1 ϵ=1e-4	4848.98	4850.73	0.036	1.4e-4	31	1
	SM2 ϵ=1e-4	4848.85	4848.97	2.5e-3	6.9e-9	15	1
ES active and reactive power bids							
3_lmbd (bus 3)	SM1 ϵ=1e-4	2016.89	2017.78	0.044	5.7e-9	1.3	3
	SM2 ϵ=1e-4	2016.89	2017.78	0.044	5.7e-9	0.19	1
24_ieee_rts (bus 3)	SM1 ϵ=1e-4	5035.02	5035.54	0.010	4.7e-9	93	3
	SM2 ϵ=1e-4	5035.05	5036.23	0.023	5.5e-5	71	3

first case and 9.3e-4% in the latter. CS-AR performs similar to CS-R, with CS-AR being somewhat faster, but also with larger duality gaps. The primal-dual counterpart techniques still offer the best tractability, but also low accuracy due to high duality gaps (ES profit errors are approximately 5% for active and reactive power bids and 2% for active-power-only bids). The two smoothing techniques perform reasonably tractable in both cases finishing in 10 seconds and taking 89 iterations on average. Meanwhile, they achieve close to zero (order of magnitude 1e-8%) duality gaps. The remaining techniques SD, CS and CS-A all either converge to an infeasible point or display serious numerical tractability issues.

This case study is completed by evaluating the performance of the two smoothing techniques with included binary variables that forbid simultaneous ES charging and discharging, as displayed in Table VIII. Both smoothing techniques achieve comparable and high accuracy. The displayed tractability of SM2 technique is marginally better.

IV. CONCLUSION

The approach presented in Part I and Part II papers avoids the lower level linearization and can be used to effectively solve a strategic energy storage bilevel transmission-network-constrained market participation problem. Both active and reactive power bids are considered. The model utilises a convex polar second-order Taylor approximation [8] of AC OPF in the lower level thanks to which the KKT-based single-level reduction is possible while achieving extremely high AC OPF accuracy. The resulting complementary conditions are transformed using the smoothing technique to achieve numerical tractability.

Results indicate very high and consistent model accuracy tested on eight meshed transmission system networks for ES placement at every bus. For active-power-only bids, the mean and median upper-level profit errors are mainly in the range 0.01%–1e-3% with the maximum observed error of 0.19% within 232 optimizations. However, this error is reduced to 4.0e-3% in the second iteration of the Algorithm 1. When both the active and reactive power bids are considered, the upper-level profit errors are slightly higher, but still very low and mainly in the range 0.10%–1e-3% with the maximum observed error 3.97% within 190 optimizations. Again, the high errors can be further reduced by iteratively running the algorithm. Already the second iteration reduces the 3.97% error down to 0.035%. The lower-level objective function mean and median errors (AC OPF errors) are mainly in range the 1e-3%–1e-6%.

Economical benefits of ES reactive power bids significantly depend on the network and ES bus placement. The highest ES profit increases are in between 26%–47%, while for the majority of cases up to 5%. ES profit increase normally also reduces the system expenses, but at a greater amount. The average ES profit increase and average system savings ratio mostly ranges from 1.4–1.6.

The smoothing techniques achieve close-to-zero duality gaps, i.e. in the range 1e-6%–1e-8%, while outperforming in terms of tractability all other classical KKT-based single-level duality gap closure-enforcing reduction techniques.

The presented approach is also applicable to the various upper-level problems, e.g. generator, load or aggregator bidding or investment problems. It is also applicable for bilevel reserve procurement problems considering reactive power. Finally, it should benefit the system operators to assess the effect of their network investments, e.g. lines or energy storage, on the social welfare. Thus, utilization of this tool may benefit the market operators to achieve a revenue adequate and a more complete and fair market design.

REFERENCES

[1]J. Fortuny-Amat and B. McCarl, "A Representation and Economic Interpretation of a Two-Level Programming Problem," *The Journal of the Operational Research Society*, vol. 32, no. 9, pp. 783–792, Sept. 1981.

[2]M. H. Zare, J.S. Borrero, B. Zeng and O.A. Prokopyev, "A note on linearized reformulations for a class of bilevel linear integer problems," *Annals of Operations Research*, vol. 272, no. 4, pp. 99—117, Nov. 2019.

[3]H. Hijazi, "Perspective Envelopes for Bilinear Functions". Accessed: Oct. 12 2021. [Online]. Available: http://www.optimization-online.org/DB_FILE/2015/03/4841.pdf.

[4]X. M. Hu and D. Ralph, "Convergence of a Penalty Method for Mathematical Programming with Complementarity Constraints," *Journal of Optimization Theory and Applications*, vol. 123, no. 2, pp. 365-–390, Nov. 2004.

[5]S. Dempe and A. Zemkoho, *Bilevel Optimization: Advances and Next Challenges*, Cham, Switzerland, Springer, 2020.

[6]A. Gupte, S. Ahmed, M. S. Cheon and S. Dey, "Solving Mixed Integer Bilinear Problems Using MILP Formulations," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 721–744, April. 2013.

[7]M. Fukushima, Z. Luo and P. Tseng, "Smoothing Functions for Second-Order-Cone Complementarity Problems," *SIAM Journal on Optimization*, vol. 12, no. 2, pp. 436–460, Jan. 2002.

[8]K. Šepetanc and H. Pandžić, "Convex Polar Second-Order Taylor Approximation of AC Power Flows: A Unit Commitment Study," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3585–3594, July 2021.

[9]C. Coffrin *et al. PGLib-OPF v19.05*, GitHub, May 13, 2019. Accessed on: March 23, 2021. [Online] Available at: github.com/power-grid-lib/pglib-opf/tree/v19.05

[10]C. Grigg *et al*., "The IEEE Reliability Test System-1996," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1010–1020, Aug. 1999.

[11]R. A. Jabr, "Radial distribution load flow using conic programming," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1458–1459, Aug. 2006.

[12]K. Pand žić, H. Pandžić and I. Kuzle, "Virtual storage plant offering strategy in the day-ahead electricity market," *International Journal of Electrical Power & Energy Systems*, vol. 104, pp. 401–413, Jan. 2019.

[13]A. Venzke, S. Chatzivasileiadis and D. K. Molzahn, "Inexact convex relaxations for AC optimal power flow: Towards AC feasibility," *Electric Power Systems Research*, vol. 187, Oct. 2020.

Solving Bilevel Optimal Bidding Problems Using Deep Convolutional Neural Networks

Domagoj Vlah, Karlo Šepetanc, *Student Member, IEEE*, and Hrvoje Pandžić, *Senior Member, IEEE*

***Abstract*—Current state-of-the-art solution techniques for solving bilevel optimization problems either assume strong problem regularity criteria or are computationally intractable. In this article, we address power system problems of bilevel structure, commonly arising after the deregulation of the power industry. Such problems are predominantly solved by converting the lower level problem into a set of equivalent constraints using the Karush–Kuhn–Tucker optimality conditions at an expense of binary variables. Furthermore, in case the lower level problem is nonconvex, the strong duality does not hold rendering the single-level reduction techniques inapplicable. To overcome this, we propose an effective numerical scheme based on bypassing the lower level completely using an approximation function that replicates the relevant lower level effect on the upper level. The approximation function is constructed by training a deep convolutional neural network. The numerical procedure is run iteratively to enhance the accuracy. As a case study, the proposed method is applied to a price-maker energy storage optimal bidding problem that considers an ac power flow-based market clearing in the lower level. The results indicate that greater actual profits are achieved as compared to the less accurate dc market representation.**

***Index Terms*—Bilevel optimization, deep convolutional neural network, optimal power flow.**

I. Introduction

A. Background and Article Scope

DEREGULATION and liberalization of the power sector worldwide dislodged large monopolistic power utilities, allowing for private companies to become important players in the sector. However, each of the newly created entities have their own goal, e.g., generating companies want to maximize their profit, system operators maximize the security of supply, while market operators maximize social welfare. Because of the increased number of players with conflicting objectives, they need to consider each other's goals and objective functions when optimizing their own utility. To accommodate the interaction between own and other player's actions, the researchers commonly resort to bilevel models, where own optimization problem [the upper level (UL) problem] is constrained by another optimization problem [the lower level (LL) problem]. This setting assumes that the lower level behavior is known, which is the case when considering the market operator conducting its market-clearing procedure or any other regulated entity that behaves according to some widely known rules.

As of October 5, 2021, the IEEE Xplore database [1] indicates that three flagship IEEE Power and Energy Society journals published 182 journals with word *bilevel* in the title (115 such papers in IEEE TRANSACTIONS ON POWER SYSTEMS, 45 in IEEE TRANSACTIONS ON SMART GRID and 22 in IEEE TRANSACTIONS ON SUSTAINABLE ENERGY). These papers cover a wide range of bilevel problems. Some of the most common topics include protection of a power system against a terrorist attack, e.g., [2], pricing schemes, e.g., [3], maintenance scheduling, e.g., [4], expansion planning, e.g., [5], or optimal bidding in one or more energy [6], or financial markets [7].

Although bilevel models have been used extensively in the literature, they often suffer from two drawbacks. The first one is related to linearization, as commonly one or more variables from the lower level problem appear multiply an upper level variable in the upper level objective function. Although in many cases this can be linearized using the strong duality theorem and some of the Karush–Kuhn–Tucker (KKT) optimality conditions [8], see e.g., [9], in some cases this is not possible. In such cases the authors commonly resort to the binary expansion method (see [6, Appendix B]). However, besides being an approximation, this method can result in intolerable computational times, bringing us to the second drawback, i.e., computational (in)tractability. Issues with tractability often arise when the lower level problem is stochastic or has many inequality constraints, resulting in a large number of binary variables. Some authors thus resort to an iterative procedure that considers the complicating dual variables in the problematic bilinear terms as parameters, and updating their values in the following iteration [10].

The aim of this article is to present a numerical scheme based on deep convolutional neural networks (NN) paired with state-of-the-art training procedure for solving complex bilevel problems arising in the power systems community. As a representative of such problems, we solve a bilevel problem of optimal participation of an energy storage in the day-ahead energy market. We assume an ac-optimal power flow (OPF)-based market clearing algorithm in the lower level. AC OPF is a challenging problem with numerous simplification attempts, e.g., by convexification [11]. However, to this date there is still no known exact finite convex ac OPF formulation that classical approaches could solve to optimality.

Manuscript received 5 March 2022; revised 5 August 2022 and 9 November 2022; accepted 22 December 2022. The work of K. Šepetanc at UNIZG-FER was supported by the Croatian Science Foundation under Grant DOK-2018-09. The work of Domagoj Vlah was supported by the Croatian Science Foundation (HRZZ) funded by European Social Fund through the Research Cooperability Program under Grant PZS-2019-02-3055. This work was supported by the European Unions Horizon 2020 Research and Innovation Programme under Grant 863876 through Project FLEXGRID and Grant 864298 through Project ATTEST. *(Corresponding author: Karlo Šepetanc.)*

The authors are with the Innovation Centre Nikola Tesla and University of Zagreb Faculty of Electrical Engineering and Computing, 10000 Zagreb, Croatia (e-mail: domagoj.vlah@fer.hr; karlo.sepetanc@fer.hr; hrvoje.pandzic@ieee.org).

Digital Object Identifier 10.1109/JSYST.2022.3232942

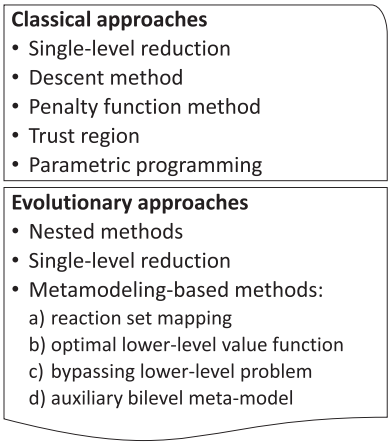

Fig. 1. Classification of the bilevel problem solution techniques.

Optimal bidding problems consist of two interlinked optimizations. The first problem, also called the leader or the upper level problem, represents the market participant that maximizes the agent's profit due to arbitrage, while the second problem, also called the follower or the lower level problem, is the market clearing that maximizes the social welfare and determines the electricity prices that depend on the bids from the upper level problem. The described bilevel optimization can not be directly solved using commercial off-the-shelf solvers, thus they are usually converted into a single-level equivalent optimization problem. However, such conversion is difficult since the exact ac OPF models, that appear in the lower level, are nonconvex and thus render many existing techniques inapplicable. The authors of this article have already explored the single-level reduction approach in a two-part paper [12] and [13], where ac OPF is modeled using a convex quadratically constrained quadratic approximation [11], but even the most computationally efficient solution techniques start diverging for large systems, i.e., systems with over 70 buses. Here, we explore a NN metamodeling approach that is known to require significant computational time and resources, but can compute for large systems and even allows for discrete variables in the lower level (at even greater computational cost). On the other hand, the KKT-based single-level reduction techniques solve the upper and the lower level simultaneously so the solution process can diverge for large systems.

Proving the global optimality of solution is not within the scope of this article. Generally, numerical optimization methods that can find global optimal solutions require much stronger conditions on the optimizing goal function and domain of optimization, for instance, in the case of optimizing a linear function on a convex domain. Our method is general in a sense that it imposes no mathematical conditions on the class of the lower level problem except that, using reasonable time and resources, the lower level can be evaluated a number of times to create a dataset for the NN training.

B. Literature Review

As depicted in Fig. 1, the existing literature on bilevel solution techniques branches out in two main directions: classical and evolutionary approaches. Due to computational difficulty of bilevel problems, the classical approaches can only tackle well-behaved problems with strong assumptions, such as linearity or convex quadraticity and continuity of the lower level problem, as strong duality generally does not hold for other types of problems. By far the most common classical approach is a single-level reduction based on the KKT conditions and the duality theory. It has been widely used to solve bilevel problems with linear constraints and either linear [14] or convex quadratic [15] objective functions, as well as problems with convex quadratic constraints when the interaction between the two levels is discrete [16]. The resulting formulations contain complementarity constraints which are combinatorial in their nature and thus can be modeled using binary variables making the final problems mixed-integer linear (MILP) or mixed-integer quadratic (MIQP). The existing state-of-the-art solvers generally handle well these types of optimizations using the branch-and-bound method for binary search tree and simplex for search-tree node subproblems, despite the exponential complexity in the worst case. The single-level reduction technique has also been successfully applied to a case where the lower level is a convex quadratically constrained quadratic problem (QCQP), as in [17]. Other classical approaches are the descent method, the penalty function method, the trust-region method, and the parametric programming method. The descent method determines the most favorable variable change for the objective function, as demonstrated in [18], so the model stays feasible. However, since the model is feasible only when the lower level is optimal, finding the descent direction is very difficult. The penalty function method replaces the lower level [19] or both-level [20] constraints with penalty terms for constraint violations in the objective function. The trust region algorithms iteratively approximate the lower level around the operating point with linear problem (LP) or quadratic problem (QP) [21]. Both the penalty and the trust-region methods as the next step apply a KKT-based single-level reduction to the lower level and thus inherit the same applicability limits. A recent research thrust in parametric programming has resulted in an alternative approach to solving bilevel programs to global optimality by exploiting the notion of *critical regions*. To this point, solution approaches based on parametric programming have been proposed to handle bilevel programs with LP [22], QP [23], MILP, and MIQP [24] lower levels.

As opposed to the classical ones, the evolutionary approaches are inspired by the biological evolution principle where candidate solutions are evaluated using a fitness function, e.g., objective function, to form the next generation of candidate solutions by reproducing, mutating, recombining, and selecting processes. Evolutionary approaches are very effective at finding good approximate solutions of numerically very difficult problems with fewer regularity assumptions than the classical approaches. For bilevel problems, the evolution is commonly applied in a nested form where the lower level needs to be solved separately for every upper level solution candidate, as explained and analyzed in [25]. The upper level solution candidates are obtained using an evolution, e.g., particle swarm optimization [26] or differential evolution [27], while the lower level can be solved using classical approaches such as interior point method [28] or as well using an evolution, as in [27]. Despite applicability to nonconvex problems, where classical approaches do not hold, the nested evolutionary method does not scale well with the number of upper level variables as they exponentially increase the number of lower level optimizations that need to be performed. A single-level reduction technique

can also be utilized in the context of evolutionary approaches where numerical evolution concept is applied to the reduced formulation. Due to KKT conditions, this technique inherits regularity assumptions of the classical approaches for the lower level problem, but allows a more irregular upper level. The work ofHejazi et al. [29] is one of the first works where this technique was employed. Evolutionary approaches can also be applied in tandem with the metamodeling method. Metamodel, or surrogate model, is an easy to evaluate, typically iteratively enhanceable, approximation of the original model. For bilevel modeling, the lower level can be metamodeled using the reaction set mapping, optimal lower level value function, by bypassing the lower level problem completely and by using an auxiliary bilevel metamodel. The reaction set method maps the lower level variable values as a response to the upper level variables, as demonstrated in [30]. On the other hand, the optimal lower level function method replaces the lower objective statement of minimization or maximization with a constraint requiring that the objective is at least as good as the optimal lower level function [31]. Both the reaction set map and the optimal lower level function are generally difficult to obtain even in an approximated form. Bypassing the lower level completely is based on the principle that the lower level variables are basically functions of the upper level variables, which allows for the upper level reformulation not to include the lower level. Similar to the trust-region method, bilevel problems can be replaced with auxiliary metamodels. As of current, we are not aware of any works based on bypassing the lower level problem or the auxiliary metamodels methods. A broader bilevel solution techniques research field review, for both the classical and the evolutionary approaches, can be found in [32].

