THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Optimization of low-cost integration of wind and solar power in multi-node electricity systems

Mathematical modelling and dual solution approaches

Caroline Granfeldt

Department of Mathematical Sciences
Chalmers University of Technology
Gothenburg, Sweden 2023

This work was supported by
the Swedish Energy Agency (project number 39907-1) and
Chalmers University of Technology

Optimization of low-cost integration of wind and solar power in multi-node electricity systems
Mathematical modelling and dual solution approaches
Caroline Granfeldt
Gothenburg 2023
ISBN 978-91-7905-892-0

© Caroline Granfeldt, 2023

Doktorshavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5358
ISSN 0346-718X

Department of Mathematical Sciences
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31 772 1000

Typeset with LaTeX
Printed by Chalmers digitaltryck
Printed in Gothenburg, Sweden, 2023

Optimization of low-cost integration of wind and solar power in multi-node electricity systems

Mathematical modelling and dual solution approaches

Caroline Granfeldt

Department of Mathematical Sciences
Chalmers University of Technology

Abstract

The global production of electricity contributes significantly to the release of CO_2 emissions. Therefore, a transformation of the electricity system is of vital importance in order to restrict global warming. This thesis concerns modelling and methodology of electricity systems which contain a large share of variable renewable electricity generation (i.e. wind and solar power).

The two models developed in this thesis concern optimization of long-term investments in the electricity system. They aim at minimizing investment and production costs under electricity production constraints, using different spatial resolutions and technical detail, while meeting the electricity demand. These models are very large in nature due to the 1) high temporal resolution needed to capture the wind and solar variations while maintaining chronology in time, and 2) need to cover a large geographical scope in order to represent strategies to manage these variations (e.g. electricity trade). Thus, different decomposition methods are applied to reduce computation times. We develop three different decomposition methods: Lagrangian relaxation combined with variable splitting solved using either i) a subgradient algorithm or ii) an ADMM algorithm, and iii) a heuristic decomposition using a consensus algorithm. In all three cases, the decomposition is done with respect to the temporal resolution by dividing the year into 2-week periods. The decomposition methods are tested and evaluated for cases involving regions with different energy mixes and conditions for wind and solar power. Numerical results show faster computation times compared to the non-decomposed models and capacity investment options similar to the optimal solutions given by the latter models. However, the reduction in computation time may not be sufficient to motivate the increase in complexity and uncertainty of the decomposed models.

Keywords: variable renewable electricity, variation management, electricity system modelling, capacity expansion, cost optimization, Lagrangian relaxation, variable splitting, subgradient algorithm, ADMM, consensus algorithm.

List of publications

This thesis is based on the work represented by the following papers:

Paper I **Granfeldt, C.**, Strömberg, A.-B., and Göransson, L. (2023)
A Lagrangian relaxation approach to an electricity system investment model with a high temporal resolution
Under review for OR Spectrum after minor revision

Paper II **Granfeldt, C.**, Strömberg, A.-B., and Göransson, L. (2023)
Managing the temporal resolution in a multi-node electricity system investment model – Parallel computations by variable splitting and Lagrangian relaxation
Manuscript

Paper III **Granfeldt, C.**, Strömberg, A.-B., and Mattsson, N. (2023)
An approximate consensus ADMM approach to a multi-node electricity system investment problem with a high temporal resolution
Manuscript

Paper IV **Göransson, L.**, Granfeldt, C., and Strömberg, A.-B. (2021)
Management of wind power variations in electricity system investment models – A parallel computing strategy.
SN Operations Research Forum, **2**:25,
doi: 10.1007/s43069-021-00065-0.

Author contributions

Paper I I developed the main algorithm, formulated the model and decomposition approach, implemented the code, ran the experiments, performed the analysis and wrote the manuscript. The co-authors contributed with ideas, provided insights of the results and reviewed the final manuscript.

Paper II I developed the extension of the model and the main algorithm, derived the extended decomposition, implemented the code, ran the experiments, performed the analysis and wrote the manuscript. The co-authors contributed with ideas, provided insights of the results and reviewed the final manuscript.

Paper III I developed the model, formulated the ADMM decomposition, implemented the code, ran the experiments, performed the analysis and wrote the manuscript. Strömberg contributed with ideas for the approach and reviewed the manuscript. Mattsson provided the initial idea to use ADMM and contributed with programming expertise.

Paper IV Göransson developed the model and decomposition approach, implemented the model, ran the tests, and wrote the manuscript. I aided with formulating the mathematical model, improved the presentation and reviewed the final manuscript.

Acknowledgements

There are many people that have accompanied me on this journey, both inside and outside of academia. First and foremost, my deepest gratitude goes to the extraordinary **Ann-Brith Strömberg**, my supervisor at the Department of Mathematical Sciences at Chalmers University of Technology. Thank you for all your encouragement, immense support, amazing attention to details and for *always* taking the time for discussions and feedback. With your vast knowledge and skill, you truly are a fantastic supervisor and have been a wonderful role model for me throughout the years, both professionally and personally. In short, your mentorship has been an experience that I will always cherish in my heart. Thank you for everything!

I would also like to thank everyone at the mathematical department for a pleasant work environment, especially some current and former members of the optimization group: **Gabrijela Obradović**, **Quanjiang Yu**, **Zuzana Nedělková**, **Sunney Fotedar**, **Edvin Åblad**, **Axel Ringh**, **Emil Gustavsson**, **Efraim Laksman**, and **Kin Cheong Sou**. Thank you **Michael Patriksson** for your cheerful spirit and many enjoyable corridor talks. It has been a privilege to know you all and to call you my friends and colleagues.

Further, a huge thank you to my co-supervisor **Lisa Göransson** at the Department of Space, Earth and Environment, Chalmers University of Technology. You have provided me with tremendous support and have helped me grow with both your professional expertise and valuable input, as well as all your advice on a personal level. I also wish to thank everyone in the energy systems group for your support and interesting discussions. Thank you **Lina Reichenberg**, **Viktor Walter**, and **Jonathan Ullmark** for answering all my novice questions about the electricity system and interpretation of data. Thank you **Niclas Mattsson** for your enthusiasm and advice on modelling implementation. I also wish to thank **Filip Johnson** for your contributions to the project application.

I moreover wish to express my appreciation to colleagues at the Department of Mathematics at Linköping University of Technology, especially **Torbjörn Larsson** for raising my very first interest in mathematical optimization. Thank you **Elina Rönnberg** and **Nils-Hassan Quttineh** for your encouragement throughout the years.

Thank you also to everyone I've met and gotten to know through SOAF activities.

I'm not sure I would have managed to accomplish this thesis without the support from my wonderful friends and family. Thank you **Gabriella**—my bosom friend, partner in

crime and the sister I never had—and my favourite teacher and "second mom" **Marie** for being there for me when I felt lost. Thank you **Leo**, **Stefan**, and **Oscar** for the fun times we had in Linköping during my undergraduate studies. And a big thank you to all my friends in the Gothenburg area, especially **Arya**, **Therese**, and **Emina**, who made my life here much more enjoyable.

Thank you **Mariette** and **Bo**, fondly known as mom and dad, for raising me to be who I am and encouraging me to follow my dreams. I thank **Christofer**—the best brother and popcorn-maker in the world—and your fantastic family **Alexandra**, **Aprilia**, and **Fabian**, for your support and for always giving me something to look forward to. I also wish to thank all my family and relatives, too many to mention, who have provided me with joy, love and laughs throughout my life. I appreciate all of you!

Finally, I would like to thank my husband and very best friend in the world, **Oscar**, and our wonderful children **Viola** and **Agnes**. You are the sunshine of my life, the adventure that never ends, and I would have never managed this journey without you. Thank you Oscar for your endless support, patience and love. And as sung by Marvin Berry and the Starlighters, I love you forever, and ever more. I'm just a fool, a fool in love with you.

Caroline Granfeldt
Mölndal 2023

"1.21 gigawatts. Great Scott!"

"What the hell is a gigawatt?"

EMMETT "DOC" BROWN
& MARTY MCFLY,
Back to the Future, 1985

Contents

Abstract **iii**

List of publications **v**

Acknowledgements **vii**

Contents **xi**

1 Introduction **1**

1.1 Purpose and Aims 2

1.2 Limitations 2

1.3 Outline 2

2 The Electricity system and modelling **3**

2.1 Electricity generation technologies 3

2.2 Energy storage technologies 4

2.3 Electricity system modelling 5

2.4 Variations in the electricity system 7

2.5 Previous work on electricity system modelling 12

3 Mathematical modelling **15**

3.1 Full-scale electricity system investment model 17

3.2 Hours-to-Decades model 27

3.3 Comparison of the different models 34

4 Mathematical methods and their theory **35**

4.1 Linear programming 35
4.2 Lagrangian dual concepts 36
4.2.1 Lagrangian relaxation 36
4.2.2 Augmented Lagrangian 37
4.2.3 Variable splitting 38
4.3 Lagrangian dual solution methods 43
4.3.1 Subgradient algorithm 44
4.3.2 Alternating direction method of multipliers (ADMM) . . 51
4.4 Heuristic consensus algorithm 57

5 A summary of the appended papers 63

Paper I: A Lagrangian relaxation approach to an electricity system investment model with a high temporal resolution. 64

Paper II: Managing the temporal resolution in a multi-node electricity system investment model: Parallel computations by variable splitting and Lagrangian relaxation 65

Paper III: An approximate consensus ADMM approach to a multi-node electricity system investment problem with a high temporal resolution 66

Paper IV: Management of wind power variations in electricity system investment models: A parallel computing strategy 67

6 Conclusions and ideas for future research 69
6.1 Additional reflections 70
6.2 Future perspective 71

Bibliography 73

Appendices 79

A Nomenclature 79
A.1 Full-scale model 79
A.2 Hours-to-Decades model 83

B Dual multiplier projection examples 87

1 Introduction

EU's roadmap 2050 establishes that the greenhouse gas emissions must decrease by some 85% until the year 2050 in order for global warming to be restricted to 2°C [9]. The electricity system contributes significantly to the emissions of carbon dioxide, both in the EU and globally (see for example Ritchie and Roser [49] and underlying data sources). A transformation of the electricity system is therefore needed, and electricity investment models can be used as a tool to make informed decisions regarding future electricity generation, storage, and transmission capacity. The mathematical optimization models describing the electricity system minimize the investment and production costs of the system under electricity production constraints, while meeting the electricity demand.

The existing systems mostly consist of thermal power [31], and thus the traditional models are designed with this in mind. The characterization of such a system, dominated by dispatchable generation, include the ability to regulate the electricity production to meet instant demand. However, a cost-efficient reduction of greenhouse gas emissions from the electricity system is expected to imply a large-scale implementation of varying renewable electricity generation (VRE), such as wind and solar power. To be able to capture the variability in electricity generation from VRE, a realistic mathematical modeling of future electricity systems must include a fine discretization of time [32]. Furthermore, one key strategy to reduce variability of wind power is geographical smoothing through trade. Thus, it is desired to consider a large geographical area in the modeling of the electricity system while accounting for the transmission bottlenecks within such an area.

There is, however, a conflict between a high temporal and spatial resolution and reasonable computing times for electricity system models. For real problem instances on the European scale, the challenge lies in finding this proper balance in the mathematical modelling.

1.1 Purpose and Aims

The purpose of the project is to formulate and analyze mathematical optimization models that capture strategies to manage the variability of variable renewable electricity generation. The research focuses on techniques and methodologies for decomposing and solving long-term investment models, i.e. capacity expansion models, including investments in electricity generation, energy storage and transmission capacity. A key objective is to examine how the temporal and spatial resolution impacts the solution times using these decomposition methods, and—when possible—compare the solutions and computation times of the decomposed models to the non-decomposed model optimal solutions.

1.2 Limitations

The models are used to examine long-term capacity investments over large regions (i.e. entire countries), and are not suited to make decisions regarding how separate power plants should operate since the models use an aggregated continuous electricity generation and storage capacity for each region. Furthermore, electricity transmissions within each region are not considered in detail. We assume a perfect forecast of electricity demand and weather. Thus, there is no stochasticity in the models and they are purely deterministic. Demand and weather profiles in terms of, for example, wind speed and water inflow (from rain and melted ice) to hydropower turbines are based on data from previous years. Electricity used for heating, and CHP plants (combined heat and power) which deliver both electricity and heat, are not included in this work.

1.3 Outline

The remainder of this thesis is organized as follows. In Chapter 2, the electricity system and some of its modelling difficulties are described, along with a review of some previous modelling work. Chapter 3 presents and compares two different mathematical optimization models, including variables, constraints and objective functions. In Chapter 4, the scientific areas and methods used to solve these models are presented. A summary of the appended papers are given in Chapter 5, and then finally Chapter 6 discusses conclusions and the main contributions of this thesis, as well as topics for future research.

2 The Electricity system and modelling

This chapter discusses the electricity system and some modelling difficulties that comes with it. It also gives an overview of some previous work done on electricity system modelling.

Electricity is an energy carrier which can be characterized using different properties, e.g. voltage, current, energy, or power. In an electricity system, electric energy is produced in power plants and then transferred to electricity consumers connected to an electrical grid. Each power plant has an installed and available production capacity, typically measured in GW, which controls the amount of electric power that can be produced at any instant. (Specifically, one watt is defined as one joule per second and thus measures the rate of transfer.) The electrical energy produced in a power plant is often measured in GWh, and is thus the product of the power in gigawatts multiplied by the running time in hours.

The electricity system inside a region typically consists of different types of sources for electricity production. These can for example be coal power, nuclear power, hydropower, wind power, solar power, or natural gas turbines.

2.1 Electricity generation technologies

Thermal power plants are, as the name indicates, power plants where heat is converted to electricity. In steam turbines, water is heated to steam which is then used to rotate turbines and generate electricity. For gas turbines, the gas is combusted directly and expanded over the turbine to rotate it. Some different fuels used as heat sources are fossil fuels, nuclear energy, biofuels,

and waste incineration. Some thermal power plants are combined to generate both electricity and heat (e.g. district heating) to consumers, but in this thesis only electricity is considered. The concept of *thermal cycling* refers to generating electricity at different demand levels. As the demand for electricity varies, some electricity generating units need to be turned on/off in response to these variations. However, every time a thermal power plant is turned off and on, the different components (e.g. boiler, steam lines, turbine) are exposed to stress caused by the large thermal and pressure variations, which then leads to maintenance costs.

In a *hydropower* plant, electricity is generated by leading water through turbines. The power extracted depends on the water volume and the height difference between the water's in- and outflow.

Wind power generates electricity by using the wind to provide mechanical power through turbines. Wind power is typically divided into onshore and offshore wind power. Investment and maintenance costs for offshore wind power are higher compared to onshore wind power [41], but offshore wind is stronger and steadier compared to onshore wind.

Solar power is the technique to convert the energy from sunlight into electricity. This can be done by using photovoltaics (PV) or concentrated solar power (CSP). PVs use solar panels which contain photovoltaic cells that convert light into an electric current by the use of the photovoltaic effect. CSP is a technique which uses lenses or mirrors to concentrate sunlight into a small beam, and then use the resulting heat to generate electricity from steam turbines.

2.2 Energy storage technologies

Energy storage technologies can, as the name suggests, be used to store energy for later use. This is especially useful for variable renewable energy sources, since this flexibility allows production to shift according to availability of the resource and the demand for electricity.

Battery storage systems mostly consist of lithium-ion batteries and are typically employed for sub-hourly, hourly and daily balancing of the electricity grid. The relatively high cost of energy storage capacity however limits the applicability of batteries to manage variations of long duration.

Hydrogen storage uses an electrolysis process, which is based on using electricity to split water into hydrogen and oxygen. This process takes place inside an electrolyzer. Unlike batteries, it is possible to use hydrogen storage as a

seasonal storage which implies that it can store energy in one season and then discharge in another when there is high demand. There is, however, a low overall efficiency for storing and regenerating electricity from hydrogen. Hence, this thesis does not consider the use of fuel cells to reproduce electricity; hydrogen is in this work mainly used as a *demand-side management strategy*. Shortly put, investing in hydrogen allows the demand for electricity to be moved in time (see also Section 2.4 for a discussion on variation management). There is currently a growing demand for hydrogen from industry; e.g. in steel production to remove oxygen from the iron ore. Previously, carbon and coke have been used for this process and thereby producing carbon dioxide as a byproduct, while the use of hydrogen only generates water vapor; see also the project HYBRIT [30].

Hydropower storage in terms of reservoirs are also included in this work. We do not consider *pumped hydropower* systems though, where excess energy is used to pump water into reservoirs for later use. Dependent on the reservoir capacity and other environmental and location constraints, pumped hydropower can store its energy up to a week.

2.3 Electricity system modelling

The modelling of the electricity system can be done on different system levels, which vary with the types of questions that are asked. For example, the unit commitment problem studies how a set of electricity generators (e.g. power plants in a country, or turbines inside a power plant) should operate (i.e. dispatch electricity) in order to meet the demand at the lowest system cost (or highest revenue). Investment models (i.e. capacity expansion models) look at cost-efficient investments in new capacity to meet future demand. These models typically consider a longer time horizon compared to the unit commitment problem, but instead lack in system details. This thesis addresses investment models, and therefore the focus here will be on those types of models.

In general, the investment models vary in scope and resolution in three different dimensions: temporal, spatial, and technological system detail. Figure 2.1 gives an illustration of the different model complexities, where a larger volume of the cube indicates a larger, and typically more complex, problem. The models tend to become very large if all dimensions have a high resolution/detail, and therefore electricity system investment models typically make resolution sacrifices in at least one dimension.

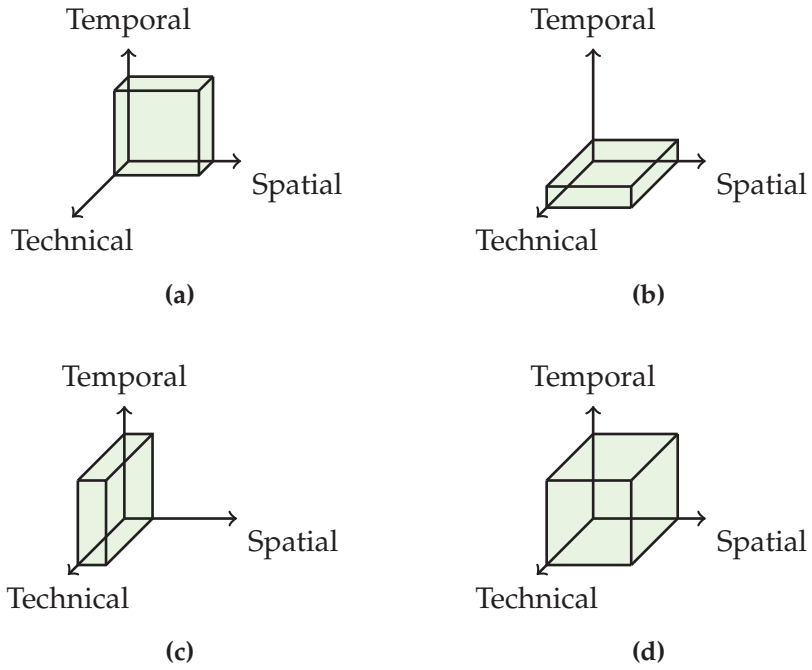


Figure 2.1: Illustrations of the problem complexity depending on the temporal resolution, spatial resolution, and technological system detail.
a) High temporal and spatial resolution, but low technological system detail.
b) High spatial resolution and technological system detail, but low temporal resolution.
c) High temporal resolution and technological system detail, but low spatial resolution.
d) High temporal and spatial resolution, and high technological system detail.

The differences in temporal scope is related to the size of the time step, and the length of the time horizon. For models with a large share of VRE, a fine time resolution is needed in order to capture the variations, and thus the size of the time step needs to be in magnitude of a few hours. The time horizon typically ranges from a single year to a longer time span, for example all years up until 2050. The longer time horizon can be motivated by the lifetime of power plants, which is approximately 40 years for thermal power plants. The spatial scope can be everything from the unit commitment problem on a single power plant, to models which contain several countries. For the latter case, the spatial scope relates to the size of the interconnected transmission grids. Typically, for a high spatial resolution to be feasible, some simplifications in the technological system details need to be done. As mentioned earlier, electricity is produced by different electricity generation technologies, e.g. coal power, hydropower, and wind power. The production capacity, measured in GW, determines an upper limit for how much electricity can be produced during a time instant. Instead of looking at separate power plants, assuming an aggregated capacity in each region will cause a loss of some system detail, but reduce the problem size significantly. Previous work by Göransson [18] has, however, shown that the loss is marginal for the total system cost. The author also showed that the loss is marginal for the average full load hours for each electricity production type, including wind power. In terms of technological system details, besides using aggregated capacity, system details typically vary with the constraints included in the model. For example, using different types of storages such as batteries or hydrogen, which connect several time steps with each other, increases the model complexity. Also, hydropower connects over several time steps and is in that sense similar to storage constraints. Another complicating feature is caused by including thermal cycling in the modelling since this typically is modelled by integer variables (that are also connected over time). As will be seen in Section 3.1, however, this latter complication can be linearized to reduce complexity [61].

2.4 Variations in the electricity system

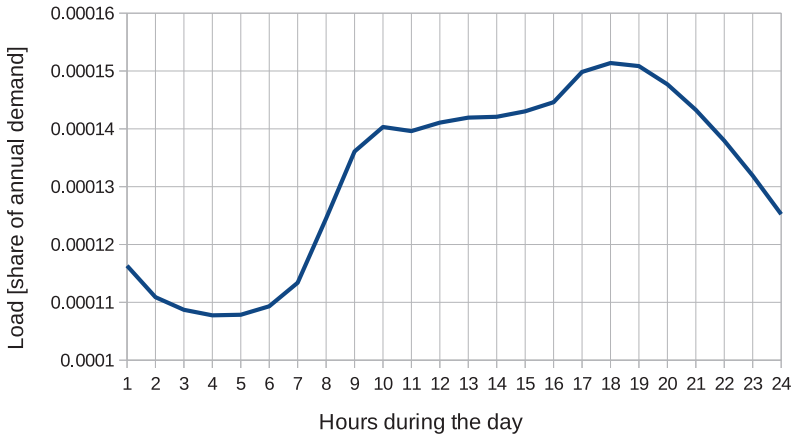
The variation of electricity demand is regular and related to our behavior as consumers. It can be divided into three main categories: seasonal variations, weekly variations, and diurnal variations. The seasonal variations has an annual cycle, where the demand varies throughout the year. In the northern European countries, the demand is higher during the darker and colder seasons due to electric heating, and lower in the warmer seasons when heating is not necessary. The contrary holds true in southern Europe, where the demand

is higher during the summer months due to the need for air conditioning in buildings. Moreover, less electricity is typically used during the nights compared to the days, and thus the demand also follows a diurnal cycle. However, the pattern of electricity use on weekends compared to workdays also differs, and therefore a weekly cycle exists. Figure 2.2 shows the electricity demand variations according to the described cycles.

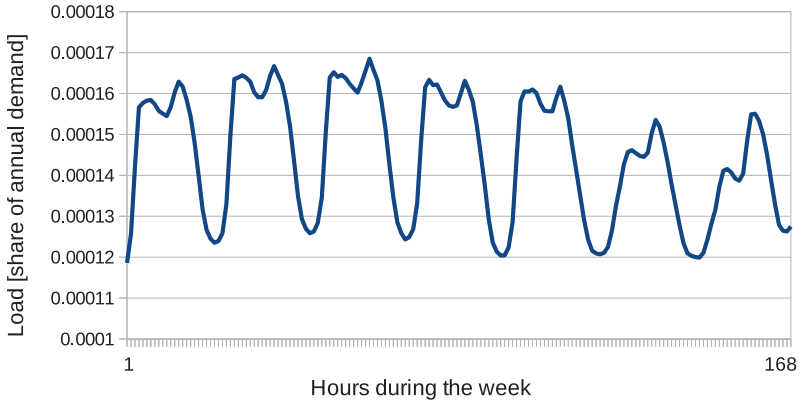
Traditionally, the electricity demand (usually referred to as load) is divided up in three different groups: base load, intermediate load, and peak load. Figure 2.3 demonstrates a load duration curve, where the load in a region has been sorted over all the hours of the year, from highest to lowest demand. The base load is consistent throughout the year, and typically the electricity production technologies that satisfy it run at full capacity during all hours. These production technology types generally have high investment costs, but low running costs. Examples include nuclear power and steam-engined power run by fossil fuels (e.g. coal). The peak load period has a significantly higher demand compared to the average load level and is fulfilled by for example gas turbines, i.e. technologies that are expensive to operate but have comparatively low investment costs. Intermediate load corresponds to the period in between base load and peak load, and can for example be covered by combined cycle gas turbines, which combine several heat engines that all use the same source of heat, e.g. natural gas or biogas. The engines then work in tandem which allows them to extract heat energy from each other. These types of plants are better at following the load curve changes compared to base load production technologies.

The electricity system has historically been designed to meet the above mentioned load variations. However, variable renewable energy (VRE) sources such as wind and solar power are intermittent and non-dispatchable. This means that they are not always available as they depend on factors which can not be controlled. These factors include the weather and the location of wind turbines and solar panels, and thus different regions have different conditions. Nonetheless, while the wind variations are irregular, they can under some conditions still be fairly slow. If a large geographical scope (e.g. a group of wind farms, or all wind turbines in a country) is considered, there can be several days of high wind power production and then several days of low wind power production [28]. Hence, a key strategy to manage variations from wind power includes a large geographical scope, and thus also electricity trade between regions to smoothen the effect of wind variability.

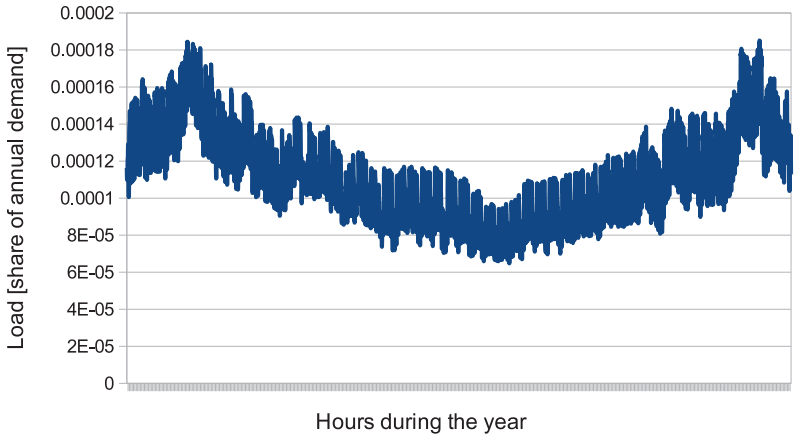
Figure 2.4 illustrates an example of the wind power production at different wind penetration levels during a week. For the lower level case, wind power will start to compete with base load production during times when the electric-



(a) Diurnal variations of the electricity demand in a region in Sweden.



(b) Weekly variations of the electricity demand in a region in Sweden.



(c) Seasonal variations of the electricity demand in a region in Sweden.

Figure 2.2: Load variations on different time scales.

