UNIVERSIDAD
DE MÁLAGA

ESCUELA DE
INGENIERÍAS
INDUSTRIALES

ESCUELA DE INGENIERÍAS INDUSTRIALES

Departamento de Ingeniería Eléctrica
Área de conocimiento: Ingeniería Eléctrica

PROYECTO/TRABAJO FIN DE MÁSTER

APROXIMACIÓN LINEAL BASADA EN DATOS DEL FLUJO DE CARGAS ÓPTIMO EN SISTEMAS ELÉCTRICOS

MÁSTER EN INGENIERÍA INDUSTRIAL

Autor: Manuel Garrido Martín
Tutor: Salvador Pineda Morente
Cotutor: Juan Miguel Morales González

Málaga, 8 de febrero de 2024

APROXIMACIÓN LINEAL BASADA EN DATOS DEL FLUJO DE CARGAS ÓPTIMO EN SISTEMAS ELÉCTRICOS

Manuel Garrido Martín

Agradecimientos

A mi familia, por su apoyo y paciencia incondicional, con mención especial a María, por su ayuda con las figuras de este trabajo.

A todo el Grupo OASYS, por hacer de cada almuerzo en la oficina toda una experiencia, y en especial a Salva y a Juanmi, por su insaciable espíritu investigador y comprensión, y a José, Conchi y Antonio, por todos esos cafés que se quieren alargar toda la mañana.

A Rocío, por demostrarme que cuando uno lo quiere lo suficiente, dos más dos suman cinco.

Gracias.

UNIVERSIDAD DE MÁLAGA

ESCUELA DE INGENIERÍAS INDUSTRIALES

RESUMEN

El Flujo de Cargas Óptimo (OPF) es un problema de optimización cuyo objetivo es determinar el despacho de potencia de las centrales generadoras de una red eléctrica para satisfacer la demanda al mínimo coste, cumpliendo al mismo tiempo las leyes físicas que rigen los flujos de potencia a lo largo de la red (Ley de Kirchoff). Adicionalmente, se asegura en su formulación el cumplimiento de los límites técnicos impuestos por las centrales generadoras y los límites de transmisión por temperatura en las líneas de transporte. Así, la relevancia del OPF radica en la determinación de un punto de operación de un sistema eléctrico, presente o futuro, de forma que facilite la toma de decisiones al planificador u operador de la red.

La formulación clásica de un OPF conlleva la resolución de un conjunto de ecuaciones no lineales, correspondiente a los balances de potencia, y un conjunto de inecuaciones, correspondientes a los límites de operación tanto de las centrales (lineales) como de las líneas de transporte (no lineales). Así, un OPF debe ser formulado matemáticamente como un problema de optimización no lineal, lo que dificulta su resolución en ámbitos como el control y operación de redes eléctricas en tiempo real. En este proyecto, se propone una familia de modelos basados en datos que utilizan el importante número de mediciones históricas recogidas por las redes energéticas inteligentes altamente monitorizadas, de forma que permitan realizar una linearización de las ecuaciones no lineales que caracterizan al OPF y así acelerar su resolución numérica. En particular, el objetivo consiste en la obtención de un modelo que establezca una relación lineal entre las potencias activas inyectadas en los nodos y los flujos de potencia activa que circulan por las líneas, permitiendo la sustitución de las complejas ecuaciones no lineales del OPF asociadas a las líneas de transporte por simples ecuaciones lineales.

Así, se pretende obtener un modelo lineal de resolución del problema OPF que ayude a operar las redes eléctricas de forma más eficiente, manteniendo un margen de error reducido y facilitando la incorporación de fuentes renovables de una manera rentable, fiable y segura.

Palabras clave: sistemas de energía eléctrica, flujo de cargas óptimo, programación matemática.

UNIVERSIDAD DE MÁLAGA

ESCUELA DE INGENIERÍAS INDUSTRIALES

ABSTRACT

The Optimal Power Flow (OPF) is an optimization problem that provides the power dispatch of the generating units of an electrical power system in order to satisfy the loads at minimum cost, fulfilling at the same time the physical laws that govern the electrical power flows through the net lines (Kirchhoff's Laws). Aditionally, its formulation ensures the compliance with both technical limits imposed by generating units and temperature limits present in transport lines. Thus, the relevance of the OPF lies in determining a feasible operating point of an electrical system, present or future, in such a way as to facilitate the decision-making process for the network planner or operator.

The classic OPF formulation implies the resolution of a set of non linear equations, corresponding to power balances, and a set of inequalities, corresponding to the operational limits for both the generating units (linear) and transport lines (non linear). Consequently, the OPF problem should be expressed as a non linear optimization problem, which challenges its resolution in areas such as control and operation of power systems in real time. In this work, a set of data based models, trained with a relevant number of historical measurements collected by highly monitored smart power systems, are used to obtain a linearization of the non linear equations that characterize the OPF formulation, accelerating thus its numerical resolution. In particular, the purpose is to obtain a model that establishes a linear relation between the injected power in the nodes and the power flows through the lines, allowing its use as a substitute for the complex non linear equations associated to the classic OPF.

The objective of this work is thus to obtain a linear model that allows the resolution of the Optimal Power Flow, improving the efficiency of the operation of power systems, maintaining a reduced margin of error and enabling the implementation of renewable energy sources to achieve more efficient, secure and sustainable energy systems.

Key words: electric power systems, optimal power flow, mathematical programming.

Contents

List of Figures

Página

List of Tables

Página

Chapter 1

Introduction

1.1 Optimal Power Flow

A power system, also known as an electrical grid, is a complex network of interconnected components designed to generate, transmit, and distribute electrical energy to consumers. It consists of power plants, transformers, transmission lines, distribution lines, substations, and various control and protection devices. All power systems must be balanced, meaning that the electricity supply and demand are equal and in equilibrium. This balance between generation and consumption is critical for maintaining the stability and reliability of the power system. However, in recent years, this task has become increasingly more challenging, as the penetration of renewable energy sources introduces a high level of variability and uncertainty into the power generation profile of the power systems.

The suboptimal operation of electric power grids results in annual losses amounting to billions of dollars [1]. Global estimates of carbon dioxide equivalents attributed to transmission and distribution losses reach one billion metric tons annually [2]. In this context, it is imperative to research on the optimal operation of grid assets and leverage accurate, physically realizable mathematical models that reduce the financial and environmental impact of grid operations.

The Optimal Power Flow (OPF) is a fundamental mathematical tool for the operation of current power systems. The first formulation of the Optimal Power Flow dates back to 1962, when J. Carpentier extended the Economic Dispatch by including the electric Power Flow equations [3]. Currently, there is a vast number of articles and academic texts that discuss diverse mathematical OPF formulations. These encompass the polar power-voltage formulation, the rectangular power-voltage formulation, and the rectangular current-voltage formulation.

As depicted in Figure 1.1, the optimization of power system operation is commonly carried out through incremental planning strategies. Long-term planning procedures involve making high-level decisions based on simplified system models that provide a coarse representation of the system. On the other hand, short-term procedures refine the earlier decisions by utilizing more detailed models, albeit with more restricted decision spaces. Optimal Power Flow can be applied to decision-making processes across various planning horizons, ranging from long-term planning for transmission network capacity to minute-by-minute adjustments of real and reactive power dispatch.

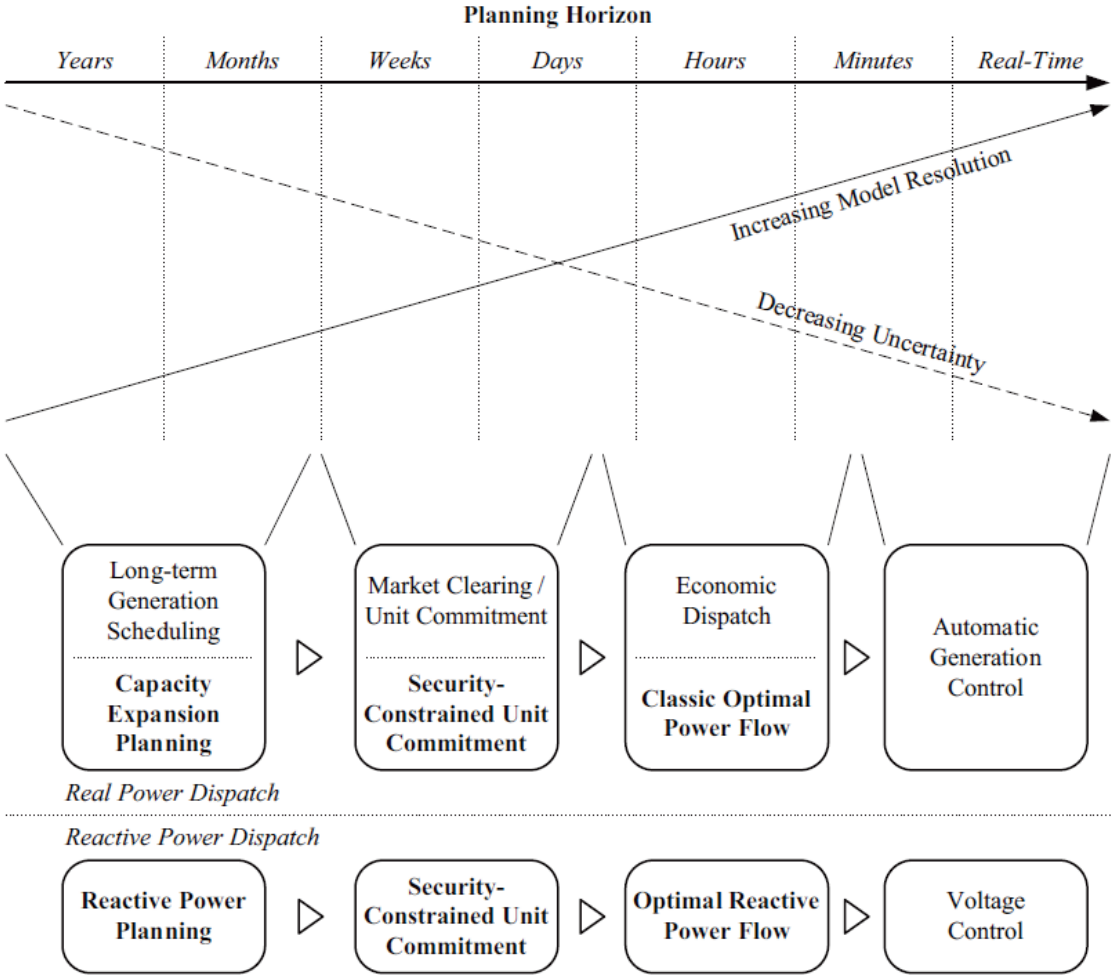


Figure 1.1: Optimization and control procedures for incremental planning of power system operation. Bold text indicates procedures that incorporate variants of Optimal Power Flow. Figure extracted from [4].

The Optimal Power Flow problem aims at determining the optimal operating levels of generating units to satisfy the electricity demand while complying with network constraints, usually with the objective of minimizing operating cost. For this purpose, the OPF problem integrates an objective function with the AC Power

Flow (AC-PF) equations to create an optimization problem. The objective function in an optimization problem serves as a quantifiable measure guiding the search for optimal solutions through the feasible space by defining the goal to be either maximized or minimized. On the other hand, the feasible or solution space can be defined as the set of points where an optimization problem presents a solution and all restrictions are met. In the present case of the Optimal Power Flow problem, the feasible space would encompass the set of points where the power flow equations have a solution and all system values (e.g. active and reactive generation levels, line power flows, bus voltages) are within their operational limits.

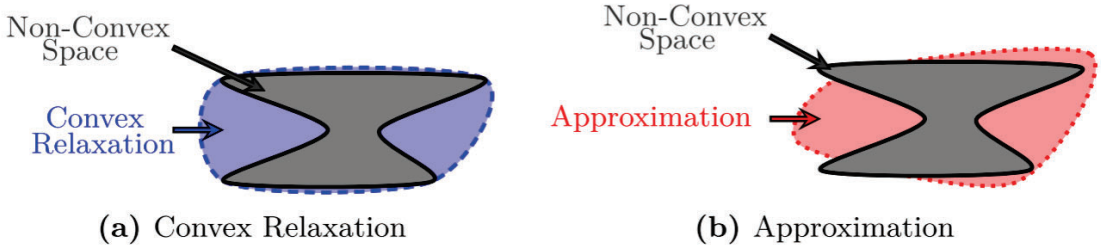
The inclusion of the AC-PF equations, which model the flow of electric power through a network of interconnected power systems, convert the AC Optimal Power Flow (AC-OPF) in a non-convex optimization problem. This non-convexity leads to multiple local optima in the solution space, making it challenging to find the global optimal solution using traditional optimization techniques. Furthermore, because of these non-convexities, the AC-OPF formulation constitutes a NP-hard problem, a type of computational problem that belongs to a class of challenges known for their high level of computational complexity. These problems are associated with the NP (Non-deterministic Polynomial-time) class and are characterized by the property that if a polynomial-time solution algorithm were to be found for any one of them, it could be used to efficiently solve all problems in the NP class. Proof of the strong NP-hardness of AC-OPF can be found in [5], while reference [6] encapsulates an extensive survey of the Optimal Power Flow problem spanning the past half-century and different approaches for reliably finding solutions. Among all of them, many non-linear optimization algorithms arise, including gradient methods, Newton-type methods, sequential linear programming, sequential quadratic programming both linear and nonlinear interior methods and semi-definite programming.

1.2 Relaxations and approximations

The non-linearity inherent in the conventional AC-PF equations present in the AC-OPF formulation poses computational challenges in various optimization and control problems. Despite providing accurate results, the non-linearity of the AC-PF equations introduces difficulties in achieving numerical convergence. These drawbacks restrict the widespread application of the AC-PF equations in system optimization, particularly for large-scale systems.

There are several alternative representations of the Power Flow equations that attempt to deal with the aforementioned non-linearity and non-convexity of the original AC-OPF problem. These can be categorized as either relaxations or approximations. Figure 1.2 shows conceptual examples of a relaxation and an approximation

of a non-convex feasible space.



(a) Convex Relaxation **(b)** Approximation

Figure 1.2: Conceptual illustrations showing a (a) convex relaxation (blue region) and (b) an approximation (red region) for the gray non-convex space. Figure extracted from [7].

Relaxations encompass the non-convex feasible spaces associated with the power flow equations within a larger space. This larger space is typically chosen to be convex, facilitating the application of theory and algorithms developed for convex optimization problems. Approximations, on the other hand, involve making assumptions about certain quantities to simplify the Power Flow equations. Power flow approximations can accurately represent system behavior when the associated assumptions are valid. Many power flow approximations are reasonably accurate for typical operating conditions.

In general, solutions to optimization problems that utilize the Power Flow relaxations and approximations do not precisely satisfy the actual power flow equations. Instead, these relaxations and approximations are typically employed to obtain tractable formulations that adequately represent the actual power flow physics. Optimization problems using convex relaxations additionally provide bounds on the optimal objective value for the original non-convex problem, as well as sufficient conditions for certifying problem infeasibility. Some convex relaxations also have associated sufficient conditions guaranteeing their ability to provide global optima for certain limited classes of power system optimization problems. Some of these sufficient conditions can be evaluated before solving the relaxation based solely on the problem parameters and network topology, while other conditions are checked after solving a relaxation. In contrast, it is important to note that approximations do not provide any of the aforementioned theoretical guarantees offered by relaxations.

One of the most common alternative representations of the Power Flow equations that attempt to deal with the aforementioned non-convexity of the original AC-OPF problem is the DC-OPF approximation. The DC-OPF formulation is a physics-driven (also known as model-driven) linearization technique that linearize the AC Optimal Power Flow model by utilizing assumptions and approximations derived from physical knowledge concerning voltage magnitudes, voltage phase angles, and line parameters. The assumptions (e.g., neglecting line losses, reactive power, fixed voltage magnitudes, and assuming small phase angle differences, presented in detail

in Section 2.2) result in computational advantages that enable operators to solve the AC-OPF problem in real-time, typically within intervals of 5-15 minutes. In this context, we assess various networks across different loading scenarios, comparing the time required to solve both the AC-OPF and DC-OPF problems associated in Figure 1.3:

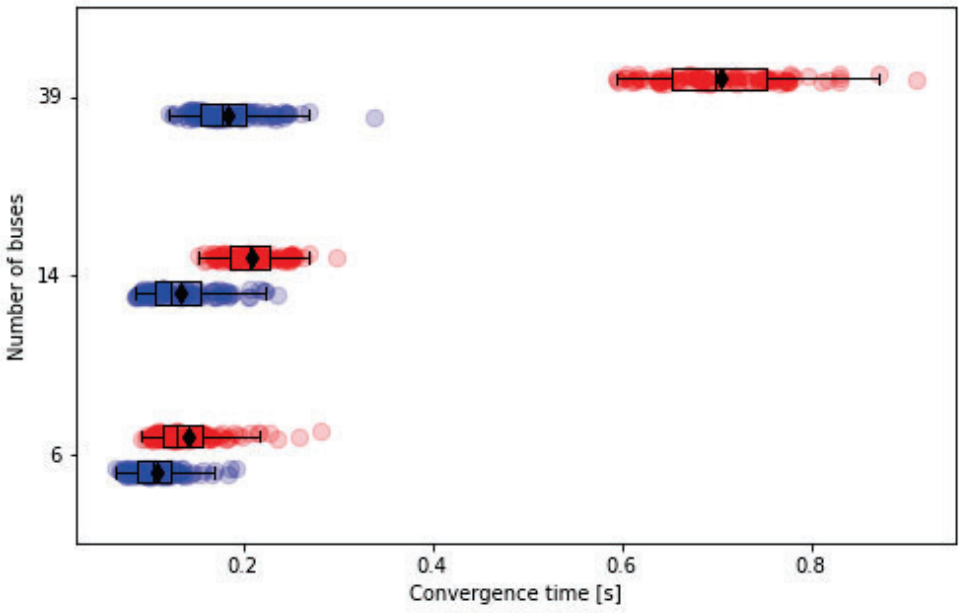


Figure 1.3: Comparison of the computational time needed for the resolution of 100 cases of DC-OPF (blue) and AC-OPF (red) for three different networks with 6, 14 and 39 buses.

The computational convergence time is plotted in the horizontal axis, while the different OPF formulations and power systems are ordered in the vertical axis. The black diamonds represent the average value of each of the variables presented, while the box and whisker plots help to understand the distribution of the data. As it can be observed, the difference in convergence time between the AC-OPF and DC-OPF formulations increases for networks with a higher number of buses, closer to the practical cases that system operators must solve in real time.

Nevertheless, the main advantage that linear formulations such as DC-OPF provide are the transformation of the NP-hard optimization problem into a P-hard optimization problem, which can be solved in polinomial time, and ensuring that the optimal solutions found are indeed global optimal points. A simple representation of a function with multiple optimal points is presented in Figure 1.4.

However, while the linearization of AC Power Flow equations may provide computational advantages, it introduces approximation and model errors into power flow

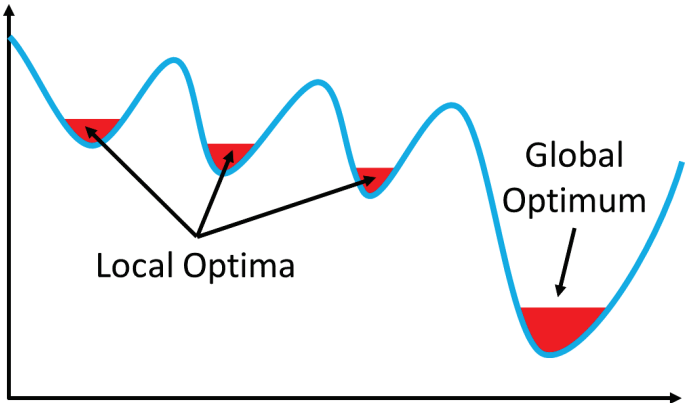


Figure 1.4: Representation of multiple local optimal points and the global optimum of a simple one-dimensional function.

solutions, which will result in the infeasibility of the solution for the real physical system. Among its deficiencies, it is important to remark the DC-OPF formulation does not account for network losses, reactive power or voltage constraints [8]. While the DC approximation may be reasonable for some networks, technical constraints such as voltage constraints become relevant especially in heavily congested systems as they implicitly limit power flow way below the available transmission capacity. As a result, traditional linear power flow models may be less reliable and more prone to breakdown in future power grids.

1.3 Data-based Power Flow Linearization

Another alternative to deal with the non-convexity of AC-OPF is to leverage data-based power flow linearization techniques, with the objective of deriving linear expressions for branch power flow equations. Data-driven Power Flow Linearization (DPFL) methods operate without prior physical knowledge of the power grid [9]. These methods rely solely on data extracted from the system to train linear models. Machine learning approaches, a subset of DPFL methods, shift computing complexity offline, enabling the discovery of near-optimal AC Optimal Power Flow solutions in real time. This facilitates rapid solving of the AC-OPF problem, allowing operators to economically manage larger systems with more decision variables.