The approach presented in this article can be classified as an evolutionary metamodeling method that bypasses the lower level problem completely, as given by [32]. This bypassing of the lower level problem is achieved by approximating the solution of the lower level, which depends only on the upper-level variables, using a carefully designed NN, see [33] and [34]. As a NN is simply a function composed of elementary functions, it can be substituted directly into the upper-level objective function. This way, the original bilevel optimization problem is reduced into a single-level optimization problem, which approximates the solution of the original problem. The main difficulty of our framework is the design and training of a NN that efficiently and accurately approximates the lower level problem.

C. Contribution

In this work, we develop a general numerical solution technique for bilevel problems and apply it to the energy storage (ES) bidding problem on an AC-OPF-constrained energy market. The technique is applicable to any other upper level subject, but we chose the ES due to modeling simplicity and clarity of presentation. The numerical and mathematical difficulty of solving the considered bilevel optimization arises from insufficient problem regularity due to nonconvexity of the exact ac OPF formulations. Current modeling practice is to avoid the difficulties by using a simpler linear dc OPF [35] network representation as in [36]. To the authors knowledge, there are very few attempts to solve bilevel problems with an ac OPF in the lower level. We have not found any with the exact ac OPF, thus we single out two papers with ac OPF relaxations, [16] and [17]. Scalability and tractability issues are not discussed in these papers which is also one of the important points of this work.

The contribution of this article consists of the following.

1) We introduce a novel numerical scheme for solving bilevel optimization problems based on deep convolutional NNs. It is an evolutionary metamodeling method that completely bypasses the lower level problem. Our method successfully works with previously intractable, i.e., nonconvex, classes of the lower level problems. As opposed to the existing techniques, solution times are basically independent on the upper level problem size and scale well with the lower level problem size.
2) We demonstrate the solution technique effectiveness by solving a price-maker energy storage AC-OPF-constrained market bidding problem. The results demonstrate higher achieved profits than with the dc market representation.

The article is organized as follows. Section II provides mathematical foundation of the work and is divided in six sections. Section II-A states the optimization problem, Section II-B explains how to approximate the lower level using a NN, Section II-C describes the concept of fully connected NNs, Section II-D explains the advantages, concept, and our choice of hyperparameters of the used convolutional NN, Section II-E describes the NN training algorithm and Section II-F explains our iterative numerical scheme to solve the optimization problem at hand. The case study is presented in Section III with implementation details stated in Section III-A and results in Section III-B. The final Section IV concludes the article.

II. Mathematical Modeling

A. Optimization Model

In the following model, we solve optimal ES bidding problem in the ac OPF network-constrained electricity market. The problem is of bilevel structure, i.e., the upper level maximizes the ES profit while the lower level maximizes social welfare due to supply and demand market bids. In the lower level, we consider an exact nonconvex quadratic ac OPF formulation based on rectangular coordinates [37] notation written out in the Appendix, however, other notations such as polar [37] or current–voltage [38] are also applicable.

The upper level problem consists of objective function (1), where λ_t is the electricity price in each hour indexed by t, and p_t^{ES} is the average ES power during 1 h, i.e., energy, at the interface. Constraint (2) models the ES (dis)charging process, i.e., change in its state-of-energy SoE_t considering charging and discharging efficiencies $\boldsymbol{\eta}^{\mathbf{ch}}$ and $\boldsymbol{\eta}^{\mathbf{dis}}$. Constraint (3) sets limits to the state-of-energy (SoE), with $\overline{\boldsymbol{SoE}}$ being the maximum value. Constraints (4) and (5) limit the ES (dis)charged energy to $\overline{\boldsymbol{q}}^{\mathbf{ch}}$ for charging and $\overline{\boldsymbol{q}}^{\mathbf{dis}}$ for discharging. Binary variable x_t^{ch} disables simultaneous charging and discharging. Finally, (6) combines charging and discharging into a single variable p_t^{ES}. Optimization variables are written in formulas in normal font and contained in the variables set Ξ, while the parameters are written in bold font.

$$\underset{\Xi}{\mathrm{Max}} \quad -\sum_t p_t^{\mathrm{ES}} \times \lambda_t \tag{1}$$

$$SoE_t = SoE_{t-1} + p_t^{\mathrm{ch}} \times \boldsymbol{\eta}^{\mathbf{ch}} - p_t^{\mathrm{dis}} / \boldsymbol{\eta}^{\mathbf{dis}} \qquad \forall t \tag{2}$$

$$0 \leq SoE_t \leq \overline{\boldsymbol{SoE}}, \quad \forall t \tag{3}$$

$$0 \leq p_t^{\text{ch}} \leq \overline{\boldsymbol{q}}^{\mathbf{ch}} \times x_t^{\text{ch}} \qquad \forall t \tag{4}$$

$$0 \leq p_t^{\text{dis}} \leq \overline{\boldsymbol{q}}^{\mathbf{dis}} \times (1 - x_t^{\text{ch}}) \qquad \forall t \tag{5}$$

$$p_t^{\text{ES}} = p_t^{\text{ch}} - p_t^{\text{dis}} \qquad \forall t \tag{6}$$

The lower level is only textually explained and not written here since we are bypassing it completely. The rectangular ac OPF consists of the objective function, the bus power balance constraints, the power flow equations, the line apparent power limits, the bus voltage limits, the generator production limits, and the reference bus constraints. Mathematically challenging are the power flow equations and the lower bus voltage limit constraints, which do not conform to the traditional single-level reduction technique as they are nonconvex. Moreover, there are two additional convex, but nonlinear parts of the formulation. The considered objective function has quadratic cost coefficients and line apparent power limit constraints are of second-order cone form. Broader insights of different ac OPF formulations can be found in tutorial works such as [37].

Bypassing the lower level is based on the fact that the locational marginal prices λ_t are essentially a function of the upper level p_t^{ES} variables, i.e., the objective can be expressed as $F(p_1^{\text{ES}}, p_2^{\text{ES}}, \ldots, p_{|\tau|}^{\text{ES}})$, where $|\tau|$ is cardinality of the time steps set τ. However, F can not be expressed explicitly, so we replace the objective function F with the approximation $\hat{F}$ from (7). The approximating function $\hat{F}$ is given as the feed-forward NN, so it can be expressed explicitly in terms of elementary mathematical functions. Essentially, we are solving a single-level optimization metamodel that maximizes (7) subject to constraints (2)–(6).

$$\underset{\Xi}{\text{Max}} \ \hat{F}\left(p_1^{\text{ES}}, p_2^{\text{ES}}, \ldots, p_{|\tau|}^{\text{ES}}\right) \tag{7}$$

The problem belongs to the mixed-integer nonlinear optimization class due to the nonlinear NN function $\hat{F}$ and due to x_t^{ch} being binary variables.

B. Lower Level Approximation Using Neural Networks

For any function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, there exists a NN that uniformly approximates the given function, see [39] and [40]. Typically, it is unknown how exactly to construct a specific NN, approximating the function f to the desired accuracy, and using the smallest possible number of neurons. The first problem we encountered is the limited size of dataset used to train such NN. More precisely, each element in the dataset must be constructed by solving a single instance of the lower level optimization problem for chosen values of the upper level variables. Solving too many instances of the lower level problem would take too long. On the other hand, the size of the dataset limits the maximum network size (the number of neurons), by limiting the number of parameters that define that particular network. NNs trained on a dataset that is small compared to the number of network parameters tend to overfit the training data and are poor in generalization on unseen data. In our case that would lead to lower accuracy of approximation of the lower level problem solutions. Basically, the size of the dataset limits the accuracy of the NN approximation.

The second issue is in determining an optimal topology of a NN for a given network size, in order to achieve the greatest possible approximation accuracy. The optimal topology is dependent on an unknown function, which we are trying to approximate. Our first approach, using fully connected NN with only few hidden layers, led to poor approximation accuracy. By carefully analyzing the properties of the lower level optimization problem, the choice of a network topology was settled on a convolutional neural network (CNN) [41]. As the CNN architecture shares the same values of parameters between different parts of the network, the cumulative number of parameters is much smaller for the network of the same size, so the CNN architecture can be trained to approximate the original optimization problem to a higher accuracy. The first big success of the CNN architecture was in the area of computer vision, in the image classification problems [42].

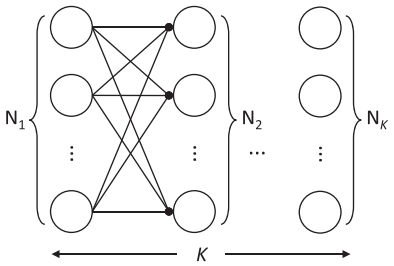

Fig. 2. Example of a fully connected NN.

The third obstacle we encountered was the generation of a dataset for the CNN training. Our first idea was to generate the dataset by uniform random sampling of the independent upper level variables, only in intervals of their permissible values. Then for each sample, we solved the associated lower level optimization problem. This strategy proved to be inefficient as the near-optimal values of the upper level variables, which solve our bilevel optimization problem, are poorly represented by sampling these variables independently from the uniform distribution. It resulted in much higher approximation error of the CNN on the optimal solution than on the generated dataset. The solution proved to be in iterative refining of the generated dataset. In the first iteration, we generate a uniform dataset on the whole permissible domain and find the solution of the approximation for the bilevel optimization problem. In each additional iteration, we restrict the domain to an even smaller neighborhood around the approximated solution from the previous iteration. Then, we generate a new uniform dataset on this smaller domain, train a new instance of the CNN, and using this new trained network, we again find a solution of the approximating problem. In each iteration, we verify the quality of the current solution by computing the upper level objective function exactly on optimal variables approximate problem values. We stop iterating when the actual value of the upper-level objective function stops improving.

C. Feed-Forward Fully Connected NNs

A feed-forward fully connected NN (see Fig. 2) consists of K layers, where each layer consists of a number of neurons [43]. The first layer is referred to as the input layer, the last layer as the output layer, while the intermediate layers are called hidden layers. Neurons in each layer are connected only to the neurons in the neighboring layers. Feed-forward means that the data flows from the input layer to the output layer, strictly from one layer to the next one and in only one direction. Fully connected means that each neuron is connected to every neuron in the

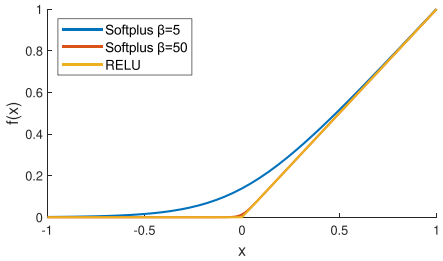

Fig. 3. Softplus and RELU plot.

neighboring layers. Finally, each neuron in every hidden layer performs a nonlinear transformation on the data by applying the so-called activation function. More precisely, a feed-forward fully connected NN is a function $\mathcal{F} : \mathbb{R}^{N_1} \to \mathbb{R}^{N_K}$, where N_1 is the number of neurons in the input layer, and N_K is the number of neurons in the output layer. Function $\mathcal{F}$ is a composition of the alternating affine maps $\mathcal{A}_k : \mathbb{R}^{N_{k-1}} \to \mathbb{R}^{N_k}$, $k = 2, \ldots, K$ and the elementwise nonlinear activation functions $\mathcal{N}_k : \mathbb{R}^{N_k} \to \mathbb{R}^{N_k}$, $k = 2, \ldots, K-1$, such that

$$\mathcal{F} = \mathcal{A}_K \circ \mathcal{N}_{K-1} \circ \mathcal{A}_{K-1} \circ \cdots \circ \mathcal{N}_3 \circ \mathcal{A}_3 \circ \mathcal{N}_2 \circ \mathcal{A}_2$$

where N_k is the number of neurons in the kth layer.

Affine map $\mathcal{A}_k$ can be written in a matrix form as

$$\hat{\mathbf{z}}_k := \mathcal{A}_k(\mathbf{z}_{k-1}) = \mathbf{W}_k \mathbf{z}_{k-1} + \mathbf{b}_k \quad \forall k = 1, \ldots, K$$

where weight matrix $\mathbf{W}_k$ has dimension $N_k \times N_{k-1}$ and bias vector $\mathbf{b}_k$ has dimension N_k.

For the activation functions, we elementwise use the Softplus function

$$z_k^i := \mathcal{N}_k^i(\hat{\mathbf{z}}_k) = \frac{\ln(1 + \exp(\beta \times \hat{z}_k^i))}{\beta} \quad \begin{matrix} \forall k = 1, \ldots, K \\ \forall i = 1, \ldots, N_k \end{matrix}$$

where $\hat{z}_k^i$ is the ith component of vector $\hat{\mathbf{z}}_k$ and β is the hyperparameter of the Softplus function. Notice that for large values of β, Softplus uniformly converges to a rectified linear unit (ReLU) activation function (see Fig. 3), which is given elementwise by

$$z_k^i = \max(\hat{z}_k^i, 0) \quad \forall k = 1, \ldots, K \quad \forall i = 1, \ldots, N_k.$$

ReLU activation function is commonly used in recent NN applications. The reason why we decided to use Softplus will be become clear in Section II-F.

D. Convolutional Neural Networks

A CNN can be regarded as a subtype of a feedforward NN. It is generally not fully connected, and a large number of weight and bias elements of matrices $\mathbf{W}_k$ and vectors $\mathbf{b}_k$ share the same values, as affine maps $\mathcal{A}_k$ are defined using the operation of matrix convolution [41].

Fig. 4 depicts a deep CNN, describing the exact NN topology used in approximating function $\hat{F}$ of our problem. The structure of the NN is determined by an educated guess of the authors and by experimentation. Besides the input and output layers, we have six additional hidden layers. Unlike in a general NN, each layer in our CNN is described using the layer length L_k and the number of channels C_k. The number of neurons in each layer is given by $N_k = L_k \times C_k$ and neurons are grouped in L_k groups of size C_k. In Fig. 4, a single square depicts one group of neurons. The exact number of neurons within each group (the number of channels C_k) is written inside each square. The number of groups in each layer is given by a number written just below each layer. For instance, the total number of neurons in the second layer is equal to $24 \times 32 = 768$.

To define affine map $\mathcal{A}_k$, each layer in a CNN has an additional integer hyperparameter called the kernel size S_k. In Fig. 4, only the first (input) layer and the last of the hidden layers have kernel sizes $S_1 = 1$ and $S_7 = 1$. All the other hidden layers have kernel size equal to 3. Notice that the kernel size is not applicable to the output layer, as the output layer only collects the output of the last hidden layer and is not applying any further affine maps. All kernels are depicted by a number of empty squares equal to the kernel size S_k. A downward arrow indicates that a kernel window is sliding over the layer in steps, performing a computation of the affine map. This means that the convolution operation can be in each step regarded as a smaller affine map $\hat{\mathcal{A}}_k$ that is defined only between the neurons in the groups covered by the kernel window in layer $k-1$ and the neurons in the single output group in layer k. In each step of the convolution operation on layer $k-1$ we use the same map $\hat{\mathcal{A}}_k$.

Each convolution layer has two additional integer hyperparameters called a stride and a padding size. The stride is the number of groups by which each kernel window moves in every step of the computing convolution operation. The first and second layers have the stride equal to 1 and the third to fifth layers have the stride equal to 2. This is the reason why the lengths L_k of the fourth to sixth layer are decreasing by a factor of 2. For the sixth and seventh layers, the stride is not applicable as the convolution operation is trivially performed only in the single possible position. The padding controls whether the kernel window can slide over the side of the layer or not. If we let the kernel windows slide over the side of the layer, as for the second to fifth layer, the padding is equal to 1 and we substitute zeros for the input in the convolution operation in place of the nonexisting data. For the first, sixth, and seventh layer, the padding is equal to 0, which means we do not let the kernel window slide over the side of the layer.