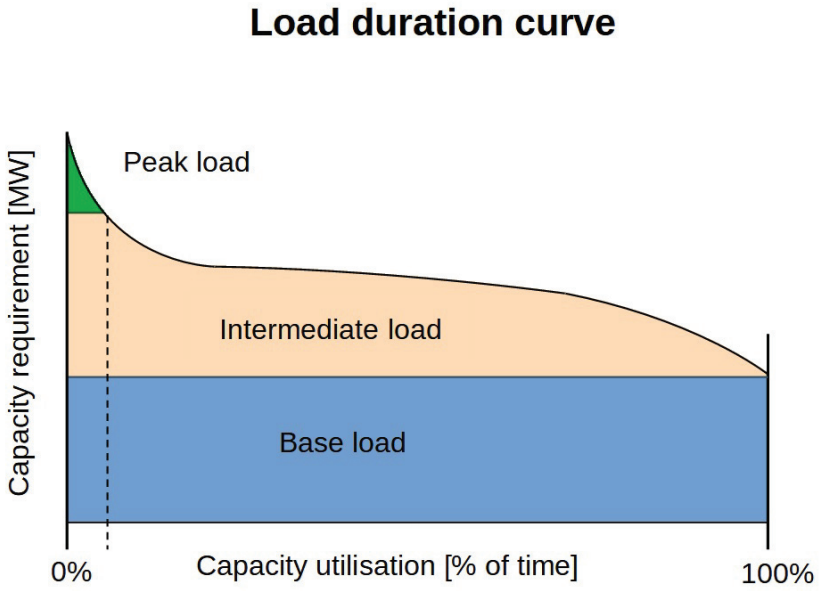


Figure 2.3: A load duration curve to illustrate how base, intermediate and peak load are used in the electricity system. The curve corresponds to the annual load, sorted according to the size of the load.

ity demand is low. As discussed earlier, it is expensive to modify the output for base load production and thus wind power production will likely be *curtailed* during those hours. This means that the output from wind power production is deliberately decreased (even though more could have been produced) in order to balance the electricity supply and demand. If instead the case when the wind penetration level is higher is considered, there will be situations when the wind power production meets a larger share of the demand. In some situations, it could meet a large share of the electricity demand for a long time. This implies that there are longer periods when base load generation is not needed in the electricity system. Curtailing all the wind power production would in that case be too costly, and the alternative is then to turn off the base load production during these time periods. Furthermore, for the shorter time periods when demand can't be met entirely by wind power production only, it is not favorable to use base load production technologies. The reason for this is that since base load production is costly to invest in, and also expensive to start up, it is not cost-efficient to only use it during bursts of a few days a time. Instead, during these time periods intermediate and peak load production types will likely be used, since they better complement the wind power production pattern. However, to manage variations such as these in the long and short term requires the solution of an optimization problem which should be captured by the electricity system investment models.

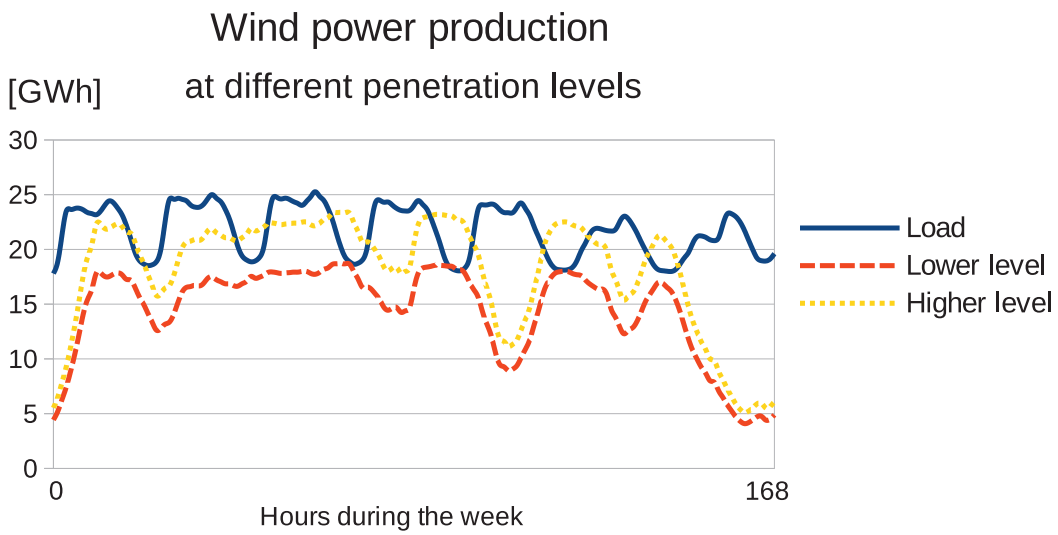


Figure 2.4: Example of wind power production at different penetration levels in a load curve during a week.

Solar power, just like the load, has regular variations that directly relates to the sunlight such that the production is highest in the middle of the day. The amplitude of the solar power production varies between days, but the general production pattern coincides with the demand curve; see Figure 2.5. This means that for low solar power penetration levels, it can replace peak load production. However, for higher penetration levels the solar power starts to compete with the base load production during the day. It would not be cost-efficient to only run base load production during the night, and therefore intermediate load production types would typically be used for these time periods. Moreover, the solar peak production is more narrow than the demand peak. This implies that when the sun is setting in the afternoon, the demand for electricity is still high and thus other electricity production need to cover that peak demand.

The operation of power plants in response to variations is a variation management strategy. It should be noted, however, that other variation management strategies for VRE integration exist. Typically, these can be found on the demand side, the production side, and by the use of energy storage units. In terms of the latter, batteries or hydrogen storage can be used to store energy when production is high, and discharged at a later point when production is lower. For example, storage units could be used to complement solar power to manage the peak demand in the afternoons. This flexibility also helps to avoid curtailing electricity since excess electricity from VRE production can be stored for later use. On the demand side, load shifting can be used to shift some of the

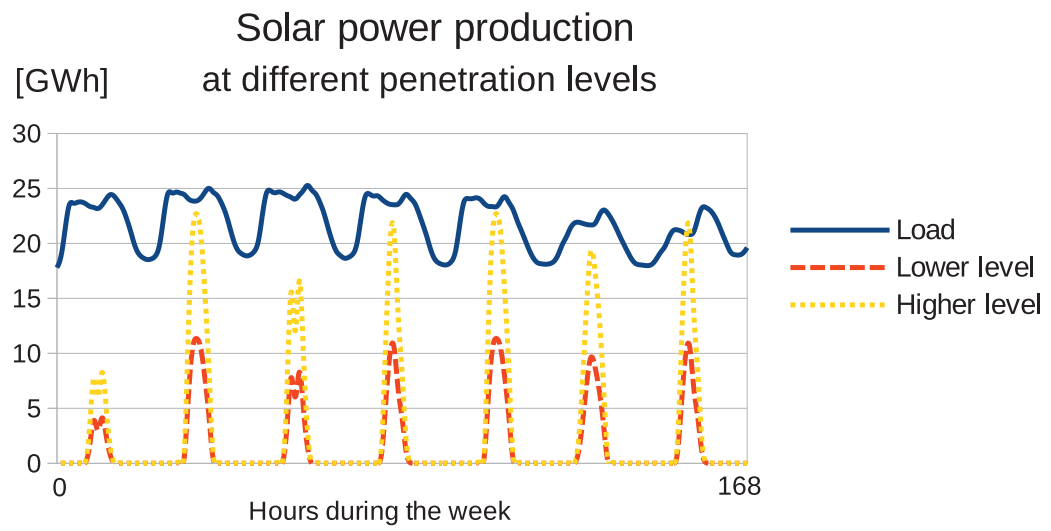


Figure 2.5: Example of solar power production at different penetration levels in a load curve during a week.

load from hours with high demand to hours when there is less demand. This could for example be smart charging of electric vehicles, which means that the vehicles are charged (or discharged during peak load) when it is beneficial to the system in terms of load and electricity production. Peak load generation or reduced base load generation are examples of supply side variation management. Some of the mentioned variation management strategies have been evaluated by van Ackooij et al. [60] using multi-objective optimization for costs and emissions on energy systems comprised of different energy mixes.

In conclusion, incorporating a high share of VRE sources, such as wind or solar power, into the energy mix requires additional variation management strategies compared to a system with a low share of VRE sources. For this to be successful, it is likely necessary to step away from the traditional way of designing the electricity system.

2.5 Previous work on electricity system modelling

In general, the computational requirements can be classified into two categories such that they are either 1) solver-based (mathematical optimization of the solving process) or 2) model-based (simplification of the real-world problem in the model) [35]. A common solver-based approach is to use a two-step simulation, where the first step is to determine a cost-effective pathway for capacity

investments. The second step is then to optimize the dispatch assuming the investments determined in the first step. The method allows highly detailed systems where the temporal and spatial scope are extensive; see for example Breyer et al. [4]. However, the drawback of this two-step approach is that it reduces the accuracy of the results. Different decomposition methods have also been employed for large-scale capacity expansion planning. For instance, Lara et al. [36] consider a Benders decomposition algorithm, where at each iteration a Benders master problem determines the investments decisions, and then each Benders subproblem (one for each year during the planning horizon) is solved separate.

Other work aiming to reduce computation time and computer memory requirements are model-based as they focus on simplifying the time representation itself. Ringkjøb et al. [48] have reviewed 75 different modelling tools used for analysing energy and electricity systems with large shares of VRE. The authors identify one of the remaining challenges: how to represent short-term variability in long-term studies. A methodological review of strategies to integrate short-term variations is given by Collins et al. [8], where the authors discuss methods to improve the time representation in long-term electricity system investment models that use the traditional ways of time representation. Pfenninger et al. [44] review several articles that discuss time representation for energy system models which contain a substantial level of VRE. Hoffmann et al. [27] have reviewed different time series aggregation methods for energy system models, and Teichgraeber and Brandt [58] recently published a comprehensive overview of existing time series aggregation methods in energy systems. Traditional time representation methods for electricity system investment models typically belong to a family of methods denoted as *time slices*. Integral time slices can for example be a single time slice per year or a small set of seasonal and daily time slices to represent the differences in demand dependent on season, weekday, or time of day. Time slicing methods that are based on approximating the joint probability distribution of the load and VRE generation are developed by for example Wogrin et al. [62] and Lehtveer et al. [39]. Another time slicing method is the *representative days* method, as suggested by Nahmmacher et al. [42], which identifies a number of 24-hour segments based on load and VRE patterns over a day. Time reduction methods based on these principles have been implemented and shown promising results for long-term investment model, see for example Mai et al. [40], Gils [16], Gerbaulet and Lorenz [15] and Frew and Jacobson [12]. The methods have been compared and evaluated in Reichenberg et al. [47].

However, the integral time slicing methods have traditionally not worked when considering a larger geographical scope which includes regional trade. The reason for this is that approximating the joint probability distribution is

challenging since, unlike variations in load, variations in VRE generation does not follow a common pattern across a wide geographical scope. Thus, the integral time slicing methods can not properly account for wind and solar variations in models where a large geographical scope is considered. Furthermore, smoothing effects through trade is an important variation management strategy for VRE and thus, as is also concluded by Reichenberg et al. [47], the integral time slicing method is not ideal for a multi-node electricity system model with large shares of VRE. The representative days approach on the other hand can be employed in network models and therefore incorporate trade (see for example [11]). This time representation can also handle short term storage, but as the representative days typically consist of diurnal slices, it does not account for storage between days which requires interconnected time steps. An alternative is to model over longer time periods, i.e. weeks, but this increases model complexity and thus computation time. Hence, simplification in the spatial or technological system detail dimensions might be necessary to compensate for the increased complexity. A comparison of different clustering methods to reduce the spatial scope is done by Frysztacki et al. [14]. However, as concluded by Frysztacki et al. [13], a low spatial resolution can lead to sub-optimal investment decisions for wind and solar generation, while higher spatial resolution provides better results; this is especially important for energy systems with a high share of VRE.

3 Mathematical modelling

The problem studied in this thesis, for which we have developed two mathematical models (and three decomposition methods), consists of minimizing investment and operational costs while meeting the demand for electricity in a European electricity system. Europe has here been divided into several regions, chosen according to country borders and, if existent, infrastructural bottlenecks within the countries; see Figure 3.1. The variation management strategies accounted for in the models include electricity trade between regions, flexible electricity production, and energy storage.

A time period stretching from 2020 to 2050 is studied where we consider both existing production capacity [GW] as well as new investments. The total time period is divided into different investment periods in order to account for the technological lifetime of different power plants, i.e. the production capacity lifetime. Moreover, it allows the capture of policy changes and technology advancements such as increased efficiency. To reduce model complexity and problem size, we assume aggregated capacity in each region at the cost of losing some system detail. Each region produces electricity (measured in GWh/h) to meet its electricity demand. Trade along the electricity grid is possible between regions, where both existing transmission lines and investments into new transmission capacity are considered. Similar to that of electricity production capacity, the transmission capacity works as an upper limit for the electricity transmission.

Thermal cycling is included as previous work has shown that is has a substantial impact on the cost-optimal electricity system composition [19]. Furthermore, to keep the model linear, thermal cycling is accounted for using a relaxed unit commitment approach as described by Weber [61]. This method is explained in greater detail in Section 3.1, but to briefly summarize, variables are used to represent active (hot) production capacity in thermal power plants that is available for electricity generation in each time step. Moreover, there are some special constraints for renewable energy production, such as production

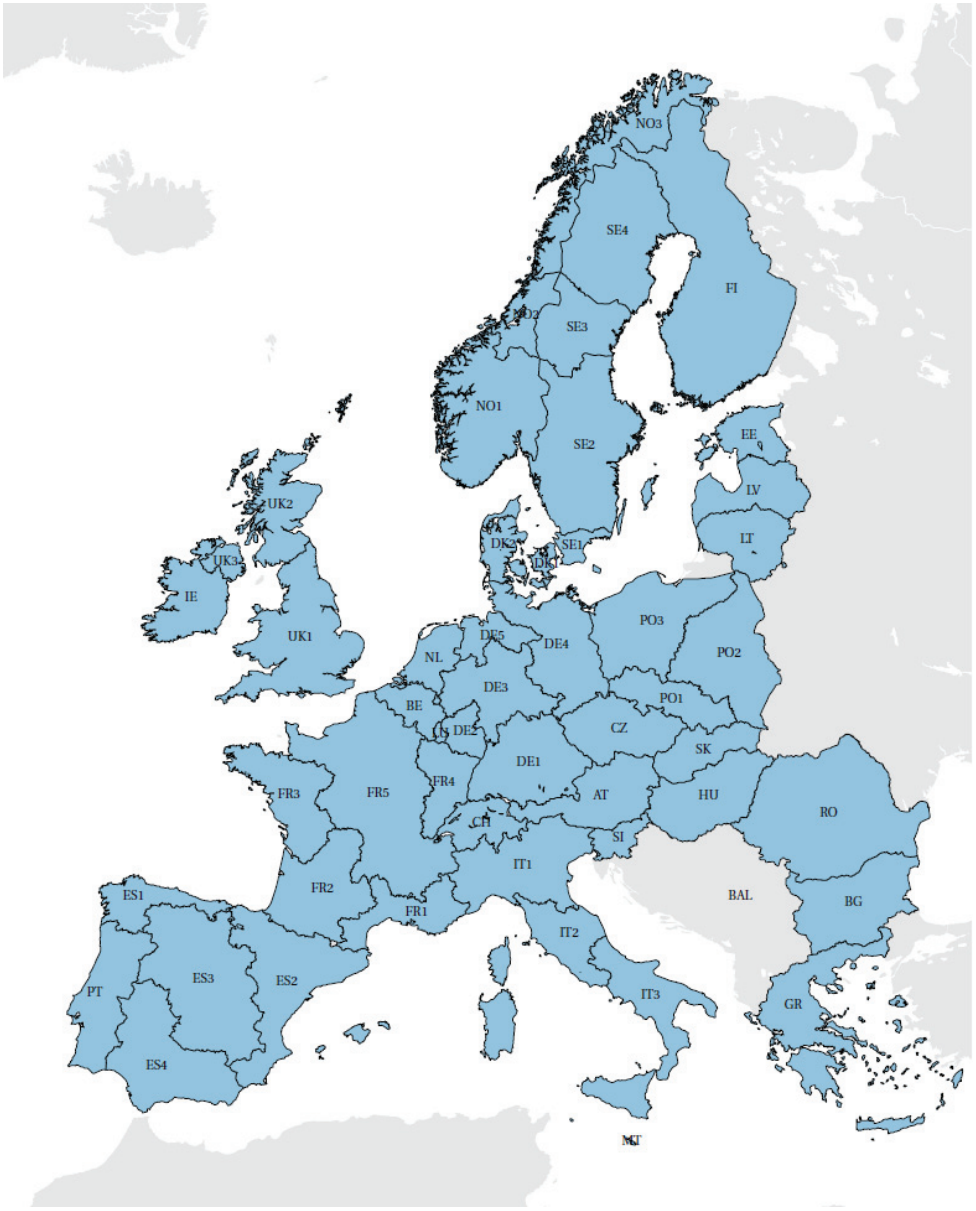


Figure 3.1: The different regions in Europe used in the model.

and capacity limits from weather, competing land use, and population density. Hydropower has constraints for ramp rates, i.e. the rate at which the electricity output can increase or decrease, and balance constraints for its energy storage. Other energy storage in the model include hydrogen and batteries. Lastly, we should account for the total system carbon dioxide emissions. This can be represented by a hard constraint, or by using a cost (and thus penalize carbon-emitting technologies) in the objective. For reasons relating to the solution methods, we have chosen to use the cost representation; see also Chapter 6 for further discussion regarding this.

3.1 Full-scale electricity system investment model

The problem is represented using a linear network model with additional side constraints. The sets used in this model are given by Table 3.1. The different regions are defined by the set $\mathcal{R}$. Transmission lines between regions are given by the set $\mathcal{A} \subseteq \mathcal{R} \times \mathcal{R}$. Thus, from region $q \in \mathcal{R}$ to region $r \in \mathcal{R}$, there exists a transmission line if $(q, r) \in \mathcal{A}$. The model uses different types of cables for transmission, and they are given by the set $\mathcal{K}$. The set $\mathcal{P}$ represents all the electricity production technologies, and it consists of three subsets: $\mathcal{P}_{\text{thermal}}$, $\mathcal{P}_{\text{ren}}$, and $\mathcal{P}_{\text{e-lysis}}$. The first subset is the set of thermal power technologies, the second subset is a set of renewable electricity generation technologies, and the last subset is the set of electrolysis technologies $\mathcal{P}_{\text{e-lysis}}$. The set of renewables is defined as $\mathcal{P}_{\text{ren}} := \mathcal{P}_{\text{wind}} \cup \mathcal{P}_{\text{solar}} \cup \mathcal{P}_{\text{hydro}}$, which are generation technologies for wind power, solar power, and hydropower, respectively. Different storage technologies are denoted by $\mathcal{L}$. It contains the set of battery technologies $\mathcal{L}_{\text{bat}}$, and the set of hydrogen technologies $\mathcal{L}_{\text{H}_2}$.

The modelling years are given by the set $\mathcal{S}$. Furthermore, the model uses the concept of investment periods, denoted $\mathcal{I}$, which are used to know at what year an investment in production capacity was made. This is relevant for two reasons: firstly, the model covers several years. Hence, if an investment is made at year s of technology type p, it should not be possible to use that invested capacity prior to the year s. Secondly, since all production technologies have an expected technical lifetime (dependent on the specific technology type $p \in \mathcal{P}$), then in order to know for how long that invested capacity can be used it is crucial to know when the investment was made. Therefore, the set $\mathcal{I}^{\mathcal{P}}_{\text{active}}(s, p)$ contains the investment periods for each technology type $p \in \mathcal{P}$ (with its own lifespan U_p) that are active at year $s \in \mathcal{S}$. Note that $\mathcal{S} \subset \mathcal{I}$. Similar to the set of active investment periods for the electricity generation technologies, a set of active periods for the transmission technologies is needed. This set is denoted $\mathcal{I}^{\mathcal{K}}_{\text{active}}(s)$, and it contains the active investment periods at year $s \in \mathcal{S}$. It is assumed that the different types of transmission technologies have a technical lifetime that outlives the model years, and therefore the set is not dependent on the transmission types $k \in \mathcal{K}$. The set of time steps within a year is denoted $\mathcal{T}$, and it is defined as $\mathcal{T} = \{\tau, 2\tau, ..., T\}$ such that τ is the time step length of the model. Then, an element in the set $\mathcal{T}$ is given by $t_\sigma = \sigma\tau$, where $\sigma \in \{1, 2, \ldots, \lfloor \frac{T}{\tau} \rfloor\}$. Note that $t_{\sigma+1} = t_\sigma + \tau$ holds. Lastly, since thermal cycling is considered in this model, $\mathcal{T}_{\text{start}}(p)$, is the set of consecutive time steps in the start-up interval for technology $p \in \mathcal{P}$.

A full nomenclature list is given in Appendix A.1. The mathematical constraints and objective for the problem are described below. All the decision variables in

the model, beside some auxiliary variables, are non-negative.

Table 3.1: The index sets used in the full-scale model

symbol		representation	member
$\mathcal{R}$		regions	r
$\mathcal{A}$	$\subseteq \mathcal{R} \times \mathcal{R}$;	transmission lines between regions	q, r
$\mathcal{K}$		technologies for transmission	k
$\mathcal{P}$	$:= \mathcal{P}_{\text{thermal}} \cup \mathcal{P}_{\text{ren}} \cup \mathcal{P}_{\text{e-lysis}}$;	electricity generation/consumption technologies	p
$\mathcal{P}_{\text{thermal}}$		thermal power technologies	p
$\mathcal{P}_{\text{ren}}$	$:= \mathcal{P}_{\text{wind}} \cup \mathcal{P}_{\text{solar}} \cup \mathcal{P}_{\text{hydro}}$;	renewable technologies	p
$\mathcal{P}_{\text{wind}}$		wind technologies	p
$\mathcal{P}_{\text{solar}}$		solar technologies	p
$\mathcal{P}_{\text{hydro}}$		hydropower technologies	p
$\mathcal{P}_{\text{e-lysis}}$		electrolyser technologies	h
$\mathcal{L}$	$:= \mathcal{L}_{\text{bat}} \cup \mathcal{L}_{\text{H}_2}$;	electricity storage technologies	ℓ
$\mathcal{L}_{\text{bat}}$		battery technologies	ℓ
$\mathcal{L}_{\text{H}_2}$		hydrogen storage technologies	ℓ
$\mathcal{I}$	$:= \{1960, 1970, \ldots, 2050\}$;	investment years, defining investment periods	i
$\mathcal{S}$	$:= \{2020, 2030, \ldots, 2050\}$;	new capacity investment years; $\mathcal{S} \subset \mathcal{I}$	s
$\mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$	$:= \mathcal{I} \cap \{s - U_p, \ldots, s\}$;	investment periods for each technology type $p \in \mathcal{P}$ with lifespan U_p that is active at year $s \in \mathcal{S}$	i
$\mathcal{I}^{\mathcal{K}}_{\text{active}}(s)$	$:= \mathcal{I} \cap \{1960, \ldots, s\}$;	investment periods for transmission technologies that are active at year $s \in \mathcal{S}$	i
$\mathcal{T}$	$:= \{\tau, 2\tau, \ldots, T\}$;	time steps within a year, where τ is the step length	t
$\mathcal{T}_{\text{start}}(p)$	$\subset \mathcal{T} \cup \{0\}$;	consecutive time steps in the start-up interval for technology $p \in \mathcal{P}$	t

Transmission and meeting the demand

We begin by introducing the decision variables (both measured in GWh/h)

x_{prist_σ} = generated electricity of technology type $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ at year $s \in \mathcal{S}$, investment period $i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$ and time step $t_\sigma \in \mathcal{T}$;

v_{kqrst_σ} = electricity traded with transmission type $k \in \mathcal{K}$ from region $q \in \mathcal{R}$ to region $r \in \mathcal{R}$, $(q,r) \in \mathcal{A}$, at year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$.

Define $v^{\text{net}}_{rst_\sigma}$ as the auxiliary variables used to express the net import [GWh/h] to region $r \in \mathcal{R}$ in year $s \in \mathcal{S}$ at time step $t_\sigma \in \mathcal{T}$, i.e.

$$v^{\text{net}}_{rst_\sigma} := \sum_{k\in\mathcal{K}} \sum_{q:(q,r)\in\mathcal{A}} v_{kqrst_\sigma} - \sum_{k\in\mathcal{K}} \sum_{j:(r,j)\in\mathcal{A}} v_{krjst_\sigma}, \qquad r \in \mathcal{R},\, s \in \mathcal{S},\, t_\sigma \in \mathcal{T}. \tag{3.1.1}$$

Now, define the decision variables

$$\begin{aligned} b^{\text{charge}}_{\ell rst_\sigma} &= \text{charging [GWh/h] of storage technology } \ell \in \mathcal{L} \text{ in region } r \in \mathcal{R} \\ &\quad \text{in year } s \in \mathcal{S} \text{ at time step } t_\sigma \in \mathcal{T}; \\ b^{\text{discharge}}_{\ell rst_\sigma} &= \text{discharging [GWh/h] of storage technology } \ell \in \mathcal{L} \text{ in region } r \in \mathcal{R} \\ &\quad \text{in year } s \in \mathcal{S} \text{ at time step } t_\sigma \in \mathcal{T}; \end{aligned}$$

and let $b^{\text{net}}_{rst_\sigma}$ be the auxiliary variables representing the net battery charge [GWh/h] in region r in year s at time step t_σ:

$$b^{\text{net}}_{rst_\sigma} := \sum_{\ell\in\mathcal{L}_{\text{bat}}} \left(b^{\text{charge}}_{\ell rst_\sigma} - b^{\text{discharge}}_{\ell rst_\sigma} \right), \qquad r \in \mathcal{R},\, s \in \mathcal{S},\, t_\sigma \in \mathcal{T}. \tag{3.1.2}$$

Define

$$\begin{aligned} h^{\text{consumption}}_{rst_\sigma} &= \text{aggregated electricity consumption [GWh/h] in the} \\ &\quad \text{electrolysers in region } r \in \mathcal{R} \text{ in year } s \in \mathcal{S} \text{ at time step } t_\sigma \in \mathcal{T}, \end{aligned}$$

and let d_{rst_σ} denote the demand for electricity in region $r \in \mathcal{R}$ at year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$. The electricity demand balance constraint can then for $r \in \mathcal{R}$, $s \in \mathcal{S}$ and $t_\sigma \in \mathcal{T}$ be formulated as

$$\sum_{p\in\mathcal{P}} \sum_{i\in\mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)} x_{prist_\sigma} + v^{\text{net}}_{rst_\sigma} \geq d_{rst_\sigma} + b^{\text{net}}_{rst_\sigma} + h^{\text{consumption}}_{rst_\sigma}. \tag{3.1.3}$$

Thus, all electricity generation and net import has to meet the demand, net battery charge and any electricity needed for hydrogen electrolysers.

Generation and transmission limits

For each region $r \in \mathcal{R}$ and investment period $i \in \mathcal{I}$, define the decision variables

$$\begin{aligned} y_{pri} &= \text{installed capacity [GW] of electricity generation technology } p \in \mathcal{P}; \\ y_{\ell ri} &= \text{installed capacity [GWh] of electricity storage technology } \ell \in \mathcal{L}. \end{aligned}$$

Let a^{gen}_{pri} and $a^{\text{sto}}_{\ell ri}$ be existing electricity generation and storage capacity, respectively, where $p \in \mathcal{P}$, $\ell \in \mathcal{L}$, $r \in \mathcal{R}$ and $i \in \mathcal{I} \setminus \mathcal{S}$. Fix the investment variable to existing capacity such that no new investments can be made prior to "now":

$$y_{pri} = a^{\text{gen}}_{pri}, \qquad p \in \mathcal{P},\, r \in \mathcal{R},\, i \in \mathcal{I} \setminus \mathcal{S}, \tag{3.1.4a}$$

$$y_{\ell ri} = a^{\text{sto}}_{\ell ri}, \qquad \ell \in \mathcal{L},\, r \in \mathcal{R},\, i \in \mathcal{I} \setminus \mathcal{S}. \tag{3.1.4b}$$

The transmission is limited by the transmission capacity. It is possible to invest into new capacity, but there is however some capacity previously installed. Let a^{tra}_{kqr} be a parameter which denotes the installed transmission capacity on transmission line $(q, r) \in \mathcal{A}$ of transmission type $k \in \mathcal{K}$. Define

$$\begin{aligned} u_{kqri} = {} & \text{installed transmission capacity [GW] for transmission type } k \in \mathcal{K} \\ & \text{from region } q \text{ to region } r,\, (q, r) \in \mathcal{A}, \text{ at investment period } i \in \mathcal{I}. \end{aligned}$$

and similar to (3.1.4), fix them to the existing capacity such that

$$u_{kqri} = a^{\text{tra}}_{kqri}, \qquad k \in \mathcal{K},\, (q, r) \in \mathcal{A},\, i \in \mathcal{I} \setminus \mathcal{S}. \tag{3.1.5}$$

Furthermore, for symmetry reasons, the transmission capacity for arcs $(q, r) \in \mathcal{A}$ and $(r, q) \in \mathcal{A}$ should be the same:

$$u_{kqri} = u_{krqi}, \qquad k \in \mathcal{K},\, (q, r) \in \mathcal{A},\, i \in \mathcal{I}. \tag{3.1.6}$$

This assumes that transmission is always possible in both directions, but for each solution it will only occur trade in one direction due to it otherwise being inefficient. The directed arcs are defined such that if $(q, r) \in \mathcal{A}$ then $(r, q) \in \mathcal{A}$. Moreover, there is an upper limit on the amount of transmission capacity:

$$\sum_{i \in \mathcal{I}^{\mathcal{K}}_{\text{active}}(s)} u_{kqri} \leq u^{\max}_{kqr}, \qquad k \in \mathcal{K},\, (q, r) \in \mathcal{A},\, s \in \mathcal{S}. \tag{3.1.7}$$

(Note that using the definition of $\mathcal{I}^{\mathcal{K}}_{\text{active}}(s)$ as in Table 3.1, these constraints are redundant for $s \in \mathcal{S} \setminus \{2050\}$.)