Various factors drive the research and utilization of Data-Driven Power Flow Linearization techniques. Some of these motivations are outlined in this section:

1. **Data availability**: The fast evolution and extensive application of advanced data measuring devices has led to the availability of abundant, high-fidelity,

high-resolution, and synchronized electrical measurements. This abundance of data facilitates data-driven training for linear power flow models.

2. **High accuracy**: Recent studies indicate that DPFL models generally exhibit higher accuracy compared to the linear models derived from state-of-the-art physics-driven approaches. This can be attributed to the following reasons:

   - Assumption-free: Physical assumptions, e.g. $cos\theta_{ij} \approx 1$, $sin\theta_{ij} \approx \theta_{ij}$, and $V_i \approx 1$, are commonly used to linearize nonlinear terms in power flow formulas. However, these assumptions may not always be valid in practice, resulting in noticeable errors. On the other hand, DPFL models are learned from historical measurements that strictly adhere to physical principles. As a result, DPFL methods are free from errors caused by inappropriate assumptions.
   - State-independent: Among existing physics-driven approaches, Taylor-expansion-based algorithms are the most common. These methods use Taylor expansions to approximate the AC manifold by constructing a tangent hyperplane around a chosen point. The resulting model tends to be accurate when the chosen point is near the real but unknown operating point, but may be less accurate when it is far from the operating point. In contrast, DPFL methods do not rely on any specific point (a.k.a., system state). Instead of using tangent hyperplanes, DPFL methods seek to find a secant hyperplane that uniformly approximates the AC manifold as closely as possible. As a result, DPFL approaches enjoy higher accuracy on average, as it can be observed in Figure 1.5.

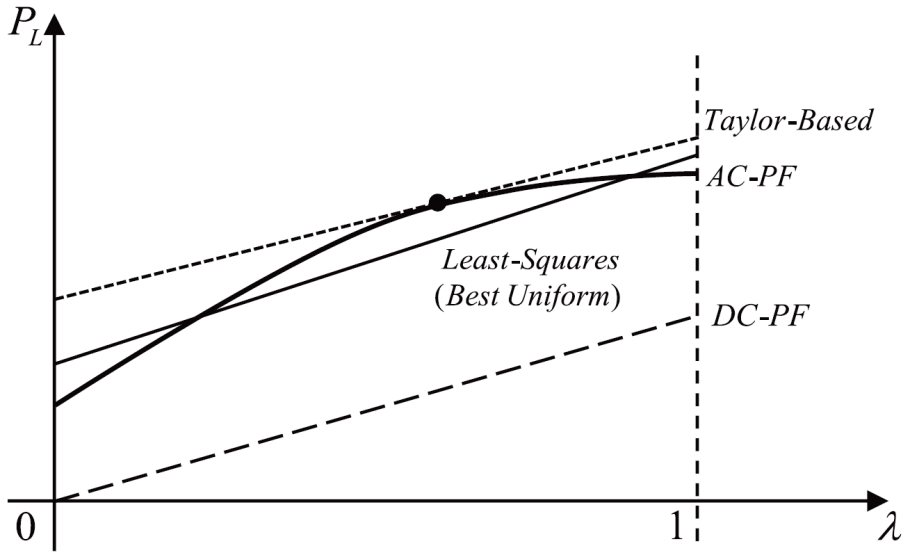


Figure 1.5: Conceptual example of different power flow linearization methods. Figure extracted from [10].

On the vertical axis, the active power flow of a certain branch is plotted

against the factor λ on the horizontal axis, defined as a proportional factor that regulates the power demand on the branch, limiting this value between its lower and upper operational limit. The DC Power Flow yields a discernible linear trend. On the other hand, the Taylor-based method produces a tangent line, with the accuracy of this approximation dependent on the choice of the expansion point, represented by the black dot. In contrast, the least-squares or the best uniform linear approximation results in a secant line, which can effectively model the power flow over a broad range of conditions.

- System-customized: In physics-driven approaches, the objective is to handle all systems in a unified manner. However, this approach inevitably applies the same assumptions to every system of interest. On the contrary, DPFL methods customize the linear model of each system based on the operational data of that specific system, resulting in higher accuracy tailored to that particular system.

3. **Wide applicability**: Implementing physics-driven linearization algorithms requires perfect knowledge of the system, including details such as the grid topology and line parameters. However, obtaining such perfect physical knowledge is challenging, especially in low-voltage distribution grids without adequate measuring devices. This challenge arises due to several factors:

   - Dynamic system topologies: Real-time tracking of system topologies is difficult because of frequent changes in devices such as lines and Distributed Energy Resources (DERs).
   - Deviation in line parameters: Actual line parameters can deviate from default values over time due to aging and temperature variations.
   - Unknown control rules of DERs: The control rules of individually-owned DERs are often unknown to operators.

   In contrast, DPFL methods do not require perfect knowledge of the power system. Instead, they only rely on historical measurements of the system, which are readily accessible to operators. Most importantly, DPFL methods implicitly incorporate the effects of atmospheric conditions and aging issues present in the measurements. Therefore, even when system parameters are unavailable, DPFL methods remain applicable and provide a practical alternative.

In spite of all the advantages and application opportunities stated above, research in DPFL techniques encounters several challenges, which can be categorized as either data-related or power-system-related. Data-related challenges include issues such as data multicollinearity, outliers, measurement noise, temporal correlation among observations, and asynchronous data. On the other hand, power-system-related

challenges are linked to the operation of power grids and AC Power Flow models. These challenges encompass frequent topological variations, variations in bus types, limited observability of distribution grids, the inherent nonlinearity of the AC Power Flow model, and the integration of physical knowledge into the models. A thorough analysis on the state of the art of DPFL can be found in [11].

For all of these reasons, there is a critical need to develop linearization methods that can effectively handle systems with large perturbations while maintaining satisfactory approximation accuracy. Additional difficulties are introduced by heightened levels of variable and uncertain power generation from renewable energy sources, which require faster decision making during the operation of grid assets.

1.4 Objectives

The main purpose of this Final Master Thesis is to develop and evaluate the numerical performance of a proposed data-based linear least squares regression model that enables an alternative resolution to the classical non-convex and non-linear AC-OPF formulation. For a brief explanation on the fundamentals of least squares regression, see Appendix B. The proposed model, Linear-OPF, is presented in two variants:

- **Linear$_{PF}$-OPF**, a linear model trained with data obtained from the resolution of multiple AC-PF problems.
- **Linear$_{OPF}$-OPF**, a linear model trained with data obtained from the resolution of multiple AC-OPF problems.

Both Linear-OPF models proposed establish a linear relationship between the power injections at the nodes and the power flows through the lines. These models are trained with case scenarios particular to the case of study, in contrast to the PTDF-OPF model, in which the PTDF approximation matrix solely depends on network parameters. In order to assess the performance of the proposed model, the approximate results obtained are compared to those from PTDF-OPF, an analogous optimization problem that serves as a benchmark.

By developing a linear Optimal Power Flow formulation that enables its faster and broader application, the proposed modifications aim to enhance the integration of renewable energy sources and improve the reliability and stability of power systems under various operating conditions, ensuring the effectiveness and stability of future power grids. The proposed linear models can serve as valuable tools for power system

operators and policymakers, enabling them to optimize power generation schedules and make informed decisions regarding power system development. Furthermore, they can guide future research in the field of power systems engineering, fostering a more sustainable and efficient energy future.

1.5 Organization of the document

The present document is divided in 5 chapters. In the following lines, a brief outline of each chapter is provided.

- **Chapter 1:** *Introduction.* In this chapter, an introduction to the state-of-the-art research and advancements in the resolution of AC Optimal Power Flow is provided. This overview aims to offer a summary of the key concepts, methodologies and challenges that persist on the resolution of non-convex optimization problems in the context of electrical power systems.
- **Chapter 2:** *Methodology.* In this chapter, the optimization problem formulations proposed in this study are presented in detail, as well as a description of the evaluation methods used to assess the performance of the linear models.
- **Chapter 3:** *Illustrative example.* In this chapter, a simple power system with 6 buses, *case6ww*, is presented as a small illustrative example. The various methods and approaches considered in Chapter 2 are applied to test the accuracy of the diverse linear Optimal Power Flow models.
- **Chapter 4:** *Case study.* In this chapter, the tools and techniques previously explored in the context of the illustrative example are applied to a more realistic case of study, the IEEE 39-bus system. Additional evaluation criteria is applied to determine the computational performance of the Linear-OPF method proposed with respect to the PTDF-OPF.
- **Chapter 5:** *Conclusions.* In this chapter, the key conclussions on the performance and regression capabilities of the proposed Linear-OPF model are presented, along with a brief discussion on the potential future lines of work that stem from the results found here.
- **Appendices**. In this last chapter, all considered OPF formulations, as well as some additional information on optimization problems and least squares regression methods, are presented.

Chapter 2

Methodology

This chapter presents a summary of the optimization formulations and procedures that are designed to test the regression capabilities and accuracy of the linear models proposed for evaluation.

2.1 Notation

In the first place, several notation rules are stated below in order to clarify the mathematical formulation presented. Even if the physical properties and parameters expressed below, as well as the numerical results obtained in this work, are presented in dimensional units, the calculations are computed in the per-unit (p.u.) system. For a brief introduction into the p.u. system and its key concepts, please refer to Appendix C.

Sets and indices

- N: Set of buses of the power system considered.
- i, j: Index of system buses, $i, j \in 1, 2, ..., N$.
- M_i: Subset of the buses of the power system considered N that are directly connected to bus i.
- L: Set of bus pairs (i,j) that are directly connected by a line.
- G: Set of generation units of the power system considered.
- K_{Tr}: Number of system states/scenarios in the training data set.

- K_{Ts}: Number of system states/scenarios in the testing data set.
- k: Index of system states/scenarios in data set, both training and test sets.

Variables

- P_i: Active power injection of bus i in [MW].
- Q_i: Reactive power injection of bus i in [MVAr].
- V_i: Voltage magnitude/angle of bus i in [V].
- θ_i: Voltage angle of bus i in [rad].
- P_{ij}^L: Active power flow from bus i to bus j in [MW].
- Q_{ij}^L: Reactive power flow from bus i to bus j in [MVAr].
- P_i^G: Active power generation on bus i in [MW].
- Q_i^G: Reactive power generation on bus i in [MVAr].
- P_i^D: Active power demand/load on bus i in [MW].
- Q_i^D: Reactive power demand/load on bus i in [MVAr].

Vectors and functions

- $\boldsymbol{P}$: Vector of the active power injections in [MW], where $\boldsymbol{P} = [P_i,\ i \in N]$.
- $\boldsymbol{Q}$: Vector of the reactive power injections in [MVAr], where $\boldsymbol{Q} = [Q_i,\ i \in N]$.
- $\boldsymbol{P^L}$: Vector of the active power flows in [MW], where $\boldsymbol{P^L} = [P_{ij}^L,\ i,j \in L]$.
- $\boldsymbol{Q^L}$: Vector of the reactive power flows in [MVAr], where $\boldsymbol{Q^L} = [Q_{ij}^L,\ i,j \in L]$.
- $\boldsymbol{V}$: Vector of voltage magnitudes in [V], where $\boldsymbol{V} = [V_i,\ i \in N]$.
- $\boldsymbol{\theta}$: Vector of voltage angles in [rad], where $\boldsymbol{\theta} = [\theta_i,\ i \in N]$.
- X_{PTDF}: PTDF matrix that linearly relates active power injections and active power flows.
- X_{P-PF}: Data-based linear approximation matrix that relates active power injections and active power flows, trained with AC-PF data.
- X_{P-OPF}: Data-based linear approximation matrix that relates active power injections and active power flows, trained with AC-OPF data.

- X_{Q-PF}: Data-based linear approximation matrix that relates reactive power injections and reactive power flows, trained with AC-PF data.
- X_{Q-OPF}: Data-based linear approximation matrix that relates reactive power injections and reactive power flows, trained with AC-OPF data.
- X_{P-loss}: Data-based linear approximation matrix that relates active power loads and active power losses, trained with AC-OPF data.
- X_{Q-loss}: Data-based linear approximation matrix that relates reactive power loads and reactive power losses, trained with AC-OPF data.

Parameters

- $P_i^{G,min}/P_i^{G,max}$: Lower/Upper limit of active power generation for generation unit i in [MW].
- $Q_i^{G,min}/Q_i^{G,max}$: Lower/Upper limit of reactive power generation for generation unit i in [MVAr].
- V_i^{min}/V_i^{max}: Lower/Upper limit of voltage magnitude for bus i in [V].
- $\theta_i^{min}/\theta_i^{max}$: Lower/Upper limit of voltage angle for bus i in [rad].
- P_{ij}^{max}: Maximum value for the active power flow through the line that connects bus i to bus j in [MW].
- S_{ij}^{max}: Maximum value for the apparent power flow through the line that connects bus i to bus j in [MVA].
- r_{ij}: Resistance of the line that connects bus i to bus j in [Ω].
- x_{ij}: Reactance of the line that connects bus i to bus j in [Ω].
- g_{ij}: Conductance of the line that connects bus i to bus j in [S].
- b_{ij}: Susceptance of the line that connects bus i to bus j in [S].
- b_{ij}^s: Shunt susceptance of the line that connects bus i to bus j in [S].
- Y: Admittance matrix of the system in [S].
- G: Conductance matrix of the system in [S].
- B: Susceptance matrix of the system in [S].

2.2 Power Flow equations

Most of the mathematical models applied to the operation and management of power grids incorporate the Power Flow equations as the fundamental governing equations of electrical systems. A power system is represented as a collection of buses denoted as N, interconnected by a set of branches labeled as L. The conventional AC Power Flow (AC-PF) problem aims to obtain a deterministic solution to the following network equations:

$$\begin{cases} P_i(V,\theta) = P_i^G - P_i^L & \forall i \in N \\ Q_i(V,\theta) = Q_i^G - Q_i^L & \forall i \in N \end{cases} \tag{2.1}$$

where P_i and Q_i represent the injected active and reactive power at each bus i, P_i^G and Q_i^G are the active and reactive generated powers at each bus, P_i^L and Q_i^L are the active and reactive power loads at each bus, and V and θ correspond to the voltage magnitude and phase angle, respectively. The choice of the specific functions that relate the inyected active and reactive powers with the voltage magnitudes and angles lead to different Power Flow formulations. In the AC-PF specific case, and considering the admittance matrix Y of the system (composed by the conductance matrix G and the susceptance matrix B that correspond to its real and imaginary part, respectively, and the shunt susceptances b^s; more information on the construction of these matrices can be found at Appendix D), the expressions that characterize the power flows through the lines are presented in the non-linear equations below:

$$P_{ij}^L = V_i V_j [G_{ij}\, cos(\theta_i - \theta_j) + B_{ij}\, sin(\theta_i - \theta_j)] - G_{ij} V_i^2 \quad \forall i,j \in N \tag{2.2}$$

$$Q_{ij}^L = V_i V_j [G_{ij}\, sin(\theta_i - \theta_j) - B_{ij}\, cos(\theta_i - \theta_j)] + (B_{ij} - b_{ij}^s) V_i^2 \quad \forall i,j \in N \tag{2.3}$$

In the conventional AC-PF analysis, each bus is fully represented by 4 variables: voltage magnitude V, voltage angle θ, active power injection P and reactive power injection Q. For each bus in the network, two of these variables must be known, while the other two remain as unknowns. The focus of the AC-PF problem is then to calculate the voltages and power injections for all buses within the system. This is achieved through the resolution of a system of $2N$ equations and $2N$ unknowns. Reference [12] presents a compilation of classical power system analysis within the framework of currently restructured electric energy systems, in which the resolution of problems such as the AC Power Flow are discussed in detail. Depending on which of these variables are known at each bus, three different bus types can be categorized:

- Slack Bus: the slack bus, also known as the swing bus, serves two main purposes within a power system model:
  - Voltage Reference: At the slack bus, both the voltage magnitude ($|V|$) and angle (θ) are fixed. Typically, the voltage magnitude is set to 1.0 per unit (p.u.) and the angle is set to 0 radians. This fixed voltage reference allows for the determination of the voltage magnitudes and angles at all other buses in the system.
  - Power Injection Freedom: The slack bus is the only bus where the real power injection (P) and reactive power injection (Q) are free to vary. This flexibility ensures that the power flow equations have a feasible solution. By allowing the real power at the slack bus to vary, the system can accommodate the power demand and generation requirements. It is important to note that there is only one slack bus present in a power system model.
- Load Bus: at a load bus, also known as a 'PQ' bus, the power injections (P and Q) are known, representing the specified real power demand and reactive power demand at that bus. On the other hand, the voltage magnitude ($|V|$) and voltage angle (θ) at the load bus are variable.
- Voltage-controlled Bus: At a voltage-controlled bus, also known as a 'PV' bus, the real power injection (P) and voltage magnitude ($|V|$) are known, representing the specified real power generation or consumption and the desired voltage level at that bus. On the other hand, the reactive power injection (Q) and voltage angle (θ) at the PV bus are variables. The presence of a PV bus allows for local control of reactive power to regulate the voltage to a desired setpoint. By fixing the real power injection and voltage magnitude, the PV bus ensures that the power flow equations are satisfied while enabling the adjustment of reactive power and voltage angle to maintain voltage stability.

Assigning buses by means of these criteria establishes an equal number of equations and unknowns ($2N$), which allows to solve the non-linear system of equations by numerical methods such as the Newton-Raphson.

All Power Flow formulations can be found at Appendix E.

2.3 AC-OPF formulation

The primary objective of the AC-OPF optimization problem is to minimize the overall cost of electricity generation while ensuring the safe operation of the electrical

power system. Before dealing with the details of the AC-OPF formulation, a summary of fundamental concepts regarding optimization problems and their general formulations can be found in Appendix A.

While the active generation power of the generating units are known in the AC-PF, the AC-OPF formulation considers these values as variables to be optimized in order to minimize the operating cost of the system, constrained by specific minimum and maximum operational limits. The controllable generators are located at a subset of the system buses, designated as G. The cost of operation for each generating unit is typically expressed as a quadratic function of its real output power, denoted as $C(P^G)$. For simplicity, we assume that there is only one generator unit and one power load per bus $i \in N$. The classic polar power-voltage formulation of the AC-OPF optimization problem [13] is presented below.

$$
\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & P_{ij}^L = V_i V_j [G_{ij}\, cos(\theta_i - \theta_j) + B_{ij}\, sin(\theta_i - \theta_j)] - G_{ij} V_i^2 \quad \forall i, j \in N \\
& Q_{ij}^L = V_i V_j [G_{ij}\, sin(\theta_i - \theta_j) - B_{ij}\, cos(\theta_i - \theta_j)] + (B_{ij} - b_{ij}^s) V_i^2 \quad \forall i, j \in N \\
& P_i^G - P_i^D = \sum_{j \in M_i} P_{ij}^L \quad \forall i \in N \\
& Q_i^G - Q_i^D = \sum_{j \in M_i} Q_{ij}^L \quad \forall i \in N \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& Q_i^{G,min} \leq Q_i^G \leq Q_i^{G,max} \quad \forall i \in N \\
& V_i^{min} \leq V_i \leq V_i^{max} \quad \forall i \in N \\
& \theta_i^{min} \leq \theta_i \leq \theta_i^{max} \quad \forall i \in N \\
& (P_{ij}^L)^2 + (Q_{ij}^L)^2 \leq (S_{ij}^{max})^2 \quad \forall i, j \in N
\end{aligned}
\tag{2.4}
$$

In the following lines, some remarks are expressed about the different equations and restrictions that compose the classic AC-OPF formulation.

Objective function

The objective function is a mathematical expression that quantifies the cost associated with power generation, including fuel costs, operating and maintenance expenses, and environmental factors, among others. In the present document, only the active generation power is accounted for, obtaining a measure of its economic cost via a quadratic polynomial.

$$\sum_{i \in G} C_i(P_i^G) = \sum_{i \in G} (a_i \cdot (T \cdot P_i^G)^2 + b_i \cdot (T \cdot P_i^G) + c_i \cdot T) \tag{2.5}$$

Where a_i [$/$MW^2h$], b_i [$/MWh] and c_i [$/h] are the economic factors that define the generation costs for all units, and T is the time period considered for the duration of the OPF operation, considered as $T = 1$h in this study.

Injected power

The injected power at each node, both active and reactive, are defined as follows:

$$P_i = P_i^G - P_i^D \quad \forall i \in N \tag{2.6}$$
$$Q_i = Q_i^G - Q_i^D \quad \forall i \in N \tag{2.7}$$

This expression, defined as the difference between the total generated power and the total load present in each of the buses of the net, is fundamental in the determination of the power balance equations.