In matrix representation $\mathbf{W}_k$ of affine map $\mathcal{A}_k$, lot of matrix components are equal to zero and lot of other nonzero matrix components share the same values. We actually have, for the kernel size equal to 1 the block diagonal matrix $\mathbf{W}_k$, and for the kernel size equal to 3 the block tridiagonal matrix $\mathbf{W}_k$, see Fig. 5. Every block is of size $C_k \times C_{k-1}$. For the kernel size equal to 1, every block in the block diagonal matrix $\mathbf{W}_k$ representing affine map $\mathcal{A}_k$ is exactly the same block. For the kernel size equal to 3, every 3 vertical blocks in the block tridiagonal matrix representation are exactly the same blocks. The bias vector $\mathbf{b}_k$ also has repeating components. Regardless on the kernel size, components of $\mathbf{b}_k$ repeat every C_k entries, which means each neuron group in a single layer shares the same biases.

Notice that a CNN, for the same number of neurons, typically has much lower number of parameters defining affine maps $\mathcal{A}_k$, than a fully connected NN. For instance, in our case the number of parameters defining map $\mathcal{A}_3$ is $C_2 S_2 C_3 + C_3 = 3104$, whether the number of parameters defining map $\mathcal{A}_3$ in a fully connected NN with the same number of neurons would be $N_2 N_3 + N_3 = 590592$.

E. Training Feed-Forward NNs

To train a NN simply means to optimize matrices $\mathbf{W}_k$ and vectors $\mathbf{b}_k$ in order to minimize a chosen loss function over a

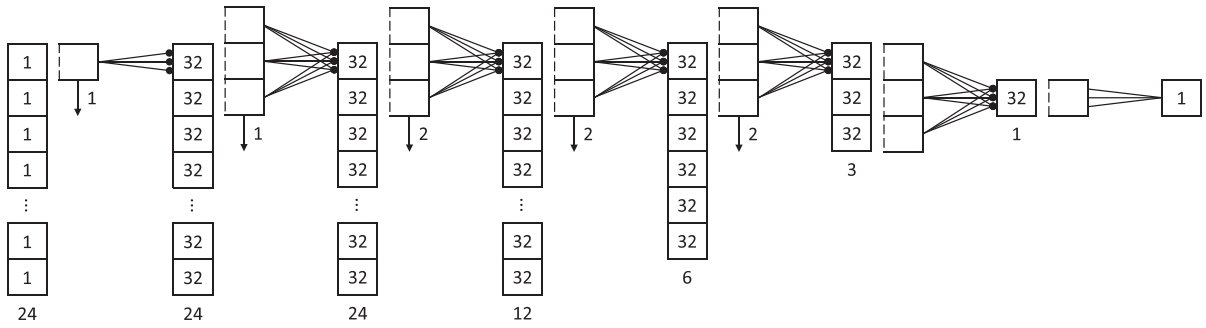

Fig. 4. Deep convolutional NN structure.

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n \end{pmatrix} \qquad \begin{pmatrix} B_1 & C_1 & 0 & \cdots & 0 \\ A_2 & B_2 & C_2 & & \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & A_{n-1} & B_{n-1} & C_{n-1} \\ 0 & \cdots & 0 & A_n & B_n \end{pmatrix}$$

(a) (b)

Fig. 5. Block matrices. (a) Diagonal. (b) Tridiagonal.

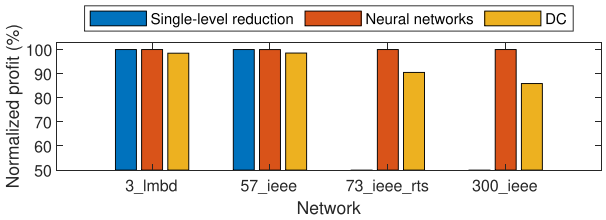

Fig. 6. Normalized profits with different approaches over four different transmission system meshed networks.

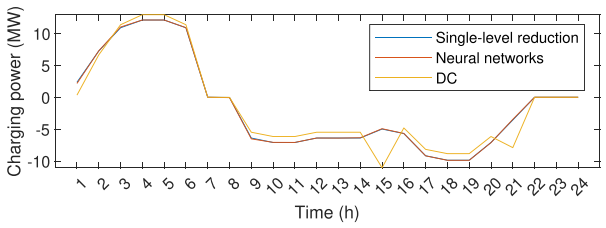

Fig. 7. Charging profile for 3_lmbd network at bus 3.

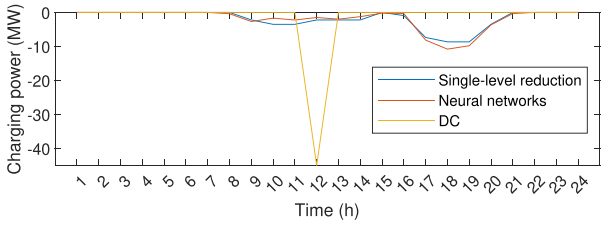

Fig. 8. Charging profile for 57_ieee network at bus 1.

dataset. Since we use CNNs, we must respect the block diagonal or tridiagonal structure of matrices $\mathbf{W}_k$ having many shared values, as discussed in Section II-D.

The dataset is generated by solving a number of instances of the lower level optimization problem, for as many random samples of vector $(p_1^{\mathrm{ES}}, p_2^{\mathrm{ES}}, \ldots, p_{|\tau|}^{\mathrm{ES}})$, and subsequently evaluating value of the target objective function (1), i.e., $F(p_1^{\mathrm{ES}}, p_2^{\mathrm{ES}}, \ldots, p_{|\tau|}^{\mathrm{ES}})$.

The loss function is the mean squared error between the NN computed value $\hat{F}(p_1^{\mathrm{ES}}, p_2^{\mathrm{ES}}, \ldots, p_{|\tau|}^{\mathrm{ES}})$ and the lower level exact solution $F(p_1^{\mathrm{ES}}, p_2^{\mathrm{ES}}, \ldots, p_{|\tau|}^{\mathrm{ES}})$.

As customary in NN training, the optimization of $\mathbf{W}_k$ and $\mathbf{b}_k$ is performed using a variant of a gradient descent optimizer. The dataset is split into training and validation datasets, using $80:20$ percent split ratio. The optimization is done in multiple epochs over the training dataset, where gradients are computed using backpropagation [44] and automatic differentiation [45] of NN. The validation dataset is used only for evaluating the loss function after every epoch of training and not for a gradient computation. Computed values of the loss function on the validation dataset are used to asses numerical viability of the training process and to select the best trained network, having the lowest value of the validation loss. Before the start of the first epoch of training, $\mathbf{W}_k$ and $\mathbf{b}_k$ are initialized to random values.

More details about our exact training procedure, together with all optimizer and training hyperparameter values are given in Section III-A.

F. Metaoptimization Numerical Scheme

We devised an iterative numerical scheme for computing a sequence of CNN approximations $\hat{F}_i$ of the otherwise intractable objective function F. In each iteration i, we first generate dataset $\mathcal{D}_i$ of random sampled vectors $(p_{1,i}^{\mathrm{ES}}, p_{2,i}^{\mathrm{ES}}, \ldots, p_{|\tau|,i}^{\mathrm{ES}})$. For each $t \in \tau$, values of $p_{t,i}^{\mathrm{ES}}$ are independently and uniformly random sampled from the interval centered around $p_{t,i}^{\mathrm{ES,cnt}}$ of length at most $2p_{t,i}^{\mathrm{ES,rad}}$, respecting the condition $-p_t^{\mathrm{dis}} \leq p_{t,i}^{\mathrm{ES}} \leq p_t^{\mathrm{ch}}$. In the first iteration, we set $p_{t,i}^{\mathrm{ES,cnt}}$ equal to zero and $p_{t,i}^{\mathrm{ES,rad}}$ such that it allows all permissible values of p_t^{ES}, i.e., $-p_t^{\mathrm{dis}} \leq p_{t,i}^{\mathrm{ES}} \leq p_t^{\mathrm{ch}}$.

In practice, we noticed that our numerical scheme runs better if we introduce an additional small relative tolerance $\epsilon > 0$ on the dataset creation. Thus, for the maximum length of the sampling interval we actually use $2p_{t,i}^{\mathrm{ES,rad}}(1+\epsilon)$ and allow all permissible values of p_t^{ES} to be from the interval $-p_t^{\mathrm{dis}}(1+\epsilon) \leq p_{t,i}^{\mathrm{ES}} \leq p_t^{\mathrm{ch}}(1+\epsilon)$. The intuition behind introducing the tolerance is that our CNN would better approximate objective function F for parameter values $p_{t,i}^{\mathrm{ES}}$ near the boundary values $-p_t^{\mathrm{dis}}$ and p_t^{ch}, if the dataset is allowed to include values $p_{t,i}^{\mathrm{ES}}$ a bit outside the interval $[-p_t^{\mathrm{dis}}, p_t^{\mathrm{ch}}]$.

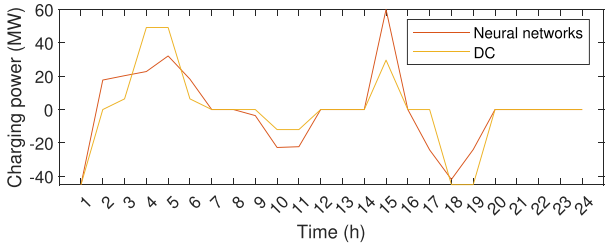

Fig. 9. Charging profile for 73_ieee_rts network at bus 101.

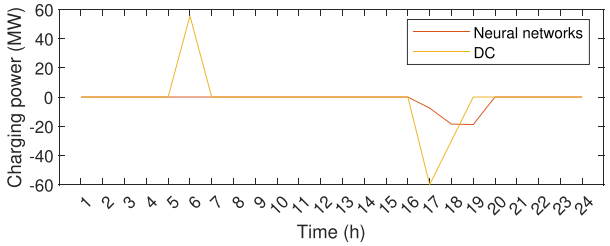

Fig. 10. Charging profile for 300_ieee network at bus 1.

Each element in dataset $\mathcal{D}_i$ is now a pair of random vector $(p^{\text{ES}}_{1,i}, p^{\text{ES}}_{2,i}, \ldots, p^{\text{ES}}_{|\tau|,i})$ and a value $F(p^{\text{ES}}_{1,i}, p^{\text{ES}}_{2,i}, \ldots, p^{\text{ES}}_{|\tau|,i})$ of the target objective function (1), computed by solving a single instance of the lower level problem.

Next, in each iteration we separately train a number of CNNs, $\hat{F}_{i,n}$ indexed by $n \in \{1, \ldots, M\}$, to approximate function F. For every additional training on the same dataset $\mathcal{D}_i$, we obtain a subtly different CNN, as the training process is intrinsically stochastic (random initialized network weights and random sampled stochastic gradient descent mini batches).

Now, as every trained CNN is a function $\hat{F}_{i,n}$, we symbolically insert function $\hat{F}_{i,n}$ into the upper level problem objective function. For this, we use our modeling environment's, which is AMPL, defining variables feature. Defining variables are a type of variables that are substituted out, potentially in a nested way, by their declaration expression before reaching the solver. This results in solving a single-level optimization metamodel which maximizes

$$\underset{\Xi}{\text{Max}} \ \hat{F}_{i,n}\left(p^{\text{ES}}_{1,i}, p^{\text{ES}}_{2,i}, \ldots, p^{\text{ES}}_{|\tau|,i}\right) \tag{8}$$

subject to constraints (2)–(3) and

$$0 \le p^{\text{ch}}_{t,i} \le \overline{\boldsymbol{q}}^{\mathbf{ch}} \times x^{\text{ch}}_{t,i} \qquad \forall t \tag{9}$$

$$0 \le p^{\text{dis}}_{t,i} \le \overline{\boldsymbol{q}}^{\mathbf{dis}} \times (1 - x^{\text{ch}}_{t,i}) \qquad \forall t \tag{10}$$

$$p^{\text{ES}}_{t,i} = p^{\text{ch}}_{t,i} - p^{\text{dis}}_{t,i} \qquad \forall t \tag{11}$$

$$p^{\text{ES,cnt}}_{t,i} - p^{\text{ES,rad}}_{t,i} \le p^{\text{ES}}_{t,i} \le p^{\text{ES,cnt}}_{t,i} + p^{\text{ES,rad}}_{t,i} \qquad \forall t \tag{12}$$

for every trained CNN indexed by n. The additional constraint (12) is used to respect that dataset $\mathcal{D}_i$ is created centered around $p^{\text{ES,cnt}}_{t,i}$ using an interval length at most $2p^{\text{ES,rad}}_{t,i}$.

Notice, as we used the differentiable Softplus activation function in our CNN, and as a CNN is just a composition of affine maps and elementwise activation functions, that $\hat{F}_{i,n}$ are differentiable functions. In case of using a ReLU activation function, the resulting $\hat{F}_{i,n}$ would not be differentiable. We experimentally established that lower values of the Softplus hyperparameter β result in lower overall NN approximation accuracy, and higher values give rise to solver instabilities in the single-level metamodel optimization step. We also tried to use ReLU instead of Softplus, but we were plagued with solver instabilities and slowdown. Using Softplus showed to be much more efficient.

For every trained CNN, we now have the computed profit $\mathcal{C}_{i,n} = \underset{\Xi}{\text{Max}} \ \hat{F}_{i,n}(p^{\text{ES}}_{t,i})$ and the computed optimal ES (dis)charged energy $(p^{\text{ES}}_{1,i,n}, p^{\text{ES}}_{2,i,n}, \ldots, p^{\text{ES}}_{|\tau|,i,n})$. Now we can verify what are the actual profits obtained for the computed optimal ES (dis)charging schedule. This is done by optimizing the lower level problem independently of the upper level with fixed ES (dis)charging schedule. The actual profit is determined as $\mathcal{V}_{i,n} = \underset{\Xi}{\text{Max}} \ -\sum_t p^{\text{ES}}_{t,i,n} \times \lambda_t$, where λ_t is in this case the bus balance constraint marginal, computed by default by many interior point solvers.

Additionally, we compute mean optimal ES energy exchange quantities

$$\overline{p^{\text{ES}}}_{t,i} = \frac{1}{M} \sum_{n=1}^{M} p^{\text{ES}}_{t,i,n} \quad \forall t \in \tau$$

and for the computed mean vector $(\overline{p^{\text{ES}}_{1,i}}, \overline{p^{\text{ES}}_{2,i}}, \ldots, \overline{p^{\text{ES}}_{|\tau|,i}})$ we again optimize the lower level problem with fixed ES charging values to the obtained mean optimal vector, arriving at the mean actual profit $\overline{\mathcal{V}_i} = -\sum_t \overline{p^{\text{ES}}}_{t,i} \times \lambda_t$. Considering the mean actual profit is justified, because averaging over optimal solutions of many different CNN models $\hat{F}_{i,n}$, each one approximating an intractable objective function F, it can result in a better mean solution. By looking at the actual test results (i.e., Tables III and IV), we see that this approach is justified in practice, i.e., in some iterations the mean actual profit $\overline{\mathcal{V}_i}$ can be higher than any of the actual profits $\mathcal{V}_{i,n}$.

Finally, we have to choose values of $p^{\text{ES,cnt}}_{t,i+1}$ and $p^{\text{ES,rad}}_{t,i+1}$ for a next iteration of the metaoptimization scheme. For $p^{\text{ES,cnt}}_{t,i+1}$, we either chose the optimal ES (dis)charging quantities $(p^{\text{ES}}_{1,i,n}, p^{\text{ES}}_{2,i,n}, \ldots, p^{\text{ES}}_{|\tau|,i,n})$ from CNN that achieved the highest actual profit $\mathcal{V}_{i,n}$, or in the case $\overline{\mathcal{V}_i}$ is the highest profit, we chose the mean optimal ES (dis)charging quantities $(\overline{p^{\text{ES}}_{1,i}}, \overline{p^{\text{ES}}_{2,i}}, \ldots, \overline{p^{\text{ES}}_{|\tau|,i}})$.

To choose $p^{\text{ES,rad}}_{t,i+1}$, we first compute the maximum over all $t \in \tau$ of the standard deviations of samples $\{p^{\text{ES}}_{t,i,n} : 1 \le n \le M\}$. More precisely, we compute

$$\sigma_i := \underset{t \in \tau}{\text{Max}} \ \text{std}\left\{p^{\text{ES}}_{t,i,n} : 1 \le n \le M\right\}.$$

For the next iteration, we take a smaller value between the current $p^{\text{ES,rad}}_{t,i}$ and a new estimate

$$p^{\text{ES,rad}}_{t,i+1} := \text{Min}\{p^{\text{ES,rad}}_{t,i}, \ \gamma \times \sigma_i\}$$

for every $t \in \tau$, where γ is a hyperparameter. Note that $p^{\text{ES,rad}}_{t,i+1}$ does not depend on t. It has the same value for every t.