Lastly, the transmission should not exceed the total transmission capacity—both new and old—modelled as

$$v_{kqrst_\sigma} \leq \sum_{i \in \mathcal{I}^{\mathcal{K}}_{\text{active}}(s)} u_{kqri}, \qquad k \in \mathcal{K},\, (q, r) \in \mathcal{A},\, s \in \mathcal{S},\, t_\sigma \in \mathcal{T}. \tag{3.1.8}$$

Storage technologies

Define the decision variables

$$b^{\text{storage}}_{\ell r s t_\sigma} = \text{electricity storage level [GWh] in storage technology } \ell \in \mathcal{L} \text{ in region } r \in \mathcal{R} \text{ in year } s \in \mathcal{S} \text{ at time step } t_\sigma \in \mathcal{T}.$$

The charging and discharging of electricity in storage technologies should not exceed the total installed storage capacity. Let δ^{inj}_ℓ and $\delta^{\text{with}}_\ell$ be the injection and withdrawal rate, respectively, of storage technology $\ell \in \mathcal{L}$. Furthermore, the battery storage level is also limited by the installed battery capacity. These conditions are summarized in Constraints (3.1.9a)–(3.1.9c):

$$b^{\text{charge}}_{\ell r s t_\sigma} \leq \sum_{j \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,\ell)} \delta^{\text{inj}}_\ell y_{\ell r j}, \qquad \ell \in \mathcal{L},\ r \in \mathcal{R},\ s \in \mathcal{S},\ t_\sigma \in \mathcal{T}, \tag{3.1.9a}$$

$$b^{\text{discharge}}_{\ell r s t_\sigma} \leq \sum_{j \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,\ell)} \delta^{\text{with}}_\ell y_{\ell r j}, \qquad \ell \in \mathcal{L},\ r \in \mathcal{R},\ s \in \mathcal{S},\ t_\sigma \in \mathcal{T}, \tag{3.1.9b}$$

$$b^{\text{storage}}_{\ell r s t_\sigma} \leq \sum_{j \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,\ell)} y_{\ell r j}, \qquad \ell \in \mathcal{L},\ r \in \mathcal{R},\ s \in \mathcal{S},\ t_\sigma \in \mathcal{T}. \tag{3.1.9c}$$

For $\ell \in \mathcal{L}$, $r \in \mathcal{R}$ and $s \in \mathcal{S}$, the electricity storage level depends on the previous level and net charging of the storage, i.e.

$$b^{\text{storage}}_{\ell r s t_\sigma} + \eta^{\text{charge}}_{\ell s} \tau b^{\text{charge}}_{\ell r s t_\sigma} - \frac{\tau}{\eta^{\text{discharge}}_{\ell s}} b^{\text{discharge}}_{\ell r s t_\sigma} \geq \begin{cases} b^{\text{storage}}_{\ell,r,s,t_{\sigma+1}}, & t_\sigma \in \mathcal{T} \setminus \{T\}, \\ b^{\text{storage}}_{\ell,r,s,\tau}, & t_\sigma = T, \end{cases} \tag{3.1.10}$$

where $\eta^{\text{charge}}_{\ell s}$ and $\eta^{\text{discharge}}_{\ell s}$ is the efficiency for charging and discharging, respectively, storage technology $\ell \in \mathcal{L}$ in year $s \in \mathcal{S}$.

Hydrogen storage uses an electrolysis process, which is based on using electricity to split water into hydrogen and oxygen. This process takes place inside an electrolyser. Let $h^{\text{consumption}}_{rst}$ be a variable representing the electricity consumption of the electrolyser [GWh/h] in region $r \in \mathcal{R}$, year $s \in \mathcal{S}$ time step $t \in \mathcal{T}$. The electrolysis process is limited by the installed electrolyser capacity:

$$h^{\text{consumption}}_{r s t_\sigma} \leq \sum_{p \in \mathcal{P}_{\text{e-lysis}}} \sum_{j \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)} y_{prj}, \qquad r \in \mathcal{R},\ s \in \mathcal{S},\ t_\sigma \in \mathcal{T}. \tag{3.1.11}$$

Hydrogen storage investments are stimulated by introducing a demand for

electricity in hydrogen production for industry corresponding to 20% of the annual electricity demand evenly distributed over the year. Let d_{rs}^{hydrogen} be this hydrogen demand in region $r \in \mathcal{R}$, year $s \in \mathcal{S}$. Moreover, let η_{es} be the efficiency of electrolyser $p \in \mathcal{P}_{\text{e-lysis}}$ in year $s \in \mathcal{S}$. The hydrogen demand must be met by hydrogen production and net discharge from the hydrogen storage:

$$\eta_{ps} h_{rst_\sigma}^{\text{consumption}} + \sum_{\ell \in \mathcal{L}_{\text{H}_2}} \left(b_{\ell rst_\sigma}^{\text{discharge}} - b_{\ell rst_\sigma}^{\text{charge}} \right) \geq d_{rs}^{\text{hydrogen}}, \tag{3.1.12}$$

for $t_\sigma \in \mathcal{T}$, $p \in \mathcal{P}_{\text{e-lysis}}$, $r \in \mathcal{R}$ and $s \in \mathcal{S}$.

Thermal cycling

The idea of thermal cycling is that for thermal power plants, the capacity that has been taken out of operation has a minimum downtime that corresponds to the time it takes to start-up the capacity before it can generate electricity once again. Furthermore, and more importantly, the start-up cost for a unit is typically high and thermal cycling constraints can be used to capture this property. Introduce the decision variables

z_{prist_σ} = available hot capacity [GWh/h] of technology type $p \in \mathcal{P}_{\text{thermal}}$ in region $r \in \mathcal{R}$ at year $s \in \mathcal{S}$ for investment period $i \in \mathcal{I}_{\text{active}}^{\mathcal{P}}(s,p)$ and time step $t_\sigma \in \mathcal{T}$;

$z_{prist_\sigma}^{+}$ = started hot capacity [GWh/h] of technology type $p \in \mathcal{P}_{\text{thermal}}$ in region $r \in \mathcal{R}$ at year $s \in \mathcal{S}$ for investment period $i \in \mathcal{I}_{\text{active}}^{\mathcal{P}}(s,p)$ from time step $t_{\sigma-1} \in \mathcal{T}$ to time step $t_\sigma \in \mathcal{T}$.

The capacity that is currently up and running in a thermal power plant is referred to as *hot* or *available capacity*. The electricity generation should never exceed the available capacity. Likewise, it is required to generate a minimum level of electricity depending on the available capacity in order for it to stay hot. Letting ϕ_p denote the percentage corresponding to the minimum load level, this yields the constraints

$$\phi_p z_{prist_\sigma} \leq x_{prist_\sigma} \leq z_{prist_\sigma}, \tag{3.1.13}$$

for $i \in \mathcal{I}_{\text{active}}^{\mathcal{P}}(s,p)$, $p \in \mathcal{P}_{\text{thermal}}$, $r \in \mathcal{R}$, $s \in \mathcal{S}$, and $t_\sigma \in \mathcal{T}$.

To connect the started capacity to the available capacity, the following con-

straints are used for $i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$, $p \in \mathcal{P}_{\text{thermal}}$, $r \in \mathcal{R}$, and $s \in \mathcal{S}$:

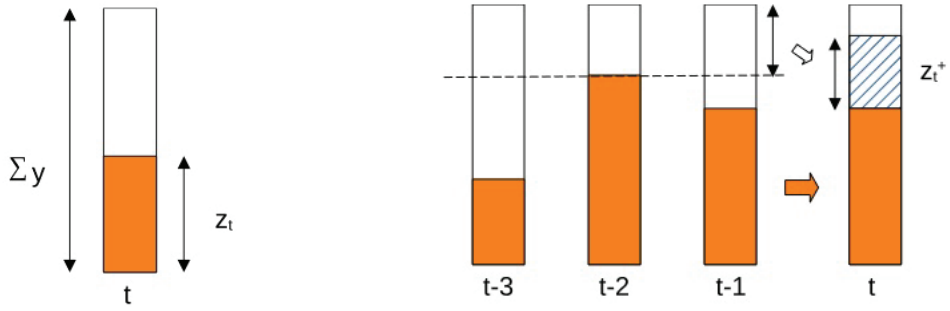
$$z^{+}_{prist_\sigma} \geq \begin{cases} z_{prist_\sigma} - z_{p,r,i,s,t_{\sigma-1}}, & t_\sigma \in \mathcal{T} \setminus \{1\}, \\ z_{prist_\sigma} - z_{prisT}, & t_\sigma = \tau. \end{cases} \tag{3.1.14}$$

The difference in available capacity between the time steps t_σ and $t_{\sigma-1}$ will then correspond to the started capacity. Since it is costly to start new capacity the variable $z^{+}_{prist_\sigma}$ is penalized in the objective function. Hence, $z^{+}_{prist_\sigma}$ equals zero whenever $z_{prist_\sigma} \leq z_{p,r,i,s,t_{\sigma-1}}$ holds. Moreover, $m \in \mathcal{T}_{\text{start}}(p)$, $p \in \mathcal{P}_{\text{thermal}}$, hours back in time, the started hot capacity is limited by the available capacity z_{prist_σ}. This yields, for $i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$, $p \in \mathcal{P}_{\text{thermal}}$, $r \in \mathcal{R}$, $s \in \mathcal{S}$, and $t_\sigma \in \mathcal{T}$, the constraints

$$\sum_{j \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)} y_{prj} - z^{+}_{prist_\sigma} \geq \begin{cases} z_{p,r,i,s,t_\sigma - m}, & m \in \mathcal{T}_{\text{start}}(p) \setminus \{t_\sigma, \ldots, T\}, \\ z_{p,r,i,s,T+t_\sigma-m}, & m \in \mathcal{T}_{\text{start}}(p) \setminus \{0, \ldots, t_{\sigma-1}\}. \end{cases} \tag{3.1.15}$$

These constraints are a linearization of start-up constraints more intuitively modeled as integers. The idea is that capacity that has been taken out of operation has a minimum downtime before it can be started again; see Weber [61] for further details.

As a simple example, consider Figure 3.2b. Here, the minimum down-time is three; thus we look at the hot capacity three time steps back. The amount of capacity that is available for start-up is capacity that has not been used in any of these time steps. The hot capacity is the largest for step $t-2$; therefore it will limit z^{+}_{t} the most.



(a) Hot capacity is limited by the total amount of installed capacity in time step t.

(b) The minimum down-time for hot capacity limits the possible start-up capacity.

Figure 3.2: Illustrations for thermal cycling.

Renewables

Letting θ_{prt_σ} denote a profile for technology $p \in \mathcal{P}_{\text{wind}} \cup \mathcal{P}_{\text{solar}}$ in region $r \in \mathcal{R}$ at time step $t_\sigma \in \mathcal{T}$, in terms of share of total installed capacity, the upper production limit of wind and solar power due to weather and climate is modelled as

$$x_{prist_\sigma} \leq \theta_{prt_\sigma} \sum_{j \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)} y_{prj}, \tag{3.1.16}$$

for $p \in \mathcal{P}_{\text{wind}} \cup \mathcal{P}_{\text{solar}}$, $r \in \mathcal{R}$, $i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$, $s \in \mathcal{S}$, and $t_\sigma \in \mathcal{T}$.

Not all areas are suitable for installations of wind farms since wind speed and terrain varies across the regions. Thus, due to reasons regarding land exploitation, there is an upper limit on the possible investments into wind capacity in each region; let W_{pr} be that upper limit for wind type $p \in \mathcal{P}_{\text{wind}}$ in region $r \in \mathcal{R}$. Then, this limit is modelled as

$$\sum_{i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)} y_{pri} \leq W_{pr}, \qquad p \in \mathcal{P}_{\text{wind}},\ r \in \mathcal{R},\ s \in \mathcal{S}. \tag{3.1.17}$$

Introduce the decision variables

$$w_{rst_\sigma} = \text{hydropower storage [GWh] in region } r \in \mathcal{R} \text{ at year } s \in \mathcal{S} \text{ at time step } t_\sigma \in \mathcal{T},$$

and let g_{rt_σ} be the inflow into the reservoirs in region $r \in \mathcal{R}$ at time step $t_\sigma \in \mathcal{T}$, assumed to be the same over the years. The hydropower balance constraint is then, for $p \in \mathcal{P}_{\text{hydro}}$, $r \in \mathcal{R}$ and $s \in \mathcal{S}$, modelled as

$$w_{rst_\sigma} + g_{rt_\sigma} - \sum_{i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)} \tau \cdot x_{prist_\sigma} \geq \begin{cases} w_{r,s,t_{\sigma+1}}, & t_\sigma \in \mathcal{T} \setminus \{T\}, \\ w_{rs\tau}, & t_\sigma = T. \end{cases} \tag{3.1.18}$$

where τ [h] denotes the length of the time step. The upper limit for the hydropower storage, denoted H_r for each $r \in \mathcal{R}$, is modelled as

$$w_{rst_\sigma} \leq H_r, \qquad r \in \mathcal{R},\ s \in \mathcal{S},\ t_\sigma \in \mathcal{T}. \tag{3.1.19}$$

The production level for hydropower cannot change too quickly, which is modelled by ramping rate constraints. Let δ_r^{inc} and δ_r^{dec} denote shares corresponding to the maximum change level in region $r \in \mathcal{R}$. For $i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$, $p \in \mathcal{P}_{\text{hydro}}$, $r \in \mathcal{R}$, and $s \in \mathcal{S}$, the following constraints imply an upper limit on the rate of

increase and decrease, respectively, of the storage level:

$$(1+\delta_r^{\text{inc}})x_{prist_\sigma} \geq \begin{cases} x_{p,r,i,s,t_{\sigma+1}}, & t_\sigma \in \mathcal{T}\setminus\{T\}, \\ x_{pris\tau}, & t_\sigma = T. \end{cases} \tag{3.1.20a}$$

$$(1+\delta_r^{\text{dec}})x_{prist_\sigma} \leq \begin{cases} x_{p,r,i,s,t_{\sigma+1}}, & t_\sigma \in \mathcal{T}\setminus\{T\}, \\ x_{pris\tau}, & t_\sigma = T. \end{cases} \tag{3.1.20b}$$

Lastly, no new investments in hydropower capacity are allowed. Thus,

$$y_{prs} = 0, \qquad p \in \mathcal{P}_{\text{hydro}},\, r \in \mathcal{R},\, s \in \mathcal{S}. \tag{3.1.21}$$

Emissions

The emissions arise from running the power plants (e.g. fuel), but also from start-ups of plants since fuel is needed for this. Furthermore, there are extra emissions when not running on full capacity, due to reduced efficiency. Let e_{pri} denote the emissions released [CO_2/MWh] by technology type $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ for capacity made in investment period $i \in \mathcal{I}$. Let e^+_{pri} and $\tilde{e}_{pri}$ denote the emissions released from start-ups and from running on part-load, respectively, for technology type $p \in \mathcal{P}_{\text{thermal}}$ in region $r \in \mathcal{R}$ and investment period $i \in \mathcal{I}$. The total emissions for year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$ is then expressed as

$$\begin{aligned} e^{\text{tot}}_{st_\sigma} := \sum_{r\in\mathcal{R}} \Bigg(& \sum_{p\in\mathcal{P}} \sum_{i\in\mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)} \tau e_{pri} x_{prist_\sigma} \\ & + \sum_{p\in\mathcal{P}_{\text{thermal}}} \sum_{i\in\mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)} \left(e^+_{pri} z^+_{prist_\sigma} + \tau\tilde{e}_{pri}(z_{prist_\sigma} - x_{prist_\sigma})\right) \Bigg). \end{aligned} \tag{3.1.22}$$

Objective

The objective is to minimize the total system costs. The sum (3.1.23a) considers the investment costs in electricity production technologies, with c_{ps}^{invtech} representing the investment cost (including annuity costs) for technology type $p \in \mathcal{P}$ in year $s \in \mathcal{S}$, and c_p^{omf} is the fixed operation and maintenance costs for technology type $p \in \mathcal{P}$. The analogous investment costs for storage technologies are given by (3.1.23b). The sum (3.1.23c) considers the costs of electricity production, with c_{pri}^{run} denoting the run costs for technology type $p \in \mathcal{P}$ in region $r \in \mathcal{R}$, where the investment is done in investment period $i \in \mathcal{I}$. The sum (3.1.23d) describes the additional costs for thermal power technology types $p \in \mathcal{P}_{\text{thermal}}$ using c_{prs}^{+} as the start-up cost for thermal capacity in region $r \in \mathcal{R}$ and year $s \in \mathcal{S}$, and $\tilde{c}_{prs}$ as the additional cost for running on part-load capacity, and the sum (3.1.23e) corresponds to the costs carbon dioxide emissions. The sum (3.1.23f) corresponds to the investment costs in new transmission capacity, with c_{kqr}^{invtra} being the investment cost in transmission capacity of type $k \in \mathcal{K}$ between regions $(q, r) \in \mathcal{A}$. This cost parameter is halved compared to the real cost in order to compensate for the network representation of using directed arcs, while in reality electricity is sent in either direction on the same transmission line. Finally, the sum (3.1.23g) covers the transmission cost of sending electricity using transmission technology $k \in \mathcal{K}$ on transmission line $(q, r) \in \mathcal{A}$, where the corresponding cost parameter is denoted c_{kqr}^{tra}. The objective is thus to minimize

$$
\begin{aligned}
&\sum_{s\in\mathcal{S}}\Bigg(\sum_{p\in\mathcal{P}}\sum_{r\in\mathcal{R}}\left(c_{ps}^{\text{invtech}}+c_{p}^{\text{omf}}\right)y_{prs} && \text{(3.1.23a)}\\
&\quad+\sum_{\ell\in\mathcal{L}}\sum_{r\in\mathcal{R}}(c_{\ell s}^{\text{invsto}}+c_{\ell}^{\text{omf}})y_{\ell rs} && \text{(3.1.23b)}\\
&\quad+\sum_{p\in\mathcal{P}}\sum_{r\in\mathcal{R}}\sum_{i\in\mathcal{I}_{\text{active}}^{\mathcal{P}}(s,p)}\sum_{t_\sigma\in\mathcal{T}}\tau c_{pri}^{\text{run}}x_{prist_\sigma} && \text{(3.1.23c)}\\
&\quad+\sum_{p\in\mathcal{P}_{\text{thermal}}}\sum_{r\in\mathcal{R}}\sum_{i\in\mathcal{I}_{\text{active}}^{\mathcal{P}}(s,p)}\sum_{t_\sigma\in\mathcal{T}}\left(c_{prs}^{+}z_{prist_\sigma}^{+}+\tau\tilde{c}_{prs}(z_{prist_\sigma}-x_{prist_\sigma})\right) && \text{(3.1.23d)}\\
&\quad+\sum_{t_\sigma\in\mathcal{T}}c_{s}^{\text{CO}_2}e_{st_\sigma}^{\text{tot}} && \text{(3.1.23e)}\\
&\quad+\sum_{k\in\mathcal{K}}\sum_{(q,r)\in\mathcal{A}}c_{kqr}^{\text{invtra}}u_{kqrs} && \text{(3.1.23f)}\\
&\quad+\sum_{k\in\mathcal{K}}\sum_{(q,r)\in\mathcal{A}}\sum_{t_\sigma\in\mathcal{T}}\tau c_{kqr}^{\text{tra}}v_{kqrst_\sigma}.\Bigg) && \text{(3.1.23g)}
\end{aligned}
$$

3.2 Hours-to-Decades model

The major difference between the full-scale model presented in Section 3.1 and the Hours-to-Decades model is that the latter model 1) disregards explicit yearly connections and instead links the years implicitly by adding operational costs to the objective, and 2) is decomposed into 2-week segments, within each of which the chronology is retained, providing 26 separate submodels. The purpose of this decomposition is to take advantage of the possibility to solve the submodels in parallel, which thus would shorten the computing times. The Hours-to-Decades model is solved by the use of a heuristic algorithm, i.e. a *consensus loop* which enables information to be exchanged between the submodels. Each separate subproblem represents the problem of meeting the demand for electricity while minimizing investment and operational costs in its 2-week segment. In the consensus loop, information from the solutions is gathered in capacity–cost curves in which the capacity invested in all 2-week segments have the lowest cost, while additional capacity invested in a subset of the segments is more expensive. The smaller the subset of segments, the more expensive the capacity. The solution process is iterated until there is consensus, i.e. until the capacity–cost curves are unchanged between iterations. The methodology for solving this model will be further discussed in Chapter 4, while the remainder of this section will focus on the electricity system investment model.

Table 3.2 provides a list of the sets used in this model. The notation here differs from the previous model as it was originally primarily aimed at an energy system modelling audience.

The set $\mathcal{I}$ is the set of all regions. The set $\mathcal{P}$ contains electricity generation technologies $\mathcal{P}^{\text{gen}} := \mathcal{P}^{\text{wind}} \cup \mathcal{P}^{\text{therm}} \cup \mathcal{P}^{\text{solar}}$, i.e. wind power, thermal power, and solar power. It also contains $\mathcal{P}^{\text{bat}}$, $\mathcal{P}^{\text{electrolysis}}$, and $\mathcal{P}^{\text{hydrogen}}$, which are the sets of battery technology, electrolyzer technology, and hydrogen storage technology, respectively. The set $\mathcal{Q}$ is the set of all transmission technologies, which are the same as in the full-scale model. The set $\mathcal{S} := \{1, \ldots, S\}$ contains all 2-week segments, where $S = 26$. The set of time steps for each $s \in \mathcal{S}$ is denoted $\mathcal{T}_s := \{(s-1)T + 1, \ldots, sT\}$. The model uses thermal cycling constraints similar to the previous model, and therefore $\mathcal{K}_p$ denotes the set of hours in the start-up interval for technology $p \in \mathcal{P}_{\text{thermal}}$. The set of cost classes $\mathcal{R}$ are calculated by the capacity–cost curves in the consensus loop, and are used to determine the cost of an investment.

Table 3.2: The index sets used in the Hours-to-Decades model

symbol	representation	member
$\mathcal{I}$	set of all regions	i, j
$\mathcal{P}$	$:= \mathcal{P}^{\text{bat}} \cup \mathcal{P}^{\text{electrolysis}} \cup \mathcal{P}^{\text{hydrogen}} \cup \mathcal{P}^{\text{gen}}$; set of all technology aggregates	p
$\mathcal{P}^{\text{bat}}$	set of all battery technologies	p
$\mathcal{P}^{\text{electrolysis}}$	set of all electrolyzer technologies	p
$\mathcal{P}^{\text{hydrogen}}$	set of all hydrogen storage technologies	p
$\mathcal{P}^{\text{gen}}$	$:= \mathcal{P}^{\text{wind}} \cup \mathcal{P}^{\text{therm}} \cup \mathcal{P}^{\text{solar}}$; set of all electricity generation technologies	p
$\mathcal{P}^{\text{wind}}$	set of all wind technologies	p
$\mathcal{P}^{\text{therm}}$	set of all thermal technologies	p
$\mathcal{P}^{\text{solar}}$	set of all solar technologies	p
$\mathcal{Q}$	set of technologies for transmission	q
$\mathcal{S}$	$:= \{1, \ldots, S\}$; set of all 2-week segments (typically, $S = 26$)	s
$\mathcal{T}_s$	$:= \{(s-1)T + 1, \ldots, sT\}$; set of all time steps in the 2-week segment $s \in \mathcal{S}$	t
$\mathcal{K}_p$	$:= \{0, \ldots\}$; set of hours in the start-up interval for technology $p \in \mathcal{P}_{\text{thermal}}$	k
$\mathcal{R}$	set of cost classes, i.e. the steps in the capacity–cost curve	r

A full nomenclature list is given in Appendix A.2. The constraints and objective function for the model are given below. All variables are non-negative besides the variable for electricity export.

Network balance and generation limits

Define the decision variables

w_{ipr} = installed electricity generation and storage capacity in technology $p \in \mathcal{P}$ in region $i \in \mathcal{I}$ and cost class $r \in \mathcal{R}$,

h_{ijqr} = installed transmission capacity between regions $i \in \mathcal{I}$ and $j \in \mathcal{I} \setminus \{i\}$, using transmission technology $q \in \mathcal{Q}$ in cost class $r \in \mathcal{R}$.

The consensus loop to be described in Chapter 4 computes cost class potentials, denoted M^{e}_{ipr} and M^{h}_{ijqr}, which acts as upper limit for the respective investment variables. This yields the constraints

$$w_{ipr} \leq M^{\mathrm{e}}_{ipr}, \qquad i \in \mathcal{I},\ p \in \mathcal{P},\ r \in \mathcal{R}, \tag{3.2.1}$$

$$h_{ijqr} \leq M^{\mathrm{h}}_{ijqr}, \qquad i \in \mathcal{I} \setminus \{j\},\ j \in \mathcal{I},\ q \in \mathcal{Q},\ r \in \mathcal{R}. \tag{3.2.2}$$

Introduce the decision variables

g_{ipt} = electricity generation in region $i \in \mathcal{I}$ of production type $p \in \mathcal{P}^{\mathrm{gen}}$ during time step $t \in \mathcal{T}_s,\ s \in \mathcal{S}$,

g_{ipt} = energy storage in region $i \in \mathcal{I}$ in battery type $p \in \mathcal{P}^{\mathrm{bat}}$ during time step $t \in \mathcal{T}_s,\ s \in \mathcal{S}$,

g_{ipt} = hydrogen storage in region $i \in \mathcal{I}$ in battery type $p \in \mathcal{P}^{\mathrm{hydrogen}}$ during time step $t \in \mathcal{T}_s,\ s \in \mathcal{S}$,

e_{ijt} = exported electricity from region $i \in \mathcal{I}$ to region $j \in \mathcal{I} \setminus \{i\}$ in time step $t \in \mathcal{T}_s,\ s \in \mathcal{S}$,

$b^{\mathrm{charge}}_{ipt}$ = charging of battery technology $p \in \mathcal{P}^{\mathrm{bat}}$ in region $i \in \mathcal{I}$ and time step $t \in \mathcal{T}_s,\ s \in \mathcal{S}$,

$b^{\mathrm{discharge}}_{ipt}$ = discharging of battery technology $p \in \mathcal{P}^{\mathrm{bat}}$ in region $i \in \mathcal{I}$ and time step $t \in \mathcal{T}_s,\ s \in \mathcal{S}$.