Power flows through the lines

The bus injection model, often referred to as the 'AC Power Flow equations', serves as a concise representation of power system behavior by describing the active and reactive power injections at each bus in the system. Given an admittance matrix Y, a set of two equations express the active and reactive power injections. These pair of equations can be expressed in various equivalent forms, depending on whether voltage and admittance matrix elements are represented in polar or rectangular coordinates. In this work, the polar coordinates are selected, shown below:

$$P_{ij} = V_i V_j [G_{ij}\, cos(\theta_i - \theta_j) + B_{ij}\, sin(\theta_i - \theta_j)] - G_{ij} V_i^2 \quad \forall i, j \in N \tag{2.8}$$
$$Q_{ij} = V_i V_j [G_{ij}\, sin(\theta_i - \theta_j) - B_{ij}\, cos(\theta_i - \theta_j)] + (B_{ij} - b_{ij}^s) V_i^2 \quad \forall i, j \in N \tag{2.9}$$

Power balance equations

The power balance on each bus of a power system is a critical aspect of ensuring the stable and efficient operation of the entire network. Each bus in the system represents a specific point of connection, such as a generator, load, or an interconnection with another system. To maintain equilibrium, the power injected into a

bus must equal the power drawn from it, accounting for losses in the transmission and distribution system. This condition is expressed below:

$$P_i = P_i^G - P_i^D = \sum_{j \in M_i} P_{ij}^L \quad \forall i \in N \tag{2.10}$$

$$Q_i = Q_i^G - Q_i^D = \sum_{j \in M_i} Q_{ij}^L \quad \forall i \in N \tag{2.11}$$

Operational limits

The active and reactive generation limits represent the maximum active and reactive power that a generator can contribute to the system while staying within its operational constraints. The active power limit ensures that generation remains within the physical capabilities of the generators, whereas the reactive power limit regulates the flow of reactive power, essential for maintaining voltage levels within acceptable bounds. If there is no generator unit in bus i, then the operational limits $P_i^{G,min}$, $P_i^{G,max}$, $Q_i^{G,min}$ and $Q_i^{G,max}$ are considered to have a value of 0.

$$P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \tag{2.12}$$

$$Q_i^{G,min} \leq Q_i^G \leq Q_i^{G,max} \quad \forall i \in N \tag{2.13}$$

Voltage limits

Voltage magnitude limits ensure that the system operates within a safe voltage range to avoid equipment damage and maintain the quality of power supplied to consumers. Voltage angle limits, on the other hand, are crucial for preserving the stability of the grid by preventing excessive phase differences between buses. Deviating from these limits can result in voltage instability and, in extreme cases, lead to cascading outages.

$$V_i^{min} \leq V_i \leq V_i^{max} \quad \forall i \in N \tag{2.14}$$

$$\theta_i^{min} \leq \theta_i \leq \theta_i^{max} \quad \forall i \in N \tag{2.15}$$

Thermal limits on the lines

As line current is dependent on voltage, establishing an exact bound necessitates a function inequality constraint. According to Ohm's law, the magnitude of current in line ij is:

$$I_{ij} = |V_i - V_j| y_{ij} \quad \forall i, j \in N \tag{2.16}$$

Where y_{ij} is the magnitude of the line admittance. This equation can be used to develop an exact bound as:

$$\begin{aligned} |V_i - V_j| y_{ij} \leq I_{ij}^{max} \longrightarrow \\ \longrightarrow \sqrt{(V_i cos\theta_i - V_k cos\theta_k)^2 + (V_i sin\theta_i - V_k sin\theta_k)^2} \leq \frac{I_{ij}^{max}}{y_{ij}} \longrightarrow \\ \longrightarrow (V_i cos\theta_i - V_k cos\theta_k)^2 + (V_i sin\theta_i - V_k sin\theta_k)^2 \leq \frac{(I_{ij}^{max})^2}{y_{ij}^2} \quad \forall i, j \in N \end{aligned} \tag{2.17}$$

Instead of bounding the square of the current, most formulations opt to establish bounds for the real and reactive power flow in the line. In nearly all instances, this simplification proves to be accurate enough.

$$(P_{ij}^L)^2 + (Q_{ij}^L)^2 \leq (S_{ij}^{max})^2 \quad \forall i, j \in N \tag{2.18}$$

2.4 DC-OPF formulation

Despite the nonlinearity of both P_{ij} and Q_{ij} with respect to V and θ in the PF equations, it is possible to find a reasonable linear approximation between P_{ij} and θ, resulting in the so-called DC load flow. This model operates under the assumption that $V_i = 1$ at all buses, thereby losing the capability to track reactive power flows or any other voltage-related data. With this hypothesis, the active power flows, as given by Equation 2.8, simplifies to:

$$P_{ij}^L = G_{ij}(cos\theta_{ij} - 1) + B_{ij} sin\theta_{ij} \tag{2.19}$$

Under the assumption that phase angle differences corresponding to adjacent buses are small ($cos\theta_{ij} \approx 1$ and $sin\theta_{ij} \approx \theta_i - \theta_j$) leads to:

$$P_{ij}^L = B_{ij}(\theta_i - \theta_j) \tag{2.20}$$

Refer to Appendix D for more information on how the admittance matrix Y is constructed and its components, the conductance matrix G and the susceptance

matrix B. For values $r/x < 3$, common in transmission networks, the error arising when the susceptance matrix element B_{ij} is replaced by $1/x_{ij}$ is lower than 1%, resulting in:

$$P_{ij}^L = \frac{1}{x_{ij}}(\theta_i - \theta_j) \rightarrow \theta_i - \theta_j = x_{ij} P_{ij}^L \tag{2.21}$$

With this approximate expression for the active power flows, the following linear OPF model, known as DC-OPF, is presented below.

$$\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & P_{ij}^L = \frac{(\theta_i - \theta_j)}{x_{ij}} \quad \forall i, j \in N \\
& P_i^G - P_i^D = \sum_{j \in M_i} P_{ij}^L \quad \forall i, j \in N \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& \theta_i^{min} \leq \theta_i \leq \theta_i^{max} \quad \forall i \in N \\
& -P_{ij}^{max} \leq P_{ij}^L \leq P_{ij}^{max} \quad \forall i, j \in N
\end{aligned} \tag{2.22}$$

2.5 PTDF-OPF formulation

Let A denote the line-to-node incidence matrix, $\boldsymbol{\theta}$ the vector of bus phase angles, both of them reduced by removing the slack row, and X a diagonal matrix of line reactances; then, Equation 2.21 can be rewritten in matrix form as:

$$A^T \boldsymbol{\theta} = X \boldsymbol{P^L} \rightarrow \boldsymbol{P^L} = [X^{-1} A^T] \boldsymbol{\theta} \tag{2.23}$$

Ignoring the line resistances, the sum of all active powers amounts to zero. This means that the slack power constitutes a linear combination of the active powers. If $\boldsymbol{P}$ denotes the vector of the net injected powers, except that of the slack bus, the application of the Kirchhoff's current law applied to power flow leads to:

$$\boldsymbol{P} = A \boldsymbol{P^L} \tag{2.24}$$

The elimination of power flows thanks to Equation 2.23 allows to obtain the

following linear relationship between active power injections and phase angles:

$$\boldsymbol{P} = [AX^{-1}A^T]\boldsymbol{\theta} = C\boldsymbol{\theta} \tag{2.25}$$

where C is a symmetric and sparse matrix, with a simmilar structure to that of the admittance matrix. The values of C only depend on the line reactances.

Alternatively, we can remove phase angles, leading to a linear relationship between power flows and power injections:

$$\boldsymbol{P^L} = [X^{-1}A^TC^{-1}]\boldsymbol{P} = X_{PTDF} \cdot \boldsymbol{P} \tag{2.26}$$

Under the DC Power Flow approximation, each of the elements of X_{PTDF} provides the sensitivity between the respective active power flow and active power injection, assuming that the slack bus compensates for changes in any bus injection. The obtained matrix X_{PTDF} is dense, but many of its elements relating electrically distant line-bus pairs can be neglected. This matrix is called the Power Transfer Distribution Factors (PTDF) matrix, which helps to quickly analyze power flow changes arising from branch or generator outages, and constitutes an essential tool for the determination of power transfer capabilities in competitive electricity markets.

Each of the PTDF matrix elements consists on a factor that represents the percentage of change in power flow through network branch (i-j) due to the existence of a new transaction in the system (from bus X to bus Y). The simple calculation procedure of the PTDF matrix, only dependant on the topology of the network and its system parameters, provides the method with a strong robustness and independency of the choice of the balancing node. This linear approximation of the active power flows is usually implemented in the PTDF-OPF optimization problem presented below, that serves as a surrogate for the classic AC-OPF problem.

$$\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & \boldsymbol{P^L} = X_{PTDF} \cdot \boldsymbol{P} \\
& \sum_{i \in N} P_i^G = \sum_{i \in N} P_i^D \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& -P_{ij}^{max} \leq P_{ij}^L \leq P_{ij}^{max} \quad \forall i,j \in N
\end{aligned} \tag{2.27}$$

While both DC-OPF and PTDF-OPF formulations are linear, there is a key

distinction lying in the consideration of voltage angles: while the DC-OPF accounts for voltage angles, the PTDF-OPF does not incorporate them, as it can be observed in the formulation presented above. The reliance of the PTDF-OPF model on static network assumptions hinders its effectiveness in capturing real-time grid dynamics, particularly in heavily interconnected systems prone to frequent topology changes and transient phenomena.

2.6 Linear-OPF formulations

This sections presents the data-based Linear-OPF model as an alternative for the PTDF-OPF formulation. While similar in structure, the proposed Linear-OPF formulation leverages a matrix determined by means of historical data of the particular network considered. This method aims to address the limitations inherent in the existing PTDF-OPF approach, offering a data-based novel solution to overcome its static dependance on the network parameters. The accuracy and performance of the proposed Linear-OPF models is tested against the results obtained from the PTDF-OPF model for the same testing set.

2.6.1 Base formulation

The most simple Linear-OPF implemented to illustrate the method proposed is presented in the following lines.

$$\begin{aligned} \min \quad & \sum_{i \in G} C_i(P_i^G) \\ \text{s.t.} \quad & \boldsymbol{P^L} = X_P \cdot \boldsymbol{P} \\ & \sum_{i \in N} P_i^G = \sum_{i \in N} P_i^D \\ & P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\ & -P_{ij}^{max} \leq P_{ij}^L \leq P_{ij}^{max} \quad \forall i,j \in N \end{aligned} \tag{2.28}$$

In order to compose it, the PTDF-OPF formulation is used as a blueprint. For this reason, this Linear-OPF model ignores the voltages (both magnitude and angle) and the reactive power. The only aspect that differs from the PTDF-OPF formulation is the expression that calculates the active power flow through the lines:

$$\boldsymbol{P^L} = X_P \cdot \boldsymbol{P} \tag{2.29}$$

Instead of using the PTDF matrix for the calculation of the active power flows, the data-based linear matrix X_P is instead used. The matrix X_P relates in a linear fashion the active power flows through the lines and the injected active power at the buses. This linear model is trained from either AC-PF data or AC-OPF data, resulting in two different linear models: X_{P-PF} and X_{P-OPF}.

2.6.2 Upgraded formulation

In this section, an upgraded version of the proposed Linear-OPF models is presented. This formulation includes two additional modifications to the first Linear-OPF proposed in the previous section:

- **Reactive flows**: an approximation of the reactive power flows in the same fashion as the method proposed for the approximation of the active power flows, a linear relationship that can be expressed as:

$$\boldsymbol{Q^L} = X_Q \cdot \boldsymbol{Q} \tag{2.30}$$

- **Loss estimation**: an approximation for the active and reactive losses on the net is included on the upgraded Linear-OPF model. Up to this point, the Linear-OPF model has considered a lossless power system, as expressed in its balance equations:

$$\sum_{i \in N} P_i^G = \sum_{i \in N} P_i^D \tag{2.31}$$

$$\sum_{i \in N} Q_i^G = \sum_{i \in N} Q_i^D \tag{2.32}$$

Now, an approximation of the losses is considered via linear approximation matrices, X_{P-loss} and X_{Q-loss}, that relates the power losses on the network with the power loads at the buses obtained from the resolution of multiple AC-OPF problems. These considerations results in the following balance equations:

$$\sum_{i \in N} P_i^G - \sum_{i \in N} P_i^D = X_{P-loss} \cdot \boldsymbol{P^D} \tag{2.33}$$

$$\sum_{i \in N} Q_i^G - \sum_{i \in N} Q_i^D = X_{Q-loss} \cdot \boldsymbol{Q^D} \tag{2.34}$$

Including these additions, the Linear-OPF model complete formulation is presented below:

$$\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & \boldsymbol{P^L} = X_P \cdot \boldsymbol{P} \\
& \boldsymbol{Q^L} = X_Q \cdot \boldsymbol{Q} \\
& \sum_{i \in N} P_i^G - \sum_{i \in N} P_i^D = X_{P-loss} \cdot \boldsymbol{P^D} \\
& \sum_{i \in N} Q_i^G - \sum_{i \in N} Q_i^D = X_{Q-loss} \cdot \boldsymbol{Q^D} \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& Q_i^{G,min} \leq Q_i^G \leq Q_i^{G,max} \quad \forall i \in N \\
& (P_{ij}^L)^2 + (Q_{ij}^L)^2 \leq (S_{ij}^{max})^2 \quad \forall i,j \in N
\end{aligned} \tag{2.35}$$

For an equivalent comparison of the benchmark PTDF-OPF model and the upgraded Linear-OPF models, the additions presented in this section are applied as well to the PTDF-OPF model, resulting in an upgraded PTDF-OPF model. However, as the PTDF matrix only considers the active power flows, the upgrade on the PTDF-OPF model can only be applied to the active loss estimation. The formulation for the upgraded PTDF-OPF model is presented below:

$$\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & \boldsymbol{P^L} = X_{PTDF} \cdot \boldsymbol{P} \\
& \sum_{i \in N} P_i^G - \sum_{i \in N} P_i^D = X_{P-loss} \cdot \boldsymbol{P^D} \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& -P_{ij}^{max} \leq P_{ij}^L \leq P_{ij}^{max} \quad \forall i,j \in N
\end{aligned} \tag{2.36}$$

All Optimal Power Flow formulations can be found at Appendix F.

2.7 Learning procedure

The reliability of the Machine Learning methods heavily depends on the underlying dataset used to train the model, emphasizing the importance of representing

a wide variety of operating conditions in the dataset. In this section, the details of the generation of case scenarios for both the training and testing of the data-based linear models is presented.

Due to the lack of historical data for the nets considered in this document, both the training and testing data are obtained via numerical simulations. For this purpose, the original net parameters are modified. In particular, the modified parameters that allow the achievement of different scenarios are the active power loads. Both the regressors (power injections in the buses) and the objectives (power flows through the lines) that are required for the training of the data-based linear model proposed are obtained after the resolution of both the Power Flow and the Optimal Power Flow for each loading scenario. Depending on the training set used for the computation of the Linear-OPF model matrices, X_P and X_Q, two different models are obtained. The data-based learning algorithm proposed is independent of the training set (active or reactive power injections and power flows) and the origin of the data (AC-PF or AC-OPF data), presenting the same steps. The algorithm is stated below for the active power case without loss of generality:

Data-based learning algorithm

1. **Step 0:** Let $\boldsymbol{P_k}$ and $\boldsymbol{P_k^L}$ be the active injection and active power flow vectors for the training scenario k, obtained either from the resolution of the AC-PF or AC-OPF formulation.
2. **Step 1:** The estimation of active line power flows is denoted by $\hat{\boldsymbol{P}}_{\boldsymbol{k}}^{\boldsymbol{L}}$, and can be computed from the X_P linear regression matrix as:

$$\hat{\boldsymbol{P}}_{\boldsymbol{k}}^{\boldsymbol{L}} = X_P \cdot P_k \tag{2.37}$$

3. **Step 2:** The linear regression matrix is computed by solving the following least-squares optimization problem:

$$X_P \in \arg\min \sum_k ||\boldsymbol{P_k^L} - X_P \cdot \boldsymbol{P_k}||_2^2 \tag{2.38}$$

This least squares regression method is implemented in Python thanks to the function *linalg.lstsq* from the *Numpy* package.

The notation and characteristics of both Linear-OPF models are stated below for the clarity of presentation:

- Linear$_{PF}$-OPF: Proposed optimization problem that uses the linear matrix X_{P-PF} for the approximation of the active power flows. In the upgraded

Linear$_{PF}$-OPF model, the linear matrix X_{Q-PF} is used as well for the approximation of the reactive power flows. These approximation matrices are computed via Power Flow training data.

- Linear$_{OPF}$-OPF: Proposed optimization problem that uses the linear matrix X_{P-OPF} for the approximation of the active power flows. In the upgraded Linear$_{OPF}$-OPF model, the linear matrix X_{Q-OPF} is used as well for the approximation of the reactive power flows. These approximation matrices are computed via Optimal Power Flow training data.

For the linear approximation model to be accurate and robust under a wide range of operation points, the training dataset must include a representative wide spectrum of the AC Optimal Power Flow feasible region. In order to do this, numerous techniques are described in the literature. One option is the continuation power flow [14], which attempts to capture the full range of the possible load profile employing a predictor-corrector scheme to trace a solution path of a set of power flow equations, which have been reformulated to include a load parameter.

In the present study, the method used for the generation of case scenarios is presented in the following equations.

$$\begin{cases} P_i^{D(k)} = P_i^D \times \eta_{A,k} \times \eta_i^P; \quad i = 1, 2, \cdots, N \\ Q_i^{D(k)} = Q_i^D \times \eta_{A,k} \times \eta_i^Q; \quad i = 1, 2, \cdots, N \\ \eta_{A,k} \sim U(0.8, 1.8); \quad k = 1, 2, \cdots, K \\ \eta_i^P, \eta_i^Q \sim U(0.95, 1.05). \end{cases}$$

This model incorporates an uniform random load coefficient $\eta_{A,k}$, sampled from an uniform distribution centered in 1, for each of the case scenarios. If $\eta_{A,k} \approx 1$, then the case scenario generated is close to the nominal operating point of the power system. Additionally, a different value of the factors η_i^P and η_i^Q, also sampled from an uniform distribution, is applied to each bus for each case scenario, with the intention of avoiding collinearity in the training and testing sets.

By means of this load variation method, any number of case scenarios can be generated. In particular, a total number of $K_{Tr} = 100 \cdot N$ case scenarios are considered for training and $K_{Ts} = 10 \cdot N$ case scenarios are considered for the testing. This scaling choice stems from the need to adjust the number both the training and testing sets needed for the faithful representation of the complexity of the power system of study.

We are only interested in those case scenarios that have a feasible solution for all the Optimal Power Flow formulations of study. An infeasible solution would

not serve as neither a training nor a test set, as its solution is neither physical nor realizable. For this reason, once a potential case scenario is generated, the various OPF formulations are tested on the case. Only if all of the OPF formulations are feasible, the case scenario proposed is included into either the training or the testin set.

For the training procedure, both AC-PF and AC-OPF are solved for each case scenario, and the predictor variables (power injections in the buses) and the response variables (power flows through the lines) obtained are saved into matrices. These matrices are then used to compute the least squares regression model and obtain the linear approximation matrices X_{PF} and X_{OPF} that characterize the proposed linear Optimal Power Flow models Linear$_{PF}$-OPF and Linear$_{OPF}$-OPF, respectively.

2.8 Testing methods

In the following section, three different testing procedures are presented, each of them assessing a different performance aspect from the linear Optimal Power Flow models considered in this study.

2.8.1 Regression Testing

The first of them, named 'Regression Testing', consists on solving multiple AC-PFs or AC-OPFs, and attempting to approximate the active power flows obtained (the predictor variables) by means of multiplying accordingly each linear regression model matrix and the active power injections (the regressor variables). Depending on the choice of the problem to obtain both the predictor and regressor datasets, the Regression Testing can be distinguished as PF-based or OPF-based Regression Testing. This method is presented in the paper of study [10], and aims to test in particular the regression capabilities of the linear models considered.

The errors that characterize the Regression Testing are computed measuring the L2-norm distance between the active power flows through the lines obtained either via a Power Flow or an Optimal Power Flow resolution, and the solution obtained from the linear approximation methods. These errors are presented in the following sections in absolute value, assuming that the deviation has the same impact on the solution no matter if there is an undererstimation or overestimation of the real value.

Another useful measure of the performance of each linear approximation in the Regression Testing is presented through the PI index, defined as:

$$PI = \frac{1}{L} \cdot \sum_{l=1}^{L} \frac{Avg.Error_l}{Max.Error_l} \cdot 100 \tag{2.39}$$

This value expresses the ratio between the average error relative to the maximum error. A low value of PI suggests that the Average Error (Avg. Err) is significantly smaller than the Maximum Error (Max. Err). This indicates that the linear models can provide high performance for the majority of test samples but may perform less effectively for a few challenging cases.

The Regression Testing procedure is extracted from Reference [10], and constitutes the basis for the validation of the first set of results obtained in this study.

