UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Advanced signal processing techniques for robust state estimation applications in smart grids

Achilleas Tsitsimelis

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Advanced Signal Processing Techniques for Robust State Estimation Applications in Smart Grids

Ph.D. Dissertation

Author

Achilleas Tsitsimelis

Advisors

Dr. Carles Antón-Haro

Dr. Javier Matamoros

Department of Signal Theory and Communications

Universitat Politécnica de Catalunya

Barcelona, November 2019

This work was supported by the ADVANTAGE Project (FP7-PEOPLE-2013-ITN) under Grant 607774.

Abstract

Since their inception, more than one century ago, electrical grids have played the role of a critical infrastructure. During the majority of this time, power systems have not faced radical changes and, further, they have been mostly based on their traditional design exclusively allowing *one-way* power flows (i.e., from power plants to consumers). In contrast, over the last two decades this paradigm has rapidly changed. On the one hand, the environmental need for de-carbonization has stimulated the introduction of (i) green energy through renewable energy sources (RES) such as photovoltaic plants or wind turbines often exhibiting an intermittent power generation profile; and (ii) Distributed Energy Resources (DER) such as biomass generators, wind turbines of smaller scale, battery storage, or electric vehicles. On the other, the de-regulation of energy market has raised the necessity for substantial cooperation between the energy utilities towards the exchange of electric power over large geographical areas along tie lines. The latter entails that a totally independent operation of such areas is not viable anymore, yet a centralized approach is not applicable either due to the privacy concerns of regional utilities.

All the above implies that, to start with, power grids must be able to support *bi-directional* power flows. And, further, that variations in power generation and consumption must be timely and accurately monitored in order to keep a real-time balance between power supply and demand. To that aim, engineers and researchers can exploit recent innovations in measurement technology such as Phase Measurement Units (PMUs) providing more accurate, frequent and time-synchronized measurements; advanced signal processing algorithms and optimization tools, and a widespread use of wired and wireless communication technologies. By doing so, extended intelligence and enhanced management capabilities can be provided to power grids. This, clearly, brings the notion of Smart Grids (SG) into play which can be defined as a "*a modernized grid that enables bidirectional flows of energy and uses two-way communication and control capabilities that will lead to an array of new functionalities and applications*". The accomplishment of this modern paradigm requires the re-design of a number of classical management and

control strategies running in the operation centers of traditional grids.

Specifically, the main objective of this PhD dissertation is the re-formulation of a key functionality for the efficient monitoring, control and optimization of electrical networks: State Estimation (SE). SE can be defined as the determination of complex voltages at *all* the buses in a power system from a reduced number of system variables measured in *selected* nodes and branches. As one of the main milestones in the smart grid roadmap is to achieve a tighter interaction between transmission and distribution system operators, our research has been accordingly divided in two parts. In the first, the study is focused on the (high-voltage) Transmission Grids (TG), taking into consideration their specificities. The second, is dedicated to the (medium voltage) Distribution Grid (DG). The latter presents a number of additional challenges in terms of network topology (radial vs. meshed), availability of measurement infrastructure (low vs. high), etc; which often renders the SE problem ill-conditioned.

With respect to TGs, we propose a *hybrid* SE scheme exploiting both PMU *and* legacy measurements. This is a realistic scenario due to the increasing penetration of PMUs in TGs. The problem suffers from an inherent non-convexity and, thus, we adopt a successive convex approximation framework (SCA-SE) to iteratively solve it. Our goal is to attain increased accuracy and faster convergence rate when compared to other classical solvers from the literature. To that aim, we exploit second order information from the initial objective function. Going one step beyond and in line with the aforementioned operational needs, we consider a *multi-area* scenario too. Hence, we pose the SCA-SE problem in a *decentralized* setting, as opposed to SE hierarchical approaches that can be found in the literature. Here, our aim is to attain an accuracy comparable to that of its centralized counterpart in a reduced number of iterations. For the solution, we resort to the so-called Alternating Direction Method of Multipliers (ADMM). The data exchange required between the different network areas is very limited and, hence, it preserves privacy between transmission system operators while reducing its computational cost. Finally, we take into consideration the presence of bad data in the measurement sets. In this case, we exploit sparsity-promoting techniques and reformulate the problem in a Least Absolute Shrinkage and Selection Operator (LASSO) optimization framework. Specifically, we provide *joint* state estimation and bad data detection (i.e., outlier cleansing), which is in stark contrast with other *sequential* approaches (e.g., Largest Normalized Residual Test, LNRT) that can be found in the literature. By doing so, the robustness of the proposed SE method can be substantially enhanced in variety of scenarios.

In the second part of this dissertation, we address the problem of SE for the *distribution* grid. Our aim is to propose an algorithm capable of tracking the rapid variations over the voltage profile resulting from the introduction of renewable energy sources (which often exhibit and intermittent/non-stationary behaviour), or other components, such as distributed energy sources/storage or electrical vehicles. To do so, we leverage on the

recently introduced micro-PMUs (μPMUs) for distribution grids. Specifically, we present a regularized SE scheme operating at two different time-scales: (i) a robust state estimator that operates at the main time instants when legacy measurements are collected (e.g., every 15 minutes); and (ii) a regularized SE scheme for a number of intermediate time instants. For the former, we formulate the estimator as a regularized non-linear least squares optimization problem, which refers to the regularized version of the Normal-Equations based SE solution (R-NESE). As for the latter, we present a Weighted Total Variation State Estimation (WTVSE) scheme which exploits the fact that voltage variation (drop) in adjacent buses is necessarily limited. The latter allows to overcome the limited availability of measurements in intermediate time instants (few μPMU data and possibly zero-power injections, vs. the adequate amount of legacy, μPMU and pseudo-measurements available at the main time instants) which, otherwise, would render the SE problem underdetermined. In order to solve the WTVSE problem, we resort to the ADMM. Further, our goal is to attain high estimation accuracy in feeders of different characteristics, considering also the presence of active control devices (In-Line Voltage Regulator, ILVRs). To do so, first we design ad-hoc branch weights for each DG case and, then, we decompose the WTVSE problem (D-WTVSE) in a number of smaller areas. Besides, we study the problem of μPMU placement (μPP). Here, our main objective is to optimize the conditioning of the R-NESE scheme. The problem is posed as a mixed integer semidefinite programming (MISDP) model and, thus, it can be efficiently solved.

Resumen

Desde su origen, hace más de un siglo, las redes eléctricas se han erigido como una de las principales infraestructuras criticas de nuestra sociedad. Durante la mayor parte de este periodo, dichas redes no han experimentado cambios sustanciales en su arquitectura. De hecho, se siguen basando principalmente en el diseño original cuyo objetivo era llevar a cabo una transferencia unidireccional del flujo de energía desde las plantas generación hasta los consumidores finales. Por el contrario, este paradigma ha cambiado rápidamente en las dos últimas décadas. Por un lado, la necesidad de reducir la emisión de gases de efecto invernadero derivados del uso de combustibles fósiles ha estimulado la introducción de (1) fuentes de energía renovable (FER), tales como plantas fotovoltaicas o turbinas eólicas, a menudo con un patrón intermitente de generación de energía; y (2) recursos energéticos distribuidos (DER), tales como generadores de biomasa, turbinas eólicas de pequeña escala, sistemas de almacenamiento en baterías o vehículos eléctricos. Por otro lado, la desregulación del mercado energético ha planteado la necesidad de que los operadores cooperen de manera más estrecha en el intercambio de energía eléctrica en grandes áreas geográficas mediante las denominadas líneas de enlace. A consecuencia de ello, ya no es viable gestionar cada una de las redes eléctricas desplegadas en dicha área geográfica de manera totalmente independiente. Sin embargo, una gestión totalmente centralizada tampoco es posible habida cuenta de las restricciones de privacidad asociadas a la gestión de los diferentes servicios públicos regionales.

Por todo lo anterior, es imprescindible que las redes eléctricas acometan un proceso de modernización. En primer lugar, las redes deben ser capaces de soportar flujos bidireccionales de energía. Además, para mantener el equilibrio entre la oferta y la demanda energética es necesario también que las variaciones en la generación y el consumo de energía puedan ser monitorizadas en tiempo real y de manera precisa. A tal efecto, ingenieros e investigadores disponen de innovaciones recientes en tecnologías de medición, como los denominados sincrofasores (PMUs, del inglés Phase Measurement Units), que son capaces de proporcionar mediciones más precisas, sincronizadas y a una mayor frecuencia; de algoritmos avanzados de procesado de la señal y otras herramientas de optimización;

y de una disponibilidad generalizada de tecnologías de comunicación inalámbricas y cableadas. De esta manera, es posible aumentar el nivel de inteligencia de dichas redes eléctricas así como mejorar la gestión de las mismas. Todo ello se traduce en la llegada de las denominadas redes eléctricas inteligentes (REI), o *Smart Grids* (SG) en inglés, que pueden definirse como "una red que permite flujos de bidireccionales de energía en base a una comunicación y control bidireccionales, lo que permite el despliegue de nuevas funcionalidades y aplicaciones". La realización de dicho paradigma conlleva el rediseño de las estrategias clásicas de gestión y control que se ejecutan en los centros de operación de las redes tradicionales.

En concreto, el objetivo de esta tesis doctoral es la reformulación de una funcionalidad clave para la monitorización, control y optimización de las redes eléctricas: la estimación de estados (SE, por sus siglas en inglés). SE puede definirse como la determinación de los niveles de tensión (voltaje) complejos en todos los buses del sistema a partir de un numero reducido de medidas obtenidas en ramas y nodos seleccionados. En la medida que uno de los hitos principales en la hoja de ruta de la red inteligente es lograr una interacción más estrecha entre los operadores de sistemas de transmisión y distribución, nuestra investigación se ha dividido en dos partes. En la primera, el estudio se centra en las Redes de Transmisión (TG, del inglés *Transmission Grids*) de alta tensión en la que se han tenido en cuenta sus especifidades. La segunda parte está dedicada a la Redes de Distribución (DG, del inglés *Distribution Grids*) de media tensión. Estas ultimas presentan una serie de retos adicionales en términos de tipología de la red (radial, en vez de mallada), y del bajo grado de disponibilidad de infraestructura de medición (que es elevado en el caso de las TGs), etc. Todo ello hace que el problema de SE, a menudo, esté mal condicionado en términos algebraicos.

Con respecto a las TG, en esta tesis proponemos un esquema de SE hibrido que aprovecha tanto las mediciones obtenidas en los modernos PMUs como en los dispositivos de medida tradicionales. Este es un escenario realista debido a la creciente penetración de las PMU en las TG. El problema no es convexo por lo que debemos recurrir a métodos de aproximaciones convexas sucesivas (SCA-SE) para poder resolver el problema de la SE de forma iterativa. Nuestro objetivo es lograr una mayor precisión y una velocidad de convergencia más rápida en comparación con otras soluciones clásicas. Con este objetivo, nos apoyamos en la información de segundo orden disponible en la función objetivo inicial. Yendo un paso más allá y en línea con las necesidades operativas anteriormente mencionadas, también consideramos un escenario de redes eléctricas formadas por multiples áreas. Por lo tanto, planteamos el problema SCA-SE en un entorno descentralizado, a diferencia de otros enfoques jerárquicos de SE disponibles en el estado del arte. Nuestro objetivo es lograr una precisión comparable a la de la versión centralizada en un numero reducido de iteraciones. Para solucionar esto, recurrimos al llamado Método de Dirección Alterna de Multiplicadores (ADMM, por sus siglas en inglés). El intercambio de datos

requeridos entre las diferentes áreas de la red es muy limitado, y, por lo tanto, garantiza la privacidad entre los operadores del sistema de transmisión al tiempo que reduce su coste computacional. Finalmente, tomamos en cuenta la posible presencia de datos incorrectos en los conjuntos de medidas. En este caso, recurrimos a técnicas de *sparsity promotion* y reformulamos el problema en un marco optimización del método LASSO (Shrinkage and Selection Operator, por sus siglas en inglés). De este modo, somos capaces de obtener de manera *simultanea* una estimación del estado y una detección de los datos incorrectos (es decir, llevar a cabo una limpieza de datos atápicos). Esto representa una clara novedad respecto a otros enfoques *secuenciales* como, por ejemplo, el denominado Largest Normalized Residual test (LNRT), entre otros. De este modo, es posible mejorar sustancialmente la robustez del método de SE propuesto en diversos escenarios.