Let $d^{\mathrm{hydrogen}}_{it}$ denote the electricity consumption of the electrolyzer. The demand for electricity, D_{it}, must be met in all regions at all times. This is expressed, for $i \in \mathcal{I}$ and $t \in \mathcal{T}_s,\ s \in \mathcal{S}$, by the constraints

$$\sum_{p \in \mathcal{P}^{\mathrm{gen}}} g_{ipt} \geq D_{it} + d^{\mathrm{hydrogen}}_{it} + \sum_{j \in \mathcal{I} \setminus \{i\}} e_{ijt} + \sum_{p \in \mathcal{P}^{\mathrm{bat}}} \left(b^{\mathrm{charge}}_{ipt} - b^{\mathrm{discharge}}_{ipt} \right). \tag{3.2.3}$$

The import and export of electricity are required to be balanced, and the export may not exceed the installed transmission capacity, as expressed by the relations

$$-e_{ijt} = e_{jit} \leq \sum_{m \in \mathcal{I} \setminus \{i\}} \sum_{q \in \mathcal{Q}} \sum_{r \in \mathcal{R}} h_{imqr}, \quad i \in \mathcal{I} \setminus \{j\},\ j \in \mathcal{I},\ t \in \mathcal{T}_s,\ s \in \mathcal{S}, \tag{3.2.4}$$

$$e^{\mathrm{pos}}_{ijt} = |e_{ijt}| = \max\left\{e_{ijt}, e_{jit}\right\}, \qquad i \in \mathcal{I} \setminus \{j\},\ j \in \mathcal{I},\ t \in \mathcal{T}_s,\ s \in \mathcal{S}. \tag{3.2.5}$$

Let $\theta_{ipt} \in [0,1]$ be a profile, which is weather-dependent for wind and solar power but equals 1 for all $p \in \mathcal{P}^{\text{therm}}$. The level of electricity generation may not exceed the installed capacity, which is weighted by the weather profile:

$$g_{ipt} \leq \sum_{r \in \mathcal{R}} w_{ipr}\theta_{ipt}, \qquad i \in \mathcal{I},\ p \in \mathcal{P} \setminus \mathcal{P}^{\text{electrolysis}},\ t \in \mathcal{T}_s,\ s \in \mathcal{S}. \tag{3.2.6}$$

Battery storage

Flow batteries and lithium ion batteries are amongst the investment options in the model. An energy balance constraint is needed to manage the storage of each battery type. The battery storage level during the last time step of each 2-week segment $s \in \mathcal{S}$ constrains the battery storage level in the first time step of the same 2-week segment. Here, g_{ipt} is the storage level of the battery, $\eta_p b_{ipt}^{\text{charge}}$ is the charging of the battery where η_p is the efficiency of battery type $p \in \mathcal{P}^{\text{bat}}$, and $b_{ipt}^{\text{discharge}}$ is the discharging of the battery. For each $i \in \mathcal{I},\ p \in \mathcal{P}^{\text{bat}}$ and $s \in \mathcal{S}$ the constraints are expressed as

$$g_{ipt} + \eta_p b_{ipt}^{\text{charge}} - b_{ipt}^{\text{discharge}} \geq \begin{cases} g_{i,p,t+1}, & t \in \mathcal{T}_s \setminus \{sT\}, \\ g_{i,p,t-(T-1)}, & t = sT. \end{cases} \tag{3.2.7}$$

Each battery type has an installed storage capacity, and the charging and discharging of batteries may not exceed this limit:

$$b_{ipt}^{\text{charge}} \leq \sum_{r \in \mathcal{R}} w_{ipr}, \qquad i \in \mathcal{I},\ p \in \mathcal{P}^{\text{bat}},\ t \in \mathcal{T}_s,\ s \in \mathcal{S}, \tag{3.2.8}$$

$$b_{ipt}^{\text{discharge}} \leq \sum_{r \in \mathcal{R}} w_{ipr}, \qquad i \in \mathcal{I},\ p \in \mathcal{P}^{\text{bat}},\ t \in \mathcal{T}_s,\ s \in \mathcal{S}. \tag{3.2.9}$$

Hydrogen storage

Hydrogen storage uses an electrolysis process, which is based on using electricity to split water into hydrogen and oxygen. This process takes place inside an electrolyzer. The investments in hydrogen storage are stimulated by introducing a demand for electricity in hydrogen production for industry. Let D_i^{hydrogen} be the industry demand for hydrogen which is evenly distributed over the year, in region $i \in \mathcal{I}$. The hydrogen production in the electrolyzer is given by $\eta_p d_{it}^{\text{hydrogen}}$ for $p \in \mathcal{P}^{\text{hydrogen}}$, $i \in \mathcal{I}$ and $t \in \mathcal{T}_s,\ s \in \mathcal{S}$, where η_p denotes the efficiency of charging the hydrogen storage. Furthermore, the storage level and

the charging and discharging of the hydrogen storage during the last time step of the 2-week segment are used to constrain the hydrogen storage level in the first time step of the same 2-week segment. Thus, for $i \in \mathcal{I},\ p \in \mathcal{P}^{\text{hydrogen}}$ and $s \in \mathcal{S}$, this is modelled as

$$g_{ipt} + \eta_p d_{it}^{\text{hydrogen}} - D_i^{\text{hydrogen}} \geq \begin{cases} g_{i,p,t+1}, & t \in \mathcal{T}_s \setminus \{sT\}, \\ g_{i,p,t-(T-1)}, & t = sT. \end{cases} \tag{3.2.10}$$

The electricity consumption of the electrolyzer, d_{it}^{hydrogen}, may not exceed the installed electrolyzer capacity:

$$d_{it}^{\text{hydrogen}} \leq \sum_{r \in \mathcal{R}} w_{ipr}, \qquad i \in \mathcal{I},\ p \in \mathcal{P}^{\text{electrolysis}},\ t \in \mathcal{T}_s,\ s \in \mathcal{S}. \tag{3.2.11}$$

Wind and solar power

The wind power capacity in technology $p \in \mathcal{P}^{\text{wind}}$ and region $i \in \mathcal{I}$ is limited by the regional resources A_{ip}, which implies an upper bound on wind power investments:

$$\sum_{r \in \mathcal{R}} w_{ipr} \leq A_{ip}, \qquad i \in \mathcal{I},\ p \in \mathcal{P}^{\text{wind}}. \tag{3.2.12}$$

For solar power, there is a total resource constraint for each modeled region $i \in \mathcal{I}$:

$$\sum_{r \in \mathcal{R}} \sum_{p \in \mathcal{P}^{\text{solar}}} w_{ipr} \leq \sum_{p \in \mathcal{P}^{\text{solar}}} A_{ip}, \qquad i \in \mathcal{I}. \tag{3.2.13}$$

Thermal cycling

As in the previous model, thermal cycling is here accounted for by applying the relaxed unit commitment approach suggested by Weber [61]. Let

g_{ipt}^{active} = capacity that is active and available for generation in each time step $t \in \mathcal{T}_s,\ s \in \mathcal{S}$, in region $i \in \mathcal{I}$ and within each technology aggregate $p \in \mathcal{P}^{\text{therm}}$.

g_{ipt}^{on} = capacity started in each time step $t \in \mathcal{T}_s,\ s \in \mathcal{S}$, in region $i \in \mathcal{I}$ and within each technology aggregate $p \in \mathcal{P}^{\text{therm}}$;

The electricity generation should stay below the active capacity in each time step. Moreover, the minimum load share of the active capacity for technology $p \in \mathcal{P}^{\text{therm}}$ is given by ϕ_p, and the electricity generation is not allowed to be below this level. Hence, the following inequalities are included in the model:

$$\xi_p^{\min} g_{ipt}^{\text{active}} \leq g_{ipt} \leq g_{ipt}^{\text{active}}, \qquad i \in \mathcal{I},\ p \in \mathcal{P}^{\text{therm}},\ t \in \mathcal{T}_s,\ s \in \mathcal{S}. \tag{3.2.14}$$

As in the previous model, the active capacity is limited in each time step by the sum of the started capacity and the active capacity in the previous time step. However, for the first time step of each 2-week segment, except the first segment, the active capacity in the previous time step is represented by the active capacity in the last time step of the previous segment, as given by the previous iteration of the consensus loop (see Chapter 4.4), i.e. by $G_{i,p,t-1}^{\text{active}}$. Moreover, for the first time step of the first segment (i.e. for $t = 1$), the active capacity in the last time step of the last segment is used, as given by the previous iteration of the consensus loop, i.e. by $G_{i,p,ST}^{\text{active}}$. For $i \in \mathcal{I}$ and $p \in \mathcal{P}^{\text{therm}}$, these relations are modelled by the inequalities

$$g_{ipt}^{\text{active}} \leq g_{ipt}^{\text{on}} + \begin{cases} g_{i,p,t-1}^{\text{active}}, & t \in \mathcal{T}_s \setminus \{(s-1)T+1\}, \quad s \in \mathcal{S}, \\ G_{i,p,t-1}^{\text{active}}, & t = (s-1)T+1, \qquad\qquad\quad s \in \mathcal{S} \setminus \{1\}, \\ G_{i,p,ST}^{\text{active}}, & t = 1. \end{cases} \tag{3.2.15}$$

Define the variable

$$c_{ipt}^{\text{cycl}} = \text{resulting thermal cycling costs in region } i \in \mathcal{I} \text{ for technology type } p \in \mathcal{P}^{\text{therm}} \text{ in time step } t \in \mathcal{T}_s,\ s \in \mathcal{S}.$$

As in the previous model, the start-up cost is proportional to the started capacity g_{ipt}^{on}, while the part-load cost is proportional to the difference between the active generation capacity and the generation level. In order to avoid boundary effects on the last time step of the 2-week segment, we include a value for the active capacity which is proportional to the solution given in the previous iteration. This scaling is based on the start-up cost $C_{i,p,t+1}^{\text{on}} G_{i,p,t+1}^{\text{on}}$ and active capacity $G_{i,p,t+1}^{\text{active}}$ paid in the first hour of the following 2-week segment. For each $i \in \mathcal{I}$ and $p \in \mathcal{P}^{\text{therm}}$, these constraints are expressed as

$$c_{ipt}^{\text{cycl}} \geq C_{ipt}^{\text{on}} g_{ipt}^{\text{on}} + C_{ipt}^{\text{part}} \left(g_{ipt}^{\text{active}} - g_{ipt} \right) - G_{ipt}, \qquad t \in \mathcal{T}_s,\ s \in \mathcal{S}, \tag{3.2.16}$$

where

$$G_{ipt} := \frac{g^{\text{active}}_{ipt}}{2} \cdot \begin{cases} 0, & t \in \mathcal{T}_s \setminus \{sT\}, \quad s \in \mathcal{S}, \\ \frac{C^{\text{on}}_{i,p,t+1} G^{\text{on}}_{i,p,t+1}}{G^{\text{active}}_{i,p,t+1}}, & t = sT, \quad s \in \mathcal{S} \setminus \{S\}, \\ \frac{C^{\text{on}}_{ip1} G^{\text{on}}_{ip1}}{G^{\text{active}}_{ip1}}, & t = ST. \end{cases} \tag{3.2.17}$$

Hence, if thermal capacity is active in the end of one 2-week segment and also in the beginning of the subsequent 2-week segment, the start-up cost for that capacity is shared equally between the segments.

As explained previously, thermal generation is subject to a start-up time, i.e. it takes some time for a thermal power plant to heat up before it can deliver electricity. Thus, in the model, once capacity is deactivated, it cannot become active again during the interval $\mathcal{K}_p$, which encompasses the time-steps k in the start-up interval. For $i \in \mathcal{I}$ and $p \in \mathcal{P}^{\text{therm}}$, this is expressed as

$$g^{\text{on}}_{ipt} \leq \sum_{r \in \mathcal{R}} w_{ipr} - g^{\text{active}}_{i,p,t-k}, \qquad t \in \mathcal{T}_s,\ s \in \mathcal{S},\ k \in \mathcal{K}_p \setminus \{t, \ldots, sT\}. \tag{3.2.18}$$

Objective

For each 2-week segment $s \in \mathcal{S}$, the objective function to be minimized is expressed as

$$c^{\text{tot}}_s := \sum_{i \in \mathcal{I}} \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}} C^{\text{inv}}_p \lambda^{\text{e}}_{iprs} w_{ipr} \tag{3.2.19a}$$

$$+ \sum_{i \in \mathcal{I}} \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}_s} \left(C^{\text{run}}_{pt} g_{ipt} + c^{\text{cycl}}_{ipt} \right) \tag{3.2.19b}$$

$$+ \sum_{q \in \mathcal{Q}} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I} \setminus \{i\}} \sum_{r \in \mathcal{R}} C^{\text{h-inv}}_q \lambda^{\text{h}}_{ijqrs} h_{ijqr} \tag{3.2.19c}$$

$$+ \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I} \setminus \{i\}} \sum_{t \in \mathcal{T}_s} C^{\text{exp}}_t e^{\text{pos}}_{jit}, \tag{3.2.19d}$$

where (3.2.19a) represents the costs for investments in the different technologies in the different regions, (3.2.19b) the running costs of the different technologies in the different regions at all time steps within the 2-week segment, (3.2.19c) the costs for investments in technologies for transmission of electricity between the regions, and (3.2.19d) the costs of transmitting electricity between the regions in each time step within the 2-week segment.

Here, for the cost class $r \in \mathcal{R}$ and segment $s \in \mathcal{S}$, the parameters $\lambda^{\mathrm{e}}_{iprs}$ and $\lambda^{\mathrm{h}}_{ijqrs}$ represent shares of the investment costs for electricity generating technologies $p \in \mathcal{P}$ in region $i \in \mathcal{I}$ and transmission technologies $q \in \mathcal{Q}$ between regions $i \in \mathcal{I}$ and $j \in \mathcal{I} \setminus \{i\}$, respectively.

3.3 Comparison of the different models

The two models presented in Sections 3.1 and 3.2, respectively, are very alike in terms of structure and type of constraints. There are however some key differences as the models serve different purposes and can be used to answer different questions.

The full-scale model can be seen as a starting point as it contains many of the features necessary to analyze for variable renewable electricity integration, but is demanding to solve for high resolution data. Thus, for large problem instances with multiple regions, it is necessary to employ some decomposition method to be able to solve the model within reasonable computation times, or even at all as problems of that size will struggle with memory management. If solved with a smaller spatial resolution however, i.e. a single region, it can be used to provide feasible solutions to the full-scale model. This is especially useful to evaluate results from both the full-scale model and the Hours-to-Decades model, but also to provide us with upper bounds necessary in the subgradient algorithm described in Section 4.3.1.

The Hours-to-Decades model from Paper IV was the initial idea on how to split the model over the temporal dimension, and it showed very promising results as it converged in only a few iterations of the consensus algorithm. However, the decomposition method lacks in some aspects as it is not a pure mathematical decomposition of the problem, but rather a heuristic. Moreover, for storage that typically extends two weeks (i.e. seasonal storage such as hydrogen), the method is unable to properly dimension the needed capacity.

However, based on the results from the Hours-to-Decades model, a mathematical decomposition based on Lagrangian relaxation that uses the same 2-week approach was tested and evaluated on the full-scale model as well; see Chapter 4 and Papers I–III. The advantage of this method is that it—in theory—should be able to dimension the seasonal storage, but this unfortunately proved difficult in practice.

4 Mathematical methods and their theory

The models introduced in Chapter 3 are by nature very large and therefore practically impossible to solve in reasonable computation times. To counteract this, different mathematical methods need to be used. This chapter provides a mathematical background to modelling and the optimization methods utilized for solving the electricity systems models discussed in Chapter 3.

4.1 Linear programming

A linear program (LP) is a convex optimization problem with affine objective and constraint functions, where the variables $\boldsymbol{x}$ are restricted to be non-negative. All LPs can be written on standard form, i.e.

$$\begin{array}{lrrl} \text{minimize} & z := & \boldsymbol{c}^\top \boldsymbol{x}, & \\ \text{subject to} & & \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b}, \\ & & \boldsymbol{x} &\geq \boldsymbol{0}, \end{array} \tag{4.1.1}$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, and m, $n \in \mathbb{Z}_+$. Hence, m is the number of constraints and n is the dimension of the variable space.

Many problems can be described as linear programs, and several clever solution methodologies exist to solve them, e.g. the simplex algorithm or interior penalty methods (see Boyd and Vandenberghe [3]). However, for very large-scale models these methods typically have to be complemented with decomposition techniques to make the problems solvable within reasonable computation times. Several such methods use Lagrangian duality and Lagrangian relaxation as a foundation.

4.2 Lagrangian dual concepts

The model structure for very large optimization problems can often be exploited by using Lagrangian duality and Lagrangian relaxation, which first came to light by the seminal work of Held and Karp [25]. Some decomposition methods relating to electrical energy applications are also covered by Sagastizábal [53].

4.2.1 Lagrangian relaxation

Many large optimization problems are structured such that they consist of several smaller and separate problems connected by some overlapping, typically complicating, constraints. Each separate problem is, however, often more easily solvable in comparison to the full problem. Lagrangian dual methods, such as Lagrangian relaxation, can take advantage of this problem structure; see Guignard [21].

The idea of Lagrangian relaxation is to relax the connecting constraints such that the remaining problem is separable into several subproblems. Consider the linear optimization problem to

$$\begin{array}{lrll} \text{minimize} & z := & \boldsymbol{c}^\top \boldsymbol{x}, & \\ \text{subject to} & & \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}^m, & \\ & & \boldsymbol{x} \in \boldsymbol{X}, & \end{array} \tag{4.2.1}$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{X} \subset \mathbb{R}^n$, m, $n \in \mathbb{Z}_+$. We also assume that $\{\boldsymbol{x} \in \boldsymbol{X} \,|\, \boldsymbol{g}(\boldsymbol{x})) \leq \boldsymbol{0}^m\} \neq \emptyset$ such that there exists a feasible solution. Here, $\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}^m$ are connecting constraints while the remaining set $\boldsymbol{x} \in \boldsymbol{X}$ is separable; see Figure 4.1.

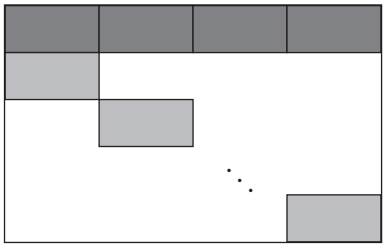


Figure 4.1: A block diagonal matrix besides the rows representing the connecting constraints. Here, the dark area corresponds to the connecting constraints $\boldsymbol{g}(\boldsymbol{x})$, while the lighter area is the separable set $\boldsymbol{x} \in \boldsymbol{X}$.

Define the Lagrangian function $L : \mathbb{R}^{m+n} \to \mathbb{R}$ such that

$$L(\boldsymbol{x}, \boldsymbol{\pi}) := \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\pi}^\top \boldsymbol{g}(\boldsymbol{x}), \tag{4.2.2}$$

using the Lagrangian multipliers $\boldsymbol{\pi} \in \mathbb{R}^m$. The Lagrangian dual problem is then defined as

$$\max_{\boldsymbol{\pi} \geq \mathbf{0}} \quad h(\boldsymbol{\pi}), \tag{4.2.3}$$

where

$$h(\boldsymbol{\pi}) \quad := \quad \min_{\boldsymbol{x} \in \boldsymbol{X}} L(\boldsymbol{x}, \boldsymbol{\pi}) \; = \min_{\boldsymbol{x} \in \boldsymbol{X}} \left(\boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\pi}^\top \boldsymbol{g}(\boldsymbol{x}) \right) \tag{4.2.4}$$

and $h : \mathbb{R}^m \to \mathbb{R}$ is denoted the Lagrangian dual function. Here, for some $\boldsymbol{\pi} \geq \mathbf{0}^m$, the problem of minimizing the Lagrangian function L over its first argument $\boldsymbol{x} \in X$ is referred to as the subproblem, and since we assumed a block diagonal structure of the set $\boldsymbol{x} \in \boldsymbol{X}$, this problem is separable.

Let z^* denote the optimal objective function value in problem (4.2.1). By weak duality, $h(\boldsymbol{\pi}) \leq z^*$ then holds for any $\boldsymbol{\pi} \geq \mathbf{0}$; that is, any feasible solution to the dual problem provides a lower bound on the optimal objective value of the original problem. Furthermore, any feasible solution $\bar{\boldsymbol{x}}$ provides an upper bound for z^* since the inequality $z^* \leq z(\bar{\boldsymbol{x}})$ holds. Moreover, strong duality implies that the equality $z^* = h^*$ holds.

4.2.2 Augmented Lagrangian

The augmented Lagrangian method adds an additional penalty term (i.e. the augmentation) to the Lagrangian function, penalizing the constraint equations. The idea was independently introduced by Hestenes [26] and Powell [46], with the purpose to combine the advantages of the penalty method and the multiplier method. It was later on extended to the case of optimization with inequality constraints by Rockafellar [50, 51], but we will for this section consider the equality case. Thus, consider once more the linear optimization problem (4.2.1), but now with equality constraints:

$$\begin{array}{lrll} \text{minimize} & z := & \boldsymbol{c}^\top \boldsymbol{x}, & \\ \text{subject to} & & \boldsymbol{g}(\boldsymbol{x}) & = \mathbf{0}^m, \\ & & \boldsymbol{x} & \in \boldsymbol{X}, \end{array} \tag{4.2.5}$$

where the corresponding Lagrangian function is defined as in (4.2.2). We can modify problem (4.2.5) by adding a quadratic penalty term to the objective.

Then, for some penalty parameter $\rho > 0$, the modified problem is

$$\begin{array}{lll} \text{minimize} & z := & \boldsymbol{c}^\top \boldsymbol{x} + \dfrac{\rho}{2} || \, \boldsymbol{g}(\boldsymbol{x}) ||_2^2, \\ \text{subject to} & & \boldsymbol{g}(\boldsymbol{x}) \;=\; \boldsymbol{0}^m, \\ & & \boldsymbol{x} \;\in\; \boldsymbol{X}. \end{array} \tag{4.2.6}$$

Note that this problem is equivalent to the original problem (4.2.5), since for any feasible $\boldsymbol{x}$ the term added to the objective is zero. Hence, the optimal solution and objective function value have not changed. The augmented Lagrangian function is then defined as

$$L_\rho(\boldsymbol{x}, \boldsymbol{\pi}) := \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\pi}^\top \boldsymbol{g}(\boldsymbol{x}) + \frac{\rho}{2} || \, \boldsymbol{g}(\boldsymbol{x}) ||_2^2. \tag{4.2.7}$$

The associated dual problem is to

$$\max_{\boldsymbol{\pi} \in \mathbb{R}^m} \quad h_\rho(\boldsymbol{\pi}), \tag{4.2.8}$$

where the augmented dual function is defined as

$$h_\rho(\boldsymbol{\pi}) \quad := \quad \min_{\boldsymbol{x} \in \boldsymbol{X}} L_\rho(\boldsymbol{x}, \boldsymbol{\pi}) \;= \min_{\boldsymbol{x} \in \boldsymbol{X}} \left(\boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\pi}^\top \boldsymbol{g}(\boldsymbol{x}) + \frac{\rho}{2} || \, \boldsymbol{g}(\boldsymbol{x}) ||_2^2 \right). \tag{4.2.9}$$

The augmented Lagrangian has several beneficial properties, including that the dual function (4.2.9) can be shown to be differentiable under rather mild conditions on the original problem [2]. For large values of the penalty parameter ρ, the augmented primal objective becomes strongly convex, which is especially useful for nonlinear objective functions. Moreover, the penalty used here is softer than penalties used in, for example, interior point methods such as the barrier method. Thus, the iterates of penalty methods that use the augmented Lagrangian are not confined to interior points.

4.2.3 Variable splitting

For the two models presented in Chapter 3, relaxing over the time dimension is tricky since some variables are time-independent. A clever workaround for this exploits *variable splitting*, introduced by Jörnsten and Näsberg [34]. Let us demonstrate this method in Example 1. Note that since this method was developed simultaneously by different researchers, it is therefore also referred to as *Lagrangian decomposition* [22] or *variable layering* [17].

Example 1. Divide the annual time steps $\mathcal{T}$ into time periods (i.e. segments of M weeks) over the year such that the subset $\mathcal{T}_n$ defines the time steps in period

$n \in \mathcal{N} := \{1, ..., N\}$. Hence, $\mathcal{T} := \{\mathcal{T}_1, \ldots, \mathcal{T}_N\} = \bigcup_{n=1}^{N} \mathcal{T}_n$. Moreover, define $\mathcal{T}_n := \{s_n, \ldots, t_n\}$, and let $t_{n-1} = s_n - 1$ for $n \neq 1$. It then follows, that for any $m, n \in \mathcal{N}$ where $m \neq n$, the intersection $\mathcal{T}_m \cap \mathcal{T}_n = \emptyset$ must hold.

Define the variables $x_t \in \mathbb{R}_+$ and $y \in \mathbb{R}_+$, representing the electricity generation in time step $t \in \mathcal{T}_n$, $n \in \mathcal{N}$, and the invested capacity, respectively. Let X be the feasible set for $\boldsymbol{x}$ such that $x_t \in X := \bigtimes_{n=1}^{N} X_n$ for $t \in \mathcal{T}$. Then, consider the problem to

$$\begin{aligned}
\underset{y, x_t}{\text{minimize}} \quad & c^{\text{inv}} y + \sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}_n} c_t^{\text{run}} x_t, && && \text{(4.2.10a)} \\
\text{subject to} \quad & (x_t)_{t \in \mathcal{T}_n} \in X_n, && n \in \mathcal{N}, && \text{(4.2.10b)} \\
& a_{n-1} x_{t_{n-1}} + b_n x_{s_n} \geq d_n, && n \in \mathcal{N} \setminus \{1\}, && \text{(4.2.10c)} \\
& x_t \leq y, && t \in \mathcal{T}_n,\ n \in \mathcal{N}, && \text{(4.2.10d)} \\
& y \in [0, Y], && && \text{(4.2.10e)}
\end{aligned}$$

where c^{inv} and c_t^{run}, $t \in \mathcal{T}_n$, $n \in \mathcal{N}$, are investment costs and run costs, respectively, and Y represents some large number acting as an upper limit for the investment variable y. The block structure corresponding to model (4.2.10) is given in Figure 4.2. Note that this problem is a simplified version of the

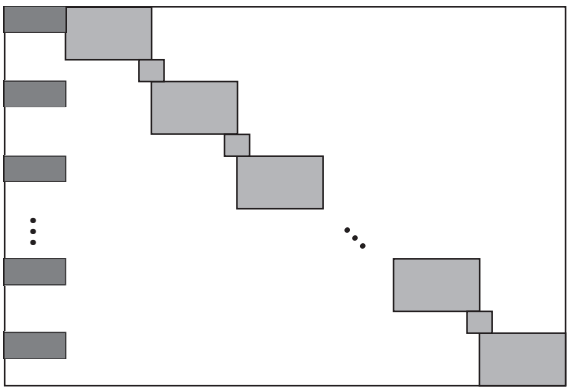


Figure 4.2: The model structure of problem (4.2.10). The dark boxes represent the variable investment variable y, and the rows they are included in correspond to the constraints (4.2.10d). The lighter small boxes represent constraints (4.2.10c), and the larger light boxes illustrate constraints (4.2.10b).

models introduced in Chapter 3. Here, (4.2.10b) represent all constraints which do not span over multiple time periods. The constraints (4.2.10c) on the other hand correspond to constraints that cover multiple time periods, such that $a_{n-1} x_{t_{n-1}}$ belongs to period $n-1$ while $b_n x_{s_n}$ and d_n belong to period n. These

constraints can be relaxed without issues to make the problem separable over n. However, the constraints (4.2.10d) are a bit more problematic since the variable y is time independent. The constraints $x_t \leq y$ can be relaxed, but they represent an important property of the electricity system (i.e. that you can not produce more electricity than the installed capacity). Thus, they are very important for the model structure and should preferably not be relaxed.

This is where we let variable splitting work its magic. The main idea is to reformulate the problem such that copies of some primal variables are introduced (in this example, the variable y). Constraints are added to ensure consistency between the original variables and the copies; then these consistency constraints are Lagrangian relaxed. Therefore, let us introduce the splitting variables for the investment variable in problem (4.2.10) by letting $y_n^{\text{split}} = y$, $n \in \mathcal{N}$. The equivalent problem formulation is given by

$$\begin{aligned}
\underset{y, y_n^{\text{split}}, x_t}{\text{minimize}} \quad & \frac{c^{\text{inv}}}{N} \sum_{n \in \mathcal{N}} y_n^{\text{split}} + \sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}_n} c_t^{\text{run}} x_t, && && \text{(4.2.11a)} \\
\text{subject to} \quad & (x_t)_{t \in \mathcal{T}_n} \in X_n, && n \in \mathcal{N}, && \text{(4.2.11b)} \\
& a_{n-1} x_{t_{n-1}} + b_n x_{s_n} \geq d_n, && n \in \mathcal{N} \setminus \{1\}, && \text{(4.2.11c)} \\
& y_n^{\text{split}} = y, && n \in \mathcal{N}, && \text{(4.2.11d)} \\
& x_t \leq y_n^{\text{split}} \leq Y, && t \in \mathcal{T}_n,\, n \in \mathcal{N}, && \text{(4.2.11e)} \\
& y \in [0, Y]. && && \text{(4.2.11f)}
\end{aligned}$$

See Figure 4.3 for an illustration of the new matrix structure.

