Regression-based inverter control for power flow and voltage regulation

Oscar Sondermeijer
Regression-based inverter control for power flow and voltage regulation

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

Oscar Sondermeijer

November 23, 2015

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology
The work in this survey is the result of a collaboration with University of California Berkeley
Abstract

Rapid and substantial voltage changes can occur in distribution networks with a high penetration of photovoltaic (PV) systems, due to their unpredictable nature. Electronic power inverters are capable of delivering fast reactive power support to maintain customer voltages within operating tolerances and reduce system losses in distribution feeders. While optimization based paradigms have been proposed to control the reactive power output of inverters, these methods typically rely on the presence of an extensive and fast communication infrastructure which is currently not in place and would be expensive to build. On the other hand, approaches that utilize completely local data require the design of a relation between local measurements and inverter output (i.e. a Volt-VAr curve). These relationships are often naively designed and typically do not yield optimal results. In this work, a systematic and data driven approach is presented to determine PV inverter output as a function of locally obtained measurements in a manner that obtains near optimal results. First, a network model and historic information are used to compute globally optimal settings \textit{a posteriori} for all controllable inverters in the network. Subsequently, a regression approach is used to find a function for each inverter that maps the solely local historical data to an approximation of the globally optimal inverter output. The resulting functions are then employed as decentralized controllers of the inverters and approximate the globally optimal reactive power outputs based on local measurements only. Simulation results on real feeder models demonstrate that this method achieves near optimal results when performing voltage- and capacity-constrained loss minimization and voltage flattening. This method paves the way to an efficient voltage optimization scheme in which legacy control equipment collaborates with existing inverters to facilitate safe operation of distribution networks with higher levels of distributed generation.
# Table of Contents

Abstract i

Acknowledgements xi

1 Introduction 1

1-1 Challenges in power systems .......................... 1

1-1-1 Distribution operation objectives .................. 1

1-1-2 Traditional voltage regulation .................... 3

1-1-3 Challenge of evolving distribution systems ........ 3

1-1-4 New technological solutions ...................... 4

1-2 Project introduction ................................. 6

1-2-1 Research problem ................................ 6

1-2-2 Outline of this thesis ............................ 6

2 Inverter control approaches 7

2-1 Distribution System Modeling and Operation .......... 7

2-1-1 Modeling Power Flow .................................. 7

2-1-2 Traditional Operation and Control .................. 11

2-2 Control for Integration of Renewable Energy .......... 12

2-2-1 Inverter capabilities ................................ 12

2-2-2 Local control methods ............................... 14

2-2-3 Optimization-based control methods ................ 17

2-2-4 Comparison of methods ............................. 20

2-3 Concluding remarks and problem formulation .......... 22

Master of Science Thesis Oscar Sondermeijer
# Table of Contents

## 3 Regression-based inverter control

3-1 Approach: local method inspired by optimal control .......................................................... 23
3-2 Optimal Power Flow .................................................................................................................. 24
   3-2-1 Objective functions .............................................................................................................. 25
   3-2-2 Overall problem .................................................................................................................. 26
   3-2-3 Suggested matrix formulation for implementation ............................................................ 27
3-3 Regression framework .............................................................................................................. 29
   3-3-1 Input data .......................................................................................................................... 30
   3-3-2 Regression method ............................................................................................................. 32
   3-3-3 Validation of obtained models ............................................................................................ 34
   3-3-4 Details of stepwise selection .............................................................................................. 35
3-4 Implementation of controllers ................................................................................................. 37
   3-4-1 Implementation in Simulation ............................................................................................ 37
   3-4-2 Challenge of voltage information ....................................................................................... 38
   3-4-3 Considerations for implementation on physical testbed ..................................................... 40
3-5 Concluding remarks .................................................................................................................. 40

## 4 Results and analysis of two case studies

4-1 Testbeds .................................................................................................................................. 42
   4-1-1 Distribution feeder models .................................................................................................. 42
   4-1-2 Consumption and PV generation data ............................................................................... 44
   4-1-3 Concluding remark ............................................................................................................ 48
4-2 Optimal power flow results ...................................................................................................... 49
   4-2-1 SB feeder optimal power flow ............................................................................................ 49
   4-2-2 IEEE feeder optimal power flow ......................................................................................... 54
   4-2-3 Concluding remark ............................................................................................................ 56
4-3 Regression results and analysis ............................................................................................... 56
   4-3-1 Stepwise settings ............................................................................................................... 59
   4-3-2 Differences in inverter structure ......................................................................................... 62
   4-3-3 Voltage information challenge .......................................................................................... 65
   4-3-4 Available input variables and objective function ............................................................... 69
   4-3-5 Correlation of loads ........................................................................................................... 71
   4-3-6 Timescale of data ............................................................................................................... 74
4-4 Simulation with regression-based controllers ........................................................................... 75
   4-4-1 Cost function values ........................................................................................................... 76
   4-4-2 Voltage and losses ............................................................................................................ 78
   4-4-3 Voltage optimization ......................................................................................................... 81
   4-4-4 Concluding remark ........................................................................................................... 82
4-5 Concluding remarks .................................................................................................................. 83

---

Oscar Sondermeijer  
Master of Science Thesis
# Table of Contents

5 Conclusions and recommendations 85
   5-1 Conclusions ......................................................... 85
   5-2 Recommendations for future research .......................... 87
      5-2-1 Within the proposed approach .......................... 87
      5-2-2 Extending the proposed framework .................... 88

A Effect of reactive power injection 91

B Regression-based control in MATLAB 93

C Feeder models 95

Bibliography 99

Glossary 105
   List of Acronyms .................................................... 105
   List of Symbols ..................................................... 105
# List of Figures

1-1 ANSI C84.1 voltage ranges. Image adopted from [1]. .................................................. 3
1-2 Effect of a capacitor bank on system voltage. The purple area represents the desired range of voltages. Loads are assumed to be inductive. Image adopted from [2]. ........................................... 4
1-3 The duck curve shows typical net load profiles during day for increasing penetration of PV systems in accordance with Californian energy goals. Image adopted from [3] 5
1-4 Overview of the thesis. .......................................................... 6
2-1 Four bus network. ................................................................. 10
2-2 Inverter complex power capabilities. Image adopted from [4]. ................................. 13
2-3 VDE code of practice local control curve with piece-wise linear relation to PV generation. Image adopted from [5]. ................................................................. 16
2-4 Local control curve with piece-wise linear relation to voltage. Image adopted from [6]. ....................................................................................................................... 16
2-5 Feasibility regions of (relaxed) optimal power flow problems. Figure inspired by [7] 20
2-6 Comparison of optimization based control methods and local methods ................ 20
3-1 Four bus network. ................................................................. 28
3-2 Overview of regression approach. ..................................................................................... 30
3-3 Training prediction error, training optimism, and true prediction error for increasing model complexity for hypothetical example. Image from [8] ................ 34
3-4 Scatter plots of standardized optimal inverter reactive power output versus optimal voltage and reference voltage. .............................................................. 39
4-1 Overview of results. ................................................................. 41
4-2 Diagrams of the (a) IEEE feeder [9] and (b) SB feeder. .................................................. 43
4-3 Consumption and generation data from ID 585 in February and July 2014. .............. 45
4-4 Average load correlation for IEEE feeder for varying IDs per load. ............................... 47

Master of Science Thesis Oscar Sondermeijer
4-5  SB feeder 15 minute interval sample data of (a) inverter 83 and (b) inverter 42.

4-6  SB feeder (a) losses and (b) multi-objective cost function values for the strategy A)–C).

4-7  SB feeder substation (a) reactive and (b) real power injection for the strategy A)–C).

4-8  All SB feeder network voltages for strategy A)–C). Colored planes represent the range between the maximum and minimum voltages at any node in the network. Lower voltage bound of 0.95 p.u. is indicated with red dashed line.

4-9  IEEE feeder (a) losses and (b) multi-objective cost function values for strategy A)–C).

4-10 All IEEE feeder network voltages for strategy A)–C). Colored planes represent the range between the maximum and minimum voltages at any node in the network. Lower voltage bound of 0.95 p.u. is indicated with red dashed line.

4-11 Lowest $R^2$ value of all inverter for varying training sample size.

4-12 Scatter plots of optimal reactive power output versus standardized input variables $X_1$–$X_4$ of inverters (a) 634, (b) 645, (c) 671, and (d) 675. Data is obtained from multi-objective optimization.

4-13 Effect of load correlation on prediction accuracy for (a) loss minimizing optimal control data and (b) multi-objective optimal control data.

4-14 SB feeder (a) multi-objective cost function values for strategy A)–D), and (b) relative optimality of B).

4-15 IEEE feeder (a) multi-objective cost function values for strategy A)–D), and (b) relative optimality of B).

4-16 SB feeder network losses for strategy A)–D).

4-17 SB feeder network voltages for strategy A)–D). Colored planes represent the range between the maximum and minimum voltages of all nodes in the network. Lower voltage bound of 0.95 p.u. is indicated with red dashed line.

4-18 SB feeder (a) minimum and maximum voltages and (b) losses for strategy B) and E).

A-1  Distribution feeder schematic with substation and the first three nodes.

C-1  SB feeder diagram with inverter and load locations.
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-1</td>
<td>Feeder characteristics.</td>
</tr>
<tr>
<td>4-2</td>
<td>Average validation $R^2$ values of four IEEE feeder inverter regression models with fixed model structure for different training and validation data.</td>
</tr>
<tr>
<td>4-3</td>
<td>Inverter location selection considerations for IEEE feeder.</td>
</tr>
<tr>
<td>4-4</td>
<td>Comparison of losses and multi-objective cost function values for different approaches. All values are in kW.</td>
</tr>
<tr>
<td>4-5</td>
<td>Experiment overview.</td>
</tr>
<tr>
<td>4-6</td>
<td>Effect of selection criterion on average number of variables and average $R^2$ values of inverter regression models.</td>
</tr>
<tr>
<td>4-7</td>
<td>Effect of initialization on regression model of IEEE feeder inverter 671, and average number of variables of all four IEEE feeder inverters. Average $R^2$ values of were 0.976 for all initializations.</td>
</tr>
<tr>
<td>4-8</td>
<td>Effect of model complexity on regression model of IEEE feeder inverter 671, and average number of variables and average $R^2$ values of all four IEEE feeder inverters.</td>
</tr>
<tr>
<td>4-9</td>
<td>Complete models of all four IEEE feeder inverters for multi-objective optimization data.</td>
</tr>
<tr>
<td>4-10</td>
<td>Correlation coefficients between optimal reactive power output versus standardized input variables $X_1$–$X_4$ of all inverters for multi-objective optimal control data.</td>
</tr>
<tr>
<td>4-11</td>
<td>$R^2$ values of inverter models with voltage obtained from simulation with regression-based control implemented based on multi-objective optimal control data.</td>
</tr>
<tr>
<td>4-12</td>
<td>$R^2$ values of inverter models where a single input variable is removed for multi-objective optimal control data and loss minimizing optimal control data.</td>
</tr>
<tr>
<td>4-13</td>
<td>Relation between independent variables.</td>
</tr>
<tr>
<td>4-14</td>
<td>$R^2$ values of SB feeder inverter models based on multi-objective optimal power flow data with and without voltage information.</td>
</tr>
<tr>
<td>4-15</td>
<td>$R^2$ values for 1- and 15 minutes interval data with and without voltage information. Data obtained with multi-objective optimal power flow on SB feeder.</td>
</tr>
<tr>
<td>C-1</td>
<td>Network data of the IEEE feeder.</td>
</tr>
<tr>
<td>C-2</td>
<td>Network data of the SB feeder.</td>
</tr>
</tbody>
</table>
During the past eleven months I had the opportunity to work on a project that allowed me to have a truly amazing year in California and the Netherlands. I would like to thank everyone who had a part in this experience, and specifically the people whom I have collaborated with. First of all, Roel, my direct and often daily advisor, for all his support since arranging this master project until the very last read of this thesis. His confidence and support allowed me to pursue my sometimes farfetched ideas, while I could always rely on him to help me bring back some structure and warrant the quality of the result. Daniel Arnold, for the motivating weekly meetings that shaped the research, challenged my ideas, and eventually resulted in a paper. Tamás Keviczky, for his always crystal clear advice and support throughout my stay in Berkeley and in Delft, even if that required evening Skype sessions. Claire Tomlin, for facilitating this project and adopting me in the Hybrid System group, which allowed me to learn from truly inspiring people. Sascha von Meier for letting me participate in her class and the discussions in the weekly smart grid meetings, both have helped me get a better appreciation of power systems. Kyle Brady and Michael Sankur, for their help with the simulation testbeds, and John Mead and his team at PG&E Applied Technology Services for the fruitful discussions and site visit.

Furthermore, I would like to thank many fellow students in Berkeley and in Delft for the fruitful discussions and coffee chats, especially my roommates Luis, Joris, and Frank. Finally, I want to express my gratitude towards my family and girlfriend whom have granted me the opportunity and encouraged me to perform to the best of my ability.

Delft, University of Technology
November 23, 2015

O. Sondermeijer
Acknowledgements
Chapter 1

Introduction

This thesis presents a framework to design controllers for electronic power inverters that enhance power flow and power quality in distribution networks. Section 1-1 introduces the challenges of the present-day distribution system, and Section 1-2 describes the research problem and thesis outline.

1-1 Challenges in power systems

In the traditional power system paradigm, electricity is produced at large scale power plants, transported over long distances through the transmission network, and distributed to millions of individual customers through distribution networks. This paradigm is changing because of the increasing penetration of renewable distributed energy resources: California installed 0.5 MW of solar electric capacity every hour over 2014 [10]. Moving from a system that is mainly powered by large fossil fueled electricity plants towards a system with more renewable energy resources has many advantages, but comes with new technical challenges. A direct benefit is the reduction of pollution and CO₂ emissions, which contribute to global rising temperatures, and climate change [11]. More fundamentally, development is absolutely necessary to assure future generations of electricity as traditional, fossil, energy sources are being depleted [12]. The immediate challenge is maintaining high quality electrical power throughout the network, while renewable distributed energy resources have uncontrollable and volatile power output [4].

1-1-1 Distribution operation objectives

Distribution systems link the high voltage transmission system to thousands of individual loads. The primary function of the distribution network is lowering the voltage from the 60–500 kV transmission network range, to the 4–35 kV distribution network range, and finally to the service voltage that allows you to plug your laptop charger in any socket. You expect to be able to charge your laptop – whenever and wherever – without damaging it. Therefore, the
overall goal of system operators and electric utility companies is to provide reliable, secure, and stable power of high quality to all customers. Although system reliability, security, and stability are primarily enforced at the transmission level, a brief explanation and the main implications for the distribution system operator (DSO) follow. A more detailed overview of these concepts, and the relation between them, is presented in [13].

1. **Reliability** concerns sourcing sufficient generation, and the ability to deliver the power to all customers [2, Ch.8]. In a distribution network this implies having sufficient line capacity for the expected power flow, maintaining equipment to prevent outages, and adequately operating switches to prevent propagation of faults.

2. **Security** refers to the number of contingencies that can occur before equipment is damaged, or loads are interrupted [2, Ch.8]. In this context, it is the responsibility of a DSO to have the possibility to reconfigure the distribution network in case of a contingency, which also requires sufficient line capacity and switches.

3. **Stability** is the ability of alternating current (AC) power to regain equilibrium conditions after a disturbance [13]. It is typically separated in angle, frequency, and voltage stability. Voltage stability is most relevant to a DSO, and implies that more power is delivered as loads increase, instead of a reduction of voltages. DSOs can enhance voltage stability with local reactive power compensation and voltage support [2, Ch.8].

4. **Power quality** is the combination of having:
   - (a) Voltage magnitude that is close to the nominal value. Voltage regulators prevent the voltage drop in the network line to become too large, as low (or high) voltages damage loads.
   - (b) Frequency that is close to nominal value. Frequency deviations are caused by unbalance between generation and consumption, and mainly affect synchronous machines [2, Ch.8].
   - (c) Sine waveform with low harmonic content. Harmonics are caused by transient effects of both loads, as well as generators, and damage electric motors and transformers [2, Ch.8].

If all objectives are met, system operators have additional (economical) incentives to:

5. Minimize the cost of operating the system, by e.g., minimizing total line loss, or optimal sizing and placement of capacitors [14, 15].

6. Either *increase service voltage* to reduce line losses and increase revenues by increasing power consumption [2, Ch.8]. Or, conversely, reduce energy consumption by *lowering service voltage*, which is referred to as conservation voltage reduction (CVR). This is performed with a combination of reactive power compensation by capacitors and voltage optimization with tap-changing transformers [16]. Optimizing voltages with reactive power support is often referred to as Volt-VAr optimization (VVO).

Regarding voltage level regulation, ANSI Standard C84.1 [1] specifies that, under normal operation, service voltage of a 120 V nominal system falls in Range A of Figure 1-1. In 120 V
1-1 Challenges in power systems

systems, this corresponds to nominal value plus or minus 5%. Occurrence of system voltages in Range B should be infrequent and limited in duration, otherwise corrective measures are undertaken to bring voltages back to Range A. Although undesirable, voltages outside Range B occur, and can trip protective devices when the duration is too long [1].

1-1-2 Traditional voltage regulation

Traditional voltage regulation is performed with transformers and capacitors [2, Ch.6]. Tap-changing transformers at the substation, or in the network, can raise or lower voltages within a prescribed range. Capacitor banks in the network generate reactive power locally to reduce losses and voltage drop across distribution lines. Figure 1-2 compares the voltage profile of a feeder with and without a capacitor bank; the capacitor bank alters the voltage profile to stay within the desired bound. Reactive power injection at the location of the capacitor reduces the reactive power flow between the substation and capacitor, which reduces losses and voltage drop on that section.

Legacy voltage regulators were developed with the traditional power system paradigm in mind. The functionality assumes unidirectional power flow from the substation to the loads, with corresponding monotonic voltage drops over the feeder, as depicted in Figure 1-2. Furthermore, the design assumes predictable load characteristics, and infrequent switching.

1-1-3 Challenge of evolving distribution systems

Where the power output of a power plant can be accurately controlled to meet demand, this is not the case for PV systems. Passing clouds lead to instantaneous and temporary sags in irradiation and power output of up to 90% [17]. Also, distributed renewable generation

Figure 1-1: ANSI C84.1 voltage ranges. Image adopted from [1].
can exceed local consumption, which causes bidirectional power flow. Simultaneously, the historical overcapacity of the network reduces by electrification of loads, e.g., the large scale adoption of electric vehicles. As a result, distribution lines and equipment are operated closer to their capacity limits, and the risk of overload during peak consumption and contingencies increases. Finally, increasing number of smart-devices affect the predictability of loads, and change overall load characteristics.

Despite the rapid development and implementation of PV in the power system, power system operation has yet to transform. First, power plants are unable to cope with the intermittent PV output; therefore, electric utilities resort to legacy distribution equipment to prevent voltage excursions from ANSI Range A. This results in increased operation and switching of transformers and capacitors, which significantly reduces their lifetime [17]. Second, bidirectional power flow damages certain types of transformers, and reverse current can falsely trip protective equipment. Third, reverse power flow violates the monotonically decreasing voltage drop assumption, and reactive power compensation of a capacitor bank can have unexpected (and undesirable) effect on downstream voltage. The fourth anticipated effect of increased penetration of PV systems is larger ramping of power plants around sunset. This effect is shown in the so called duck curve of Figure 1-3 [3]. The California independent system operator (CAISO) foresees low output of power plants during the day, when PV systems produce a large share of consumption and potentially generate more than local consumption. But, PV generation diminishes as the sun sets, which coincides with the rise in consumption towards the peak at 8pm. In the example of Figure 1-3 this causes the power plant to approximately double its output in only three hours. Because of these effects, the limits of the existing U.S. distribution network are expected to be reached when the penetration of renewable energy sources increases to government goals [17].

1-1-4 New technological solutions

The challenges of a transition to more sustainable power systems require a combination of solutions to meet the power system objectives. Examples are, storing abundant PV power in batteries to prevent reverse power flow; making consumption more flexible and controllable by demand response to cope with uncontrollable renewable energy sources; using electronic power inverter control. Modern inverters are capable of curtailing real power, absorbing or
injecting reactive power, and can be interfaced with batteries to reduce the intermittent effect of PV systems.

Research shows that recruiting electronic power inverters potentially improves power quality throughout the system, and prevents damage to existing equipment, e.g., [4, 6, 18, 19]. The primary task of electronic power inverters is to convert real power generated by PV systems from direct current (DC) to the grid’s alternating current (AC). However, modern inverters have the capability to generate or absorb reactive power simultaneously with the processing of real power. Reactive power support can be used to change the voltage profile similar to the effect of the capacitor bank, as depicted in Figure 1-2. Active voltage regulation functionalities of inverters are not yet employed in accordance with the IEEE 1547 standard, but extension of this standard to allow for supplemental support is topic of serious consideration and might be released in the near future [20].

Recruiting inverters for reactive power support has several advantages. First, it can relieve legacy voltage regulator operation, and prevent increased wear. Second, reactive power output of inverters is more continuous than the binary effect of capacitor bank switching. Third, contrary to capacitor banks, inverters can absorb reactive power to reduce voltage rise caused by high PV output. Fourth, PV inverters are located at the source of the increased uncertainty and variability, where reactive power support is most effective [2]. Finally, new, utility scale inverter-interfaced reactive power compensators are developed as the modern alternative to capacitor banks; these offer the same functionality, and require novel control strategies. Inverter-based reactive power support control strategies are topic of recent work, but many challenges remain.
1-2 Project introduction

1-2-1 Research problem

Reactive power support provided by electronic power inverters requires controllers that determine the amount of generated or absorbed reactive power. The goal of this thesis is to formulate a structured approach to design inverter controllers that use reactive power to maintain and enhance the efficiency and quality of electric distribution systems with high penetration of solar generation. Specifically, the control must have close-to-optimal results while respecting practical limitations of distribution systems.

A more rigorous formulation requires more insight on practical limitations and globally optimal power flow (OPF) is included in Chapter 2-3.

1-2-2 Outline of this thesis

Figure 1-4 depicts the organization of this thesis. Chapter 2 introduces the required background on power system modeling and operation, including an overview of proposed inverter control approaches. Chapter 3 presents the control framework proposed by this thesis. This includes details on the offline optimization, the regression method, and the envisioned implementation. Chapter 4 evaluates the framework based on two case studies. The results are extensively analyzed under different scenarios to gain more insight on the method’s strengths and limitations. The first case study is an IEEE standard distribution test feeder, with few nodes, but relatively large loads. The second case study uses an American distribution model with more nodes and smaller loads. The load data used for both feeders are obtained from a database that monitored real houses with PV systems in Phoenix, Texas.
Chapter 2

Inverter control approaches

The present-day challenges in the operation of the distribution system were introduced in Section 1-1. Traditional voltage regulators are used to warrant power quality in distribution networks. However, due to a rapid adoption of distributed and renewable generation such as photovoltaic (PV) systems, power flow in distribution networks significantly evolved from the traditional paradigm. The reactive power capabilities of inverters are suggested to improve power quality and power flow in a feeder. It is important to understand power flow in a distribution system and existing proposed inverter control methods before the control framework of this thesis is presented. Therefore, Section 2-1 introduces distribution system power flow modeling, and subsequently presents more insight into traditional voltage regulation. Section 2-2 introduces and compares two categories of inverter control to cope with distributed renewable generation in the distribution system.

2-1 Distribution System Modeling and Operation

The goal of this section is to develop more insight into distribution system modeling techniques, and with this intuition further elaborate on the effect of reactive power support. To that end, Subsection 2-1-1 motivates the power flow models that are used in this work, and Subsection 2-1-2 discusses the traditional operation and control of the distribution network.

2-1-1 Modeling Power Flow

The first step in inverter control design for enhancing power flow and voltages is to understand – and model – the underlying relation between different states of the network. Power systems are constantly subjected to small changes in the network, e.g., loads change, switches are operated, transformers change taps, and the output of PV systems change. Each change produces transient effects on the sinusoidal currents, however the time varying effects are generally small and last for a few 60 Hz cycles only [21, Ch.3.2]. For the purpose of this thesis – controlling the magnitude of inverter reactive power output – equilibrium states
Inverter control approaches

are assumed. Analyzing steady state operation of power systems is extensively described in literature [2, 21, 22], and uses static power flow models, which are less involved than dynamic models. Peak load analysis is a traditional application of steady state power flow models which uses a single load profile to assess if all components in the network are sized correctly. However, inverters can be controlled at all times, and under very different scenarios. Quasi-static modeling incorporates different load profiles in the system, but considers changes in the system to occur slowly so that the system is in equilibrium at all times. Effectively, models for static and quasi-static analysis are identical and assume steady state sinusoidal current.

Balanced and unbalanced power flow

Three phase transmission networks have roughly equal loads on all phases and transposed lines. This justifies the assumption that power flow on each phase is balanced, and can be represented with a single phase model [23], commonly referred to as balanced power flow. Conversely, distribution feeders have inherently unbalanced power flow because of the large number of single-phase loads, and because feeders consist of single-, two-, or three-phase line segments [23]. Not surprisingly, unbalanced power flow models are more complex than balanced models. Because of this, and the availability of well developed balanced power flow methods from the transmission network, balanced power flow is commonly assumed for distribution modeling and analysis. Whether or not this is desirable is debatable [23, 24, 25, 26], and, to the author’s knowledge, guidelines do not exist. Ideally, a comparison between a balanced and unbalanced power flow solution is used to assess if a balanced power flow model is sufficiently accurate. This thesis assumes balanced power flow, but the extension to unbalanced models is possible and included as a recommendation for future research in Chapter 5.