En la segunda parte de esta tesis, abordamos el problema de SE para redes de distribución. Nuestro objetivo es proponer un algoritmo capaz de monitorizar las variaciones rápidas de tensión que resultan de la introducción de fuentes de energía renovables (que, a menudo, presentan un patrón de comportamiento intermitente/no estacionario), y de otros componentes como fuentes y sistemas de almacenamiento distribuidos de energía o vehículos eléctricos. Para ello, recurrimos al uso de micro-PMUs (μPMUs) de reciente introducción en redes de distribución. Así, presentamos un doble esquema de SE regularizado que opera en dos escalas de tiempo diferentes. Por una parte, un estimador robusto de estados para los instantes de tiempo principales, es decir, cuando se recogen las medidas en los dispositivos tradicionales (por ejemplo, cada quince minutos). Y, por otra, un esquema SE regularizado para los instantes de tiempo intermedios. Para el primero, formulamos el estimador como un problema de optimización no lineal de mínimos cuadrados, que constituye la versión regularizada de solución SE basada en ecuaciones normales (R-NESE). Para el segundo, presentamos un esquema *Weighted Total Variation State Estimation* (WTVSE) que explota el hecho de que la variación (caída) de tensión en buses adyacentes es necesariamente limitada. Esto ultimo permite superar la disponibilidad limitada de mediciones en instantes de tiempo intermedios (pocos datos μPMU y posiblemente inyecciones de potencia cero, frente a una cantidad adecuada de medidas de dispositivos tradicionales, μPMUs, así como las pseudomedidas disponibles en los instantes de tiempo principales) que, de lo contrario, generaría un problema SE indeterminado. Para resolver el problema de WTVSE, recurrimos de nuevo al método ADMM. Además, nuestro objetivo es lograr una alta precisión de estimación en alimentadores de diferentes características, considerando también la presencia de dispositivos de control activos (los denominados reguladores de voltaje en línea, ILVR). Para ello, proponemos un diseño *ad-hoc* para los pesos de las diferentes ramas de una DG y, a continuación, descomponemos el problema WTVSE (D-WTVSE) en áreas más pequeñas. Por ultimo, estudiamos el problema de la ubicación optimizada de PMUs (μPP, μPMU placement, por sus siglas en inglés). Aquí, nuestro objetivo principal es optimizar el condicionamiento del sistema

de ecuaciones del esquema R-NESE anteriormente mencionado. El problema se plantea en base a un modelo MISDP (Mixed Integer Semidefinite Programming) y, por lo tanto, puede resolverse de manera eficiente.

Acknowledgements

The last few years had been an intensive, educative and productive journey. The end of this journey is established by this dissertation. This last written part can be regarded as an effort to gather all memories and moments and thank all the people who, with one or another way, contributed to this accomplishment and my improvement as a person.

First of all, I would like to express my sincere gratitude to my advisors Dr. Carles Antón-Haro and Dr. Javier Matamoros. I would like to thank them for the opportunity to work in a multidisciplinary research project which helped me to broad my knowledge and to enlighten my horizons in engineering. Their guidance, the freedom I was given in our work and their constructive criticism were essential for the development of this dissertation and my progress as a researcher.

I would like to acknowledge ADVANTAGE project for the financial support of my doctoral studies and my research visits. Further, I would like to thank the ADVANTAGE team for our constructive collaboration. It was really a pleasure to meet all these interesting people, sharing different scientific background, ambitions, but mostly cultures. I would like to especially thank Prof. Dejan Vukobratovic and Dr. Mirsad Cosovic for our collaboration and their hospitality during my research stay in the University of Novi Sad. A special acknowledgment here goes to Dr. Charalambos Kalalas for our collaboration and friendship.

Further, I would like to take the opportunity to thank Prof. Gonzalo Seco-Granados and Dr. Juan Luis Navarro-Mesa for conducting the preliminary assessment and quality review of the thesis and for agreeing to serve in this thesis' examining committee. For the same reason my gratitude also extends to Dr. Miquel Payaró, Prof. Dejan Vukobratovic and Prof. Miguel Ángel Lagunas.

Typically, this part ends with a general acknowledgment to all people who contributed to the accomplishment of this work. However, except if someone goes for a second PhD (which is not the case), there is only one opportunity to mention a number of persons who played an important role during the last years. To start with, I would like to thank my colleagues in CTTC, Dr. Juan Manuel Arvizu, Dr. Arturo Mayoral, Dr. Miguel

Calvo Fullana and Dr. Onur Tan for their friendship, the long-term opportunity to practice my Spanish and the vermuth enlightenment. Further, I would like to express my deepest gratitude to Alex, Alkioni, Apostolis, Carlos, Christos, Iraklis, Nikos, Panos and Spyros for being my "family" in Barcelona all these years. I am also indebted to my "old-school" colleagues Iasonas K. L., Nasos V., Giorgis M., Dimitris T. and Alex R. for all the stimulating technical discussions. Not to forget to mention my "old" good friends Athanasios, Michail, Ioannis, Ioanna and of course, Stelios and Eri, who made my life enjoyable with their visits in Barcelona, their support and... the summers in Ikaria. Finally, I would like to refer to my loyal friend "Kaltsa", who made easier the decision to start this beautiful journey.

Last but not least, I would like to express my deepest gratitude to my family for all the support and solidarity, especially during the tough moments.

This dissertation is dedicated to the free public education.

Achilleas Tsitsimelis,
Barcelona,
November 2019

List of Acronyms

AC	Alternating Current
ADMM	Alternating Direction Method of Multipliers
AMI	Advanced Metering Infrastructure
BDD	Bad Data Detection
BC	Branch Current
BSDP	Binary Semidefinite Programming
DC	Direct Current
DER	Distributed Energy Resources
DG	Distribution Grid
DMS	Distribution Management System
DSO	Distribution System Operator
D-WTVSE	Distributed Weighted Total Variation Estimation
EMS	Energy Management System
GPS	Global Positioning System
i.i.d.	Independent and Identically Distributed
ILVR	In-Line Voltage Regulator
ISO	Independent System Operator
LASSO	Least Absolute Value Shrinkage Operator
LMI	Linear Matrix Inequalities
LMI	Linear Matrix Inequalities
LNRT	Largest Normalized Residual Test
MISDP	Mixed Integer Semidefinite Programming
MLE	Maximum Likelihood Estimation

μPMU	μicro Phasor Measurement Unit
μPP	μicro Phasor Measurement Unit Placement
MV	Medium Voltage
OPP	Optimal Phasor Measurement Unit Placement
pdf	Probability Density Function
PMU	Phasor Measurement Unit
PV	Photovoltaic
RES	Renewable Energy Sources
RMSE	Root Mean Squared Error
R-NESE	Regularized Normal Equations State Estimation
RSCA-SE	Robust Successive Convex Approximation State Estimation
RTUs	Remote Terminal Units
SCA	Successive Convex Approximation
SCADA	Supervisory Control And Data Acquisition
SCA-SE	Successive Convex Approximation State Estimation
SE	State Estimation
SDP	Semidefinite Programming
SDR	Semidefinite Relaxation
TG	Transmission Grid
TSO	Transmission System Operator
WLS	Weighted Least Squares
WTVSE	Weighted Total Variation Estimation

Notation

x	A scalar.
$\mathbf{x}$	A column vector.
$\mathbf{X}$	A matrix.
$\mathscr{X}$	A set.
$\mathbb{R}$	The set of real numbers.
$\mathbb{R}^n$	The set of real n vectors.
$\mathbb{R}^{m \times n}$	The set of real $m \times n$ matrices.
$\mathbb{S}^k$	The set of positive semidefinite $k \times k$ matrices.
relint $\mathscr{X}$	Relative interior of set $\mathscr{X}$.
dom f	Domain of function f.
$\|\cdot\|$	Norm.
$(\cdot)^T$	Transpose operator.
$\mathbf{X}^{-1}$	Inverse of matrix $\mathbf{X}$.
$\mathbf{I}$	Identity matrix.
max	Maximum.
min	Minimum.
inf	Infimum.
log	Logarithm.
arg	Argument.
$\triangleq$	Defined as.
$\nabla_{\mathbf{x}} f(\mathbf{x})$	Gradient of function f with respect to $\mathbf{x}$.
$\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Real Gaussian distributed vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

The established electrical grids, which have been developed over the past 80 years, are intricate interconnected networks designed to deliver electricity from its generation to the consumers. The generation comprises large central generating stations, commonly fossil fueled, nuclear and hydroelectric, producing electrical power rating of up to 1000 MW, which is fed through generator transformers to the so-called Transmission Grid (TG). The high-voltage transmission lines carry the electric power from the distant sources to the distribution substations, that in turn, through the distribution transformers and lines, namely the Distribution Grid (DG), connects the final circuits to the individual customers. That is, electrical power flows in one direction: from power generation plants to consumers, as Figure 1.1 illustrates. The TG, as well as the central power plants, have

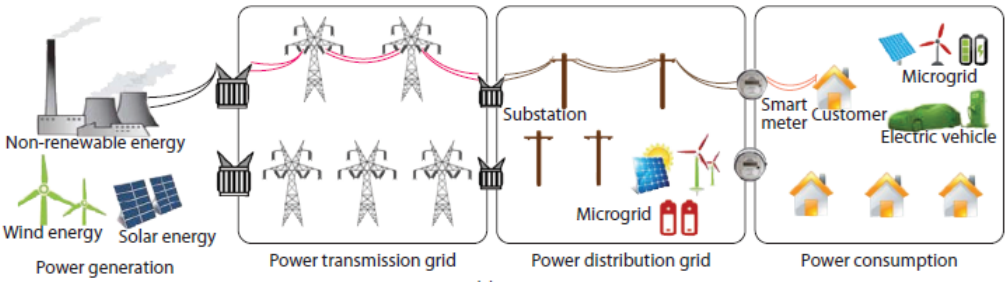

Figure 1.1: Segments in a electrical grid.

advanced communication infrastructure to guarantee the security and reliable operation of the system and to enable market transactions. For the same reasons, this part of the grid is enhanced with extended automatic control systems. In terms of topology, TGs tend to be more complex, with redundant pathways forming a meshed network. This network operates at higher voltages (in the order of 150, 220 and 400 kV) mainly due to

their ability to keep losses over copper lines lower. On the other hand, the DG presents a number of differences compared to the TG. Primarily, despite the fact that distribution lines are very extensive, DG is almost passive in terms of communication and control systems. In other words, the only interactive action is the supply of electric power to the loads. Further, distribution systems typically use a radial configuration for their operation, with feeders taking electricity from substations to households exhibiting a tree structure. These lines typically operate at lower voltages (e.g., 11 or 33 kV). This means that step-down transformers are needed in between the transmission and distribution substations.

Typically, the power system operates in three different states: normal, emergency and restorative. The system operates in the normal state when there is no disruption in the power supply and no operational limits, such as upper and lower bounds on bus voltage magnitudes, are violated. When this happens due to unexpected operational circumstances but the power supply is not yet disrupted, the system enters the emergency state. In this case, a number of control actions have to be taken (e.g., topology reconfiguration, load deferral, etc) in order to restore normal operation. During such rebalancing effort between power supply and demand, the system operates in the so-called restorative state. All the above lies into the situational awareness, which in turn, lies into the grid monitoring. The key functionality for the electrical network monitoring, control and optimization is State Estimation (SE). SE is defined as the determination of complex voltages at *all* the buses (lines) in a power system from a reduced number of system variables measured in *selected* nodes and branches. With such voltage information and by means of the power flow equations one can compute every magnitude of interest: currents, losses on the lines, power injections, etc.