2.8.2 Optimization Testing

The second method is a novel approach that diverges from the testing procedures found in the literature, and motivates the current study. The Optimal Power Flow is resolved, but its power flows through the lines, whose calculation is essential to determine the line capacity restrictions, are not calculated by means of the non-linear equations that characterize the AC-OPF. Instead, the formulations of both the PTDF-OPF and the Linear-OPF are used.

In order to evaluate the performance of the different linear Optimal Power Flow models with the Optimization Testing procedure, two complementary criteria are considered. For the sake of illustrating these different criteria, the Figure 2.1 displays these metrics in a visual way.

The 2D space representation of Figure 2.1 represents the space of all possible active generation profiles that characterize any potential solution of the AC-OPF problem. This white space is composed by both AC-feasible and non AC-feasible solutions of the AC-OPF problem.

The approximated solutions that are obtained after the resolution of the linear Optimal Power Flow models proposed are represented with the $P1$ and $P2$ points. The solution of an AC-OPF, characterized by the same load parameters, is represented by $P3$. For each of the points (each of which represent an active generation profile), there is a active generation cost value associated, captured by the value of the objective function of the optimization problem. Each of the elipsoidal black dotted lines in Figure 2.1 represents an isocontour of the active generating cost value, with an outwards increasing gradient pointed by the black arrow.

The purpose of the Optimization Testing is to asssess the quality of each of the

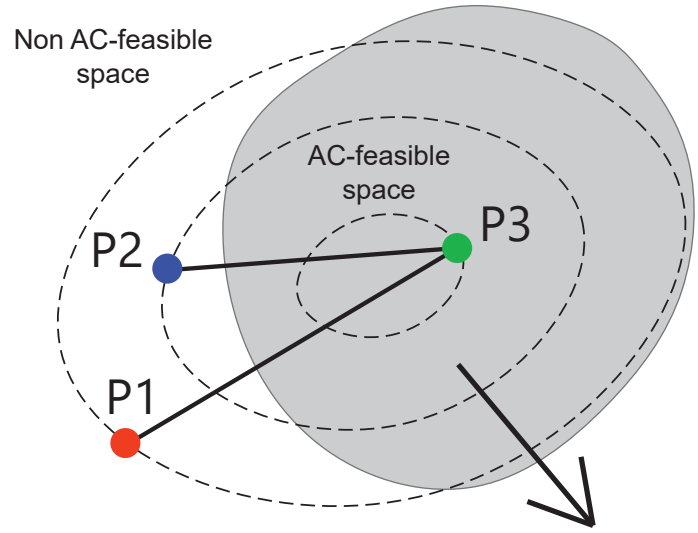


Figure 2.1: Visual plot of the Optimization Testing criteria.

approximated solutions by means of two complementary criteria:

- **Objective function:** The value of the objective function determines the active generation cost associated with the active generation profile obtained. For each of the Linear Optimal Power Flow models, the one that produces an approximated solution with a lower value for the objective function is capable of providing cheaper approximated solutions. Considering this criterion, point $P2$ would be the optimal approximate solution, as it is located in an isocontour with a lower value of the objective function.
- **L2-norm of the $P^{G,OPF}$ with respect to $P^{G,AC-OPF}$:** The distance, in the space of active power generation, between the approximated solution and the one representing the AC-OPF solution gives an estimate of how close the approximated solution is to the best solution, considered to be the one obtained from AC-OPF. This distance is represented by the segments that connect each approximate solution to the AC-OPF point $P3$. Considering this metric, point $P2$ would be as well the optimal choice, as it lays closer to point $P3$.

2.8.3 Feasibility Testing

Electricity spot market operators employ direct current based models, such as DC-OPF presented in Appendix F.2, for security-constrained unit commitment (SCUC) and security-constrained economic dispatch (SCED) in resource scheduling. The linear programming formulation offers the advantage of optimality and convergence, making it particularly suitable for large-scale systems. However, the

DC-OPF model primarily optimizes active power, simplifying considerations related to voltage magnitude and reactive power. This deficiency can lead to issues with uncontrolled voltage magnitude and power flow violations. Specifically, since voltage magnitudes are not taken into consideration during the market clearing process, DC market solutions may experience voltage magnitude violations, particularly in regions with high load demands.

Consequently, there is a notable disparity between approximated market solutions and AC-feasible points. This disparity becomes more pronounced with the increasing integration of renewable resources and electronic equipment into the power system, further emphasizing the interdependence between active power and reactive power. In practical industrial operations, reliability coordinators take on the responsibility of monitoring and managing voltage or reactive power concerns before and after contingencies. This approach ensures that the costs associated with correcting approximations solutions to ensure AC feasibility are not factored into the day-ahead market pricing [15].

The failure of conventional numerical methods for solving the AC-OPF problem may be consequence of several different factors [16]. They may either fail to converge due to the infeasibility of the problem or diverge due to the inability to find proper initial values. The distinction between problem infeasibility and solution infeasibility, as well as non-convergence and divergence of a numerical algorithm, can be defined as follows:

- Divergence: sometimes referred to as a "blow-up" of the solution algorithm, a divergence occurs when the solution trajectory is not reaching any final point, even after a significant number of iterations. When an algorithm diverges, it cannot output any real or physically feasible solution.

- Non-Convergence: A numerical algorithm is considered to not converge when the numerical calculation of underlying matrices (e.g., Jacobians and Hessians) has failed, meaning the matrices are very close to singular. In this case, the algorithm halts the iteration process, as it cannot progress further.

- Infeasible OPF problem: An OPF problem is infeasible or overdetermined when there is no solution satisfying all the specified operational constraints.

- Infeasible OPF solution: An AC-OPF solution is feasible if and only if all the AC-OPF constraints are satisfied at that solution. Accordingly, a solution is infeasible if at least one constraint is not satisfied at that solution. All mathematically infeasible solutions are physically infeasible, but the reverse is not true. A mathematically feasible solution is physically infeasible if the solution cannot be realized as the practical system operating condition.

In [17], it is demonstrated that the solution to the DC-OPF problem is incapable of meeting the requirements of AC Optimal Power Flow constraints. Even when generation adjustments are taken into account to consider the effect of losses, the DC-OPF solutions are still not feasible. Consequently, it is crucial to develop techniques that can identify the causes of infeasibility and offer recommendations for returning the system to a feasible region, aiding in restoration of AC-feasibility for the obtained approximated solutions.

Despite numerous mathematical formulations and algorithms designed to enhance computational efficiency, there is still no standard or widely accepted methodology for identifying the causes of infeasibilities in nonlinear optimization problems in general, and the AC-OPF problem in particular. There are several publications that attempt to adress this need by diverse means: in [18], the authors propose a novel solution approach to the AC-OPF problem, involving successive linear approximation of power flow equations, while [19] formulates a restoration procedure based on an unconstrained optimization problem that considers information regarding voltage phasors, power injections, and line flows from relaxed or approximated AC-OPF solutions.

In the present work, the restoration procedure proposed to obtain AC-feasible solutions from the approximate solutions of the linear OPF models is called PgenSlack-OPF and is presented below:

$$
\begin{aligned}
\min \quad & \sum_{i \in G} (\Delta P_i^G)^2 \\
\text{s.t.} \quad & P_i^G = P_i^{G,OPF} + \Delta P_i^G \quad \forall i \in N \\
& P_{ij}^L = V_i V_j [G_{ij}\, cos(\theta_i - \theta_j) + B_{ij}\, sin(\theta_i - \theta_j)] - G_{ij} V_i^2 \quad \forall i,j \in N \\
& Q_{ij}^L = V_i V_j [G_{ij}\, sin(\theta_i - \theta_j) - B_{ij}\, cos(\theta_i - \theta_j)] + (B_{ij} - b_{ij}^s) V_i^2 \quad \forall i,j \in N \\
& P_i^G - P_i^D = \sum_{j \in M_i} P_{ij} \quad \forall i \in N \\
& Q_i^G - Q_i^D = \sum_{j \in M_i} Q_{ij} \quad \forall i \in N \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& Q_i^{G,min} \leq Q_i^G \leq Q_i^{G,max} \quad \forall i \in N \\
& V_i^{min} \leq V_i \leq V_i^{max} \quad \forall i \in N \\
& \theta_i^{min} \leq \theta_i \leq \theta_i^{max} \quad \forall i \in N \\
& (P_{ij}^L)^2 + (Q_{ij}^L)^2 \leq (S_{ij}^{max})^2 \quad \forall i,j \in N
\end{aligned}
\tag{2.40}
$$

This approach consists on solving an optimization problem which shares almost

the same structure as the classic AC-OPF. The difference is found in the objective function: considering a base active generation profile $P^{G,OPF}$ obtained from the application of any of the linear Optimal Power Flow models, a slack variable ΔP^G is defined. This variable allows for a variation of the active generation around the base generation profile $P^{G,OPF}$, so that an AC-feasible generation profile P^G is obtained:

$$P^G = P^{G,OPF} + \Delta P^G \tag{2.41}$$

The values of the slack variable ΔP^G are limited by the objective function, which minimizes the sum of the squares of this magnitude for all generation units. The rest of restrictions and equations in the AC-restoration formulation are identical to those present in the classic AC-OPF formulation.

From this AC-restoration procedure, two different evaluation criteria are considered in order to assess the performance of the AC-feasible solutions recovered from the approximated solutions of each of the various linear OPF formulations used. To clarify the definition and implications of these criteria, Figure 2.2 visually presents these metrics.

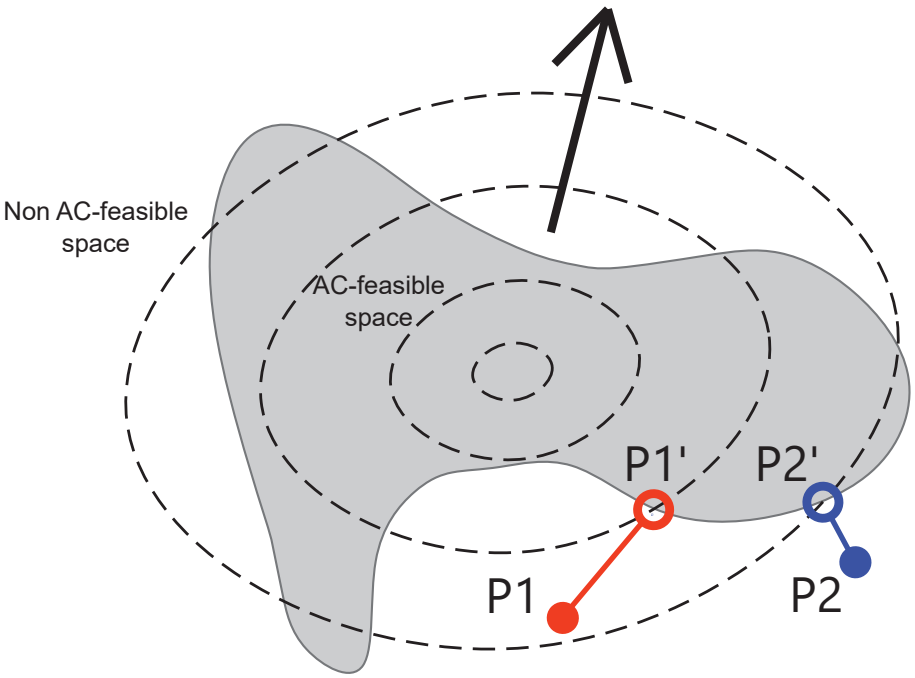


Figure 2.2: Visual plot of the AC-restored approximated solutions.

The 2D space representation of Figure 2.2 represents the space of all possible active generation profiles that characterize any potential solution of the AC-OPF problem. This space is divided into two areas: in a shade of gray, and bordered

by a thin black line, we can observe the AC-feasible space, while the rest of the space, in white, represents the non AC-feasible space. If a point in space is located inside the gray area, the OPF solution that this point represents would be a feasible solution of the problem. On the other hand, any other point located in the exterior of this closed figure would represent a generation profile that characterizes a non AC-feasible solution of the problem.

The approximate solutions that we obtain after the resolution of the linear OPF problems proposed are represented with the P1 and P2 points. These points lie in the non AC-feasible space. The purpose of the proposed AC-restoration method is to find a path (represented by a straight line) that leads from these non AC-feasible points to the closest AC-feasible point. These final points are named $P1'$ and $P2'$, and are located in the border of the AC-feasible space. Note that, without any loss of generality, a closed and non-convex shape has been chosen for the non-convex space. Note as well that, for each of the points in the 2D space presented, there exists a value for the active generating cost of the potential solution that each point represents. Each of the elipsoidal black dotted lines in Figure 2.2 represents an isocontour of the generating cost value, with an outwards increasing gradient pointed by the black arrow. Both evaluation criteria from the AC-restored solutions are presented next:

- **Distance from the initial to the AC-feasible solution:** One potential measure of the degree of unfeasibility is determined by the distance in parameter space between the desired operating point (given by the approximated solution) and the closest AC-feasible operating point. The difference between these two points is identified as the optimal change in the system parameters, corresponding to the generating profile, necessary to achieve feasibility. From Figure 2.2, one can observe that the distance (in active generation space) between the non AC-feasible point $P1$ and the closest AC-feasible point $P1'$ is shorter than the distance analogous to $P2$ and $P2'$. This represents the fact that, in order to go from the active generation profile defined by $P1$ to the active generating profile represented by $P1'$, the 'sum' of the changes needed in the generation units are greater than the corresponding changes for $P2$ to go to $P2'$. By this criterion, the non AC-feasible approximated solution of $P2$ is closer to being AC-feasible than $P1$.

- **Cost generation value of the AC-feasible solutions:** In addition to the necessary changes to be made so as to obtain an AC-feasible solution, the final cost of the AC-restored solution must be taken into account. In particular, the AC-restored solution $P2'$ lies in an isocontour with a higher value of the active generation cost associated than the isocontour in which $P1'$ is located. In other words, the resulting AC-feasible solution $P1'$ obtained from the AC-restoration of the approximated solution $P1$ is cheaper than $P2'$, AC-feasible point corresponding to $P2$.

These criteria are as presented not strictly dependent on one another, because even if the needed change in MW is greater, it may result in a cheaper generation configuration. With this figure, we intend to show the potential casuistics that may arise when evaluating the AC-restored solutions from the approximated points obtained by any of the OPF models used. Both criteria have to be taken into account to properly determine the performance of the AC-restoration procedure and the quality of each approximated solution. For a Linear-OPF approximated solution to be strictly superior, it must be projected into a closer AC-feasible point and with a lower active generation cost associated than that obtained from the approximated solution of PTDF-OPF.

2.8.4 Summary of the testing methods

In order to present a summarized view of the different testing approaches considered to evaluate the performance of each of the linear Optimal Power Flow models in this work, the following lines encapsulate a brief description of the objectives and testing criteria for each of the testing methods.

REGRESSION TESTING

Objective: Evaluate the regression capabilities of the linear model matrices.

Evaluation criteria:

1. Maximum and average approximation errors of the active power flows.
2. PI index.

OPTIMIZATION TESTING

Objective: Evaluate the closeness of the approximate solution obtained from the linear models to the AC-OPF solution.

Evaluation criteria:

1. Difference in the objective function cost value.
2. L2-distance of the active generation profile from linear models.

FEASIBILITY TESTING

Objective: Evaluate the quality of the closest AC-feasible solution to the approximate one.

Evaluation criteria:

1. Cost of the closest AC-feasible solution.
2. Changes in the active generation profile necessary to converge to the closest AC-feasible solution.

2.9 Hardware and software configuration

In this section, we present the hardware and software characteristics of the setup used in the computation of the numerical simulations and results discussed in Chapters 3 and 4. The numerical computations were conducted on a desktop computer with the following specifications:

- Processor: 12th Gen Intel(R) Core(TM) i7-12700H, 2.70 GHz
- RAM: 16GB DDR4
- Graphics Card: NVIDIA GeForce GTX 3050 Ti
- Storage: 512GB SSD

All calculations and numerical results presented in this document have been obtained using Python 3 programming language. Python 3 is a open-source high-level programming language that is widely used for scientific computing, data analysis, and machine learning applications. One of the main advantages of using free software like Python 3 is its accessibility, as it can be downloaded and used by anyone at no cost. Additionally, free software promotes collaboration and knowledge-sharing, as users can modify and redistribute the software to meet their specific needs, allowing an easier identification and fixing of potential bugs and security vulnerabilities.

Among all the Python libraries used for the code (e.g. NumPy, Pandas, Matplotlib...), PandaPower, an easily accessible open-source tool designed for power system modeling, analysis, and optimization, has been utilized in the current work to acquire power networks and other power system tools. It features a high degree of automation and is implemented in Python. The optimization problems presented

in this document are solved with IPOPT (Interior Point OPTimizer), which is a optimization solver included in software package for large-scale nonlinear optimization of continuous systems. IPOPT is written in C++ and adapted to Python 3, and is released as open source code under the Eclipse Public License (EPL), available from the COIN-OR initiative.

Chapter 3

Illustrative example

Using a small illustrative example, such as a simplified power system with only a few buses and transmission lines, is a crucial preliminary step in the study of methods and techniques in the field of power systems. These simplified models provide a controlled and less complex environment for researchers to experiment with new methods and algorithms. This approach offers several advantages: it allows for quick testing and validation of concepts, reduces the risk of unintended consequences, and helps in identifying potential flaws or shortcomings in a safe and manageable setting. By employing an illustrative example as a starting point, researchers can refine their methods and gain valuable insights before transitioning to the complexities of real-world power systems.

3.1 Illustrative example data

With these considerations in mind, the present study showcases a reduced network as an illustrative example for the demonstration of the linear approximation methods proposed. This network is extracted from [20]. Referred to as *case6ww* by PandaPower (which imports the power system parametrizations from Matpower [21]), its schematic can be observed in Figure 3.1:

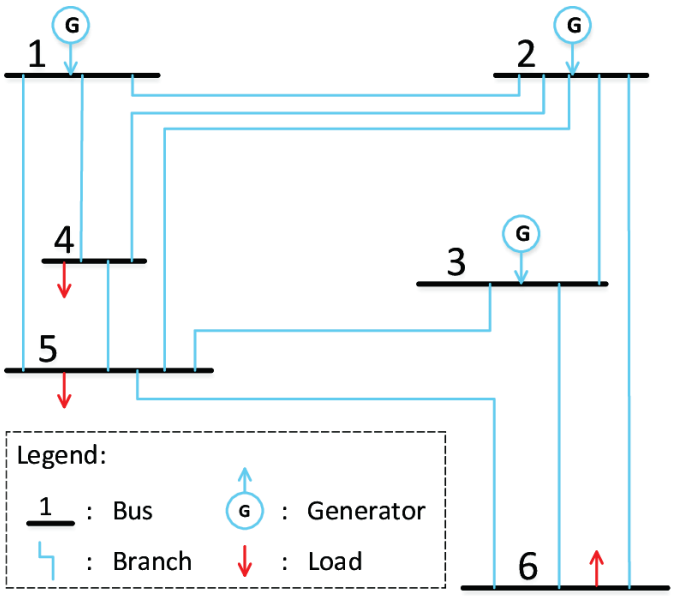


Figure 3.1: Diagram of the PandaPower test case 'case6ww' used as illustrative example.

The network presents a total of 6 buses (or nodes), interconnected by 11 transmission lines. Of these 6 buses, 3 of them (buses from 1 to 3) are characterized by a generating unit, whereas the other 3 buses (buses from 4 to 6) only contain a loading unit. In reality, the nature of each bus is not limited to a particular unit, but for demonstration purposes this configuration serves as an adequate illustrative example.

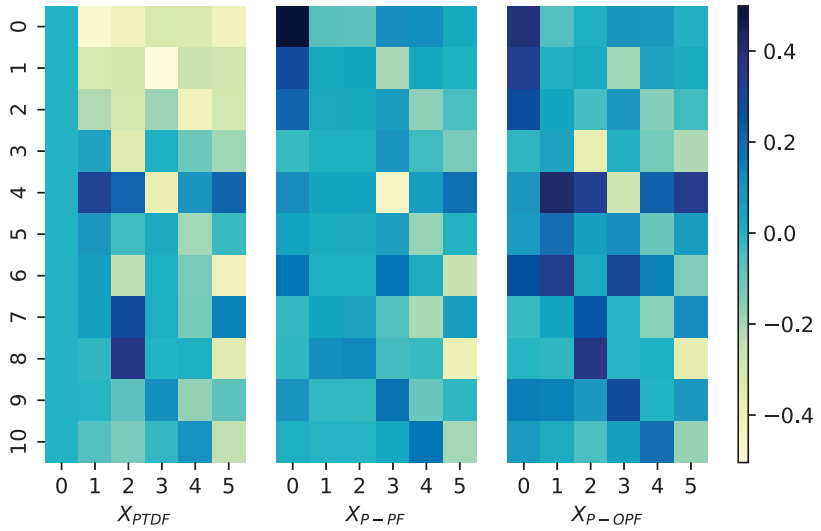
In order to depict a simplified network operation point, some changes have been made into the parameters that characterize the network. In particular, all loads have been set to 70 MW and 10 MVAr of active and reactive loads, respectively. The operational limits of the generating units, as well as the thermal limits of the lines, have been left as default. The parameters characterizing *case6ww* network can be found at the official web of Matpower `https://matpower.org/docs/ref/matpower5.0/case6ww.html`.