Conditions in electric networks are described with Ohm’s (2-1), Kirchhoff’s current (2-2), and Kirchhoff’s voltage (2-3) laws. Ohm’s law states that the difference in potential, or the voltage drop, across a conductor is the product of current and impedance. Applying Kirchhoff’s laws to current and voltage phasors states that the sum of all \( m \) phasor currents entering any node is zero, and the sum of the \( m \) phasor-voltage drops around any closed path is zero [21, Ch.2.4].

\[
V = Iz \quad (2-1)
\]

\[
\sum_{k=1}^{m} I_k = 0 \quad (2-2)
\]

\[
\sum_{k=1}^{m} V_k = 0 \quad (2-3)
\]

Multi-phase unbalanced power flow models are build from these principles. Complex impedance \( z = r + jx \) of a three-phase line section consists of self-impedance of a conductor and mutual-impedance caused by the presence of the other conductors. Multiphase impedance is therefore commonly specified as the matrix \( z_{abc} \) (2-4), where the diagonal entries denote the self-impedance and non-diagonal entries denote the coupling between pairs of conductors [23, Ch.4]. In a balanced system all self- and mutual-impedances are equal, and denoted with
$z_s$ and $z_m$ respectively. The impedance matrix can then be expressed as self- and mutual-impedances only (2-4). A single impedance value for the corresponding one-line diagram can be expressed with the sequence component method [23, Ch.4]. This method states that the self- and mutual-impedance values can be lumped into the positive sequence impedance $z_1$ (2-5), which represents the impedance of the corresponding single-phase network.

$$z_{abc} = \begin{bmatrix} z_{aa} & z_{ab} & z_{ac} \\ z_{ba} & z_{bb} & z_{bc} \\ z_{ca} & z_{cb} & z_{cc} \end{bmatrix} = \begin{bmatrix} z_s & z_m & z_m \\ z_m & z_s & z_m \\ z_m & z_m & z_s \end{bmatrix}$$ (2-4)

$$z_1 = z_s - z_m.$$ (2-5)

For unbalanced systems, however, the impedance values are unequal, mainly because unbalanced power flow results in different mutual reactance between the lines. Therefore, applying the sequence component method requires averaging of the self- and mutual-impedance values [23, Ch.4]. This is illustrated with an example (2-6) of a three-phase impedance matrix from the original model of the IEEE 13 node distribution test feeder model that is used in Chapter 4. The impedance values of example (2-6) have units of $\Omega$/mile.

$$z_{abc} = \begin{bmatrix} 0.3465 + j1.0179 & 0.1560 + j0.5017 & 0.1580 + j0.4236 \\ 0.1560 + j0.5017 & 0.3375 + j1.0478 & 0.1535 + j0.3849 \\ 0.1580 + j0.4236 & 0.1535 + j0.3849 & 0.3414 + j1.0348 \end{bmatrix}$$ (2-6a)

$$z_s = \frac{1}{3} ((0.3465 + j1.0179) + (0.3375 + j1.0478) + (0.3414 + j1.0348))$$ (2-6b)

$$= 0.3418 + j1.0335$$

$$z_m = \frac{1}{3} ((0.1560 + j0.5017) + (0.1580 + j0.4236) + (0.1535 + j0.3849))$$ (2-6c)

$$= 0.1558 + j0.4367$$

$$z_1 = (0.3418 + j1.0335) - (0.1558 + j0.4367)$$ (2-6d)

$$= 0.1860 + j0.5968$$

**Branch flow equations**

$$P_i = p_i + \sum_{j:(i,j)\in E} (P_j + r_j f_j)$$ (2-7a)

$$Q_i = q_i + \sum_{j:(i,j)\in E} (Q_j + x_j f_j)$$ (2-7b)

$$v_i = v_j + 2 (r_j P_j + x_j Q_j) + \left( r_j^2 + x_j^2 \right) f_j \quad \forall j : (i,j) \in E$$ (2-7c)

$$f_i = \frac{P_i^2 + Q_i^2}{v_i}$$ (2-7d)

The **branch flow equations** (2-7) are the result of applying basic physical laws to a radial single-phase network. They were first presented in [14, 15], and are also known as the DistFlow...
equations. A unique power flow solution with feasible voltage magnitude always exists for practical radial distribution systems when the substation voltage is given and demand at all nodes is known [27]. The equations are best explained with the schematic line segment of Figure 2-1, and introducing graph notation \( G(\mathcal{N}, \mathcal{E}) \) for a balanced radial distribution feeder. The notation from [28] is adopted with small alterations; the set \( \mathcal{N} \) represents the nodes of the feeder indexed by \( i = 0, 1, \ldots, n \), where the substation is indexed 0. \( \mathcal{E} \) is the set of edges (line segments) \((i, j)\) that connect nodes \( i \) and \( j \). The substation is modeled to have constant voltage, which is a common approach to solve power flow and decouple the network from the rest of a power system [2, 23]. This notation denotes real and reactive power demand at node \( i \) as \( p_i \) and \( q_i \), respectively. Power consumption is treated as positive demand and power injection (e.g., generation from PV systems) as negative demand. Hence, the net real power demand is consumption minus generation \( p_i = p_i^c - p_i^g \). Capitals \( P_j \) and \( Q_j \) are used for real and reactive power flow on the line to node \( j \). The phase angle of the complex voltage is eliminated by using the squared voltage magnitude \( v_i = |V_i|^2 \). Similarly \( \ell_j = |I_j|^2 \) represents the squared magnitude of the current flowing into node \( j \). The resistance and reactance of the complex impedance of line \((i, j)\) are denoted by \( r_j \) and \( x_j \).

The interpretation of (2-7a) and (2-7b) is that the power flowing into node \( i \) equals the net demand at node \( i \) plus the sum of: all powers flowing into the nodes \( j \) that are downstream of and connected to node \( i \); and the power that is lost as heat on the lines \((i, j)\) connecting node \( i \) and the downstream connected nodes \( j \). For the example of Figure 2-1, the real power flowing into node \( i \) is the sum of demand \( p_i \), real power flowing into nodes \( j \) and \( k \), and the losses on lines \((i, j)\) and \((i, k)\).

The voltage drop between nodes \( i \) and \( j \) in (2-7c) equals \( 2(r_j P_j + x_j Q_j) + (r_j^2 + x_j^2) \ell_j \). The first term expresses voltage drop as a function of the power that flows into node \( j \); the second term accounts for the power that is lost progressively on the line, but still contributes to the voltage drop. Typically, the second term is much smaller, and is sometimes neglected to obtain linearized power flow equations, e.g., in [29].

Although this thesis uses the branch flow equations, it is worth mentioning that it is not the only formulation for power flow in balanced networks. Another popular representation, the
bus injection model, is first described in 1962 to solve an optimal power flow problem [30]. This model is traditionally used for power flow analysis and optimization, and focuses on the current phasors, and voltages and complex injections at the nodes, instead of power flows on lines. Both models are equivalent for a radial network [31, 32]. The choice for one of these models can be based on the preferred method for solving power flow, which, in this thesis, is performed by solving a convex optimization problem. Or in a practical setting, the choice can be based on the availability of measurements.

2-1-2 Traditional Operation and Control

The operation objectives for distribution systems were introduced in Section 1-1. The overall goal of system operators is to provide reliable, secure, and stable power of high quality to all customers. The most pressing concern (with regard to PV systems) is to perform voltage regulation with an aging infrastructure of equipment [21] to meet ANSI Standard C84.1 [1]. With a better understanding of power flow from Section 2-1, it is possible to gain more insight in the underlying physics of reactive power support.

Recall that transformers and capacitors are traditionally used for voltage regulation. Transformers raise or lower downstream voltages, and can be placed at the substation or in the network. Capacitors are used for reactive power support which was introduced in Section 1-1. Intuitively the voltage effect of reactive power injection by a capacitor is described as follows. Typical loads are mostly inductive, and the reactive power they consume must be generated at a power plant. In the setting of a distribution network this implies that reactive power is delivered by the substation, and travels over the distribution lines to the node at which the reactive power is consumed. Reactive power flow increases the current through that line, which increases loss and voltage drop. Locally supplying the reactive power with a capacitor bank reduces the distance that reactive power has to flow, and hence reduces the loss and voltage drop. Reactive power injection thus has a positive effect on losses and voltage drop in the system, which is also shown in Figure 1-2.

The branch flow equations can aid this intuition of the voltage effect of reactive power injection. Reactive power injection of a capacitor can be regarded as negative demand at a node \(i\), which reduces the reactive power flowing into that node (2-7b). Equality (2-7d) states that this reduces the current that flows into node \(i\), and (2-7c) shows that the terms \(2(r_j P_j + x_j Q_j) + (r_j^2 + x_j^2) \ell_j\) decrease. Hence, the voltage drop \(v_i - v_j\) over the line \((i, j)\) is reduced. In a similar fashion, the injection of reactive power at node \(i\) reduces the voltage drop on all lines between the feeder head and node \(i\).

Traditional voltage regulators have been able to guarantee a high quality power in distribution networks. The combination of transformers and capacitors allowed changing overall voltages, reducing voltage drop, and reducing losses. However, the existing equipment in the United States of America (U.S.) is aging, and was designed for predictable and slowly changing load profiles. As a result, conventional capacitor bank control is limited to on/off. Similarly, tap changing transformer control is limited to a few settings, which are meant to change infrequently (a few times a day) [17, 21]. This old paradigm is ill-equipped to cope with the increased variability caused by renewables. Simultaneously, an increasing amount of reactive power capacity in the form of residential scale inverters is installed – but unused. This enables new control paradigms to optimally recruit the new, and more flexible reactive power capacity.
2-2 Control for Integration of Renewable Energy

Currently, voltage regulation and reactive power support is provided with traditional hardware and methods. However, the rapid adoption to uncontrollable renewable energy resources changes the nature of power flow in the distribution network. These characteristics were not anticipated when the traditional voltage regulators were designed. Therefore, it is necessary to think of new control approaches to prevent deterioration of power quality, and increased wear of traditional equipment which necessitates earlier replacement and increases cost of operation.

As of now, IEEE standard 1547 [33] prevents PV inverters to actively control their reactive power output in the U.S.. However, the potential of PV inverter reactive power support is acknowledged: an IEEE workgroup studies opportunities to adapt this standard [20]; and the Californian Energy Commission’s "Smart Inverter Working Group" develops advanced inverter functionality [34]. Contrary to the U.S., Germany has already implemented inverter control [5]. The majority of proposed solutions found in recent literature, including the German paradigm, can be split in two categories:

Local control methods: controlling reactive power using measurements at the inverter node with limited or no communication requirements.

Optimization-based control methods: solving an optimization problem at centralized or distributed agents and communicating the optimal solution to inverters.

First, Subsection 2-2-1 elaborates on the functionalities of electronic power inverters in Germany and the proposed functionalities in the U.S.. Subsequently, Subsection 2-2-2 discusses local control methods, Subsection 2-2-3 presents optimization-based control methods, and Subsection 2-2-4 compares the two methods.

2-2-1 Inverter capabilities

Before both categories are discussed in more detail, the reactive power capability of inverters is presented. Inverter modeling is an active field of research in itself, but for this thesis it is sufficient to use the following model for reactive power capacity (2-8).

$$q^g = \sqrt{s^2 - (p^g)^2}, \quad (2-8)$$

where \( s \) is the inverter apparent power capacity and \( p^g \) the real power output of the PV system. This is best illustrated in the complex power plane with Figure 2-2. The reactive power capacity of an inverter depends on the amount of real power produced by the PV system. Although it is technically possible to curtail the real power production of a PV panel to increase the reactive power capacity, and the proposed framework naturally allows for it, this thesis focuses on controlling inverter reactive power. As such, it respects the current trend of injecting all surplus real power into the grid.

The German VDE 4105 code of practice [35], and the EEG renewable energy sources act [36], set four requirements for PV systems and other generators with less than 100 kW of capacity at low voltage levels of distribution networks in Germany:
1. Participation in feed-in management or general limitation of feed-in power to 70% of installed PV capacity (EEG), aimed at reducing real power generation when it exceeds demand.

2. Real power reduction in case of over frequency (VDE), aimed at actively controlling power frequency.

3. Maximum permissible unbalanced load (VDE), limits single phase PV systems to 4.6 kVA to assist in distribution network load balancing.

4. Provision of reactive power (VDE, EEG) to prevent high voltages under high generation scenarios. Residential scale inverters must have sufficient capacity to operate at a power factor of 0.95 during peak PV generation, which implies an inverter overcapacity of approximately 5%. PV systems larger than 13.8 kVA must support a power factor of 0.9, implying an overcapacity of approximately 10%.

For a concise overview of these requirements the reader is referred to a publication of a leading inverter manufacturer [5]. These requirements can be monitored and enforced simultaneously. For example, real power generation is reduced when the frequency becomes too high, while reactive power is absorbed to reduce voltages. This reactive power support control scheme falls within the local control category.

The Californian Rule 21 Smart Inverter Working Group [34] proposes a three phase implementation of advanced inverter functionality. Only the most relevant aspects are mentioned below; the interested reader is referred to [34] for more details.

1. Phase 1 - autonomous (local) inverter functions, such as:
   (a) Reactive power provision based on voltage information ("Volt-VAr" control).
   (b) Reactive power provision based on PV output (constant power factor control).

2. Phase 2 - communication infrastructure between three levels:
(a) Autonomous PV controllers.
(b) Facility (residence) generation and consumption energy management.
(c) DSO and retail energy provider communication system.

3. Phase 3 - additional functionality, such as monitoring and controlling distributed generation status and real power output.

The functionalities of Phase 1 are comparable to the German functionalities, but the functionalities described in Phase 2 and 3 are more advanced. However, recall that the German functionalities are operational, and the Rule 21 functionalities are currently being developed. Furthermore, the economical compensation for these grid support functions is considered an open issue [34].

2-2-2 Local control methods

The first category of proposed inverter control methods uses local information and decentralized controllers to determine the reactive power output of an inverter. With local measurements and information, this thesis refers to information at one location of interest in the network. Hence, a local inverter control paradigm implies control that is based on information at the inverter node only. This section discusses what local information is available, and presents the German control paradigm and other proposed methods for the U.S. power system.

Depending on the level of detail of an analysis, the notion of local can have a rather different practical interpretation. If a node represents the combination, or aggregation, of 100 residential customers, local refers to aggregated information at the point of interconnection with the network. In this scenario, a local measurement of real power demand is interpreted as the total demand of all 100 customers. This requires either access to the point of interconnection of that node, or a communication infrastructure between all 100 customers. Conversely, when a more detailed feeder model is used, a local measurement can refer to information of a single residence. Examples of local information are:

- Voltage phasor.
- Real and reactive power consumption.
- Real power generation (of PV system).
- Inverter capacity.
- Frequency.

Theoretically, no communication infrastructure is required when reactive power output is controlled with local information only. However, German EEG renewable energy sources act [36] and Californian working groups [34] require inverters to respond to broadcast signals. Similar functionality is described in [6], where inverters can be remotely controlled by a DSO to switch between several modes of operation, with corresponding separate control laws. These scenarios encompass a light, one-way, communication infrastructure.
German local control

Several local control strategies have been proposed that differ not only in the measurements they use, but also in their goals. German approaches are primarily aimed at reducing the effect of excessive generation on voltages. The amount of reactive power provision can be specified in two ways according to the VDE 4105 code [35]:

1. Fixed power factor specified by the grid operator.
2. Variable power factor specified by standard characteristic.

The standard characteristic is shown in Figure 2-3, and specifies that inverters absorb reactive power when the PV system generates more than 50% of its capacity. PV systems with less than 13.8 kVA capacity are required to support a minimum power factor of 0.95, whereas larger systems support a minimum power factor of 0.90. This approach is designed for the situation where excessive distributed PV generation results in voltages rises in the network, and absorbing reactive power reduces the voltage rise.

Although this approach has a beneficial effect during high voltage scenarios, it increases line losses and does not provide reactive power support when PV generation is less than 50% of its capacity. Whether this is the most suitable controller for a given network depends on several conditions. One of the underlying assumptions of this method is that PV systems generate significantly more than is locally consumed: German rooftop PV systems typically have a peak generation of 20 kW [37], whereas the average energy consumption of a household in Germany is 0.3 kW [38]. This local control method is tailored to the German situation, and might not have the desired effect elsewhere. For example, the leading PV system installation company in the U.S. suggests a 4.5 kW system for an average household in the U.S. [39], while the average consumption in the U.S. is 1.4 kW [38]. When local consumption and peak generation are closer together, the voltage rise will be smaller, and the German local control paradigm has potentially undesirable effect. The control method of Figure 2-3 is simulated in Section 4-4.

Volt-VAr control

One of the leading proposed solutions for the U.S. power system [6] advocates using Volt-VAr control with the droop curve shown in Figure 2-4 to balance voltage sags and swells that are caused by passing clouds. The intuition behind this method is that a sudden reduction in PV real power generation can be observed as a sag in voltage. A carefully tuned droop curve responds to low voltages by generation of reactive power. In this example, the reactive power generation counters the loss of real power generation and reduces the sag in voltage. On the contrary, if real power injection of PV systems causes voltages to rise above their typical −or nominal− value, reactive power absorption can lower voltages.

Other proposed local control

Reducing voltage variation between two nodes and minimizing line loss requires opposite reactive power control action in specific situations [4]. Pursuing a multi-objective goal benefits
Figure 2-3: VDE code of practice local control curve with piece-wise linear relation to PV generation. Image adopted from [5].

Figure 2-4: Local control curve with piece-wise linear relation to voltage. Image adopted from [6].
from a controller that uses more information, such as voltage and reactive power consumption measurements. An example of such a multi-objective controller is presented in [4], and is completely aimed at loss reduction when voltage is 1 p.u., but performs voltage regulation when voltages are close to the upper or lower voltage bounds. This multi-objective approach can not be obtained with a single voltage measurement, and requires access to real power consumption and generation, and reactive power consumption.

A more systematic local inverter control paradigm is proposed in [28], which uses an extremum seeking approach to minimize the real power injected by the substation. This approach uses local measurements and a local optimization that minimizes losses. The results in [28] advocate using alternatives to global optimization based inverter control methods, but acknowledge that further development of the method is required to deal with voltage regulation.

Commercial solutions exist to enhance power flow and power quality in distribution grids [40, 41]. These solutions include inverter-based functionality with a separate communication and control infrastructure; the devices can operate with local or centralized control. However, these solutions require DSOs to invest in additional hardware, while the already installed reactive power capacity of PV inverters remains unused.

Research has shown that local inverter control strategies have a beneficial effect on distribution system operation in a variety of conditions and scenarios. An essential aspect is the selection of the most suitable control approach and correct tuning of each controller depending on network conditions. In general, however, local controllers do not obtain a system-wide and globally optimal objective while satisfying system constraints.

### 2-2-3 Optimization-based control methods

The other category of proposed methods to control inverter reactive power output does exactly that: obtain system-wide and globally optimal performance. Optimization-based control solves an optimal power flow problem where reactive power output of all inverters is used as optimization variables. Solving such a problem results in cooperation to achieve optimal system states, where optimal refers to a predefined goal that is expressed as an objective function. This section presents an overview of optimization based methods; where these originate from, what they pursue, how the optimization problem is formulated, and how it is efficiently solved.

An optimal power flow problem was first formulated for the transmission network in 1962 [30] to find the lowest economical cost to supply all loads in the network. Since then, optimal power flow problems have become an important part of operation and economic dispatch scheduling on the transmission level [2, Ch.7]. Moreover, an optimal power flow problem was introduced in 1989 [14, 15] for the optimal sizing and placement of capacitor banks in distribution networks. Each optimization problem finds an optimal solution for a single load profile (the demand at all nodes in a network at one time instance). In these settings, optimization problems are solved infrequently and based on a single, or a small number of power flow scenarios, but optimization based control approaches suggest solving optimal power flow problems throughout the day and use the solutions as setpoints for inverters.

Although optimal power flow problems are commonly used in the operation of the transmission network, it is challenging to implement such approaches in the distribution grid. The transmission system has far fewer nodes and larger power flow than a distribution system, which
makes it economically feasible to have a supervisory control and data acquisition (SCADA) system at transmission stations and other key points in the network. The SCADA system of the transmission network provides system operators with sufficient information to automatically reconfigure the network topology to meet the operation objectives [2, Ch.9]. This communication infrastructure makes the transmission network more observable and controllable than the distribution network. However, large scale installation of advanced metering infrastructure (AMI), or smart meters, in the U.S. and Europe is a recent initiative to increase the observability of distribution systems [42]. Another initiative to increase the observability of distribution networks is the recent development of micro synchrophasor measuring units [43]. Many distribution networks are now equipped with smart meters that measure, store, and communicate the real and reactive power demand, and voltages. However, the overarching communication infrastructure does not support real-time access to smart meter data, which is only communicated at 15 or 60 minute intervals.

**Optimal power flow**

The goal or objective function (2-9a) of an optimal power flow problem can vary. Therefore, it can be adapted to best reflect specific goals of a DSO, e.g., minimizing the cost of generation including losses, reducing voltage drop over the feeder, or minimizing voltage rise during high PV output scenarios, e.g., [4, 19, 44]. Next to finding a global minimum of the objective function, the optimization problem must adhere to the laws of physics. The effect of inverter’s reactive power injection can be described with power flow models. Therefore, the objective function is subjected to a model such as the branch flow equations as a constraint (2-9b)–(2-9e). Furthermore, the reactive power injection or consumption of an inverter is constrained by the capacity (2-9f), which in this setting is a function of the real power generation of the PV system. The responsibility of American utilities to maintain service voltage within ±5% of 120V as specified by ANSI Standard C84.1 is expressed as another constraint in the optimization problem (2-9g). Other constraints can be formulated to express other physical and safety constraints such as thermal line flow limits [32]. The overall optimization problem now has the structure of:

\[
\begin{align*}
\min_u & \quad J(u) \\
\text{s.t.} & \quad P_i = (p_i^c - p_i^g) + \sum_{j: (i,j) \in E} (P_j + r_j \ell_j) \\
& \quad Q_i = (q_i^c - q_i^g) + \sum_{j: (i,j) \in E} (Q_j + x_j \ell_j) \\
& \quad v_i = v_j + 2 (r_j P_j + x_j Q_j) + \left( r_j^2 + x_j^2 \right) \ell_j, \quad \forall j : (i, j) \in E \\
& \quad \ell_i = \frac{P_i^2 + Q_i^2}{v_i} \\
& \quad |q_i^g| \leq \sqrt{s_i^2 - (p_i^g)^2}, \quad \forall i \in N \\
& \quad \underline{v} \leq v_i \leq \bar{v}, \quad \forall i \in N \\
& \quad u = (v, P, Q, \ell, q^g).
\end{align*}
\] (2-9a)-(2-9h)

Although the formulation specifies the optimization variables to include all inverter outputs, voltages, real and reactive power flows, and currents (2-9h), in practice the goal is to find the
optimal $q^g$, while the other variables are determined by physical laws (2-9b) – (2-9e) [45]. This is possible because a unique practical solution for the power flow exists for a given substation voltage [27]. In this specific formulation, where we assume objective function (2-9a) to be convex, the nonlinear terms in (2-9e) cause the problem to be non-convex [32]. Generally, non-convex optimization problems are hard to solve in reasonable time, and it is impossible to guarantee that the global optimum is found.

**Convex relaxation**

Convex problems are better understood [46], and have a globally optimal solution that can be attained in reasonable time. Many dedicated software packages exist to obtain the global optimum of convex problems; this work uses the Convex Optimization toolbox CVX for MATLAB [47, 48]. The advantages of convex optimization problems have motivated researchers to study formulations of convex relaxation for the originally non-convex optimal power flow problem. Semidefinite programming (SDP) relaxations have been proposed for balanced [32, 44, 49], and unbalanced networks [25]. Another approach suitable for balanced networks involves second-order cone programming (SOCP), and is described by [7, 19, 32, 49]. Moreover, the branch flow model and bus injection model, and their convex relaxations are equivalent [31].

An optimal power flow problem formulated with the branch flow models (2-9) is adapted with the SOCP convex relaxation first presented in [19]. The SOCP problem is obtained by relaxing equality (2-9e) to inequality (2-10). This is equivalent to allowing the current magnitude to go above its physical equality constraint. In [19], the authors show that under reasonable conditions, the solution to the relaxed problem respects the equality constraint. Realize that the inequality can be formulated as a standard SOCP constraint (2-11).

\[
\ell_i \geq \frac{P_i^2 + Q_i^2}{v_i},
\]

\[
\ell_i^2 - \ell_i^2 + v_i^2 - v_i^2 + 4\ell_i v_i \geq 4P_i^2 + 4Q_i^2
\]

\[
(\ell_i + v_i)^2 - (\ell_i - v_i)^2 \geq (2P_i)^2 + (2Q_i)^2
\]

\[
(\ell_i + v_i)^2 \geq (2P_i)^2 + (2Q_i)^2 + (\ell_i - v_i)^2
\]

\[
2P_i \leq \ell_i + v_i.
\]

The convex relaxation is schematically depicted in Figure 2-5, where the dark blue star represents the non-convex feasibility region of an optimal power flow (OPF) problem, the light blue star represents the feasibility region of the branch flow optimal power flow problem (2-9) that uses squared voltage magnitude, and the green circle represents the convex SOCP problem. The relaxed problem is proved to be convex and exact in [19], which implies that an optimal solution $u^*$ of the green circle is also the optimum of the non-convex blue stars in Figure 2-5. Several conditions that guarantee exactness are described in literature. The first allows for over-satisfaction of loads [19], i.e., $p_i$ and $q_i$ are allowed to be larger than the actual demand. However, for non-decreasing objective functions this does not change the solution of optimal power flow problems [50]. Over-satisfaction of loads will increase losses,
20 Inverter control approaches

Figure 2-5: Feasibility regions of (relaxed) optimal power flow problems. Figure inspired by [7]

Figure 2-6: Comparison of optimization based control methods and local methods

and is therefore driven to zero by the optimization. However, if the solution leads to over-satisfaction of loads, the optimization problem is either infeasible or impractical [51]. Second, [49] proves that the solution of the SOCP problem is exact if it lies within the feasible set of the original problem and the phase angles can be recovered from \( v = |V|^2 \) and \( \ell = |I|^2 \). Sufficient conditions for when this is possible are presented in [7, 19, 45, 52]. Finally, the exactness conditions in [52] have the most realistic physical interpretation. It states that, as a rule of thumb, the solution is exact if \( v \sim 1 \) p.u., \( p, q < 1 \) p.u., \( r, x << 1 \) p.u., and \( r/x \in [0.1, 10] \), which holds for many real feeders.

2-2-4 Comparison of methods

Two classes of controllers for inverter reactive power provision have been described: local methods, and optimization based methods. Both are based on very different paradigms: local control typically relies on heuristics to infer a power flow scenario and drives system states, such as voltages, in the desired direction; optimization based control expresses the
goal as an objective function and allows an optimization to determine all inverter outputs. Strengths and weaknesses of both methods are presented in the graphic of Figure 2-6, and are elaborated next.

**Local control:**

+ Little communication infrastructure. The local control that is implemented in Germany, and is suggested by [6, 34] require inverters to support switching modes through broadcast signals. This communication infrastructure is significantly lighter and less vulnerable than the infrastructure that is required for optimization based methods.