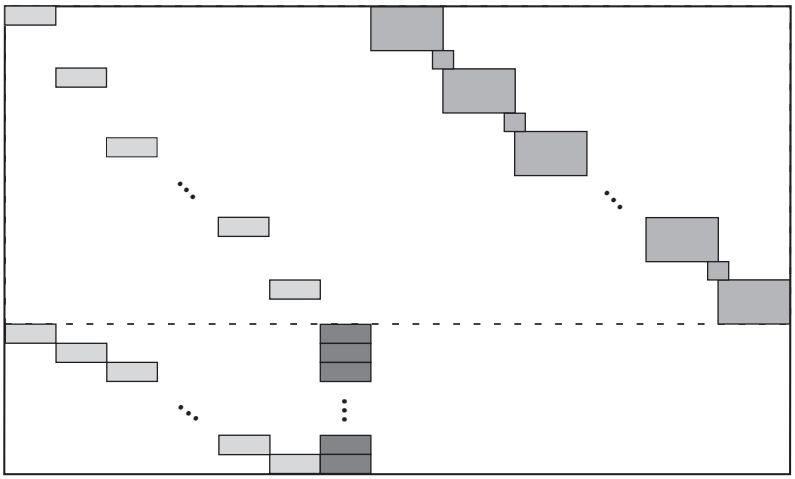
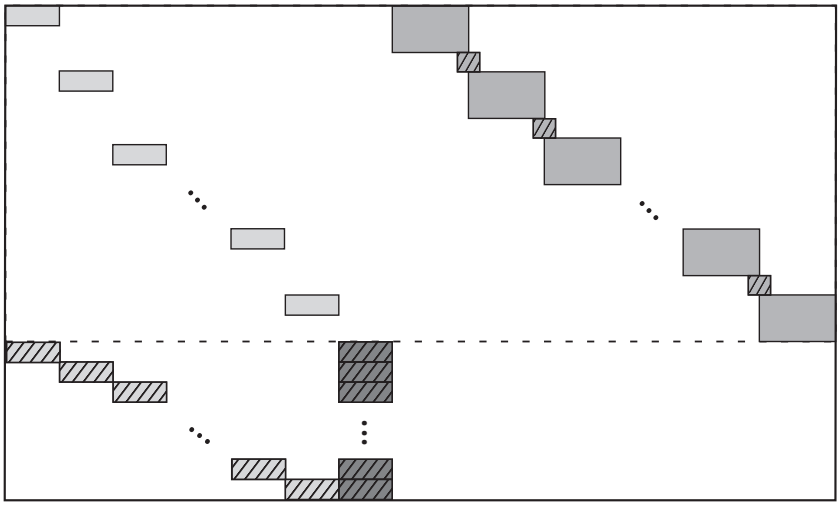
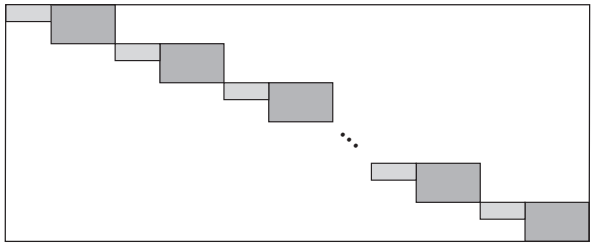


Figure 4.3: The model structure of problem (4.2.11), after the variable splitting reformulation. The constraints above the dashed line are the same as in Figure 4.2 but has had y replaced by y_n^{split}, and below the dashed line we see the constraints (4.2.11d).

Here, we can relax (4.2.11c) and (4.2.11d) to make the problem separable while letting the model structure in each subproblem $n \in \mathcal{N}$ remain; see also Figure 4.4.



(a) The model structure of problem (4.2.11). The striped boxes correspond to the constraints that will be relaxed.



(b) The resulting model structure when constraints (4.2.11c) and (4.2.11d) have been relaxed.

Figure 4.4: The new model structure when constraints (4.2.11c) and (4.2.11d) have been relaxed in problem (4.2.11).

□

As illustrated by Example 1, variable splitting is an innovative and resourceful way to accomplish consensus using Lagrangian duality. For problems with the specific structure concluded in the example, the dual problem has some nice properties which simplify its solution process. Let us contemplate this while we review Example 2.

Example 2. Consider once more problem (4.2.11) from Example 1. Now, let π_n and λ_n, respectively, denote the Lagrangian dual variables corresponding to the relaxation of the constraints (4.2.11d) and (4.2.11c). To simplify the notation in this example, we introduce the variable $x_{t_0} \geq 0$ and define $a_0 := -1$, $b_1 := 0$,

$d_1 := 0$, and let $x_{t_0} \geq 0$. Next, we extend the constraints (4.2.11c) to the case $n = 1$, with the result that x_{t_0} is forced to the value 0. The Lagrangian dual function is then defined by the minimization of the subproblem, according to

$$h(\boldsymbol{\pi}) := \left(\begin{array}{lll} \min\limits_{y, y_n^{\text{split}}, x_t} & \left[\sum\limits_{n \in \mathcal{N}} \frac{c^{\text{inv}}}{N} y_n^{\text{split}} + \sum\limits_{n \in \mathcal{N}} \sum\limits_{t \in \mathcal{T}_n} c_t^{\text{run}} x_t + \sum\limits_{n \in \mathcal{N}} \pi_n (y - y_n^{\text{split}}) \right. & \\ & \left. + \sum\limits_{n \in \mathcal{N}} \lambda_n \left(d_n - (a_{n-1} x_{t_{n-1}} + b_n x_{s_n}) \right) \right] & \\ \text{s.t.} & (x_t)_{t \in \mathcal{T}_n} \in X_n, & n \in \mathcal{N}, \\ & x_t \leq y_n^{\text{split}} \leq Y, & t \in \mathcal{T}_n,\ n \in \mathcal{N}, \\ & y \in [0, Y] & \end{array} \right) \tag{4.2.12a}$$

$$= \min_{y \in [0,Y]} \sum_{n \in \mathcal{N}} \pi_n y \tag{4.2.12b}$$

$$+ \sum_{n \in \mathcal{N}} \left(\begin{array}{lll} \min\limits_{y_n^{\text{split}}, x_t} & \left[\left(\frac{c^{\text{inv}}}{N} - \pi_n \right) y_n^{\text{split}} + \sum\limits_{t \in \mathcal{T}_n} c_t^{\text{run}} x_t \right. & \\ & \left. + \lambda_n \left(d_n - (a_{n-1} x_{t_{n-1}} + b_n x_{s_n}) \right) \right] & \\ \text{s.t.} & (x_t)_{t \in \mathcal{T}_n} \in X_n, & \\ & x_t \leq y_n^{\text{split}} \leq Y, & t \in \mathcal{T}_n \end{array} \right). \tag{4.2.12c}$$

The corresponding dual problem is

$$\begin{array}{lll} \text{maximize} & h(\boldsymbol{\pi}), & \\ \text{subject to} & \pi_n \in \mathbb{R}, & n \in \mathcal{N}, \\ & \lambda_n \geq 0, & n \in \mathcal{N}, \end{array} \tag{4.2.13}$$

and we can see that it has some additional properties:

$$\min_{y \in [0,Y]} \sum_{n \in \mathcal{N}} \pi_n y \implies \begin{cases} y = 0, & \text{when } \sum\limits_{n \in \mathcal{N}} \pi_n \geq 0, \\ y = Y, & \text{when } \sum\limits_{n \in \mathcal{N}} \pi_n < 0. \end{cases} \tag{4.2.14}$$

Thus, the solution to the problem to $\min\limits_{y \in [0,Y]} \sum\limits_{n \in \mathcal{N}} \pi_n y$ will always either be 0 or Y, dependent on the value of $\boldsymbol{\pi}$. It could, however, be argued that the variable y should be unbounded as we lack a reasonable upper bound. Thus,

consider the special case that $Y \to \infty$, such that y is not bounded. Then, for $\sum_{n\in\mathcal{N}} \pi_n < 0$, the subproblem (4.2.12b) is unbounded, thus providing an infeasible primal problem. Therefore, this implicit constraint on the dual variables can be expressed explicitly in the dual problem. Furthermore, since the inequality $\sum_{n\in\mathcal{N}} \pi_n \geq 0$ implies that $y = 0$ in the solution to (4.2.12b), we can remove the variable y in the subproblem when the dual problem is subject to this constraint. Thus, with $y = 0$ the subproblem (4.2.12a) is separable over $n \in \mathcal{N}$, which means that the solution process can be parallelized.

The analogous reasoning for the subproblem in (4.2.12c) yields that if the inequality $\pi_n > \frac{c^{\text{inv}}}{N}$ holds, the nth subproblem will be unbounded, $n \in \mathcal{N}$. Hence we include the explicit constraints $\pi_n \leq \frac{c^{\text{inv}}}{N}$, $n \in \mathcal{N}$, in the dual problem. Under these conditions, the Lagrangian dual problem can thus be stated as to

$$\begin{aligned} \text{maximize} \quad & \hat{h}(\boldsymbol{\pi}), \\ \text{subject to} \quad & \sum_{n\in\mathcal{N}} \pi_n \geq 0, \\ & \pi_n \leq \frac{c^{\text{inv}}}{N}, \qquad n \in \mathcal{N}, \end{aligned} \tag{4.2.15}$$

where

$$\hat{h}(\boldsymbol{\pi}) = \sum_{n\in\mathcal{N}} \left(\begin{array}{ll} \min\limits_{y_n^{\text{split}}, x_t} & \left[\left(\frac{c^{\text{inv}}}{N} - \pi_n \right) y_n^{\text{split}} + \sum\limits_{t\in\mathcal{T}_n} c_t^{\text{run}} x_t \right. \\ & \left. + \lambda_n \left(d_n - (a_{n-1} x_{t_{n-1}} + b_n x_{s_n}) \right) \right] \\ \text{s.t.} & (x_t)_{t\in\mathcal{T}_n} \in X_n, \\ & x_t \leq y_n^{\text{split}}, \quad t \in \mathcal{T}_n \end{array} \right). \tag{4.2.16}$$

□

As concluded in the above examples, combining variable splitting with Lagrangian relaxation yields subproblems which are separable. In the next section, we will discuss different methodologies for solving the decomposed problem.

4.3 Lagrangian dual solution methods

After a problem has been decomposed, it can be solved using different solution methodologies. This section covers the two Lagrangian dual methods that have been evaluated in this work: a subgradient algorithm, and an approximate

consensus ADMM approach.

4.3.1 Subgradient algorithm

The subgradient algorithm was developed by N. Z. Shor in 1962; see Shor [56] for a full review of the early history of nonsmooth optimization. This method has often been applied to solve optimization problems, especially together with Lagrangian duality. Larsson et al. [38] formulated the conditional subgradient method, which combines subgradient methods with subgradient projection methods. However, we begin by a definition.

Definition 1. A vector $\boldsymbol{\gamma} \in \mathbb{R}^n$ is a *subgradient* of the concave function h at $\bar{\boldsymbol{\pi}} \in \mathbb{R}^n$ if the inequality

$$h(\boldsymbol{\pi}) \leq h(\bar{\boldsymbol{\pi}}) + \boldsymbol{\gamma}^\top (\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}) \tag{4.3.1}$$

holds for all $\boldsymbol{\pi} \in \mathbb{R}^n$. The set of subgradients of h at $\bar{\boldsymbol{\pi}}$ is called the *subdifferential*, denoted $\partial h(\bar{\boldsymbol{\pi}})$. □

Geometrically, a subgradient to the function h at the point $\bar{\boldsymbol{\pi}}$ is a vector defining a supporting hyperplane to the epigraph of h at $\bar{\boldsymbol{\pi}}$. The subgradient algorithm, assuming a dual problem on the form as in (4.2.3), is provided in Algorithm 1. Here, we assume that $\boldsymbol{\pi} \in \Pi$, i.e. Π is the feasible set for the multipliers $\boldsymbol{\pi}$.

Algorithm 1 Subgradient algorithm

1: Initiate $\boldsymbol{\pi}^0 \in \Pi$ and $h^0_{\text{best}} = h(\boldsymbol{\pi}^0)$. Let $k := 0$.
2: Find a subgradient to h at the point $\boldsymbol{\pi}^k$
$\implies$ solve the subproblem $\min_{\boldsymbol{x} \in \boldsymbol{X}} L(\boldsymbol{x}, \boldsymbol{\pi}^k)$,
which gives a solution $\boldsymbol{x}(\boldsymbol{\pi}^k)$.
A subgradient to h is then given by $\boldsymbol{\gamma}^k := \boldsymbol{g}(\boldsymbol{x}(\boldsymbol{\pi}^k))$.
3: For some $\alpha_k > 0$, calculate the new point $\boldsymbol{\pi}^{k+1} := \text{Proj}_\Pi\{\boldsymbol{\pi}^k + \alpha_k \boldsymbol{\gamma}^k\}$.
4: Update $h^{k+1}_{\text{best}} := \max\{h^k_{\text{best}}, h(\boldsymbol{\pi}^{k+1})\}$.
5: Termination criteria fulfilled $\implies$ stop.
Otherwise, let $k := k + 1$ and go to 2.

The step lengths α_k are chosen according to some rule which guarantees convergence. The Polyak step length rule [45] has seen much use in practice, and is used in Papers I and II. It is defined as

$$\alpha_k = \frac{\theta_k \left(h^* - h(\boldsymbol{\pi}^k)\right)}{|| \boldsymbol{g}(\boldsymbol{x}(\boldsymbol{\pi}^k))||_2^2} \tag{4.3.2}$$

under the conditions

$$0 < \epsilon_1 \leq \theta_k \leq 2 - \epsilon_2 < 2, \quad k = 0, 1, 2, ... \tag{4.3.3}$$

Here, θ_k acts as a scaling parameter for the step length, and the parameters ϵ_1 and ϵ_2 define positive limits for the scaling parameter. Polyak showed that using (4.3.2) together with (4.3.3) guarantees theoretical convergence to an optimal solution to the dual problem (4.2.3). However, the dual optimal value h^* is typically not known. If so, an upper bound $\bar{h} \geq h^*$ can be used instead to achieve finite convergence to an ϵ-optimal solution, where we define ϵ-optimal as $h(\boldsymbol{\pi}^k) \geq 2h^* - \bar{h} - \epsilon$ for any $\epsilon > 0$; see [45, Theorem 4].

In Papers I and II, the upper bound is given by solving the full-scale model but with isolated regions and thus without trade. This provides a feasible solution, but not optimal due to the lack of trade (unless the original problem instance only considers a single region). The upper bound is then the sum of the total system costs for all the included regions. Another alternative, as examined in Paper II, is to solve the full-scale model with trade but with disconnected years. The model is then solved iteratively for each year, where the investments from previous years are fixed to the solution provided by the previous iteration.

The value of the scaling parameter θ_k can be chosen in different ways. One method that in practice has been shown to give fast convergence to an optimal solution is presented by Caprara et al. [5]. The authors present the use of an adaptive method, where the value of the parameter is updated every few subgradient iterations; see Paper I for further details on how this was implemented.

Recovery of feasible solutions

In general, subgradient optimization methods often identify near-optimal dual solutions, but do not directly provide solutions to the primal problem. The conditional subgradient method constructs a sequence $\{\boldsymbol{x}(\boldsymbol{\pi}^k)\}$ of solutions to the Lagrangian subproblem, but these solutions are typically not feasible in the original primal problem since they do not have to satisfy the relaxed constraints. Thus, the sequence $\{\boldsymbol{x}(\boldsymbol{\pi}^k)\}$ does not converge to the optimal primal solution. To remedy this, we generated in Papers I and II *ergodic sequences of subproblem solutions*. As first presented by Larsson et al. [37], ergodic sequences create approximations of primal solutions by averaging the solutions from the subproblems. The authors showed that the ergodic sequences in the limit produce optimal solutions to the original problem. An enchanced version in terms of convergence speed was introduced by Gustavsson et al. [24]. This

version exploits more information from later subproblem solutions than from earlier ones. See also Paper I for additional details on this approach.

Termination criterion

In Papers I and II, we use as a termination criterion for the duality gap such that the algorithm terminates when the following criterion is satisfied:

$$\frac{\text{UBD} - \text{LBD}}{\text{UBD}} < \epsilon^{\text{gap}}. \tag{4.3.4}$$

Here, UBD is an upper bound on the objective value of the primal problem and LBD is a lower bound on the optimal objective value of the primal problem. The left–hand–side is defined as the relative duality gap, and we say that it should be smaller than some small positive number ϵ^{gap}. Since we are working with an LP, strong duality implies that the duality gap is zero in the optimal solution. For numerical reasons however, we let $\epsilon^{\text{gap}} := 10^{-4}$ in our implementation. Moreover, convergence is typically sluggish and thus a large number of iterations are necessary to fully (or approximately) close the dual gap. We noticed however in our implementation that the solutions became reasonable after only a few iterations. Therefore, the algorithm was terminated after a fixed number of iterations.

Lower bounds on the optimal objective value of the primal problem are given by the Lagrangian function evaluated for the Lagrangian multipliers, and thus a result of the subgradient algorithm. In Paper II, the upper bound is given by either 1) solving the full-scale model with isolated regions (and thus solve it once for every region, without any option for trade), or 2) by disconnecting the years and solve the full-scale model sequentially, where we in each year use the investments made previous years.

Updating the dual multipliers

If the dual problem has constraints on the dual variables, then the update from the subgradient algorithm may lead to values on the multipliers that are infeasible. If so, the multiplier needs to be projected on the feasible region. Let

$$\bar{\boldsymbol{\pi}} := \boldsymbol{\pi}^k - \alpha_k \boldsymbol{\gamma}^k$$

as in the subgradient algorithm update. For most multiplier vectors $\boldsymbol{\pi}$, the projection according to step 3 in Algorithm 1 can be made separately for each

element of each vector. Thus, the projection is given by

$$\pi_j^{k+1} := \begin{cases} \bar{\pi}_j, & \bar{\pi}_j \geq 0, \\ 0, & \bar{\pi}_j < 0, \end{cases}$$

where j refers to element j in the respective multiplier vector.

For the multiplier π_n in the dual problem (4.2.15) in Example 2, with $c := \dfrac{c^{\text{inv}}}{N}$, the projection corresponds to solving the problem to

$$\text{minimize} \qquad \tfrac{1}{2}||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||^2 = \tfrac{1}{2} \sum_{n \in \mathcal{N}} (\pi_n - \bar{\pi}_n)^2, \tag{4.3.5a}$$

$$\text{subject to} \qquad \sum_{n \in \mathcal{N}} \pi_n \geq 0, \tag{4.3.5b}$$

$$\pi_n \leq c, \quad n \in \mathcal{N}, \tag{4.3.5c}$$

with the KKT conditions (where λ and ψ_n are multipliers of the constraints (4.3.5b) and (4.3.5c), respectively)

$$\begin{aligned} \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_N \end{pmatrix} - \begin{pmatrix} \bar{\pi}_1 \\ \vdots \\ \bar{\pi}_N \end{pmatrix} - \lambda \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \sum_{n \in \mathcal{N}} \psi_n \boldsymbol{e}_n &= 0, \\ \lambda(-\sum_{n \in \mathcal{N}} \pi_n) &= 0, \\ \psi_n(\pi_n - c) &= 0, \qquad n \in \mathcal{N}. \end{aligned} \tag{4.3.6}$$

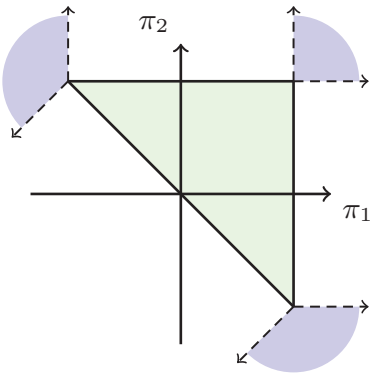


Figure 4.5: The projection problem for $N = 2$. Here, the green area is the feasible region. For points inside the blue cones, the projection should be onto the corresponding extreme points. For any other points outside the green area, the projection is defined by the corresponding shortest Euclidean distance to the triangle. The blue cones are spanned by the gradients of the active constraints in the extreme points.

The projection problem, illustrated in Figure 4.5 for the case $N = 2$, has the property that it is a convex quadratic program. Therefore, the Lagrangian dual

of this problem is also a convex quadratic program. To see this, Lagrangian relax the constraint (4.3.5b). Then, using the Lagrange multiplier λ, the Lagrangian function is

$$L(\boldsymbol{\pi}, \lambda) := \tfrac{1}{2} \sum_{n \in \mathcal{N}} (\pi_n - \bar{\pi}_n)^2 - \lambda \sum_{n \in \mathcal{N}} \pi_n = \sum_{n \in \mathcal{N}} \left(\tfrac{1}{2} (\pi_n - \bar{\pi}_n)^2 - \lambda \pi_n \right).$$

Let $L_n(\pi_n, \lambda) := \frac{1}{2}(\pi_n - \bar{\pi}_n)^2 - \lambda \pi_n$, such that $L(\boldsymbol{\pi}, \lambda) = \sum_{n \in \mathcal{N}} L_n(\pi_n, \lambda)$ holds. It follows that the function L is separable with respect to $n \in \mathcal{N}$, where L_n is a quadratic, convex, and differentiable function with respect to its first argument.

Now, define

$$h_n(\lambda) := \begin{array}{ll} \text{minimum} & L_n(\pi_n, \lambda) \\ \text{subject to} & \pi_n \leq c \end{array} = \begin{array}{ll} \text{minimum} & \frac{1}{2}(\pi_n - \bar{\pi}_n)^2 - \lambda \pi_n \\ \text{subject to} & \pi_n \leq c \end{array}$$

so that the Lagrangian dual function is given by

$$h(\lambda) = \sum_{n \in \mathcal{N}} h_n(\lambda) = \sum_{n \in \mathcal{N}} \left(\begin{array}{ll} \text{minimum} & \frac{1}{2}(\pi_n - \bar{\pi}_n)^2 - \lambda \pi_n \\ \text{subject to} & \pi_n \leq c \end{array} \right).$$

For a constant value of λ, the minimum for L_n over $\pi_n \in [-\infty, c]$ is attained when either $\frac{\partial L_n(\pi_n, \lambda)}{\partial \pi_n} = 0$ or $\pi_n = c$. Since

$$\frac{\partial L_n(\pi_n, \lambda)}{\partial \pi_n} = \pi_n - \bar{\pi}_n - \lambda = 0 \iff \pi_n = \bar{\pi}_n + \lambda,$$

it follows that $L_n(\cdot, \lambda)$ attains its minimum for

$$\pi_n = \begin{cases} \bar{\pi}_n + \lambda, & \bar{\pi}_n \leq c - \lambda, \\ c, & \bar{\pi}_n > c - \lambda. \end{cases} \tag{4.3.7}$$

Hence, the function h can be expressed according to the following:

$$\begin{aligned} h(\lambda) &= \sum_{n \in \mathcal{N} : \bar{\pi}_n \leq c - \lambda} \left(\tfrac{1}{2}\lambda^2 - \lambda(\bar{\pi}_n + \lambda) \right) + \sum_{n \in \mathcal{N} : \bar{\pi}_n > c - \lambda} \left(\tfrac{1}{2}(c - \bar{\pi}_n)^2 - \lambda c \right) \\ &= \sum_{n \in \mathcal{N} : \bar{\pi}_n \leq c - \lambda} \left(-\tfrac{1}{2}\lambda^2 - \lambda \bar{\pi}_n \right) + \sum_{n \in \mathcal{N} : \bar{\pi}_n > c - \lambda} \left(\tfrac{1}{2}(c - \bar{\pi}_n)^2 - \lambda c \right). \end{aligned}$$

The function h is clearly a quadratic and concave function of λ.

The Lagrangian dual problem is given by

$$\begin{array}{ll} \text{maximum} & h(\lambda), \\ \text{subject to} & \lambda \geq 0. \end{array}$$

The dual function is differentiable, and thus the derivative can be used to find the maximum. Hence, the dual problem is maximized when $\frac{\partial h(\lambda)}{\partial \lambda} = 0$ or $\lambda = 0$:

$$\begin{aligned} \frac{\partial h(\lambda)}{\partial \lambda}\Big|_{\lambda} &= -\sum_{n\in\mathcal{N}:\bar{\pi}_n+\lambda\leq c} (\lambda + \bar{\pi}_n) - \sum_{n\in\mathcal{N}:\bar{\pi}_n+\lambda>c} c \\ \implies \frac{\partial h(\lambda)}{\partial \lambda} &= -\sum_{n\in\mathcal{N}} \min\{\bar{\pi}_n + \lambda\,; c\} = 0. \end{aligned} \tag{4.3.8}$$

The partial derivative $\frac{\partial h(\lambda)}{\partial \lambda}$ is decreasing (though not strictly) for increasing values of λ. Hence, it will be zero for a specific value of λ. As an example, consider Figure 4.6, where we use the property

$$-\min\{\bar{\pi}_n + \lambda\,; c\} = \max\{-(\bar{\pi}_n + \lambda)\,; -c\}.$$

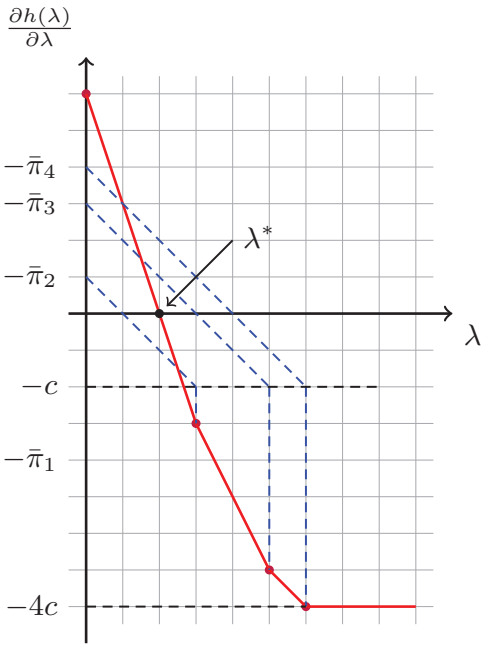


Figure 4.6: The partial derivative as a function of λ. The optimal value λ^* is where the partial derivative is equal to zero.

The gradient of the function h changes in the break points corresponding to $\lambda = c - \bar{\pi}_n$, $n \in \mathcal{N}$. Thus, between two break points, the point at which $\frac{\partial h(\lambda)}{\partial \lambda} = 0$ holds will be passed, and this corresponds to the optimal value λ^*.

Using the optimal value λ^*, the optimal values π_n^*, $n \in \mathcal{N}$, are then given by (4.3.7). The method discussed here to find the projection point is described in Algorithm 2.