3.2 Illustrative example results

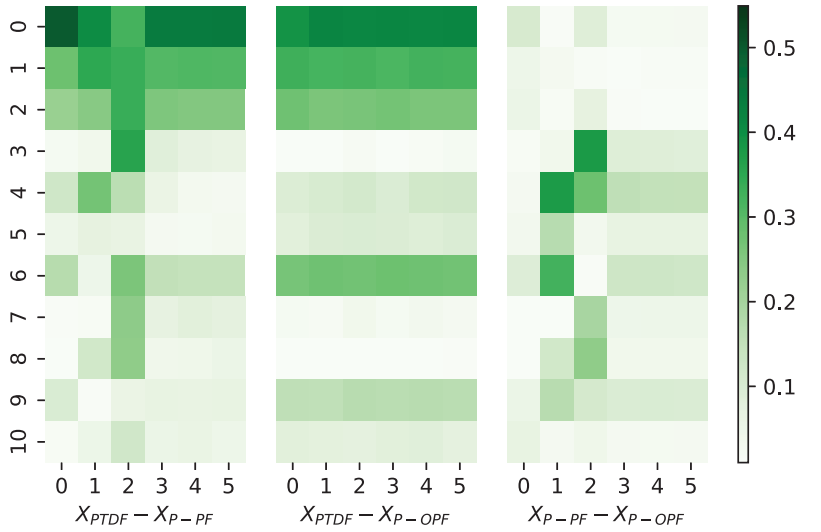
In this section, the numerical results obtained from the application of the different performance testing approaches to the Linear-OPF on the illustrative example network are presented.

3.3 Base formulations

To begin with, both the base Linear-OPF and PTDF-OPF formulations that only consider active power flows are considered. The different linear approximation matrices obtained are shown in Figure 3.2.



(a)



(b)

Figure 3.2: Heatmaps of the approximation matrices (a) and differences between each other (b).

Each of the colored cells depicted in Figure 3.2 (a) represents the factor each of the regressors (active power injections) with the predictors (active power flows). In particular, the element in the $i-j$ row-column position reflect the influence on active power flow in the i-th position of the active power injection in the j-th position. The colormap legend on the right states the meaning of each color: darker colors depict a strong positive correlation, while light colors depict a strong negative correlation.

The distribution of the factors in all linear approximation matrices are relatively similar. The main difference resides in the first column of the PTDF matrix, which is null by construction, corresponding to that of the slack bus. This column is colored in a light blue (corresponding to values near 0), and is a result of the balance equations in a lossless model, in which the power demand exactly equals the power generation. In this case, the power balance equation of one of the buses can be deduced from the results of the rest of the buses, not being necessary for the computation of the matrix.

However, this column on the linear model matrices X_{P-PF} and X_{P-OPF} is not null, as that restriction is not imposed by construction in the least squares method used. This allows the proposed Linear-OPF models to not be restricted in the matrix formulation to the assumption of a lossless network, as the imposition of this information in a net with non-negligible line losses may result in a less efficient linear regression model.

On the other hand, each of the colored cells depicted in Figure 3.2 (b) represent the difference in absolute value of corresponding factors between any pair of two matrices considered. This way, dark green colors highlight a strong discrepancy between factors on both matrices, while light green colors indicate the proximity of the values. While the X_{PTDF} matrix presents the most number of value differences with both X_{P-PF} and X_{P-OPF} in the first three rows (corresponding to the active power flows through the first three lines), the X_{P-PF} and X_{P-OPF} matrices differences are located sparsely on the active power injections at buses located in the second and third positions.

3.3.1 Regression Testing

The Regression Testing results of the benchmark PTDF-OPF model and the two proposed Linear-OPF models, Linear$_{PF}$-OPF and Linear$_{OPF}$-OPF, are shown in Figure 3.3.

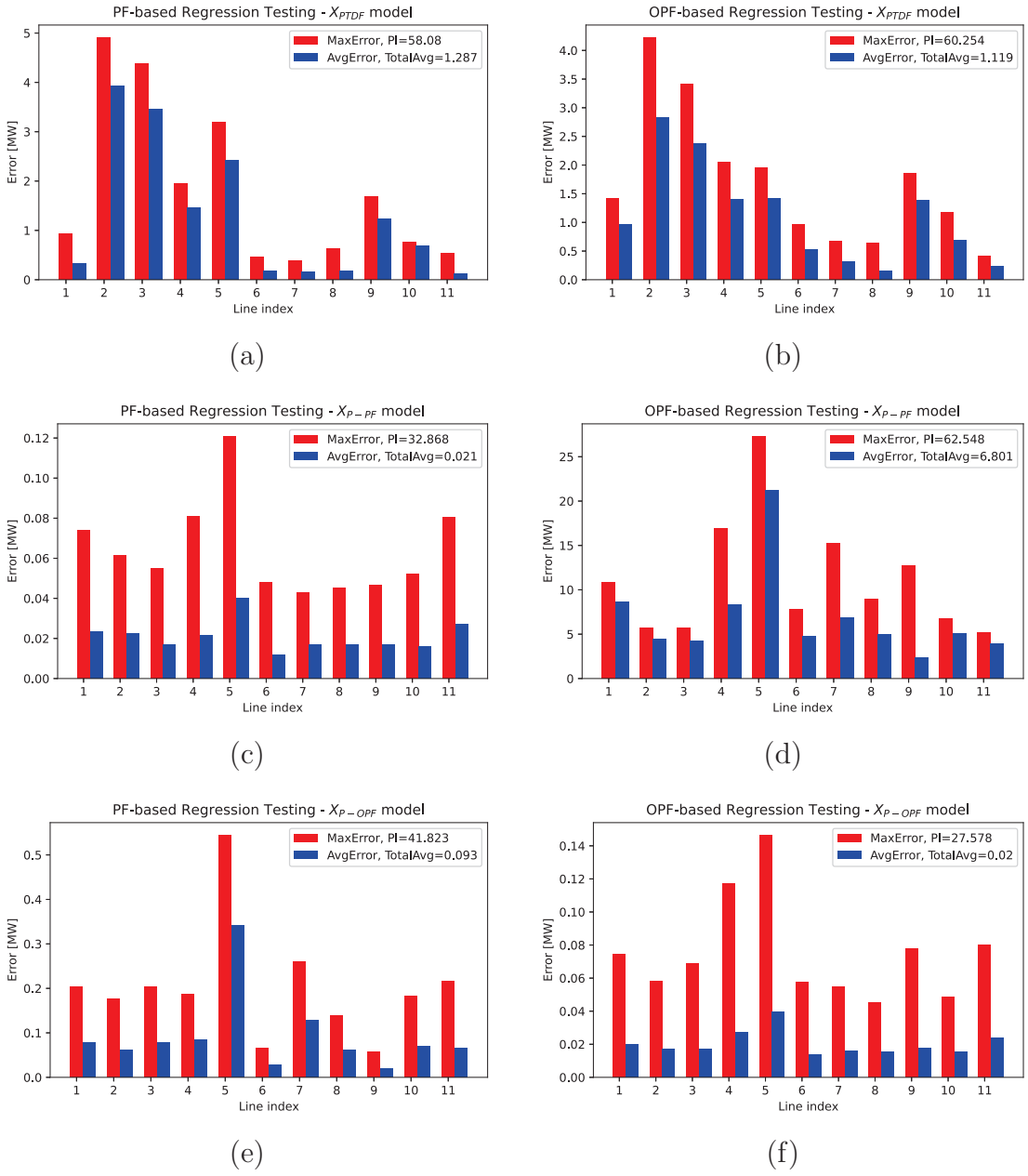


Figure 3.3: Regression Testing for the different base linear OPF models: PF-based on the left column and OPF-based on the right column. Linear model X_{PTDF} is represented in the first row, while linear models X_{P-PF} and X_{P-OPF} are represented in the second and third column, respectively.

On each of the six subfigures that compose Figure 3.3, the horizontal axis represents each of the 11 lines that compose the *case6ww* network, while in the vertical axis a measure of the active power flow approximation error is presented. Red bars represent the maximum approximation error for each line, while blue bars represent the average approximation error over all test scenarios for each line. In the legend located at the upper-right corner of each subfigure, additional information on the error distribution can be found, specifically the PI value and the total average approximation error value for all 11 lines. Each of the six subfigures represent a different linear approximation model tested against a different Regression Testing dataset. For the first and second columns, the PF-based and OPF-based Regression Testing are presented, respectively. On the first, second and third row, the different linear approximation models are ordered, being X_{PTDF}, X_{P-PF} and X_{P-OPF}, respectively.

To facilitate the comparison, the key results from the Regression Testing have been summarized in the tables below, for both the PF-based (Table 3.1) and OPF-based (Table 3.2) Regression Testing.

Table 3.1: Summary table of the performance of the various linear approximation methods analyzed for the PF-based Regression Testing of the *case6ww* power system.

PF-based Regression Testing			
OPF Formulation	**Avg.Error**	**Improvement over PTDF**	**PI index**
PTDF-OPF	1.287	-	58.080
Linear$_{PF}$-OPF	0.021	1.63 %	32.868
Linear$_{OPF}$-OPF	0.093	7.23 %	41.823

Table 3.2: Summary table of the performance of the various linear approximation methods analyzed for the OPF-based Regression Testing of the *case6ww* power system.

OPF-based Regression Testing			
OPF Formulation	**Avg.Error**	**Improvement over PTDF**	**PI index**
PTDF-OPF	1.119	-	60.254
Linear$_{PF}$-OPF	6.801	607.77 %	62.548
Linear$_{OPF}$-OPF	0.020	1.79 %	27.578

The first highlight that results from the Regression Testing is that the Linear-OPF models perform significantly better than the PTDF-OPF when the type of Regression Testing matches the training: the PF-trained Linear$_{PF}$-OPF shows an average error of only 1.63% of that of the PTDF-OPF for the PF-based Regression Testing, while the OPF-trained Linear$_{OPF}$-OPF shows an average error of only 1.79% of that of the PTDF-OPF for the OPF-based Regression Testing. On the

other hand, when the training does not match the type of Regression Testing, the results worsen for the Linear-OPF models: whereas the PF-trained Linear$_{PF}$-OPF shows an average error of 607.77% of that of the PTDF-OPF for the OPF-based Regression Testing, the OPF-trained Linear$_{OPF}$-OPF only performs slightly worse in the corresponding PF-based Regression Testing, with an average error of 7.23% of that of PTDF-OPF.

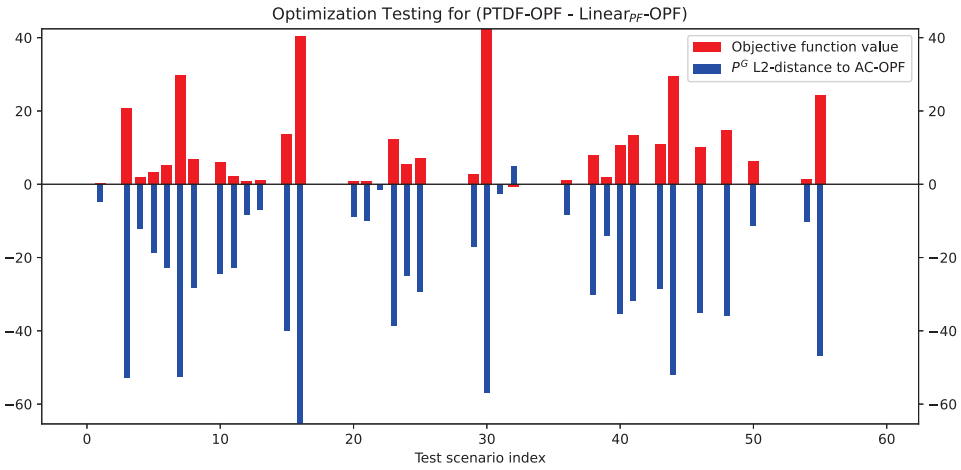
This trend is confirmed with the values of the PI index: the PTDF-OPF method presents a PI index over 60, indicating an average error at the same scale as the maximum error for all lines. On the other hand, the Linear-OPF models show a noticeably lower value, highlighting the value of around 27 for the Linear$_{OPF}$-OPF model for the OPF-based Regression Testing.

The results obtained in this section are validated with those presented in the literature in [10], where the direct least squares method is used to obtain a linear model designed for addressing collinear and massive data. The method is efficient for large systems with broad load variations, outperforming the PTDF-OPF model by several orders of magnitude as shown in Figs.5, 6 and 7 of the referenced paper.

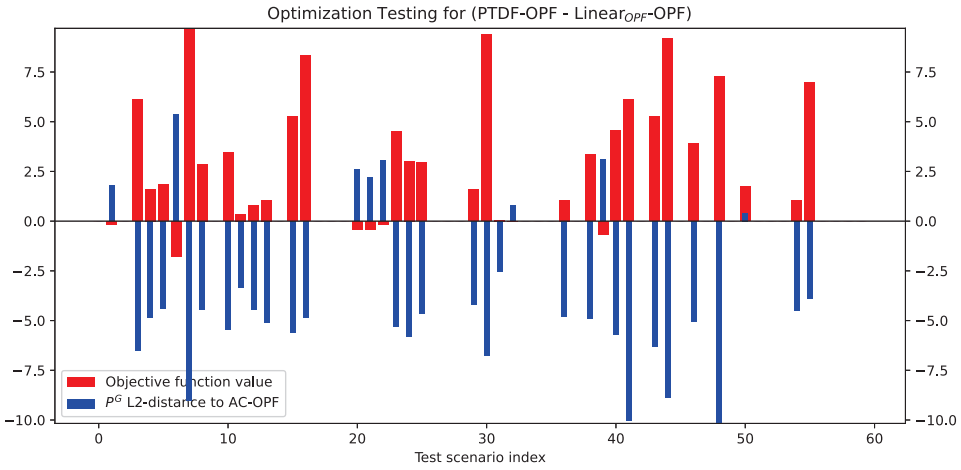
The Regression Testing analysis is not repeated for upgraded formulations of both Linear-OPF and PTDF-OPF problems. This is because the upgraded formulations only include modifications in the optimization problem (such as the consideration of reactive flows and power losses). Only the linear approximation matrices that relate the active power flows and the active power injections (X_{PTDF}, X_{P-PF} and X_{P-OPF}) are needed for the Regression Testing, and these remain the same for both base and upgraded formulations.

3.3.2 Optimization Testing

The Optimization Testing of the benchmark base PTDF-OPF model and the two proposed base Linear-OPF models, Linear$_{PF}$-OPF and Linear$_{OPF}$-OPF, is shown in Figure 3.4.



(a)



(b)

Figure 3.4: Optimization Testing for the base Linear-OPF models.

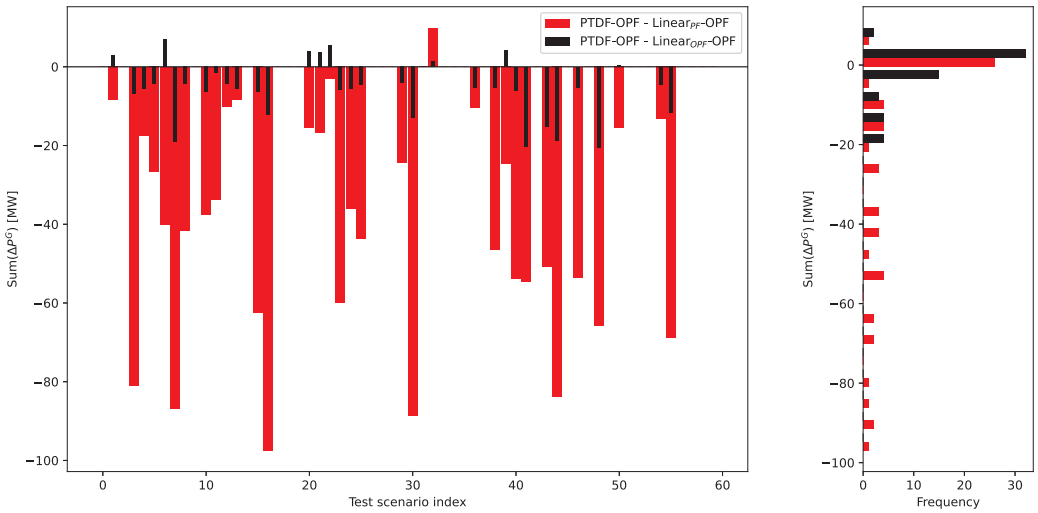
On the horizontal axis of subfigures 3.4 (a) and (b), the 60 test scenarios considered for the illustrative example case are presented. On the vertical axis, the two criteria pertaining to the Optimization Testing are plotted. The red bars represent the difference (between the approximate solutions of PTDF-OPF and Linear-OPF models, in this order) of the objective function values in [$]. Conversely, the blue bars show the difference (between the approximate solutions of PTDF-OPF and Linear-OPF models, in this order as well) of the L2-distance of the active generation profile to the one obtained from AC-OPF, in $[MW]$. The differences are displayed in this fashion so that a positive value on each test scenario shows an improvement of the Linear-OPF models over the PTDF-OPF model in any of both criteria.

We can observe that for both Linear-OPF models, the criteria of the objective function and the L2-norm distance of P^G are not consistent: while the PTDF-OPF shows a higher objective function value (cost of generation) than the Linear OPF models (presented by positive red bars), the L2-norm distance to the solution of the AC-OPF is higher for the Linear OPF models (presented by negative blue bars). In other words, approximate solutions from Linear-OPF models are cheaper than the ones obtained from PTDF-OPF but result in a active generation profile most distant to that of AC-OPF. It is important to extract as well an estimate of the percentage of improvement that the objective function criterion offers to the Linear-OPF models. Even if Figures 3.4 (a) and (b) show that the Linear-OPF models results in a solution less expensive than the one corresponding to the PTDF-OPF (\$40 for the Linear$_{PF}$-OPF and \$10 for the Linear$_{OPF}$-OPF at maximum), these improvements are almost negligible, corresponding to no more than 1% of the corresponding PTDF-OPF solution cost.

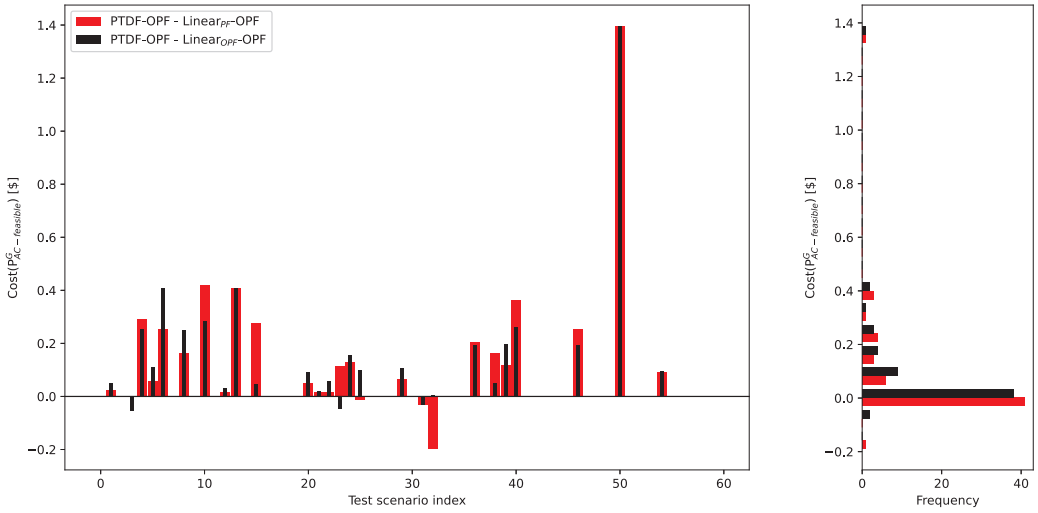
It should be emphasized as well that these solutions are not AC-feasible, and thus their results and the conclusions derived from them must be taken with a certain level of reservation.

3.3.3 Feasibility Testing

Once we have obtained the active generation profiles from each of the linear models considered, we solve the corresponding AC-restoration optimization problem PgenSlack-OPF and obtain AC-feasible solutions for each of these models.



(a)



(b)

Figure 3.5: Feasibility Testing for both base Linear$_{PF}$-OPF and Linear$_{OPF}$-OPF models.

Similarly to the figures from the Online Testing, the Feasibility Testing plots on Figure 3.5 show the 60 test scenarios considered for the illustrative example in the horizontal axis. On the vertical axis, the two criteria (one for each subfigure) pertaining to the Feasibility Testing are plotted. Note that red bars represent the Linear$_{PF}$-OPF model, whereas black bars represent Linear$_{OPF}$-OPF model. Similar to Figures 3.4 (a) and (b), positive values represent an improvement of the Linear-OPF models over PTDF-OPF model. On the right of each subfigure, a histogram of the values is presented to aid in the identification of trends on the data.