+ Model free. Once implemented, this control method requires no network models. However, the tuning of inverters might benefit from prior simulations with feeder models.

+ Computationally light. Local controllers typically require modest computations to compute the reactive power output [4, 6].

− No measure for optimality. Local control methods generally rely on heuristics to infer a power flow scenario from available information, and react with corresponding control action [4]. However, the lack of information and lighter communication infrastructure results in sub-optimal results.

− Prone to tuning. It can be challenging to select and tune the most appropriate control scheme for a given inverter. A systematic tuning approach typically is not provided, e.g. [4, 6].

− Constraints are challenging to include. Local controllers typically cannot guarantee satisfaction of network constraints.

**Optimization based control:**

+ Globally optimal power flow solution. The method is able to find globally optimal power flow solutions, and the conditions that guarantee the global optimum are well described [25, 32, 44, 49]. Moreover, the convex relaxations are formulated for balanced and unbalanced power flow models.

+ Flexible control objective. The objective function of optimization based control can be changed to reflect the goals and priorities of DSOs.

+ Constraints are easily incorporated in the optimization problem, and given that the problem is feasible, the control will satisfy the constraints.

+ Scalability with distributed algorithms. It is possible to solve the global optimization problem with distributed algorithms. Distributed algorithms require more modest communication infrastructures, and are required to control larger networks [25, 44].

− Extensive communication infrastructure. Solving an optimal power flow problem requires information from all nodes in the network [25]. Subsequently, the optimal inverter output must be communicated to the inverters. The existing network does not
support this functionality, and required updates are expensive. Moreover, this system is potentially vulnerable to interruptions, and latency in communication and control poses a limit to the timescale on which the method can be implemented [4].

- Computationally heavy. Although dedicated software packages exist to solve convex problems, and the problem can be solved with distributed algorithms, optimization based methods are computationally heavier than local controllers.

- Model knowledge. Optimization based methods require accurate distribution feeder models, which are not necessarily available.

2.3 Concluding remarks and problem formulation

This chapter thoroughly discussed the balanced branch flow model that is used for power flow modeling in this thesis. Sequence components can be used to convert unbalanced multi-phase models to balanced one-line models. The effect of reactive power support was discussed and supported with the branch flow equations. This showed that electronic power inverters of PV systems can improve power quality and reduce system losses. Moreover, they are better suited to cope with the variability of renewable generation than traditional voltage regulators. Subsequently, two categories of inverter controllers were introduced and their potential and limitations were compared. Optimization-based methods have promising potential, but are challenging to implement on present-day distribution systems due to a lack of communication infrastructure. Conversely, local methods adhere to practical limitations, but lack a systematic design approach such that system wide objectives and constraints are met.

With the additional information that was presented in this chapter, the original research problem from Section 1-2 is reformulated. The overarching goal of this thesis is to:

Formulate a structured approach to design inverter controllers that use reactive power to maintain and enhance the efficiency and quality of electric distribution systems with high penetration of solar generation.

More specifically, the inverter reactive power control framework has the following requirements:

1. Controllers respect practical limitations of present-day distribution systems. Therefore, controllers cannot rely on communication to compute the reactive power output of an inverter.

2. Controllers must cooperate and pursue system-wide and globally optimal results, with the flexibility to represent different control objectives and respect critical system constraints.

3. Control design must be performed systematically, require little or no heuristics, and leverage available data.
Chapter 3

Regression-based inverter control

The rapid adoption of distributed and renewable generation in the distribution network diversifies power flow and increases the time variability of power and voltage, necessitating new operation paradigms for the distribution system operator (DSO). In areas with a high penetration of photovoltaic (PV) systems, the variability leads to increased switching and accelerated wear of legacy equipment and degraded power quality [17]. Electronic power inverters in PV systems are capable of providing reactive power support, but this potential is currently not utilized by American DSOs. Chapter 2-2 presented two paradigms that control voltage and power flow through inverter reactive power injection or absorption. The first, optimization-based methods, yield globally optimal power injections and incorporates network constraints, but relies on the availability of an extensive communication infrastructure, which is far from practical for most present day distribution networks. The second category of proposed methods uses local information and relies on realistic system capabilities [34], but suffers from extensive tuning. In addition, these methods yield suboptimal results and cannot guarantee satisfaction of critical system constraints. Therefore, the goal of this chapter is to propose a framework to systematically design local controllers that obtain near optimal results and incorporates system constraints. Moreover, the method has to account for practical limitations of present-day distribution systems. Regression-based inverter control is such a framework.

This chapter presents an overview of the regression-based inverter control framework in Section 3-1. Subsequently, specific aspects are discussed more rigorously. Section 3-2 presents details of the optimization problem, Section 3-3 covers the proposed regression method, and finally Section 3-4 includes suggestions for implementation in simulation- and practical testbeds.

3-1 Approach: local method inspired by optimal control

Regression-based control is inspired by the benefits of both described categories of inverter control. It applies techniques from optimization-based methods to historical data to obtain
globally optimal inverter reactive power output for representative power flow scenarios. Subsequently regression is used to understand how all inverters in a network cooperate to obtain optimal system states. The regression analysis results in models that are tailored to each individual inverter, which are employed as controllers to predict optimal inverter output based on local information only. The approach consists of four steps:

1. Obtain a balanced feeder model, and collect \( N \) historical data points of consumption and generation at each load in the network. The information can be collected with advanced metering infrastructure (AMI) over an extended period of time.

2. Solve the convex optimal power flow problem for all \( N \) scenarios. The solutions contain globally optimal reactive power output for all inverters and all \( N \) power flow scenarios.

3. Perform regression analysis for each inverter to obtain the function that most accurately predicts its optimal reactive power output based on its historical local information only.

4. Employ the functions as local inverter controllers in a simulation environment to determine the reactive power output for new measurements.

Some of the characteristics of this approach are:

+ **System wide globally optimal results** are approached by learning local controllers from globally optimal power flow data.

+ **No communication** between inverters is required to determine the reactive power output for new measurements. The only assumed communication is collection of historical measurements from AMI, and infrequent updating of the control function.

+ **Little tuning** is required by system operators. After the objective function of the optimal power flow problem and options of the regression algorithm are selected, the control functions of all inverters are automatically constructed with the data-driven approach. The structured approach is therefore easily incorporated on new feeders.

+ The **control objective is flexible** by adapting the objective function of the convex optimal power flow problem.

+ System **constraints** are easily incorporated in the optimal power flow problem.

+ **Systematic** method that leverages available data to design local controllers.

+/− **A feeder model** and the location of all loads are assumed to be available for systematic tuning of controllers. The formulation in this work assumes balanced power flow, but the framework can be extended to unbalanced power flow in future research.

### 3-2 Optimal Power Flow

The optimal power flow problem that is considered in this work finds the reactive power output for each inverter in the network that collectively minimizes a given objective. It is possible to solve the optimal power flow problem when a model and measurements of consumption and...
generation at each load of a distribution feeder are available. With a single optimization the optimal output of all inverters in a network for that load profile is found. The regression-based control framework solves the optimization problem for $N$ load profiles that represent a wide variety of historic load (and thus power flow) scenarios for that feeder. Subsequently it uses regression to express the relation between the optimal reactive power output, and local information of the inverter.

This section presents more details of the convex optimal power flow problem that is used for regression-based control. The optimization problem is based on problem (2-9) that was discussed in Section 2-2 and was originally described in [19]. First, Subsection 3-2-1 motivates the inverter control objectives used in this work. Second, Subsection 3-2-2 presents the overall optimal power flow problem. Finally, Subsection 3-2-3 suggests a matrix formulation of this optimization problem for any radial balanced power flow network, which can be solved in MATLAB using the convex optimization toolbox CVX [47, 48]. The suggested formulation in Subsection 3-2-3 is not essential in understanding regression-based control, and is mainly intended for readers who wish to implement this work.

### 3-2-1 Objective functions

In the setting of this thesis, the objective function determines what goal the inverters pursue with their reactive power output. As introduced in Section 1-1, the primary purpose of reactive power support for DSOs is to maintain voltages within range A of ANSI Standard C84.1 [1] is included as a constraint to the optimization problem (3-6a). Therefore, the objective function represents the operational goal of inverters given that voltages are in the desired range. Another concern that was introduced in Section 1-1, is the increased variability of voltage levels, which leads to increased switching and accelerated wear of legacy equipment and degraded power quality. Finally, minimizing real power losses reduces the economical cost of supplying all loads in the network. The objective function is easily adapted to reflect specific concerns for a feeder. Two objective functions that are tailored to these concerns are evaluated in this thesis. The first minimizes losses (3-1), and the second also reduces voltage variability by extending the objective with a penalty on voltage deviations from a nominal value (3-2).

\[
J_1 = \sum_{i \in E} r_i \ell_i \quad (3-1)
\]

\[
J_2 = \sum_{i \in E} r_i \ell_i + \gamma \sum_{i \in N} |v_i - v_{\text{nom}}| \quad (3-2)
\]

The branch flow equations (2-7) show that losses are minimized by supplying reactive power demand at node $i$ with reactive power from the inverter at node $i$. This reduces reactive power flow on the lines connecting node $i$ to the substation, and hence reduces losses. Similarly, if the total resistance between node $k$ and the substation is larger than the total resistance between nodes $i$ and $k$, losses are reduced when the inverter at node $i$ supplies the consumption at node $k$. However, this is only possible when the inverter has sufficient capacity. The loss minimizing objective function has additional benefits for the systems: less reactive power flow in the network also reduces the voltage drop. However, when voltages are high, they are further raised by minimizing losses. This can be seen when PV generation exceeds consumption, and
real power flows back to the substation (reverse power flow). Although, the voltage bounds are included as constraints, this objective function does not necessarily reduce variations in voltage.

Optimization with the second objective function (3-2) will be referred to as multi-objective optimization. It extends the loss minimizing objective function with a term that accounts for the difference between the voltage in the network and a nominal voltage. The variable $\gamma$ can be used to emphasize one of the two goals of the objective function; for $\gamma = 0$ the objective function reduces to loss minimization. The $\gamma$ also depends on the base values that are used to convert variables to the per unit system. When $\gamma$ is large, the optimal solution lowers the voltage magnitude during high reverse power flow scenarios, and raises voltages when the consumption is high. With more room to the voltage bounds of ANSI Standard C84.1, DSOs can use legacy equipment to optimize the overall voltage.

### 3-2-2 Overall problem

With these objective functions, the overall optimization problem has the form of (3-3), where the trade-off parameter $\gamma$ selects loss minimizing or multi-objective optimal power flow. The constraints represent the physical power flow in the network (3-3b)–(3-3d), the second-order cone programming (SOCP) relaxation (3-3e), the inverter reactive power capacity (3-3f), and the ANSI voltage constraint (3-3g). The radial network is represented as the graph $G(N, E)$ with $N$ the set of $|N| = n$ nodes and $E$ the set of edges. Furthermore, $C \subset N$ is the subset of $|C| = c$ nodes that are equipped with controllable PV inverters.

The optimization is performed over the variables in (3-3h), where the vector $v \in \mathbb{R}^n$ contains the squared voltage magnitude at each node, $v_i, \forall i \in N$. Similarly, the vectors $P, Q, L \in \mathbb{R}^n$ are defined for the variables $P_i, Q_i, \ell_i, \forall i \in N$ that were introduced in Section 2-1-1, and $q \in \mathbb{R}^c$ contains the reactive power output of all inverters $q_i^g, \forall i \in C$. Effectively, the only optimization variables are the reactive power outputs of the $c$ inverters $q$, other variables are uniquely determined by the consumption and substation voltage through the branch flow equations.

\[
\begin{align*}
\min_u & \quad \sum_{i \in N} r_i \ell_i + \gamma \sum_{i \in N} |v_i - v_{nom}| \\
\text{s.t.} & \quad P_i = (p_i^c - p_i^g) + \sum_{j: (i,j) \in E} (P_j + r_j \ell_j) \\
& \quad Q_i = (q_i^c - q_i^g) + \sum_{j: (i,j) \in E} (Q_j + x_j \ell_j) \\
& \quad v_i = v_j + 2 (r_j P_j + x_j Q_j) + \left(v_j^2 + x_j^2\right) \ell_j, \quad \forall j : (i,j) \in E \\
& \quad \ell_i \leq \frac{P_i^2 + Q_i^2}{v_i} \\
& \quad |q_i^g| \leq \sqrt{s_i^2 - (p_i^g)^2}, \quad \forall i \in C \\
& \quad \underline{v} \leq v_i \leq \overline{v}, \quad \forall i \in N \\
& \quad u = \begin{bmatrix} v & P & Q & L & q \end{bmatrix}^T 
\end{align*}
\]
3-2-3 Suggested matrix formulation for implementation

The above described iterative formulation (3-3) is reformulated to matrix (in)equalities that can be implemented in the convex optimization toolbox CVX [47, 48] for Matlab. This section is mainly intended for readers who want to implement optimal power flow problem (3-3). The optimization variable $u$ is defined as (3-4) and consists of the vectors $v, P, Q, L \in \mathbb{R}^n$ and $q^g \in \mathbb{R}^c$ that were defined for the variables $v_i, P_i, Q_i, \ell_i, \forall i \in N$ and $q^g_i, \forall i \in C$. Also define $p, q^c, r, x, z \in \mathbb{R}^n$ for the real power demand, reactive power consumption, resistance, reactance and complex impedance, $p_i, q^c_i, r_i, x_i, z_i, \forall i \in N$.

$$u = \begin{bmatrix} v & P & Q & L & q^g \end{bmatrix}^T \tag{3-4}$$

The suggested formulation for objective function (3-3a) is given by (3-5), where $r_1 = 0$ by definition. The branch flow equations (3-3b)–(3-3d) that account for the physical power flow in the network can be expressed as (3-6), where interconnection matrix $A$ represents the network connectivity, and $\circ$ denotes the Hadamard product. For a network with $n$ nodes, the $n \times n$ matrix $A$ has entries $a_{i,j} = 1$ and $a_{j,i} = -1$ when power flows from node $i$ to node $j$.

$$J = \begin{bmatrix} 0^{1\times 3n} & r^{1\times n} & 0^{1\times c} \end{bmatrix} u + \gamma \left\| \begin{bmatrix} I^{n\times n} & 0^{n\times (3n+c)} \end{bmatrix} u - v_{\text{nom}} \right\|_1 \tag{3-5}$$

$$\begin{bmatrix} C^{n\times n} & \text{diag}(2r) & \text{diag}(2x) & \text{diag}(z \circ z) & 0^{n\times c} \end{bmatrix} u = \begin{bmatrix} v_0 \end{bmatrix} \tag{3-6a}$$

$$\begin{bmatrix} 0^{n\times n} & D^{n\times n} & 0^{n\times c} & F^{n\times n} & 0^{n\times c} \end{bmatrix} u = -p \tag{3-6b}$$

$$\begin{bmatrix} 0^{n\times n} & 0^{n\times n} & D^{n\times n} & F^{n\times n} & G^{n\times c} \end{bmatrix} u = -q^c \tag{3-6c}$$

$$\begin{bmatrix} 0^{1\times 3n} & 1 & 0^{1\times (n-1+c)} \end{bmatrix} u = 0 \tag{3-6d}$$

with

$$C = I^{n\times n} + A^-$$
$$D = -I^{n\times n} + A^+$$

where $A^-$ denotes the matrix with only the negative values of interconnection matrix $A$, and similarly $A^+$ only contains its positive values. Matrix $E$ has non-zero entries $e_{i,j} = a_{i,j}^+ x_j$, whereas matrix $F$ has non-zero entries $e_{i,j} = a_{i,j}^+ x_j$. Matrix $G$ has a single non-zero entry $g_{i,j} = -1$ in each column at the rows of nodes with inverters, for example, if the first inverter is located at node 2, $g_{2,1} = -1$. Equality (3-6d) is included to constrain the current into the substation to zero, $\ell_1 = 0$, as defined by the branch flow equations.

The voltage bounds specified by ANSI standard C84.1 are expressed in constraint (2-9g), and is reformulated in matrix form (3-7). Furthermore, the reactive output of each inverter is

Master of Science Thesis
Oscar Sondermeijer
limited to its capacity by inequality constraint (2-9f) and in matrix form by (3-8). Finally, the SOCP relaxation for each node $i$, formulated as (2-11), is included as inequality (3-9).

\[
\begin{bmatrix}
I^{n \times n} & 0^{n \times (3n+c)} \\
-I^{n \times n} & 0^{n \times (3n+c)}
\end{bmatrix}
\begin{bmatrix} u \\ v \end{bmatrix} \leq \begin{bmatrix} \nu \\ \gamma \end{bmatrix} \quad (3-7)
\]

\[
\begin{bmatrix}
0^{c \times 4n} & I^{c \times c} \\
0^{c \times 4n} & -I^{c \times c}
\end{bmatrix}
\begin{bmatrix} u \\ q \end{bmatrix} \leq \begin{bmatrix} q^e \\ q^d \end{bmatrix} \quad (3-8)
\]

The notation of matrices $A$, $C$, $D$, $E$, $F$, and $G$ is illustrated for the four bus network example of Figure 3-1, where the unlabeled node is indexed 1, node $i = 2$, $j = 3$, and $k = 4$, and nodes $i$ and $j$ (or node 2 and 3) are assumed to have inverters. The figure is repeated for reference.
With the matrix notation introduced in this section, the optimal reactive power output of all inverters \( q^g \) can be found by solving the optimal power flow problem with the Convex Optimization Toolbox CVX for MATLAB [47, 48].

### 3-3 Regression framework

The goal of regression in this setting is to find a model that approximates the optimal inverter outputs based on local measurements only. The optimal inverter reactive power outputs are computed for \( N \) historical load scenarios, chosen to be a representative set of power flow scenarios. Every set of real and reactive power demand results in a unique globally optimal solution. Although it is challenging to solve the optimal power flow problem in real time, it is possible to solve the problem with historic smart-meter information. Subsequently, regression captures the relation between locally obtained measurements and the optimal output of an inverter; regression is performed for each individual inverter.

This section presents the regression framework that is introduced in Figure 3-2. First, Subsection 3-3-1 discusses the input data for the regression. Second, Subsection 3-3-2 motivates the regression method used in this work. Third, Subsection 3-3-3 presents the validation method for the regression models. The overall regression is a combination of a hypothesis and data driven approach: inputs and the regression method are chosen with hypotheses; coefficient and variable selection is performed with a data driven approach. For the interested reader, details of the stepwise selection method are included in Subsection 3-3-4.

The regression notation from [53, 54] is adopted with small alterations. Regression estimates a function \( f \) with \( n \) samples (or observations) of \( k \) input variables\(^1\) \( X_j \) and corresponding output \( y \). Therefore, with some abuse of notation, let \( x_{ij} \) represent the \( j \)th predictor, or input variable, for observation \( i \), where \( i = 1, 2, \ldots, n \) and \( j = 1, 2, \ldots, k \). Lower case normal font column vector \( x_i \in \mathbb{R}^k \) denotes all \( k \) predictors of the \( i \)th observation; lower case bold

\(^1\)This work denotes the number of predictors with \( k \) instead of \( p \) to prevent confusion with real power demand.

\(^2\)In the regression framework \( x_{ij} \) is used for an input variable, whereas the power flow section uses it to denote resistance of a line segment.
3-3-1 Input data

Regression estimates a model that best describes the relation between input and output data. In this setting, the regression output is the optimal reactive power output of inverters for historical power flow scenarios. This section discusses the regression input variables from which the optimal reactive output can be approximated. A key aspect in this work is to construct...
inverter controllers that account for limitations of the physical distribution network: there is no fast communication infrastructure available. Therefore, regression input variables are limited to local measurements from the node of the inverters. The goal of the regression analysis is to approximate the reactive power output that was generated in a convex optimization problem with local information only; the first step is to carefully choose the regression input variables.

Since the power flow is modeled and simulated with the branch flow equations, the only variables that are assumed to be "measurable", or known, in simulation at node $i$ are:

- squared voltage magnitude $v_i$
- real power generation $p_i^g$
- real power consumption $p_i^c$
- reactive power consumption $q_i^c$
- inverter capacity $s_i$.

The variables $P_i, Q_i$, and $\ell_i$ are not assumed to be available because in practice this requires measurements on the distribution line instead of measurements at the location of AMI\(^3\).

With these local measurements it is possible to construct the following four input variables for regression:

- $X_1 =$ squared voltage magnitude $v_i$.
- $X_2 =$ real power demand $p_i^c - p_i^g$
- $X_3 =$ reactive power consumption $q_i^c$
- $X_4 =$ inverter reactive power capacity $\bar{q}_i^g = \sqrt{s_i^2 - (p_i^g)^2}$

The construction of these four variables is based on physical intuition of the optimal power flow problem. First, voltage magnitude is a function of all real and reactive demand in the network, thus this information forms a link to surrounding nodes. Moreover, it is included in the objective function for multi-objective optimization. Therefore, $X_1$ is expected to have predictive power to estimate the reactive power output. Second, real power demand governs real power flow in the network, and reverse real power flow can cause high voltages. Therefore, $X_2$ is expected to have a relation with the optimal inverter output. Third, losses are minimized by local sourcing of reactive power consumption $X_3$. Finally, the reactive power output is constrained by the reactive power capacity of an inverter $X_4$. Note that the analysis in Chapter 4 assumes a constant power factor, therefore real and reactive power consumption are linearly dependent. Including both leads to singularity issues in the linear least squares approach used to find the regression coefficients.

\(^3\)The physical interpretation of a local measurement depends on the level of detail of a network model. For example, when five houses are aggregated and modeled as a single node, the notion of local refers to the voltage and demand as seen by the network. Yet, in reality there is not necessarily a measuring device present at that location.

Master of Science Thesis  
Oscar Sondermeijer
3-3-2 Regression method

Regression includes a wide range of methods, models, and approaches to best describe the relation between input and output data. This thesis uses *parametric multiple-linear models with interaction and quadratic variables*, and uses *stepwise selection* to determine which variables are included in the models. Details on the regression method and model class are presented next. The motivation for the stepwise selection algorithm is included in Subsection 3-3-4.

Regression methods generally fall in one out of two classes: parametric, or non-parametric regression [53, Ch.2]. Subclasses, however, exist, and some techniques can be formulated to fall within both [55]. Parametric regression implies that an unknown function \( \hat{f}(x, \beta) \) is characterized by a finite number of parameters \( \beta \). It involves a two-step approach: first, assume a structure that best corresponds to the data; second, train the model, or estimate the parameters \( \beta \). Conversely, non-parametric methods do not make an assumption about the structure of the model, but typically require significantly more observations to obtain an accurate estimate of the model [53]. A parametric method is preferred for two reasons. First, once the model coefficients are trained, prediction of inverter output for new measurements is straightforward. Second, parametric models are easier to interpret; although the solution of the optimal power flow problem is complex, it is desirable to gain intuition on the structure.

Regression model class

A multiple-linear relation (3-12) is one of the simplest forms of parametric regression. As the name implies, the function \( \hat{f} \) is linear in each of the predictors \( X_j \) for \( j = 1, 2, \ldots, k \). Although (3-12) is linear in each variable \( X_j \), the variables themselves can be more complex terms, such as interaction or quadratic terms. Interaction terms are constructed by the product of two other variables, for example, \( X_5 = X_1 X_2 \). Similarly, an example of a quadratic variable is \( X_6 = X_2^2 \). It is impossible to have an accurate estimation of \( f \) when the assumed model of parametric regression \( \hat{f} \) is too different from the real system; the solution of the underlying optimal power flow problem is not linear in the four variables \( X_1 \sim X_4 \), but the model becomes more flexible by including the higher order variables.

\[
\hat{f}(\beta, x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_K
\]  

(3-12)

Stepwise selection

A hybrid forward- and backward-stepwise selection method selects a subset of input variables [54]. The algorithm is initialized with a multiple linear model of the base variables only. At each iteration, the variable that improves model the most, and sufficiently, is added to the model. Subsequently, the variable with the lowest, and sufficiently little, contribution is removed. These two steps are iterated until no variables meet the entrance or exit threshold of the algorithm. Traditionally, stepwise selection adds and removes variables from the model, based on its significance expressed as F-statistic [54, Ch.3]. Other commonly used criteria are the adjusted \( R^2 \) value, and information theoretic approaches Akaike information criterion (AIC) [53, Ch.6] and Bayesian information criterion (BIC) [53, Ch.6]. More details of the selection criteria are included in Subsection 3-3-4.
The hybrid forward- and backward-stepwise selection algorithm included in the Matlab Statistics and Machine Learning Toolbox is given by:

1. Fit the initial model.

2. If any terms that are not in the model have criteria that meet the entrance tolerance, add the one with the best value and repeat this step; otherwise, go to step 3.

3. If any terms in the model have criteria that meet the exit tolerance, remove the one with the worst value and go to step 2; otherwise, end.

Although the goal of the stepwise selection algorithm is to select the subset of variables that yields the lowest prediction accuracy for new measurements, the globally best subset is not guaranteed to be found. The method depends on the initialization and the order in which terms are added or removed. The results of the case studies in Chapter 4 include an analysis on final models for different initializations and criteria.