For the next iteration of our metaoptimization scheme, dataset D_{i+1} is created around $p^{\text{ES,cnt}}_{t,i+1}$, which is the best optimal ES (dis)charging schedule computed in the current iteration, i.e., ES charging and discharging energy quantities that produce the highest profit. The dataset width, which is decided by $p^{\text{ES,rad}}_{t,i+1}$,

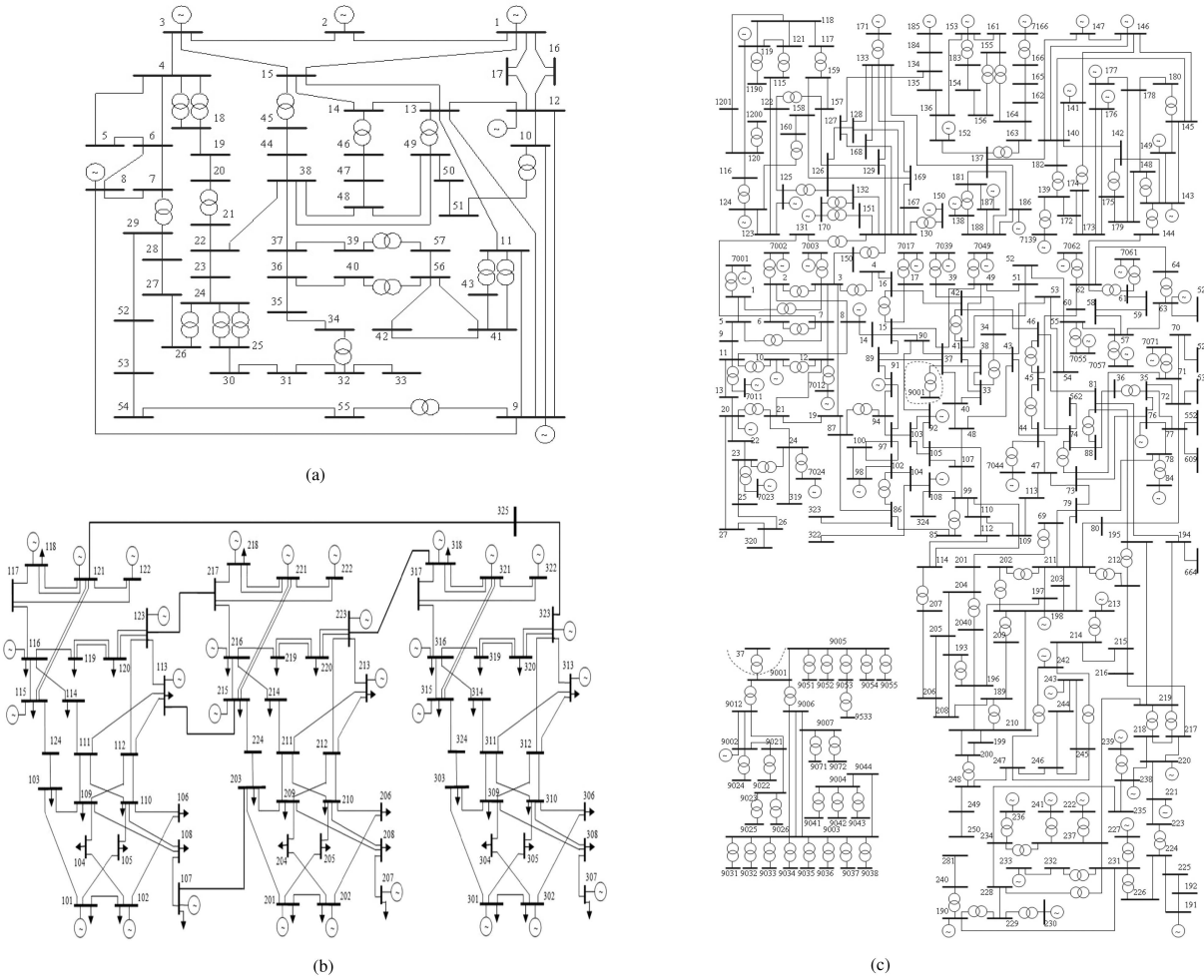

Fig. 11. Network topologies. (a) 57 bus system [59]. (b) 73 bus system [58]. (c) 300 bus system [59].

is influenced by how close together are optimal ES schedules predicted by different CNNs trained in the current iteration. In case different CNNs produce relatively close optimums, the computed standard deviation σ_i is relatively small, and a next iteration dataset is going to be concentrated around a smaller neighborhood of $p_{t,i+1}^{\mathrm{ES,cnt}}$. On the contrary, in case different CNNs produce optimums that are more apart, the computed standard deviation σ_i is relatively large, and a next iteration dataset is going to span over a bigger neighborhood of $p_{t,i+1}^{\mathrm{ES,cnt}}$. Notice that $p_{t,i}^{\mathrm{ES,rad}}$ is nonincreasing between iterations.

In the end, we have to prescribe a stopping criterion for our metaoptimization scheme. We choose to stop further iterations if there is no improvement in the actual profit of the current iteration compared to the previous one. We empirically conclude (see Section III-B) that the convergence of our metaoptimization scheme is achieved in few iterations (see Tables III–V). An overview of the numerical optimization scheme is provided in Algorithm 1.

Algorithm 1: Numerical Optimization Scheme.

1: **repeat**
2: Generate a new random dataset (10^5 entries)
3: Evaluate LL response for the dataset
4: Train 60 NNs to approximate LL response
5: Optimize the ULs with inserted NNs into objective function
6: Determine actual profits by optimizing LL with fixed ES (dis)charging schedule
7: Select the best actual solution out of:
- the best direct result;
- the result obtained averaging decisions from all optimized NNs;

8: For the next iteration, reduce and concentrate the dataset spatial size in the neighborhood of the best solution found from this iteration
9: **until** The best solution is worse than in the preceding iteration

III. Case Study

A. Implementation Details

For the dataset creation, each instance of the lower level problem, one for every dataset entry, was solved using AMPL running KNITRO 12.3 solver. A single dataset entry consists of a 24-D floating point vector $(p_1^{\mathrm{ES}}, p_2^{\mathrm{ES}}, \ldots, p_{|\tau|}^{\mathrm{ES}})$ as an input and a single floating point value as an output, which corresponds to the computed upper-level profit for given $(p_1^{\mathrm{ES}}, p_2^{\mathrm{ES}}, \ldots, p_{|\tau|}^{\mathrm{ES}})$

TABLE I
VALUES OF $p_{t,i}^{\mathrm{ES,rad}}$ FOR DIFFERENT TRANSMISSION NETWORKS AND ITERATIONS

Iter	3_lmbd	57_ieee	73_ieee_rts	300_ieee
1	0.6	0.6	0.6	0.6
2	0.0618	0.3265	0.4894	0.2302
3	0.0319	0.2123	0.3465	
4	0.0129	0.1876	0.3465	
5	0.0129			
6	0.0129			

TABLE II
AVERAGE PROCESSING TIME IN SECONDS PER SINGLE ITERATION OF THE METAOPTIMIZATION SCHEME

	Dataset creation [s]	NN training [s]	Solving meta-models (mean $\pm$ std) [s]	Component total time [s]
3_lmbd	471	5118	$60\times(7.7\pm10.0)$	6051
57_ieee	1774	5129	$60\times(11.7\pm10.6)$	7605
73_ieee_rts	3644	5125	$60\times(4.0\pm2.6)$	9009
300_ieee	28256	5245	$60\times(2.0\pm1.9)$	33621

values (computed as $-\sum_t p_t^{\mathrm{ES}} \times \lambda_t$, where λ_t is the marginal of an active power bus balance constraint at the ES location). In total, we used 10^5 dataset entries and thus the same number of independent lower levels to be solved. To reduce the computation time, computations were carried out in parallel running on a dual Intel Xeon CPU computer system over a total of 40 physical cores.

We implemented the CNN depicted in Fig. 4 in Python using PyTorch library [46]. CNN hyperparameters are described in Section II-D, and the hyperparameter β of the Softplus activation function is set to 50.

Training of the CNN was implemented using the fast.ai library [47] using training procedure similar as in [48]. We used the Ranger algorithm [49], employing the RAdam optimizer [50], the parameter lookahead [51], and the flat-cosine one-cycle policy [52]. After experimenting with training thousands of models, we set RAdam hyperparameters to the following values: the number of training epochs to 500, the maximum learning rate to 0.003, the training batch size to 128, the weight decay factor to 0.01, and the exponential decay rates of the first and second moments to values 0.95 and 0.85.

In every iteration of the metaoptimization scheme, we trained a total of $M = 60$ CNNs on the same dataset. Our computer system was equipped with 6 Nvidia Quadro RTX 6000 GPUs, each having 16 GB of RAM, so we could train all CNNs in parallel, training 10 CNNs per GPU.

In the metaoptimization scheme, there are two hyperparameters to consider. After experimenting with different values, for the relative tolerance of the dataset creation we take $\epsilon = 0.1$, and for the other hyperparameter we take $\gamma = 5$. A value of hyperparameter γ influences the decreasing rate of $p_{t,i}^{\mathrm{ES,rad}}$ through subsequent iterations. Using lower values of γ produces lower values of $p_{t,i}^{\mathrm{ES,rad}}$ in later iterations, which can lead to a suboptimal optimization result in the end, and a higher value tends to slow down the speed of convergence. Table I shows that $p_{t,i}^{\mathrm{ES,rad}}$ decreases throughout all iterations from the case study. Notice the difference in speed and intensity of $p_{t,i}^{\mathrm{ES,rad}}$ decrease in different transmission networks. More significant and faster decrease, as seen in, e.g., 3_lmbd, suggests that all 60 trained CNNs yield closer optimums, which can be explained by inherently tamer underlying optimization landscape. On the contrary, much less significant decrease in $p_{t,i}^{\mathrm{ES,rad}}$, as seen in, e.g., 73_ieee_rts, suggests that CNNs are struggling more to approximate the optimums, which is probably induced by more demanding optimization landscape.

B. Results

We tested our method on four separate transmission system meshed networks from PGLib-OPF [53] library: 3_lmbd, 57_ieee, 73_ieee_rts, and 300_ieee. Topologies of the three large networks are provided in Fig. 11. The three-bus network is of typical triangle topology. A time dimension was added to the data by applying the load scaling factors for winter workdays available from IEEE RTS-96 [54]. Set of time steps τ has 24 elements for different hours in a single day. In case of 73_ieee_rts network, we also applied 0.85 scaling factor to the transmission lines capacities to induce congestion. The networks were otherwise unmodified. For an ES to have an impact on the energy market prices, a feature for which the bilevel modeling is used for, it has to be very large. Thus, we model the ES with 100 MWh (1 p.u.) capacity. Charging and discharging efficiencies were both set to 90% and maximum ES (dis)charging power to 60 MW.

Table II presents an average wall time per iteration for each step of the proposed metaoptimization scheme. Dataset creation is a cumulative time for three substeps: generation of 10^5 random vectors of p_t^{ES}, solving lower level problems for every dataset entry, and data format postprocessing of the generated dataset, which mostly include disk input–output (IO) operations. Most of the time is consumed for solving 10^5 lower level problems. Notice that larger transmission system networks require more solver time. NN training is the time consumed for parallel CNN models training. This time does not depend on the transmission system network size, as it depends solely on the CNN and training hyperparameters. The last step is solving the metamodels, which is performed sequentially for all 60 trained CNNs. A possible speedup of using parallel computations in this step would not be significant compared to the total time used per iteration. Time for solving the metamodels can vary greatly between different CNNs and different iterations. We consider this to be a normal solver behavior due to binary variables x_t^{ch}. For the dataset creation and the NN training the average wall time is pretty much unchanged between different iterations, so we supply only the mean times without the standard deviation. Component total time is a sum of the dataset creation time, the NN training time, and the mean time for solving meta-models. Our model is highly scalable as long as the lower level problem can be evaluated a number of times under reasonable time and resources. An alternative method from our two-part paper [12] and [13] has scalability issues when using lower levels with larger networks since it computes the lower and the upper level simultaneously so the solution process can diverge. Table II indicates that our method scales reasonably even for 73 and 300 bus systems.

Tables III–VI present optimization results in terms of the ES computed and actual profits acquired in four different transmission systems. Actual profits are obtained in the verification Step 6 of Algorithm 1 by optimizing the lower level with fixed ES charging decisions as explained in Section II-F. We also compare actual profits achieved by our method to actual profits

TABLE III
PROFITS FOR 3_LMBD NETWORK (ES AT BUS 3)

Iter	Total time [s]	Mean p_t^{ES} actual profit	Best NN actual profit	Best NN computed profit	Single-level reduction actual profit [12], [13]	DC OPF actual profit
1	8386	2016.0583	**2016.2954**	2011.5968	2016.8762	1986.4979
2	7158	**2016.8395**	2016.7695	2018.1744		
3	6965	**2016.8414**	2016.8271	2017.5707		
4	6910	**2016.8416**	2016.8339	2016.9533		
5	6903	**2016.8505**	2016.8015	2017.2495		
6	6882	2016.8452	2016.8096	2017.1350		

TABLE IV
PROFITS FOR 57_IEEE NETWORK (ES AT BUS 1)

Iter	Total time [s]	Mean p_t^{ES} actual profit	Best NN actual profit	Best NN computed profit	Single-level reduction actual profit [12], [13]	DC OPF actual profit
1	8305	1564.0254	**1564.1351**	1564.3315	1565.2053	1542.5985
2	8724	1564.5929	**1564.6218**	1572.8308		
3	8886	**1564.9676**	1564.8713	1568.1618		
4	9088	1564.9433	1564.9370	1570.1936		

TABLE V
PROFITS FOR 73_IEEE_RTS NETWORK (ES AT BUS 101)

Iter	Total time [s]	Mean p_t^{ES} actual profit	Best NN actual profit	Best NN computed profit	Single-level reduction actual profit [12], [13]	DC OPF actual profit
1	10234	5503.2526	**5512.4143**	5525.5078	No solution	4990.1637
2	10186	5503.2608	**5512.6547**	5520.6102		
3	10159	5503.8406	**5512.7558**	5519.5750		
4	10192	5503.7870	5512.1032	5513.6247		

TABLE VI
PROFITS FOR 300_IEEE NETWORK (ES AT BUS 1)

Iter	Total time [s]	Mean p_t^{ES} actual profit	Best NN actual profit	Best NN computed profit	Single-level reduction actual profit [12], [13]	DC OPF actual profit
1	33482	1396.5797	**1397.0175**	1377.6627	No solution	1199.9745
2	36481	1396.7220	1397.0172	1399.9574		

achieved by solving a bilevel ES market optimal bidding using 1) the ac OPF model and single-level reduction approach from the two-part paper [12] and [13] and 2) a standard dc OPF [35] model in the lower level. AC OPF single-level reduction approach results in slightly higher actual profits compared to the NN approach, but its solution process fails to converge for 73- and 300-bus networks. Actual dc OPF profits are profits that would occur in the ac OPF market, but by using bidding decisions from the dc OPF bilevel model. In our tables, the best NN computed profit is the maximum value of $\mathcal{C}_{i,n}$ over all 60 NNs, the best NN actual profit is the maximum value of $\mathcal{V}_{i,n}$ over all 60 NNs, and the mean p_t^{ES} actual profit is $\overline{\mathcal{V}_i}$, where i is the iteration number and NNs are indexed by n. By design, the best profit is always achieved in the penultimate iteration of our method, as a worse profit in the last iteration actually triggers the stopping criterion. The number of required iterations differs between the transmission systems and ranges from 2 to 6. Tables III–VI demonstrate we also achieved a high first iteration accuracy, since the greatest second iteration improvement of the actual profit is only 0.03%. Note that the total time per iteration presented is somewhat higher than a component total time in Table II, as it includes an additional overhead for some data reformatting and IO disk operations. Also, note that we decided to present profits using up to four decimal places, so that small improvements between subsequent iterations in Table III become visible, which also reaffirms the optimality of the first iteration result. Relative differences in profits using all three approaches (single-level reduction, NN, and bilevel dc OPF) are clearly presented in Fig. 6 where the highest profits are normalized to 100%. On 3- and 57-bus networks, the NN approach achieved 99.9987% and 99.9848% of ES profits of the single-level reduction. Profit increase over the dc OPF model was 1.5% for 3_lmbd, 1.5% for 57_ieee, 10.5% for 73_ieee network, and 16.4% for 300_ieee network.