In Figure 3.5 (a), the sum of the absolute values of the slack variable ΔP^G, which expresses the change in MW that needs to be applied to the generation units so that an AC-feasible solution is obtained, is presented on the vertical axis. Almost all test scenarios report a higher value of the $Sum(\Delta P^G)$ for the Linear-OPF models than for the PTDF-OPF model, which result in negative values. However, it is important to note that the Linear$_{OPF}$-OPF model (black bars) clearly shows a better performance than the Linear$_{PF}$-OPF model (red bars), even reporting in some test cases a slight improvement over the PTDF-OPF model (positive black bars). This trend can be clearly seen on the histogram at the right of Figure 3.5 (a): almost al instances of Linear$_{OPF}$-OPF model present a value near 0, whereas many of the test cases for the Linear$_{PF}$-OPF model present the most negative values.

On the other hand, in Figure 3.5 (b) the other Feasibility Testing criterion is shown, corresponding to the economic cost of generation in [$] of the AC-feasible solution. Under this perspective, both the Linear$_{PF}$-OPF and Linear$_{OPF}$-OPF models perform almost equivalently, both showing an improvement over the PTDF-OPF model, represented by positive values on both red and black bars. Even if this economic improvement is only marginal, with a cost reduction of less than 0.1% of the total generation cost, we have to take into account that these solutions are actually AC-feasible and represent a realistic profile to be implemented in the power system, and that they consistently outperform the PTDF-OPF in the test scenarios studied.

3.4 Upgraded formulations

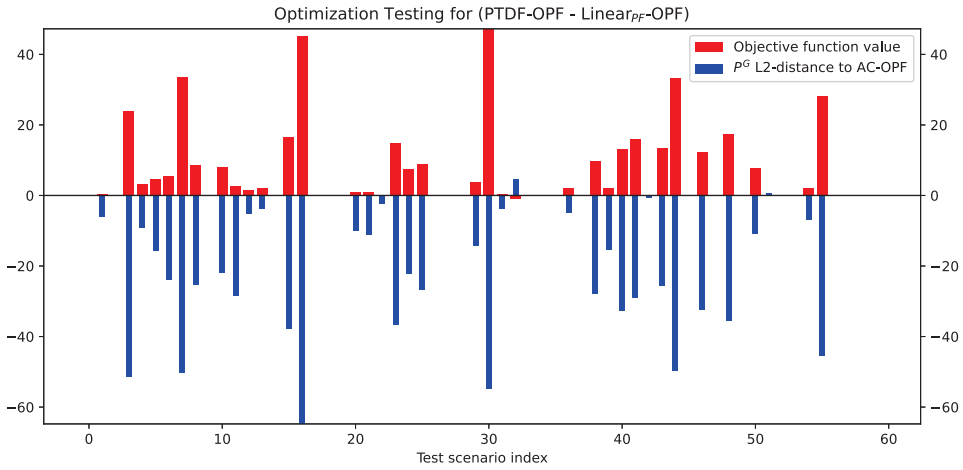
In this section, both the upgraded Linear-OPF (also considering reactive powers and power loss estimation) and PTDF-OPF (considering as well active power loss estimation) formulations are considered. The Regression Testing results are not shown again for the upgraded Linear-OPF model, as the values are exactly the same to those obtained for the base Linear-OPF model. This is because the linear approximation model used, X_P, is the only factor that determines the active power flow approximations from the active power flow injections, and neither the inclusion in the Linear-OPF optimization problem of a reactive power flow approximation X_Q nor the approximation of the active losses plays a role in determining the Regression Testing results. However, these latter factors do determine the performance of the model for the Optimization Testing and Feasibility Testing, whose results are shown below.

3.4.1 Optimization Testing

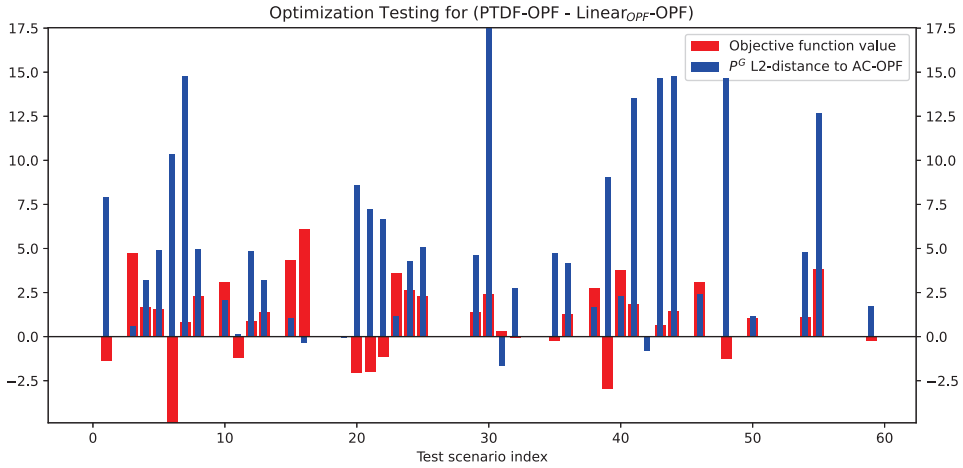
The Optimization Testing of the benchmark upgraded PTDF-OPF model and the two proposed upgraded Linear-OPF models, Linear$_{PF}$-OPF and Linear$_{OPF}$-OPF, is shown in Figure 3.4.

The inclusion of reactive power and loss estimation for the Linear$_{PF}$-OPF model does not change its performance at all, showing the same distributions in Figures 3.6 (a) and (b) as in the previous chapters: while the objective function value is lower than that of PTDF-OPF (positive red bars), the active generation profile is more distant to that of the AC-OPF (negative blue bars).

However, there is a radical change in the performance of the Linear$_{OPF}$-OPF model, which can be seen in Figure 3.6 (c). In this case, both Optimization Testing criteria are consistent and show that the solutions from the Linear$_{OPF}$-OPF model outperform the ones obtained from PTDF-OPF: not only they present a lower value for the objective function (positive red bars) but now the L2-norm distance from their active generation profile to that of AC-OPF is closer than the corresponding one proposed by PTDF-OPF (positive blue bars).



(a)

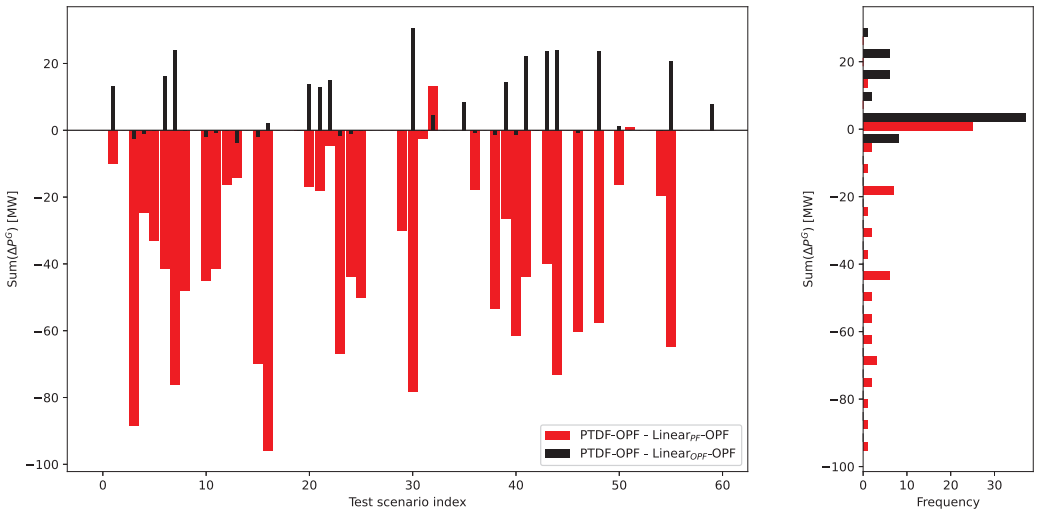


(b)

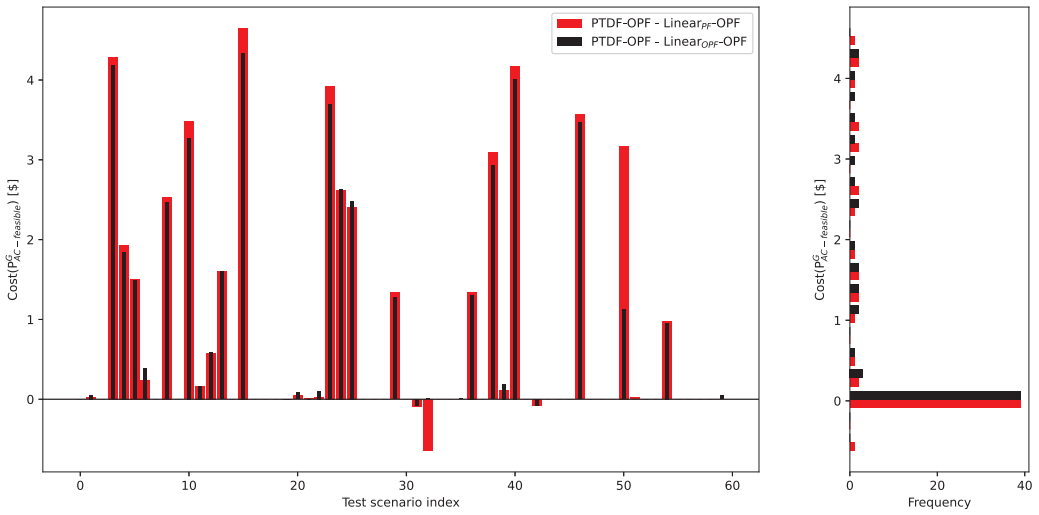
Figure 3.6: Optimization Testing for the upgraded linear models.

3.4.2 Feasibility Testing

Once obtained the active generation profiles from each of the upgraded linear models considered, we solve the corresponding AC-restoration optimization problem PgenSlack-OPF and obtain AC-feasible solutions for each of these models.



(a)



(b)

Figure 3.7: Feasibility Testing for the upgraded linear models.

For the upgraded linear models, a significant improvement can be observed in Figure 3.7 (a): for almost all test scenarios it is shown that the restored AC-feasible solution from the Linear$_{OPF}$-OPF model needs a lower change in active generation than PTDF-OPF (positive black bars). Compare this result to the less impressive performance of the base models in Figure 3.5 (a), where the Linear$_{OPF}$-OPF model provides a barely inferior active generation profile to that of PTDF-OPF (slightly negative black bars in most test scenarios). Despite that, the performance of the Linear$_{PF}$-OPF model is in this aspect still worse than PTDF-OPF (negative red bars).

On the other hand, the generation cost of the restored AC-feasible Linear-OPF model solutions are both significantly lower for both models than the PTDF-OPF, as it is presented in Figure 3.7 (b). The difference in generation cost between the Linear-OPF models and the PTDF-OPF has multiplied roughly by a factor of 4, showing no significant difference between the Linear$_{PF}$-OPF and Linear$_{OPF}$-OPF models.

3.5 Summary of the illustrative example results

In this section, we provide a brief overview of the key findings and outcomes obtained from the analysis of the *case6ww* illustrative example network with the benchmark PTDF-OPF model and the proposed Linear-OPF models.

The Regression Testing has proved the regression capabilities of the Linear-OPF models over the PTDF-OPF. The most favorable scenario for the Linear$_{OPF}$-OPF corresponds to the OPF-based Regression Testing in Figure 3.3 (f), where it scores an average error value of 0.02. Even in the most unfavorable scenario for the Linear$_{OPF}$-OPF (corresponding to the PF-based Regression Testing in Figure 3.3 (e)) its average error (with a value of 0.093) is almost two orders of magnitude inferior to the average error of the PTDF-OPF model for the same scenario (corresponding to the PF-based Regression Testing in Figure 3.3 (a), with an average value of 1.287). The Linear$_{PF}$-OPF shows a similar trend, although with less impressive upgrade margins to the PTDF-OPF.

After the inclusion of reactive powers and loss estimation in the upgraded linear models, the efficiency of the Linear$_{OPF}$-OPF is highlighted even more over the results presented by PTDF-OPF, as it can be observed in both the Optimization Testing (Figure 3.6 (b)) and the AC-feasible testing (Figure 3.7), where all evaluating criteria are favorable for the Linear$_{OPF}$-OPF with respect to the PTDF-OPF model.

On the other hand, the Linear$_{PF}$-OPF model does not maintain the same level of performance after these updates in the model, as it can be observed in the Optimization Testing (Figure 3.6 (a) and (b)) and the AC-feasible testing (Figure 3.7). For this reason, the rest of the document only considers the Linear$_{OPF}$-OPF for the evaluation of its performance in a larger power system.

Chapter 4

Case study

In this section, the performance of the proposed linear approximation models is analyzed for a larger electrical power system, more representative of the realistic cases solved in power system operations and management.

As stated in the previous chapter, the upgraded formulation of the Linear$_{OPF}$-OPF model resulted in the best performance results against the (upgraded) PTDF-OPF, outperforming the benchmark for each evaluation criterion. The proposed linear method trained with PF-data, Linear$_{PF}$-OPF model, does not perform as well against the PTDF-OPF, as it can be observed in the Optimization Testing (Figure 3.6 (a)) and the Feasibility Testing (Figure 3.7). For this reason, this chapter only considers the upgraded Linear$_{OPF}$-OPF model (with reactive powers and power loss estimation) for the evaluation of its performance against the upgraded PTDF-OPF formulation in a larger power system.

4.1 Case study data

The IEEE 39-bus system is a well-known power network that models the New England area of the United States. The model includes, as its name suggests, 39 buses, 10 generators and 46 lines. Generator 1 represents the aggregation of a large number of generators. The original paper with the IEEE 39-bus system can be found at [22]. The parameters characterizing the IEEE 39-bus system network can be found at the official web of Matpower `https://matpower.org/docs/ref/matpower5.0/case39.html`. The topology diagram of the system is presented in Figure 4.1.

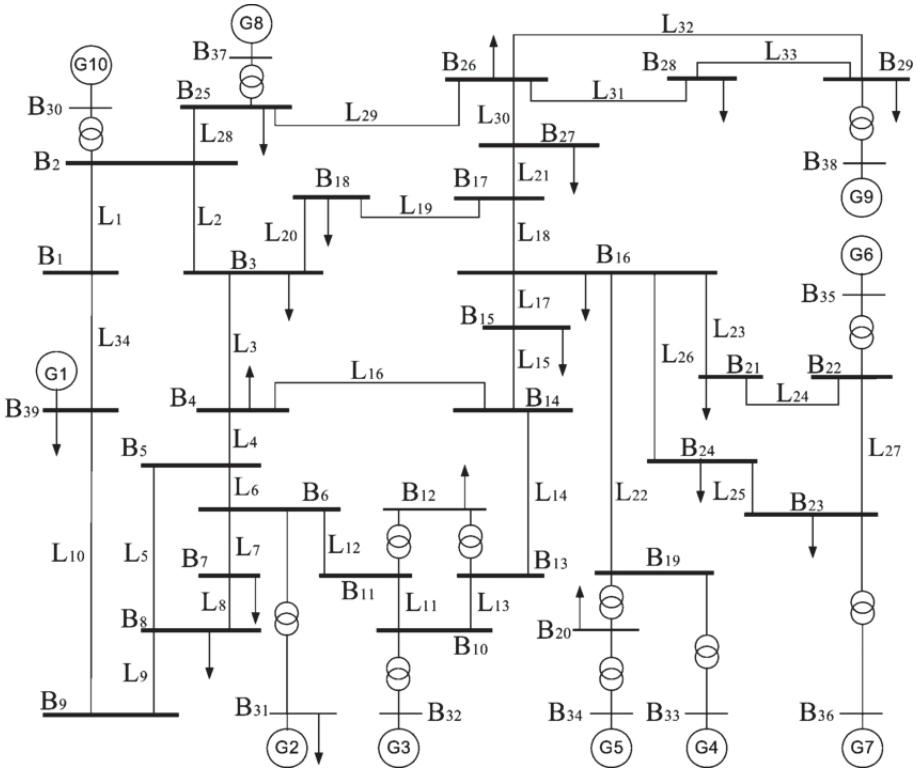


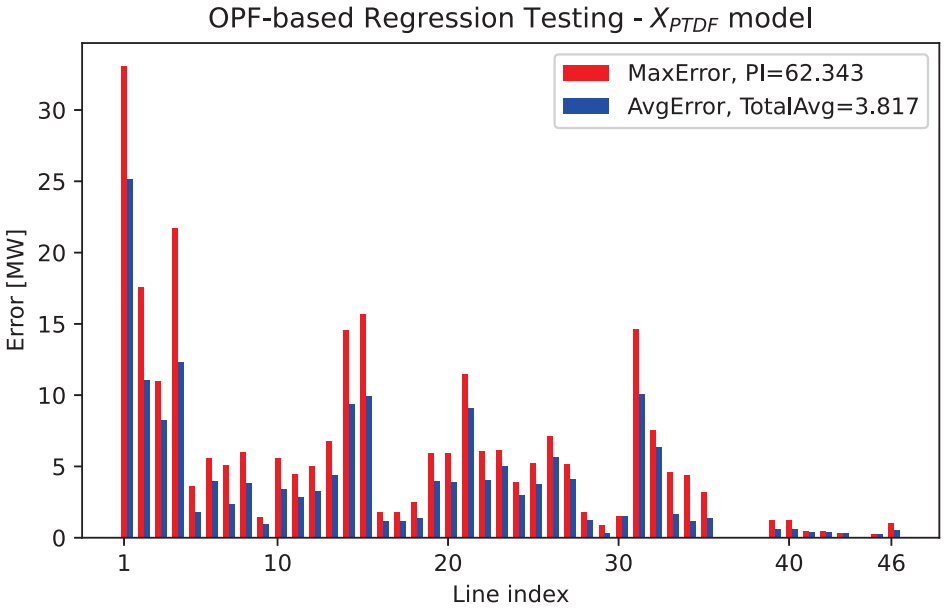
Figure 4.1: Diagram of the IEEE 'case39' network with 39 buses and 10 generating units, used as case study for the evaluation of the performance of the Linear$_{OPF}$-OPF method.

It can be observed that the detailed structure of the IEEE 39-bus system is much more intricate and meshed than the one presented by the *case6ww* illustrative example network considered in previous sections. Hopefully, this configuration will serve to test the regression capabilities of both models and give a glimpse of the potential that DPFL methods offer to the resolution of the AC-OPF problem.

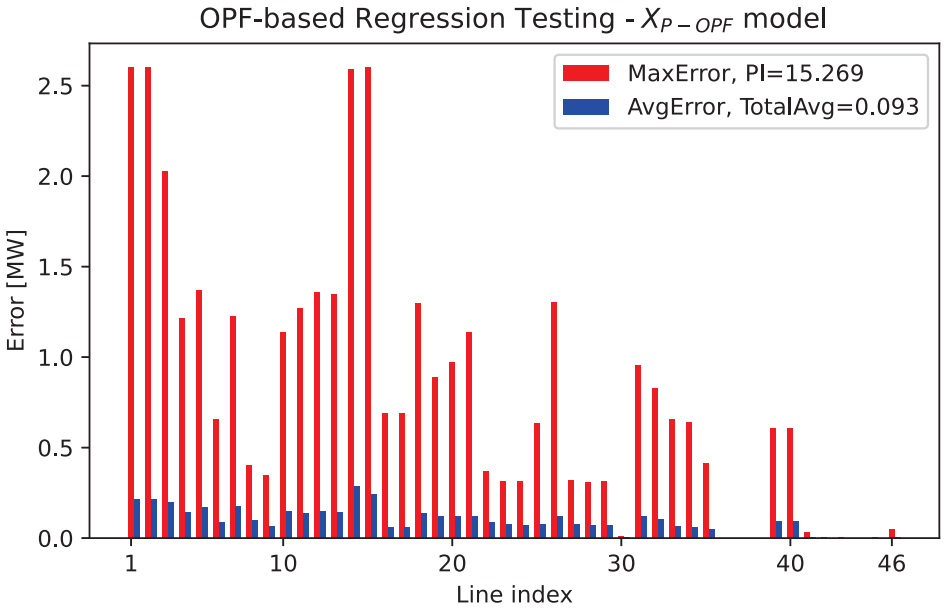
4.2 Case study results

In this section, the numerical results obtained from the performance comparison (via Online and Feasibility Testing) of Linear$_{OPF}$-OPF and the PTDF-OPF (both upgraded) formulations are presented for the case study IEEE 39-bus system.

In particular, the OPF-based Regression Testing of the benchmark PTDF-OPF model and the proposed Linear$_{OPF}$-OPF model is shown in Figure 4.2.



(a)



(b)

Figure 4.2: Offline OPF testing results of the case39 network for (a) PTDF and (b) XP-OPF.