**Regression coefficients**

The regression coefficients $\beta$ are found by solving a least-squares problem. When a column of ones is added from the left to the input data matrix $X$, the multiple linear model (3-12) can be written as (3-13), with error $\epsilon$. The error of one observation denotes the difference between the true output $y_i$ and the predicted output $\hat{y}_i = x_i \beta$. Least-squares finds the coefficients $\beta$ that minimize the sum of squared errors $\|X \beta - y\|^2$ [56]. The textbook version of the solution is given by (3-14).

\[
\begin{align*}
\mathbf{y} &= \mathbf{X}\beta + \epsilon \quad (3-13) \\
\beta &= \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{y} \quad (3-14)
\end{align*}
\]

The obtained regression model can be interpreted by its coefficients. The coefficient $\beta_j$ is a measure for how much the output changes when input variable $X_j$ changes one unit. Units of two variables can greatly differ. For example, voltage in p.u. has to be between 0.95–1.05, whereas real power demand can vary between several kilowatts of generation or consumption. Standardizing regression data scales it to have zero mean, $\mu = 0$, and unity variance, $\sigma^2 = 1$. Removing the difference in units from a model gives a different interpretation to the coefficients; a coefficient represents how much the output changes when a variable changes one standard deviation [53]. This interpretation is a sensitivity analysis for additive models. It can, however, not deduce how much the output changes for a change in the base variables $X_1$–$X_4$ because these are also included in the higher order terms, hence the model is no longer additive. Although the regression model does not quantify the output’s sensitivity to the base variables, it does present qualitative insight into the structure of the optimal power flow solutions.
3-3-3 Validation of obtained models

The goal of this regression is to accurately predict new outputs, instead of accurately predicting training data. Yet, the model is optimized to the training set. Therefore, the expected true prediction error will always be higher than the expected prediction error of the training data. A measure for how much worse a model performs on new data can be called training optimism [54, Ch.7], see Figure 3-3. Validation of the regression results is needed to verify the training data results, because the prediction error of this data does not contain training optimism. This work uses 70% of a data set for training and 30% of the data for validation unless mentioned otherwise; data is randomly assigned to one of the two sets. Furthermore, regression in this work is always performed to standardized data. The prediction error of the validation data is expressed as its $R^2$ value (3-15), which is based on the residual sum of squares (RSS) of a model, and the RSS of a null model, where the null model is a constant value (mean of $y$). If the sum of squared errors of a model is equal to the result of simply taking the average of the training data, the $R^2$ value is 0. Conversely, if it perfectly predicts the data, the errors are zero, and the $R^2$ value is 1.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \frac{\sum_{i=1}^{n} y_i}{n})^2} \quad (3-15)$$

Regression results are extensively validated in Chapter 4-3.

---

The model in Figure 3-3 is purely hypothetical and serves as illustration only.

Oscar Sondermeijer
Master of Science Thesis
3-3-4 Details of stepwise selection

This section is mainly intended for readers who wish to better understand why a subset of variables is selected, why this thesis uses the stepwise selection algorithm, and what the different selection criteria are.

Relevance of subset selection

Although the training error always decreases for higher complexity, the prediction error for new measurements might actually increase. Training errors always decrease as more variables are included in the model because least squares minimizes the sum of squared errors (or RSS) for training data. The claim is best illustrated with a hypothetical example of two models, \( A \) and \( B \), where model \( A \) contains all the predictors of the model \( B \), and has one additional predictor. Model \( A \) always fits the training data at least as good as model \( B \), because the structures are identical if the extra information has no predictive power, and the corresponding coefficient is assigned a zero value. Simultaneously, the prediction error for new data might increase. This is often referred to as overfitting, and is seen when models that are too flexible are optimized to the training data and lose the ability to describe the general trend, see the upper right plot of Figure 3-3. For this application, it is known that the optimal reactive power output is not obtained with a multiple-linear model, but is the solution of a system wide optimization problem. Therefore, the regression model must best describe the overall structure between local measurements and the optimal reactive power output.

With four base variables, their four quadratic variables, and six possible interaction variables, a total of 14 variables are constructed. Although this presents a lot more flexibility to a multiple-linear model than the four base variables, it is likely that not all variables are meaningful to predict the output. With four base variables, the model potentially underfits the data because the assumed model might not sufficiently represent the relation between input and output data. On the contrary, a model that includes all 14 variables potentially overfits the training data and might perform poorly for new measurements. Two measures are performed to prevent this. First, the regression is trained with sufficient data to prevent optimizing to individual observations. Second, stepwise selection is used to choose a subset of variables that accurately predict the optimal reactive power output for new measurements.

Preventing overfitting

Selecting the variables that best describe the structure between input and output data is achieved with stepwise selection. In general there are three methods to prevent overfitting. A brief discussion of these methods follows and is based on [53, Ch.6] and [54, Ch.3], to which the interested reader is referred for more detail.

- Subset selection finds the subset of predictors that result in the most accurate model. With 14 variables, a total of \( 2^{14} = 16384 \) different models can be constructed; although perhaps feasible, it is not desirable to train and validate all of possible models. Two structured approaches are: best subset selection, and stepwise selection. Although both
methods can be applied to 14 variables, stepwise selection is more efficient when the number of variables increases.

- *Shrinkage* penalizes the sum of the coefficients, which shrinks the size of coefficients with respective to coefficients that are obtained with least squares. One popular shrinkage technique is least absolute shrinkage and selection operator (lasso) [57], which penalizes the \( L_1 \) norm of the coefficients. As a result, some coefficients become zero, and this technique thus selects variables. Lasso regression is similar to best subset selection, but computationally more feasible [53, Ch.6]. Lasso requires tuning of the shrinkage penalty for each regression that is applied. Therefore stepwise selection is preferred because tuning the algorithm is more straightforward which is an advantage if the method has to be repeated for many inverters and distribution networks.

- *Dimension reduction* applies a transformation to the variables, and performs least squares on a lower dimensional space. A transformed variable is constructed as linear combinations of the original variables. However, this potentially makes the regression models harder to interpret because the four variables \( X_1 \)–\( X_4 \) are carefully selected, and the ten other variables are already constructed from these.

**Stepwise selection criteria**

A brief overview of the four criteria follows, and is based on [53, Ch.3] [8]. The F-statistic tests the null hypothesis that variable \( X_j \) has a zero coefficient versus the alternative hypothesis that the coefficient is not zero:

\[
H_0 : \beta_j = 0 \\
H_1 : \beta_j \neq 0.
\]

The \( p \) value is the outcome of the test, and is the probability that the null hypothesis is true. A significance level of 5\%, or a \( p \) value 0.05, is commonly used.

The adjusted \( R^2 \), AIC, and BIC [58] criteria adjust the training prediction error to account for training optimism. As the name suggests, the adjusted-\( R^2 \) value is a direct extension to the \( R^2 \) value and is expressed as (3-16). AIC (3-17) and BIC (3-18) are information theoretic techniques to measure how much information is lost between an estimated parametric model and the true model. The formulation in this thesis is specific for fitted least squares models [53, Ch.6], but the AIC and BIC are also commonly used for fitted maximum likelihood models. Both consist of a term for the training error expressed as RSS, and a penalty for training optimism based on \( k \), the number of variables of the model. Additionally, BIC also depends on the number of observations \( n \); the penalty for model complexity of BIC is larger for \( n > 7 \), and generally results in smaller estimated models.

\[
\text{Adjusted}-R^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1} 
\]

\[
\text{AIC} = \frac{1}{n\hat{\sigma}^2} \left( \text{RSS} + 2k\hat{\sigma}^2 \right) 
\]

\[
\text{BIC} = \frac{1}{n} \left( \text{RSS} + \log (n) k\hat{\sigma}^2 \right) 
\]
Because these criteria account for training optimism, the entrance criteria is simply: an improvement of the criterion value. But, in general it is desirable to have less complex models, therefore the exit criteria allows for a minor decline in criterion value. The standard exit criteria values of the Matlab Statistics and Machine Learning Toolbox are applied, an increase of 0.01 for AIC and BIC and a decrease of 0.05 of the adjusted-$R^2$ value.

3-4 Implementation of controllers

The two key components of the proposed method, optimization and regression, were described in Section 3-2 and Section 3-3. Under specific assumptions, it is possible to simulate the proposed method with realistic data. First, Subsection 3-4-1 includes suggestions for implementation in a simulation setting, including an overview of the made assumptions. Second, Subsection 3-4-2 discusses the challenge of using voltage information with the regression approach. Finally, Subsection 3-4-3 discusses considerations for implementation of this framework on a physical testbed.

3-4-1 Implementation in Simulation

The regression models are employed as local inverter controllers to evaluate the effect of regression-based control on the power flow, and evaluate the difference with respect to the optimal solutions. Once the regression model structure is selected and coefficients are estimated, it is straightforward to predict an output for new input measurements through (3-12). Using voltage information is challenging, as will be elaborated in Subsection 3-4-2. Therefore, let us assume that regression is performed with the three remaining base variables $X_2$–$X_4$ only. These variables are independent of the power flow in the network, and can be constructed prior to simulating the power flow in a network. Subsequently, power flow in the network can be simulated with inverter’s reactive power output determined by the regression models. Appendix B presents the detailed steps of implementing regression-based control in simulation.

When network variables (e.g., voltage) are used as input variables to the regression models, the regression-based control is suggested to be implemented in an iterative fashion, e.g.:

1. Initialize inverter output from regression model without network variables (voltage).
2. Simulate power flow with inverter reactive power output from step 1.
3. Update inverter output with regression model including network variables from step 2.
4. Repeat steps 2 and 3 until power flow solution and inverter output converges.

The overall assumptions of this work are listed below, when additional assumptions are made for the case studies they are explicitly mentioned in the text.

- Power flow is in equilibrium at all times, i.e., transient effects are neglected.
- Power flow is balanced.
• Network topology is radial and known.
• Line impedances are known.
• Single positive sequence impedance values can represent multi-phase impedances through (2-5).
• Historical generation data is available for PV systems in the network.
• Historical real and reactive consumption is available for all loads in the network.
• Loads are not voltage sensitive.
• Inverter reactive power capacity is given by (2-8).
• When inverter capacity is not specified, an overcapacity of 5% is assumed.

3-4-2 Challenge of voltage information

Two challenges of voltage information arise in the described approach. First, in the case of the multi-objective optimization (3-2), voltage information saved from the optimization solution is always close to the reference value. As a result, the predictive power of voltage information is low. The second and more important challenge is that voltage is a network variable, which is also a function of the reactive power output of other inverters. Both are discussed in more detail.

Optimal voltage

As a consequence of penalizing deviation from a nominal voltage, the optimal reactive power output of all inverters drives the voltage, or optimal voltage, to the nominal value. Nonetheless, intuitively voltage information is strongly related to the magnitude of reactive power output: low voltages require more reactive power than voltages that are closer to the nominal value. However, this intuition refers not to optimal voltage, but to a voltage that is observed without reactive power support of an inverter, or reference voltage. A reference voltage can be thought of in two ways: voltage without any reactive power support in the network; or voltage without the reactive power support of a single inverter, while other inverters are operational.

A comparison of optimal voltage and reference voltage where a single inverter provides no reactive power support is included in Figure 3-4. The data in the figure is standardized and constructed in a case study on the IEEE 13 node feeder with consumption and PV data from Pecan Street during July 2014.

In simulations, the first notion of reference voltage is obtained by solving the power flow equations without reactive power provision of inverters. In practice, this is challenging for the same reason that optimization-based control is challenging to implement: a lack of extensive and fast communication infrastructure. The only option (to the author’s knowledge) to measure this reference voltage is by turning off all inverters simultaneously at each time step that inverter output is computed. Although perhaps technically feasible, the accompanying large fluctuations in power flow and voltage are undesirable. A more feasible alternative is the second concept of reference voltage.
The second concept of a reference voltage refers to the voltage that is measured by an inverter if that inverter is turned off, while the rest of the network is in normal operation. Another way to think of this concept is: estimating the effect of inverter reactive power support on the voltage that it measures. Based on the branch flow equations, an analytical expression for this effect is derived in Appendix A, and results in (3-20), where \( v_{i,y} \) and \( v_{i,n} \) denote the voltage with and without reactive power support respectively. Then, \( v_{i,y} - v_{i,n} \) denotes the difference in the square of voltage magnitude at node \( i \) caused by the generation of reactive power \( q_i^g \), and \( \mathcal{L} \) is the set of nodes that are between the substation and node \( i \). The relation states that the difference in squared voltage is the product of reactive power injection and twice the total reactance between the substation and the inverter. The derivation is obtained separately but the result shows resemblance to a formulation in [29].

\[
v_{i,y} - v_{i,n} = 2 \sum_{j \in \mathcal{L}} (x_j) q_i^g
\]  
(3-20)

**Voltage as network variable**

The second challenge is that voltage is a network variable. That is, voltage at neighboring nodes are related to each other through laws of physics. Contrarily, real and reactive power consumption, and PV generation are not coupled\(^5\) with other inverters because they are assumed to be independent of voltage. Although voltage is not the only variable that holds information about other nodes, it is the easiest to measure. Intuitively, this makes voltage information extremely relevant to a local controller. However, for regression-based control it comes with a challenge.

The training data of regression models is obtained from optimal power flow problems; therefore, the voltage information captures the effect of optimal reactive power output of all inverters. If regression-based control is active and relies on voltage information, this information no longer captures the effect of optimal reactive power output, but an approximation thereof. Although the resulting difference in voltages might seem small, it can result in larger differences in regression output of all inverters and an unstable closed-loop system. This effect is analyzed in Chapter 4-3.

\(^5\)These variables are unrelated to other nodes in power flow equations, but in practice can be statistically related to other nodes, e.g. the effect of clouds on PV generation is likely to be observed at neighboring nodes in a network.

---

3-4 Implementation of controllers

---

Master of Science Thesis  
Oscar Sondermeijer
3-4-3 Considerations for implementation on physical testbed

A key requirement of regression-based control is that it accounts for practical limitations. Therefore, this subsection discusses some considerations for implementation on a physical testbed.

- Representative historic real and reactive power demand measurements can be collected with AMI (smart meters). These devices typically do not communicate the information in real time, but at regular time intervals. This poses no problem to this method, because the optimal power flow problems are solved offline with historic data.

- Residential PV generation measurements are not necessarily available to DSOs (these typically measure net demand). However, modern PV systems generally have advanced monitoring capabilities [39]. Thus, obtaining generation information is technically feasible, but practical concerns should be carefully considered.

- Feeder models in the possession of DSOs contain more details that the one-line balanced power flow models that were discussed in this chapter. Moreover, the power flow in a distribution feeder can be significantly unbalanced [23]. Therefore, it is recommended to extend this method to unbalanced optimal power flow problems, such as described in [25]. This recommendation is further discussed in Chapter 5-2.

- DSOs need to solve optimal power flow problems with historical data and perform regression to construct local inverter controllers. The practical requirements to perform this efficiently have not been considered in this work, but aspects that come to mind are specialized software to solve the optimal power flow problems, and sufficient computing resources.

- The obtained multiple-linear regression model structure and coefficients have to be communicated to inverters. Initially this can be envisioned as a firmware update, but updating of coefficients is comparable to the requirements described in [34].

- The time scale at which inverters update their reactive power output must be selected by DSOs. It is advised to perform optimization and regression with data from the same time scale.

3-5 Concluding remarks

The regression-based inverter control framework was presented in this Chapter. As Figure 2-6 shows, it is clearly inspired by both optimization-based control methods, as well as local control methods. Regression-based control uses available data and existing convex optimal power flow methods to provide input and output data to train local controllers. The local control functions are obtained with regression, and are optimized to predict the reactive power output that minimizes a system wide objective. The framework is a systematic and data driven approach to design local inverter controllers that account for system constraints and the practical limitations of present-day distribution systems.
This chapter evaluates the proposed regression-based method on realistic virtual testbeds that are constructed from two independent sources: distribution feeder models based on real networks, and residential consumption and photovoltaic (PV) generation measurements from Pecan Street [59]. The goal of these analyses is to gain better insight in the performance and potential limitations of regression-based control. The organization of this chapter is graphically depicted in Figure 4-1. Section 4-1 introduces the two testbeds: their network characteristics, the consumption data, and the PV generation data. Section 4-2 presents the effect of optimal power flow on both testbeds. Section 4-3 studies the prediction accuracy and structure of several regression models that are constructed with optimal power flow results. Section 4-4 presents the effect of regression-based control on the two testbeds, and compares this with optimal power flow and two other local controllers. Finally, Section 4-5 summarizes the chapter with concluding remarks.
4-1 Testbeds

This section presents the testbeds on which regression-based control is analyzed. The goal is to evaluate the proposed method in simulations that reflect realistic settings. To that end, the two testbeds are carefully constructed from a combination of real U.S. distribution feeder models, and historic consumption and generation measurements from residential customers in Austin, Texas. Constructing a realistic testbed requires the combination of two independent sources because the feeder models include the demand at all nodes for a single time instance only. By populating the network with residential demand data from Pecan Street [59], a network is constructed for which the consumption and PV generation is known between February–December 2014\(^1\). Characteristics of the two distribution feeder models are discussed in Section 4-1-1; details of consumption and generation data are included in Sections 4-1-2.

4-1-1 Distribution feeder models

The original feeder models include the network topology, multi-phase line models, and a sample load profile. Furthermore, the networks contain switches, shunt capacitors, transformers, and protection equipment such as circuit breakers, reclosers and fuses. Multi-phase line models are expressed as single impedance values with the sequence component method described in Subsection 2-1-1. Closed switches are included as two nodes, connected with a zero impedance line segment, which are replaced with small (but non-zero) values for modeling purposes. Open switches are simply modeled as the absence of a line; shunt capacitors are modeled as constant reactive power generators, or negative consumption. Finally, transformers are neglected, and voltage is modeled in the per unit system.

The first distribution feeder is the publicly available 13 node IEEE Distribution Test Feeder, that was originally published to evaluate the accuracy of three-phase unbalanced power flow simulation algorithms [9]. This feeder is referred to as IEEE feeder. The second feeder is referred to as SB feeder, and is based on a real 127 node distribution feeder in Arizona; contrary to the first, it is not publicly available. Schematic representations of both feeders are presented in Figure 4-2, characteristics are included in Table 4-1, and more detailed model data is included in Appendix C.

IEEE feeder

The IEEE feeder, depicted in Figure 4-2a, is a reduced-order model of an actual distribution network operating at 4.16 kV. Although it has a relatively small number of nodes, with only 9 loads, the total sample load is 3822 kVA. The feeder has two large loads at node 671 and 675, of 1446 kVA and of 961 kVA, while the seven other loads are between 115–494 kVA with an average of 236 kVA. This work assumes that all consumption in a network comes from residences, which implies that the smallest node is comparable to the aggregation of 12 residences\(^2\), and the largest node can be thought of as the simultaneous peak consumption of

\(^1\)The time period used in this thesis, more Pecan Street data is available.

\(^2\)Average peak consumption in July 2014 of a residential load, that has a PV system, from the Pecan Street database is 8.4 kW. With a power factor of 0.9, the peak apparent power is 9.3 kVA. This number does not account for the fact that the peak of several houses does not necessarily coincide in time.
Figure 4-2: Diagrams of the (a) IEEE feeder [9] and (b) SB feeder.

Table 4-1: Feeder characteristics.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>IEEE feeder</th>
<th>SB feeder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>13</td>
<td>129</td>
</tr>
<tr>
<td>Number of loads</td>
<td>9</td>
<td>53</td>
</tr>
<tr>
<td>Number of inverters</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>Total sample load</td>
<td>3822 kVA</td>
<td>1230 kVA</td>
</tr>
<tr>
<td>Individual load size</td>
<td>115–1446 kVA</td>
<td>4–301 kVA</td>
</tr>
<tr>
<td>Voltage level</td>
<td>4.16 kV</td>
<td>0.70–16.30 kV</td>
</tr>
<tr>
<td>Average line length</td>
<td>820 feet</td>
<td>132 feet</td>
</tr>
<tr>
<td>Average line impedance</td>
<td>$0.51 + j0.65 \Omega$/mile.</td>
<td>$1.97 + j1.20 \Omega$/mile.</td>
</tr>
<tr>
<td></td>
<td>$0.08 + j0.10 \Omega$</td>
<td>$0.05 + j0.03 \Omega$</td>
</tr>
<tr>
<td>Average $r/x$ ratio</td>
<td>0.78</td>
<td>1.65</td>
</tr>
<tr>
<td>Number of optimization variables</td>
<td>56</td>
<td>543</td>
</tr>
</tbody>
</table>
155 houses. Therefore, this feeder model represents a real distribution feeder on an aggregated level.

The feeder model neglects the most granular part of the distribution network that connects to individual houses. Nonetheless, this network is deemed relevant because the majority of voltage drop and power loss occurs on this level. If voltage magnitude at the nodes of this network are in the ANSI specified range of 0.95–1.05V p.u., it is likely that the voltage at individual houses of the real feeder are also in the desired range. Furthermore, the manageable number of nodes are useful for gaining insight of the method because it simplifies intuition, analysis, and presentation of inverter results.

The interpretation of inverter control for this network deserves clarification. A total of four nodes are selected to have PV systems, and the proposed control method uses total consumption and generation at a node to compute the desired reactive power generation for that node. The translation to individual houses and inverters is not considered for this testbed, but the second model represents a network with more detail.

**SB feeder**

The SB model is based on a real 129 node distribution feeder in Arizona, Figure 4-2b. Again, a single set of sample data is included, with a total consumption of 1230 kVA, which is divided over 54 loads ranging between 4–301 kVA. Clearly, the loads are significantly smaller than the loads of the IEEE feeder; hence, they represent the aggregation of fewer individual residences. Combined with the larger number of nodes, it implies that this model includes more details of the feeder, e.g., the lines up to a service transformer which is connected to a single or several houses. A total of 27 nodes of the SB feeder are randomly selected to have PV systems. The interpretation of inverter control is more realistic because most nodes correspond to a single or a small number of residences. Evaluation of regression-based control on this testbed also tests the scalability to larger networks.

Note that the voltage level of the SB feeder in Table 4-1 ranges between 0.70–16.30kV. This analysis neglects transformers, and uses a single base voltage to obtain voltage in the per unit system. Initial power flow solutions with varying voltage levels indicated that the voltage drop in the system is small\(^3\) For demonstration purposes, the base voltage is set to 2.77 kV, and impedances are increased with by factor 2.5, which results in a small violation of the lower voltage bound during a normal evening peak.

### 4-1-2 Consumption and PV generation data

The sample load profile accompanying the models consists of real and reactive demand for each node, which are assumed to be peak values. The sample load profile has a single power flow solution, and it is possible to solve a single optimal power flow problem. However, the proposed approach requires sufficient optimal power flow solutions to train the regression models. Moreover, the models contain no information about distributed generation (from PV systems). Because more load and PV information of the feeders is not available, the feeders

\(^3\)If the voltage drop (or rise) in the system is small, the line loss is also low, and the effect of reactive power support is small.
are populated with residential consumption and PV generation data from Pecan Street [59].

This subsection first discusses the period of time over which consumption and generation data is used. Subsequently, the approach to construct consumption and generation profiles is presented.

The Pecan Street data that is used in this work is measured at 113 individual residences in Austin, Texas. All houses are equipped with PV systems; real power consumption and generation data are measured separately. This information is available at 1 and 15 minute time intervals, over which the consumption and generation are averaged. As a result, 1 minute data is more variable than 15 minute data. Consequently, the average correlation between two residences is larger for 15 minute interval data than for 1 minute interval data.

**Timespan**

Pecan Street data of the selected houses (or IDs) is available for February–December 2014 in 15 minute time intervals, and during July 2014 in 1 minute time intervals. It is infeasible to visually evaluate which timespan contains the most interesting or relevant consumption and generation data, because there are simply too many data sets. Moreover, it is unclear which demand scenarios result in the most interesting and relevant optimal power flow solutions. One month of 15 minute interval data is assumed to contain sufficient (relevant) load scenarios for regression. However, the demand characteristics in February can differ significantly from the demand in e.g., July. Figure 4-3 illustrates this difference in consumption and generation data from the same residence in February and July. The consumption in July shows a daily pattern that is likely caused by an electric air-conditioning unit. Furthermore, generation in February clearly shows the difference between a sunny day on 13/02, and a completely clouded day on 11/02.
Results and analysis of two case studies

An experiment is conducted to evaluate differences in the consumption and generation data of the months February–December 2014. Regression models are trained with optimal power flow results from one month, and cross-validated with optimal power flow results from other months. For completeness, the regression models are also trained and validated with data from the same month based on a holdout-set of 30%. The optimal power flow results are obtained for the IEEE feeder model with four inverters. The prediction accuracy of the validation data of the regression is reported as the average of $R^2$ values of four inverters in Table 4-2. In this experiment, contrary to the stepwise selection technique, the structure of regression models are fixed and coefficients are trained with least squares. The fixed model structures are constructed beforehand with stepwise selection and training data from all months; therefore, the model structure is optimized to all data instead of to a single month. The results in Table 4-2 show that:

- On average, each month of Pecan Street data contains sufficiently representative load profiles to train accurate regression models.
- All models that are trained and validated with data from the same month (70% and 30% of the data respectively) have $R^2$ values of 0.97.
- Surprisingly, the largest difference in data characteristics is observed between March and April.
- The lowest $R^2$ value is reported for models that are trained with optimal power flow data from March and validated with data from April.

The remainder of this work uses optimal power flow data from the month July 2014 for analyses.

Consumption data

Especially for the IEEE feeder, demand of the accompanying sample data is several orders of magnitude larger than the demand of a single residence from the Pecan Street database. Using the data of a single residence results in unrealistic power flow scenarios; therefore, data from multiple residences, or IDs, are aggregated until the coinciding peak consumption equals the sample load of a node. Aggregation is performed by randomly selecting an ID from the set of 131 IDs that have a complete data set of 15 minute interval measurements in July 2014.

Approximately 330 consumption profiles are aggregated to match the largest demand of the IEEE feeder sample data. This is only possible by using IDs multiple times. Although the composition of each aggregated load profile is unique, they are constructed with the same data. Combined with aggregation, this results in highly correlated consumption profiles. Load correlation is relevant to the prediction error of regression models because these use local information to predict a globally optimal reactive power output. High load correlation

---

4Originally, 137 IDs in Austin with PV systems were enrolled in the Pecan Street program between February 2014 and May 2015, when the list was made. However, six IDs have missing entries in the data of July 2014, and are discarded from the analysis.

Oscar Sondermeijer

Master of Science Thesis
Table 4-2: Average validation $R^2$ values of four IEEE feeder inverter regression models with fixed model structure for different training and validation data.

<table>
<thead>
<tr>
<th></th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>February</td>
<td>0.97</td>
<td>0.96</td>
<td>0.86</td>
<td>0.94</td>
<td>0.95</td>
<td>0.94</td>
<td>0.93</td>
<td>0.92</td>
<td>0.96</td>
<td>0.97</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>March</td>
<td>0.97</td>
<td>0.97</td>
<td>0.79</td>
<td>0.92</td>
<td>0.96</td>
<td>0.94</td>
<td>0.92</td>
<td>0.89</td>
<td>0.96</td>
<td>0.97</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>April</td>
<td>0.95</td>
<td>0.94</td>
<td>0.97</td>
<td>0.95</td>
<td>0.91</td>
<td>0.83</td>
<td>0.89</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>May</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>June</td>
<td>0.96</td>
<td>0.95</td>
<td>0.92</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>0.94</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>July</td>
<td>0.94</td>
<td>0.93</td>
<td>0.91</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>August</td>
<td>0.96</td>
<td>0.95</td>
<td>0.90</td>
<td>0.95</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>September</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>October</td>
<td>0.95</td>
<td>0.94</td>
<td>0.96</td>
<td>0.97</td>
<td>0.95</td>
<td>0.94</td>
<td>0.93</td>
<td>0.95</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>November</td>
<td>0.96</td>
<td>0.96</td>
<td>0.88</td>
<td>0.94</td>
<td>0.94</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
<td>0.97</td>
<td>0.97</td>
<td>0.94</td>
</tr>
<tr>
<td>December</td>
<td>0.96</td>
<td>0.96</td>
<td>0.83</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
<td>0.92</td>
<td>0.96</td>
<td>0.97</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Figure 4-4: Average load correlation for IEEE feeder for varying IDs per load.

implies that local information is representative for other nodes in the network; the effect of load correlation on the results of regression-based control is further discussed in Section 4-3.