1.1.1 Limitations of the Traditional Electrical Grid

The traditional power system had been designed decades ago. Thus, in its current form it is not possible to support more recent concepts and paradigms, such as low-carbon technologies, energy efficiency and the connection with modern energy markets. More specifically, today's electrical grid raises a number of technical limitations related with a number of issues. Primarily, the system provides limited support for the Distributed Energy Resources (DER). These small scale generators create two-way power flows in contrast with the traditional one-way approach (from the power plants to the consumers). The same applies for the introduction of the Renewable Energy Sources (RES). RES often exhibit an *intermittent* power generation profile (due to day/night cycles, seasons, wind/no wind conditions, etc.). Such an intermittent energy injection makes it more difficult to keep the balance between power supply and demand in real-time which is mandatory since large-scale electricity storage is not economically viable yet. Another important issue is the system inefficiency to manage peak loads. As the electricity demand

varies, supply must be adjusted continuously to track demand. To do so, utilities always maintain excess power supply, which is mainly produced from fossil fuel plants and so, results to higher costs and emissions. Last but not least, the current power system does not count with a robust two-way communication infrastructure. This is related with a number of crucial system functionalities. For instance, the control actions, which are required to prevent system outages, need a reliable and rapid information flow from the network to the energy control center. Further, the system monitoring shall be based on a robust communication system able to support different measuring technologies and to deal with a large volume of data.

All the above shortcomings may be addressed by key emerging technologies able to provide extended intelligence and enhanced management capabilities that can improve operational efficiency and performance. Primarily, existing and enhanced communication technologies, including wide area wireless internet and cellular networks, are able to provide high-speed, standardized two-way information flows for robust system monitoring and decision-making. Totally related, the last years a number of measurement advancements are capable of producing in real-time a high volume of information. An example is the Phasor Measurement Units (PMUs), extensively installed to the transmission system. On the other hand, with the massive deployment of Advanced Metering Infrastructure (AMI) households can monitor their energy demand. Finally, advanced power system components will enhance the operation of the grid, at both transmission (e.g., power electronics) and distribution (e.g., digital voltage regulators) level.

1.1.2 The Advent of Smart Grid

All the above brings *Smart Grid* into play. A smart grid can be defined as a "a modernized grid that enables bidirectional flows of energy and uses two-way communication and control capabilities that will lead to an array of new functionalities and applications" (NIST, National Institute of Standards and Technology). In particular, this concept refers to the introduction of digital and communication technologies to the electric power sector in order to improve reliability, reduce cost, increase efficiency and enable new components and modern applications. In contrast to the existing electrical grid, the smart grid promises improvements in power quality, two-way information flow, and more importantly, an improved support for DER and RES. In addition, based on the advanced sensors and monitoring devices, smart grid technologies will form a new framework where consumers will be able to participate actively into play.

However, in order to cope with the evolving operational environment of the smart grid, a re-design of the classical management strategies and control functionalities, which have been used for decades, is needed. This holds in particular for SE algorithms which are typically run based on the SCADA (Supervision, Control and Data Acquisition) system. SE is the main topic addressed in this PhD thesis. Specifically, we aim to design

novel SE schemes on the basis of advanced signal processing techniques, such as convex optimization or sparse and statistical signal processing tools. SE algorithms rely on the measurements collected by legacy and modernized metering infrastructure (e.g., phasor measurement technology), on statistical models and, also, an in-depth knowledge of power system operation. Hence, this PhD thesis lies in the crossroad of a number of scientific fields: electric power engineering, signal processing and communications.

1.2 Challenges

The target of this PhD thesis is to design novel SE algorithms for power systems, leveraging on advanced signal processing techniques and the recent advancements on the sensing/measurement technology. This is needed in order to cope with the new operational environment that the advent of smart grid brings, which includes (i) a bidirectional flow of energy and data along transmission and distribution systems; and (ii) a larger degree of variability and intermittency typically associated to renewable and distributed energy resources.

The specific challenges that we plan to address in this PhD thesis include the following:

– **To model state estimation as a convex optimization problem**. As many optimization problems in the literature, SE for power systems suffers from inherent non-convexity. In particular, the SE problem formulation is based on the non-linear power flow equations. This unavoidably leads to a non-convex optimization problem that in turn, leads the traditional solvers, such as the Gauss-Newton method, to converge to local minima. Instead, we can exploit the rich literature on convex optimization, and resorting to convex approximation techniques, we plan to formulate SE as a convex optimization problem. In that way, we aim to provide improved SE accuracy and convergence rate comparing with benchmarks from the literature.

– **To design hybrid state estimation schemes exploiting both legacy and PMU measurements**. The challenge here is three-fold: (i) the fact that such measurements are of a different nature. Specifically, SCADA system and PMUs relate the state variables with the power flow equations in non-linear and linear fashion, respectively; (ii) their different time-scales, which is much shorter in the case of PMU data (i.e., millisecond vs. second timescales); and (iii) the non-uniform availability of PMU devices. In particular, in the case of TG, where has been observed a rapidly-growing deployment of phasor meters, still their presence can not guarantee system observability. This, inevitably leads us to hybrid SCADA/PMU-based SE schemes. Even worse, PMUs are seldom found in the case of distribution systems.

– **To derive decentralized versions of state estimation algorithms suitable**

for multiarea scenarios. The challenge here is to derive decentralized SE algorithms which: (i) preserve the maximum privacy between different system operators; and (ii) provide reduced computational complexity. To that aim, we will resort to optimization algorithms able to perform efficiently in a distributed environment, such as the dual ascent decomposition and the Alternating Direction Method of Multipliers (ADMM). Further, we will analyze the performance of the resulting algorithms using as benchmarks their centralized counterparts, investigating their convergence behavior and the resulting estimation accuracy.

– **To design advanced schemes for bad data analysis and detection**. The most common bad data detection method for SE, namely the Largest Normalized Residual Test (LNRT), constitutes an a posteriori process implemented after the convergence of the SE algorithm and consequently, leads to performance degradation. Additionally, in specific occasions (e.g., multiple and conforming bad data) this method is subject to fail. To circumvent that, we plan to investigate methods to conduct bad data detection and elimination jointly with the SE procedure. More specifically, we plan to leverage on the sparse presence of such outliers and resort to respective techniques from the signal processing literature.

– **To investigate state estimation schemes particularly suitable for distribution networks**. In contrast with the TG, distribution networks present a number of peculiarities in terms of their topological uniqueness and the existing restricted measurement infrastructure. These characteristics are usually translated into ill-posed SE models. To overcome this issue, we plan to resort to regularization techniques, such as the Tikhonof regularization and the total variation regularization algorithm. Further, we will exploit the recent measuring technology advancements in order to built reliable and robust SE schemes with increased temporal granularity. The main target is to customize these strategies according to the evolving operational environment of the medium voltage networks.

– **To investigate optimal μPMU placement methods particularly suitable for distribution networks**. As mentioned, the measurement infrastructure in the case of the DG is limited. An answer to the pressing need for the medium voltage sytem monitoring can be the recent measuring technology advancements, such as the low-cost μicro PMUs. In particular, we plan to investigate optimal placement methods with respect to the DG specificities, taking into account the already limited existing measurements. More precisely, beyond the fundamental criterion of system observability, we plan to pose further objectives related with the DG SE problem robustness. The final target will be to improve the SE algorithmic accuracy and convergence rate. To do so, we will resort to linear and matrix algebra analysis, and techniques from the convex optimization literature.

1.3 Outline of the Thesis

As discussed above, in this PhD dissertation we propose novel SE algorithms with respect to the new operational environment of the electrical grid. More specifically, our concentration has been dedicated separately to the different voltage levels, taking into consideration in each case the different specificities and issues under discussion. In general, Chapters 4 and 5 refer to the SE schemes built for the transmission and distribution system, respectively. In details, this thesis is organized as follows:

In Chapter 2, we provide insight on the state of the art of SE algorithms. Specifically, we start with SE for the transmission system and its distributed formulations according to the current and future needs. These models are mainly divided into hierarchical and decentralized. Later on, we focus on the DG SE and all the recent advancements. In addition to that, we give a brief overview of the phasor measurement technology as well as the phasor meter placement methods, both constituting important aspects related with the SE algorithmic behavior.

In Chapter 3, we review a number of important mathematical concepts over which the majority of this study has been built. In details, we first provide the basic theory behind convex optimization. Then, we focus on convex approximation schemes which deal with the non-convexity of many optimization problems. Later on, we give a brief description on a special class of problems, namely, the binary semidefinite programming models. Finally, we mention some interesting applications of convex optimization with emphasis on estimation and regularization techniques.

We begin the contributions of this dissertation in Chapter 4. The scenario under consideration assumes a hybrid communication infrastructure at the transmission system level, where both SCADA and PMU measurements are available. As already mentioned, the SE optimization problem suffers from inherent non-convexity. Thus, we resort to an iterative framework that follows the spirit of successive convex approximation algorithms applied efficiently in other fields, such as communications and signal processing. Among others features, this framework provides solution feasibility at each iteration and leads easily to distributed implementation. Our goal is to conclude to an improved SE algorithm, which outperforms in terms of accuracy compared to the traditional solvers. Further, we take into consideration a multiarea scenario, commonly adopted for the SE functionality between different system operators. Here, we aim to conclude in a decentralized version of the model with negligible penalty in terms of accuracy. For this purpose, we leverage on the ADMM. In both centralized and distributed cases, we exploit second-order information of the original non-convex objective function in order to improve the convergence rate of the algorithm. Finally, we examine a scenario where a number of meters transmit corrupted data (i.e., outliers). In this case, we present the robust version of the algorithm, capable of performing jointly SE and bad data detection. To do so,

we reformulate the SE problem in a Least Absolute Shrinkage and Selection Operator (LASSO) optimization framework. The underlying idea is to promote sparsity in the vector of corrupted (PMU and legacy) measurements. The latter is accompanied from an algorithmic analysis with respect to the outliers cleansing.

Next, in Chapter 5 we deal with the SE problem in the case of DGs. This scenario presents a number of additional challenges, such as the restricted available information, the peculiar radial topologies and the increasing penetration of DER. More precisely, we plan to build a novel SE scheme which operates in a two-time scale fashion. The ultimate goal is to track accurately the system state at a faster time scale with increased reliability, according to the needs of the new operational environment. Thus, the SE algorithm will be capable of tracking short-term off-limit system conditions. At first, we plan to enhance robustness to the classical SE schemes that operate every 15 minutes. The observations comprise SCADA measurements and pseudomeasurements. On top of these, a limited number of precise synchrophasors is added. This low redundant measurement set, characterized by a noise variation diversity, results to an ill-conditioned optimization problem and affects significantly the SE accuracy. To overcome this barrier, we pose the SE as a regularized non-linear least squares optimization problem. To do so, we resort to the well-known Tikhonof regularization exploiting prior system information, that is, the last computed state estimate. Later on, we construct a state estimation scheme that operates on the intermediate time instants, i.e., between the main computed estimates. In this case, we leverage on the positioned μPMUs at a subset of buses and on the possible zero power injection buses. As unavoidably the number of measurements is low, observability is not guaranteed. To circumvent that, we exploit prior information on selected DG features. Specifically, we formulate the SE as a constrained weighted *total variation* estimation problem which limits variations in voltage estimates in adjacent buses. This stems from the fact that, in DGs, electrical lengths and current flows tend to be lower than in TGs and so are voltage drops in adjacent buses. On that basis, we also propose a rule to define branch-specific weights for the regularizer. Finally, we take into consideration the presence of In-Line Voltage Regulators (ILVRs) across the feeders which forces us to re-formulate the problem as a constrained *decomposed* model. Similarly to Chapter 4, we effectively solve the problem by finding an iterative solution based on the ADMM. Finally, we present a μPMU placement method in order to enhance robustness to the SE scheme. The scenario assumes the existence of a number of legacy measurements. To do so, we follow a similar rationale with already developed algorithms for sensor scheduling purposes and we pose a mixed integer semidefinite (MISDP) optimization problem. In particular, the objectives of the problem are to ensure system observability and to improve the accuracy of the SE scheme.