Figs. 7–10 present ES (dis)charging profiles for all four test cases. The 3-bus network is characterized with relatively small charging and discharging powers (up to 10 MW). In such small network, the ES (dis)charging would change the marginal producer, thus significantly affecting the market prices. In the 57-bus network the dc model discharges at over 40 MW in hour 12. Since it is a lossless model, it does not predict any price change due to ES arbitrage. On the other hand, the ac model captures price changes incurred by the ES and charges more evenly throughout the day. The 73-bus network features high price volatility, thus the ES makes the most cycles. In the 300-bus network, the dc model also assumes no price impact by its actions, which is incorrect, while the NN approach, which better captures the price changes due to the ES (dis)charging, is more conservative and discharges at low power in ours 17-19. Figs. 7–8 also show that the NN approach produces almost identical ES charging profile as the single-level reduction approach from [12] and [13].

IV. CONCLUSION

This article presents a novel numerical method, which utilizes deep convolutional NNs to efficiently solve a wide class of bilevel optimization problems arising in deregulated power systems. The method uses evolutionary metamodeling to bypass the lower level problem, thus it is insensitive to the lower level complexity, which is the main culprit in rendering bilevel optimization problems intractable. The model successfully deals with nonlinear nonconvex lower levels that include binary variables, as long as the lower level can be efficiently solved as a single-level problem by treating all upper level variables as parameters. We demonstrate the application of the method to solve the ES market participation problem using ac OPF in the lower level, which enables electricity market operators to perform highly accurate market clearing procedure. However, the proposed framework is generally applicable for any other bilevel optimization problem in the power systems domain.

Additionally, our method is scalable in terms of the required precision versus the run time. Using larger training datasets, we could train an NN having more parameters than we used here. We would obtain more accurate optimums, but would also require longer run time. On the other hand, if we are satisfied with lower precision, we could use a smaller dataset for training a smaller NN, and our method would run faster.

Finally, we note that this procedure can be easily implemented to trilevel models, as they are generally solved by first merging the middle- and the lower level problems into a mixed-integer

problem with equilibrium constraints, see, e.g., [55], which is a direct application of our proposed procedure. The obtained single-level problem then acts as a lower level problem to the original upper level problem. The resulting bilevel structure is commonly iteratively solved using a cutting plane algorithm. However, a direct implementation of our procedure to trilevel problems will be explored in future.

Acknowledgment

The sole responsibility for the content of this document lies with the authors. It does not necessarily reflect the opinion of the Innovation and Networks Executive Agency (INEA) or the European Commission (EC). INEA or the EC are not responsible for any use that may be made of the information contained therein.

Appendix
Lower Level

This section includes the formulation of the lower level, i.e., the exact ac OPF in the rectangular coordinates. Objective function (A.1) minimizes production costs, (A.2) and (A.3) are bus balances, (A.4)–(A.7) are power flow equations, (A.8) and (A.9) are generator production limits, (A.10) is line thermal limit, (A.11) is voltage limit, and (A.12) is reference bus constraint. $V^{\boldsymbol{r}}_{t,i}$ and $V^{\boldsymbol{i}}_{t,i}$, $\boldsymbol{\tau}^{\mathbf{r}}_e$ and $\boldsymbol{\tau}^{\mathbf{i}}_e$ are real and imaginary parts of voltage magnitude and transformer tap ratio, respectively. All other notations are the same as in our previous paper [11]. The simulated bidding process assumes the strategic market participant is the one in the upper level, while bids of all other participants are deterministic. Such modeling is a common practice and is described in detail in [56]. In reality, sufficient historical offering data to derive other participants' offering curves are available in some markets, e.g., see [57] for the Alberta market.

$$\underset{\Xi^{\mathrm{ll}}}{\mathrm{Min}} \sum_{t,k}(\ddot{\boldsymbol{c}}_k \times (P^{\mathrm{g}}_{t,k})^2 + \dot{\boldsymbol{c}}_k \times P^{\mathrm{g}}_{t,k} + \boldsymbol{c}_k) \tag{A.1}$$

$$\sum_{k\in G_i} P^{\mathrm{g}}_{t,k} - \sum_{l\in L_i} \boldsymbol{P}^{\mathbf{d}}_{t,l} - \sum_{(e,i,j)\in E\cup E^{\mathrm{R}}} P_{t,e,i,j} - p^{\mathrm{ES}}_{t\ :i\in\beta} - ((V^{\boldsymbol{r}}_{t,i})^2 + (V^{\boldsymbol{i}}_{t,i})^2) \times \sum_{s\in S_i} \boldsymbol{g}^{\mathbf{sh}}_s = 0 \quad \forall t,i \ : \lambda_{t,i} \tag{A.2}$$

$$\sum_{k\in G_i} Q^{\mathrm{g}}_{t,k} - \sum_{l\in L_i} \boldsymbol{Q}^{\mathbf{d}}_{t,l} - \sum_{(e,i,j)\in E\cup E^{\mathrm{R}}} Q_{t,e,i,j} + ((V^{\boldsymbol{r}}_{t,i})^2 + (V^{\boldsymbol{i}}_{t,i})^2) \times \sum_{s\in S_i} \boldsymbol{b}^{\mathbf{sh}}_s = 0 \quad \forall t,i \tag{A.3}$$

$$\begin{aligned} P_{t,e,i,j} &= ((V^{\boldsymbol{r}}_{t,i})^2 + (V^{\boldsymbol{i}}_{t,i})^2) \times (\boldsymbol{g}_e + \boldsymbol{g}^{\mathbf{fr}}_e)/\boldsymbol{\tau}^2_e \\ &+ ((-\boldsymbol{g}_e \times \boldsymbol{\tau}^{\mathbf{r}}_e + \boldsymbol{b}_e \times \boldsymbol{\tau}^{\mathbf{i}}_e) \times (V^{\boldsymbol{r}}_{t,i} \times V^{\boldsymbol{r}}_{t,j} + V^{\boldsymbol{i}}_{t,i} \times V^{\boldsymbol{i}}_{t,j}) \\ &+ (\boldsymbol{b}_e \times \boldsymbol{\tau}^{\mathbf{r}}_e + \boldsymbol{g}_e \times \boldsymbol{\tau}^{\mathbf{i}}_e) \times (V^{\boldsymbol{r}}_{t,i} \times V^{\boldsymbol{i}}_{t,j} - V^{\boldsymbol{i}}_{t,i} \times V^{\boldsymbol{r}}_{t,j}))/\boldsymbol{\tau}^2_e, \\ &\forall t,(e,i,j) \in E \end{aligned} \tag{A.4}$$

$$\begin{aligned} P_{t,e,i,j} &= ((V^{\boldsymbol{r}}_{t,i})^2 + (V^{\boldsymbol{i}}_{t,i})^2) \times (\boldsymbol{g}_e + \boldsymbol{g}^{\mathbf{to}}_e) + \\ &((-\boldsymbol{g}_e \times \boldsymbol{\tau}^{\mathbf{r}}_e - \boldsymbol{b}_e \times \boldsymbol{\tau}^{\mathbf{i}}_e) \times (V^{\boldsymbol{r}}_{t,i} \times V^{\boldsymbol{r}}_{t,j} + V^{\boldsymbol{i}}_{t,i} \times V^{\boldsymbol{i}}_{t,j}) \\ &+ (\boldsymbol{b}_e \times \boldsymbol{\tau}^{\mathbf{r}}_e - \boldsymbol{g}_e \times \boldsymbol{\tau}^{\mathbf{i}}_e) \times (V^{\boldsymbol{r}}_{t,i} \times V^{\boldsymbol{i}}_{t,j} - V^{\boldsymbol{i}}_{t,i} \times V^{\boldsymbol{r}}_{t,j}))/\boldsymbol{\tau}^2_e, \\ &\forall t,(e,i,j) \in E^{\mathrm{R}} \end{aligned} \tag{A.5}$$

$$\begin{aligned} Q_{t,e,i,j} &= -((V^{\boldsymbol{r}}_{t,i})^2 + (V^{\boldsymbol{i}}_{t,i})^2) \times (\boldsymbol{b}_e + \boldsymbol{b}^{\mathbf{fr}}_e)/\boldsymbol{\tau}^2_e \\ &+ ((\boldsymbol{b}_e \times \boldsymbol{\tau}^{\mathbf{r}}_e + \boldsymbol{g}_e \times \boldsymbol{\tau}^{\mathbf{i}}_e) \times (V^{\boldsymbol{r}}_{t,i} \times V^{\boldsymbol{r}}_{t,j} + V^{\boldsymbol{i}}_{t,i} \times V^{\boldsymbol{i}}_{t,j}) \\ &+ (\boldsymbol{g}_e \times \boldsymbol{\tau}^{\mathbf{r}}_e - \boldsymbol{b}_e \times \boldsymbol{\tau}^{\mathbf{i}}_e) \times (V^{\boldsymbol{r}}_{t,i} \times V^{\boldsymbol{i}}_{t,j} - V^{\boldsymbol{i}}_{t,i} \times V^{\boldsymbol{r}}_{t,j}))/\boldsymbol{\tau}^2_e, \\ &\forall t,(e,i,j) \in E \end{aligned} \tag{A.6}$$

$$\begin{aligned} Q_{t,e,i,j} &= -((V^{\boldsymbol{r}}_{t,i})^2 + (V^{\boldsymbol{i}}_{t,i})^2) \times (\boldsymbol{b}_e + \boldsymbol{b}^{\mathbf{fr}}_e)/\boldsymbol{\tau}^2_e \\ &+ ((\boldsymbol{b}_e \times \boldsymbol{\tau}^{\mathbf{r}}_e + \boldsymbol{g}_e \times \boldsymbol{\tau}^{\mathbf{i}}_e) \times (V^{\boldsymbol{r}}_{t,i} \times V^{\boldsymbol{r}}_{t,j} + V^{\boldsymbol{i}}_{t,i} \times V^{\boldsymbol{i}}_{t,j}) \\ &+ (\boldsymbol{g}_e \times \boldsymbol{\tau}^{\mathbf{r}}_e - \boldsymbol{b}_e \times \boldsymbol{\tau}^{\mathbf{i}}_e) \times (V^{\boldsymbol{r}}_{t,i} \times V^{\boldsymbol{i}}_{t,j} - V^{\boldsymbol{i}}_{t,i} \times V^{\boldsymbol{r}}_{t,j}))/\boldsymbol{\tau}^2_e, \\ &\forall t,(e,i,j) \in E \end{aligned} \tag{A.7}$$

$$\underline{\boldsymbol{P}}^{\mathbf{g}}_k \le P^{\mathrm{g}}_{t,k} \le \overline{\boldsymbol{P}}^{\mathbf{g}}_k \qquad \forall t,k \tag{A.8}$$

$$\underline{\boldsymbol{Q}}^{\mathbf{g}}_k \le Q^{\mathrm{g}}_{t,k} \le \overline{\boldsymbol{Q}}^{\mathbf{g}}_k \qquad \forall t,k \tag{A.9}$$

$$P^2_{t,e,i,j} + Q^2_{t,e,i,j} \le \overline{\boldsymbol{S}}^2_e \qquad \forall t,(e,i,j) \in E\cup E^{\mathrm{R}} \tag{A.10}$$

$$\underline{\boldsymbol{V}}^2_i \le (V^{\boldsymbol{r}}_{t,i})^2 + (V^{\boldsymbol{i}}_{t,i})^2 \le \overline{\boldsymbol{V}}^2_i \qquad \forall t,i \tag{A.11}$$

$$V^{\boldsymbol{r}}_{t,i} \ge 0, \quad V^{\boldsymbol{i}}_{t,i} = 0 \quad \forall t, i \in R \tag{A.12}$$

References

[1] IEEE Xplore. Accessed: Oct. 4, 2021. [Online]. Available: ieeexplore.ieee.org/Xplore/home.jsp

[2] J. M. Arroyo and F. D. Galiana, "On the solution of the bilevel programming formulation of the terrorist threat problem," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 789–797, May 2005.

[3] I. Momber, S. Wogrin, and T. G. San Román, "Retail pricing: A bilevel program for PEV aggregator decisions using indirect load control," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 464–473, Jan. 2016.

[4] C. Wang, Z. Wang, Y. Hou, and K. Ma, "Dynamic game-based maintenance scheduling of integrated electric and natural gas grids with a bilevel approach," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4958–4971, Sep. 2018.

[5] L. P. Garces, A. J. Conejo, R. Garcia-Bertrand, and R. Romero, "A bilevel approach to transmission expansion planning within a market environment," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1513–1522, Aug. 2009.

[6] E. Nasrolahpour, J. Kazempour, H. Zareipour, and W. D. Rosehart, "A bilevel model for participation of a storage system in energy and reserve markets," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 582–598, Apr. 2018.

[7] M. Carrion, J. M. Arroyo, and A. J. Conejo, "A bilevel stochastic programming approach for retailer futures market trading," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1446–1456, Aug. 2009.

[8] S. A. Gabriel, A. J. Conejo, J. D. Fuller, B. F. Hobbs, and C. Ruiz, *Complementarity Modeling in Energy Markets*. New York, NY, USA: Springer, 2013.

[9] H. Pandžić, A. J. Conejo, I. Kuzle, and E. Caro, "Yearly maintenance scheduling of transmission lines within a market environment," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 407–415, Feb. 2012.

[10] K. Pandžić, K. Bruninx, and H. Pandžić, "Managing risks faced by strategic battery storage in joint energy-reserve markets," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4355–4365, Sep. 2021.

[11] K. Šepetanc and H. Pandžić, "Convex polar second-order taylor approximation of AC power flows: A unit commitment study," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3585–3594, Jul. 2021.

[12] K. Šepetanc, H. Pandžić, and T. Capuder, "Solving bilevel AC OPF problems by smoothing the complementary conditions–Part I: Model description and the algorithm," *IEEE Trans. Power Syst.*, early access, doi: 10.1109/TPWRS.2022.3207088.

[13] K. Šepetanc, H. Pandžić, and T. Capuder, "Solving bilevel AC OPF problems by smoothing the complementary conditions–Part II: Solution techniques and case study," *IEEE Trans. Power Syst.*, early access, doi: 10.1109/TPWRS.2022.3207097.

[14] J. Fortuny-Amat and B. McCarl, "A representation and economic interpretation of a two-level programming problem," *J. Oper. Res. Soc.*, vol. 32, no. 9, pp. 783–792, Sep. 1981.

[15] T. A. Edmunds and J. F. Bard, "Algorithms for nonlinear bilevel mathematical programs," *IEEE Trans. Systems, Man, Cybern.*, vol. 21, no. 1, pp. 83–89, Jan./Feb. 1991.

[16] S. Zolfaghari and T. Akbari, "Bilevel transmission expansion planning using second-order cone programming considering wind investment," *Energy*, vol. 154, pp. 455–465,Jul. 2018.

[17] B. Dandurand, K. Kim, and S. Leyffer, "A bilevel approach for identifying the worst contingencies for nonconvex alternating current power systems," *SIAM J. Optim.*, vol. 31, no. 1, pp. 702–726, Feb. 2021.

[18] C. D. Kolstad and L. S. Lasdon, "Derivative evaluation and computational experience with large bilevel mathematical programs," *J. Optim. Theory Appl.*, vol. 65, no. 3, pp. 485–499, Jun. 1990.

[19] E. Aiyoshi and K. Shimizu, "A solution method for the static constrained Stackelberg problem via penalty method," *IEEE Trans. Autom. Control*, vol. 29, no. 12, pp. 1111–1114, Dec. 1984.

[20] Y. Ishizuka and E. Aiyoshi, "Double penalty method for bilevel optimization problems," *Ann. Oper. Res.*, vol. 34, no. 1, pp. 73–88, Dec. 1992.

[21] B. Colson, P. Marcotte, and G. Savard, "A trust-region method for nonlinear bilevel programming: Algorithm and computational experience," *Comput. Optim. Appl.*, vol. 30, no. 3, pp. 211–227, Mar. 2005.

[22] H. C. Bylling, S. A. Gabriel, and T. K. Boomsma, "A parametric programming approach to bilevel optimisation with lower-level variables in the upper level," *J. Oper. Res. Soc.*, vol. 71, no. 5, pp. 846–865, 2020.

[23] N. P. Faísca et al., "Parametric global optimisation for bilevel programming," *J. Glob. Optim.*, vol. 38, pp. 609–623, Aug. 2007.

[24] S. Avraamidou and E. N. Pistikopoulos, "A multi-parametric optimization approach for bilevel mixed-integer linear and quadratic programming problems," *Comput. Chem. Eng.*, vol. 125, pp. 98–113, Jun. 2019.