Algorithm 2 Algorithm to solve the projection problem

if $\bar{\pi}_n > 0$ for all $n \in \mathcal{N}$ **then**
    $\lambda^* = 0$
else
    Calculate all break points $\lambda_n := c - \bar{\pi}_n$, $n \in \mathcal{N}$, and define $\lambda_0 := 0$.
    Let $\mathcal{N}_0 := \mathcal{N} \cup \{0\}$
    Calculate the partial derivative for all break points, i.e. $\frac{\partial h(\lambda_n)}{\partial \lambda}$, $n \in \mathcal{N}_0$
    Find λ_i and λ_j such that

$$\frac{\partial h(\lambda_i)}{\partial \lambda} = \min_{n \in \mathcal{N}_0 : \frac{\partial h(\lambda_n)}{\partial \lambda} \geq 0} \left\{ \frac{\partial h(\lambda_n)}{\partial \lambda} \right\} \quad \text{and} \quad \frac{\partial h(\lambda_j)}{\partial \lambda} = \max_{n \in \mathcal{N}_0 : \frac{\partial h(\lambda_n)}{\partial \lambda} \leq 0} \left\{ \frac{\partial h(\lambda_n)}{\partial \lambda} \right\}$$

    if $\{n \in \mathcal{N}_0 : \frac{\partial h(\lambda_n)}{\partial \lambda} \geq 0\} = \emptyset$ **then**
        $\lambda^* := 0$
    else if $\lambda_i = \lambda_j$ **then**
        $\lambda^* := \lambda_i = \lambda_j$
    else
        Linear interpolation gives

$$\lambda^* := \frac{(0 - \frac{\partial h(\lambda_i)}{\partial \lambda})(\lambda_j - \lambda_i)}{\frac{\partial h(\lambda_j)}{\partial \lambda} - \frac{\partial h(\lambda_i)}{\partial \lambda}} + \lambda_i$$

        if $\lambda^* < 0$ **then**
            $\lambda^* := 0$

Use λ^* to calculate π_n^*, $n \in \mathcal{N}$, according to (4.3.7)

If there does not exist a point $\lambda \geq 0$, for which $\frac{\partial h(\lambda)}{\partial \lambda} = 0$ holds, then this implies that $\frac{\partial h(\lambda)}{\partial \lambda} = -\sum_{n \in \mathcal{N}} \min\{\bar{\pi}_n + \lambda\,; c\} \neq 0$. This can happen when $\bar{\pi}_n > 0$ for all $n \in \mathcal{N}$, which implies that $\lambda^* = 0$ and the function $h(\lambda)$ has no zero derivative for $\lambda \geq 0$. There are also some other special cases that can occur which we need to consider; see Examples B.1 and B.2 in Appendix B.

It should be noted that if thermal cycling constraints are included, the dual problem and the corresponding projection problem will be to

$$\begin{aligned}
\text{minimize} \quad & \tfrac{1}{2}||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||^2 = \tfrac{1}{2}\sum_{n\in\mathcal{N}}(\pi_n - \bar{\pi}_n)^2 + \frac{1}{2}\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}}(\mu_{kn} - \bar{\mu}_{kn})^2, \\
\text{subject to} \quad & \sum_{n\in\mathcal{N}} \pi_n \geq 0, \\
& \pi_n + \sum_{k\in\mathcal{K}} \mu_{kn} \leq c, \quad n \in \mathcal{N}, \\
& \mu_{kn} \geq 0, \quad k \in \mathcal{K},\, n \in \mathcal{N},
\end{aligned}$$

which is a bit more complicated than problem (4.3.5) above. We noticed however that, at least in our implementation, the constraints $\pi_n + \sum_{k\in\mathcal{K}} \mu_{kn} \leq c$, $n \in \mathcal{N}$, was consistently satisfied throughout each subgradient iteration of the examined problem instances. Hence, although our problems included thermal cycling, Algorithm 2 was still applied successfully to solve the projection problem for the dual multipliers.

4.3.2 Alternating direction method of multipliers (ADMM)

Augmented Lagrangian methods add an additional penalty term (i.e. the augmentation) to the Lagrangian function. In particular, the alternating direction method of multipliers (ADMM) is an algorithm which uses partial updates for the dual variables. It combines properties from dual ascent methods, which typically allow problems to be decomposed and thus solved in parallel, with good convergence properties from the method of multipliers. See Boyd et al. [2] for extensive details on ADMM and its background.

Consider the convex optimization problem to

$$\begin{aligned}
\text{minimize} \quad & v^* := \quad f(\boldsymbol{z}) + g(\boldsymbol{y}), \\
\text{subject to} \quad & \qquad \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{c}
\end{aligned} \tag{4.3.9}$$

with variables $\boldsymbol{z} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^\ell$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} \in \mathbb{R}^{m\times \ell}$, $\boldsymbol{c} \in \mathbb{R}^m$, and $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^\ell \to \mathbb{R}$ are convex functions. We also assume that $\{\boldsymbol{z} \in \mathbb{R}^n, \boldsymbol{y} \in \mathbb{R}^\ell \mid \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{c}\} \neq \emptyset$ such that there exists a feasible solution.

The augmented Lagrangian to problem (4.3.9), where $\boldsymbol{\pi} \in \mathbb{R}^m$ is the dual variable, is expressed as

$$L_\rho(\boldsymbol{z}, \boldsymbol{y}; \boldsymbol{\pi}) := f(\boldsymbol{z}) + g(\boldsymbol{y}) + \boldsymbol{\pi}^\top(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{c}) + \frac{\rho}{2}||\,\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{c}||_2^2, \tag{4.3.10}$$

and we define the ADMM-algorithm as the iterative updates

$$\boldsymbol{z}^{k+1} := \underset{\boldsymbol{z}}{\operatorname{argmin}}\, L_\rho(\boldsymbol{z}, \boldsymbol{y}^k; \boldsymbol{\pi}^k), \tag{4.3.11a}$$

$$\boldsymbol{y}^{k+1} := \underset{\boldsymbol{y}}{\operatorname{argmin}}\, L_\rho(\boldsymbol{z}^{k+1}, \boldsymbol{y}; \boldsymbol{\pi}^k), \tag{4.3.11b}$$

$$\boldsymbol{\pi}^{k+1} := \boldsymbol{\pi}^k + \rho\left(\boldsymbol{A}\boldsymbol{z}^{k+1} + \boldsymbol{B}\boldsymbol{y}^{k+1} - \boldsymbol{c}\right), \tag{4.3.11c}$$

where $k > 0$ is the iterator and $\rho > 0$ is a constant parameter. Here, (4.3.11a) is a $\boldsymbol{z}$-minimization step, (4.3.11b) is a $\boldsymbol{y}$-minimization step, and (4.3.11c) is a dual variable update with the step size equal to the augmented Lagrangian parameter ρ. Typically, $\boldsymbol{z}^0$ and $\boldsymbol{y}^0$ are some feasible starting points, with some arbitrarily value of $\boldsymbol{\pi}^0 \in \mathbb{R}^m$.

Multi-block ADMM

The two-block ADMM can be extended to a multi-block approach, where the problem is divided into more than two parts, hence yielding further sequential updates each iteration. Revisit once more problem (4.2.11) in Example 1, but with added slack variables $\boldsymbol{u}$ for the inequality constraints (4.2.11c):

$$\underset{y, y_n^{\text{split}}, x_t}{\text{minimize}} \quad \frac{c^{\text{inv}}}{N} \sum_{n\in\mathcal{N}} y_n^{\text{split}} + \sum_{n\in\mathcal{N}} \sum_{t\in\mathcal{T}_n} c_t^{\text{run}} x_t, \tag{4.3.12a}$$

$$\text{subject to} \quad (x_t)_{t\in\mathcal{T}_n} \in X_n, \qquad n \in \mathcal{N}, \tag{4.3.12b}$$

$$a_{n-1}x_{t_{n-1}} + b_n x_{s_n} = u_n + d_n, \qquad n \in \mathcal{N}\setminus\{1\}, \tag{4.3.12c}$$

$$y_n^{\text{split}} = y, \qquad n \in \mathcal{N}, \tag{4.3.12d}$$

$$x_t \le y_n^{\text{split}} \le Y, \qquad t \in \mathcal{T}_n, n \in \mathcal{N}, \tag{4.3.12e}$$

$$y \in [0, Y]. \tag{4.3.12f}$$

Let $\bar{\boldsymbol{x}}_n := ((x_t)_{t\in\mathcal{T}_n})$ and define $\boldsymbol{z}_n := (y_n^{\text{split}}, \bar{\boldsymbol{x}}_n, u_n)$. For $n \in \mathcal{N}$, let Z_n denote the remaining feasible set for the variables $\boldsymbol{z}_n$ when constraints (4.3.12c) and (4.3.12d) are relaxed. We can then rewrite problem (4.3.12) in matrix form, where we then consider the problem to

$$\underset{0\le y\le Y,\, \boldsymbol{z}_n\in Z_n}{\text{minimize}} \sum_{n\in\mathcal{N}} f_n(\boldsymbol{z}_n) + g(y), \tag{4.3.13a}$$

$$\text{subject to} \sum_{n\in\mathcal{N}} \boldsymbol{A}_n\boldsymbol{z}_n + \boldsymbol{b}y - \boldsymbol{c} = \boldsymbol{0}. \tag{4.3.13b}$$

Here, we have $|\mathcal{N}| + 1 = N + 1$ blocks. Define the augmented Lagrangian as

$$\begin{aligned}\hat{L}_\rho := \sum_{n\in\mathcal{N}} f_n(\boldsymbol{z}_n) + g(y) + \boldsymbol{\pi}^\top \left(\sum_{n\in\mathcal{N}} \boldsymbol{A}_n \boldsymbol{z}_n + \boldsymbol{b}y - \boldsymbol{c}\right) \\ + \frac{\rho}{2}\left\|\sum_{n\in\mathcal{N}} \boldsymbol{A}_n \boldsymbol{z}_n + \boldsymbol{b}y - \boldsymbol{c}\right\|_2^2,\end{aligned} \tag{4.3.14}$$

and the corresponding multi-block ADMM algorithm is as follows:

$$\boldsymbol{z}_n^{k+1} := \underset{\boldsymbol{z}_n}{\operatorname{argmin}}\, \hat{L}_\rho(\boldsymbol{z}_1^{k+1}, \ldots, \boldsymbol{z}_n, \ldots, \boldsymbol{z}_N^k, y^k; \boldsymbol{\pi}^k), \qquad n \in \mathcal{N}, \tag{4.3.15a}$$

$$y^{k+1} := \underset{y}{\operatorname{argmin}}\, \hat{L}_\rho(\boldsymbol{z}_1^{k+1}, \ldots, \boldsymbol{z}_N^{k+1}, y; \boldsymbol{\pi}^k), \tag{4.3.15b}$$

$$\boldsymbol{\pi}^{k+1} := \boldsymbol{\pi}^k + \rho\left(\sum_{n\in\mathcal{N}} \boldsymbol{A}_n \boldsymbol{z}_n^{k+1} + \boldsymbol{b}y^{k+1} - \boldsymbol{c}\right). \tag{4.3.15c}$$

Approximate consensus ADMM

Define $f(\boldsymbol{x}) := \sum_{n\in\mathcal{N}} f_n(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^m$ and consider the problem to

$$\underset{\boldsymbol{x}_n \in X_n, y\in Y}{\text{minimize}} \quad \sum_{n\in\mathcal{N}} f_n(\boldsymbol{x}_n), \tag{4.3.16a}$$

$$\text{subject to} \quad \boldsymbol{x}_n - \boldsymbol{y} = \boldsymbol{0}, \qquad n \in \mathcal{N}, \tag{4.3.16b}$$

such that $\boldsymbol{y} \in \mathbb{R}^m$ is a common global variable, $\boldsymbol{x}_n \in \mathbb{R}^m$ with $\mathcal{N} := \{1, \ldots, N\}$ are local variables and $f_n : \mathbb{R}^m \to \mathbb{R}$ are convex functions. The constraint expresses that all the variables should be equal, i.e. reach consensus. As stated by Boyd et al. [2], it can be seen as a technique for turning additive objectives $\sum_{n\in\mathcal{N}} f_n(\boldsymbol{x})$, which do not split across terms, into separable objectives $\sum_{n\in\mathcal{N}} f_n(\boldsymbol{x}_n)$ which do split.

Now, the augmented Lagrangian is given by

$$\begin{aligned}\tilde{L}_\rho(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N, \boldsymbol{y}; \boldsymbol{\pi}) &:= \sum_{n\in\mathcal{N}} \tilde{L}_{\rho,n}(\boldsymbol{x}_n, \boldsymbol{y}; \boldsymbol{\pi}_n) \\ &:= \sum_{n\in\mathcal{N}} \left(f_n(\boldsymbol{x}_n) + \boldsymbol{\pi}_n^\top(\boldsymbol{x}_n - \boldsymbol{y}) + \frac{\rho}{2}||\,\boldsymbol{x}_n - \boldsymbol{y}\,||_2^2\right).\end{aligned}$$

The corresponding ADMM algorithm, referred to as *consensus ADMM*, is

$$\boldsymbol{x}_n^{k+1} := \operatorname*{argmin}_{\boldsymbol{x}_n} \tilde{L}_{\rho,n}(\boldsymbol{x}_n, \boldsymbol{y}^k; \boldsymbol{\pi}^k), \qquad n \in \mathcal{N}, \tag{4.3.17a}$$

$$\boldsymbol{y}^{k+1} := \frac{1}{N} \sum_{n \in \mathcal{N}} \left(\boldsymbol{x}_n^{k+1} + (1/\rho)\boldsymbol{\pi}_n^k \right), \tag{4.3.17b}$$

$$\boldsymbol{\pi}_n^{k+1} := \boldsymbol{\pi}_n^k + \rho \left(\boldsymbol{x}^{k+1} - \boldsymbol{y}^{k+1} \right), \qquad n \in \mathcal{N}. \tag{4.3.17c}$$

See also Parikh and Boyd [43] for additional details on consensus ADMM. Note that the consensus ADMM approach allows us to update the local variables in step (4.3.17a) in parallel.

Let us once again consider problem (4.3.12). Its structure contains blocks that are *sequentially connected*, meaning that each subproblem n shares some variables from subproblems $n-1$ and $n+1$. Hence, although our example has many similarities with problem (4.3.16), the connected constraints force the sequential update seen in (4.3.15).

However, in Paper III we consider an *approximate consensus ADMM approach*, where we—like in the consensus ADMM—solve the subproblems in parallel. Thus, using the same variable definitions as in problem (4.3.13), i.e. $\bar{\boldsymbol{x}}_n := ((x_t)_{t \in \mathcal{T}_n})$ and $\boldsymbol{z}_n := (y_n^{\text{split}}, \bar{\boldsymbol{x}}_n, u_n) \in Z_n$, and with $\hat{L}_\rho$ defined as in (4.3.14), our new scheme becomes

$$\boldsymbol{z}_n^{k+1} := \operatorname*{argmin}_{\boldsymbol{z}_n} \hat{L}_\rho(\boldsymbol{z}_1^k, \ldots, \boldsymbol{z}_n, \ldots, \boldsymbol{z}_N^k, y^k; \boldsymbol{\pi}^k), \qquad n \in \mathcal{N}, \tag{4.3.18a}$$

$$y^{k+1} := \operatorname*{argmin}_{y} \hat{L}_\rho(\boldsymbol{z}_1^{k+1}, \ldots, \boldsymbol{z}_N^{k+1}, y; \boldsymbol{\pi}^k), \tag{4.3.18b}$$

$$\boldsymbol{\pi}^{k+1} := \boldsymbol{\pi}^k + \rho \left(\sum_{n \in \mathcal{N}} \boldsymbol{A}_n \boldsymbol{z}_n^{k+1} + \boldsymbol{b} y^{k+1} - \boldsymbol{c} \right). \tag{4.3.18c}$$

Note that unlike in problem (4.3.16) and its corresponding consensus ADMM algorithm (4.3.17), y is in this problem a scalar with the same dimension as x_t.

Convergence

The traditional two-block ADMM, i.e. the algorithm (4.3.11), converges under the assumptions that 1) the functions f and g are closed, proper and convex, and that 2) the unaugmented Lagrangian function $L_0(\boldsymbol{z}, \boldsymbol{y}; \boldsymbol{\pi})$ has a saddle point; see also Boyd et al. [2]. In practice, the ADMM algorithm tends to converge very slowly to high accuracy. However, the method often converges

to modest accuracy within a few tens of iterations and thereby provides results acceptable for most large-scale problems. Nevertheless, the choice of ρ tends to greatly influence the practical convergence of the ADMM-algorithm. If ρ is too large, then there is not enough emphasis on minimizing the objective. On the other hand, if ρ is too small, then there is not enough emphasis on feasibility. Consensus ADMM is considered robust in the choice of ρ as it is guaranteed to converge for any choice of $\rho > 0$, under the same two assumptions as the standard ADMM (Chang et al. [6]). See also Shi et al. [54, 55] and Hong and Luo [29] for additional results on convergence for consensus ADMM.

Multi-block ADMM is not necessarily convergent. According to Chen et al. [7], convergence requires the same assumptions as the two-block ADMM, and the additional condition that any two coefficient matrices in a general multi-block problem are orthogonal, i.e. in the problem to

$$\begin{aligned} &\underset{\boldsymbol{x}_n}{\text{minimize}} \sum_{n=1}^{N} f_n(\boldsymbol{x}_n), \\ &\text{subject to} \sum_{n=1}^{N} \boldsymbol{A}_n \boldsymbol{x}_n - \boldsymbol{c} = \boldsymbol{0}, \end{aligned}$$

for some $N \geq 3$, the conditions $\boldsymbol{A}_i^\top \boldsymbol{A}_j = \boldsymbol{0}$ for all $i, j = 1, \ldots, N$, $i \neq j$, should be satisfied. This latter requirement is however typically difficult to examine in practice. However, some strategies exist for dealing with lack of convergence; Hong and Luo [29] e.g. added a relaxation factor to the Lagrange-multiplier update step. Sun et al. [57] suggested to interchange the order of the variable update by, in each ADMM iteration, solve or update the blocks in a randomly permuted order. Consequently, this latter approach successfully eliminated the divergence example constructed in [7].

To summarize, the three well-established ADMM algorithms discussed here (i.e. two-block, multi-block, and consensus ADMM) have promising chances of convergence. The approximate consensus algorithm in Paper III, however, to the best of our knowledge, currently lacks proper convergence conditions.

Termination criterion

Consider again problem (4.2.11) on matrix form, i.e. the problem (4.3.13). A common stopping criterion considers the primal and dual feasibility. The primal residual, here denoted $\boldsymbol{r}^k$ at iteration k, is given by the sum of all

subgradients. For the problem (4.3.13), this implies

$$\boldsymbol{r}^k := \sum_{n\in\mathcal{N}} \boldsymbol{A}_n \boldsymbol{z}_n^k + \boldsymbol{b} y^k - \boldsymbol{c} = \boldsymbol{0}. \tag{4.3.20}$$

The dual feasibility conditions are

$$\begin{aligned} &\boldsymbol{0} \in \partial f_n(\boldsymbol{z}_n^*) + \boldsymbol{A}_n^\top \boldsymbol{\pi}^*, \quad n \in \mathcal{N}, \\ &\boldsymbol{0} \in \partial g(y^*) + \boldsymbol{b}^\top \boldsymbol{\pi}^*. \end{aligned} \tag{4.3.21}$$

The latter condition is always satisfied for y^{k+1} (see details in Boyd et al. [2]). For some $n \in \mathcal{N}$, the former conditions give us, since $\boldsymbol{z}_n^{k+1}$ minimizes $\hat{L}_\rho(\boldsymbol{z}_1^k, \ldots, \boldsymbol{z}_n, \ldots, \boldsymbol{z}_N^k, y^k; \boldsymbol{\pi}^k)$ by definition,

$$\begin{aligned} \boldsymbol{0} &\in \partial f_n(\boldsymbol{z}_n^{k+1}) + \boldsymbol{A}_n^\top \boldsymbol{\pi}^k + \rho \boldsymbol{A}_n^\top \left(\sum_{i\in\mathcal{N}\setminus n} \boldsymbol{A}_i \boldsymbol{z}_i^k + \boldsymbol{A}_n \boldsymbol{z}_n^{k+1} + \boldsymbol{b} y^k - \boldsymbol{c} \right) \\ &= \partial f_n(\boldsymbol{z}_n^{k+1}) \\ &\quad + \boldsymbol{A}_n^\top \left(\boldsymbol{\pi}^k + \rho \boldsymbol{r}^{k+1} + \rho \left(\sum_{i\in\mathcal{N}\setminus n} \boldsymbol{A}_i \left(\boldsymbol{z}_i^k - \boldsymbol{z}_i^{k+1} \right) + \boldsymbol{b} \left(y^k - y^{k+1} \right) \right) \right) \\ &= \partial f_n(\boldsymbol{z}_n^{k+1}) + \boldsymbol{A}_n^\top \boldsymbol{\pi}^{k+1} + \rho \boldsymbol{A}_n^\top \left(\sum_{i\in\mathcal{N}\setminus n} \boldsymbol{A}_i \left(\boldsymbol{z}_i^k - \boldsymbol{z}_i^{k+1} \right) + \boldsymbol{b} \left(y^k - y^{k+1} \right) \right), \end{aligned}$$

or equivalently,

$$\rho \boldsymbol{A}_n^\top \left(\sum_{i\in\mathcal{N}\setminus n} \boldsymbol{A}_i \left(\boldsymbol{z}_i^{k+1} - \boldsymbol{z}_i^k \right) + \boldsymbol{b} \left(y^{k+1} - y^k \right) \right) \in \partial f_n(\boldsymbol{z}_n^{k+1}) + \boldsymbol{A}_n^\top \boldsymbol{\pi}^{k+1}. \tag{4.3.22}$$

The dual residual at iteration k is thus defined as

$$\boldsymbol{s}_n^k := \rho \boldsymbol{A}_n^\top \left(\sum_{i\in\mathcal{N}\setminus n} \boldsymbol{A}_i \left(\boldsymbol{z}_i^k - \boldsymbol{z}_i^{k-1} \right) + \boldsymbol{b} \left(y^k - y^{k-1} \right) \right). \tag{4.3.23}$$

The algorithm is terminated when the primal and dual residuals satisfy a stopping criterion, i.e., for some tolerances $\epsilon^{\text{primal}} > 0$ and $\epsilon^{\text{dual}} > 0$,

$$|| \boldsymbol{r}^k ||_2 \leq \epsilon^{\text{primal}}, \quad || \boldsymbol{s}^k ||_2 \leq \epsilon^{\text{dual}}. \tag{4.3.24}$$

4.4 Heuristic consensus algorithm

Distributed computing and multi-agent systems in computer science often require their processes to reach consensus, thus coordinating their many candidate values to agree on a single consensus value. Different methods for this can be applied, dependent on the studied application.

In Paper IV, a heuristic consensus algorithm is used to connect the models separated by the 2-week segments. When the algorithm begins, the starting point is that all 2-week segments share the investment cost equally, i.e. the costs are weighted by $1/26$. The 26 different electricity system models are then solved, and information on investments in different types of generation, transmission, and storage capacity in each 2-week segment is collected to form one capacity–cost curve for each technology and region. The idea is that the investments form the basis for the investment cost in the subsequent solve. The cost of the capacity that is invested in all 2-week segments is weighted by $1/26$, however, if, e.g. only k problems have made the investment, the capacity is weighted by $1/k$ of the investment cost for all 2-week segments in the next iteration. Figure 4.7 illustrates an example of a capacity–cost curve. The remainder of this section presents the capacity–cost curve construction for generation technologies; the construction for transmission technologies are done analogously. Figure 4.8 presents a schematic illustration of the methodology.

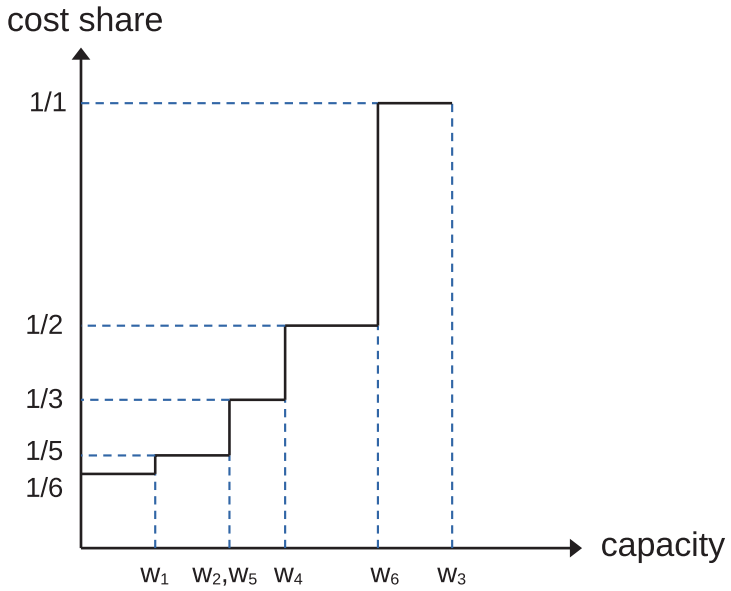


Figure 4.7: An example of a capacity–cost curve. Let $w_1 = 20$, $w_2 = 40$, $w_3 = 100$, $w_4 = 55$, $w_5 = 40$, and $w_6 = 80$. Then, $L = \{w_1, w_2, w_3, w_4, w_5, w_6\}$ and sorted as $L = \{w_1, w_2, w_5, w_4, w_6, w_3\}$

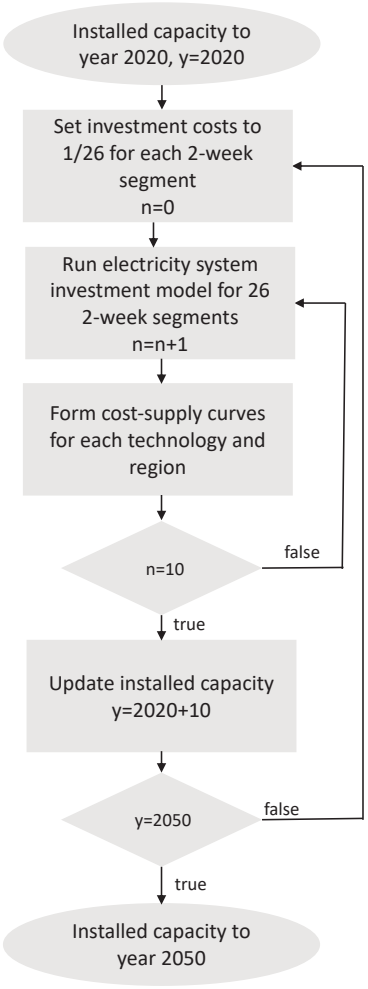


Figure 4.8: A schematic illustration of the modelling methodology.

The capacity–cost curves are composed by 26 steps, where the length of the first step corresponds to the capacity investment level that is common to all 26 subproblems. The length of the second step represents, in addition to the first step, the capacity investment which is shared by all the 2-week segments except one, and so on. In order to determine the lengths of the steps, the number R_{ips} of 2-week segments that have lower or equal levels of installed capacity of technology $p \in \mathcal{P}$ in region $i \in \mathcal{I}$ than the 2-week segment $s \in \mathcal{S}$ is calculated as[1]

$$R_{ips} = 1 + S - \sum_{u \in \mathcal{S}} [w_{ipu} \leq w_{ips}], \qquad i \in \mathcal{I},\ p \in \mathcal{P},\ s \in \mathcal{S}, \tag{4.4.1}$$

[1]The *Iverson bracket* [33] returns 1 if the expression within the brackets is true; otherwise it returns 0.

where $\mathcal{S}$ is the set of 2-week segments. It follows that the length of the first step in the capacity–cost curve M^{e}_{i,p,r_1} is given by

$$M^{\mathrm{e}}_{i,p,r_1} = \frac{\sum_{s\in\mathcal{S}}[R_{ips}=1]w_{ips}}{\sum_{s\in\mathcal{S}}[R_{ips}=1]}, \qquad i\in\mathcal{I},\ p\in\mathcal{P}, \tag{4.4.2}$$

where r_1 is the first element in the set of cost classes $\mathcal{R}$. For $i\in\mathcal{I}$ and $p\in\mathcal{P}$, the lengths of the subsequent steps in the capacity–cost curve are calculated sequentially as

$$M^{\mathrm{e}}_{i,p,r_m} = \frac{\sum_{s\in\mathcal{S}}[R_{ips}=m]w_{ips}}{\sum_{s\in\mathcal{S}}[R_{ips}=m]} - \sum_{n=1}^{m-1} M^{\mathrm{e}}_{i,p,r_n}, \qquad m\in\{2,\dots,|\mathcal{R}|\}. \tag{4.4.3}$$

The length of the last step in the capacity–cost curve is set to be very large, i.e. three times the maximum annual load in the respective region. The height of each step in the capacity–cost curve, i.e. the weight of the investment, is given by the number of 2-week segments sharing the investment, as

$$\lambda^{\mathrm{e}}_{i,p,s,r_m} = \frac{1}{S-(m-1)}, \qquad m\in\{1,\dots,|\mathcal{R}|\}. \tag{4.4.4}$$

This cost is slightly modified in two ways: 1) the cost share is lower in the first iterations in order to enable the capacity with a high investment costs to stabilize before extinction, and 2) the cost share is lower for those 2-week segments that have not invested in the capacity that other 2-week segments have. This "rebate" is then reduced with the iteration number. Hence, for $i\in\mathcal{I}$, $p\in\mathcal{P}$, and $s\in\mathcal{S}$, it holds that

$$\lambda^{\mathrm{e}}_{i,p,s,r_m} = \frac{\alpha_{nips}}{S-\beta_n(m-1)}, \qquad m\in\{1,\dots,|\mathcal{R}|\},\ n\in\{1,\dots,10\}, \tag{4.4.5}$$

where the choices for the parameters α_{nips} and β_n in each iteration n are listed in Table 4.1. The parameter α_{nips} can take on a high ($\alpha^{\mathrm{high}}_{nips}$) or low ($\alpha^{\mathrm{low}}_{nips}$) value depending on whether or not investments have been made for the corresponding region, technology, and 2-week segment (i,p,s).