Finally, in Chapter 6 we conclude this thesis by providing a summary and final remarks on the main results of this work. Also, we outline and discuss some possible lines of future

work.

1.4 Contribution

The research presented in this dissertation has also been disseminated through the following publications [10–13].

C1 J. Matamoros, A. Tsitsimelis, M. Gregori and C. Antón-Haro, "Multiarea state estimation with legacy and synchronized measurements," 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016, pp. 1-6.

J1 M. Cosovic, A. Tsitsimelis, D. Vukobratovic, J. Matamoros and C. Anton-Haro, "5G Mobile Cellular Networks: Enabling Distributed State Estimation for Smart Grids," in IEEE Communications Magazine, vol. 55, no. 10, pp. 62-69, Oct. 2017.

C2 A. Tsitsimelis, J. Matamoros and C. Antón-Haro, "An ADMM-based regularized state estimation scheme for the distribution grid," 2017 IEEE International Conference on Communications Workshops (ICC Workshops), Paris, 2017, pp. 900-905.

J2 A. Tsitsimelis and C. Antón-Haro, "A Regularized State Estimation Scheme for a Robust Monitoring of the Distribution Grid," International Journal of Electrical Power & Energy Systems, October, 2019.

Other related works not included in this dissertation [14]:

C3 A. Tsitsimelis, C. Kalalas, J. Alonso-Zarate and C. Antón-Haro, "On the impact of LTE RACH reliability on state estimation in wide-area monitoring systems," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, 2018, pp. 1-6.

Chapter 2

State of the Art in Power System State Estimation

In this chapter, we provide an overview of the current state of the art technologies in power system state estimation. First, we focus on transmission system estimation algorithms and more specifically, in distributed solutions that meet the needs of the new operational environment. Then, we focus on distribution system monitoring which is characterized by communication infrastructure deficiency and consequently, becomes more challenging. Finally, we present the current status on phasor measurement technology and the evolution of optimal phasor meter placement methods.

2.1 State Estimation for the Transmission System

Transmission systems have always been an active infrastructure equipped with advanced communication systems and actuators, such as, fault detectors, remotely operated reclosers and breakers. The available metering infrastructure, hence, provides redundant information that can be efficiently exploited by the SE scheme. Specifically, *legacy* metering devices, mainly Remote Terminal Units (RTUs), provide measurements (voltage magnitude, power injections, etc.) that are non-linearly related with the system variables (complex voltages). These devices and the supporting communication software, responsible for collecting and transmitting the measurements to the control center, constitute the Supervisory Control and Data Acquisition (SCADA) system. The latter, is traditionally considered as the cornerstone for the high voltage system monitoring. On the contrary, the so-called Phasor Measurement Units (PMUs) [15], offer faster and linear measurements (voltages and currents phasors) which, in addition, are tightly synchronized with one another. Due to cost considerations, however, only a limited number of PMUs can be deployed. Hence, such (linear) measurements are traditionally complemented by legacy (non-linear) ones for SE tasks.

SE techniques for the TG have been intensively investigated for decades. Most of the efforts focused on the numerical stability of the models and their computational efficiency. The fact that the measurements collected by legacy metering devices are non-linearly related with the complex voltages (state variables) often leads to nonconvex SE problems, typically formulated as a Weighted Least Squares (WLS) one. Hence, popular gradient-based iterative procedures, such as the Gauss-Newton method, are unavoidably subject to convergence to local optima. The classical WLS estimator was first introduced by Schweppe in [16]. Over the following decades, many improvements in terms of computational efficiency and increased numerical stability were proposed [17]. One of the recent developments in this context is the bilinear SE model introduced in [18]. By introducing auxiliary variables in the state vector, the authors transform the non-linear WLS SE problem in a sequence of two linear problems with a non-linear state variable transformation in between. As a result, the algorithm attains the same accuracy as the classic Gauss-Newton iterative method with the advantage of a decreased computational cost.

In recent years, however, two main trends can be observed in the evolution of power grids: (i) the increasing penetration of RES; and (ii) the de-regulation of energy markets. The former leads to larger dynamics (due to the intermittency of such RES) that need to be timely and accurately monitored. To that aim, the deployment of synchronized PMUs is instrumental. Still, as discussed earlier, such measurements must be complemented by legacy ones due to cost considerations. This entails the development of novel *hybrid* schemes as in [19]. In this context, the authors in [20] study the behavior of the classic WLS estimator incorporating PMUs and they examine the improvements in terms of convergence speed and accuracy. Two-stage hybrid approaches are also very popular in the literature. Specifically, in [21], the first level of the estimator exploits measurements from PMUs and solves a linear model. Next, the estimated states from this level are combined with legacy measurements in order to provide a classical non-linear estimator. On the contrary, the authors in [22] obtain in the first stage a classic WLS estimator based on measurements collected by the SCADA system. In the second stage, by using the estimated complex voltages and currents as pseudomeasurements, they also incorporate PMU data to solve a linear estimation problem. Results indicate that, by incorporating the estimated currents in the second stage as additional measurements the resulting estimation accuracy is higher.

The deregulation of energy markets makes it feasible to conduct energy trading among Transmission System Operators (TSOs). The pressing need, in recent times, to keep operational costs low has further stimulated inter-regional collaborations. In turn, this results into an increased exchange of large amounts of power along the tie-lines between adjacent areas possibly under the control of different utilities. A totally independent operation of such areas is, therefore, no longer viable. However, (classical) centralized

SE schemes are not directly applicable either. This is mainly, either due to data security and privacy concerns of regional utilities, or due to high computational complexity considerations. This, undoubtedly, substantiates the need for developing hierarchical [23, 24] and decentralized [25] SE methods, which are described next.

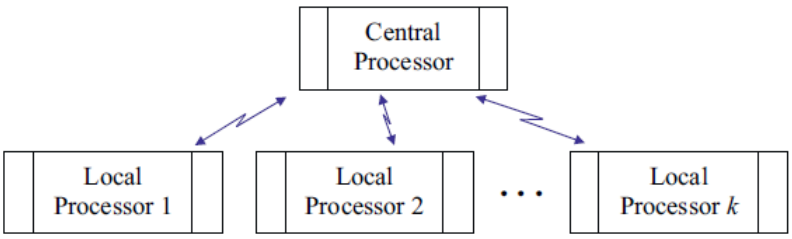

Figure 2.1: Hierarchical scheme for SE [1].

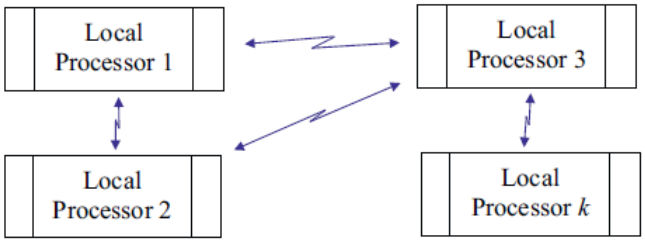

Figure 2.2: Decentralized scheme for SE [1].

2.1.1 Hierarchical State Estimation

In hierarchical schemes, each geographical area runs its own local state estimator. Each local state estimator communicates the results to the inter-area coordinator which ensures consensus between the overlapping state variables in adjacent local areas (Fig. 2.1). In decentralized schemes, on the contrary, each local operator just communicates with its neighboring network operators, without the presence of a coordinator (Fig. 2.2). The aforementioned schemes also exhibit substantial differences in terms of the amount of information that needs to be exchanged. In hierarchical approaches, each local area must send a subset of measurements or even the whole state to the coordinator. In contrast, in decentralized approaches, exchanging information from the border variables only, typically suffices.

In a hierarchical SE context, a multi-level SE paradigm is presented in [23]. Here, local state estimates are computed first at lower voltage levels. Subsequently, those estimates are transferred to upper voltage levels up to the Independent System Operator (ISO) in order to estimate the whole system state. In the same context, in [24] Korres proposes

a multiarea SE scheme. In this case, each area computes its own local state estimate and conveys information on border variables to a coordinator which, in turn, computes the system-wide state. This scheme also incorporates bad data analysis and detection, in order to identify and remove all the measurements which are in gross error. On a different key, in [26], Guo *et al* propose a hierarchical SE where the local operators communicate to the coordinator the sensitivity functions[1] of the local objective (SE) functions. Computer simulation results reveal that the local estimates converge faster than in the classic scenario (where measurements or state variables are exchanged). Another interesting work can be found in [27], where the authors pose a dynamic Kalman filter based estimator within the hierarchical framework. More specifically, a number of parallel low-processors deal with the measurement subsets producing the respective state vector for each subsystem. Then, this information is collated to the central processor that computes the global estimate. However, the possible case of measurement existence on the tie-lines between the adjacent subsystems raises the need for bi-directional communication between the central and low processors. This, in turn, increases the communication burden and computational time.

2.1.2 Decentralized State Estimation

As discussed earlier, in decentralized SE schemes, the operators running each geographical area communicate directly, with no intervention of an inter-area coordinator. Representative works of this paradigm are [28] and [25]. In [28], Xie *et al* propose a fully distributed state estimation scheme. In this work, the local control areas compute their own estimates and, by exchanging their estimates with the corresponding neighbouring areas, iteratively improve their own estimates. The proposed scheme attains the same performance as its centralized counterpart for both the DC linear and the AC models. Complementarily, in [25], Kekatos and Giannakis propose a decentralized robust state estimator, leveraging on the ADMM. Going one step beyond, in [29] the authors develop a hybrid SE scheme based on the so-called Semidefinite Relaxation (SDR) optimization technique [30], which in the sequel is referred to as SDR-SE. In addition, a distributed (approximate) version of the algorithm, suitable for multiarea scenarios, is introduced too. Interestingly, the authors prove that the SDR-SE technique is optimal for the case of noiseless measurements and voltage magnitude measurements at all buses. Unfortunately, when these conditions are not satisfied, the performance of the SDR-SE approach may degrade considerably. Further, in [31] the authors present a decentralized version of the three-stage bilinear state estimation model proposed in [18]. Each step is decoupled over the different areas, yielding a fully distributed robust bilinear state estimation using ADMM. For the two linear stages, each area solves its local SE problem, sends

[1] The term 'sensitivity function' refers here to the Taylor's expansion of the first order derivative (with respect to the boundary state variables) of the local objective function.

the latest boundary states to its neighboring areas and iterates in this way until convergence. The intermediate nonlinear transformation can be processed within each area in parallel without any need of inter-regional communication. Besides, the authors exploit the ℓ_1- regularization in order to deal with gross error measurements, enhancing performance to their algorithm. Additionally, in [32] a decentralized solution is proposed based on the Gauss-Newton method. Specifically, the authors, exploiting the inherent sparse structure of the system, leverage on matrix-splitting methods in order to provide a fully distributed SE solution. However, the bad data implementation is not straightforward here. The same applies for [33], where a decentralized SE algorithm based on probabilistic methods is proposed. In particular, the authors, first represent the system as factor graph and then they provide a SE solution based on the belief propagation algorithm. This approach is also exploited in other works from the literature [34], offering decreased computational burden compared to the classical Gauss-Newton solvers. Moreover, as in the case of hierarchical state estimation, quasi steady-state estimators based on the Kalman filter rationale can also be posed in a decentralized fashion. These approaches leverage on series of measurements observed over time, leading into an improved accuracy compared to the case where a measurement set of a single time instant is exploited. For instance, in [35] an algorithm is presented based on the extended Kalman filter, where similarly with the previous works, the local estimators only require data within their own subsystem, and information is exchanged only between adjacent subsystems. Finally, an interesting scheme can be found in [36], where the algorithm leverages on Lagrange relaxation for its decentralized operation. More importantly, the authors take into consideration the case of different types of neighboring networks. Specifically, the assumed scenario concerns the existence of AC and high voltage DC subsystems, and the estimation problem is separated accordingly.