The first aspect to take into account in the OPF-based Regression Testing presented for the IEEE 39-bus system is the order of magnitude of errors for both

formulations: in a similar fashion to the analysis of the *case6ww* illustrative example, the Linear$_{OPF}$-OPF model outperforms the PTDF-OPF model by almost two orders of magnitude in the average error evaluated over all lines.

This fact is further reinforced with the values of the PI index: the PTDF-OPF method presents a PI index over 60, indicating an average error at the same scale as the maximum error for all lines. On the contrary, the Linear$_{OPF}$-OPF models exhibit a significantly lower value of around 4 times lower ($PI \approx 15.3$), highlighting that the maximum errors at each line are not representative of the average performance of the model over all test cases.

These results align perfectly with the trends observed in Figures 5, 6 and 7 from [10]. It is also interesting to note that the error distribution along the lines for both formulations is very similar: while there are lines that exhibit a high error (such as lines 1 and 2) there are other that show almost a perfect linear correlation between the active power flows and the active injected power on the corresponding buses. The reason for the presence of noticeable error peaks in specific lines can be atributted, according to [10] to two possible factors:

- The line corresponds to a transformer branch or low-resistance branch, which tends to carry large power flows in practice, expecting a larger error value from a branch with larger power flow value.
- The susceptance b of the transformer branch or low-resistance branch is generally larger than that of transmission branch, which can amplify the nonlinear term of AC-PF equations.

In the following pages, the results pertaining to the AC-feasible restored solutions for both OPF formulations, PTDF-OPF and Linear$_{OPF}$-OPF, are presented in a condensed format with the use of histograms. This arrangement allows the clear visualization of any patterns appearing in the comparison of both formulations, without the need to check on every test case. On the vertical axis of these histograms, the number of ocurrences or frequency of each value for each of the bins appears. A total number of 40 bins has been chosen to display the results.

Firstly, the results corresponding to the L2-distance of the active generation profiles of the approximated solutions to the closest AC-feasible solution (first performance criterion of the Feasibility Testing) are presented in Figure 4.3. In particular, the difference of the distance between the PTDF-OPF formulation and the Linear$_{OPF}$-OPF formulation is chosen to help visualize the performance of the proposed method: if the value is positive, the AC-restored approximated solution from the Linear$_{OPF}$-OPF is closer (than the one obtained from PTDF-OPF) to its closest AC-feasible solution. The distance is measured as the sum of the absolute values of

the changes in the active generation units needed to restore the approximate solution to an AC-feasible state, in the same fashion as presented in the histogram of Figure 3.7 (a) from the illustrative example analysis.

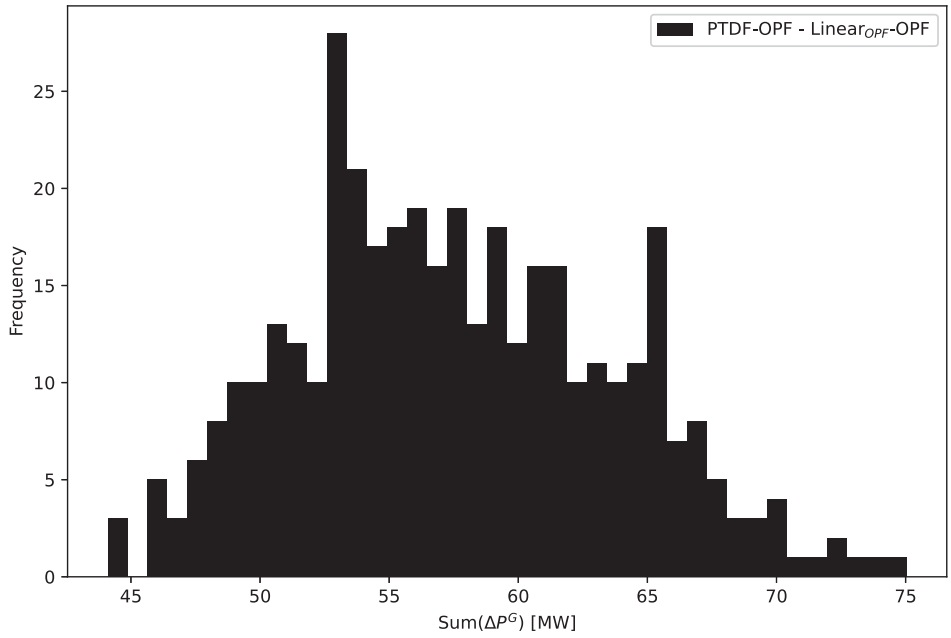


Figure 4.3: Difference between PTDF and XP-OPF of the distance of the approximated solution and the closest AC-feasible solution in the active generation profile space.

The histogram illustrates a distribution that bears a loose resemblance to a normal distribution, with all of its values falling within the positive range. This means that the approximated solution obtained from the Linear$_{OPF}$-OPF formulation is closer to its AC-feasible closest solution than the case for the PTDF-OPF formulation, favouring the approximate starting point solution of the Linear$_{OPF}$-OPF for the AC-restoration of the PgenSlack-OPF problem.

It is important to understand these results with relation to the full picture of the problem. Considering the mean value of the difference in active generation L2-distance between formulations to be around 60 MW, as presented in Figure 4.3, and the average value of the total active generation power over all AC-feasible solutions considered to be 6292 MW, this amounts to a difference of approximately 1% between formulations. In other words, the proposed Linear$_{OPF}$-OPF formulation is capable of providing starting points for the obtention of AC-feasible active power generation profiles that consistently outperform the solutions proposed by the

PTDF-OPF formulation, extensively used as a benchmark in the technical literature.

In the following lines we present a comparison of the generation costs for the AC-feasible solutions obtained from the approximated solutions from both formulations is presented in Figure 4.4 in histogram format. In particular, the difference of the generation cost between the PTDF-OPF formulation and the Linear$_{OPF}$-OPF formulation is chosen: if the value is positive, the AC-restored approximated solution from the Linear$_{OPF}$-OPF is cheaper than the one obtained from PTDF-OPF.

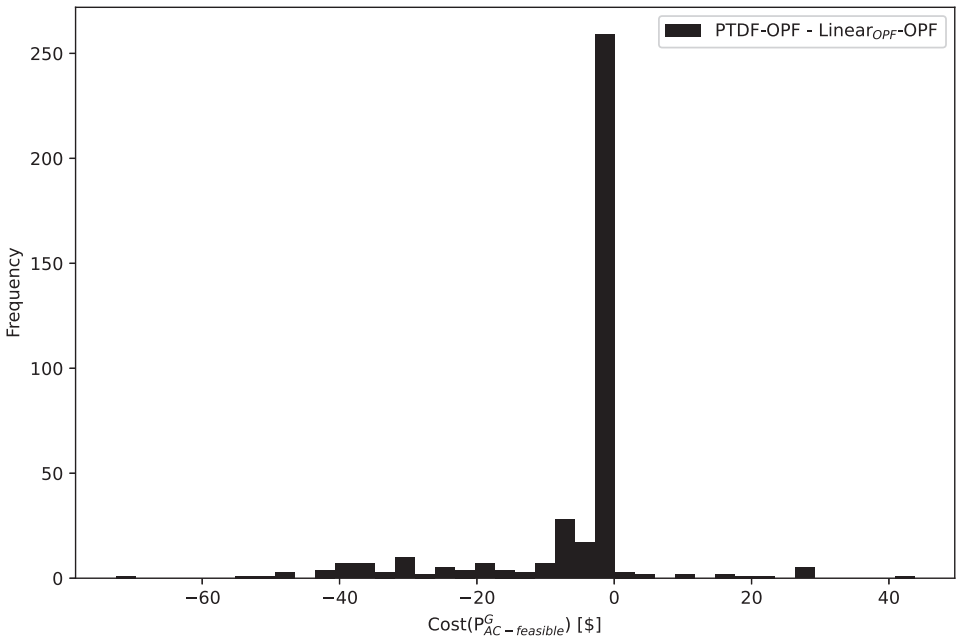


Figure 4.4: Difference between the costs of the AC-feasible active generation profiles restored from the approximate solutions of PTDF-OPF and Linear$_{OPF}$-OPF formulations.

We can observe that for most of the test cases, the difference in AC-feasible generation cost in negligible, result that is reflected in the peak appearing around 0. Even if there are some marginal cases for which the generation cost of the Linear$_{OPF}$-OPF solution is cheaper (to the right of the histogram, with positive values), we find that there are more instances where the PTDF-OPF AC-restored solution is slightly cheaper.

However, it is important to take these results into account relative to the average cost of the AC-feasible generation profile, which amounts to approximately $41991. Even if we consider the least favorable cases, in which the PTDF-OPF formulation

achieves an AC-feasible solution \$70 cheaper than the AC-feasible generation profile proposed by the Linear$_{OPF}$-OPF, that improvement only amounts to less than 0.2%. This result, along with the fact that the vast majority of AC-feasible solutions don't present any difference in the cost between both formulations, as it can be seen in the peak of the histogram at 0, suggests that the choice between both formulations has minimal impact on the overall generating cost of AC-feasible solutions.

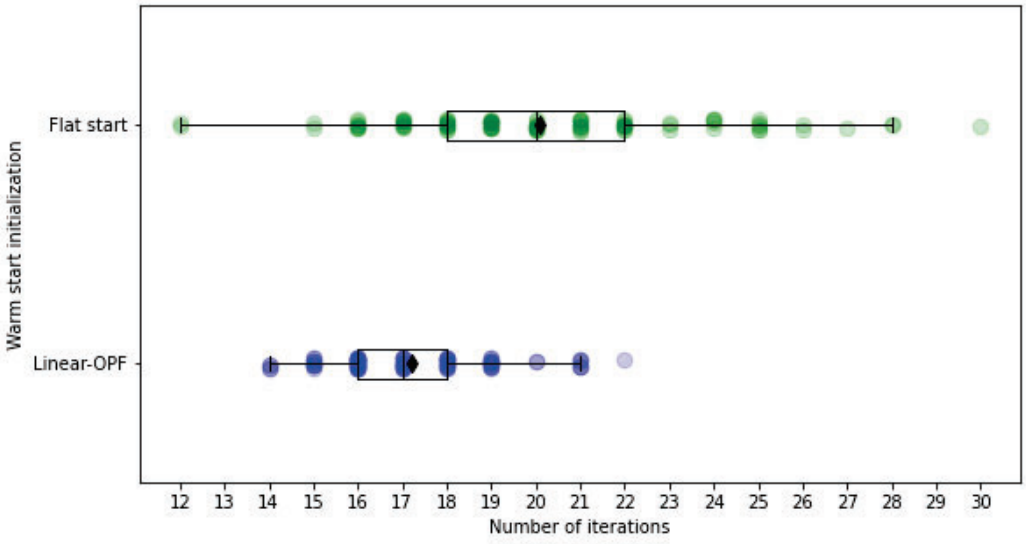
4.3 Computational comparison

Up until this point, the different criteria to assess the validity of the Linear$_{OPF}$-OPF approximated solutions have been focused on the quality of the solution itself with respect to other benchmarks (such as the PTDF-OPF approximated solution). However, when scaling the problem to a larger network, other requirements regarding the computational efficiency of the optimization problem resolution may have a more immediate effect, such as:

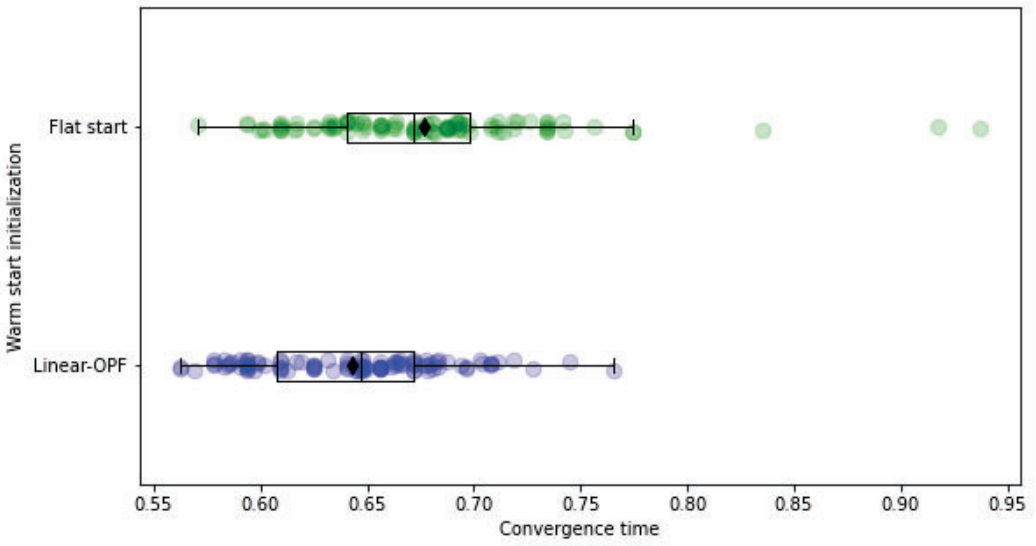
- **Convergence Time**: measuring the convergence time of the optimization solver provides a valuable indicator of its performance. By assessing the time taken to solve the problem, the computational cost can be estimated and it is possible to determine if the solver meets your requirements. Comparing convergence times across different problem sizes or solver configurations can help as well to understand scalability and efficiency.
- **Number of iterations**: the iterative nature of optimization methods implies that the algorithm refines its solution through successive updates, and the count of iterations directly reflects the computational effort invested. Monitoring the number of iterations is valuable for assessing the solver's convergence speed, enabling comparisons between different algorithms.

Due to the non-convex, non-linear nature of the AC Optimal Power Flow problem and its constraints, using iterative numerical methods to find a solution may encounter convergence failures. Employing warm start solutions to initialize the iterative procedure (typically a Newton-Raphson-based iterative method) can aid in ensuring convergence [23]. Grid operators commonly perform this warm start by using the previous timestep's AC-OPF solution to initialize the current timestep. This can be done using a flat start or by first solving an easier, linearized version of the AC-OPF equation, assuming that these solutions will be relatively close. Addressing this final aspect, we use the active generation profile obtained from the approximate Linear$_{OPF}$-OPF solution to initialize the non-linear solver before computing the AC-OPF. In Figure 4.5, we present the results of this convergence analysis, showing

both the number of iterations and convergence time needed for the flat start and Linear$_{OPF}$-OPF warm start initialization.



(a)



(b)

Figure 4.5: Comparison of the computational performance of the AC-OPF with several warm-start points: flat-start (green) and active generation profile from Linear$_{OPF}$-OPF (blue). Both the (a) number of iterations and the (b) computational time needed for the AC-OPF to converge are presented.

The black diamonds showed in Figure 4.5 represent the average value of each of the variables presented, while the box and whisker plots help to understand the distribution of the data. From Figure 4.5 (a), it can be observed that the distribution of the number of iterations for the AC-OPF with a flat start initialization is much more distributed and has a higher average value, around 20 iterations. On the other hand, the Linear$_{OPF}$-OPF warm start offers a much more robust distribution, with a lower average value of around 17 iterations, which represents a reduction of approximately 15%.

Furthermore, the distribution of computational convergence times for both initializations of the AC-OPF problem show a similar trend: the Linear$_{OPF}$-OPF warm start method offers a reduction on the average convergence time of around 5%. These results confirm that the initialization of the AC-OPF problem from the Linear$_{OPF}$-OPF approximate solution is more favorable than using the flat start initialization, often chosen as default setting for the AC-OPF resolution.

Chapter 5

Conclusions and future lines of work

This chapter summarizes the main conclusions of this study, as well as suggestions for potential future lines of research.

5.1 Key conclusions

The AC Optimal Power Flow formulation is characterized by the complex non-linear equations of AC Power Flow. In order to assess the difficulties of efficiently solving this optimization problem, many linear physics-based models have been proposed in the literature, such as the linear PTDF matrix approximation. This work presents a data-based linear model, Linear-OPF, that establishes a linear relationship between the injected active power in the nodes and the power flows through the lines. Unlike the physics-based PTDF linear approximation matrix, the proposed model learns this relationship from system-specific data, obtained from the resolution of multiple AC-PF and AC-OPF problems.

Several evaluation tests are deviced to assess the performance of the proposed model in two illustrative power systems with 6 and 39 buses, respectively. The analysis of the performed simulations shows that the proposed data-based linear approximation model provides solutions that are closer to that of the AC-OPF formulation when compared to the PTDF-OPF formulation, used as a benchmark in the technical literature. In particular, the improvement on the AC-feasible active generation setpoints found for the data-based Linear-OPF model with respect to the physics-based PTDF-OPF model is about 1%, showing promise as an alternative linear formulation to the full non-convex resolution of the classic AC-OPF.

5.2 Future lines of work

Once the main conclusions of the work have been presented, it is essential to summarize as well the future lines of work available, considering the results presented here as a foundational base for further research.

1. **Historical information as training data**: Ideally, the training of the linear model would be composed of feasible training cases derived from real operating points of the net, obtained in practice. In particular, the recently developed *OPFLearnData* database [24] was developed as a means to provide researchers with a standardized way to efficiently create AC-OPF datasets that are representative of more of the AC-OPF feasible load space compared to typical dataset creation methods. The *OPFLearnData* dataset creation method uses a relaxed AC-OPF formulation to reduce the volume of the unclassified input space throughout the dataset creation process. Implementing this tool for the generation of more complete training sets may aid in capturing the full behaviour of the power system in more diverse loading conditions.

2. **AC-restoration methods**: As stated in [25], further research is essential to identify the most effective means of restoring AC-feasibility, considering the cost differences associated with different controller actions.

3. **Application to other power systems**: Future work could as well be directed to assessing whether the results presented here remain valid, and to which extent, for other types of net distributions such as meshed distribution networks and DERs with more complex price responses.

Appendix A

Optimization problems

Numerous challenges across diverse domains such as engineering, medicine, and economics involve the task of identifying optimal decisions that align with specified objectives while adhering to predefined limitations. Mathematically, these decision-based problems in power system operations are commonly cast into the framework of optimization problems, exemplified as the following instance:

$$\begin{aligned} \min_{x} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0 \quad \forall i \\ & g_j(x) = 0 \quad \forall j \end{aligned} \tag{A.1}$$

where $x \in \mathbb{R}^n$ are the decisions variables, $f_0 : \mathbb{R}^n \longrightarrow \mathbb{R}$ is the objective function, and $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $g_j : \mathbb{R}^n \longrightarrow \mathbb{R}$ are, respectively, the inequality and equality constraints. The properties of the optimization problem stated and the mathematical techniques required to solve it vary significantly depending on the type of decision variables x considered (continuous, discrete, etc.) and the type of functions f_0; f_i; g_j (linear, convex, non-convex, etc.).

For the general optimization problem to be a convex optimization problem, the objective function f_0 and the inequality constraint f_i must be convex functions, and the equality constraints g_j must be affine functions [16]. The main advantage of convex optimization is that local optimal solutions are guaranteed to be global optimal solutions as well. On such basis, the technical literature includes a variety of interior-point methods to solve convex optimization problems in reasonable computational times [26]. Convex optimization problems can be used to formulate a vast variety of power system applications, some of which are described in detail in [27].

If any of the functions f_0 or f_i are non-convex, or any of the equality constraints g_j

are not affine, the general optimization problem becomes a non-convex optimization problem. Interior-point methods can be used to find local optimal solutions of non-convex optimization models [28]. However, local optima are not guaranteed to be global optima due to non-convexities. In some applications, this drawback is overcome by applying interior-point methods with multiple random initial solutions and by taking the global solution as the best local solution [29]. Alternatively, some of the global optimization techniques reviewed in [30] can be used.

Appendix B

Least squares regression

The least squares regression model adopts the following formulation:

$$Y = \beta X \tag{B.1}$$

Where $X = [x_1, ..., x_{Ns}]^T \in \mathbb{R}^{N_s \times N_x}$ and $Y = [y_1, ..., y_{Ns}]^T \in \mathbb{R}^{N_s \times N_y}$ are datasets of N_s samples in which each column represents a vector of measurements of N_x and N_y elements, respectively. The ordinary least squares approach finds the coefficient matrix β (also known as the regression coefficient) by minimizing the sum of squared residuals:

$$\min_{\beta} \quad ||Y - X\beta||_F^2 \tag{B.2}$$

Where $||\cdot||_F$ is the Frobenius norm, also known as Hilbert-Schmidt norm. The estimation of β, which constitutes the solution to the above unconstrained programing problem, can be expressed by:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{B.3}$$

Appendix C

Per unit system

In the study of power systems, electrical parameters are often represented as the ratio of their actual SI (International System of Units) value to a reference or base quantity. This transformation of variables is commonly referred to as the 'per unit system.' Base quantities are expressed in SI units such as Volts, Amperes, Ohms, and more, while per unit quantities are dimensionless and are typically denoted by the label 'p.u.' if required. The per unit value of an SI quantity, denoted as X with respect to a specified base value X_{Base}, is given by:

$$X_{PU} = \frac{X}{X_{Base}} \tag{C.1}$$

The accurate interpretation of the SI value of a per unit quantity necessitates having knowledge of the associated base quantity.