[25] A. Sinha, P. Malo, A. Frantsev, and K. Deb, "Finding optimal strategies in a multi-period multi-leader–follower Stackelberg game using an evolutionary algorithm," *Comput. Oper. Res.*, vol. 41, pp. 374–385, Jan. 2014.

[26] X. Li, P. Tian, and X. Min, "A hierarchical particle swarm optimization for solving bilevel programming problems," in *Artificial Intelligence and Soft Computing–ICAISC 2006 (LNCS 4029)*, L. Rutkowski et al. Eds. Heidelberg, Germany:Springer, 2006, pp. 1169–1178.

[27] J. S. Angelo, E. Krempser, and H. J. C. Barbosa, "Differential evolution for bilevel programming," in *Proc. IEEE Congr. Evol. Comput.*, 2013, pp. 470–477.

[28] X. Zhu, Q. Yu, and X. Wang, "A hybrid differential evolution algorithm for solving nonlinear bilevel programming with linear constraints," in *Proc. IEEE 5th Int. Conf. Cogn. Inform.*, 2006, pp. 126–131.

[29] S. R. Hejazi, A. Memariani, G. Jahanshahloo, and M. M. Sepehri, "Linear bilevel programming solution by genetic algorithm," *Comput. Oper. Res.*, vol. 29, no. 13, pp. 1913–1925, Nov. 2002.

[30] A. Sinha, P. Malo, and K. Deb, "Evolutionary algorithm for bilevel optimization using approximations of the lower level optimal solution mapping," *Eur. J. Oper. Res.*, vol. 257, no. 2, pp. 395–411, Mar. 2017.

[31] A. Sinha, P. Malo, and K. Deb, "Solving optimistic bilevel programs by iteratively approximating lower level optimal value function," in *Proc. IEEE Congr. Evol. Comput.*, 2016, pp. 1877–1884.

[32] A. Sinha, P. Malo, and K. Deb, "A review on bilevel optimization: From classical to evolutionary approaches and applications," *IEEE Trans. Evol. Comput.*, vol. 22, no. 2, pp. 276–295, Apr. 2018.

[33] A. Ivakhnenko and V. Lapa, *Cybernetic Predicting Devices*. New York, NY, USA: USA: CCM Information Corp., 1973.

[34] A. Ivakhnenko and V. Lapa, *Cybernetics and Forecasting Techniques*. New York, NY, USA: USA: American Elsevier, 1967.

[35] B. Stott, J. Jardim, and O. Alsac, "DC power flow revisited," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1290–1300, Aug. 2009.

[36] H. Pandžić and I. Kuzle, "Energy storage operation in the day-ahead electricity market," in *Proc. 12th Int. Conf. Eur. Energy Market*, 2015, pp. 1–6.

[37] S. Frank and S. Rebennack, "An introduction to optimal power flow: Theory, formulation, and examples," *IIE Trans.*, vol. 48, no. 12, pp. 1172–1197, May 2016.

[38] R. P. O'Neill, A. Castillo, and M. B. Cain, "The IV formulation and linear approximations of the ac optimal power flow problem," FERC, Washington, DC, USA, Tech. Rep., Dec. 2012. Accessed: Jul. 10, 2021. [Online] Available at: cms.ferc.gov/sites/default/files/2020-05/acopf-2-iv-linearization.pdf

[39] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, Mar. 1989.

[40] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control Signals Syst.*, vol. 2, no. 4, pp. 303–314, Dec. 1989.

[41] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.

[42] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3642–3649.

[43] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[44] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[45] R. B. Rall, *Automatic Differentiation: Techniques and Applications*. New York, NY, USA: Springer-Verlag, 1981.

[46] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 1–12, 2019.

[47] J. Howard et al., fastai, GitHub, 2018. Accessed: Aug. 17, 2021. [Online] Available: github.com/fastai/fastai

[48] T. Ivek and D. Vlah, "BlackBox: Generalizable reconstruction of extremal values from incomplete spatio-temporal data," *Extremes*, vol. 24, no. 1, pp. 145–162, Oct. 2020.

[49] L. Wright et al., Ranger Deep Learning Optimizer, GitHub, 2019. Accessed: Aug. 17, 2021. [Online] Available at: https://github.com/lessw2020/Ranger-Deep-Learning-Optimizer

[50] L. Liu et al., "On the variance of the adaptive learning rate and beyond," in *Proc. Int. Conf. Learn. Representations*, 2020, pp. 1–14.

[51] M. R. Zhang, J. Lucas, G. Hinton, and J. Ba, "Lookahead optimizer: K steps forward, 1 step back," *Neural Inf. Process. Syst.*, 2019.

[52] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1—Learning rate, batch size, momentum, and weight decay," US Naval Res. Lab., Tech. Rep. 5510-026, 2018.

[53] C. Coffrin et al. PGLib-OPF v20.07, GitHub, Jul. 2020. Accessed: Sep. 15, 2021. [Online] Available at: github.com/power-grid-lib/pglib-opf/tree/v20.07

[54] C. Grigg et al., "The IEEE reliability test System-1996," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1010–1020, Aug. 1999.

[55] K. Pandžić, H. Pandžić, and I. Kuzle, "Coordination of regulated and merchant energy storage investments," *IEEE Trans. Sustain. Energy*, vol. 9, no. 3, pp. 1244–1254, Jun. 2022.

[56] S. Wogrin, S. Pineda, and D. A. Tejada-Arango, "Applications of bilevel optimization in energy and electricity markets," in *Bilevel Optimization: Advances and Next Challenges*. Cham, Switzerland: Springer, 2020.

[57] Alberta Electric System Operator, "Market and system reporting," 2022. [Online] Available at: www.aeso.ca/market/market-and-system-reporting

[58] R. A. González-Fernández et al., "Composite systems reliability evaluation based on Monte Carlo simulation and cross-entropy methods," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4598–4606, Nov. 2013.

[59] Al-Roomi, "Power Systems and evolutionary algorithms—Load flow," 2015. [Online]. Available: https://al-roomi.org/power-flow/

Transmission Expansion Planning using a Highly Accurate AC Optimal Power Flow Approximation

Otto Heide
University of Zagreb
Faculty of Electrical
Engineering and Computing
Zagreb, Croatia
Email: otto.heide@fer.hr

Karlo Šepetanc
University of Zagreb
Faculty of Electrical
Engineering and Computing;
Innovation Centre N. Tesla
Zagreb, Croatia
Email: karlo.sepetanc@fer.hr

Hrvoje Pandžić
University of Zagreb
Faculty of Electrical
Engineering and Computing;
Innovation Centre N. Tesla
Zagreb, Croatia
Email: hrvoje.pandzic@fer.hr

2021 International Conference on Smart Energy Systems and Technologies (SEST) | 978-1-7281-7660-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/SEST50973.2021.9543431

***Abstract*—The objective of transmission network expansion planning is to find the optimal strategy that balances the investment and the operating costs, considering all generation and transmission constraints. Attempts to address this problem in a tractable manner have led researchers to develop different convex relaxations and approximations. Due to the constant power grid evolution, new and improved approximation models are required to successfully handle the upcoming challenges. In this paper, we present a comprehensive approach to handle this highly complex problem both tractably and accurately. The model is based on a convex polar second-order Taylor expansion approximation of the AC power flows where both the voltage magnitudes and angles are quadratically constrained. The proposed approach achieves high accuracy due to the elimination of constraint relaxation errors, as determined by the presolve, which can occur due to the convexification process. The model demonstrated superior accuracy and similar computation times as the existing approximation models. In comparison to the exact formulations, our model shows similar accuracy while improving the computation time.**

***Index Terms*—Optimal power flow approximation, transmission expansion planning, mixed-integer quadratically constrained quadratic program**

NOMENCLATURE

A. Sets and Indices

N	Set of buses, indexed by i and j.
N^{P}	Tuple set of paired buses aligned with branch E orientation, indexed by (i, j).
R	Set of reference buses, indexed by i.
E, E^{R}	Tuple set of branches, forward and reverse orientation, indexed by (e, i, j).
E_i, E_i^{R}	Array of tuple sets of branches at bus i, forward and reverse orientation, indexed by (e, i, j).
$E^{+}, E^{\mathrm{R}+}$	Tuple set of prospective expansion branches, forward and reverse orientation, indexed by (e, i, j).
$E_i{}^{+}, E_i^{\mathrm{R}+}$	Array of tuple sets of prospective expansion branches at bus i, forward and reverse orientation, indexed by (e, i, j).
G, G_i	Set of all generators and array of sets of generators at bus i, indexed by k.
L_i	Array of sets of loads at bus i, indexed by l.
S_i	Array of sets of shunts at bus i, indexed by s.
τ	Set of time steps, indexed by t.
Ω, Ω_i	Set of all wind generators and array of sets of wind generators at bus i, indexed by w.
Ξ	Set of decision variables.

B. Parameters

$\ddot{\boldsymbol{c}}_k, \dot{\boldsymbol{c}}_k, \boldsymbol{c}_k$	Generator cost coefficients.
$\boldsymbol{P}^{\mathbf{d}}_{t,l}, \boldsymbol{Q}^{\mathbf{d}}_{t,l}$	Active and reactive power load.
$\boldsymbol{g}^{\mathbf{sh}}_s, \boldsymbol{b}^{\mathbf{sh}}_s$	Bus shunt conductance and susceptance.
$\boldsymbol{g}_e, \boldsymbol{g}^{\mathbf{fr}}_e, \boldsymbol{g}^{\mathbf{to}}_e$	Branch π-section conductances.
$\boldsymbol{b}_e, \boldsymbol{b}^{\mathbf{fr}}_e, \boldsymbol{b}^{\mathbf{to}}_e$	Branch π-section susceptances.
$\boldsymbol{\tau}_e, \boldsymbol{\sigma}_e$	Branch tap magnitude and shift angle.
$\underline{\boldsymbol{P}}^{\mathbf{g}}_k, \overline{\boldsymbol{P}}^{\mathbf{g}}_k$	Generator minimum and maximum active power output.
$\underline{\boldsymbol{Q}}^{\mathbf{g}}_k, \overline{\boldsymbol{Q}}^{\mathbf{g}}_k$	Generator minimum and maximum reactive power output.
$\overline{\boldsymbol{P}}^{\omega}_w$	Wind generator maximum active power output.
$\overline{\boldsymbol{S}}_e$	Branch maximum apparent power.
$\underline{\boldsymbol{\theta}}_{i,j}, \overline{\boldsymbol{\theta}}_{i,j}$	Bus-pair minimum and maximum voltage angle difference.
$\underline{\boldsymbol{V}}_i, \overline{\boldsymbol{V}}_i$	Bus minimum and maximum voltage magnitude.
$\boldsymbol{V}^{\mathbf{op}}_{t,i}, \boldsymbol{\theta}^{\mathbf{op}}_{t,i}$	Assumed bus voltage magnitude and angle operating points.
$\Lambda_{t,e}, \Gamma_{t,i,j}$	Boolean parameters which indicate whether to use quadratic form of voltage and cosine representation respectively.
M	Disjunctive factor, a large positive number.
$cost_e$	Expansion cost coefficient.

C. Variables

$P^{\mathrm{g}}_{t,k}, Q^{\mathrm{g}}_{t,k}$	Generator active and reactive power production.
$P^{\omega}_{t,w}$	Wind generator active power production.
$P_{t,e,i,j}, Q_{t,e,i,j}$	Branch active and reactive power flow.
$V^{\Delta}_{t,i}, \theta^{\Delta}_{t,i}$	Bus voltage magnitude and angle change.
$V_{t,i}, \theta_{t,i}$	Bus voltage and magnitude.
$\widehat{cos}_{t,i,j}$	Cosine approximation.
$\breve{V}_{t,e}$	Second order Taylor series voltage magnitude term approximation.
z_e	Binary decision variable for a prospective line.

978-1-7281-7660-4/21/$31.00 ©2021 IEEE

1 . INTRODUCTION

A. Motivation and background

Transmission expansion planning (TEP) represents an important research area in the field of power systems. TEP determines the location and number of new lines that need to be installed to achieve certain goals in the transmission of electrical power. An optimal TEP solution usually consists of several targets, such as increasing reliability and ensuring the security of supply, minimizing the investment and operating costs, reducing power losses, and avoiding potential congestion. With the continuous increase in demand levels, more lines will become congested in the near future and for that reason, it is important to identify and improve potential weak spots in the transmission system to ensure system security and to maximize social welfare.

The nonlinear and non-convex nature of the exact AC TEP problem makes the computation of the globally optimal solution, in a reasonable time, difficult to achieve, especially when large-scale networks are considered. The TEP problem has been solved using mathematical optimization approximation and relaxation models [1] – [11] and heuristic optimization methods [12] – [14]. Paper [15] presents a comprehensive review and classification of available publications and models on the TEP problem. Heuristic methods, based on the power flow results, incrementally select expansion line that removes congestion in a selected part of the network. Considering the added line, power flow analysis is recalculated and the process continues step-by-step until there is no more congestion in the network. Heuristic methods rarely achieve global optimality and do not provide any optimality estimates. On the other hand, convex mathematical optimization models provide model's solution optimality guarantees, but no feasibility guarantees due to reduced accuracy due to applied relaxations or approximations. Several methods have been proposed for the TEP problem and they mostly use classical optimization techniques, such as linear programming [3] – [5], non-linear programming [6] and mixed-integer programming [7] – [8].

B. Literature review

Due to the high computational complexity of TEP, using one of the exact AC network models is not a popular approach despite the ultimate accuracy of the obtained solution. Thus, [12] presented a mixed-integer nonlinear programming (MINLP) approach for solving TEP for an AC network model using heuristic algorithms and interior point method which obtained a quality solution for the presented problem. For mathematical optimization programming, approximations such as linear DC model [10], piecewise-linearized AC model [1] and linear-programming of AC power flows (LPAC) [9] are commonly used to approximate the exact AC power flow equations.

The DC model approximation has the fastest computation time compared to any other approach, but in terms of accuracy, it results in a suboptimal and frequently infeasible solution in reality. Accuracy gap of the linear DC model compared to the exact one arises from simplifications made when neglecting reactive power flows, active power losses, and voltage drops in network optimization modeling [2], [12], [14]. Piecewise linearization of AC power flows (LACTEP) was introduced in [1], where reactive power flows, active power losses, and off-nominal bus voltage magnitudes were retained. Linearization is based on the first-order Taylor series expansion and is used to separately model network losses initially defined as non-convex constraints. The optimal solution and computation time highly correlate with the number of linear blocks used in the piecewise linearization process. The objective function in this approach varies with the number of linear blocks. To obtain the best solution, it is necessary to perform an iterative sensitivity analysis with different number of linear blocks, which results in a prolonged computation time. [9] proposes a linear approximation of the AC power flow equations (LPAC) that, contrary to the DC model, captures reactive power flows and voltage magnitudes, as well as active and reactive power losses, which means they do not have to be modeled separately as in [1]. The linearity of power flow equations in the LPAC model is highly desirable in terms of computational tractability. However, for the TEP process, the cosine approximation is modeled in its quadratically constrained formulation to better capture the voltage angle variable, thus the LPAC model in this paper is presented as a mixed-integer quadratically constrained quadratic programming (MIQCQP) model.

However, all of the previously mentioned popular approximations tend to have certain accuracy disadvantages when it comes to modeling of reactive power flows, voltage magnitudes, and losses, and thus often result in a suboptimal or even infeasible solutions. On the other hand, convex quadratic approaches of the AC power flows can achieve high accuracy when there are no constraint relaxation errors due to the convexification process. A new ACOPF approximation approach [16] is based on the convex polar second-order Taylor approximation (CPSOTA) of AC power flows, where the relaxed quadratic constraints that would likely cause relaxation errors are identified in the presolve process. The identified quadratic inequalities are substituted with linear equality constraints, significantly improving model's accuracy. Our work builds upon the CPSOTA approximation by introducing new constraints necessary for TEP.

C. Paper contribution and structure

Contribution of the paper consists of the following:

- We present new model for TEP process based on the Taylor series that approximates second-order voltage variables.
- Presolve technique is used to decide whether to use the quadratic or the linear form of power flow constraints to avoid constraint relaxation errors due to the convexification process.
- The resulting MIQCQP solution is obtained much quicker than the MINLP solution, while maintaining high accuracy

Rest of the paper is structured as follows: Section 2 presents our mathematical model for TEP problem. Subsection 2 -A in-

troduces the presolve technique and Subsection 2 -B presents model components. Case study is presented in Section 3 . It presents description and set-up of test cases, and algorithm that describes four step procedure of our TEP model. Section 4 provides relevant conclusions and guidelines for future work.