2.2 State Estimation for the Distribution Grid

In contrast with the TG, the availability of metering and communication infrastructure in DGs is rather limited [37], [2]. The main reason is that, traditionally, the DG was conceived as a passive infrastructure in charge of just carrying electricity to industrial and residential users. Over the last decade, the increasing penetration of DERs, the deployment of dynamic control elements and, to some limited extent, distributed energy storage systems, have transformed the DG into an active system [23]. As a consequence, the interest of Distribution System Operators (DSOs) in monitoring the DG has increased significantly. The DSO maintains the system operation based on the Distribution Management System (DMS), that in turn, sets crucial decisions for real time (e.g., voltage control, optimal power flow, fault management) or off-line (e.g., post-event analysis) applications on the estimated system state, Fig. 2.3. In particular, the management of

active distribution networks, to a large extent, exploits the presence of DERs, capable of providing auxiliary services through their controlled (re-)active power [38]. The latter is possible only if specific values, such as the losses and voltage drops, are measured or estimated. Additionally, the DSO should take into account the short time variability of DER power generation. To this end, beyond the fundamental need for DG monitoring, the operators are forced to rely on SE algorithms characterized by increased temporal granularity and robustness. To that aim, the DG utilities may resort to the recently introduced, low-cost, high-resolution μPMUs (micro Phasor Measurement Units) [39].

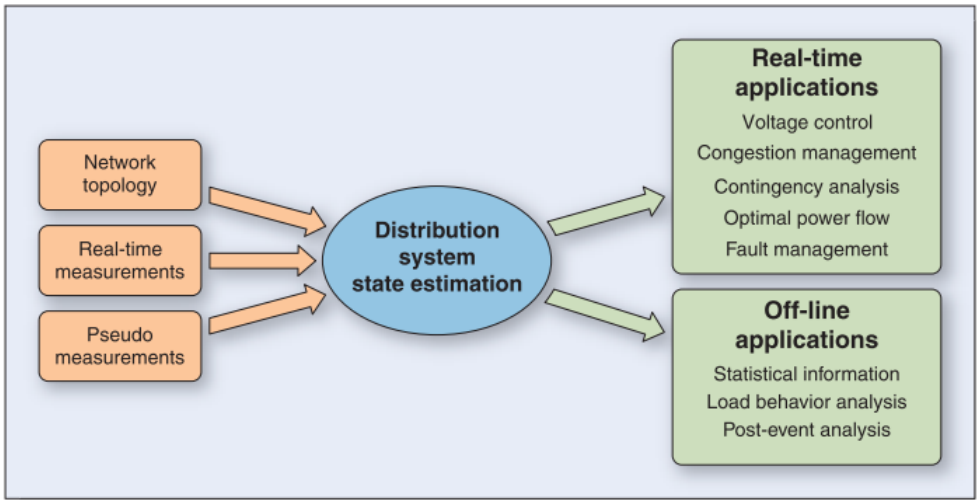

Figure 2.3: Functionalities of the distribution management system based on the estimated state [2].

SE algorithms for DGs face a number of different challenges compared to those of TGs [40]. Notably, the limited availability of measurement sets along with the specificities of radial/tree topologies often lead to ill-posed state estimation problems [41], [2]. Further, DG characteristics, such as the low reactance/resistance (X/R) ratio, make prohibitive a number of assumptions commonly adopted in the case of TG. For instance, in the high voltage system the dominant nature of inductive terms, i.e. $R << X$, allows to neglect the system resistances and exploit a constant gain matrix. This, in turn, leads to decoupled SE models which present algorithmic efficiency. In this context, in [42] the authors propose a branch current (BC) -based SE scheme. Here, instead of complex voltages, the state vector consists of complex branch currents in Cartesian coordinates. This leads to a reformulation of increased simplicity in the relationship between the state variables and the AC load flows, that, in turn, leads to a constant Jacobian matrix which is both both line-insensitive and state-independent. Thus, the resulting scheme turns out to be more appropriate than classical SE for the case of the radial-shaped DG, presenting insensitivity to the aforementioned sources of numerical instability. A BC-

based SE algorithm is also proposed in [43], where, additionally, the authors resort to the Lagrange multipliers to cope with zero power injections. A more analytical three-phase BC SE model is presented in [44], where the power on the branches and the squared branch current magnitudes are estimated. Finally, more recent works have taken into consideration the advent of active distribution networks. In [45], an efficient model is introduced, which is capable of encompassing phasor measurements and accounts for radial and weakly meshed topologies.

Concerning the limited measurement sets, in recent years, substantial efforts have been devoted to strategies aimed to deal with the lack of observability arising from insufficient measurements sets. Resorting to data mining techniques on historical load data profiles has been instrumental towards the definition of efficient strategies for the construction of pseudomeasurement sets. In [46], the authors generate pseudomeasurements from a few real measurements using artificial neural networks in combination with typical load profiles. They decompose the associated error of the generated pseudo measurements into several components through the Gaussian mixture model (GMM) in order to make it suitable for the WLS-SE problem formulation. Further, in [47] the authors present a closed loop-based robust SE. The proposed framework consists of a robust M-estimator and a machine learning algorithm that constructs pseudomeasurements, which are improved via a link that connects the SE and the machine learning function. By using such pseudomeasurements in conjunction with actual SCADA telemetry data (slightly) redundant measurement sets can be constructed. Reference [48], for instance, investigates a two-time scale SE based on weighted least squares. Separating the current and foreseen available information in the DG according to their availability in time, the authors formulate the classical SE problem in a two time scale fashion and investigate the efficiency of the resulting scheme under several load profile scenarios. Other works that deal with the limited measurement infrastructure (i.e., lack of system observability) are also presented in the literature. To name a few, in [49] a DG SE is presented which is based only on PMUs and leverages on a model reduction method. In particular, the method selects an arbitrary number of lines and buses and reduces them in one. The latter results into guaranteed system observability with a lower number of meters. In addition, in reference [50] the authors exploit the presence of stationary loads (nodes) in the DG and present a power flow solution over consecutive time instants. From this coupled formulation, they finally prove that the grid state can be recovered by leveraging on fewer meters.

Beyond classical Maximum Likelihood Estimation (MLE) approaches, the authors in [51], propose a Belief Propagation based SE scheme suitable for both TG and DG. Under the assumption that the system state can be characterized by a set of stochastic variables and by resorting to a set of prior distributions, the posterior distributions of the state variables can be efficiently computed using real-time measurements from both legacy

and high resolution smart metering data. Borrowing inspiration from [29], the authors in [52] propose a semi-definite programming (SDP) formulation for the SE problem as well as a convex relaxation. The authors prove that classic gradient methods present convergence problems even with small test cases. In contrast, their proposed method attains the globally optimum solution without suffering from numerical issues caused by the peculiarities of the DG. In the same context, [52] presents a multi-phase SE scheme for the distribution grid using the recently introduced low-cost micro-synchrophasors, also referred to as μPMUs[2] [53]. As their scheme is sensitive to noisy measurements, they separate the network in several parts including the μPMUs to mitigate the noise and solve the respective subproblems. As another approach, also beyond the MLE method, can be considered frameworks which are based on Kalman filter [54]. These estimators are referred in the literature as dynamic SE or forecasting-aided SE and their method involves the exploitation of a time-sequence of measurement snapshots [55].

2.3 Phasor Measurement Unit Technology

Over the recent years, SE has been transformed from an important application into a critical one. Inevitably, the interest in phasor measurement technology has reached its peak [15]. Leveraging on the global positioning system (GPS), the phasor measurement untis (PMUs) are able to provide time-synchronized and accurate phasor measurements at a very fast rate (30-60 samples per second) [56]. A PMU, installed at a specific bus, is capable of measuring the voltage phasor and the current phasors of a number (depending on the available communication channels) of incident branches. Because of their nature, these measurements are instrumental for a number of energy management system functionalities. Primarily, the phasor meters can significantly enhance monitoring features, either in a non-networked form (e.g., detection of power system disturbances) either communicating real time data (e.g., power system SE). Additionally, PMUs are expected to constitute the cornerstone for wide-area systems which will support the control centers to their corrective decisions according to the operational system conditions.

In the case of SE, these complex valued measurements, i.e. bus voltages and branch currents, can be related with the state variables in a linear fashion, in contrast with traditional SCADA measurements, such as power injections. This can be proved to be beneficial for a number of related research lines, such as power system observability, SE accuracy, algorithmic convergence and bad data detection. For now, PMUs augment the already existing SCADA measurements and consequently, the energy management centers leverage on hybrid SCADA/PMU data systems. However, their fast-growing deployment is possible to lead in the near future to PMU-based only state estimators.

[2] μPMUs are particularly suited for distribution networks since, given their extension, the use of the standard transmission-grid PMUs would be prohibitive.

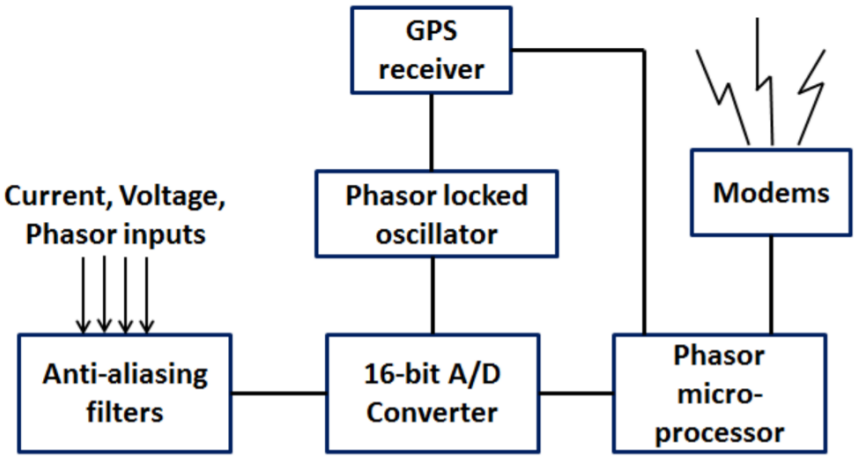

Figure 2.4: Phasor Measurement Units (PMU) block diagram [3].

With an increasing number of new PMU installations scheduled for the next years, naturally, arises the problem of the optimal PMU placement (OPP). That means, the optimal positioning for the phasor meters according to specific criteria, e.g., system observability, financial cost, communication constraints. The fundamental objective around OPP is of course system observability, that means, to guarantee observability with a minimum number of PMUs. In order to solve the OPP problem, the majority of the earlier works follow, either integer-programming-based methods [57], or heuristic solutions, such as genetic algorithms [58], simulating annealing [59] and particle swarm optimization [60]. Moreover, the placement of PMUs has been examined under different scenarios (with or without existing SCADA measurements) and several criteria (installation cost [61], channel limitation [62]). In the most recent years, interesting approaches based on convex optimization have also been proposed for the OPP problem [63]. In [64] for instance, the authors present an SDP formulation of the problem taking into concern the impact of PMU channel limits. Further, in [65] they propose a binary SDP based placement method which, beyond the classical installation purposes, can be exploited in order to identify temporarily unobservable areas of the system and the minimum needed number of pseudomeasurements to restore observability.

2.3.1 μPMUs for the Distribution Grid

In contrast with the dynamic high voltage systems, where the PMUs have already experienced a dramatic installation deployment, in the case of DG the phasor meters have only a limited presence in the primary distribution substations. As main reasons we can identify: (i) the high installation cost and; (ii) the fact that, as already mentioned, the DG in the past had only the role of transferring the energy from higher to low voltages. Conse-

quently, just the evaluation of the design conditions (e.g., fault currents) was necessary. This situation has changed recently, as we mentioned earlier in Chapter 1, through the advent of the so-called Smart Grid. The rapid integration of renewable and distributed energy resources has introduced a system behavioral uncertainty but also, opportunities for ancillary services (e.g., reactive power support). This, in turn, has raised the need for robust and real-time monitoring of the medium voltage system.