The construction of the capacity–cost curve is summarized in Algorithm 3.

Table 4.1: Parameter values used in the consensus loop

iteration number (n)	$\alpha^{\text{low}}_{nips}$	β_n	$\alpha^{\text{high}}_{nips}$
1	0.5	0.5	0.1
2	0.6	0.6	0.1
3	0.7	0.7	0.2
4	0.8	0.8	0.2
5	0.8	0.9	0.3
6	0.8	1.0	0.4
7	0.8	1.0	0.5
≥ 8	0.8	1.0	0.6

Algorithm 3 Creating the capacity–cost curve

1: Create a list L of the capacities such that $L := (w_{ip1}, w_{ip2}, \ldots, w_{ipS})$.
2: Sort the list L in ascending capacity size order. Each unique element represents a step in the capacity–cost curve.
3: The height of each step in the capacity–cost curve, i.e. $\lambda^{\text{e}}_{iprs}$, is determined by the number of 2-week segments sharing the investment. For each element, the number of 2-week segments sharing the investment corresponds to $\mathcal{S}$ reduced by the order of the element in the list L.
4: **if** $\exists$ duplicates in list L **then**
    remove duplicates from the list L
5: The length of the steps corresponds to capacity, such that each new step occurs at the values present in the reduced list $L*$. The potential of each cost class, M^{e}_{i,p,r_m}, is given by the capacity in the capacity–cost curve reduced by the capacity of the prior step.

Yearly linkages

In traditional electricity system investment models, the represented years are linked by the investment variables. However, the Paper IV model disregards any possible influence that future years might have on investments. This is based on the hypothesis that investments are made only to meet exactly the demand for electricity in the cost-optimal system, largely ignoring future needs in terms of capacity.

The cost of CO_2 emissions, investment costs (due to learning), efficiencies and discount rate can change between years and may influence the investment decisions. For scenarios with gradually increasing costs for generation capacity or operation over the years, this increase is likely to impact investments and needs to be transferred to prior years. Electricity generation technologies

that rely on fossil fuels are for example typically subject to a gradual increase in operational costs over the decades considered, which reduces the cost-competitiveness of these technologies in the long-term perspective. Under the assumption that the total cost for investments and operation of a power plant is evenly distributed across all of its hours of operation, some of the operational costs from later years need to be shifted to earlier years. The net present value of these future operational cost (with interest rate δ) is added to the objective function. Thus, for $p \in \mathcal{P}$ and $t \in \mathcal{T}_s$, $s \in \mathcal{S}$, we define the additional operational costs, C^{add}_{pty}, as

$$C^{\text{add}}_{pty} := \frac{1}{Z_p} \sum_{n=y}^{y+Z_p} \frac{1}{(1+\delta)^{(n-y)}} (C^{\text{run},n}_{pt} - C^{\text{run},y}_{pt}), \qquad y \in \mathcal{Y}, \tag{4.4.6}$$

where $y \in \mathcal{Y}$ is the year considered, i.e. the year in which investments are made, and Z_p is the lifetime of technology $p \in \mathcal{P}$. The costs (4.4.6) are added to the running cost C^{run}_{pt}, in the objective function (3.1.23) for the respective years.

5 A summary of the appended papers

This section summarizes the appended papers and highlights their different models and decomposition methods. Figure 5.1 visualizes their interrelations. Note that the papers are not numbered chronologically.

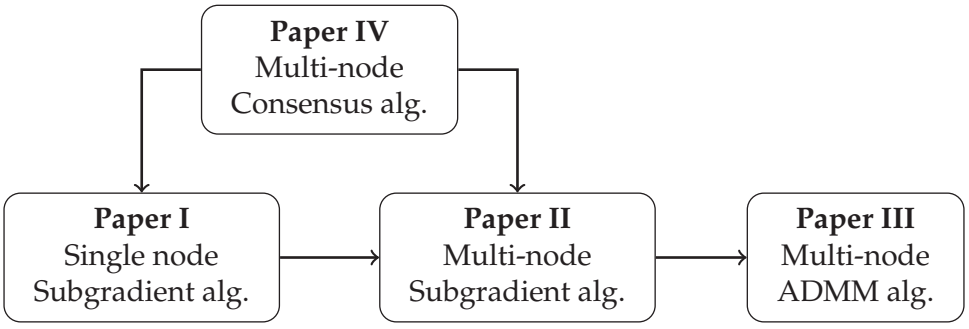


Figure 5.1: Illustration of the relation between the appended papers. Directions indicate that ideas/methods in the origin paper are generalized or applied in the destination paper.

Paper IV introduces the initial idea to decompose the temporal representation of an electricity system investment model into 2-week periods through a heuristic consensus algorithm. Paper I explores the same temporal decomposition through the use of Lagrangian duality methods combined with a subgradient solution algorithm, but on a model with a smaller spatial scope. Paper II presents an extension to the model in Paper I by introducing trade and storage, thereby making it similar to the model in Paper IV. Finally, Paper III considers the same model and decomposition as in Paper II, but utilizes an approximate consensus ADMM solution algorithm.

Paper I: A Lagrangian relaxation approach to an electricity system investment model with a high temporal resolution

In this paper, we formulate a long-term electricity system investment model that accounts for some variation management strategies to capture the variations from intermittent electricity production technologies. The model is decomposed using Lagrangian relaxation in combination with variable splitting. The decomposition results in 26 subproblems—each representing a time period of two weeks—which can be solved in parallel. The Lagrangian dual problem is solved by using a subgradient algorithm, which leads to 26 different subproblem solutions each iteration of the algorithm. Ergodic sequences are used to create a single solution in terms of production technology investments for each of the 26 subproblems, and these 26 investment solutions are then combined through a heuristic algorithm.

The decomposed model is implemented and solved using `Julia` [1] and `JuMP` [10] together with the `Gurobi` [23] solver. Three different cases, all involving single and isolated regions without the possibility for trade, are examined: 1) Hungary (HU), a region with poor conditions for wind power and medium conditions for solar PV, 2) Ireland (IE), a region with good wind power conditions, and 3) southern Sweden (SE3), a region with a large share of existing hydropower. The solutions found by the decomposed model provide capacity investments similar to the optimal solution for investments provided by the non-decomposed model. The heuristic used to combine the subproblem solutions can be further developed as it tends to overestimate the wind power capacity which increases the total system costs. The most important extension to this work is to include the possibility to trade with neighbouring regions, and also to include energy storage options in terms of hydrogen and batteries.

This paper has been resubmitted to *OR Spectrum* after it received minor revisions. The initial ideas were presented at the Swedish Operations Research Conference, Linköping (2017), and some later ideas on the Swedish Operations Research Conference, Nyköping (2019) and virtually at EUROPT 2021 Workshop on Advances in Continuous Optimization, Toulouse (2021).

Paper II: Managing the temporal resolution in a multi-node electricity system investment model: Parallel computations by variable splitting and Lagrangian relaxation

We formulate an extension to the long-term electricity system investment model in Paper I, where we in this paper consider a multi-node model with electricity trade and energy storage in terms of batteries and hydrogen storage. The model uses the same 2-week decomposition and subgradient algorithm as in Paper I and is implemented and solved using `Julia` [1] and `JuMP` [10] together with the `Gurobi` [23] solver. The subproblems are separable, which allows a parallel computing process. The parallel implementation furthermore has potential to reduce computation times and memory requirements compared to a non-decomposed model, although the extent of this reduction depends on the computer(s) and problem properties.

The method is tested on several cases (the British Isles, Iberian Peninsula and Sweden) with and without variation management options, with mixed results. The non-decomposed model for the multi-node cases involving electricity trade is unfortunately too large to be solvable on the used hardware. Thus, to evaluate the presented method, the results are compared with outcomes from the full-scale model but with either i) isolated regions, thus removing the option for trade but keeping the perfect foresight, or ii) disconnected years, which allows trade but has limited foresight. For most instances, our method provides capacity investments similar to the solutions provided by the other model implementations where it represents a reasonable "middle ground" in terms of capacity investments. However, due to the chosen costs of carbon-dioxide emissions, natural gas technology is often picked for new investments in favour of biomass plants. The computation times for the decomposed model challenges especially the model with disconnected years, where it performs better in all the tested instances. The computations for the model with isolated regions can on the other hand also be performed in parallel, which hence provides it with substantially quicker computation times than the other options.

It should be noted that the decomposed model fails to provide reasonable results for seasonal storage, which is to store energy during summer or winter, and discharge it during the other respective season. In this paper, this corresponds to hydrogen storage and hydropower reservoirs. The most important addition to this work however would be to compare this subgradient solution methodology to the popular ADMM-approach.

This paper is a complete manuscript and will shortly be submitted to a journal. The main ideas have been presented at the Stockholm Optimization Days 2022, Stockholm (2022).

Paper III: An approximate consensus ADMM approach to a multi-node electricity system investment problem with a high temporal resolution

We consider in this paper the same multi-node model with electricity trade and energy storage as in Paper II, but solve it using an approximate consensus ADMM-approach. The temporal resolution is decomposed into 2-week periods throughout the year, with any overlapping constraints Lagrangian relaxed. The algorithm is implemented in `Julia` [1] and `JuMP` [10] and solved using the `Gurobi` [23] solver.

As in Paper II, the method is evaluated on the British Isles, Iberian Peninsula and Sweden, with different combinations of variation management techniques. The method's efficiency is assessed by comparing it to the results from the subgradient method developed in Papers I–II. The solutions from the subgradient method and ADMM algorithm share the same production pattern in terms of base, peak and intermediate power. The total system costs are consistently a few percent higher for the ADMM algorithm compared to the subgradient method in the trade cases. For a single region with different energy storage options, ADMM occasionally outperforms the subgradient method in terms of costs, but both methods are worse than the optimum. The computation times in both methods are better if compared to a corresponding non-decomposed multi-node model, but especially the solution times are exceptionally better for the ADMM algorithm. However, this is primarily the result of the ADMM algorithm using a 6h time step (due to otherwise reaching the memory limit), compared to the 3h time step used in the subgradient algorithm. Moreover, the time it takes to build the model and perform the algorithm outside the Gurobi solver is a big drawback of the method, as the result is a drastically increased summarized computation time which extends beyond the computation time for the subgradient method. The main advantage however of the ADMM algorithm compared to the subgradient method is an easier implementation, especially since it circumvents 1) the requirement to recover primal feasible solutions and the need to use ergodic sequences, and 2) the need for an upper bound on the system cost to calculate the step length. Both methods however fail to dimension the need for seasonal storage, i.e. hydrogen storage and hydropower reservoirs. Thus, the 2-week decomposition might not be the best

approach if energy storage over longer time-scales than two weeks is of vital importance to the application.

The quality of the work in this paper can initially be raised by revisiting implementation of the model and ADMM algorithm, which potentially can shorten the computation times. The most important extension however would be to further examine if convergence conditions can be expressed for our approximate consensus ADMM algorithm.

This paper have neither been presented nor submitted to a journal. Submission is expected early autumn.

Paper IV: Management of wind power variations in electricity system investment models: A parallel computing strategy

In this paper, we develop a mathematical model and a heuristic method which account for variation management strategies in a long-term electricity system investment model. The Hours-to-Decades model discretizes the time dimension into 2-week segments and solves the resulting 26 separate problems in parallel. Information between the segments is then exchanged in a consensus loop, and the main idea is that the investments from the different solutions form the basis for the investment costs in the subsequent solve. This process is then iterated until consensus for the investments made by all 26 problems is reached.

The model is implemented in `GAMS` and then solved using `CPLEX`. Different cases are considered, with some variation management options including energy storage and trade. The different regions considered are: (1) Ireland, which is a region with good conditions for wind power, (2) Ireland and UK, for the case when trade for Ireland is considered, (3) central Spain, a region with good conditions for solar power, and (4) Iberian Peninsula for central Spain trade. The solutions found by the Hours-to-Decades model possess an increased total system cost of approximately 1 % compared to the same electricity investment model but with connected time. The resulting energy mix shows that the Hours-to-Decades model responds to variation management similar to the connected-time model. When it comes to the computation times, the Hours-to-Decades model are able to solve the problem faster than a time-connected model. This is even more prominent when several regions with trade are included. As an example, when two regions are considered the

Hours-to-Decades model takes around half an hour to solve while the run time for the time-connected model is several days. A drawback of the method is, however, that it cannot dimension seasonal storages. The results of the cases studied indicates, however, that seasonal storage capacity can be dimensioned post-process.

Our heuristic method targets the combination of wind variation management and trade in electricity system models. Nevertheless, if wind power or trade are not of relevance for the investigated regions, representative days or integral time slicing are likely more efficient modelling methodologies.

This paper was initially presented by Lisa Göransson at the International Energy Workshop, Paris (2019), and later published in *SN Operations Research Forum, **2**:25* (see [20]).

6 Conclusions and ideas for future research

To summarize, we have developed two electricity system models that can be used as a tool to analyze long-term investments in an electricity system that contains a large share of variable renewable electricity generation. We have further suggested three decomposition methods to decrease computation times for all the models presented in this thesis. The first method, which is the one used in the Hours-to-Decades model, can be seen as a heuristic approach to variable splitting. This approach provided us the initial idea to develop the second method which uses Lagrangian relaxation combined with variable splitting, and a solving process based on a subgradient algorithm. This later led to our third method, in which we utilize the same decomposition approach with Lagrangian relaxation and variable splitting, but solve the model using an approximate consensus ADMM algorithm. These contributions target the stated objectives to (i) formulate mathematical long-term investments models that capture strategies to manage the variability of variable renewable electricity generation, (ii) develop and evaluate decomposition methods for said models, and (iii) examine the impact on the solution times for different spatial resolutions.

The main conclusion is that our different methods produce numerical results with faster computation times compared to the non-decomposed models, and capacity investment options similar to the solutions provided by the latter models. They are furthermore able to do so while allowing a parallel solution process, thus potentially reducing the necessary computer memory. However, the suggested methods are sensitive to parameter settings, and the fine-tuning of algorithm parameters to a specific problem instance likely requires some trial and error. Moreover, although the solutions are acceptable after only a few iterations, the convergence speed for the subgradient method is typically slow, requiring many iterations before satisfying traditional termination criteria. For

the approximate consensus ADMM, convergence to the optimum is, to the best of our current knowledge, not even guaranteed. Hence, the potential reduction in memory requirement and computation time may not be sufficient to motivate the increase in complexity and uncertainty of the decomposed models.

6.1 Additional reflections

Among other things, we have concluded that the most important part of the original model structure is kept when using a decomposition over 2-week periods. Indeed, in each subproblem we basically solve the entire electricity system model, but for a smaller time interval. Hence, for both the decomposed full-scale model and the Hours-to-Decades model, the subproblems have a structure which is very similar to their respective non-decomposed model. For each of the two models, this furthermore leads to subproblem solutions which are very similar to each other. Therefore, the optimal solution provided by each subproblem is decent in terms of a solution to the non-decomposed problem. As a comparison, consider a decomposition over groups of variable types instead. For example, electricity generation and investments could be one group, and hot capacity and start-up capacity another group. Here, we assume that we relax constraints that connect different groups which then makes the problem separable into a few subproblems. The most important model structures would immediately be lost, and most likely a massive number of iterations for the algorithms (i.e. subgradient method and ADMM) would be required before the method converges.

As earlier concluded, the decomposition methods developed in this thesis make the models parallelizable with the potential to reduce computation times. A nice feature of the chosen 2-week decomposition is that since the subproblems have the same model structure and several parameters with the same value (e.g. costs), they also take approximately the same amount of time to solve. There is therefore not much idle time in the solution algorithms. As a counterexample, consider the case that a single subproblem had taken much longer time to solve compared to the other subproblems. Consequently, each iteration of the algorithm would include some "waiting time" before the dual variable updates could be done. This is, however, not the case if all subproblems finish solving their model simultaneously.

Another reflection is that if we would model the emission limits as a hard constraint, it would need to be relaxed for the optimization problem to become separable. It would hence be penalized in the objective using different dual

variables for each subproblem, with the values of the dual variables being updated in the subgradient iterations. Thus, providing from the start a reasonable cost for the emissions in the original problem will in theory 1) make the Lagrangian problem separable with respect to emissions, since these are not constrained, 2) lower the complexity of the model, since fewer dual variables are required, and 3) provide better subproblem solutions, since emissions are penalized with the same costs in all of the subproblems.

The results in Papers II–IV showed us that the decomposition methods based on a 2-week time representation is not able to correctly estimate seasonal storage. The subproblems provided by this decomposition struggle to capture model structure and properties that stretch beyond their own time period. Thus, the conclusion is that our 2-week decomposition is likely not a recommended approach if properly dimensioned seasonal storage is of high importance to the user.

6.2 Future perspective

There exists several ideas for possible research directions from here. The model itself can be extended to include additional variation management strategies, such as electric vehicles. Other ideas involve further exploration of the applied solution methodologies. For instance, the convergence to optimal solutions in the limit for ergodic sequences used together with the variable splitting approach needs to be examined. Furthermore, the heuristic used to combine the subproblem solutions in Papers I and II can be further developed. In addition, possible convergence conditions for the approximate consensus ADMM used in Paper III have not been investigated.

Other methods to decompose the problem could moreover be investigated. For instance, a column generation approach would yield the same subproblems as given by our current decomposition, but with the added benefit of finite theoretical primal convergence. Tonbari and Ahmed [59] presents a fully distributed Dantzig-Wolfe decomposition algorithm, where the master problem is solved using a consensus ADMM method. If viewed from the perspective of two-stage stochastic optimization, the approach resembles the use of progressive hedging algorithms [52].

Lastly, another possible direction to pursue involves a spatial decomposition (i.e. over regions) compared to the temporal decomposition (i.e. 2-week periods) done in this thesis, as indicated by the results in Paper II.

Bibliography

[1] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98.

[2] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):122.

[3] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

[4] Breyer, C., Bogdanov, D., Ram, M., Khalili, S., Vartiainen, E., Moser, D., Román Medina, E., Masson, G., Aghahosseini, A., Mensah, T. N. O., Lopez, G., Schmela, M., Rossi, R., Hemetsberger, W., and Jäger-Waldau, A. (2022). Reflecting the energy transition from a European perspective and in the global context—Relevance of solar photovoltaics benchmarking two ambitious scenarios. *Progress in Photovoltaics: Research and Applications*.

[5] Caprara, A., Fischetti, M., and Toth, P. (1999). A heuristic method for the set covering problem. *Operations Research*, 47(5):730–743.

[6] Chang, T.-H., Hong, M., and Wang, X. (2014). Multi-agent distributed large-scale optimization by inexact consensus alternating direction method of multipliers. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6137–6141.

[7] Chen, C., He, B., Ye, Y., and Yuan, X. (2016). The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*, 155(1):57–79.

[8] Collins, S., Deane, J. P., Poncelet, K., Panos, E., Pietzcker, R. C., Delarue, E., and Ó Gallachóir, B. P. (2017). Integrating short term variations of the power system into integrated energy system models: A methodological review. *Renewable and Sustainable Energy Reviews*, 76:839–856.

[9] COM(2011) 885 (2011). Energy roadmap 2050. European Commission.

[10] Dunning, I., Huchette, J., and Lubin, M. (2017). JuMP: A modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320.

[11] Frew, B. A., Becker, S., Dvorak, M. J., Andresen, G. B., and Jacobson, M. Z. (2016). Flexibility mechanisms and pathways to a highly renewable us electricity future. *Energy*, 101:65–78.

[12] Frew, B. A. and Jacobson, M. Z. (2016). Temporal and spatial tradeoffs in power system modeling with assumptions about storage: An application of the POWER model. *Energy*, 117:198–213.

[13] Frysztacki, M. M., Hörsch, J., Hagenmeyer, V., and Brown, T. (2021). The strong effect of network resolution on electricity system models with high shares of wind and solar. *Applied Energy*, 291.

[14] Frysztacki, M. M., Recht, G., and Brown, T. (2022). A comparison of clustering methods for the spatial reduction of renewable electricity optimisation models of Europe. *Energy Informatics*, 5(1).

[15] Gerbaulet, C. and Lorenz, C. (2017). dynELMOD: A dynamic investment and dispatch model for the future European electricity market. Technical report, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin.

[16] Gils, H.-C. (2016). Energy system model REMix. Report, Deutsches Zentrum für Luft- und Raumfahrt, DLR.

[17] Glover, F. and Klingman, D. (1988). Layering strategies for creating exploitable structure in linear and integer programs. *Mathematical Programming*, 40(1):165–181.

[18] Göransson, L. (2014). *The impact of wind power variability on the least-cost dispatch of units in the electricity generation system*. PhD thesis, Chalmers University of Technology.

[19] Göransson, L., Goop, J., Odenberger, M., and Johnsson, F. (2017). Impact of thermal plant cycling on the cost-optimal composition of a regional electricity generation system. *Applied Energy*, 197:230–240.

[20] Göransson, L., Granfeldt, C., and Strömberg, A.-B. (2021). Management of wind power variations in electricity system investment models. *Operations Research Forum*, 2(2):25.

[21] Guignard, M. (2003). Lagrangean relaxation. *Top*, 11(2):151–200.

[22] Guignard, M. and Kim, S. (1987). Lagrangean decomposition: A model yielding stronger Lagrangean bounds. *Mathematical Programming*, 39(2):215–228.

[23] Gurobi Optimization, LLC (2021). Gurobi Optimizer Reference Manual.

[24] Gustavsson, E., Patriksson, M., and Strömberg, A.-B. (2015). Primal convergence from dual subgradient methods for convex optimization. *Mathematical Programming*, 150(2):365–390.

[25] Held, M. and Karp, R. M. (1971). The traveling-salesman problem and minimum spanning trees: Part II. *Mathematical Programming*, 1(1):6–25.

[26] Hestenes, M. R. (1969). Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4(5):303–320.

[27] Hoffmann, M., Kotzur, L., Stolten, D., and Robinius, M. (2020). A review on time series aggregation methods for energy system models. *Energies*, 13(3).

[28] Holttinen, H., Meibom, P., Orths, A., Van Hulle, F., Lange, B., O'Malley, M., Pierik, J., Ummels, B., Tande, J. O. G., Estanqueiro, A. I., Matos, M. A., Gomez-Lazaro, E., Söder, L., Strbac, G., Shakoor, A., Ricardo, J., Smith, J. C., Milligan, M., and Ela, E. G. (2009). *Design and operation of power systems with large amounts of wind power*. Julkaisija-Utgivare.

[29] Hong, M. and Luo, Z.-Q. (2017). On the linear convergence of the alternating direction method of multipliers. *Mathematical Programming*, 162(1):165–199.

[30] HYBRIT (2021). Hydrogen breakthrough ironmaking technology. `https://www.hybritdevelopment.se/`. Accessed: 2023-05-31.

[31] IEA (2020). Total energy supply (TES) by source, World 1990–2018. International Energy Agency.

[32] IEA Wind TCP Task 25 (2018). Wind and PV integration studies. Report, International Energy Agency.

[33] Iverson, K. E. (1962). *A Programming Language*. Wiley.

[34] Jörnsten, K. and Näsberg, M. (1986). A new Lagrangian relaxation approach to the generalized assignment problem. *European Journal of Operational Research*, 27(3):313–323.

[35] Klemm, C., Wiese, F., and Vennemann, P. (2023). Model-based run-time and memory reduction for a mixed-use multi-energy system model with high spatial resolution. *Applied Energy*, 334:120574.

[36] Lara, C. L., Mallapragada, D. S., Papageorgiou, D. J., Venkatesh, A., and Grossmann, I. E. (2018). Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. *European Journal of Operational Research*, 271(3):1037–1054.

[37] Larsson, T., Patriksson, M., and Strömberg, A.-B. (1999). Ergodic, primal convergence in dual subgradient schemes for convex programming. *Mathematical Programming*, 86(2):283–312.

[38] Larsson, T., Patriksson, M., and Strömberg, A.-B. (1996). Conditional subgradient optimization—Theory and applications. *European Journal of Operational Research*, 88(2):382–403.

[39] Lehtveer, M., Mattsson, N., and Hedenus, F. (2017). Using resource based slicing to capture the intermittency of variable renewables in energy system models. *Energy Strategy Reviews 18*.

[40] Mai, T., Drury, E., Eurek, K., Bodington, N., Lopez, A., and Perry, A. (2013). Resource planning model: An integrated resource planning and dispatch tool for regional electric systems. Report, National Renewable Energy Laboratory (NREL), Golden, CO, United States.

[41] Mone, C., Hand, M., Bolinger, M., Rand, J., Heimiller, D., and Ho, J. (2017). 2015 cost of wind energy review. Report NREL/TP-6A20-66861, National Renewable Energy Laboratory (NREL).

[42] Nahmmacher, P., Schmid, E., Hirth, L., and Knopf, B. (2016). Carpe diem: A novel approach to select representative days for long-term power system modeling. *Energy*, 112:430–442.

[43] Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239.

[44] Pfenninger, S., Hawkes, A., and Keirstead, J. (2014). Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33:74 – 86.

[45] Polyak, B.T. (1969). Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*, 9(3):14–29.

[46] Powell, M. J. D. (1969). A method for nonlinear constraints in minimization problems. (R. Fletcher, ed.), Academic Press.

[47] Reichenberg, L., Siddiqui, A. S., and Wogrin, S. (2018). Policy implications of downscaling the time dimension in power system planning models to represent variability in renewable output. *Energy*, 159:870–877.

[48] Ringkjøb, H.-C., Haugan, P. M., and Solbrekke, I. M. (2018). A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews*, 96:440–459.

[49] Ritchie, H. and Roser, M. (2020). CO_2 and greenhouse gas emissions. *Our World in Data*. `https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions`.

[50] Rockafellar, R. T. (1973a). A dual approach to solving nonlinear programming problems by unconstrained optimization. *Mathematical Programming*, 5(1):354–373.

[51] Rockafellar, R. T. (1973b). The multiplier method of Hestenes and Powell applied to convex programming. *Journal of Optimization Theory and Applications*, 12(6):555–562.

[52] Rockafellar, R. T. and Wets, R. J.-B. (1991). Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16(1):119–147.

[53] Sagastizábal, C. (2012). Divide to conquer: decomposition methods for energy optimization. *Mathematical Programming*, 134(1):187–222.

[54] Shi, W., Ling, Q., Yuan, K., Wu, G., and Yin, W. (2013). Linearly convergent decentralized consensus optimization with the alternating direction method of multipliers. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4613–4617.

[55] Shi, W., Ling, Q., Yuan, K., Wu, G., and Yin, W. (2014). On the linear convergence of the ADMM in decentralized consensus optimization. *IEEE Transactions on Signal Processing*, 62(7):1750–1761.

[56] Shor, N. Z. (1991). The development of numerical methods for nonsmooth optimization in the USSR. In Lenstra, J. K., Rinnoy Kan, A. H. G., and Schrijver, A., editors, *History of Mathematical Programming: A Collection of Personal Reminiscences*, pages 135–139. North-Holland, Amsterdam, The Netherlands.

[57] Sun, R., Luo, Z.-Q., and Ye, Y. (2019). On the efficiency of random permutation for ADMM and coordinate descent. *Mathematics of Operations Research*, 45:233–271.

[58] Teichgraeber, H. and Brandt, A. R. (2022). Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities. *Renewable and Sustainable Energy Reviews*, 157:111984.