2 . Mathematical Model

Our model is based on the convex polar second-order Taylor approximation of AC power flows where both the voltage magnitudes and angles are quadratically constrained. The proposed approach achieves high accuracy due to the elimination of constraint relaxation errors, as determined in the presolve process, which can occur due to the convexification process. Detailed algorithmic implementation of the presented model is defined in Section 3 .

A. Presolve technique

The convex quadratic approach can achieve high accuracy when there are no constraint relaxation errors that result from the convexification process. The presolve process identifies constraints that would likely cause relaxation errors and decides whether to use the quadratic or the linear form of power flow constraints to avoid relaxation errors.

$$\breve{V}_{t,e}=\frac{\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{fr}}}{\boldsymbol{\tau}_e^2}\cdot(V_{t,i}^{\Delta})^2-\frac{2\cdot\boldsymbol{g}_e}{\boldsymbol{\tau}_e}\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e)\cdot V_{t,i}^{\Delta}\cdot V_{t,j}^{\Delta} +(\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{to}})\cdot(V_{t,j}^{\Delta})^2,\quad \forall t,(e,i,j)\in E \tag{1.1}$$

$$\widehat{cos}_{t,i,j}=1-\frac{(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})^2}{2},\quad \forall t,(i,j)\in N^{\mathrm{P}} \tag{1.2}$$

Swapping the inequality constraints in equations (2.8) and (2.10) with their linear alternative in equations (2.9) and (2.11) avoids constraint relaxation errors due to convexification process. The decision for swapping inequality constraints with their linear alternative is based on the marginal value of equations (1.1) and (1.2), which are defined as quadratic equality constraints in the presolve process. Constraints' marginal values represent the sensitivity of the objective function on these constraints and they are computed by default by many solvers, e.g. IPOPT. For constraint (2.8) to be binding, due to its greater-or-equal sign, $\breve{V}_{t,e}$ should have the tendency to be as small as possible, i.e. marginal value of (1.1) needs to be positive. Oppositely, for constraint (2.10) to be binding, due to its less-or-equal sign, marginal value of (1.2) needs to be negative. Therefore, the quadratic form of constraint $\breve{V}_{t,e}$ is used only if the Boolean parameter $\Lambda_{t,e}$ is true and conductance $\boldsymbol{g}_e$ is positive, and quadratic form of constraint $\widehat{cos}_{t,i,j}$ is used only in the Boolean parameter $\Gamma_{t,i,j}$ is true.

B. Optimization model

This subsection presents the whole network-constrained TEP model.

$$\mathrm{Min}(\sum_{t,k}(\ddot{\boldsymbol{c}}_k\cdot(P_{t,k}^{\mathrm{g}})^2+\dot{\boldsymbol{c}}_k\cdot P_{t,k}^{\mathrm{g}}+\boldsymbol{c}_k)+\sum_{e\in E^+}z_e\cdot cost_e) \tag{2.1}$$

$$\sum_{k\in G_i}P_{t,k}^{\mathrm{g}}-\sum_{l\in L_i}\boldsymbol{P}_{t,l}^{\mathbf{d}}-\sum_{(e,i,j)\in E_i\cup E_i^{\mathrm{R}}}P_{t,e,i,j}-\sum_{(e,i,j)\in(E_i^{+}\cup E^{\mathrm{R}+}_i)}P_{t,e,i,j} -((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot\sum_{s\in S_i}\boldsymbol{g}_s^{\mathbf{sh}}+\sum_{w\in\Omega}P_{t,w}^{\omega}=0,\forall t,i \tag{2.2}$$

$$\sum_{k\in G_i}Q_{t,k}^{\mathrm{g}}-\sum_{l\in L_i}\boldsymbol{Q}_{t,l}^{\mathbf{d}}-\sum_{(e,i,j)\in E_i\cup E_i^{\mathrm{R}}}Q_{t,e,i,j}-\sum_{(e,i,j)\in(E_i^{+}\cup E^{\mathrm{R}+}_i)}Q_{t,e,i,j} +((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot\sum_{s\in S_i}\boldsymbol{b}_s^{\mathbf{sh}}=0,\quad\forall t,i \tag{2.3}$$

$$P_{t,e,i,j}=\frac{((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot(\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{fr}})}{\boldsymbol{\tau}_e^2}+\frac{\breve{V}_{t,e}}{2} -(\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e)+\boldsymbol{b}_e\cdot\sin(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e))\cdot (\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot\widehat{cos}_{t,i,j}+V_{t,i}^{\Delta}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}+V_{t,j}^{\Delta}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}})/\boldsymbol{\tau}_e -(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e)-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e))\cdot \boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})/\boldsymbol{\tau}_e,\quad\forall t,(e,i,j)\in E \tag{2.4}$$

$$P_{t,e,i,j}=((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot(\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{to}})+\frac{\breve{V}_{t,e}}{2} -(\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}+\boldsymbol{\sigma}_e)+\boldsymbol{b}_e\cdot\sin(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}+\boldsymbol{\sigma}_e))\cdot (\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot\widehat{cos}_{t,j,i}+V_{t,i}^{\Delta}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}+V_{t,j}^{\Delta}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}})/\boldsymbol{\tau}_e -(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}+\boldsymbol{\sigma}_e)-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}+\boldsymbol{\sigma}_e))\cdot \boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})/\boldsymbol{\tau}_e,\quad\forall t,(e,i,j)\in E^{\mathrm{R}} \tag{2.5}$$

$$Q_{t,e,i,j}=-\frac{((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot(\boldsymbol{b}_e+\boldsymbol{b}_e^{\mathbf{fr}})}{\boldsymbol{\tau}_e^2} +(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e)-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e))\cdot (\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot\widehat{cos}_{t,i,j}+V_{t,i}^{\Delta}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}+V_{t,j}^{\Delta}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}})/\boldsymbol{\tau}_e -(\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e)+\boldsymbol{b}_e\cdot\sin(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e))\cdot \boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})/\boldsymbol{\tau}_e,\quad\forall t,(e,i,j)\in E \tag{2.6}$$

$$Q_{t,e,i,j}=-((\boldsymbol{V}_{t,i}^{\mathbf{op}})^2+2\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot V_{t,i}^{\Delta})\cdot(\boldsymbol{b}_e+\boldsymbol{b}_e^{\mathbf{to}}) +(\boldsymbol{b}_e\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}+\boldsymbol{\sigma}_e)-\boldsymbol{g}_e\cdot\sin(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}+\boldsymbol{\sigma}_e))\cdot (\boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot\widehat{cos}_{t,j,i}+V_{t,i}^{\Delta}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}+V_{t,j}^{\Delta}\cdot\boldsymbol{V}_{t,i}^{\mathbf{op}})/\boldsymbol{\tau}_e -(\boldsymbol{g}_e\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}+\boldsymbol{\sigma}_e)+\boldsymbol{b}_e\cdot\sin(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}+\boldsymbol{\sigma}_e))\cdot \boldsymbol{V}_{t,i}^{\mathbf{op}}\cdot\boldsymbol{V}_{t,j}^{\mathbf{op}}\cdot(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})/\boldsymbol{\tau}_e,\quad\forall t,(e,i,j)\in E^{\mathrm{R}} \tag{2.7}$$

$$\breve{V}_{t,e}\geqslant\frac{\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{fr}}}{\boldsymbol{\tau}_e^2}\cdot(V_{t,i}^{\Delta})^2-\frac{2\cdot\boldsymbol{g}_e}{\boldsymbol{\tau}_e}\cdot\cos(\boldsymbol{\theta}_{t,i}^{\mathbf{op}}-\boldsymbol{\theta}_{t,j}^{\mathbf{op}}-\boldsymbol{\sigma}_e)\cdot V_{t,i}^{\Delta}\cdot V_{t,j}^{\Delta} +(\boldsymbol{g}_e+\boldsymbol{g}_e^{\mathbf{to}})\cdot(V_{t,j}^{\Delta})^2,\ \forall t,(e,i,j)\in(E\cup E^{+}):\boldsymbol{g}_e>0\wedge\Lambda_{t,e} \tag{2.8}$$

$$\breve{V}_{t,e}=0,\quad\forall t,(e,i,j)\in(E\cup E^{+}):\boldsymbol{g}_e\leqslant 0\vee\neg\Lambda_{t,e} \tag{2.9}$$

$$\widehat{cos}_{t,i,j}\leqslant 1-\frac{(\theta_{t,i}^{\Delta}-\theta_{t,j}^{\Delta})^2}{2},\quad\forall t,(i,j)\in N^{\mathrm{P}}:\Gamma_{t,i,j} \tag{2.10}$$

$$\widehat{cos}_{t,i,j}=1,\quad\forall t,(i,j)\in N^{\mathrm{P}}:\neg\Gamma_{t,i,j} \tag{2.11}$$

$$0\leqslant P_{t,w}^{\omega}\leqslant\overline{\boldsymbol{P}}_w^{\omega},\quad\forall t,w \tag{2.12}$$

$$\underline{\boldsymbol{P}}_k^{\mathbf{g}}\leqslant P_{t,k}^{\mathrm{g}}\leqslant\overline{\boldsymbol{P}}_k^{\mathbf{g}},\quad\forall t,k \tag{2.13}$$

$$\underline{\boldsymbol{Q}}_k^{\mathbf{g}} \leqslant Q_{t,k}^{\mathrm{g}} \leqslant \overline{\boldsymbol{Q}}_k^{\mathbf{g}}, \quad \forall t,k \tag{2.14}$$

$$P_{t,e,i,j}^2 + Q_{t,e,i,j}^2 \leqslant \overline{\boldsymbol{S}}_e^2, \quad \forall t,(e,i,j) \in (E \cup E^{\mathrm{R}}) : \exists \overline{\boldsymbol{S}}_e \tag{2.15}$$

$$\boldsymbol{\theta}_{t,i}^{\mathbf{op}} + \theta_{t,i}^{\Delta} = 0, \quad \forall t, i \in R \tag{2.16}$$

$$\underline{\boldsymbol{V}}_i \leqslant \boldsymbol{V}_{t,i}^{\mathbf{op}} + V_{t,i}^{\Delta} \leqslant \overline{\boldsymbol{V}}_i, \quad \forall t,i \tag{2.17}$$

$$\underline{\boldsymbol{\theta}}_{i,j} \leqslant (\boldsymbol{\theta}_{t,i}^{\mathbf{op}} + \theta_{t,i}^{\Delta}) - (\boldsymbol{\theta}_{t,j}^{\mathbf{op}} - \theta_{t,j}^{\Delta}) \leqslant \overline{\boldsymbol{\theta}}_{i,j}, \forall t,(i,j) \in N^{\mathrm{P}} \tag{2.18}$$

$$(z_e - 1) \cdot M \leqslant P_{t,e,i,j} - p_f(\psi) \leqslant (1 - z_e) \cdot M, \forall e \in E^{+} \tag{2.19}$$

$$(z_e - 1) \cdot M \leqslant P_{t,e,i,j} - p_t(\psi) \leqslant (1 - z_e) \cdot M, \forall e \in E^{\mathrm{R}+} \tag{2.20}$$

$$(z_e - 1) \cdot M \leqslant Q_{t,e,i,j} - q_f(\psi) \leqslant (1 - z_e) \cdot M, \forall e \in E^{+} \tag{2.21}$$

$$(z_e - 1) \cdot M \leqslant Q_{t,e,i,j} - q_t(\psi) \leqslant (1 - z_e) \cdot M, \forall e \in E^{\mathrm{R}+} \tag{2.22}$$

$$P_{t,e,i,j}^2 + Q_{t,e,i,j}^2 \leqslant \overline{\boldsymbol{S}}_e^2 \cdot z_e, \forall t,(e,i,j) \in (E^{+} \cup E^{\mathrm{R}+}) : \exists \overline{\boldsymbol{S}}_e \tag{2.23}$$

The objective function (2.1) minimizes total operating and investment cost over defined time period. For operating cost we use quadratic cost curve. Equations (2.2) and (2.3) represent bus balance constraints for active and reactive power. Constraints (2.4)–(2.7) represent power flow equations for existing lines which also contain second-order term approximation variables $\breve{V}_{t,e}$ and $\boldsymbol{\theta}_{t,i}^{\mathbf{op}}$ that are defined from (2.8) - (2.11). Wind generators output is limited by its maximum generating capacity as defined in (2.12), and output limits of active and reactive power for conventional generators are set in (2.13) and (2.14). Constraint (2.15) defines maximum branch apparent power flow in both directions. Equation (2.16) defines the reference bus angle value. Voltage magnitude and bus-pair angle constraints are defined in (2.17) and (2.18). Constraints (2.19)–(2.22) represent power flow equations for prospective lines, where $p_f(\psi)$, $p_t(\psi)$, $q_f(\psi)$, $q_t(\psi)$ are sequentially defined as the right-hand side of equations (2.4)–(2.7). Equation (2.23) is used to force the power flow on prospective lines to be equal to zero it the prospective line is not selected for the expansion process.

3 . Case Study

We demonstrate the accuracy of our model on the IEEE 24-bus and the IEEE 73-bus (RTS96) systems from the OPF benchmark [17]. Due to the limited amount of available network data for transmission network expansion planning, the presented grids are modified to capture different time intervals during the operating horizon. Wind power generation units are included in the system, and their active power production limits are defined for each time step, which have assigned occurrence frequency throughout the target years. Detailed input data on these modified power systems can be found in [18]. To incur congestion, conventional generator's minimum production limits are reduced by the factor of 50% and the line ratings are reduced by the factor of 20% as compared to the original values defined in [17]. Different time segments are used to account for different branch power flows that occur as a result of variable output of wind generator active power production at each time period. Therefore, during the different

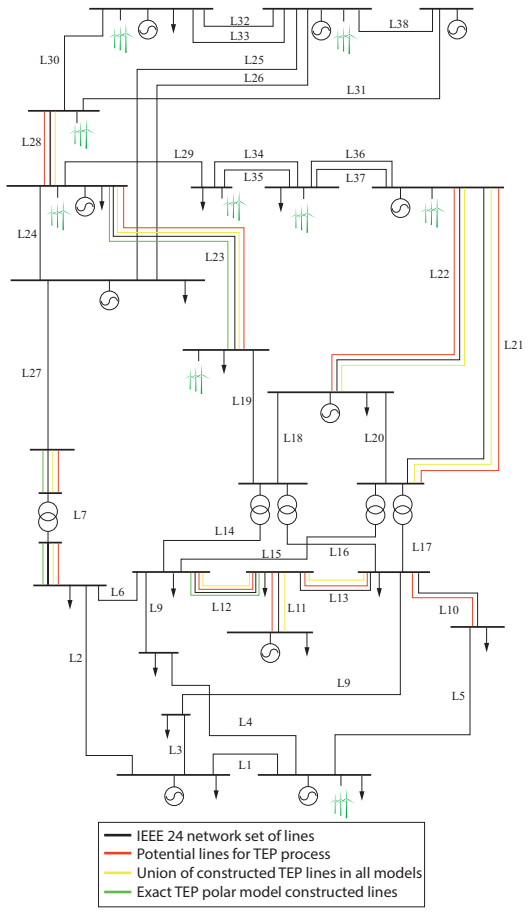

Fig. 1: IEEE 24-bus system for TEP.

time segments, different lines are identified as prospective candidates for the expansion of the transmission network. For the IEEE 24-bus system, the number of representative time steps is set to 7, and for the IEEE 73-bus (RTS96) system it is set to 5. The IEEE 73-bus (RTS96) system with 120 branches is a complex network by itself. To capture a computationally complex, yet time feasible case study, the number of prospective candidate lines for the IEEE 24-bus system with 38 branches has to be greater than for the IEEE 73-bus (RTS96) system. Set of prospective expansion branches E_i^{+} is determined after the first step of our model where the lines with the apparent flow greater than 70% of their capacity, for the IEEE 24-bus system, are defined as prospective candidates. For the IEEE 73-bus (RTS96) system, lines with the apparent flow value greater than 90% of their capacity are chosen as prospective candidates. It is assumed that at most one line is allowed to be added in each transmission corridor in the TEP process. The IEEE 24-bus system with prospective and constructed lines is visualized in Fig. 1.

The accuracy of the presented model is obtained through a four-step process. Our computation procedure starts with the exact AC polar model where the network expansion binary variables are excluded. This step provides a good Taylor expansion operating point of both the voltage magnitude and

angle, which are then used in the second step. The second step is defined as a presolve where the non-convex form of our model is run, but also without the computationally demanding binary variables. In the non-convex form, quadratic inequality constraints (2.8) and (2.10) are applied as quadratic equality constraints (1.1) and (1.2) whose marginal value signs we use to determine if the constraint would be binding if relaxed, as described in Subsection 2 -A. The third step is to run the main, full mixed-integer AC TEP using the convex approximation around the previously computed operating point, with binary variables and constraints defined in the presolve. The last step is to run the exact polar model where binary variables are considered as parameters whose values are defined in the previous step. Results of this step will determine approximation errors that were made in this model. The described procedure is itemized in Algorithm 1.