The correlation is reduced by dividing IDs in smaller sets from which a single load profile is constructed. Hence, IDs are not used in multiple load profiles. Correlation is further reduced by constructing load profiles from smaller sets of IDs; the lowest correlation is obtained by scaling the profile of a single residence. In practice, however, loads of this size likely consist of smaller loads, which averages the unique characteristics of a single profile. The effect of reducing the number of IDs from which a load is constructed on the average correlation of 15 minute interval consumption profiles of the IEEE feeder is presented in Figure 4-4. The SB feeder has far more loads in the network, which are significantly smaller than the loads of the IEEE feeder. Therefore, the load correlation is only 0.49 without dividing the 131 IDs in separate sets for the 53 loads of the SB feeder.

Another aspect of constructing realistic loads from Pecan Street data regards reactive power.
Original Pecan Street data contains real power measurements only. Reactive power consumption is related to real power $P$ and apparent power $S$ as (4-1) and (4-2), and can be constructed from real power data by assuming a power factor. The power factor is assumed to be constant: for the IEEE feeder it is assumed to be 0.9 for all loads; for the SB feeder the power factor of each load in the sample data is used.

$$\text{p.f.} = \cos(\alpha) = \frac{P}{S} \quad (4-1)$$

$$Q = P \tan(\alpha) \quad (4-2)$$

**Generation data**

The feeder models do not include information about the presence of PV systems in the network. Therefore, three choices have to be made:

- Number of PV systems.
- Location of PV systems.
- Size of PV systems.

For a thorough evaluation of the proposed method, these three aspects are selected with a different approach for the IEEE feeder and for the SB feeder. Pecan Street data is measured in the same geographical area. Generation data is highly correlated because all residences experienced the same weather conditions, which is realistic for a distribution feeder.

In IEEE feeder analyses, four PV systems are assumed to be present at nodes 634, 645, 671, and 675 of Figure 4-2a, these locations are manually selected; the selection considerations are included in Table 4-3. The relative size of PV systems with respect to consumption is denoted as $PV$ penetration (4-3) [60], where peaks are non-coincident. At each node generation profiles are aggregated until PV penetration of 80% at that node is reached. As a result the overall PV penetration in the network is 60.8%.

For the SB feeder, the locations of PV systems, and corresponding inverters, are randomly selected. A little over 50% (27 out of 53) of the loads are equipped with PV systems. In practice, a utility has little or no control over the location of PV systems in a network, and the locations can be considered random. Moreover, PV generation is constructed with the same IDs that are aggregated for the consumption data, i.e., if the consumption of a node is constructed from ID 1, ID 2, and ID 3, the PV generation profile of that node is constructed with generation data of the same three IDs. This results in an overall PV penetration on the SB feeder of 61.7%. The larger feeder is considered to be the most realistic testbed.

$$PV \text{ penetration} = \frac{\text{peak PV generation}}{\text{peak apparent power consumption}} = \frac{P^g}{S^c} \quad (4-3)$$

**4-1-3 Concluding remark**

This section presented the two realistic testbeds on which regression-based control is analyzed. The IEEE feeder testbed represents a reduced-order model of a real distribution feeder, which
results in large –aggregated– loads. This feeder is mainly intended for gaining insight into the regression models. The SB feeder contains far more nodes and smaller loads; it is a more realistic representation of a distribution feeder. Therefore, the SB feeder is mainly used to test the performance of regression-based control. Both feeders are populated with real consumption and generation data from residential loads, the measurements are gathered by Pecan Street [59].

4-2 Optimal power flow results

A key aspect of the problem formulation is close-to-optimal performance of the control approach. Section 3-2 introduced and motivated two objective functions: minimizing real power losses; simultaneously reducing real power losses and maintaining voltages close to a reference value\(^5\). Therefore, this section presents the effect of optimal power flow on both feeder models. Subsection 4-2-1 presents and compares the effect of both optimal control approaches on the power flow of the SB feeder during 4 July 2014. Subsequently, Subsection 4-2-2 discusses the effect of optimal reactive power support on the IEEE feeder, and compares it to the results of the SB feeder. The three control approaches of this section are referred to as:

A) No inverter reactive power control: results are presented in red.

B) Loss minimizing optimal control: results are presented in yellow.

C) Multi-objective optimal control: results are presented in blue.

4-2-1 SB feeder optimal power flow

The consumption and generation at two nodes with inverters is presented in Figure 4-5. Figure 4-5a shows the demand at node 83, which is constructed with data from 76 residences, whereas Figure 4-5b shows the demand of a single residence, which is used at node 42. The effect of aggregation is clearly observable from the figures: especially the consumption data of a single residence is volatile, while the relative fluctuation of aggregated data is much smaller. Figure 4-5b shows how volatile 15 minute data of a single residence can be; it is hard to see a general trend in this data.

Figure 4-5a is representative for the overall demand characteristics on 4 July 2014. PV systems start generating power around 08:00, have a sudden and temporary sag in generation

---

\(^5\)The value of \(\gamma\) is 0.0026/13 for the IEEE feeder and 0.001/129 for the Hoover feeder
Results and analysis of two case studies

of approximately 60% at 14:00, followed by peak generation, and a gradual decay in generation towards 20:00. The nearly instantaneous drop in generation is likely caused by a passing cloud. Consumption is fairly steady until 08:00, after which it gradually increases to the maximum around 18:00. The consumption remains high until 23:00, and declines to nighttime base consumption.

A comparison of the objective function values for control approach A)–C) is shown in Figure 4-6, with cost function values of (3-1) (losses) depicted in Figure 4-6a, and multi-objective cost function values of (3-2) shown in Figure 4-6b. Comparing the losses of loss minimizing optimal control to the losses of multi-objective optimal control shows that:

- As expected, both methods improve performance with respect to the situation without reactive power control. The improvement is larger when performance is expressed as the multi-objective cost function.
- The penalty on voltage deviation results in marginally higher losses. Conversely, the loss minimizing optimal control results in slightly higher multi-objective cost function values.
- Between 12.00 and 20.00 the objective function values of approach A) increases rapidly, which is caused by the increased power flow in the system due to the transition from peak real power generation to peak real consumption, as seen in Figure 4-5a and Figure 4-7. The objective function values of both optimal control approaches also increase, but significantly less.

Figure 4-7 compares the substation real and reactive power injection for strategy A)–C). The effect of optimal control on reactive power injection is clearly visible in Figure 4-7a. But the reduction in real power injection is much smaller, therefore Figure 4-7b shows the real power injection of A) on the left axis, and the reduction thereof due to optimal control B) and C) on the right axis. Some observations:

- For B), the reduction in reactive power injection is between 84–317 kVAR; for C) it is between 91–363 kVAR. The reactive power support greatly reduces the required reactive

Figure 4-5: SB feeder 15 minute interval sample data of (a) inverter 83 and (b) inverter 42.
Figure 4-6: SB feeder (a) losses and (b) multi-objective cost function values for the strategy A)–C).
Results and analysis of two case studies

power injection, and hence improves the power factor at the substation. In general all the reactive power consumption is provided locally from inverters.

- Whereas the substation injects an average of 180 kVAr for approach A), it absorbs an average of 0.3 kVAr and 12 kVAr for B) and C) respectively. The largest difference between the two optimal power flow approaches, B) and C), occurs between 16:00–21:00; approach C) causes a larger reverse reactive power flow to reduce voltage drop, as can be seen in Figure 4-8.

- The real power injection at the substation (of A)) illustrates the duck curve [3] that was introduced in Figure 1-3. It increases from an absorption of 55 kW around 14:00 to an injection of 618 kW around 20:00. The local provision of real power consumption between 10:00–16:00 contributes to the low loss and cost in Figure 4-6.

- As expected, inverter reactive power provision cannot prevent this rapid increase. The observed modest reduction of 0–3 kW for both B) and C) is caused by the reduced real power losses on the lines.

Figure 4-8 presents voltages at all nodes in the SB network for approach A)–C), which are indicated with colored surfaces. The substation voltage is assumed to be constant at 1 p.u., and is not included in the figure. Some observations:

- The result of A) shows that the voltage drop in the system is smallest between 10:00–16:00, when most real power demand is supplied by PV systems, as seen in Figure 4-5. During these hours, the combination of real power injection of PV systems and

Figure 4-7: SB feeder substation (a) reactive and (b) real power injection for the strategy A)–C).
loss minimizing reactive power generation of B) prompts a small voltage rise in the distribution feeder. The multi-objective cost function of C) achieves system voltages that are closer to the nominal value of 1 p.u., and simultaneously reduces losses as observed in the losses of Figure 4-6a.

- The transition from peak PV generation to peak consumption between 12.00 and 20.00 causes system voltages to change dramatically without control: the lower bound of the ANSI standard is violated around 20:00. Next to minimizing losses, the optimal control of B) reduces the voltage drop in the feeder during these hours, and prevents a lower bound constraint violation. The additional penalty on voltage deviation in the optimization problem’s objective function has the expected result: C) further reduces the voltage drop.

- The temporary and sudden reduction of PV generation at 14:00, observable in Figure 4-5a, causes a notable effect in voltages of A) and B), whereas the multi-objective optimal control acts appropriately to this sudden change; no significant change in system voltages is observed in Figure 4-8 for C). The penalty to voltage deviation not only improves the overall voltage drop in the system, but also balances the effect of quickly changing consumption or generation, i.e., the variability is reduced.

The results show that both optimal control approaches B) and C) have beneficial effect on power flow and voltages in the network. Inevitably the benefits of extending the objective
Results and analysis of two case studies

Table 4-4: Comparison of losses and multi-objective cost function values for different approaches. All values are in kW.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Type</th>
<th>SB [kW]</th>
<th>IEEE [kW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss reduction of B) w.r.t. A)</td>
<td>average</td>
<td>1.1</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>2.9</td>
<td>25.6</td>
</tr>
<tr>
<td>Loss reduction of B) w.r.t. C)</td>
<td>average</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>0.5</td>
<td>1.8</td>
</tr>
<tr>
<td>Cost reduction of C) w.r.t. A)</td>
<td>average</td>
<td>3.2</td>
<td>24.6</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>6.9</td>
<td>41.9</td>
</tr>
<tr>
<td>Cost reduction of C) w.r.t. B)</td>
<td>average</td>
<td>0.4</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>0.8</td>
<td>3.0</td>
</tr>
</tbody>
</table>

function of B) to the objective function of C) comes at the cost of additional losses. But, for this network, the maximum increase in loss is less than 0.5 kW, while the multi-objective optimal control simultaneously reduces losses with nearly 2.5 kW with respect to the scenario without control.

4-2-2 IEEE feeder optimal power flow

Results of the IEEE feeder model are simulated for the same day to compare the effect of optimal control on different feeders for similar circumstances. Recall that both feeders have comparable PV penetration (4-3) of approximately 61%. Therefore, the analyses mainly differ in network characteristics, total consumption, and the method in which Pecan Street data is aggregated. With respect to data aggregation, the sample data of Figure 4-5a is representative for the load and PV characteristics of the nodes in this case study. But, generation at the largest node of the IEEE feeder lacks the sag at 14:00, and shows closer resemblance with generation of Figure 4-5b.

Figure 4-9 compares the losses and multi-objective cost function values on the IEEE feeder. For this case study too, the effect of optimal control B) and C) is obvious, and results in lower losses and cost function values. Again, without reactive power provision, both measures for performance increase significantly towards 20:00. Compared to results of the SB feeder case study, the benefits of loss minimizing optimal control B) and multi-objective optimal control C) are larger. Table 4-4 quantifies the difference in loss and multi-objective function values:

- With respect to the scenario without reactive power support A), the average reduction in losses of B) is nearly 14 times larger for the IEEE feeder than for the SB feeder, and average multi-objective cost function value reduction of C) increases 7 times.
- Although the absolute difference between the two optimal control approaches B) and C) is larger for the IEEE feeder, the difference relative to the performance gain over scenario A) is significantly smaller. This is best seen in the results of B) and C) of Figure 4-9, where the lines are indistinguishable most of the time.

Figure 4-10 presents voltages at all nodes of the IEEE feeder for approaches A)–C). Once more, the substation voltage is assumed constant at 1 p.u., and is omitted from the results.
Figure 4-9: IEEE feeder (a) losses and (b) multi-objective cost function values for strategy A)–C).
The improvement in losses and multi-objective cost function values of optimal control, Figure 4-9, has clear effect on the system voltages, some observations:

- Minimizing losses results in a lower voltage drop in the network. But, it raises overall voltage around 12:00, has a strong fluctuation around 14:00, and voltage is slightly below the nominal value around 20:00.

- The effect of multi-objective optimal control is remarkable, the voltage profile is nearly constant throughout the day; all fluctuations are absorbed. Especially considering the relatively small trade-off in increased real power loss.

- The maximum values of A) show that the line segment from the substation to the second node in the network incurs a large voltage drop; the lowest value of the maximum system voltage is 0.978 p.u., which implies a voltage drop of at least 2.2% over the first line segment.

The effect of optimal control is significantly bigger for the IEEE feeder than for the SB feeder. Nonetheless, optimal control has clear benefits in both case studies. However, the differences are caused by characteristics of the feeders such as: total power flow and line impedance. One difference is the average ratio between resistance and reactance, or $r/x$ ratio, of the feeder lines, which is 1.5 for the SB feeder and 0.8 for the IEEE feeder. A smaller ratio implies relatively larger reactance component of the impedance, which results in a larger reduction of reactive power losses ($x_j\ell_j$ in (2-7b)), hence larger reduction of voltage drop ($x_j Q_j$ in (2-7c)) and larger reduction of current (2-7d) and real power loss losses. Local reactive power provision thus has a larger effect on feeders with a relatively low $r/x$ ratio.

4-2-3 Concluding remark

This section rigorously discussed the effect of optimal inverter reactive power provision on both testbeds. The results showed that optimal control reduces voltage drop over the feeder, reduces voltage variability throughout the day by absorbing (sudden) loss of PV generation, and reduces losses. However, optimal power flow problems cannot be solved in real time, due to a lack of communication infrastructure. Therefore, the next section evaluates the ability to capture the structure of the optimal power flow problem in regression models.

4-3 Regression results and analysis

Optimal power flow results in Section 4-2 demonstrated the potential of inverter reactive power control. Unfortunately, it is challenging to obtain optimal control solutions for real time demand data due to a lack of communication infrastructure. The proposed method constructs regression models based on optimal results, and subsequently employs the obtained models as local inverter controllers. This section analyzes the regression models with a number of experiments that are listed in Table 4-5. The goal of these analyses is to gain better insight in the strengths and limitations of this approach. To that end, Section 4-3-1 examines different settings of the stepwise algorithm. Section 4-3-2 interprets the differences between inverter regression models. Section 4-3-3 illustrates the challenge of using voltage information. Section
Figure 4-10: All IEEE feeder network voltages for strategy A)–C). Colored planes represent the range between the maximum and minimum voltages at any node in the network. Lower voltage bound of 0.95 p.u. is indicated with red dashed line.
Table 4-5: Experiment overview.

<table>
<thead>
<tr>
<th>Experiment (Subsection)</th>
<th>Question</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection criterion</td>
<td>Which criterion has best validation prediction accuracy?</td>
<td>BIC has largest training optimism penalty and obtains smallest models with highest prediction accuracy.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stepwise initialization</td>
<td>How stable are the final models?</td>
<td>Similar final models are obtained for different initialization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model complexity</td>
<td>What complexity has best validation prediction accuracy?</td>
<td>Highest complexity (quadratic) is most flexible, and stepwise selects subset of best variables.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>How much training data is the minimum samples that are representative for many power flow scenarios is unknown.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model interpretation</td>
<td>How different are inverter models?</td>
<td>There are significant differences in the regression models due to differences in network properties.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voltage challenge</td>
<td>How sensitive are models to voltage</td>
<td>Difference in voltage due to approximation of inverter output is small, with small effect on prediction accuracy.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leave-one-out</td>
<td>How important is each variable?</td>
<td>Voltage and reactive power consumption are most important for prediction.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective function</td>
<td>What is the effect of objective function on models?</td>
<td>Loss minimization is easier to predict, and those models show stronger relation to reactive power consumption.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Load correlation</td>
<td>How important is load correlation?</td>
<td>Without voltage information, prediction accuracy deteriorates when load correlation decreases.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timescale</td>
<td>Can local controllers be trained and implemented on different timescales?</td>
<td>Averaging over time period results in characteristic data which can not be used on a different timescale.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4-3-4 evaluates the availability of input data, and the effect of objective functions. Finally, Section 4-3-5 analyzes the impact of load correlation of prediction accuracy of regression models.

Regression data is randomly separated in training and validation data, where 70% of the data is assigned to the training set and the remaining 30% to the validation set. This section presents results for 15 minute interval optimal power flow data from July 2014; a total of 2977 samples are split in a training set of 2084 samples, and a validation set of 893 samples. All $R^2$ values in this section refer to the prediction error of validation data, and hence need no training optimism adjustment.

4-3-1 Stepwise settings

This section analyzes the prediction error and number of variables in the final regression model under varying settings of the stepwise regression technique. The first three experiments listed in Table 4-5 analyze the effect of: selection criterion, initialization, and allowed complexity.

A note on the terminology of model complexity. Four different model complexities are discussed in this section: constant, linear, interactions, and quadratic. The input variables $X_1$–$X_4$ are denoted as base variables. A constant model consists of the constant 1, or intercept, and the least squares solution for $\beta_0$ simply results in the average of all outputs $y$. Linear complexity refers to a model that is linear in the base variables and includes an intercept. Interactions extend the linear model with interactions between the base variables to construct variables as, e.g., $X_5 = X_1X_2$. Finally, quadratic models include an intercept, linear terms, interaction variables, and quadratic variables, e.g., $X_6 = X_1^2$.

Selection criterion

Table 4-6 shows the average number of variables included in an inverter model, and its prediction error for four selection criteria. The algorithm was initialized with linear models, the upper limit for model complexity was quadratic, and results of both feeders are included. For brevity the number of variables and $R^2$ values in the table are average values of all inverters in the networks. The results show that:

- IEEE inverter regression models are larger and more accurate than the SB feeder inverter models. The accuracy is the result of differences in the optimal power flow data, and is further discussed in Subsection 4-3-5.

- The prediction errors are nearly identical for all selection criteria. Therefore, the choice of criterion is based on the resulting model size, or complexity.

- Adjusted-$R^2$ obtains the largest regression models for the both feeders. While $F$-statistics (SSE) and the Akaike information criterion result in similar models for the IEEE feeder, the SSE models are smaller for SB inverters. But, as mentioned in Section 3-3-4, the Bayesian information criterion has a larger model complexity penalty for this number of observations, and leads to smaller models. This is confirmed by these results.

Because the prediction accuracy is nearly identical, BIC is selected as criterion for the remainder of this thesis.
Table 4-6: Effect of selection criterion on average number of variables and average $R^2$ values of inverter regression models.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>IEEE</th>
<th>Variables</th>
<th>$R^2$</th>
<th>SB</th>
<th>Variables</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj-$R^2$</td>
<td>13.50</td>
<td>0.976</td>
<td></td>
<td>11.63</td>
<td>0.906</td>
<td></td>
</tr>
<tr>
<td>SSE</td>
<td>12.50</td>
<td>0.976</td>
<td></td>
<td>9.78</td>
<td>0.905</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>12.50</td>
<td>0.976</td>
<td></td>
<td>10.48</td>
<td>0.906</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>11.50</td>
<td>0.976</td>
<td></td>
<td>8.96</td>
<td>0.905</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-7: Effect of initialization on regression model of IEEE feeder inverter 671, and average number of variables of all four IEEE feeder inverters. Average $R^2$ values of were 0.976 for all initializations.

<table>
<thead>
<tr>
<th>Initial model</th>
<th>Variables</th>
<th>Final model structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>12.25</td>
<td>Linear $+X_1X_2 + X_1X_4 + X_2X_4 + X_3X_4 + X_3^2 + X_4^2$</td>
</tr>
<tr>
<td>Linear</td>
<td>11.50</td>
<td>Linear $+X_1X_2 + X_1X_4 + X_2X_3 + X_3^2$</td>
</tr>
<tr>
<td>Interactions</td>
<td>12.50</td>
<td>Linear $+X_1X_2 + X_1X_4 + X_2X_3 + X_2X_4 + X_3X_4 + X_3^2$</td>
</tr>
<tr>
<td>Quadratic</td>
<td>12.00</td>
<td>Linear $+X_1X_2 + X_1X_3 + X_1X_4 + X_2X_4 + X_3X_4 + X_4^2$</td>
</tr>
</tbody>
</table>

Initialization

Although stepwise selection is an efficient algorithm to select a good subset of variables, it is not guaranteed to find the best subset. The choice of selection criterion and the initialization of the algorithm influences the final models, as shown in Tables 4-6 and 4-7. The effect of changing the initial model is presented as the average number of variables in IEEE inverter models, and illustrated with the final model structure of inverter 671. Two observations:

- Initialization of the algorithm with linear models results in the smallest models, while prediction accuracy is equal for all initializations.

- All final models structure in Table 4-7 are relatively similar. All models include: each linear variable; the interaction variables $X_1X_2$, and $X_1X_4$; and the quadratic variable $X_4^2$. These are plausibly the most relevant in explaining the output. One situation that leads to different final models is correlation between two (or more) variables; these compete to explain the same variance in output [61]. Both of the competing variables have similar predictive power, and only one needs to be included. Initialization determines which of the two competing variables is most relevant, and enters the model first. Thereafter, the other variable does not improve the selection criterion, and is not added to the model.

Because the base variables are anticipated in each model, and initialization has no clear effect on prediction accuracy, the remainder of this thesis initializes the stepwise algorithm with linear models.

6The location of inverter 671 is shown in Figure 4-2a.
Table 4-8: Effect of model complexity on regression model of IEEE feeder inverter 671, and average number of variables and average $R^2$ values of all four IEEE feeder inverters.

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Variables</th>
<th>$R^2$</th>
<th>Final model structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>4.75</td>
<td>0.932</td>
<td>$1 + X_1 + X_2 + X_3 + X_4$</td>
</tr>
<tr>
<td>Interactions</td>
<td>9.75</td>
<td>0.969</td>
<td>Linear $+X_1X_2 + X_1X_4 + X_2X_3 + X_2X_4 + X_3X_4$</td>
</tr>
<tr>
<td>Quadratic</td>
<td>11.50</td>
<td>0.976</td>
<td>Linear $+X_1X_2 + X_1X_4 + X_2X_3 + X_4^2$</td>
</tr>
</tbody>
</table>

Model complexity

Table 4-8 compares the average number of variables and the prediction error of linear, interaction, and quadratic regression models. The final model structure of inverter 671 is included as illustration. For the results in Table 4-8:

- Quadratic models have the highest $R^2$ value. This complexity includes linear and interaction models; hence, these are the most flexible.
- The least flexible models, linear models, also result in high $R^2$ values. This implies that the base variables are well chosen, and account for a large part of the output variable’s variance. However, the observed improvement of interaction and quadratic models confirm that the true optimal power flow solution is not merely linear in the base variables.

Unless specifically mentioned otherwise, the remainder of this thesis allows stepwise regression to include interaction and quadratic variables to the model. The main argument is: while the best structure of an inverter model is unknown (and can be different for each inverter), quadratic models contain the other model structures, and stepwise regression is relied on to select the variables that best describe the data. Linear models have high prediction accuracy, which is improved to near perfect prediction by including quadratic variables. Therefore, it is concluded that these models sufficiently represent the structure of the optimal solution; higher order variables are not tested.

Required number of data points for regression

The results in this section are all obtained with 15 minute interval data during the month July 2014, with a total of 2977 observations. The regression models are trained with 70% of the data, and validated with 30% of the data set, or 2084 and 893 samples respectively. The computation times of solving all 2977 optimal power flow problems was not a problem to this work, however it is relevant to evaluate the prediction error when less data is used for training. Therefore, regression models are constructed for all inverters with a varying number of training observations. The smallest training set consists of 24 samples, and the largest set contains 2084 samples. Figure 4-11 presents the lowest $R^2$ value (of all inverters) for varying training set size. Some observations:

- The $R^2$ values reach a maximum for training sets with 300 samples or more.
The regression models are only trained once. Empirical results indicate that the smallest number of randomly selected training samples that yields relatively constant $R^2$ values is 80. Using fewer training samples results in large differences in $R^2$ values depending on the samples that were selected. It is recommended to perform regression on at least 300 randomly selected samples. This number can be lower when the samples are carefully selected to be representative for a wide variety of power flow scenarios.

**Concluding remark**

For the remainder of this chapter, stepwise regression is initialized with linear models, variables are added and removed from the model based on the BIC, and both interaction and quadratic variables are allowed. These settings obtained the lowest prediction error and smallest models in the performed experiment. However, the improvement over other settings was marginal. Furthermore, regression is performed with a complete month of data, although for both case studies it is not strictly necessary to use more than 300 observations as training data.

### 4-3-2 Differences in inverter structure

This subsection analyzes differences between inverter models. The goal is to better understand the relation between the underlying optimal reactive power output solution and local measurements. Intuitively, the optimal solution is dependent on properties as: total impedance to the feeder substation, availability of PV systems at nearby nodes, size of nearby consumption, and to the local variables. These differences are expected to be reflected in the structure of inverter models. First, the full regression models of the four IEEE feeder inverters are interpreted. Second, the differences are further analyzed with scatter plots and correlation coefficients.
Model interpretation

The models of the four IEEE feeder inverters are presented in Table 4-9. These are obtained with stepwise regression for multi-objective optimal control data, and include all discussed choices: BIC criterion, linear initial models, and quadratic upper complexity. For additive models and standardized data, the regression coefficients of a multiple linear regression quantify how much output changes for a standard deviation change in the corresponding input variable. Unfortunately, this approach cannot be applied directly to this work, because the interaction and quadratic variables in this work make the models non-additive. Therefore, the coefficients do not quantify the change, but they do allow for qualitative analysis. The interpretation is visually supported with scatter plots of the optimal reactive power output versus the four base variables for all inverters in Figure 4-12. Table 4-10 further aids this with correlation coefficients between input and output variables of all inverters. Correlation coefficient is a measure for the linearity between two variables; it takes values between -1 and 1, where -1 implies a perfect negative linear relation, 0 no linear relation, and 1 a perfect positive linear relation [53, Ch.3].