[59] Tonbari, M. E. and Ahmed, S. (2023). Consensus-based Dantzig-Wolfe decomposition. *European Journal of Operational Research*, 307(3):1441–1456.

[60] van Ackooij, W., Chorobura, A. P., Sagastizábal, C., and Zidani, H. (2021). Demand response versus storage flexibility in energy: multi-objective programming considerations. *Optimization*, 70(7):1459–1486.

[61] Weber, C. (2005). *Uncertainty in the Electric Power Industry: Methods and Models for Decision Support*. Springer, New York.

[62] Wogrin, S., Duenas, P., Delgadillo, A., and Reneses, J. (2014). A new approach to model load levels in electric power systems with high renewable penetration. *IEEE Transaction on Power Systems*, 29(5):2210–2218.

A Nomenclature

A.1 Full-scale model

Table A.1: The index sets used in the full-scale model

symbol		representation	member
$\mathcal{R}$		regions	r
$\mathcal{A}$	$\subseteq \mathcal{R} \times \mathcal{R}$;	transmission lines between regions	q, r
$\mathcal{K}$		technologies for transmission	k
$\mathcal{P}$	$:= \mathcal{P}_{\text{thermal}} \cup \mathcal{P}_{\text{ren}} \cup \mathcal{P}_{\text{e-lysis}}$;	electricity generation/consumption technologies	p
$\mathcal{P}_{\text{thermal}}$		thermal power technologies	p
$\mathcal{P}_{\text{ren}}$	$:= \mathcal{P}_{\text{wind}} \cup \mathcal{P}_{\text{solar}} \cup \mathcal{P}_{\text{hydro}}$;	renewable technologies	p
$\mathcal{P}_{\text{wind}}$		wind technologies	p
$\mathcal{P}_{\text{solar}}$		solar technologies	p
$\mathcal{P}_{\text{hydro}}$		hydropower technologies	p
$\mathcal{P}_{\text{e-lysis}}$		electrolyser technologies	h
$\mathcal{L}$	$:= \mathcal{L}_{\text{bat}} \cup \mathcal{L}_{\text{H}_2}$;	electricity storage technologies	ℓ
$\mathcal{L}_{\text{bat}}$		battery technologies	ℓ
$\mathcal{L}_{\text{H}_2}$		hydrogen storage technologies	ℓ
$\mathcal{I}$	$:= \{1960, 1970, \ldots, 2050\}$;	investment years, defining investment periods	i
$\mathcal{S}$	$:= \{2020, 2030, \ldots, 2050\}$;	new capacity investment years; $\mathcal{S} \subset \mathcal{I}$	s
$\mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$	$:= \mathcal{I} \cap \{s - U_p, \ldots, s\}$;	investment periods for each technology type $p \in \mathcal{P}$ with lifespan U_p that is active at year $s \in \mathcal{S}$	i
$\mathcal{I}^{\mathcal{K}}_{\text{active}}(s)$	$:= \mathcal{I} \cap \{1960, \ldots, s\}$;	investment periods for transmission technologies that are active at year $s \in \mathcal{S}$	i
$\mathcal{T}$	$:= \{\tau, 2\tau, \ldots, T\}$;	time steps within a year, where τ denotes the step length	t

$\mathcal{T}_{\text{start}}(p)$	$\subset \mathcal{T} \cup \{0\}$;	consecutive time steps in the start-up interval for technology $p \in \mathcal{P}$	t

Table A.2: The decision variables used in the full-scale model

symbol	restriction	explanation	unit
x_{prist_σ}	≥ 0	generated electricity of technology type $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ using technology from investment period $i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$ in year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$	GWh/h
y_{pri}	≥ 0	investments in production capacity (both new and old) for technology $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ during investment period $i \in \mathcal{I}$	GW
y^{split}_{prin}	≥ 0	investments in production capacity (both new and old) for technology $p \in \mathcal{P}$ during investment period $i \in \mathcal{I}$ for subproblem $n \in \mathcal{N}$	GW
$y_{\ell ri}$	≥ 0	investments in battery and hydrogen capacity for technology $\ell \in \mathcal{L}$ in region $r \in \mathcal{R}$ during investment period $i \in \mathcal{I}$	GWh
$y^{\text{split}}_{\ell rin}$	≥ 0	investments in battery and hydrogen capacity for technology $\ell \in \mathcal{L}$ in region $r \in \mathcal{R}$ during investment period $i \in \mathcal{I}$ for subproblem $n \in \mathcal{N}$	GWh
v_{kqrst_σ}	≥ 0	electricity traded with transmission type $k \in \mathcal{K}$ from region q to region r, $(q,r) \in \mathcal{A}$, in year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$	GWh/h
u_{kqri}	≥ 0	investments in new transmission capacity for transmission type $k \in \mathcal{K}$ between regions q and r, where $(q,r) \in \mathcal{A}$, during investment period $i \in \mathcal{I}$	GW
u^{split}_{kqrin}	≥ 0	investments in new transmission capacity for transmission type $k \in \mathcal{K}$ between regions q and r, where $(q,r) \in \mathcal{A}$, during investment period $i \in \mathcal{I}$ for subproblem $n \in \mathcal{N}$	GW
$b^{\text{charge}}_{\ell rst_\sigma}$	≥ 0	charging of battery type $\ell \in \mathcal{L}_{\text{bat}}$ in region $r \in \mathcal{R}$ during year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$	GWh/h
$b^{\text{discharge}}_{\ell rst_\sigma}$	≥ 0	discharging of battery type $\ell \in \mathcal{L}_{\text{bat}}$ in region $r \in \mathcal{R}$ during year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$	GWh/h
$b^{\text{storage}}_{\ell rst_\sigma}$	≥ 0	battery storage level in battery type $\ell \in \mathcal{L}_{\text{bat}}$ in region $r \in \mathcal{R}$ during year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$	GWh
$h^{\text{consumption}}_{rst_\sigma}$	≥ 0	electricity consumption in the electrolyser used for hydrogen production in region $r \in \mathcal{R}$ during year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$	GWh/h
$h^{\text{storage}}_{\ell rst_\sigma}$	≥ 0	hydrogen storage level in technology $\ell \in \mathcal{P}_{\text{hydro}}$ in region $r \in \mathcal{R}$ during year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$	GWh
w_{rst_σ}	≥ 0	stored hydropower in region $r \in \mathcal{R}$, year $s \in \mathcal{S}$ at time step $t_\sigma \in \mathcal{T}$	GWh

z_{prist_σ}	≥ 0	hot capacity of technology type $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ at investment period $i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$, year $s \in \mathcal{S}$ at time $t_\sigma \in \mathcal{T}$	GWh/h
$z^{+}_{prist_\sigma}$	≥ 0	increase in hot capacity from time step $t_{\sigma-1} \in \mathcal{T}$ to $t_\sigma \in \mathcal{T}$ for technology type $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ at investment period $i \in \mathcal{I}^{\mathcal{P}}_{\text{active}}(s,p)$, year $s \in \mathcal{S}$	GWh/h

Table A.3: The auxiliary variables used in the full-scale model

symbol	restriction	explanation	unit
$v^{\text{net}}_{rst_\sigma}$		net electricity import to region $r \in \mathcal{R}$ in year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$ (if negative then export)	GWh/h
$b^{\text{net}}_{rst_\sigma}$		net charge of battery type $\ell \in \mathcal{L}_{\text{bat}}$ in region $r \in \mathcal{R}$ during year $s \in \mathcal{S}$ and time step $t_\sigma \in \mathcal{T}$ (if negative, then net discharge)	GWh/h
$e^{\text{tot}}_{st_\sigma}$	≥ 0	auxiliary definition variable for the total system emissions at year $s \in \mathcal{S}$ in time step $t_\sigma \in \mathcal{T}$	tonnes CO_2

Table A.4: The parameters used in the full-scale model

symbol	representation	unit
a^{gen}_{pri}	Existing electricity generation capacity of technology $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ in investment period $i \in \mathcal{I} \setminus \mathcal{S}$	GW
$a^{\text{sto}}_{\ell ri}$	Existing electricity storage capacity of technology $\ell \in \mathcal{L}$ in region $r \in \mathcal{R}$ in investment period $i \in \mathcal{I} \setminus \mathcal{S}$	GW
a^{tra}_{kqri}	Existing transmission capacity of technology $k \in \mathcal{K}$ on transmission line $(q,r) \in \mathcal{A}$ in investment period $i \in \mathcal{I} \setminus \mathcal{S}$	GW
c^{invtech}_{ps}	Investment cost for technology type $p \in \mathcal{P}$ during year $s \in \mathcal{S}$. Includes an annuity factor	k€/GW
c^{omf}_{p}	Fixed operation and maintenance costs for technology type $p \in \mathcal{P}$	k€/GW
c^{run}_{pri}	Run cost for technology type $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ using technology from investment period $i \in \mathcal{I}$	k€/GWh
$c^{\text{invsto}}_{\ell s}$	Investment cost for technology type $\ell \in \mathcal{L}$ in year $s \in \mathcal{S}$. Includes an annuity factor	k€/GWh
c^{omf}_{ℓ}	Fixed operation and maintenance costs for technology type $\ell \in \mathcal{L}$	k€/GWh
c^{invtra}_{kqrs}	Investment cost of new transmission capacity of technology $k \in \mathcal{K}$ on transmission line $(q,r) \in \mathcal{A}$ during year $s \in \mathcal{S}$. This cost is divided by two to compensate for double arcs since the network is expressed only with non-negative arcs. Includes an annuity factor	k€/GW

c^{tra}_{kqr}	Transmission cost of transmission technology $k \in \mathcal{K}$ on transmission line $(q, r) \in \mathcal{A}$	k€/GWh
c^{+}_{prs}	Upstart cost for technology type $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ in year $s \in \mathcal{S}$	k€/GW
$\tilde{c}_{prs}$	Part-load cost for technology type $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ using technology at year $s \in \mathcal{S}$	k€/GWh
$c^{\text{CO}_2}_{s}$	The costs for emissions at year $s \in \mathcal{S}$	k€/tonnes CO_2
d_{rst}	Electricity demand in region $r \in \mathcal{R}$ at year $s \in \mathcal{S}$ and time step $t \in \mathcal{T}$	GWh/h
d^{hydrogen}_{rs}	Hydrogen demand from industry in $r \in \mathcal{R}$, year $s \in \mathcal{S}$	GWh/h
e_{pri}	Emissions per produced GWh of technology $p \in \mathcal{P}$ in region $r \in \mathcal{R}$ using technology from investment period $i \in \mathcal{I}$	tonnes CO_2/GWh
e^{+}_{pri}	Upstart emissions for technology type $p \in \mathcal{P}_{\text{thermal}}$ in region $r \in \mathcal{R}$ using technology from investment period $i \in \mathcal{I}$	tonnes CO_2/GW
$\tilde{e}_{pri}$	Extra emissions when running on part-load for technology type $p \in \mathcal{P}_{\text{thermal}}$ in region $r \in \mathcal{R}$ using technology from investment period $i \in \mathcal{I}$	tonnes CO_2/GWh
g_{rt}	Inflow into hydro power from rain, ground etc. in region $r \in \mathcal{R}$ during time step $t \in \mathcal{T}$	GWh
H_r	Upper limit for hydropower storage in region $r \in \mathcal{R}$	GWh
$u^{\max}_{kqr}$	Maximum transmission capacity of technology $k \in \mathcal{K}$ that is possible on transmission line $(q, r) \in \mathcal{A}$	GW
U_p	The lifespan of technology type $p \in \mathcal{P}$	years
W_{pr}	Maximum capacity of wind, i.e. land availability, for wind technology $p \in \mathcal{P}_{\text{wind}}$ in region $r \in \mathcal{R}$	GW
δ^{inc}_{r}	The maximum ramping rate for water level increase in hydropower in region $r \in \mathcal{R}$	share
δ^{dec}_{r}	The maximum ramping rate for water level decrease in hydropower in region $r \in \mathcal{R}$	share
$\delta^{\text{inj}}_{\ell}$	injection rate of storage technology $\ell \in \mathcal{L}$	1/h
$\delta^{\text{with}}_{\ell}$	withdrawal rate of storage technology $\ell \in \mathcal{L}$	1/h
$\eta^{\text{charge}}_{\ell s}$	efficiency of charging storage technology $\ell \in \mathcal{L}$ in year $s \in \mathcal{S}$	share
$\eta^{\text{discharge}}_{\ell s}$	efficiency of discharging storage technology $\ell \in \mathcal{L}$ in year $s \in \mathcal{S}$	share
η_{ps}	efficiency of electrolyser $p \in \mathcal{P}_{\text{e-lysis}}$ in year $s \in \mathcal{S}$	share
θ_{prt}	Weather profile for renewable technologies $p \in \mathcal{P}_{\text{ren}}$ in region $r \in \mathcal{R}$ at time step $t \in \mathcal{T}$	share
τ	time step length	h
ϕ_p	Minimum load level for technology $p \in \mathcal{P}_{\text{thermal}}$	share
I	The total number of investment periods in the model	
S	The total number of years where it is possible to make new investments in capacity	

T	The number of total time steps in the model

A.2 Hours-to-Decades model

Table A.5: The index sets used in the Hours-to-Decades model

symbol	representation	member
$\mathcal{I}$	set of all regions	i, j
$\mathcal{P}$	$:= \mathcal{P}^{\text{bat}} \cup \mathcal{P}^{\text{electrolysis}} \cup \mathcal{P}^{\text{hydrogen}} \cup \mathcal{P}^{\text{gen}}$; set of all technology aggregates	p
$\mathcal{P}^{\text{bat}}$	set of all battery technologies	p
$\mathcal{P}^{\text{electrolysis}}$	set of all electrolyzer technologies	p
$\mathcal{P}^{\text{hydrogen}}$	set of all hydrogen storage technologies	p
$\mathcal{P}^{\text{gen}}$	$:= \mathcal{P}^{\text{wind}} \cup \mathcal{P}^{\text{therm}} \cup \mathcal{P}^{\text{solar}}$; set of all electricity generation technologies	p
$\mathcal{P}^{\text{wind}}$	set of all wind technologies	p
$\mathcal{P}^{\text{therm}}$	set of all thermal technologies	p
$\mathcal{P}^{\text{solar}}$	set of all solar technologies	p
$\mathcal{Q}$	set of technologies for transmission	q
$\mathcal{S}$	$:= \{1, \ldots, S\}$; set of all 2-week segments (typically, $S = 26$)	s
$\mathcal{T}_s$	$:= \{(s-1)T + 1, \ldots, sT\}$; set of all time steps in the 2-week segment $s \in \mathcal{S}$	t
$\mathcal{K}_p$	$:= \{0, \ldots\}$; set of hours in the start-up interval for technology $p \in \mathcal{P}_{\text{thermal}}$	k
$\mathcal{R}$	set of cost classes, i.e., the steps in the cost–supply curve	r
$\mathcal{Y}$	set of years	y

Table A.6: The variables used in the Hours-to-Decades model

symbol	restriction	explanation	unit
w_{ipr}	≥ 0	investment in region $i \in \mathcal{I}$ in generation technology $p \in \mathcal{P}^{\text{gen}}$ in cost class $r \in \mathcal{R}$	GW
w_{ipr}	≥ 0	investment in storage capacity in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{bat}} \cup \mathcal{P}^{\text{hydrogen}}$ in cost class $r \in \mathcal{R}$	GWh
h_{ijqr}	≥ 0	investment in transmission capacity between regions $i, j \in \mathcal{I}$ using transmission technology $q \in \mathcal{Q}$ in cost class $r \in \mathcal{R}$	GW
g_{ipt}	≥ 0	electricity generation in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{gen}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GWh/h
g_{ipt}	≥ 0	battery storage in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{bat}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GWh
g_{ipt}	≥ 0	hydrogen storage in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{hydrogen}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GWh
e_{ijt}		electricity export from region $i \in \mathcal{I}$ to region $j \in \mathcal{I}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$ ($e_{ijt} < 0$ represents import to i from j)	GWh/h
e^{pos}_{ijt}	≥ 0	absolute value of electricity export from region $i \in \mathcal{I}$ to region $j \in \mathcal{I}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GWh/h
c^{cycl}_{ipt}	≥ 0	resulting thermal cycling costs in region $i \in \mathcal{I}$ for technology $p \in \mathcal{P}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	k€/h
b^{charge}_{ipt}	≥ 0	battery charging in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{bat}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GWh/h
$b^{\text{discharge}}_{ipt}$	≥ 0	battery discharging in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{bat}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GWh/h
g^{active}_{ipt}	≥ 0	activated thermal capacity in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{therm}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GW
g^{on}_{ipt}	≥ 0	started thermal capacity in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{therm}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GW
d^{hydrogen}_{it}	≥ 0	electricity consumption in the electrolyzer in region $i \in \mathcal{I}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GWh/h

Table A.7: The parameters used in the Hours-to-Decades model

symbol	representation	unit
S	number of 2-week segments	1
T	number of time steps in each 2-week segment	1
C_p^{inv}	investment cost of technology $p \in \mathcal{P}^{\text{gen}}$	k€/GW
C_p^{inv}	investment cost of storage capacity for technology $p \in \mathcal{P}^{\text{bat}} \cup \mathcal{P}^{\text{hydrogen}}$	k€/GWh
$C_{q,i,j}^{\text{h-inv}}$	investment cost of transmission technology $q \in \mathcal{Q}$ between regions $i, j \in \mathcal{I}$	k€/GW
$\lambda_{ipsr}^{\text{e}}$	share of the investment cost for technology $p \in \mathcal{P}$ in region $i \in \mathcal{I}$ taken by cost class $r \in \mathcal{R}$ and segment $s \in \mathcal{S}$	1
α_{nips}, β_n	parameters used to compute $\lambda_{ipsr}^{\text{e}}$ in iteration n of the consensus loop	1
$\lambda_{ijqrs}^{\text{h}}$	share of the investment cost for transmission technology $q \in \mathcal{Q}$ between regions $i, j \in \mathcal{I}$ taken by cost class $r \in \mathcal{R}$ and segment $s \in \mathcal{S}$	1
C_{pt}^{run}	running cost of technology $p \in \mathcal{P}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	k€/GWh
C_t^{exp}	cost of transmitting electricity at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	k€/GWh
M_{ipr}^{e}	cost class potential for generation technology $p \in \mathcal{P}$ in region $i \in \mathcal{I}$ and cost class $r \in \mathcal{R}$	GW
M_{ijqr}^{h}	cost class potential for transmission technology $q \in \mathcal{Q}$ between regions $i, j \in \mathcal{I}$ in cost class $r \in \mathcal{R}$	GW
D_{it}	demand for electricity in region $i \in \mathcal{I}$ at time $t \in \cup_{s \in \mathcal{S}} \mathcal{T}_s$	GWh/h
D_i^{hydrogen}	electricity demand for hydrogen in region $i \in \mathcal{I}$	GWh/h
η_p	efficiency of technology $p \in \mathcal{P}$	1
A_{ip}	regional resources based on land available in region $i \in \mathcal{I}$ for technology $p \in \mathcal{P}$	GW
ξ_p^{min}	minimum share of load for $p \in \mathcal{P}^{\text{therm}}$	1
C_{ipt}^{on}	start-up cost in region $i \in \mathcal{I}$ for technology $p \in \mathcal{P}^{\text{therm}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	k€/(GW·h)
C_{ipt}^{part}	part-load cost in region $i \in \mathcal{I}$ for technology $p \in \mathcal{P}^{\text{therm}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	k€/GWh
$G_{i,p,t}^{\text{active}}$	activated thermal capacity from previous iteration in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{therm}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GW

$G^{\text{on}}_{i,p,t}$	started thermal capacity from previous iteration in region $i \in \mathcal{I}$, technology $p \in \mathcal{P}^{\text{therm}}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	GW
θ_{ipt}	weather profile for region $i \in \mathcal{I}$ of technology $p \in \mathcal{P}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$	1
C^{add}_{pty}	additional, future, running cost for technology $p \in \mathcal{P}$ at time step $t \in \mathcal{T}_s$, $s \in \mathcal{S}$ and year $y \in \mathcal{Y}$	k€/GWh
Z_p	technical lifetime of technology $p \in \mathcal{P}$	years

B Dual multiplier projection examples

This appendix contains some examples of the dual multiplier projection discussed in Section 4.3.1. Recall that in these examples, the break point for problem $n \in \mathcal{N} := \{1, \dots, N\}$ is calculated as $\lambda_n := c - \bar{\pi}_n$, and the partial derivative is given by $\frac{\partial h(\lambda_i)}{\partial \lambda} = -\sum_{n \in \mathcal{N}} \min\{\bar{\pi}_n + \lambda_i \,; c\}$.

Example B.1. We here provide a numerical example to demonstrate why it is necessary to add $\lambda_0 = 0$ as a break point (as done in step 4 in Algorithm 2).

Let $N = 4$ with $\bar{\pi}_1 = 1$, $\bar{\pi}_2 = \bar{\pi}_3 = \bar{\pi}_4 = -1$, and $c = 5$. The corresponding break points are $\lambda_1 = c - \bar{\pi}_1 = 4$ and, for $i = 1, 2, 3$, we get $\lambda_i = c - \bar{\pi}_i = 6$. Thus, to calculate the partial derivative for $\lambda \in [0, 4)$, the break point $\lambda_0 := 0$ needs to be added.

The partial derivative for each break point is then

$$\frac{\partial h(\lambda_0)}{\partial \lambda} = -\sum_{n=1}^{4} \min\{\bar{\pi}_n \,; c\} = -\left(\min\{1 \,; 5\} + 3 \cdot \min\{-1 \,; 5\}\right) = 2$$

$$\frac{\partial h(\lambda_1)}{\partial \lambda} = -\sum_{n=1}^{4} \min\{\bar{\pi}_n + \lambda_1 \,; c\} = -\left(\min\{5 \,; 5\} + 3 \cdot \min\{3 \,; 5\}\right) = -14,$$

$$\frac{\partial h(\lambda_2)}{\partial \lambda} = -\sum_{n=1}^{4} \min\{\bar{\pi}_n + \lambda_2 \,; c\} = -\left(\min\{7 \,; 5\} + 3 \cdot \min\{5 \,; 5\}\right) = -20,$$

and $\frac{\partial h(\lambda_3)}{\partial \lambda} = \frac{\partial h(\lambda_4)}{\partial \lambda} = \frac{\partial h(\lambda_2)}{\partial \lambda} = -20$.

Figure B.1 shows an illustration of the partial derivative as a function of λ, where $\lambda^* = 0.5$. Note that if λ_0 had not been added as a break point, Algorithm 2 would not catch the optimum λ^* as the partial derivative is not positive for the other break points. (Step 7 would set it to 0.)

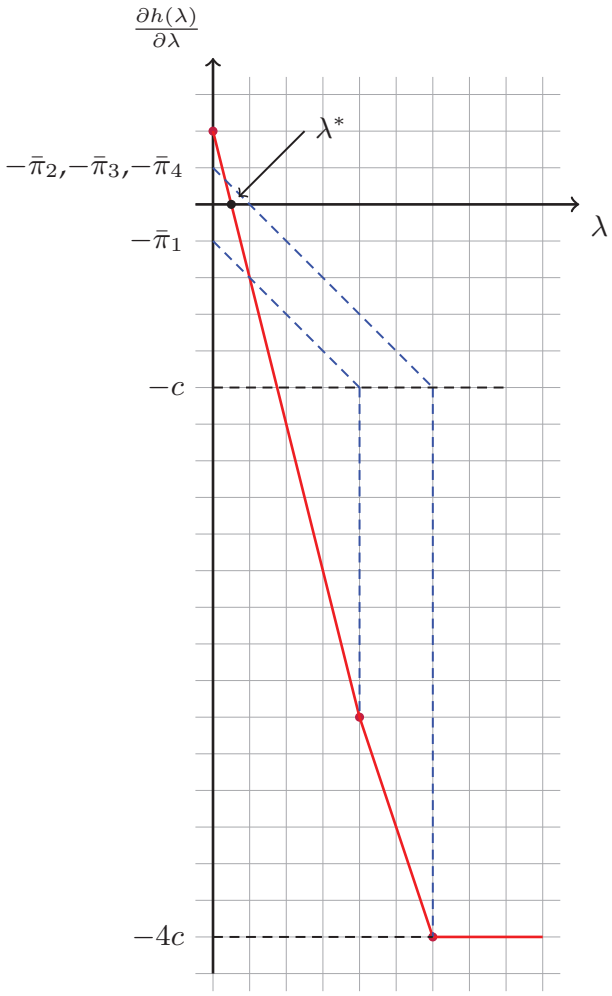


Figure B.1: The partial derivative as a function of λ. The optimal value λ^* is where the partial derivative is equal to zero. Here, this corresponds to $\lambda^* = 0.5$.

□

Example B.2. This example illustrates the case where $\lambda^* < 0$ after the linear interpolation step in Algorithm 2, which indicates that the partial derivative is negative for all $\lambda > 0$.

Let $N = 3$ with $\bar{\pi}_1 = 1$, $\bar{\pi}_2 = 3, \bar{\pi}_3 = 10$, and $c = 5$. The corresponding break points are $\lambda_1 = c - \bar{\pi}_1 = 4$, $\lambda_2 = c - \bar{\pi}_2 = 2$, and $\lambda_3 = c - \bar{\pi}_3 = -5$.

We calculate the partial derivatives:

$$\frac{\partial h(\lambda_0)}{\partial \lambda} = -\sum_{n=1}^{3} \min\{\bar{\pi}_n \,; c\} = -(1+3+5) = -9,$$

$$\frac{\partial h(\lambda_1)}{\partial \lambda} = -\sum_{n=1}^{3} \min\{\bar{\pi}_n + \lambda_1 \,; c\} = -(5+5+5) = -15,$$

$$\frac{\partial h(\lambda_2)}{\partial \lambda} = -\sum_{n=1}^{3} \min\{\bar{\pi}_n + \lambda_2 \,; c\} = -(3+5+5) = -13,$$

$\frac{\partial h(\lambda_3)}{\partial \lambda} = -\sum_{n=1}^{3} \min\{\bar{\pi}_n + \lambda_2 \, ; c\} = -\left(-4 - 2 + 5\right) = 1,$

and let Figure B.2 correspond to the partial derivative as a function of λ.

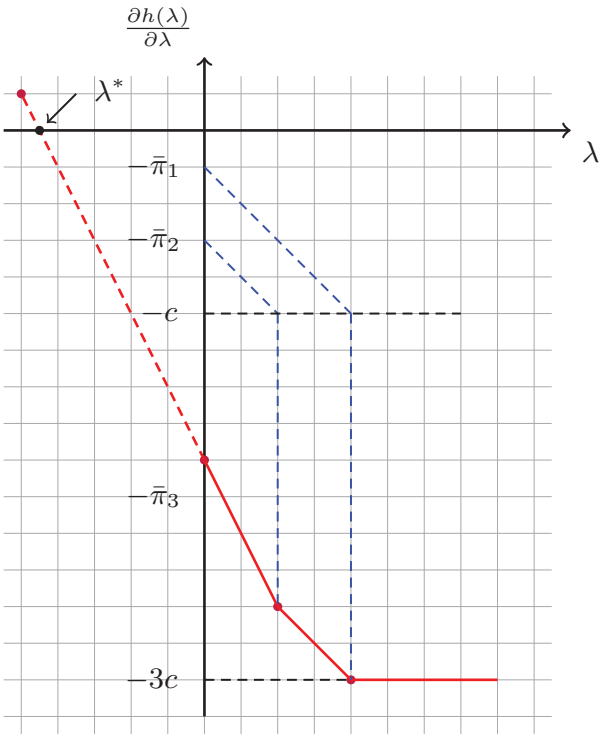


Figure B.2: The partial derivative as a function of λ. The optimal value λ^* is where the partial derivative is equal to zero. For this case, $\lambda^* = -4.5$, which is not feasible.

Here, we note that $\lambda^* = -4.5 \ngeq 0$. Thus, there does not exists a solution where the partial derivative is zero for $\lambda \geq 0$, and thus it must hold that $\lambda^* = 0$. In Algorithm 2, this corresponds to the last **if**-statement in steps 13-14.

□