Algorithm 1 Transmission expansion planning (TEP)

1: Run exact polar model without binary variables for transmission expansion (NLP)
2: Run non-convex presolve using the operating point from the previous step, also without binary variables for transmission expansion (non-convex QCQP)
{In this step the presolve selects constraints for the main TEP computation by evaluating the constraints' marginal value}
3: Run the main TEP model around the previously computed operating point with binary variables for transmission expansion (MIQCQP)
4: Run the exact polar model with fixed binary variables to TEP solution in step 3 to determine approximation error (NLP)

Simulation results are provided in Table I and Table II, and their visual representation is provided in Fig 2. and Fig 3.

Our convex approximation (CPSOTA) is compared with the exact AC polar model, linearized AC model (LPAC) [9], linear DC model, Jabr's relaxed second-order cone programming model (JABR) [19] and piecewise-linearized AC model (LACTEP) [1].

In both test cases, the presented model by far outperforms other approximations in terms of the objective function value error and, more importantly, it is the only one that yields the correct expansion plan. Construction of the new transmission

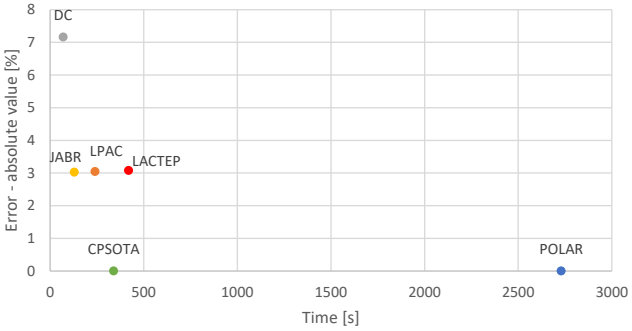

Fig. 2: Visualization of different TEP models accuracy for the IEEE 24-bus system test case.

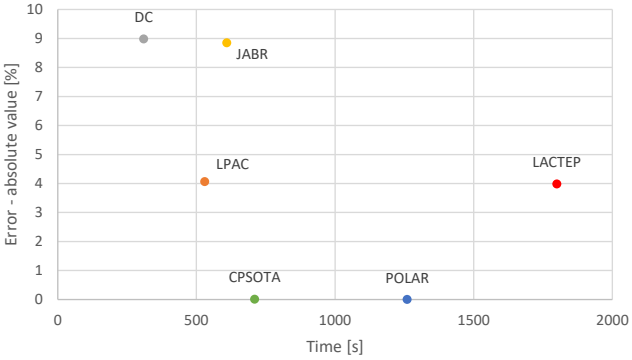

Fig. 3: Visualization of different TEP models accuracy for the IEEE 73-bus system test case.

power lines shifts the cost from operation to investment, but eventually, TEP process provides saving in total costs. Due to imprecise modeling of the reactive power flows, voltage magnitudes and power losses, the LPAC, JABR, and LACTEP models do not accurately reflect the actual solution, thus those models in both test cases result in expansion plans that consists of larger sets of newly erected lines than the exact AC polar model. The larger expansion plan of those models results in lower, realistically imprecise, total costs. Consequently, the approximation error of those models is greater than of the proposed CPSOTA model that captures the same expansion plan as the exact AC polar model. The

TABLE I
TEP RESULT COMPARISON FOR THE IEEE 24-BUS SYSTEM

Model	Time [s]	Expansion plan	Total cost	Error [%]
POLAR (MINLP)	2730	L7, L13, L23	4.0359299 E+09	0
LPAC (MIQCQP)	240	L7, L12, L13 L21, L22, L23, L28	3.9130797 E+09	-3.044
DC (MILP)	70	L12, L22, L23, L28	3.7468042 E+09	-7.164
JABR's (MISOCP)	130	L7, L12, L13 L21, L22, L23, L28	3.9138228 E+09	-3.026
LACTEP (MILP)	370	L7, L11, L12, L13 L21, L22, L23, L28	3.9116847 E+09	-3.078
CPSOTA (MIQCQP)	340	L7, L13, L23	4.0361073 E+09	0.004

TABLE II
TEP RESULT COMPARISON FOR THE IEEE 73-BUS (RTS96) SYSTEM

Model	Time [s]	Expansion plan	Total cost	Error [%]
POLAR (MINLP)	1260	L30, L90	1.390911 E+10	0
LPAC (MIQCQP)	530	L25, L53, L91, L102	1.334429 E+10	-4.061
DC (MILP)	310	L53	1.265998 E+10	-8.981
JABR's (MISOCP)	610	L30, L53 L69, L90, L91	1.267801 E+10	-8.851
LACTEP (MILP)	1800	L25, L53 L69, L90, L91	1.335497 E+10	-3.984
CPSOTA (MIQCQP)	710	L30, L90	1.390794 E+10	-0.008

DC model determines a larger set of required lines than the exact AC polar model for the expansion process of the IEEE 24-bus system, which has a large set of prospective candidates. On the other hand, the DC model for the IEEE 73-bus (RTS96) system captures a smaller set of required lines than the exact AC polar model. As expected, in both test cases the linear DC model has the highest approximation error. Computation-time-wise, the presented CPSOTA model is faster than LACTEP approximation and slower than JABR's relaxation, and DC and LPAC approximations for both test cases. The highest objective function approximation error of the CPSOTA model, in both test cases, is -0.008% as compared to the exact AC polar model. In terms of the total computation time, the CPSOTA model is 87% faster than the exact AC polar model for the IEEE 24-bus system with a larger set of prospective lines. For the IEEE 73-bus (RTS96) system, with a smaller set of prospective lines, the CPSOTA model achieves 43% faster total computation time than the exact AC polar model. The presented model is accurate around the operating point estimated by solving the exact AC polar model without binary variables for transmission expansion. The advantage of this approach is the possibility to iteratively run the main TEP model by updating the operating point and retesting the constraints in step 2 of Algorithm 1 for that new operating point. This way the approximation errors reduce even further.

4 . Conclusion

This paper utilizes the recently published convex polar second-order Taylor approximation of AC power flows [16] to deliver high modeling accuracy and tractability to the TEP. Model's accuracy is achieved by utilization of quadratically constrained voltage magnitudes and angles. In the presolve process quadratic inequality constraints, which could cause relaxation errors due to the convexification process, are identified and replaced by their linear equality constraint alternative. The method is evaluated on two modified test cases based on the PGLib-OPF benchmark [17] and compared against the existing models. The proposed model demonstrates superior accuracy at no additional computation cost, as computation times are similar to the ones achieved by using the existing approximation models. In comparison to the exact AC power flow formulations, our model shows similar accuracy and the same realistic expansion planning results, while computation time is significantly improved. The high accuracy of the presented model is desirable for further applications in more complex power systems with flexible devices, such as battery energy storage systems (BESS). The BESS will have an important role in congestion reduction, voltage control, and transition to the sustainable and secure energy system based on renewable sources. The BESS can be favorable at locations where construction of new lines is not possible, and to reduce power curtailment at locations where renewable energy sources are installed. It is also possible to coordinate TEP with the generation expansion planning (GEP) by allocating the necessary expansion investments. However, the selection of different generation units for the expansion process can affect the TEP results. The relationship between TEP and GEP can be investigated in the future work.

Acknowledgment

Employment of Karlo Šepetanc at the University of Zagreb Faculty of Electrical Engineering and Computing is funded by the Croatian Science Foundation under programme DOK-2018-09. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864298 (project ATTEST).

References

[1] H. Zhang, G. T. Heydt, V. Vittal and J. Quintero, "An Improved Network Model for Transmission Expansion Planning Considering Reactive Power and Network Losses," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3471–3479, Aug. 2013.

[2] J. Jundi, "Transmission Expansion Planning in Large Power Systems Using Power System Equivalencing Techniques," 2014.

[3] R. S. Chanda and P. K. Bhattacharjee, "Application of Computer Software in Transmission Expansion Planning Using Variable Load Structure," *Electric Power System Research*, no. 31, pp. 13–20, 1994.

[4] R. Villasana, L. L. Garver and S.L. Salon, "Transmission Network Planning Using Linear Programming," *IEEE Trans. Power Apparat. Syst.*, vol. PAS-104, pp. 349–356, Feb. 1985.

[5] L. L. Garver, "Transmission Network Estimation Using Linear Programming," *IEEE Trans. Power Apparat. Syst.*, vol. PAS-89, pp. 1688–1697, Sept./Oct. 1970.

[6] H. K. Youssef and R. Hackaam, "New Transmission Planning Model," *IEEE Trans. Power Syst.*, vol. 4, pp. 9–18, Feb. 1989.

[7] L. Bahiense, G. C. Oliveira, M. Pereira and S. Granville, "A Mixed-Integer Disjunctive Model for Transmission Network Expansion," *IEEE Trans. Power Syst.*, vol. 16, pp. 560–565, Aug. 2001.

[8] S. T. Lee, K. L. Hocks, and E. Hnyilicza, "Transmission Expansion Using Branch-and-Bound Integer Programming with Optimal Cost-Capacity Curves," *IEEE Trans. Power Apparat. Syst.*, vol. PAS-93, pp. 1390–1400, July. 1974.

[9] C. Coffrin and P. Van Hentenryck, "A Linear-Programming Approximation of AC Power Flows," *INFORMS Journal on Computing*, 2012.

[10] R. Romero, A. Monticelli, A. Garcia and S. Haffner, "Test Systems and Mathematical Models for Transmission Network Expansion Planning," *IEE Proc., Gen. Trans. Distrib.*, vol. 149, (1), pp. 27–36., 2002.

[11] Z. Luburić, H. Pandžić and M. Carrion, "Test Systems and Mathematical Models for Transmission Network Expansion Planning,"Transmission Expansion Planning Model Considering Battery Energy Storage, TCSC and Lines Using AC OPF," *IEEE Access*, vol. 8, pp. 203429–203439., Jan. 2020.

[12] M. J. Rider, A. V. Garcia and R. Romero, "Power System Transmission Network Expansion Planning Using AC Model," *IET Gener. Transm. Distrb.*, vol. 1, no. 5, Sep. 2007.

[13] R. Bent, G. L. Toole and A. Berscheid, "Transmission Network Expansion Planning With Complex Power Flow Models," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 904–912, May. 2012.

[14] R. Bent, C. Coffrin, R. Gumucio and P. Van Hentenryck, "Transmission Network Expansion Planning: Bridging the Gap between AC Heuristics and DC Approximations," *Power Systems Computation Conference, Wroclaw*, 2014., pp. 1-8.

[15] G. Latorre, R. D. Cruz, J. M. Areiza and A. Villegas, "Classification of Publications and Models on Transmission Expansion Planning," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 938–946, May. 2003.

[16] K. Šepetanc and H. Pandžić, "Convex Polar Second-Order Taylor Approximation of AC Power Flows: A Unit Commitment Study," *IEEE Trans. Power Syst.*, Dec. 2020, Early Access.

[17] C. Coffrin *et al. PGLib-OPF*, GitHub, July 29, 2020. Accessed on: Jan. 9, 2021. [Online] Available at: github.com/power-grid-lib/pglib-opf/

[18] H. Pandžić, Y. Dvorkin, T. Qiu, Y. Wang and D. S. Kirschen, "Unit Commitment Under Uncertainty - GAMS Models," *REAL Lab Library*, University of Washington, Seattle, WA, USA. Accessed: March. 2021., [Online] Avalilable at: http://labs.ece.uw.edu/real/gams_code.html

[19] D. K. Molzahn and I. A. Hiskens, "A Survey of Relaxations and Approximations of the Power Flow Equations," pp. 75–77, 2019.

Biography

Karlo Šepetanc is an assistant at the Department of Energy and Power Systems at the University of Zagreb Faculty of Electrical Engineering and Computing. His employment has been funded by the Croatian Science Foundation under programme DOK-2018-09. He is also a researcher at the Innovation Centre Nikola Tesla (ICENT). He received his bachelor's and master's degrees in electrical engineering and communication technology in July 2016 and September 2018, respectively.

His research interests include planing, operation and economics of power and energy systems with an emphasis on optimization's computational tractability. Karlo is involved in teaching activities as a teaching assistant in the Bachelor Programme course Integration of Renewable Energy Sources in Power Systems and the Master Programme course Electricity Markets. He supervises bachelor and master student thesis progress.

Karlo is a member of professional associations IEEE and CIGRE.

Journal Papers

[J_1]D. Vlah, K. Šepetanc and H. Pandži ć, "Solving Bilevel Optimal Bidding Problems Using Deep Convolutional Neural Networks," *IEEE System Journal*, 2023,
ISSN: 1937-9234, DOI: 10.1109/JSYST.2022.3232942

[J_2]K. Šepetanc, H. Pandži ć and T. Capuder, "Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part I: Model Description and the Algorithm," *IEEE Transactions on Power Systems*, 2022,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2022.3207088

[J_3]K. Šepetanc, H. Pandži ć and T. Capuder, "Solving Bilevel AC OPF Problems by Smoothing the Complementary Conditions – Part II: Solution Techniques and Case Study," *IEEE Transactions on Power Systems*, 2022,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2022.3207097

[J_4]K. Steriotis, K. Šepetanc, K. Smpoukis, N. Efthymiopoulos, P. Makris, E. Varvarigos and H. Pandžić, "Stacked Revenues Maximization of Distributed Battery Storage Units Via Emerging Flexibility Markets," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 1,

pp. 464-478, Jan. 2022,
ISSN: 1949-3037, DOI: 10.1109/TSTE.2021.3117313

[J_5]K. Šepetanc and H. Pandži ć, "A Cluster-Based Model for Charging a Single-Depot Fleet of Electric Vehicles," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3339-3352, July 2021,
ISSN: 1949-3061, DOI: 10.1109/TSG.2021.3064272

[J_6]K. Šepetanc and H. Pandži ć, "Convex Polar Second-Order Taylor Approximation of AC Power Flows: A Unit Commitment Study," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3585-3594, July 2021,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2020.304697

[J_7]K. Šepetanc and H. Pandži ć, "A Cluster-Based Operation Model of Aggregated Battery Swapping Stations," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 249-260, Jan. 2020,
ISSN: 1558-0679, DOI: 10.1109/TPWRS.2019.2934017

Conference Papers

[C_1]N. Čović, D. Badanjak, K. Šepetanc and H. Pandžić, "Cost Sensitivity Analysis to Uncertainty in Demand and Renewable Energy Sources Forecasts," in *2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON)*, 2022, pp. 860-865,
10.1109/MELECON53508.2022.9842933

[C_2]M. Gržani ć, K. Šepetanc, T. Capuder, M. Usman and F. Capitanescu, "Attest project: tools for ancillary service procurement in day-ahead operation and real-time activation in distribution grids," in *CIRED Porto Workshop 2022: E-mobility and power distribution systems*, 2022, pp. 57-61,
10.1049/icp.2022.0661

[C_3]O. Heide, K. Šepetanc, and H. Pandži ć, "Transmission Expansion Planning using a Highly Accurate AC Optimal Power Flow Approximation," in *2021 International Conference on Smart Energy Systems and Technologies (SEST)*, 2021, pp. 1-6,
10.1109/SEST50973.2021.9543431

Životopis

Karlo Šepetanc je asistent na Zavodu za visoki napon i energetiku Sveučilišta u Zagrebu Fakulteta elektrotehnike i računarstva. Njegovo zaposlenje financira Hrvatska zaklada za znanost u sklopu programa DOK-2018-09. Također je istraživač u Inovacijskom centru Nikola Tesla. Prvostupnik elektrotehnike i informacijske tehnologije postao je u lipnju 2016. godine, a diplomirao je u rujnu 2018. godine na Sveučilištu u Zagrebu Fakultetu elektrotehnike i računarstva.

Njegovi istraživački interesi obuhvaćaju planiranje, vođenje i ekonomiku elektroenergetskog sustava s naglaskom na računalnu izračunljivost optimizacija. Karlo sudjeluje u nastavi preddiplomskog studija na predmetu Integracija obnovljivih izvora energije u energetske sustave i na diplomskom studiju na predmetu Tržište električne energije. Također nadzire napredak studenata preddiplomskog i diplomskog studija tokom završnog i diplomskog rada.

Karlo je član stručnih udruga IEEE i CIGRE.