Interpretation of the inverter models in Table 4-9 show that the models are clearly different from each other, both in structure, and coefficients:

- Generally, the linear variable coefficients of the inverter models are largest; thus, these variables have the largest effect in output prediction. This observation confirms the high prediction accuracy of linear models in Table 4-8.

- Voltage $X_1$ has relatively large coefficients in all models, and this coefficient dominates the regression model of inverter 671. This implies that the optimal reactive power output is best predicted with a voltage measurement, which is visually supported by the scatter plot of Figure 4-12c.

- The coefficients of reactive power capacity $X_4$ are largest in the models of inverters 645 and 675. In the model of inverter 675, the other coefficients are notably smaller, and this variable appears to be highly relevant. Contrarily, for inverter 645, the coefficients for $X_1–X_3$ are also large, and the importance of each variable is unclear.

- The three largest coefficients for inverter 634 correspond to $X_1, X_2, X_4$; the most notable is the smaller weight for reactive power.

A more detailed discussion of each of the base variables with support of scatter plots and correlation coefficients follows.

Voltage $X_1$

The correlations coefficients of voltage ($X_1$) are large and negative, which is visually supported by the scatter plots in the first column of Figure 4-12. These are best explained by realizing that the optimal power flow data is obtained with multi-objective optimization, including a penalty on the deviation of voltage from a nominal value. The voltage information in this section refers to reference voltage, which was first discussed in Section 3-4-2, and is obtained by simulating the system with optimal reactive power output of all-but-one inverters. Also the intuition of power flow equation (2-7c) supports that voltage is strongly related to the optimal reactive power output.
Table 4-9: Complete models of all four IEEE feeder inverters for multi-objective optimization data.

<table>
<thead>
<tr>
<th>X_e</th>
<th>( \beta_j )</th>
<th>X_e</th>
<th>( \beta_j )</th>
<th>X_e</th>
<th>( \beta_j )</th>
<th>X_e</th>
<th>( \beta_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.23</td>
<td>1</td>
<td>-0.65</td>
<td>1</td>
<td>0.13</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>( X_1 )</td>
<td>-1.53</td>
<td>( X_1 )</td>
<td>-0.90</td>
<td>( X_1 )</td>
<td>-1.04</td>
<td>( X_1 )</td>
<td>-0.41</td>
</tr>
<tr>
<td>( X_2 )</td>
<td>-1.10</td>
<td>( X_2 )</td>
<td>-1.22</td>
<td>( X_2 )</td>
<td>-0.02</td>
<td>( X_2 )</td>
<td>-0.13</td>
</tr>
<tr>
<td>( X_3 )</td>
<td>0.17</td>
<td>( X_3 )</td>
<td>0.99</td>
<td>( X_3 )</td>
<td>-5.05</td>
<td>( X_3 )</td>
<td>0.16</td>
</tr>
<tr>
<td>( X_4 )</td>
<td>0.69</td>
<td>( X_4 )</td>
<td>1.67</td>
<td>( X_4 )</td>
<td>-0.07</td>
<td>( X_4 )</td>
<td>0.74</td>
</tr>
<tr>
<td>( X_1 X_2 )</td>
<td>0.17</td>
<td>( X_1 X_3 )</td>
<td>-0.15</td>
<td>( X_1 X_2 )</td>
<td>0.16</td>
<td>( X_1 X_2 )</td>
<td>0.07</td>
</tr>
<tr>
<td>( X_1 X_3 )</td>
<td>-0.42</td>
<td>( X_1 X_4 )</td>
<td>-0.31</td>
<td>( X_1 X_4 )</td>
<td>-0.27</td>
<td>( X_1 X_3 )</td>
<td>0.15</td>
</tr>
<tr>
<td>( X_1 X_4 )</td>
<td>-0.41</td>
<td>( X_2 X_4 )</td>
<td>0.78</td>
<td>( X_2 X_3 )</td>
<td>0.05</td>
<td>( X_1 X_4 )</td>
<td>-0.08</td>
</tr>
<tr>
<td>( X_3 X_4 )</td>
<td>-0.39</td>
<td>( X_3 X_4 )</td>
<td>-0.71</td>
<td>( X_4 )</td>
<td>-0.07</td>
<td>( X_2 X_3 )</td>
<td>-0.18</td>
</tr>
<tr>
<td>( X_1 )</td>
<td>-0.21</td>
<td>( X_2 )</td>
<td>-0.04</td>
<td>( X_2 X_4 )</td>
<td>-0.27</td>
<td>( X_3 X_4 )</td>
<td>0.30</td>
</tr>
<tr>
<td>( X_2 )</td>
<td>0.23</td>
<td>( X_3 )</td>
<td>-0.05</td>
<td>( X_3 X_4 )</td>
<td>0.05</td>
<td>( X_2 )</td>
<td>0.24</td>
</tr>
<tr>
<td>( X_3 )</td>
<td>-0.16</td>
<td>( X_4 )</td>
<td>-0.22</td>
<td>( X_2 )</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Real power demand \( X_2 \)

The real power demand \( (X_2) \) is much less positively correlated to the optimal control output of the first three inverters, but has a positive correlation coefficient of 0.86 for inverter 675. Inspecting the second column of scatter plots shows an interesting structure for the first three inverters where the middle of the plot shows a lighter area. Clearly, the real power demand of the fourth inverter has a correlation with optimal reactive power output, however it is not necessarily linear. It is remarkable that the correlation coefficients are positive, while the coefficients in Table 4-9 are all negative.

Reactive power consumption \( X_3 \)

The reactive power consumption \( (X_3) \) has high positive correlation coefficients between 0.85–0.90 for the first three inverters, while the last inverter is not linearly correlated. Again, the scatter plots help to understand these correlation coefficients. The first three inverters show a positive relation, which is expected because losses are minimized by locally sourcing reactive power consumption. However, scatter plots of the second, third, and especially of the fourth inverter appear to have a maximum. Moreover, the reactive power consumption scatter plot of the fourth inverter also has a diagonal line in the upper left corner. These lines are caused by the inverter capacity constraint of (2-8).

Inverter reactive power capacity \( X_4 \)

The difference in structure of the optimal solution is best seen in the relation with inverter reactive power capacity. Whereas the scatter plots of inverters 634, 645, and 671 are comparable, the scatter plot of the fourth inverter is evidently different. Two bounds are visible in
Table 4-10: Correlation coefficients between optimal reactive power output versus standardized input variables $X_1$–$X_4$ of all inverters for multi-objective optimal control data.

<table>
<thead>
<tr>
<th>Inverter</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>634</td>
<td>-0.85</td>
<td>0.19</td>
<td>0.85</td>
<td>-0.40</td>
</tr>
<tr>
<td>645</td>
<td>-0.87</td>
<td>0.46</td>
<td>0.90</td>
<td>-0.22</td>
</tr>
<tr>
<td>671</td>
<td>-0.99</td>
<td>0.59</td>
<td>0.88</td>
<td>0.07</td>
</tr>
<tr>
<td>675</td>
<td>-0.90</td>
<td>0.86</td>
<td>-0.03</td>
<td>0.96</td>
</tr>
</tbody>
</table>

The plots of the first three inverters: the optimal output is diagonally upper bounded, and the inverter capacity is bounded on the right. The first is caused by inverter capacity constraint (2-8), and the second is caused by the inverter’s apparent power capacity $\pi$, that was first introduced in Figure 2-2.

The strong relation with inverter reactive power capacity of inverter 675 indicates that this inverter often generates reactive power at maximum capacity. The data points in the upper right corner suggest that the optimal output is close to maximum capacity when PV output is zero. This can be caused by a lack of PV systems and capacitor banks at surrounding nodes, so that the optimal solution requires the inverter to provide reactive power for the surrounding nodes too. Whether it is (economically) desirable that an inverter operates at full capacity most of the time is not considered in this thesis. Perhaps it is a more desirable solution for an electric utility company to place a capacitor bank at this location. These questions, and similarly, the question of how to incentivize inverter owners to participate in inverter control schemes is an interesting research topic.

Concluding remark

Overall, it is clear that the relation between optimal reactive power output and locally measured input data as represented by the regression variables and coefficients of all four inverters is distinctly different. The structure of the optimal inverter output thus depend on the location in the network; this statement also holds for the inverters of the SB feeder, which are not included for brevity. Although some of the relations can be explained and motivated with hypotheses, it is extremely challenging to design local controllers based on heuristics only. The large differences between inverters show that a single best model structure does not exist, and the tuning of local controllers clearly benefits from a data driven approach with sufficient design freedom.

4-3-3 Voltage information challenge

This section shows that voltage information comes with two challenges. First, including both voltages leads to competing regression model coefficients. Second, when regression-based control is implemented, voltage measurements couple different inverters, and inverter control can be considered as "closing-the-loop" (of an unstable system).
Figure 4-12: Scatter plots of optimal reactive power output versus standardized input variables $X_1$–$X_4$ of inverters (a) 634, (b) 645, (c) 671, and (d) 675. Data is obtained from multi-objective optimization.
Competing regression coefficients

The concept of reference voltage was introduced in Section 3-4-2 because it shows more correlation to the optimal inverter output than the optimal voltage, and it better conforms to the intuitive notion of voltage. Optimal voltage refers to voltage that is measured when the system is in the globally optimal state; reference voltage refers to voltage seen by an inverter when that inverter has no reactive power output, and all other inverters are operational. The reference voltage can be computed from a measured optimal voltage through the derivation in Appendix A.

Because optimal and reference voltage are closely related, including both in a regression model results in opposite coefficients that are several orders of magnitude larger than other coefficients, which makes it harder to interpret the regression models. A regression model of inverter 645 with linear complexity, where $X_1$ denotes reference voltage and $X_5$ denotes optimal voltage, is included in (4-4) to illustrate this. This is caused by the correlation between the variables; they explain the same output variance, and compete to do so [61]. Therefore, unless specifically mentioned otherwise, voltage in this chapter refers to reference voltage information.

$$\hat{y} = -1.6298X_1 + 0.0011X_2 - 0.0006X_3 + 0.0013X_4 + 0.9111X_5$$ (4-4)

Implementation challenge

Capturing the structure between optimal inverter output and local measurements is at the heart of the proposed method. The reference voltage in the training data thus represents the voltage where one inverter is inactive, and the remaining inverters have optimal reactive power output. Voltage contains information about the consumption and generation of other nodes; it is a network variable, and it connects different inverters. Implementation of regression-based control approximates the optimal reactive power output. The effect of this difference is analyzed with the following experiment:

1. Train regression model with reference voltage from optimal power flow data.
2. Estimate the reactive power output with optimal power flow validation data; express the prediction error as $R^2$ values.
3. Simulate the power flow with predicted inverter reactive power output from step 2.
4. Estimate the reactive power output with voltage measurements from step 3; express the prediction error as $R^2$ values.

The experiment is performed with standardized regression data, and non-standardized data. The results are shown in Table 4-11, where $v_{\text{ref}}$ refers to reference voltage from optimization data in step 1, and $v_{\text{sim}}$ denotes the reference voltage with regression-based control from step 3. The corresponding columns contain the $R^2$ values from step 3 and 4, and minimum and maximum difference between the two voltages are included. Recall that a $R^2$ value of 0

---

7The globally optimal system state depends on the objective function, which expresses the goals.

8Standardized refers to data with mean 0 and a sample standard deviation of 1.
Table 4-11: $R^2$ values of inverter models with voltage obtained from simulation with regression-based control implemented based on multi-objective optimal control data.

<table>
<thead>
<tr>
<th>Inverter</th>
<th>$R^2$ $v_{ref}$</th>
<th>$v_{sim}$</th>
<th>$v_{ref} - v_{sim}$ Min</th>
<th>$v_{ref} - v_{sim}$ Max</th>
<th>$R^2$ $v_{ref}$</th>
<th>$v_{sim}$</th>
<th>$v_{ref} - v_{sim}$ Min</th>
<th>$v_{ref} - v_{sim}$ Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>634</td>
<td>0.956</td>
<td>-4.731</td>
<td>-3.401</td>
<td>1.884</td>
<td>0.939</td>
<td>-1.879</td>
<td>-0.006</td>
<td>0.003</td>
</tr>
<tr>
<td>645</td>
<td>0.961</td>
<td>-1.542</td>
<td>-3.995</td>
<td>1.755</td>
<td>0.960</td>
<td>-1.210</td>
<td>-0.007</td>
<td>0.003</td>
</tr>
<tr>
<td>671</td>
<td>0.992</td>
<td>0.956</td>
<td>-0.670</td>
<td>0.277</td>
<td>0.992</td>
<td>0.956</td>
<td>-0.006</td>
<td>0.002</td>
</tr>
<tr>
<td>675</td>
<td>0.997</td>
<td>0.798</td>
<td>-3.384</td>
<td>1.330</td>
<td>0.997</td>
<td>0.737</td>
<td>-0.013</td>
<td>0.005</td>
</tr>
</tbody>
</table>

corresponds to the prediction error that is obtained by predicting the average reactive power output of the training data. Some observations:

- The difference between the two voltages with non-standardized data seems small compared to the voltage magnitude of 1 p.u., but is much larger for standardized data. This is caused by low variance in the voltage information, e.g., $v_{ref}$ of inverter 645 is between 0.994V–1.004V with standard deviation $\sigma = 0.002$.

- The minimum and maximum difference between the two voltages is comparable for non-standardized data, but the difference for inverter 671 is significantly smaller with standardized data. This is caused by a wider range of voltages (0.961V–1.001V) and corresponding larger variance ($\sigma = 0.009$) at inverter 671. Hence, the relative difference is smaller.

- The prediction accuracy of inverters 634 and 645 with voltage from regression-based control simulation is extremely poor with non-standardized regression data, and even worse with standardized regression data. This is caused by the large relative difference between the two voltage measurements, and the sensitivity, indicated with large coefficients, to $X_1$ (including interactions, and quadratic term) of both models in Table 4-9.

- The prediction error of inverter 671 increases slightly when using $v_{sim}$, and is exactly the same for standardized and non-standardized data. This is remarkable considering that the coefficients in Table 4-9 imply that this inverter is sensitive to voltage. But it is caused by the lower minimum and maximum relative difference between the two voltages (the average difference is -0.140 standard deviation).

- The prediction error of inverter 675 increases significantly but remains acceptable when using $v_{sim}$, the increase in error is larger for non-standardized data. This is best accounted for with the relatively small regression coefficients corresponding to $X_1$ in Table 4-9.

Overall, the prediction accuracy of regression based control is heavily affected by the difference between optimal inverter output and regression-based inverter output. This experiment

Footnote: Data is standardized by subtracting the sample mean and dividing by the standard deviation.
4-3 Regression results and analysis

Table 4-12: $R^2$ values of inverter models where a single input variable is removed for multi-objective optimal control data and loss minimizing optimal control data.

<table>
<thead>
<tr>
<th>Inverter</th>
<th>Multi-objective</th>
<th>Loss minimization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without</td>
<td>Without</td>
</tr>
<tr>
<td>634</td>
<td>0.956</td>
<td>0.811</td>
</tr>
<tr>
<td>645</td>
<td>0.961</td>
<td>0.883</td>
</tr>
<tr>
<td>671</td>
<td>0.991</td>
<td>0.926</td>
</tr>
<tr>
<td>675</td>
<td>0.997</td>
<td>0.993</td>
</tr>
<tr>
<td>Mean</td>
<td>0.976</td>
<td>0.903</td>
</tr>
</tbody>
</table>

included a single update of the power flow after approximating inverter output; a more realistic simulation iteratively updates power flow and inverter output until the result converges. Initial experiments indicated that this system is unstable and does not converge, even when regression models are trained with voltage data from step 3 and the algorithm is perfectly initialized. Therefore, it is recommended to further study this closed-loop behavior and perhaps implement additional measures to control the interaction between controllers.

### 4-3-4 Available input variables and objective function

As presented in Subsection 4-3-2, the regression models of various inverters are distinctly different. This section further analyzes the effect of input data on the prediction error of regression models. First, the effect of removing a variable from the input data is discussed. Intuitively, reactive power consumption and voltage are most relevant for minimizing losses and maintaining voltage close to a nominal value. This is expected to be reflected in this experiment by increased prediction error when these variables are removed. Second, the prediction error of both objective functions are compared. Intuitively, one might expect that the structure between input variables and reactive power output of loss minimizing optimal power flow data is more straightforward. The reason for this hypothesis is that loss minimization requires a similar approach at all times, except when the upper voltage constraint becomes active.

#### Leave-one-out

This experiment performs stepwise regression with data sets that exclude one of the base variables at a time. The resulting increase in prediction error indicates how important that variable is for regression; it implies that the remaining variables cannot account for the variance in output originally accounted for by the removed variable. Constructing regression models without voltage information simplifies implementation because voltage information couples the inverter controllers and can cause instability. The $R^2$ values of regression models with all four base variables, and results of the leave-one-out experiment are shown in Table 4-12. Three observations on the results of Table 4-12:
Results and analysis of two case studies

Table 4-13: Relation between independent variables.

<table>
<thead>
<tr>
<th></th>
<th>X_2</th>
<th>X_3</th>
<th>X_4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase</td>
<td>p^c↑↑</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Increase</td>
<td>p^g↓</td>
<td>-</td>
<td>↓</td>
</tr>
</tbody>
</table>

- Overall, the prediction accuracy is high; nearly all $R^2$ values are well above 0.9.
- For inverters 634, 645, and 671, removing voltage ($X_1$) has the largest effect of the $R^2$ value. Intuitively, voltage is important for a local controller because it captures information from the entire network, whereas the other variables contain local information only. The effect is stronger for the multi-objective optimization data.
- Inverter 675 is hardly affected by removing voltage information from the input data. Inspection of the scatter plots of Figure 4-12d and the regression model in Table 4-9 shows that the output of this inverter is strongly related to its reactive power capacity ($X_4$). This is caused by the location of this inverter, and the characteristics of surrounding loads.
- In general, the $R^2$ values are only marginally affected by removing $X_1$, $X_2$, or $X_3$. The remaining variables, their interactions, and quadratic terms thus contain sufficient information to account for most of the output variance.

The overall high prediction accuracy can be explained by the relation between the variables. Hardly decreasing $R^2$ values suggest that the remaining variables are related to the removed variable and describe the same output variance. Voltage is a function of all consumption and generation in the network, therefore there is only a weak relation between $X_1$ to local variables. The relation between $X_2$–$X_4$ is more obvious; these are constructed from local real power consumption $p^c_i$ and PV generation $p^g_i$ as:

- $X_2^{(i)} = p^c_i - p^g_i$
- $X_3^{(i)} = q^c_i = \text{p.f. } p^c_i$
- $X_4^{(i)} = \sqrt{s_i} - p^g_i$  

with constant power factor\(^{10}\) (p.f.) and inverter apparent power capacity $s_i$. Changing the real power consumption and generation affects $X_2$–$X_4$ as shown in Table 4-13, where ↑ indicates an increase and ↓ a decrease. However, using these variables directly as input to the regression does not yield better results. This affirms the choice of the three base variables, $X_2$-$X_4$.

**Different objective functions**

The ability of regression models to accurately predict optimal reactive power output with local measurements naturally depends on the objective function. The hypothesis in Table 4-5

---

\(^{10}\)Power factor is assumed constant in constructing reactive power consumption data from real consumption data.

Oscar Sondermeijer  
Master of Science Thesis
is that prediction errors for loss minimizing optimal solutions are lower, because minimizing loss is straightforward related to compensating local reactive power consumption and reactive consumption of surrounding loads. Table 4-12 presents results of the leave-one-out experiment for both objective functions. It shows that:

- Both objective functions can be accurately predicted with variables $X_1$–$X_4$, with average $R^2$ values of 0.976 and 0.978.
- The largest difference is observed for inverter 634 when voltage is not available: the multi-objective data has a $R^2$ value of 0.811, while loss minimizing optimal data results in a value of 0.930.

The results contradict the intuition: both objective functions are accurately described, but for specific inverters the loss minimizing optimal solution is easier to predict when voltage information is not available.

**Concluding remark**

Whereas voltage holds information about the state of the network, local real power demand, reactive power consumption, and inverter reactive power capacity measurements are independent from other nodes. The high prediction accuracy of regression models without voltage information for both objective functions is remarkable. It implies that the remaining local measurements can be used to approximate globally optimal reactive power output. This is possible with regression when the input variables explain, or account for, the output variance.

**4-3-5 Correlation of loads**

Optimal inverter output reduces losses\(^{11}\) by providing local reactive power consumption – which is available as measurement – , and reactive power consumption at neighboring nodes. A key requirement of the proposed solution is the lack of communication between nodes, hence consumption at other nodes is unknown to the inverter. Voltage, however, is a function of all demand in the network, and can indirectly inform the inverter about states of other nodes in the network. The high prediction accuracy of regression models that do not use voltage information raises another question: is the accuracy the result of high load correlation? When consumption at different nodes is highly correlated, a local measurement is representative for other loads, and communication is less important. Therefore, this subsection analyzes the effect of load correlation on prediction accuracy of the IEEE and SB feeder inverters. Three key insights are derived from these analyses. Firstly, without voltage information, some inverters are significantly affected by reducing load correlation while others are completely unaffected. Secondly, the regression models of the inverters that are least affected show a strong relation to reactive power consumption or inverter reactive power capacity. Third, a strong relation to these variables can be contributed to characteristics of surrounding loads.

\(^{11}\)For loss minimizing optimal power flow, the solution for multi-objective power flow depends on the voltage magnitude too, and can result in a tradeoff.
IEEE feeder

The $R^2$ values of all four IEEE inverters for varying load correlation are presented in Figure 4-13. Each load correlation correspond to a unique set of load and generation profiles for which the optimal power flow is computed. The $R^2$ values are obtained with regression models that are trained and validated with the corresponding optimal power flow data, and do not use voltage information. The results indicate that the prediction accuracy of:

- Inverter 675 has nearly no prediction error, and is not affected by reduced load correlation at all for both objective functions.
- Inverter 634 has larger prediction errors, but the $R^2$ value remains relatively constant under decreasing load correlation.
- Inverter 671 has large variations in $R^2$ values for multi-objective optimal control.
- Inverter 645 is clearly affected by reducing load correlation: $R^2$ values decrease from 0.90 to 0.53 for loss minimization, and from 0.88 to 0.58 for multi-objective optimal control.

A definite conclusion about the effect of load correlation on prediction error of inverters can not be drawn. Some inverters are affected, while other inverters are completely unaffected by reducing load correlation. The difference between the inverters can be explained with the inverter models in Table 4-9 and the scatter plots of Figure 4-12. Inverter 675 is generating its maximum capacity at nearly all times, and the measurement of reactive power capacity is dominant in predicting the output. This optimal reactive power structure is well described with the regression model and unaffected by load correlation. However, the model of the most affected inverter, at node 645, is strongly related to voltage information $X_1$. Without voltage information and with lower load correlation the structure of this optimal output is less accurately described with a multiple linear regression model. The effect of lower load correlation for different inverter models is further analyzed with SB feeder data, which has more inverters, and smaller –less aggregated– loads.

SB feeder

The SB feeder has more inverters, and smaller loads with an average correlation coefficient of 0.49. IEEE feeder results in Figure 4-13 showed that some inverters are more affected by reducing load correlation than others. The prediction error of more inverter models are evaluated, both with and without voltage information, to better understand why inverters are affected by lower load correlation, and the role of voltage information for these inverters. $R^2$ values of all inverters trained with and without voltage information are included in Table 4-14. Scatter plots and full inverter models are not included for brevity, but interconnection and indicative load size is included in Appendix C. The location of the inverters in the feeder is indicated in Figure 4-2b. Some observations:

- Inverters with the highest $R^2$ values are hardly affected by removing $X_1$: inverters 4, 29, 42, 62, and 83, have $R^2$ values between 0.988-1.000. Inspection of the regression
models shows that inverters 4, 29, and 42 are completely governed by reactive power capacity, which implies that they have relatively small PV systems with respect to local consumption, or that they provide reactive power for surrounding loads. Inverters 62 and 83 are mainly dependent on reactive power consumption. For inverter 62, this can be ascribed to the presence of PV systems at surrounding nodes. A closer inspection of inverter 83 shows that the magnitude of consumption and generation is significantly larger than the demand at surrounding nodes (see Appendix C), which explains the strong relation with local reactive power consumption.

• Some inverters with lower $R^2$ values are not much affected by removing $X_1$: inverters 20, 26, 33, and 61. The regression models of inverters 20, 26, and 61 are not dominated by a single variable. But the output of inverter 33 is most influenced by inverter capacity. The scatter plots of inverter 20, 26, and 33 show the clearest relation to inverter capacity, while the scatter plot of 61 shows the most relation with reactive power consumption. The reasoning of the previous observation can also be applied to these inverters.

• Some regression models poorly represent the optimal solution structure with $X_1$, but worse without $X_1$: inverters 46, 90, 94, and 108 have $R^2$ values below 0.85. An explanation for inverter 46 is that the consumption is significantly smaller than surrounding nodes, and that the optimal solution is therefore governed by surrounding nodes. The other inverters (90, 94, and 108) are on the same part of the feeder (see Figure 4-2b), and neighboring loads are of comparable size and do not have PV.

• Other inverters are significantly affected by removing voltage information, and the average $R^2$ value without $X_1$ is 0.763.

This analysis shows that the combination of $R^2$ values, inverter models, scatter plots, and knowledge of the feeder can be used to interpret the optimal results. The inverters that are least affected by low load correlation tend to have a strong relation to inverter capacity or local reactive power consumption. The first type of model is typically observed for inverters
Results and analysis of two case studies

Table 4-14: $R^2$ values of SB feeder inverter models based on multi-objective optimal power flow data with and without voltage information.