Figure 2.5: The μicro Phasor Measurement Unit (μPMU) device [4].

In this context, the recently introduced μPMUs [39] can be proved to be a reliable tool to address this need, Fig. 2.5. The common synchrophasor applications of interest at the high voltage system, such as SE and stability analysis, collect measurements across large distances. In contrast, μPMUs have been designed to address functionalities with respect to the medium-voltage distribution systems in order to enhance local decisions. This distribution-centric analysis is more challenging for a number of reasons. Specifically, in the case of DGs power flows are smaller and distances shorter. Given that, voltage phase angle differences of interest are typically two orders of magnitude smaller than those across transmission systems, i.e., lie usually at the level of small fractions of a degree. In addition to that, phase angle values of voltage and currents are small compared to measurement noise. Thus, μPMUs are exclusively designed to provide ultra-precise measurements, providing higher visibility compared to the classical PMUs. From an application perspective, the goal for the μPMUs data is to support monitoring and control functionalities. In both cases, the algorithms can significantly benefit from time-stamped and precise information. A number of works based on μPMUs have been already appeared in the literature. To mention a few, in [66] the authors propose a topology detection model based on synchrophasors in order to confirm the current system configuration.

Further, phasor data have also been exploited for event detection (e.g., voltage sags) and classification [67].

The same applies for the functionality of DG state estimation. Today DG SE relies on the few available SCADA measurements coming from the primary distribution substation and mostly, on pseudomeasurements. As a result, the accuracy of the algorithms can be characterized as moderate. Given that, the results from optimization functionalities (of the DMS) which are based on the estimated steady-state conditions, such as the optimal power flows, may diverge significantly from the optimal one. Consequently, even a small number of precise micro-synchrophasor could be beneficial [68]. Similarly to the case of TG, the exact location of the μPMUs is important and should be examined considering the system observability, but also with respect to a number of other criteria related with the DG requirements, such as the SE accuracy.

Chapter 3

Mathematical Preliminaries

In this chapter, we provide the fundamental elements of the mathematical tools that have been used throughout this thesis. More specifically, the basic foundations and concepts of convex optimization, such as convexity, convex optimization problems and duality, are given briefly. Then, the definition of non-convex problems is presented accompanied from a successive convex approximation method, commonly leveraged to deal with non-convex problems. Later on, we focus on specific applications, the binary semidefinite programming models and estimation. Finally, we present the concept of the alternating direction method of multipliers, a method which is extensively used throughout this dissertation. This synopsis is based on classical texts, such as [5] and [69], where we refer the reader for a more in-depth discussion.

3.1 Convex Optimization

Convex optimization is a special field of mathematical optimization dealing with a large class of problems, that is, convex optimization problems. This class of problems provide two fundamental features. Firstly, local optimality coincides with global optimality; and secondly, there are available very effective solution methods. These properties imply that, when a practical problem can be formulated as a convex optimization one, then the original problem can be solved reliably by means of efficient algorithms (e.g. interior point methods) in the scale of few seconds. Fortunately, many problems in different fields, such as power system engineering, signal processing and control, can be formulated as convex problems or can be transformed to. In the next subsections we present the basic foundations of this active mathematical optimization field.

3.1.1 Convexity

Definition 3.1. Consider a set C. The set is *convex* if the line segment that connects two points x_1, x_2 that belong in C lies in C. That means:

$$\theta x_1 + (1-\theta)\, x_2 \in C. \tag{3.1}$$

where $\theta \in \mathbb{R}$ with $0 \leq \theta \leq 1$.

Practically, that means that if a set is convex, any two points that lie in the set can be connected with an unobstructed path.

Definition 3.2. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. This function is *convex* if $\mathbf{dom}\, f$ is a *convex set* and if for all $\mathbf{x}$ and $\mathbf{y} \in \mathbf{dom}\, f$ applies:

$$f\left(\theta \mathbf{x} + (1-\theta)\, \mathbf{y}\right) \leq \theta f\left(\mathbf{x}\right) + (1-\theta)\, f\left(\mathbf{y}\right) \tag{3.2}$$

where $\theta \in \mathbb{R}$ with $0 \leq \theta \leq 1$.

Geometrically, inequality (3.2) implies that the chord connecting $(f(\mathbf{x}), \mathbf{x})$ and $(f(\mathbf{x}), \mathbf{y})$ lies above the graph of f. Furthermore, the function f is strictly convex if strict inequality holds in (3.2) whenever $\mathbf{x} \neq \mathbf{y}$ and $0 < \theta < 1$. Finally, f is concave if $-f$ is convex, and strictly concave if $-f$ is strictly convex.

Definition 3.3. The *epigraph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\mathbf{epi}\, f = \left\{(\mathbf{x}, t) \mid \mathbf{x} \in \mathbf{dom}\, f, f(\mathbf{x}) \leq t\right\}. \tag{3.3}$$

The epigraph allows to establish a relationship between convex sets and convex functions. Namely, a function is convex if and only if its epigraph is a convex set. Now, we establish the conditions that allow us to identify convexity.

Proposition 3.4 (First-order conditions)**.** *Consider a differentiable function f. The function is convex if $\mathbf{dom}\, f$ is a convex set and:*

$$f\left(\mathbf{y}\right) \geq f\left(\mathbf{x}\right) + \nabla f\left(\mathbf{x}\right)^T \left(\mathbf{y} - \mathbf{x}\right) \tag{3.4}$$

holds for all $\mathbf{x}$ and $\mathbf{y} \in \mathbf{dom}\, f$.

Here, $f\left(\mathbf{x}\right) + \nabla f\left(\mathbf{x}\right)^T \left(\mathbf{y} - \mathbf{x}\right)$ is the first order Taylor expansion of f near $\mathbf{x}$. From (3.4) and Fig. 3.1 we observe that the first-order Taylor expansion act as a global underestimator for a convex function f. This is the most important property for convex functions, as (3.4) essentially shows that if $\nabla f\left(\mathbf{x}\right) = 0$, then $\mathbf{x}$ is a global minimizer of f and $f\left(\mathbf{y}\right) \geq f\left(\mathbf{x}\right)$ for all $\mathbf{y} \in \mathbf{dom}\, f$.

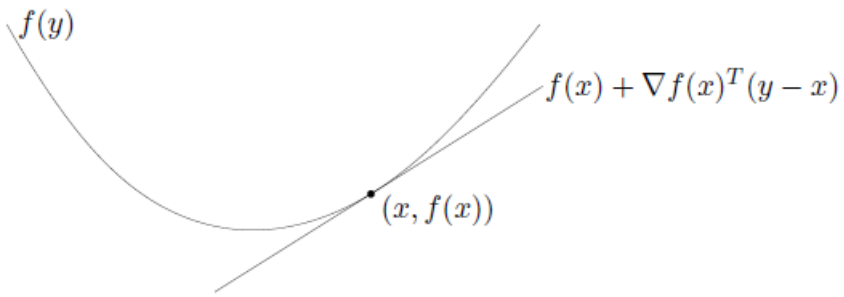

Figure 3.1: If function f is convex and differentiable, then $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{y} \in \mathbf{dom}\, f$ [5].

Proposition 3.5 (Second-order conditions)**.** *Now, consider a twice differentiable function f, that is, its second derivative (Hessian) $\nabla^2 f$ exists at each point in* $\mathbf{dom}\, f$*, which is open. Then, the function f is convex if and only if* $\mathbf{dom}\, f$ *is convex and its Hessian is positive semidefinite:*

$$\nabla^2 f(\mathbf{x}) \geq 0. \tag{3.5}$$

for all $\mathbf{x} \in \mathbf{dom}\, f$*.*

For a function f on $\mathbb{R}$ this condition reduces to $f''(x) \geq 0$, that means the derivative is nondecreasing. The condition in (3.5) can be interpreted geometrically as the requirement that the graph of the function f has positive curvature at $\mathbf{x}$.

3.1.2 Convex Optimization Problems

The basic optimization problem is formulated as:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) && \text{(3.6a)} \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m && \text{(3.6b)} \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p && \text{(3.6c)} \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the *optimization variable*, $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the *objective or cost function*, $f_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, \ldots, m$ are the *inequality constraint functions*, and $h_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, \ldots, p$ are the *equality constraint functions*. The inequalities $f_i(\mathbf{x}) \leq 0$ are called the *inequality constraints*, and the equations $h_i(\mathbf{x}) = 0$ are called the *equality constraints*. If in (3.6) there are no constraints we say that the problem in *unconstrained*.

The goal in (3.6) is to find the optimal $\mathbf{x}$ that minimizes f_0 and satisfies $f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m$ and $h_i(\mathbf{x}) = 0,\ i = 1, \ldots, p$. The optimal $\mathbf{x}$ belongs to the domain of the optimization problem and is called feasible.

Definition 3.6. We say that the optimization problem (3.6) composes a convex optimization problem, if the objective function $f_0(\mathbf{x})$ is convex, the inequality constraint functions $f_i(\mathbf{x}), i = 1, \ldots, m$ are convex and the equality constraint functions $h_i(\mathbf{x}), i = 1, \ldots, p$ are affine.

Definition 3.7. The set of points for which the objective and all constraint functions are defined,

$$\mathscr{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i. \tag{3.7}$$

is called the domain of the optimization problem (3.6).

Definition 3.8. A point $\mathbf{x} \in C$ is called *feasible* if it satisfies $f_i(\mathbf{x}) \leq 0$ for $i = 1, \ldots, m$ and $h_i(\mathbf{x}) = 0$ for $i = 1, \ldots, p$.

Definition 3.9. The *optimal value* $p^\star$ of the optimization problem is given by

$$p^\star = \inf \{ f_0(\mathbf{x}) \mid f_i(\mathbf{x}) \leq 0, i = 1, \ldots, m, h_i(\mathbf{x}) = 0, i = 1, \ldots, p \}. \tag{3.8}$$

Furthermore, we say $\mathbf{x}^\star$ is an *optimal point*, if $\mathbf{x}^\star$ is feasible and $f_0(\mathbf{x}^\star) = p^\star$.

Definition 3.10. The set of all optimal points is the optimal set and is denoted:

$$X_{opt} = \inf \{ \mathbf{x} \mid f_i(\mathbf{x}) \leq 0, i = 1, \ldots, m, h_i(\mathbf{x}) = 0, i = 1, \ldots, p, f_0(\mathbf{x}^\star) = p^\star \}. \tag{3.9}$$

If there exists an optimal point for (3.6), we say that the optimal value is achieved and the problem is solvable.

Definition 3.11. We say a feasible point $\mathbf{x}$ is locally optimal if there is an $R > 0$ such that

$$f_0(\mathbf{x}) = \inf \{ f_0(\mathbf{z}) \mid f_i(\mathbf{z}) \leq 0, i = 1, \ldots, m, h_i(\mathbf{z}) = 0, i = 1, \ldots, p, \|\mathbf{z} - \mathbf{x}\|_2 \leq R \}. \tag{3.10}$$

This is to say that a locally optimal point minimizes f_0 over nearby feasible points.

3.1.3 Duality

Definition 3.12. The Lagrange duality entails into augmenting the objective function in the optimization problem (3.6) by the weighted sum of the constraint functions $f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m$ and $h_i(\mathbf{x}) = 0, \ \ i = 1, \ldots, p$. The *Lagrangian* $\mathscr{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ of optimization problem (3.6) is defined as:

$$\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}), \tag{3.11}$$

where λ_i is referred to as the *Lagrange multiplier* associated to constraint $f_i(\mathbf{x}) \leq 0$ and ν_i as the Lagrange multiplier associated to constraint $h_i(\mathbf{x}) = 0$. The vectors $\boldsymbol{\lambda}, \boldsymbol{\nu}$ are called the *Lagrange multiplier vectors* or the *dual variables* of the optimization problem.