In power systems analysis, base quantities are defined for voltage, current, power, impedance, and admittance. When any two system bases are specified, the values of the others are uniquely determined. In three-phase power factor analysis, it is customary to define the voltage and power bases as a standard convention.

$$\begin{cases} V_{Base} = \text{Line to line voltage} \\ S_{Base} = \text{Three phase power} \end{cases} \tag{C.2}$$

For mathematical convenience and ease of calculation, the two fundamental per unit bases have been considered in this work to be $S_{Base} = 100$ MVA and $V_{Base} = 230$ kV. The remaining base system bases are thus calculated according to the following expressions:

$$I_{Base} = \frac{S_{Base}}{\sqrt{3} \cdot V_{Base}} \tag{C.3}$$

$$Z_{Base} = \frac{V_{Base}}{\sqrt{3} \cdot I_{Base}} \tag{C.4}$$

$$Y_{Base} = \frac{\sqrt{3} \cdot I_{Base}}{V_{Base}} = \frac{S_{Base}}{V_{Base}^2} = \frac{1}{Z_{Base}} \tag{C.5}$$

The use of the per unit system offers several advantages over the use of the SI system of measurement, notably:

- Convergence: per unit quantities maintain consistent magnitudes around 1.0, enhancing the numerical stability of power flow calculations.
- Standarization: per unit system eliminates the necessity to differentiate between single-phase and three-phase electrical quantities.
- Clarity: the per unit system is more straightforward to interpret at a glance. For instance, per unit voltage should consistently fall within the approximate range of 0.95 to 1.05 p.u., regardless of the SI voltage.
- Simplification: the use of per unit values eliminates the need to apply voltage and current scaling factors at the majority of system transformers.

Appendix D

Admittance matrix

In this section, the construction of the admittance matrix of a power system is presented.

For the derivation of the admittance matrix form, a generic bus i, as presented in Figure D, is considered. The shunt susceptance $y_s i$ (named in the present document as b_i^s, and equivalent to half the shunt susceptance b_{ij}^s in the line ij) models the total charging current.

The net current that is injected into bus i either by generators or loads, I_i, can be calculated through Kirchoff's current law:

$$I_i = \sum_{j \in M_i} y_{ij}(V_i - V_j) + b_i^s V_i \tag{D.1}$$

Where M_i is the subset of buses that are directly connected to bus i, and V_i represents the complex voltage at bus i. Rearranging the previous terms leads to:

$$I_i = \left[\sum_{j \in M_i} y_{ij} + b_i^s\right] V_i - \sum_{j \in M_i} y_{ij} V_j \tag{D.2}$$

The repetition of the above development for the N buses leads to the nodal equations, presented below in matrix form:

$$\boldsymbol{I} = Y\boldsymbol{V} \tag{D.3}$$

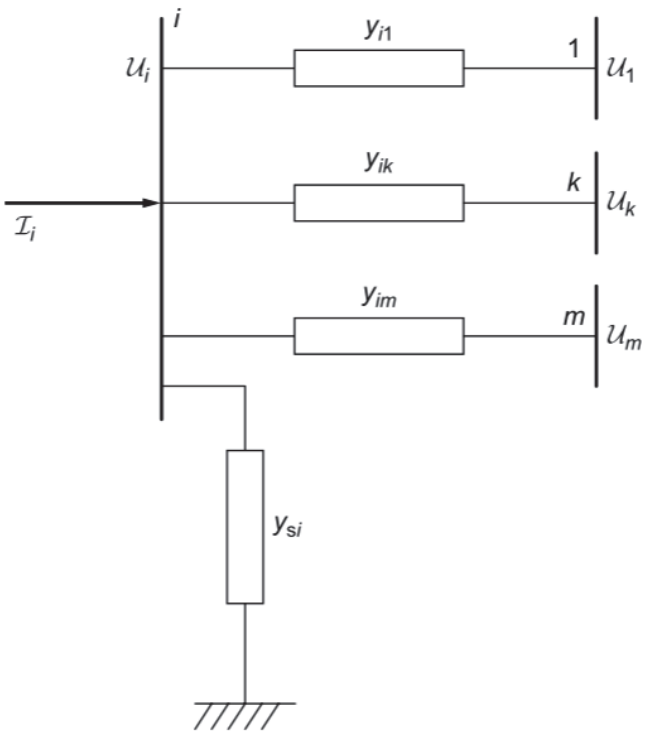


Figure D.1: Diagram of the different electric elements connected to a generic bus i. Figure extracted from [12].

The resulting $N \times N$ matrix Y is the so-call admittance matrix. Each of its elements present a real (conductance) and imaginary (susceptance) component, which may be collected in the matrices G and B, respectively. Each of the elements of Y can be computed in the following fashion:

$$Y_{ii} = \sum_{j \in M_i} y_{ij} + b_i^s \tag{D.4}$$

$$Y_{ij} = -y_{ij} \quad \forall j \in M_i \tag{D.5}$$

In essence, the construction of the admittance matrix Y can be summarized in the following points:

- Diagonal elements Y_{ii}: sum off all admittances connected to the i bus.
- Off-diagonal elements of Y_{ij}: negative admittance that connects the i and j buses.
- Several admittances in parallel: the equivalent admittance must be previously computed by calculating the sum of all parallel admittances.

Appendix E

Power Flow formulations

In this chapter the different formulations of the Power Flow problems considered are presented.

E.1 AC-PF

In this section, the formulation for the classic AC-PF equations is presented. This model serves as the governing equations for the power flow in electrical power systems. The details of this formulation can be found at Section 2.2.

$$\begin{cases} P_i^G - P_i^D = \sum_{j \in M_i} P_{ij}^L & \forall i \in N \\ Q_i^G - Q_i^D = \sum_{j \in M_i} Q_{ij}^L & \forall i \in N \\ \\ P_{ij}^L = V_i V_j [G_{ij}\ cos(\theta_i - \theta_j) + B_{ij}\ sin(\theta_i - \theta_j)] - G_{ij} V_i^2 \\ Q_{ij}^L = V_i V_j [G_{ij}\ sin(\theta_i - \theta_j) - B_{ij}\ cos(\theta_i - \theta_j)] + [B_{ij} - b_{ij}^s] V_i^2 \end{cases}$$

E.2 DC-PF

In this section, the formulation for the DC-PF model is presented. This PF model is a simplification of the classic AC-PF equations in Appendix E.1. This model does not consider neither power losses nor reactive power, and approximates the active power flows through the difference between adjacent voltage angles and line reactance. The full derivation of the DC-PF formulation can be found at Section 2.4.

$$\begin{cases} P_i^G - P_i^D = \sum_{j \in M_i} P_{ij}^L & \forall i \in N \\ \\ V_i \approx 1 \\ g_{ij} \approx 0 \\ -b_{ij} \approx \frac{1}{x_{ij}} \\ cos(\theta_i - \theta_j) \approx 1 \\ sin(\theta_i - \theta_j) \approx \theta_i - \theta_j \\ P_{ij}^L = \frac{(\theta_i - \theta_j)}{x_{ij}} \end{cases} \tag{E.1}$$

Appendix F

Optimal Power Flow formulations

In this chapter the different formulations of the Optimal Power Flow problems solved in this document are presented.

The initialization of the Optimal Power Flow variables is done by means of the flat start, setting the voltages in the buses to 1 and the angles of voltages in the buses to 0. Another option to initialize the voltages would be to consider the average value of the superior and inferior limits, but its implementation makes no significant changes to the solutions obtained for the networks considered.

The slack bus, considered here without loss of generality as the first bus listed in the PandaPower net (which also happens to be the one connected to the external grid), has an angle of voltage set to 0.

The cost which is set to be minimized in the OPF problems is obtained through the active power generation in the generator buses, multiplied by constant coefficients (dependent on the generator) up to the quadratic terms. The only exception is the PgenSlack-OPF formulation in Appendix F.7, which minimizes the active power generation deviation ΔP^G.

F.1 AC-OPF

In this section, the formulation for the classic AC-OPF model is presented. In this model, both the active and reactive power flows are calculated as a non-linear function of the voltages and angles on the buses of the network. The details of each restriction in this formulation can be found at Section 2.3.

$$
\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & P_{ij}^L = V_i V_j [G_{ij}\ cos(\theta_i - \theta_j) + B_{ij}\ sin(\theta_i - \theta_j)] - G_{ij} V_i^2 \quad \forall i, j \in N \\
& Q_{ij}^L = V_i V_j [G_{ij}\ sin(\theta_i - \theta_j) - B_{ij}\ cos(\theta_i - \theta_j)] + (B_{ij} - b_{ij}^s) V_i^2 \quad \forall i, j \in N \\
& P_i^G - P_i^D = \sum_{j \in M_i} P_{ij}^L \quad \forall i \in N \\
& Q_i^G - Q_i^D = \sum_{j \in M_i} Q_{ij}^L \quad \forall i \in N \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& Q_i^{G,min} \leq Q_i^G \leq Q_i^{G,max} \quad \forall i \in N \\
& V_i^{min} \leq V_i \leq V_i^{max} \quad \forall i \in N \\
& \theta_i^{min} \leq \theta_i \leq \theta_i^{max} \quad \forall i \in N \\
& (P_{ij}^L)^2 + (Q_{ij}^L)^2 \leq (S_{ij}^{max})^2 \quad \forall i, j \in N
\end{aligned}
\tag{F.1}
$$

F.2 DC-OPF

In this section, the formulation for the DC-OPF model is presented. This OPF model is a simplification of the classic AC-OPF model in Appendix F.1. This OPF model does not consider neither power losses nor reactive power, and approximates the active power flows through the difference between adjacent voltage angles and line reactance. The full derivation of the DC-OPF formulation can be found at Section 2.4.

$$
\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & P_{ij}^L = \frac{(\theta_i - \theta_j)}{x_{ij}} \quad \forall i, j \in N \\
& P_i^G - P_i^D = \sum_{j \in M_i} P_{ij}^L \quad \forall i, j \in N \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& \theta_i^{min} \leq \theta_i \leq \theta_i^{max} \quad \forall i \in N \\
& -P_{ij}^{max} \leq P_{ij}^L \leq P_{ij}^{max} \quad \forall i, j \in N
\end{aligned}
\tag{F.2}
$$

F.3 PTDF-OPF

In this section, the formulation for the PTDF-OPF model is presented. In this model, the PTDF approximation matrix X_{PTDF} is used to approximate the active power flows through the lines with the active power injection at the nodes. This OPF model does not consider neither reactive power nor any power loss.

$$\begin{aligned} \min \quad & \sum_{i \in G} C_i(P_i^G) \\ \text{s.t.} \quad & \boldsymbol{P^L} = X_{PTDF} \cdot \boldsymbol{P} \\ & \sum_{i \in N} P_i^G = \sum_{i \in N} P_i^D \\ & P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\ & -P_{ij}^{max} \leq P_{ij}^L \leq P_{ij}^{max} \quad \forall i,j \in N \end{aligned} \tag{F.3}$$

F.4 Upgraded PTDF-OPF

In this section, the formulation for the upgraded PTDF-OPF model is presented. In this model, the PTDF approximation matrix X_{PTDF} is used to approximate the active power flows through the lines with the active power injection at the nodes. Additionally, an estimation of the active power loss on the network is obtained via a linear approximation between the active loads and the active power loss thanks to the matrix X_{P-loss}. This OPF model does not consider reactive power.

$$\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & \boldsymbol{P^L} = X_{PTDF} \cdot \boldsymbol{P} \\
& \sum_{i \in N} P_i^G - \sum_{i \in N} P_i^D = X_{P-loss} \cdot \boldsymbol{P^D} \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& -P_{ij}^{max} \leq P_{ij}^L \leq P_{ij}^{max} \quad \forall i,j \in N
\end{aligned} \tag{F.4}$$

F.5 Linear-OPF model

In this section, the formulation for the proposed Linear-OPF model is presented. In this model, an approximation matrix X_P establishes a linear relation between the active power flows through the lines and the injected active power in the nodes. This relation can be expressed as $(P^L) = Linear(P)$. This model does not consider any power loss.

$$
\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & \boldsymbol{P^L} = X_P \cdot \boldsymbol{P} \\
& \sum_{i \in N} P_i^G = \sum_{i \in N} P_i^D \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& -P_{ij}^{max} \leq P_{ij}^L \leq P_{ij}^{max} \quad \forall i,j \in N
\end{aligned}
\tag{F.5}
$$

F.6 Upgraded Linear-OPF model

In this section, the formulation for the upgraded Linear-OPF proposed model is presented. In this model, two approximation matrices, X_P and X_Q, establish a linear relation between the power flows through the lines and the injected power in the nodes for the active and reactive powers, respectively. These relations can be expressed as $(P^L) = Linear(P)$ and $(Q^L) = Linear(Q)$. Additionally, an estimation of the active and reactive power losses on the network are obtained via a linear approximation between the loads (active and reactive) and the power losses (active and reactive) thanks to the matrices X_{P-loss} and X_{Q-loss}.

$$
\begin{aligned}
\min \quad & \sum_{i \in G} C_i(P_i^G) \\
\text{s.t.} \quad & \boldsymbol{P^L} = X_P \cdot \boldsymbol{P} \\
& \boldsymbol{Q^L} = X_Q \cdot \boldsymbol{Q} \\
& \sum_{i \in N} P_i^G - \sum_{i \in N} P_i^D = X_{P-loss} \cdot \boldsymbol{P^D} \\
& \sum_{i \in N} Q_i^G - \sum_{i \in N} Q_i^D = X_{Q-loss} \cdot \boldsymbol{Q^D} \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& Q_i^{G,min} \leq Q_i^G \leq Q_i^{G,max} \quad \forall i \in N \\
& (P_{ij}^L)^2 + (Q_{ij}^L)^2 \leq (S_{ij}^{max})^2 \quad \forall i,j \in N
\end{aligned}
\tag{F.6}
$$

F.7 PgenSlack-OPF

In this section, the AC-restoration PgenSlack-OPF procedure is presented. Given an active generation profile $P^{G,OPF}$ obtained from an approximate solution of a linear OPF model, this formulation finds the values of the slack variable ΔP that allows to recover an AC-feasible solution from an approximate solution of a linear OPF model.

$$
\begin{aligned}
\min \quad & \sum_{i \in G} (\Delta P_i^G)^2 \\
\text{s.t.} \quad & P_i^G = P_i^{G,OPF} + \Delta P_i^G \quad \forall i \in N \\
& P_{ij}^L = V_i V_j [G_{ij}\, cos(\theta_i - \theta_j) + B_{ij}\, sin(\theta_i - \theta_j)] - G_{ij} V_i^2 \quad \forall i, j \in N \\
& Q_{ij}^L = V_i V_j [G_{ij}\, sin(\theta_i - \theta_j) - B_{ij}\, cos(\theta_i - \theta_j)] + (B_{ij} - b_{ij}^s) V_i^2 \quad \forall i, j \in N \\
& P_i^G - P_i^D = \sum_{j \in M_i} P_{ij} \quad \forall i \in N \\
& Q_i^G - Q_i^D = \sum_{j \in M_i} Q_{ij} \quad \forall i \in N \\
& P_i^{G,min} \leq P_i^G \leq P_i^{G,max} \quad \forall i \in N \\
& Q_i^{G,min} \leq Q_i^G \leq Q_i^{G,max} \quad \forall i \in N \\
& V_i^{min} \leq V_i \leq V_i^{max} \quad \forall i \in N \\
& \theta_i^{min} \leq \theta_i \leq \theta_i^{max} \quad \forall i \in N \\
& (P_{ij}^L)^2 + (Q_{ij}^L)^2 \leq (S_{ij}^{max})^2 \quad \forall i, j \in N
\end{aligned}
\tag{F.7}
$$

Bibliography

[1] Mary Blockinb Blockincain. History of Optimal Power Flow and Formulations. 2012.

[2] Kavita Surana and Sarah Jordaan. The climate mitigation opportunity behind global power transmission and distribution. *Nature Climate Change*, 9:1–6, 09 2019.

[3] J. Carpentier. Contribution á l'étude du dispatching économique. *Bulletin de la Société Francaise des électriciens*, 3:431 – 447, 1962.

[4] Stephen Frank and Steffen Rebennack. An introduction to optimal power flow: Theory, formulation, and examples. *IIE Transactions*, 48(12):1172–1197, 2016.

[5] Daniel Bienstock and Abhinav Verma. Strong NP-hardness of AC power flows feasibility, 2019.

[6] Stephen M Frank, Ingrida Steponavice, and Steffen Rebennack. Optimal power flow: a bibliographic survey {I} formulations and deterministic methods. *Energy Systems*, 3(3):221 – 258, 2012.

[7] Daniel K. Molzahn and Ian A. Hiskens. *A Survey of Relaxations and Approximations of the Power Flow Equations*. 2019.

[8] Brian Stott, Jorge Jardim, and Ongun Alsac. DC Power Flow Revisited. *IEEE Transactions on Power Systems*, 24(3):1290–1300, 2009.

[9] Yitong Liu, Zhengshuo Li, and Yu Zhou. A Physics-based and Data-driven Linear Three-Phase Power Flow Model for Distribution Power Systems, 2021.

[10] Zhentong Shao, Qiaozhu Zhai, Jiang Wu, and Xiaohong Guan. Data Based Linear Power Flow Model: Investigation of a Least-Squares Based Approximation. *IEEE Transactions on Power Systems*, 36(5):4246–4258, 2021.

[11] Mengshuo Jia, Gabriela Hug, Ning Zhang, Zhaojian Wang, and Yi Wang. Tutorial on Data-driven Power Flow Linearization - Part I: Challenges and Training Algorithms. *ETH Zurich*, 04 2023.

[12] A. Gomez-Exposito, A.J. Conejo, and C. Canizares. *Electric Energy Systems: Analysis and Operation*. Electric Power Engineering Series. CRC Press, 2017.

[13] Byungkwon Park, Lisa Tang, Michael C. Ferris, and Christopher Lawson DeMarco. Examination of Three Different ACOPF Formulations With Generator Capability Curves. *IEEE Transactions on Power Systems*, 32(4):2913–2923, 2017.

[14] V. Ajjarapu and C. Christy. The continuation power flow: a tool for steady state voltage stability analysis. *IEEE Transactions on Power Systems*, 7(1):416–423, 1992.

[15] Xinxin Fang, Zhifang Yang, Juan Yu, and Yi Wang. AC Feasibility Restoration in Market Clearing: Problem Formulation and Improvement. *IEEE Transactions on Industrial Informatics*, 18(11):7597–7607, 2022.

[16] Jagadeesh Gunda, Gareth Harrison, and S.Z. Djokic. Analysis of Infeasible Cases in Optimal Power Flow Problem. *IFAC-PapersOnLine*, 49(27):23–28, 2016. IFAC Workshop on Control of Transmission and Distribution Smart Grids CTDSG 2016.

[17] Kyri Baker. Solutions of DC OPF Are Never AC Feasible. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, e-Energy '21, page 264–268, New York, NY, USA, 2021. Association for Computing Machinery.

[18] Zhifang Yang, Haiwang Zhong, Qing Xia, Anjan Bose, and Chongqing Kang. Optimal power flow based on successive linear approximation of power flow equations. *IET Generation, Transmission & Distribution*, 10(14):3654–3662, 2016.

[19] Babak Taheri and Daniel K. Molzahn. Restoring AC Power Flow Feasibility from Relaxed and Approximated Optimal Power Flow Models, 2023.

[20] Allen J. Wood and Bruce F. Wollenberg. *Power Generation, Operation, and Control, 2nd Edition*. John Wiley Sons, New Jersey, 1996.

[21] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. Matpower: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. *IEEE Transactions on Power Systems*, 26(1):12–19, 2011.

[22] T. Athay, R. Podmore, and S. Virmani. A Practical Method for the Direct Analysis of Transient Stability. *IEEE Transactions on Power Apparatus and Systems*, PAS-98(2):573–584, 1979.

[23] Kyri Baker. Learning Warm-Start Points for AC Optimal Power Flow, 2019.

[24] Trager Joswig-Jones, Kyri Baker, and Ahmed S. Zamzam. OPF-Learn: An Open-Source Framework for Creating Representative AC Optimal Power Flow Datasets. In *2022 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, 2022.

[25] T.J. Overbye. A power flow measure for unsolvable cases. *IEEE Transactions on Power Systems*, 9(3):1359–1365, 1994.

[26] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[27] Joshua Adam Taylor. *Convex Optimization of Power Systems*. Cambridge University Press, 2015.

[28] Anders Forsgren, Philip E. Gill, and Margaret H. Wright. Interior Methods for Nonlinear Optimization. *SIAM Review*, 44(4):525–597, 2002.

[29] Rafael Marti. *Multi-Start Methods*, volume 57, pages 355–368. 04 2006.

[30] C. Floudas and C. Gounaris. A Review of Recent Advances in Global Optimization. *Journal of Global Optimization*, 45:3–38, 09 2009.