<table>
<thead>
<tr>
<th>Inverter</th>
<th>With $X_1$</th>
<th>Without $X_1$</th>
<th>Inverter</th>
<th>With $X_1$</th>
<th>Without $X_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.990</td>
<td>0.990</td>
<td>69</td>
<td>0.909</td>
<td>0.751</td>
</tr>
<tr>
<td>20</td>
<td>0.804</td>
<td>0.793</td>
<td>70</td>
<td>0.923</td>
<td>0.581</td>
</tr>
<tr>
<td>26</td>
<td>0.916</td>
<td>0.914</td>
<td>83</td>
<td>0.996</td>
<td>0.986</td>
</tr>
<tr>
<td>29</td>
<td>1.000</td>
<td>1.000</td>
<td>90</td>
<td>0.832</td>
<td>0.722</td>
</tr>
<tr>
<td>33</td>
<td>0.935</td>
<td>0.932</td>
<td>91</td>
<td>0.916</td>
<td>0.846</td>
</tr>
<tr>
<td>42</td>
<td>1.000</td>
<td>1.000</td>
<td>93</td>
<td>0.943</td>
<td>0.582</td>
</tr>
<tr>
<td>46</td>
<td>0.672</td>
<td>0.555</td>
<td>94</td>
<td>0.707</td>
<td>0.594</td>
</tr>
<tr>
<td>57</td>
<td>0.963</td>
<td>0.871</td>
<td>101</td>
<td>0.913</td>
<td>0.723</td>
</tr>
<tr>
<td>59</td>
<td>0.930</td>
<td>0.694</td>
<td>105</td>
<td>0.860</td>
<td>0.680</td>
</tr>
<tr>
<td>60</td>
<td>0.941</td>
<td>0.709</td>
<td>108</td>
<td>0.790</td>
<td>0.494</td>
</tr>
<tr>
<td>61</td>
<td>0.940</td>
<td>0.916</td>
<td>110</td>
<td>0.922</td>
<td>0.799</td>
</tr>
<tr>
<td>62</td>
<td>0.988</td>
<td>0.987</td>
<td>113</td>
<td>0.867</td>
<td>0.407</td>
</tr>
<tr>
<td>67</td>
<td>0.961</td>
<td>0.825</td>
<td>129</td>
<td>0.919</td>
<td>0.657</td>
</tr>
<tr>
<td>68</td>
<td>0.903</td>
<td>0.584</td>
<td>Mean</td>
<td>0.905</td>
<td>0.763</td>
</tr>
</tbody>
</table>

that have little inverter capacity at surrounding nodes; the second type of model is observed at nodes with much significantly larger consumption than surrounding nodes, or at nodes that have sufficient PV systems at surrounding nodes. Unfortunately this presents few opportunities for improvement of the method because the location of PV systems is uncontrollable to DSOs.

4-3-6 Timescale of data

The final analysis of regression results concerns the time interval at which the consumption and generation data is sampled. Up to this point, all analysis were performed with Pecan Street data that was averaged over 15 minutes time intervals. Regression models are trained with sufficient data to represent a wide variety of power flow scenarios, and have a low prediction error for new –15 minutes averaged– measurements. However, in practice it can be desirable to control the inverters on a faster timescale, e.g., at 1 minute intervals. Measurements that are averaged over 1 minute show more variability than data that is averaged over 15 minutes; hence the average load correlation on the SB feeder is 0.36 for 1 minute data and 0.49 for 15 minutes data\(^\text{12}\). The question is if data that is averaged over 15 minutes is representative for shorter interval data, and vice versa. Therefore, Table 4-15 presents the $R^2$ values of regression models: trained with 15 minutes data and validated with 1 minute data, trained with 1 minute data and validated with 15 minute data, and trained and validated with the same data for comparison. Results are included both with, as well as without voltage information. The 1 minute data is obtained at each minute of July 5 2014. Some observations:

- It is clearly best to use similar data for training and validation, i.e., if control is im-

\(^{12}\)These correlation coefficients are obtained from identical load allocation, i.e., the 1- and 15 minutes data of the SB feeder loads are constructed with exactly the same Pecan Street IDs.
Table 4-15: $R^2$ values for 1- and 15 minutes interval data with and without voltage information. Data obtained with multi-objective optimal power flow on SB feeder.

<table>
<thead>
<tr>
<th>Training</th>
<th>Validation</th>
<th>With Voltage</th>
<th>Without Voltage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 min</td>
<td>1 min</td>
<td>0.877</td>
<td>0.693</td>
</tr>
<tr>
<td></td>
<td>15 min</td>
<td>-4.64</td>
<td>-7.49</td>
</tr>
<tr>
<td>15 min</td>
<td>15 min</td>
<td>0.905</td>
<td>0.763</td>
</tr>
<tr>
<td></td>
<td>1 min</td>
<td>0.579</td>
<td>-2.28</td>
</tr>
</tbody>
</table>

implemented on a 1 minute timescale, the regression models should be constructed with 1 minute interval data. However, models trained with 15 minutes data have higher $R^2$ values for predictions on 1 minute data than vice versa. Especially when voltage information is available, the average $R^2$ value of 0.579 is not as low as the other results on the second and fourth row (all negative values).

- The prediction accuracy of regression models trained and validated with 15 minutes data (third row) is highest; the $R^2$ values of models trained and validated with 1 minute data is slightly lower. This can be ascribed to the higher load correlation of 15 minute data. However, it could be relevant to sample the 1 minute data from a longer period of time (e.g., a month) to include a wider variety of power flow scenarios.

- The maximum prediction accuracy in all experiments is remarkably high. This implies that the relation between local measurements and optimal reactive power output is similar for 1 and 15 minutes data, and is well captured by the multiple-linear regression model.

Based on this experiment it can be said that consumption and generation measurements that are averaged over 1 and 15 minutes are not sufficiently representative to be used interchangeably. The SB feeder testbed is the most realistic: it is the largest network with many loads that represent a single or a few residences. The experiment with 1 minute data has the lowest correlation between different loads; therefore, represents the largest challenge for local controllers that pursue a system-wide objective. This is reflected in the lower prediction accuracy of these regression models, with an average $R^2$ value of 0.693 (without voltage information).

4-4 Simulation with regression-based controllers

Section 4-2 presented the potential improvement of optimal inverter reactive power output, and Section 4-3 showed low overall prediction error of inverter regression models under varying circumstances. Therefore, this section presents the effect that approximating optimal reactive power output has on power flow, and quantifies the difference with respect to optimization based control. Regression-based control is simulated on both the SB, and the IEEE feeder, and the result is expressed as cost function values, which are compared with globally optimal cost. Moreover, the approach is compared to two other local control approaches. For the SB feeder, the cost function values are supported with a visualization of system voltages.
Results and analysis of two case studies and losses for better interpretation. All simulations are performed with 15 minutes interval consumption and generation data, and optimal refers to the multi-objective cost function that reduces losses and voltage deviations. The different control strategies in this section are indicated with the letters A), B), C), and D), which refer to:

A) **No inverter reactive power control:** this corresponds to the present-day situation in the U.S.. Results are presented in red.

B) **Regression-based control:** the proposes method of this thesis, implemented *without voltage information*. Results are presented in blue.

C) **Constant power factor control:** adopted from [6]. Tuned to generate power at a constant *leading* power factor of 0.9 to reduce losses instead of reducing voltage fluctuations. Results are presented in yellow.

D) **Varying power factor control:** present-day situation in Germany. Depicted in Figure 2-3 with maximum *lagging* power factor of 0.9. Results are presented in purple.

Performance of other local control approaches, such as, the control curve proposed in [6] and shown in Figure 2-4, are prone to tuning, and are therefore not used for comparison. The average correlation coefficient between consumption is 0.49 on the SB feeder, and 0.84 on the IEEE feeder.

First, Subsection 4-4-1 evaluates the results of regression-based control in terms of the cost function values for both testbeds, and compares these to globally optimal results. Second, Subsection 4-4-2 interprets the cost function values of the SB feeder with the aid of system voltages and losses. Finally, Subsection 4-4-3 demonstrates the potential of cooperation between regression-based control and the substation transformer to pursue voltage optimization.

### 4-4-1 Cost function values

Figures 4-14a and 4-15a compare the objective function values, or cost, of approach A)–C) on 4 July 2014 for the SB feeder, and for the IEEE feeder. Figures 4-14b and 4-15b quantify how sub-optimal the system states are for regression-based control on both feeders. The relative optimality of B) with respect to optimal system states is expressed as (4-5), where \( J \) denotes the cost, superscript A) and B) refer to the corresponding control strategies, and \( J^* \) denotes the optimal cost. The relative optimality of B) is 1 when regression-based control obtains globally optimal system states; 0 when A) and B) have equal cost; negative values when the cost of regression-based control is higher (worse) than the cost without reactive power control. Negative values thus imply that control has detrimental effect on the system in terms of the goals that are articulated as the objective function. A graphic interpretation of this measure is the distance between \( J^{B)} \) and \( J^* \) divided by the distance between \( J^{A)} \) and \( J^* \) in figures as Figure 4-14. However, the optimal cost is not included in Figure 4-14 because it would overlap with the cost of B).

\[
\text{Relative optimality } \left( J^{B)} \right) = 1 - \frac{J^{B)}}{J^{A)}} - J^* \\
(4-5)
\]

Some observations on Figure 4-14:
Figure 4-14: SB feeder (a) multi-objective cost function values for strategy A)–D), and (b) relative optimality of B).
- Approach B) has the best performance of all inverters at all moments in the day, as indicated by the lowest cost in Figure 4-14a.

- Approaches C) and D) do not control inverter output at all times of the day: C) is active between approximately 08:00–20:00, and D) is active between approximately 10:00–18:00. This is as expected, because the inverter output of these control strategies is a function of the PV generation; without (sufficient) PV generation, these controllers have no reactive power output, and are effectively the same as A). The sudden loss of PV generation around 14:00 (best seen in Figures 4-5 and 4-7) causes D) to become inactive, and the cost equals approach A).

- Approach D) has a detrimental effect on the system, as indicated by the higher cost in Figure 4-14a. This is explained by the different goal that this controller was developed for. Generally, the capacity of a PV system in Germany is significantly larger than the local consumption [37, 38], which results in reverse power flow and high voltages. This controller increases the power flow and voltage drop in the system when the output of a PV system is above 50% of the maximum output. However, the reverse power flow in this network is not large enough to cause an overall voltage rise in the network, and approach D) is clearly inadequate. It is recommended to compare B) and D) on a network with an even higher PV penetration.

- Approach B) results in system states that are close to the global optimum, as implied by the high relative optimality of B) in Figure 4-14b. The lowest value of (4-5) is 0.934, while the average value is 0.986.

Similar observations are made for Figure 4-15, except that C) has the lowest value at two 15 minute intervals, the first around 12:00, and the second around 15:00. However, regression-based control systematically obtains close to optimal system states as shown by the high relative optimality in Figure 4-15b, with a minimum of 0.956 and average of 0.988. Overall, the differences between the four control strategies are larger on the IEEE feeder than on the SB feeder, which is in agreement with the previous optimal results of Figures 4-6b and 4-9b.

### 4-4-2 Voltage and losses

Figures 4-16 and 4-17 compare the effect of the four control strategies on total real power loss, and minimum and maximum voltages in the network. The substation voltage is held constant at 1 p.u. to decouple the distribution feeder from the transmission network. Some observations about the results:

- Between 00:00–08:00 the losses for A)–D) are low and close together, but the minimum and maximum voltages in the network are closer together for B). During these hours the consumption in the network is low (Figure 4-7), and reactive power provision has limited effect on losses. However, simultaneously the voltage in the network is closer together because the voltage drop is reduced.

- Losses for B) and C) are close together until 18:00, but, thereafter, losses of C) increase with respect to B). Until 18:00, PV generation is relatively high, and the injected reactive power output of C) is largest, which reduces losses. However, after 18:00, the reactive power output of B) improves losses with respect to the other strategies.
Figure 4-15: IEEE feeder (a) multi-objective cost function values for strategy A)–D), and (b) relative optimality of B).

Figure 4-16: SB feeder network losses for strategy A)–D).
Results and analysis of two case studies

Figure 4-17: SB feeder network voltages for strategy A)–D). Colored planes represent the range between the maximum and minimum voltages of all nodes in the network. Lower voltage bound of 0.95 p.u. is indicated with red dashed line.
• The lowest cost of approach B) is further supported by the voltages in Figure 4-17. The minimum and maximum system voltages are nearly constant throughout the day, all voltages in the system are between 1.007–0.985 p.u.. Moreover, B) responds adequately to the abrupt loss of PV generation around 14:00, and the voltages profile remains flat.

• Reactive power generation of C) reduces losses but causes a voltage rise between 10:00–16:00. Inverter’s reactive power injection causes the voltages to rise because the real power PV generation already results in a small reverse real power flow.

• Approaches A), C), and D) lead to a violation of the lower voltage bound around 20:00. Power factor control does not utilize inverter reactive power capacity when PV systems have no output; therefore, the constraint violation is not prevented by C) and D).

• Reactive power consumption of D) increases real power losses, and voltage drop in the network. Between 16:00–22:00 the voltage drop is largest and system voltages are approximately between 1.00–0.95 p.u..

Comparison of the losses and voltages shows that local inverter control strategies A)–D) result in large differences. It is not straightforward to choose the leading or lagging power factor that results in the most adequate trade-off between loss minimization and reducing voltage deviations. Although constant power factor approach C) at times results in losses that are comparable to regression-based control, its effect on system voltage is clearly worse. Approach D) did not have the desired effect on this feeder, which is seen as lower voltages and higher losses, because it was designed to reduce the effect of excessive PV generation on system voltages.

4-4-3 Voltage optimization

Regression-based control reduces voltage fluctuations throughout the day, and reduces the voltage drop over the feeder, i.e., the voltage magnitude at the end of the feeder is closer to the voltage magnitude at the substation. Although not simulated, an expected result is that capacitor banks and tap changing voltage regulators are operated less frequently. Moreover, it simplifies voltage optimization for a distribution system operator (DSO) and the increases the potential effect. Voltage optimization was introduced as an economical incentive for DSOs in Section 1-1, and can be interpreted as lowering voltages to reduce overall energy consumption (CVR), or raising voltage to reduce losses, which agrees with the goal of this section. Therefore, a fifth strategy is introduced:

E) Cooperation of regression-based control and substation transformer: the regression-based control of B) is used to reduce losses and flatten the voltage profile, simultaneously the substation voltage raised to 1.04 p.u. by the substation transformer to further reduce losses without violating the upper voltage constraint. Results are presented in green.

The effect of E) is compared to the system voltages and losses of B) in Figures 4-18a and 4-18b. The voltage profile of E) is as expected: it is similar to B), but 0.04 p.u. higher, without exceeding 1.05 p.u.. The voltage profiles seem identical, however, the intuition from
(2-7c) and (2-7d) implies that the voltage drop of E) is slightly smaller due to larger $v_i$ at all nodes $i$. Moreover, the loss of E) is lower at all times, with the largest absolute improvement is around 20:00. The result of E) is included to demonstrate the additional operational flexibility that regression-based control provides for DSOs. This straightforward cooperation between traditional equipment and novel inverter control is possible because this specific implementation of regression-based control does not use voltage information. If the challenge of implementing regression-based control with voltage information is overcome, the suggested cooperation likely requires communication of a reference voltage to the inverter controller to prevent hindrance.

### 4-4-4 Concluding remark

The results in this section showed that regression-based control is able to systematically obtain results that are, on average, within 2% of the global optimum with respect to no reactive power control. This result was obtained with controllers parameterized to reduce losses and maintain voltage at all nodes close to 1 p.u., but the regression results of Table 4-12 lead to believe that regression-based controllers tuned for loss minimization perform at

---

Oscar Sondermeijer  
Master of Science Thesis
Regression-based control has comparable performance on both feeders. The results of the two power factor controlled local controllers indicate that it is insufficient to monitor PV generation only, when the objective requires conflicting control action at different times. The regression-based local controller used more information, and resulted in the desired trade-off between both goals. While reducing losses, the voltage profile was flat throughout the feeder and throughout the day. Although not simulated, regression-based control would hypothetically reduce switching of legacy equipment; it allows DSOs to use the substation transformer (and potentially other voltage regulators) for voltage optimization. Furthermore, it is recommended to study the effect of increasing PV penetration on regression-based control, as well as on other local control approaches.

4-5 Concluding remarks

This chapter presented and thoroughly analyzed results of regression-based control on two realistic testbeds. The two models represented real distribution feeders: the first a reduced-order model with only 13 nodes but large loads; the second with 129 nodes and 53 smaller loads. PV systems were manually placed in the first network, while the locations were randomly selected for the second feeder. Real consumption and generation measurements from the Pecan Street database [59] were used to populate the feeder with 15 minute interval data.

For both testbeds, optimal inverter control has clear effect on power flow and the voltage profile. Inverter-based reactive power support was capable of minimizing losses, and significantly reducing voltage drop in the feeder and voltage variability throughout the day. The effectiveness of reactive power provision was greater on the smaller feeder, which can be accredited to network properties. Moreover, the increase in loss due to the additional objective of reducing voltage variability was remarkably small although the effect on voltage was significant.

The prediction error of regression models was carefully analyzed for various aspects of the regression method and underlying optimization data in Section 4-3. First of all, different settings of the stepwise algorithm had little effect on the prediction error, but showed changes in obtained model structure. Second, for different inverters, the diversity in regression models was telling: some inverters relied heavily on reactive power capacity; others showed stronger relation with local reactive power consumption. This shows that pursuing system-wide objectives requires different control from each inverter. Additionally, it motivates careful selection and tuning of each controller, which can be an arduous effort for other local control methods. Third, although the prediction error of regression benefits from voltage information, it is challenging to implement regression-based controllers that rely on network variables to predict the reactive power output. This challenge is mainly caused by the combination of regression model’s high sensitivity to voltage information, and the differences between voltage in optimal power flow simulation and voltage in regression-based control simulation. Fourth, the load correlation between nodes is a relevant aspect for the prediction error of some inverters, while others are unaffected by it. Lastly, regression models should be constructed with optimal power flow data that is averaged over the same time intervals as the timescale at which controllers are implemented.

Finally, the simulation results of regression-based control were close to optimal for both
testbeds. When learned from multi-objective optimization data, local regression-based control reduced losses, reduced voltage drop on the feeder, and reduced voltage variability throughout the day. This could drastically reduce the number of legacy equipment operations and prolong the lifetime of these devices. Moreover, an initial simulation demonstrated the potential of cooperation between regression-based controlled inverters and substation transformer to perform voltage optimization. Finally, compared to a scenario without reactive power provision, regression-based control prevented a violation of the voltage constraint. Regression-based control has the potential to enhance power quality and power flow in distribution feeders.
Conclusions and recommendations

In this thesis, inverter reactive power provision was controlled with a regression-based method to enhance power quality and power flow in distribution systems. This chapter summarizes the conclusions of this thesis and lists recommendations for future research.

5-1 Conclusions

The overall goal of this thesis was to formulate a structured approach to design inverter controllers that use reactive power to maintain and enhance the efficiency and quality of electric distribution systems with high penetration of solar generation. Therefore, the advantages and limitations of local and optimization-based methods (Chapter 2), were combined in the regression-based control method (Chapter 3). This method exploits available historic data and convex optimal power flow techniques to find the globally optimal reactive power output of each inverter in a network for representative power flow scenarios. For each inverter, regression is used to construct a model that describes the relation between optimal inverter output and local measurements. These models are then employed as local inverter controllers. The performance of regression-based control was analyzed in simulation on two testbeds (Chapter 4); the most important conclusions are included below, organized per design requirement from Section 2-3.

1. **Controllers must respect practical limitations of distribution systems:**
   The resulting controllers rely on local measurements only to compute reactive power output, and thus falls in the category of local control methods. The novelty of regression-based control lies in the design process, and thereby is a local alternative for optimization-based methods. Although regression-based control only approximates globally optimal results, it respects the limitations of distribution systems. Therefore, the presented method is an alternative that is one step closer to implementation on physical systems.

2. **Controllers must cooperate and pursue system-wide and globally optimal results, with the flexibility to represent different control objectives and respect critical system constraints:**
Simulation of regression-based control (without voltage information) resulted in near-optimal results. For both testbeds, simulation achieved optimality on the order of 98% compared to a scenario without reactive power control. The results of regression models constructed from loss minimizing optimal data and multi-objective optimal data were comparable, thereby demonstrating the possibility to use different objective functions. Although the objective function can be tailored to reflect the priorities of a distribution system operator (DSO), they are limited to objectives that are compatible with convex optimal power flow problems. Furthermore, system constraints can be incorporated in the optimization problem, but constraint satisfaction of regression-based controllers cannot be formally guaranteed.

3. Control design must be performed systematically, require little or no heuristics, and leverage available data:
   The novelty of regression-based control is in the systematic and data driven approach to local control design. The only manually selected aspects of the method were the objective function of the optimization problem, the local variables used for regression, and the stepwise regression settings. The choices were thoroughly discussed and motivated, and especially the last two aspects require little tuning for new distribution networks. Design of the local regression based controllers requires sufficient historic data from advanced metering infrastructure (AMI) (smart meters): empirical results indicate that the regression requires at least 300 randomly selected load and power flow scenarios.

4. Other conclusions:
   - Inverter models that collectively obtain near-optimal system states are distinctly different. This implies that careful local control design is required, which benefits from a systematic and data driven approach to prevent arduous tuning.
   - Availability of voltage information improves prediction accuracy of regression models. But implementing network-dependent variables (voltage) is more challenging than using network-independent variables (consumption, photovoltaic (PV) generation).
   - Without voltage information, load correlation is important for the prediction error of inverter models that are not strongly related to reactive power consumption or inverter reactive power capacity.

This thesis presented a novel methodology to design local controllers for power flow and voltage regulation in distribution networks. The main contribution of this method is the structured and data driven design approach that results in near optimal results. This makes it straightforward to apply the framework to different feeders, and avoids having to rely on heuristics and arduous tuning to obtain desirable effects. The developed insights motivate the need for controllers that are tailored to individual inverters to approach optimal results. Finally, the two testbeds demonstrated the vast potential of inverter-based reactive power provision for the enhancing power flow and voltages in distribution systems. Clearly, there are hurdles to be overcome; therefore the next section presents recommendations for future research.

Oscar Sondermeijer
Master of Science Thesis
5-2 **Recommendations for future research**

This section presents a list of the most relevant topics for future research. The recommendations are separated in suggestions that fall within the proposed framework, and suggestions that further extend this work.

5-2-1 **Within the proposed approach**

- *Reduce sensitivity to voltage information:*
  
  The challenge of using voltage information with this framework was introduced in Section 3-4 and empirical results were shown in Section 4-3. Voltage information is different from the other input variables of the regression models because it carries information about the rest of the network. The voltage magnitude at a specific node, node \( i \), in the network can be expressed as the voltage at the substation (which is assumed to be constant) and the voltage drop in the network. Voltage drop is a function of the line impedance and the total power flow on the lines between node \( i \) and the substation. The total power flow between node \( i \) and the substation is governed by the demand (consumption minus generation) of all nodes that are connected to the substation through at least one line segment between node \( i \) and the substation. Therefore, a voltage measurement represents the state of the network, but also couples nearly all inverters in a network.

  The regression models are trained with optimal power flow data, and the voltage information represents the network state when all inverters have optimal power output. However, when regression-based control is implemented, the controller *approximates* the optimal power output; therefore, the voltage measurement available to an inverter is slightly different from the voltage used in training data. Moreover, the inverters now form a closed loop in which they influence each other’s output.

  Empirical results showed that the inverter regression models in the IEEE case study were very sensitive to voltage information. One reason is that voltage measurements in the training data are in a small range (between 0.96–0.99). Therefore, a perceived small change, e.g., caused by the prediction error of all inverters, is more significant relative to the small range in the training data. Initial experiments where the power flow and regression output of all inverters were iteratively solved showed that this can lead to instability of the inverter controllers. It is recommended to further study this effect. One possible solution is to incorporate the anticipated uncertainty of the voltage measurement in the regression model to lower the voltage sensitivity. Another relevant direction is to limit the maximum change in inverter output between the iterative steps, this solution is inspired by the work of [6, 62]. The author deems it a high priority to carefully study this phenomenon before decentralized inverter controllers that act on voltage information are implemented, especially when inverter reactive output has significant impact on network voltages as in the case studies in Chapter 4.

- *Incorporate voltage dependent load models:*
  
  The described framework was tested with voltage independent consumption data in Chapter 4. Realistic loads do not necessarily have constant power characteristics: some loads have constant current, constant impedance, or a combination of all three. Such
loads can be described with the ZIP-model, which can be incorporated in the optimal power flow problem [19]. Modeling and simulating voltage dependent loads closer resembles reality and shows the potential energy saving of voltage optimization approaches such as CVR.

- **Study different model structures for regression:**
  Especially when the optimal reactive power output is obtained from data with a low (<0.5) average load correlation coefficient and voltage information is unavailable, the prediction error of most inverter models increased. The prediction error increases because the structure of the optimal solution is poorly described by multiple linear models with interactions and quadratic variables. More flexible models potentially improve the result. Careful consideration is required to prevent selecting overly complex methods that require extensive tuning, making it more laborious for DSOs to apply this framework.

- **Study the performance of this framework for a larger total installed capacity of PV generation:**
  Higher PV penetration in a network brings other challenges for the operation of a distribution network. Large reverse power flow in the network results in voltage increase instead of voltage drop in the network, and minimizing losses further raises voltages in the network. The two goals in the multi-objective optimization (reducing losses and maintaining system voltage close to a nominal value) then require opposing control action. It is recommended to study the performance of the discussed objective functions, and the ability of this regression approach to describe the structure of the optimal solution.

5-2-2 Extending the proposed framework

- **Incorporate unbalanced power flow modeling, simulation, and optimization:**
  Power flow in distribution systems is inherently unbalanced due to the numerous unequal loads connected to a single phase, and the unbalanced spacing of three-phase line segments [23]. Multi-phase models describe the unbalanced power flow in a network, but are more involved than their one-line counterparts that assume balanced power flow. Unbalanced power flow can be simulated with specialized commercial e.g., [63], or open-source software e.g., [64]. This software can better analyze the expected effect of this method on a real distribution feeder. An anticipated outcome of that analysis is that regression models that are trained with data from a balanced optimal power flow problem do not yield the desired performance. The framework can then be further extended with the convex relaxation of the unbalanced optimal power flow problem described in [25].