Definition 3.13. The *Lagrange dual function* $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is defined as

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathscr{D}} \mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathscr{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right). \tag{3.12}$$

Since the dual function $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is given by the pointwise infimum of a family of affine functions, it is concave, even when the original problem (3.6) is not convex.

Proposition 3.14 (Lower bound on optimal value)**.** *The dual function yields a lower bound on the optimal value $p^\star$ of the problem* (3.6)*. For any $\boldsymbol{\lambda} \geq 0$ and $\boldsymbol{\nu}$ we have*

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star. \tag{3.13}$$

Any pair $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ with $\boldsymbol{\lambda} \geq 0$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathbf{dom}\, g$ is denoted as *dual feasible.*

Since for each pair $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ with $\boldsymbol{\lambda} \geq 0$, the dual function provides a lower bound on the optimal value $p^\star$, one might want to obtain the best lower bound possible. This is the dual problem.

Definition 3.15. The *Lagrange dual problem* is given by

$$\underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{maximize}} \quad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \tag{3.14a}$$

$$\text{subject to} \quad \boldsymbol{\lambda} \geq 0 \tag{3.14b}$$

We denote by $d^\star$ the optimal value of the Lagrange dual problem and the respective dual variables $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ As previously said, the Lagrange dual problem allows us to find the best lower bound possible. The Lagrange dual problem (3.14) is a convex optimization problem, since the objective to be maximized is concave and the constraint is convex, even if the problem (3.6) is not convex.

Proposition 3.16 (Duality gap)**.** *The optimal value of the Lagrange dual problem, i.e., $d^\star$, is the best lower bound with respect to $p^\star$. The inequality*

$$d^\star \leq p^\star, \tag{3.15}$$

is called weak duality *and the difference $p^\star - d^\star$ is defined as the* optimal duality gap*. The optimal duality gap is always nonnegative. If $p^\star = d^\star$, then* strong duality *holds.*

In general, strong duality does not hold. When the primal problem is a convex optimization problem, usually (but not always), strong duality holds. There exist several results that establish conditions under which strong duality holds. These conditions are called *constraint qualifications.* One important constraint qualification under which strong duality holds is *Slater's condition.*

Proposition 3.17 (Slater's condition)**.** *There exists an* $\mathbf{x} \in \mathbf{relint}\, \mathscr{D}$ *such that*

$$f_i(\mathbf{x}) < 0, i = 1, \ldots, m \tag{3.16}$$

and

$$h_i(\mathbf{x}) = 0, i = 1, \ldots, p. \tag{3.17}$$

Furthermore, a point satisfying these conditions is called an strictly feasible *point.*

Slater's condition not only implies strong duality for convex problems. It also implies the existence of a dual feasible pair $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ with $g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = p^\star = d^\star$.

3.2 Non-convex Optimization

In many cases, the formulation of the problem under consideration leads to a non-convex optimization model. That means, several local optima exist and additionally, there are no effective methods for solving the problem.

A first approach to deal with non-convex models is to leverage on global optimization methods. However, in that case the computational cost is often an exponential function of the number of the decision variables (e.g., sensor selection problem with binary decision vector variable). Another approach is local optimization. In this case, a compromise is made to find a locally optimal solution. That is, a solution that minimizes the objective over nearby feasible points. The method entails the replacement of the original problem with an appropriate approximation, which is able to be solved by existing fast methods. The price one has to pay is that there is no guarantee that the solution is globally optimal. However, these methods are widely used as in many practical problems an efficient local optima is sufficient (e.g., power system state estimation). Further, these types of methods are highly dependent on the initialization point, which practically determines the solution to which the method will converge. A common way to choose the initial point is by a simple randomization. Another approach, which can possibly lead to an improved initial guess, is to primarily solve a convex relaxation of the original non-convex problem.

3.2.1 Successive Convex Approximation

The successive convex approximation (SCA) constitutes a framework which is leveraged to solve non-convex optimization problems [70]. More specifically, the underlying idea is to replace the initial non-convex problem with an approximated convex one, solved in an iterative manner. The final goal of solving this sequence of strongly convex approximations is to efficiently compute locally optimal solutions. A number of different approaches can be found in the literature, providing several features, such as parallel and distributed implementations of the SCA framework [71] and solution feasibility over the iterations.

For instance, the latter can be exploited for on-line implementations of the SCA method, which is relevant for optimization problems, such as the power system state estimation.

Definition 3.18. Consider the following optimization problem:

$$\underset{\mathbf{x}}{\text{minimize}} \quad f_0(\mathbf{x}) \tag{3.18a}$$

$$\text{subject to} \quad f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \tag{3.18b}$$

$$x \in \mathscr{X} \tag{3.18c}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the *optimization variable*, $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the *smooth non-convex* objective function subject to the non-convex inequality constraints $f_i : \mathbb{R}^n \to \mathbb{R}$.

First, we make the following assumptions for the non-convex problem in (3.18):

A.1 $\mathscr{X} \subseteq \mathbb{R}^n$ is closed and convex;

A.2 f_0 and each f_i are continuously differentiable on $\mathscr{X}$;

A.3 $\nabla_x f_0$ is Lipschitz continuous on $\mathscr{X}$;

A.4 f_0 is coercive on $\mathscr{X}$;

Under these assumptions, which are quite standard and are satisfied by a large class of problems, the successive convex approximation framework entails into solving the following sequence of strongly convex approximations of (3.18):

$$\underset{\mathbf{x}}{\text{minimize}} \quad \tilde{f}_0(\mathbf{x}; \mathbf{x}^\nu) \tag{3.19a}$$

$$\text{subject to} \quad \tilde{f}_i(\mathbf{x}; \mathbf{x}^\nu) \leq 0, \quad i = 1, \ldots, m \tag{3.19b}$$

$$x \in \mathscr{X} \tag{3.19c}$$

where $\tilde{f}_0(\mathbf{x}; \mathbf{x}^\nu)$ and $\tilde{f}_i(\mathbf{x}; \mathbf{x}^\nu)$ represent the approximations of $f_0(\mathbf{x})$ and $f_i(\mathbf{x})$ at the current iterate $\mathbf{x}^\nu$, respectively.

Here, a number of further assumptions should be made for the approximated problem in (3.19):

B.1 $\tilde{f}_0(\bullet; \mathbf{y})$ is convex on $\mathscr{X}$ for all $\mathbf{y} \in \mathscr{X}$;

B.2 $\nabla_x \tilde{f}_0(\mathbf{y}; \mathbf{y}) = \nabla_x f_0(\mathbf{y})$ for all $\mathbf{y} \in \mathscr{X}$;

B.3 $\tilde{f}_i(\mathbf{x}; \mathbf{x}^\nu)$ is convex on $\mathscr{X}$ for all $\mathbf{y}$;

B.4 $f_i(\mathbf{x}) \leq \tilde{f}_i(\mathbf{x}; \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathscr{X}$;

B.5 $\nabla_x \tilde{f}_i(\mathbf{y}; \mathbf{y}) = \nabla_x f_i(\mathbf{y})$ for all $\mathbf{y} \in \mathscr{X}$;

Algorithm 3.1 Successive Convex Approximation framework

1: Initialize $\nu = 0$, $\gamma^{(\nu)} \in (0, 1]$, $\{\mathbf{x}^{(0)} \in \mathscr{X}\}$
2: Compute $\{\hat{\mathbf{x}}^{(\nu)}\}$, the solution of (3.19)
3: $\{\mathbf{x}^{(\nu+1)} \leftarrow \mathbf{x}^{(\nu)} + \gamma^{(\nu)}(\hat{\mathbf{x}}^{(\nu)} - \mathbf{x}^{(\nu)})\}$
4: If $\mathbf{x}^{(\nu+1)}$ is a stationary solution of (3.18) STOP
5: $\nu \leftarrow \nu + 1$ and go to step 2

Here, assumptions B.1 and B.3 guarantee that the approximated problem (3.19) is strongly convex. Further, assumption B.4 guarantees iterate feasibility and, B.2 and B.5 that the approximated objective function and constraints have the same first order behavior with the original ones.

Algorithm 3.1 summarizes the corresponding iterative procedure.

3.3 Binary Semidefinite Programming

In this subsection, we aim to describe gradually a special class of mathematical optimization problems, namely, the binary semidefinite programming (BSDP) models. BSDP problems have gathered lots of interest, mostly because of their combinatorial nature, with problems arising in systems and control, communications and signal processing.

First, consider the following problem:

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T\mathbf{x} && \text{(3.20a)} \\ \text{subject to} \quad & x_1\mathbf{F1} + \cdots + x_n\mathbf{Fn} + \mathbf{G} \succeq \mathbf{0} && \text{(3.20b)} \\ & \mathbf{Ax} = \mathbf{b} && \text{(3.20c)} \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{G}, \mathbf{F1}, \ldots, \mathbf{Fn} \in \mathbb{S}^k$, with $\mathbb{S}^k$ standing for the space of positive semidefinite $k \times k$ matrices.

The problem in (3.20) amounts to minimize a linear objective function under equality and linear matrix inequality (LMI) constraints and is called a semidefinite programming problem (SDP) [5].

Further, in plenty of cases, depending on the nature of the problem, the decision variable is possible to be binary, i.e., $\mathbf{x} \in \mathbb{Z}_2^n$. In that case, the problem (3.20) takes the following form:

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T\mathbf{x} && \text{(3.21a)} \\ \text{subject to} \quad & x_1\mathbf{F1} + \cdots + x_n\mathbf{Fn} + \mathbf{G} \succeq \mathbf{0} && \text{(3.21b)} \\ & \mathbf{Ax} = \mathbf{b} && \text{(3.21c)} \\ & \mathbf{x} \in \{0, 1\} && \text{(3.21d)} \end{aligned}$$

and constitutes a non-convex binary semidefinite programming (BSDP) problem.

The model in (3.21) may arise for example in a sensor selection problem where the goal is to construct a binary $\mathbf{x}$ that selects to activate a specific number of sensors. Commonly, this kind of problems can be implemented and solved from commercial mathematical softwares that utilize optimization techniques, such as the branch and bound method. Another approach, based on convex optimization, is to relax the non-convex constraint $\mathbf{x} \in \{0, 1\}$ with its convex counterpart, i.e., $\mathbf{x} \in [0, 1]$. As the latter does not produce a binary solution, a rounding procedure and a local search should follow.

3.4 Convex Optimization for Estimation Problems

3.4.1 Estimation

Consider the following linear measurement model:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{e} \tag{3.22}$$

where $\mathbf{z} \in \mathbb{R}^m$ is a vector of noisy measurements, $\mathbf{x} \in \mathbb{R}^n$ is a vector of problem parameters to be estimated and $\mathbf{A} \in \mathbb{R}^{m \times n}$ stands for a matrix that describes the relationship between $\mathbf{z} \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$ with $m \geq n$. In addition, $\mathbf{e} \in \mathbb{R}^m$ stands for the independent and identically distributed measurement noise. The estimation problem entails to compute the optimal $\hat{\mathbf{x}}$ with given $\mathbf{z} \in \mathbb{R}^m$:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \quad \|\mathbf{z} - \mathbf{A}\mathbf{x}\| \tag{3.23}$$

in order to minimize the residual vector:

$$\mathbf{r} = \mathbf{A}\hat{\mathbf{x}} - \mathbf{y} \tag{3.24}$$

Now, if we express $\mathbf{A}\mathbf{x}$ as:

$$\mathbf{A}\mathbf{x} = x_1\mathbf{a_1} + \cdots + x_n\mathbf{a_n} \tag{3.25}$$

where $x_1\mathbf{a_1}+, \cdots + x_n\mathbf{a_n}$ are the columns of $\mathbf{A} \in \mathbb{R}^{m \times n}$, the problem in (3.23) can be interpreted as the approximation of the vector $\mathbf{z}$ by a linear combination of the columns of $\mathbf{A}$, with the minimum error and deviation measured in the norm $\| \ \|_x$. This problem is also known as the regression problem in the field of statistics.