- **Study the formulation of a learning algorithm:**
  The regression models in this thesis are optimized to have low prediction error on training data. Therefore, the results in this thesis rely on training the regression models with representative data. The error between the regression prediction and the optimal solution likely increases when the situation in the network changes. Results in Table 4-2 showed that regression models that are trained with data from one month accurately
predicted the optimal solution throughout the year when nothing else changed. However, the performance of this network cannot be guaranteed when the optimal power flow solution changes significantly because e.g., more residences install PV systems, more electric vehicles are charged, or the physical distribution network is changed. This framework is likely more robust to changes when extended with a learning algorithm. This could encompass regular and automatic adaption of regression models to include the most recent data. Practically this can involve DSOs that solve optimal power flow problems on a daily basis with the most recent AMI data. The regression models can be automatically updated, or the new data can be used for validation, and models are only updated when the prediction error exceeds a certain threshold.

- **Tailor this framework to the distribution system of the future:**
  This thesis presented a formulation of regression-based control that accounts for the practical limitations of the present-day distribution network. However, the rapid installation of AMI makes it reasonable to assume that a faster communication infrastructure will be available in the future. For example, one can envision a scenario in which inverters have access to AMI data of neighboring nodes, which likely increases the accuracy of regression models. Alternatively, let us assume that the required communication infrastructure for optimization-based control methods exists. The regression-based control framework can be extended and the potential reduction of computational effort could be one of the main arguments to prefer it over (distributed) optimization-based control. This recommendation is mainly intended as a thought starter, but opens up many new research directions.

- **Study options to incentivize customers to participate in inverter-based reactive power support:**
  This thesis completely ignored the issue of incentivizing customers or third-parties to participate in inverter-based reactive power support. Germany enforced their local inverter control system through legislation, and inverter companies ensured that their products support the requirements [5]. In the U.S., the leading PV system installation company presents customers the option to lease the system [39]. This presents the opportunity to implement this type of control in collaboration with a few suppliers and potentially simplifies the economical aspect of this solution. Furthermore, it is relevant to study the impact of providing reactive power support on the performance and expected lifetime of an inverter.
Appendix A

Effect of reactive power injection

This appendix derives the result of (3-20) for a three bus line segment example of Figure A-1. With the branch flow equations (A-1) the power flow in a network is described. Under the assumption that all loads are constant, an analytic expression is formulated for the change in voltage due to injecting reactive power with an inverter. Subscript \( y \) denotes variables including the effect of control, while subscript \( n \) denotes variables without control. Equation A-2 expresses difference in voltage at node 3 in terms of power flow on, and impedance of the line segment between nodes 2 and 3. Equations A-3 iterate this result for nodes 1 and 2.

\[
\begin{align*}
    v_{i+1} &= v_i - 2 (r_{i+1} P_{i+1} + x_{i+1} Q_{i+1}) - \left( \frac{P_{i+1}^2 + Q_{i+1}^2}{v_{i+1}} \right) \\
    P_{i+1} &= P_i - r_{i+1} (\frac{P_{i+1} + Q_{i+1}}{v_{i+1}}) - p_i^c + p_i^g \\
    Q_{i+1} &= Q_i - x_{i+1} (\frac{P_{i+1} + Q_{i+1}}{v_{i+1}}) - q_i^c + q_i^g
\end{align*}
\]

(A-1a) \hspace{2cm} (A-1b) \hspace{2cm} (A-1c)

Figure A-1: Distribution feeder schematic with substation and the first three nodes.

Master of Science Thesis
Oscar Sondermeijer
The difference in reactive power flow as a result of injecting reactive power is expressed as:

\[ v_{3,y} - v_{3,n} = v_{2,y} - 2 (r_3 P_3 + x_3 Q_{3,y}) - \left( r_3^2 + x_3^2 \right) \left( \frac{P_{3,y}^2 + Q_{3,y}^2}{v_{3,y}} \right) - \left( v_{2,n} - 2 (r_3 P_3 + x_3 Q_{3,n}) - \left( r_3^2 + x_3^2 \right) \left( \frac{P_{3,n}^2 + Q_{3,n}^2}{v_{3,n}} \right) \right) \]  

(A-2)

\[ = (v_{2,y} - v_{2,n}) - 2 x_3 (Q_{3,y} - Q_{3,n}) - \left( r_3^2 + x_3^2 \right) \left( \frac{Q_{3,y}^2}{v_{3,y}} - \frac{Q_{3,n}^2}{v_{3,n}} \right) \]

where similarly,

\[ v_{2,y} - v_{2,n} = (v_{1,y} - v_{1,n}) - 2 x_2 (Q_{2,y} - Q_{2,n}) - \left( r_2^2 + x_2^2 \right) \left( \frac{Q_{2,y}^2}{v_{2,y}} - \frac{Q_{2,n}^2}{v_{2,n}} \right) \]  

(A-3)

\[ v_{1,y} - v_{1,n} = (v_{0,y} - v_{0,n}) - 2 x_1 (Q_{1,y} - Q_{1,n}) - \left( r_1^2 + x_1^2 \right) \left( \frac{Q_{1,y}^2}{v_{1,y}} - \frac{Q_{1,n}^2}{v_{1,n}} \right) \]

and by definition,

\[ v_{0,y} - v_{0,n} = 0. \]  

(A-4)

Recall that the second term in the voltage equation of (A-1) accounts for the voltage drop due to loss on the line. The third term compensates for the progressive loss on the line; therefore, assume \( 2 x_1 (Q_{1,y} - Q_{1,n}) \gg (r_1^2 + x_1^2) \left( \frac{Q_{1,y}^2}{v_{1,y}} - \frac{Q_{1,n}^2}{v_{1,n}} \right) \), which gives the following approximation of the voltage difference.

\[ v_{3,y} - v_{3,n} = - 2 x_1 (Q_{1,y} - Q_{1,n}) - 2 x_2 (Q_{2,y} - Q_{2,n}) - 2 x_3 (Q_{3,y} - Q_{3,n}) \]  

(A-5)

The difference in reactive power flow as a result of injecting reactive power is expressed as:

\[ Q_{3,y} - Q_{3,n} = Q_3 + x_3 \left( \frac{P_3 + Q_{3,y}}{v_{2,y}} \right) + q_3^c - q_3^g \]

\[ - \left( Q_3 + x_3 \left( \frac{P_3 + Q_{3,n}}{v_{2,n}} \right) + q_3^c \right) \]

\[ = x_3 \left( \frac{Q_{3,y}^2}{v_{2,y}} - \frac{Q_{3,n}^2}{v_{2,n}} \right) - q_3^g \]  

(A-6)

similarly,

\[ Q_{2,y} - Q_{2,n} = (Q_{3,y} - Q_{3,n}) + x_2 \left( \frac{Q_{2,y}^2}{v_{1,y}} - \frac{Q_{2,n}^2}{v_{1,n}} \right) \]

\[ Q_{1,y} - Q_{1,n} = (Q_{2,y} - Q_{2,n}) + x_1 \left( \frac{Q_{1,y}^2}{v_{0,y}} - \frac{Q_{1,n}^2}{v_{0,n}} \right) \]  

(A-7)

Again neglect the term \( x_1 \left( \frac{Q_{0,y}^2}{v_{1,y}} - \frac{Q_{0,n}^2}{v_{1,n}} \right) \), and recursively use the result of (A-6) to write (A-8), and substitute in (A-5) to obtain (A-9).

\[ Q_{3,y} - Q_{3,n} = Q_{2,y} - Q_{2,n} = Q_{1,y} - Q_{1,n} = -q_3^g \]  

(A-8)

\[ v_{3,y} - v_{3,n} = 2 (x_1 + x_2 + x_3) q_3^g \]  

(A-9)

More generally, this intuition shows that the voltage effect of reactive power injection is a product with the reactance between the feeder head and the inverter node. This result has been empirically verified in simulation.

Oscar Sondermeijer  
Master of Science Thesis
Regression-based control in MATLAB

The regression-based control framework is described in Chapter 3. The summary that was described in Chapter 3-1 is repeated for reference:

1. Obtain a balanced feeder model, and collect $N$ historical data points of consumption and generation at each load in the network. The information can be collected with advanced metering infrastructure (AMI) over an extended period of time.

2. Solve the convex optimal power flow problem for all $N$ scenarios. The solutions contain globally optimal reactive power output for all inverters and all $N$ power flow scenarios. Details in Section 3-2.

3. Perform regression analysis for each inverter to obtain the function that most accurately predicts its optimal reactive power output with its historical local information only. Details in Section 3-3.

4. Employ the functions as local inverter controllers in a simulation environment to determine the reactive power output based on new measurements.

A formulation of the convex optimal power flow problem that can be solved with the Convex Optimization Toolbox CVX [47] for MATLAB was suggested in Chapter 3-2-3.

The summary of the proposed approach can be further tailored for simulation experiments in Matlab.

1. Obtain a balanced feeder model:
   
   (a) If line impedances are multi phase values, construct one-line values with (2-5).
   
   (b) Express network topology as interconnection matrix $A$, use same indexing for impedances.

2. Construct optimization problem:
(a) Construct objective function (3-5) and select trade-off variable $\gamma$.

(b) Construct branch flow equality constraint (3-6) with the interconnection matrix, impedance data, PV generation data, real and reactive power consumption, and substation voltage.

(c) Construct additional constraints for voltage (3-7), inverter capacity (3-8), and SOCP inequality (3-9).

3. Solve the optimization problem with CVX toolbox [47, 48], store solution, update the branch flow equality constraint with demand from another time step, and repeat.

4. Perform regression for each inverter:

   (a) Construct regression input and output data matrices (3-10) and (3-11), and split into training and validation set (70% and 30%).

   (b) For each inverter: perform stepwise regression on training data for all inverters.

   (c) Predict inverter reactive power output for validation data.

5. Use the predicted reactive power output and simulate power flow problem with CVX toolbox.

6. Compare optimal objective function values with objective function value of regression-based control simulation to evaluate performance.
Appendix C

Feeder models

The tables in this appendix contain the network data of the IEEE feeder (Table C-1) and the SB feeder (Table C-2 and Figure C-1). Column B denotes the node number, and column A denotes the first upstream node, i.e. power typically flows from A to B. Column \( p \) and \( q \) denote the sample demand at node B, and \( r \) and \( x \) denote the resistance and impedance of the line between A and B. The complete IEEE feeder data is available through [9].

Table C-1: Network data of the IEEE feeder.

<table>
<thead>
<tr>
<th>B</th>
<th>A</th>
<th>( p ) kW</th>
<th>( q ) kVAR</th>
<th>( r ) ( \Omega )</th>
<th>( x ) ( \Omega )</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>650</td>
<td>650</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>632</td>
<td>650</td>
<td>100</td>
<td>58</td>
<td>0.0705</td>
<td>0.2261</td>
<td>0</td>
</tr>
<tr>
<td>633</td>
<td>632</td>
<td>0</td>
<td>0</td>
<td>0.0561</td>
<td>0.072</td>
<td>0</td>
</tr>
<tr>
<td>634</td>
<td>632</td>
<td>400</td>
<td>290</td>
<td>0.0017</td>
<td>0.0017</td>
<td>1</td>
</tr>
<tr>
<td>645</td>
<td>632</td>
<td>170</td>
<td>125</td>
<td>0.1061</td>
<td>0.0798</td>
<td>1</td>
</tr>
<tr>
<td>646</td>
<td>633</td>
<td>230</td>
<td>132</td>
<td>0.0636</td>
<td>0.0479</td>
<td>0</td>
</tr>
<tr>
<td>671</td>
<td>645</td>
<td>1255</td>
<td>718</td>
<td>0.0705</td>
<td>0.2261</td>
<td>1</td>
</tr>
<tr>
<td>684</td>
<td>671</td>
<td>0</td>
<td>0</td>
<td>0.0636</td>
<td>0.0507</td>
<td>0</td>
</tr>
<tr>
<td>652</td>
<td>671</td>
<td>128</td>
<td>86</td>
<td>0.2034</td>
<td>0.0776</td>
<td>0</td>
</tr>
<tr>
<td>680</td>
<td>671</td>
<td>0</td>
<td>0</td>
<td>0.0352</td>
<td>0.113</td>
<td>0</td>
</tr>
<tr>
<td>611</td>
<td>684</td>
<td>170</td>
<td>80</td>
<td>0.0755</td>
<td>0.0766</td>
<td>0</td>
</tr>
<tr>
<td>692</td>
<td>684</td>
<td>170</td>
<td>151</td>
<td>0.0017</td>
<td>0.0017</td>
<td>0</td>
</tr>
<tr>
<td>675</td>
<td>692</td>
<td>843</td>
<td>462</td>
<td>0.0462</td>
<td>0.0393</td>
<td>1</td>
</tr>
</tbody>
</table>

\(^1\)When there is no reverse power flow
Table C-2: Network data of the SB feeder.

<table>
<thead>
<tr>
<th>B</th>
<th>A</th>
<th>p kW</th>
<th>q kVar</th>
<th>r Ω</th>
<th>x Ω</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td>0.000</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.005</td>
<td>0.000</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td>0.005</td>
<td>46</td>
<td>5.6</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>13.1</td>
<td>0.003</td>
<td>0.015</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td>0.005</td>
<td>48</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0</td>
<td>0.000</td>
<td>0.003</td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0.003</td>
<td>0.013</td>
<td>50</td>
<td>47</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>0</td>
<td>0.000</td>
<td>0.003</td>
<td>51</td>
<td>47</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>0</td>
<td>0.008</td>
<td>0.008</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>0</td>
<td>0.028</td>
<td>0.033</td>
<td>53</td>
<td>48</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>0</td>
<td>0.015</td>
<td>0.015</td>
<td>54</td>
<td>49</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>0</td>
<td>0.005</td>
<td>0.005</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>0</td>
<td>0.005</td>
<td>0.003</td>
<td>56</td>
<td>50</td>
</tr>
<tr>
<td>14</td>
<td>12</td>
<td>0</td>
<td>0.005</td>
<td>0.003</td>
<td>57</td>
<td>51</td>
</tr>
<tr>
<td>15</td>
<td>13</td>
<td>111</td>
<td>46.5</td>
<td>0.013</td>
<td>0.005</td>
<td>58</td>
</tr>
<tr>
<td>16</td>
<td>14</td>
<td>0</td>
<td>0.003</td>
<td>0.018</td>
<td>59</td>
<td>54</td>
</tr>
<tr>
<td>17</td>
<td>16</td>
<td>0</td>
<td>0.000</td>
<td>0.003</td>
<td>60</td>
<td>55</td>
</tr>
<tr>
<td>18</td>
<td>16</td>
<td>0</td>
<td>0.000</td>
<td>0.003</td>
<td>61</td>
<td>56</td>
</tr>
<tr>
<td>19</td>
<td>18</td>
<td>0</td>
<td>0.005</td>
<td>0.018</td>
<td>62</td>
<td>57</td>
</tr>
<tr>
<td>20</td>
<td>19</td>
<td>13.9</td>
<td>5.8</td>
<td>0.000</td>
<td>0.003</td>
<td>63</td>
</tr>
<tr>
<td>21</td>
<td>20</td>
<td>0</td>
<td>0.010</td>
<td>0.040</td>
<td>64</td>
<td>59</td>
</tr>
<tr>
<td>22</td>
<td>21</td>
<td>0</td>
<td>0.015</td>
<td>0.068</td>
<td>65</td>
<td>60</td>
</tr>
<tr>
<td>23</td>
<td>22</td>
<td>3.7</td>
<td>1.5</td>
<td>0.000</td>
<td>0.003</td>
<td>66</td>
</tr>
<tr>
<td>24</td>
<td>22</td>
<td>0</td>
<td>0.030</td>
<td>0.138</td>
<td>67</td>
<td>63</td>
</tr>
<tr>
<td>25</td>
<td>23</td>
<td>0</td>
<td>0.023</td>
<td>0.015</td>
<td>68</td>
<td>64</td>
</tr>
<tr>
<td>26</td>
<td>24</td>
<td>9.3</td>
<td>3.9</td>
<td>0.238</td>
<td>0.148</td>
<td>69</td>
</tr>
<tr>
<td>27</td>
<td>24</td>
<td>3.7</td>
<td>1.5</td>
<td>0.198</td>
<td>0.123</td>
<td>70</td>
</tr>
<tr>
<td>28</td>
<td>25</td>
<td>0</td>
<td>0.013</td>
<td>0.055</td>
<td>71</td>
<td>67</td>
</tr>
<tr>
<td>29</td>
<td>26</td>
<td>5.6</td>
<td>2.3</td>
<td>0.010</td>
<td>0.050</td>
<td>72</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>9.3</td>
<td>3.9</td>
<td>0.008</td>
<td>0.035</td>
<td>73</td>
</tr>
<tr>
<td>31</td>
<td>28</td>
<td>0</td>
<td>0.003</td>
<td>0.010</td>
<td>74</td>
<td>70</td>
</tr>
<tr>
<td>32</td>
<td>29</td>
<td>0</td>
<td>0.003</td>
<td>0.005</td>
<td>75</td>
<td>71</td>
</tr>
<tr>
<td>33</td>
<td>30</td>
<td>3.7</td>
<td>1.5</td>
<td>0.018</td>
<td>0.078</td>
<td>76</td>
</tr>
<tr>
<td>34</td>
<td>31</td>
<td>0</td>
<td>0.030</td>
<td>0.075</td>
<td>77</td>
<td>74</td>
</tr>
<tr>
<td>35</td>
<td>32</td>
<td>0</td>
<td>0.003</td>
<td>0.005</td>
<td>78</td>
<td>75</td>
</tr>
<tr>
<td>36</td>
<td>33</td>
<td>0</td>
<td>0.020</td>
<td>0.013</td>
<td>79</td>
<td>76</td>
</tr>
<tr>
<td>37</td>
<td>34</td>
<td>0</td>
<td>0.020</td>
<td>0.013</td>
<td>80</td>
<td>77</td>
</tr>
<tr>
<td>38</td>
<td>35</td>
<td>18.5</td>
<td>7.7</td>
<td>0.100</td>
<td>0.063</td>
<td>81</td>
</tr>
<tr>
<td>39</td>
<td>35</td>
<td>13.9</td>
<td>5.8</td>
<td>0.275</td>
<td>0.170</td>
<td>82</td>
</tr>
<tr>
<td>40</td>
<td>36</td>
<td>13.9</td>
<td>5.8</td>
<td>0.228</td>
<td>0.140</td>
<td>83</td>
</tr>
<tr>
<td>41</td>
<td>37</td>
<td>13.9</td>
<td>5.8</td>
<td>0.288</td>
<td>0.178</td>
<td>84</td>
</tr>
<tr>
<td>42</td>
<td>38</td>
<td>5.6</td>
<td>2.3</td>
<td>0.263</td>
<td>0.163</td>
<td>85</td>
</tr>
<tr>
<td>43</td>
<td>39</td>
<td>0</td>
<td>0.025</td>
<td>0.065</td>
<td>86</td>
<td>84</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>p kW</td>
<td>q kVar</td>
<td>r Ω</td>
<td>x Ω</td>
<td>PV</td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>87</td>
<td>85</td>
<td>0</td>
<td>0</td>
<td>0.183</td>
<td>0.113</td>
<td>0</td>
</tr>
<tr>
<td>88</td>
<td>86</td>
<td>9.3</td>
<td>3.9</td>
<td>0.145</td>
<td>0.088</td>
<td>0</td>
</tr>
<tr>
<td>89</td>
<td>87</td>
<td>9.3</td>
<td>3.9</td>
<td>0.235</td>
<td>0.138</td>
<td>0</td>
</tr>
<tr>
<td>90</td>
<td>87</td>
<td>9.3</td>
<td>3.9</td>
<td>0.123</td>
<td>0.073</td>
<td>1</td>
</tr>
<tr>
<td>91</td>
<td>87</td>
<td>9.3</td>
<td>3.9</td>
<td>0.218</td>
<td>0.128</td>
<td>1</td>
</tr>
<tr>
<td>92</td>
<td>88</td>
<td>0</td>
<td>0</td>
<td>0.128</td>
<td>0.078</td>
<td>0</td>
</tr>
<tr>
<td>93</td>
<td>89</td>
<td>9.3</td>
<td>3.9</td>
<td>0.233</td>
<td>0.145</td>
<td>1</td>
</tr>
<tr>
<td>94</td>
<td>90</td>
<td>9.3</td>
<td>3.9</td>
<td>0.230</td>
<td>0.143</td>
<td>1</td>
</tr>
<tr>
<td>95</td>
<td>92</td>
<td>0</td>
<td>0</td>
<td>0.135</td>
<td>0.085</td>
<td>0</td>
</tr>
<tr>
<td>96</td>
<td>93</td>
<td>9.3</td>
<td>3.9</td>
<td>0.185</td>
<td>0.105</td>
<td>0</td>
</tr>
<tr>
<td>97</td>
<td>94</td>
<td>9.3</td>
<td>3.9</td>
<td>0.170</td>
<td>0.105</td>
<td>0</td>
</tr>
<tr>
<td>98</td>
<td>95</td>
<td>0</td>
<td>0</td>
<td>0.020</td>
<td>0.013</td>
<td>0</td>
</tr>
<tr>
<td>99</td>
<td>97</td>
<td>0</td>
<td>0</td>
<td>0.175</td>
<td>0.108</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>98</td>
<td>9.3</td>
<td>3.9</td>
<td>0.163</td>
<td>0.095</td>
<td>0</td>
</tr>
<tr>
<td>101</td>
<td>99</td>
<td>13.9</td>
<td>5.8</td>
<td>0.223</td>
<td>0.133</td>
<td>1</td>
</tr>
<tr>
<td>102</td>
<td>99</td>
<td>13.9</td>
<td>5.8</td>
<td>0.240</td>
<td>0.143</td>
<td>0</td>
</tr>
<tr>
<td>103</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td>0.375</td>
<td>0.233</td>
<td>0</td>
</tr>
<tr>
<td>104</td>
<td>100</td>
<td>13.9</td>
<td>5.8</td>
<td>0.155</td>
<td>0.095</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101</td>
<td>9.3</td>
<td>3.9</td>
<td>0.240</td>
<td>0.148</td>
<td>1</td>
</tr>
<tr>
<td>106</td>
<td>103</td>
<td>13.9</td>
<td>5.8</td>
<td>0.223</td>
<td>0.138</td>
<td>0</td>
</tr>
<tr>
<td>107</td>
<td>103</td>
<td>13.9</td>
<td>5.8</td>
<td>0.240</td>
<td>0.143</td>
<td>0</td>
</tr>
<tr>
<td>108</td>
<td>104</td>
<td>5.6</td>
<td>2.3</td>
<td>0.263</td>
<td>0.155</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure C-1: SB feeder diagram with inverter and load locations.
Bibliography


Master of Science Thesis Oscar Sondermeijer


## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>AMI</td>
<td>advanced metering infrastructure</td>
</tr>
<tr>
<td>AC</td>
<td>alternating current</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian information criterion</td>
</tr>
<tr>
<td>CAISO</td>
<td>California independent system operator</td>
</tr>
<tr>
<td>CVR</td>
<td>conservation voltage reduction</td>
</tr>
<tr>
<td>DC</td>
<td>direct current</td>
</tr>
<tr>
<td>DSO</td>
<td>distribution system operator</td>
</tr>
<tr>
<td>lasso</td>
<td>least absolute shrinkage and selection operator</td>
</tr>
<tr>
<td>OPF</td>
<td>optimal power flow</td>
</tr>
<tr>
<td>PV</td>
<td>photovoltaic</td>
</tr>
<tr>
<td>RSS</td>
<td>residual sum of squares</td>
</tr>
<tr>
<td>SCADA</td>
<td>supervisory control and data acquisition</td>
</tr>
<tr>
<td>SDP</td>
<td>semidefinite programming</td>
</tr>
<tr>
<td>SOCP</td>
<td>second-order cone programming</td>
</tr>
<tr>
<td>U.S.</td>
<td>United States of America</td>
</tr>
<tr>
<td>VVO</td>
<td>Volt-VAr optimization</td>
</tr>
</tbody>
</table>
## Glossary

### List of Symbols

**Power flow**

- $i, j, k$: Indices of nodes
- $\ell$: Squared current magnitude
- $p$: Real power demand
- $q$: Reactive power demand
- $r$: Resistance
- $s$: Apparent power
- $v$: Squared voltage magnitude
- $z$: Complex impedance
- $I$: Current phasor
- $J$: Objective function
- $P$: Real power line flow
- $Q$: Reactive power line flow
- $V$: Voltage phasor
- $0$: Null matrix
- $q^g$: Vector of $q^g$ of all inverters in $C$
- $u$: Vector of optimization variables
- $v$: Vector of $v$ at all nodes in $N$
- $A$: Interconnection matrix
- $C, D, E, F, G$: Auxiliary matrices
- $I$: Identity matrix
- $L$: Vector of $\ell$ into all nodes in $N$
- $P$: Vector of $P$ into all nodes in $N$
- $Q$: Vector of $Q$ into all nodes in $N$
- $C$: Set of nodes with inverter
- $E$: Set of edges
- $G(N, E)$: Graph defined by nodes $N$ and edges $E$
- $N$: Set of nodes
- $\gamma$: Tradeoff variable
- $(\cdot)^c$: Consumption of input $(\cdot)$
- $(\cdot)^g$: Generation of input $(\cdot)$
- $(\cdot)_{i,j,k}$: Indices of input $(\cdot)$
- $(\cdot)^{nom}$: Nominal value of input $(\cdot)$
- $(\cdot)^{ref}$: Reference value of input $(\cdot)$
- $(\cdot)^*$: Optimal value of input $(\cdot)$
- $(\cdot)^\star$: Maximum of input $(\cdot)$
- $(\cdot)^\star$: Minimum of input $(\cdot)$
- $\text{diag}(\cdot)$: Matrix with entries of input $(\cdot)$ on the diagonal
Regression

\( f (\cdot) \)  Function of input (\cdot)
\( k \)  Number of input variables
\( n \)  Number of observations
\( x_{i,j} \)  \( i \)th observation of \( j \)th input variable
\( x_i \)  Vector of \( i \)th observation of \( k \) input variables
\( y_i \)  \( i \)th observation of the output variable
\( X_j \)  \( j \)th input variable
\( x_j \)  Vector of \( n \) observations of \( j \)th input variable
\( X \)  Matrix of all \( x_{i,j} \)
\( y \)  Vector of all observations of the output variable
(\cdot)_{(i)}  Index of inverter
(\cdot)_{(i)}  Index of observations/samples
(\cdot)_{(j)}  Index of input variable