Commonly, the problem in (3.23) leverages on the Euclidean norm and points out to the so-called least squares optimization problem:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \quad \|\mathbf{z} - \mathbf{A}\mathbf{x}\|_2^2 \tag{3.26}$$

that constitutes a convex problem, i.e., we can find the global optimal solution; and since $m \geq n$ the solution is given by:

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{z} \tag{3.27}$$

3.4.2 Regularized Estimation

The regularized estimation refers to a general bi-criterion problem that constitutes a weighted sum. The goal is to minimize the following objective function:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \quad \|\mathbf{z} - \mathbf{A}\mathbf{x}\| + \tau\|\mathbf{x}\| \tag{3.28}$$

where τ is a positive parameter that controls the trade-off between data fitting and the regularization term. The problem in (3.28) is used in several contexts. A case of special interest is when $\mathbf{A}$ is ill-conditioned or even singular and through the regularization term is exploited prior knowledge on the unknown parameters of $\mathbf{x}$. The general case of (3.28) can take several forms depending the application. Below we present specific formulations that have been used across this dissertation.

3.4.2.1 Tikhonof Regularization

One of the most common approaches is Tikhonof regularization which combines two Euclidean norms:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \quad \|\mathbf{z} - \mathbf{A}\mathbf{x}\|_2^2 + \tau\|\mathbf{x}\|_2^2 \tag{3.29}$$

The solution to the above quadratic optimization problem is given by:

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A} + \tau\mathbf{I})^{-1}\mathbf{A}^T\mathbf{z} \tag{3.30}$$

One can easily notice that $\mathbf{A}^T\mathbf{A} + \tau\mathbf{I} \succ \mathbf{0}$ for any positive τ. The latter makes problem (3.29) solvable without any requirements on the rank of $\mathbf{A}$.

3.4.2.2 ℓ_1- Regularization

In the case where someone seeks for a sparse solution, the Euclidean norm in the regularization term is replaced from the ℓ_1-norm:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \quad \|\mathbf{z} - \mathbf{A}\mathbf{x}\|_2^2 + \tau\|\mathbf{x}\|_1 \tag{3.31}$$

Here, τ controls the trade-off between the fidelity term and the sparsity of the estimated vector $\hat{\mathbf{x}}$.

3.4.2.3 Total Variation Regularization

Another case of regularization is the so-called total variation estimation which assigns penalty to large deviations between the neighboring parameters of the estimated vector $\hat{\mathbf{x}}$:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \quad \|\mathbf{z} - \mathbf{A}\mathbf{x}\|_2^2 + \tau\sum_{x=1}^{n}|x_n - x_{n-1}| \tag{3.32}$$

In this case, τ controls the trade-off between the fidelity term and the total variation of the vector $\hat{\mathbf{x}}$.

3.5 Alternating Direction Method of Multipliers

Here, we briefly discuss the alternating direction method of multipliers (ADMM), an efficient algorithm which is well suited for distributed convex optimization problems in applied statistics and machine learning [72]. The ADMM offers a number of benefits, as practically it combines the benefits of its precursors. Specifically, the advantage of decomposition when the objective function is separable, as in the dual ascent approach, and the robustness in terms of convergence, as in the case of augmented Lagrangian methods.

First, consider the following equality-constrained convex optimization problem:

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) \tag{3.33a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b} \tag{3.33b}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the optimization variable, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$ and, $f : \mathbb{R}^n \to \mathbb{R}$ is a convex objective function.

Now, consider the following optimization problem:

$$\underset{\mathbf{x},\mathbf{z}}{\text{minimize}} \quad f(\mathbf{x}) + g(\mathbf{z}) \tag{3.34a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c} \tag{3.34b}$$

where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^m$ are the optimization variables, $\mathbf{A} \in \mathbb{R}^{p \times n}$, $\mathbf{B} \in \mathbb{R}^{p \times m}$ and $\mathbf{c} \in \mathbb{R}^p$. Let us assume that $f(\mathbf{x})$ and $g(\mathbf{z})$ are convex functions. Compared to the general convex optimization problem in (3.33), the difference in (3.34) is that the optimization variable $\mathbf{x}$ has been split in two parts, $\mathbf{x}$ and $\mathbf{z}$, assuming that the objective function $f(\mathbf{x})$ in (3.33) is separable across the splitting.

The optimal value $p^\star$ of the optimization problem (3.34) is denoted by:

$$p^\star = \inf \{ f(\mathbf{x}) + g(\mathbf{z}) \mid \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c} \}. \tag{3.35}$$

and the Lagrangian of (3.34) is given by:

$$\mathscr{L}(\mathbf{x}, \mathbf{z}, \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^T(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}) + (\rho/2)\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\|_2^2 \tag{3.36}$$

In a glance, ADMM consists of the following steps:

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \mathscr{L}\big(\mathbf{x}, \mathbf{z}^k, \mathbf{y}^k\big) \tag{3.37}$$

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \mathscr{L}\big(\mathbf{x}^{k+1}, \mathbf{z}, \mathbf{y}^k\big) \tag{3.38}$$

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \rho\big(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}\big) \tag{3.39}$$

where $\rho > 0$ is the positive control parameter for the dual variable update step size. One can observe that ADMM consists of an $\mathbf{x}$ minimization step (3.37), a $\mathbf{z}$ minimization step (3.38), and the dual variable $\mathbf{y}$ update (3.39). The main difference of ADMM with respect to its precursors is that the optimization variables $\mathbf{x}$ and $\mathbf{z}$ are updated in a sequential fashion, which accounts for the term alternating direction. This is exactly what allows for problem decomposition when the objective function is separable.

Chapter 4

SCA-based Robust and Distributed State Estimation

In this chapter, we present a *hybrid* SE scheme exploiting both PMU and legacy measurements. To solve the resulting non-convex SE problem, we resort to a Successive Convex Approximation (SCA) scheme where the objective function is replaced by a sequence of strongly convex ones solved in an iterative manner. Then, we improve the convergence rate of the algorithm by exploiting second-order information from the original objective function. We also show that, for multi-area scenarios, the problem can be solved in a distributed fashion reaching the accuracy of its centralized counterpart. To do so, we propose an algorithm based on the ADMM. Finally, we present a robust version capable of performing SE and (sparse) bad data detection by means of a LASSO formulation. The performance of the proposed schemes is assessed via computer simulations for a number of IEEE standard test cases. Other state-of-the art methods are used as benchmarks.

4.1 Introduction

As already mentioned in Chapter 2, in recent years, the high voltage system presents a new operational environment due to the increasing penetration of RES and the evolution of the energy markets [73]. The RES power production, mainly photovoltaic and wind energy plants, is characterized from intermittency, which in turn, has introduced further uncertainty to the system state. Nevertheless, the system stability requierements and the energy market operation demand a timely and accurate monitoring. To that aim, the deployment of PMUs is instrumental [74], as they can significantly improve the accuracy of the SE via precise and time-synchronized measurements. However, for the time being such measurements are complemented by observations collected from the SCADA system, due to cost considerations. Thus, the operators set their decisions on hybrid PMU-SCADA systems, which bring into the game the development of novel *hybrid* schemes [75], [19].

In [76] for instance, the authors provide a detailed study on classic Gauss-Newton-based SE aided by PMUs.

On the other hand, the evolution of the energy market has stimulated the energy trading among transmission system operators. This has resulted into an increased exchange of power along the tie-lines between adjacent areas or countries, possibly under the control of different utilities. A totally *independent* operation of such areas is, thus, no longer viable. Accordingly, *centralized* SE schemes are not directly applicable either, due to data security and privacy concerns of regional utilities, or computational complexity considerations. This, has raised the need for hierarchical [23] and decentralized [28] SE frameworks, offering accurate monitoring and reduced computational burden. With regard to the *hierarchical* SE, the main drawback is the vulnerability of a centralized point failure that could be proven vital. In contrast, the decentralized rationale can overcome this barrier, leveraging on suitable algorithms for distributed optimization [25]. In this context, a number of interesting works appear in the literature [29], based on optimization methods [30] which have already been used in other fields, such as communications and signal processing [77]. An extended discussion on the implementation of distributed SE schemes in the framework of 5G communications systems is given in [11]. Other efficient SE algorithms in the same context can be also found [32], [78]. Nevertheless, as plenty of these schemes are based on the Gauss-Newton method, their distributed implementation is not straightforward [79]. The same holds for bad data detection (BDD) schemes.

In many occasions the measurements might be corrupted with gross error beyond the usual white Gaussian noise. Therefore, BDD schemes are leveraged in order to remove the outliers that bias the SE. However, the commonly preferred BDD frameworks, namely, the chi-squares test and the LNRT [80], are performed a posteriori and more importantly, are subject to BD identification failure. This holds in particular for the presence of multiple corrupted observations where the error is interacting between the residuals of different measurements [81, Ch.5,6]. In addition to that, the incorporation of new type of meters to the SE functionality, i.e. PMUs, introduces further obstacles. More specifically, the fact that historical data and experience on phasor measurements is limited has raised a discussion with respect to the appropriate weight introduction to these values. A number of works have proposed ad-hoc methods to overcome this issue [82], [83]. However, the algorithmic complexity is significantly increased.

4.1.1 Contribution

Considering all the above, in this chapter we present a robust hybrid SE scheme for the TG. Our final goal is to present a SE model which is capable of: (i) encompassing efficiently legacy and PMU measurements, (ii) being implemented in a distributed fashion with limited accuracy degradation compared to its *centralized* counterpart and, (iii) offering extended robustness against outliers. Specifically, the contributions of this chapter

can be summarized as follows:

1) A hybrid SE scheme based on a successive convex approximation (SCA). Specifically, as (already mentioned) the SE problem suffers from inherent non-convexity. To overcome this barrier we resort to a framework that follows the spirit of SCA algorithms [70], [84] applied successfully in other fields, such as communications and signal processing [85]. Specifically, the non-convex part (preserving as much from the original convexity) of the objective function is replaced by a sequence of strongly convex ones and the problem is solved in an iterative manner. In addition to that, we exploit second-order information from the non-convex objective function, keeping the computational complexity affordable. By doing so, we improve significantly the convergence rate of the algorithm, both in *centralized* and distributed settings. In contrast with other SE schemes [29], [81], the proposed scheme is able to attain high accuracy over a diversity of measurement set scenarios and power system test cases. This comes in a relatively fast convergence rate due to the exploitation of second-order information from the initial objective function.

2) A decentralized version of the SCA-SE algorithm. The adopted iterative SCA framework [71] can be easily implemented in a distributed fashion, as opposed to other similar schemes [32]. This makes it suitable for multi-area SE settings, which are typically exploited for inter-regional collaborations between different system operators. For the decentralized implementation of the SCA-SE scheme we leverage on the ADMM and its superior properties. In contrast with other similar works [24], the proposed scheme needs a limited information exchange, preserving privacy between transmission system operators. More importantly, unlike other interesting approaches [29], there is a limited SE accuracy penalty associated to such decentralized solution.

3) A robust version of the hybrid SCA-SE scheme conducting joint state estimation and bad data detection. To that aim, we reformulate the SE problem in a LASSO optimization framework [86]. The underlying idea is to promote sparsity in the vector of corrupted measurements. This applies for legacy and PMU measurements. In addition to that, we derive an upper bound of the residual error after bad data cleansing and conclude that the proposed method can be regarded as an instantiation of Huber's estimator. Beyond classical BDD methods, such as the largest normalized residual test [81, Ch.5,6], the proposed scheme performs jointly SE and BDD and so, the convergence rate of the algorithm does not deteriorate. More importantly, opposed to other schemes [81, Ch.5,6], [87], the robust SCA-SE is able to perform well even in scenarios where multiple corrupted measurements are present.

The rest of the chapter is organized as follows. In the next section we provide the SE system model. The multiarea SE is the subject of Section 4.3 and next, in Section 4.4, we analyze the successive convex approximation framework in addition with our approach to incorporate second order information. The distributed counterpart of the algorithm is

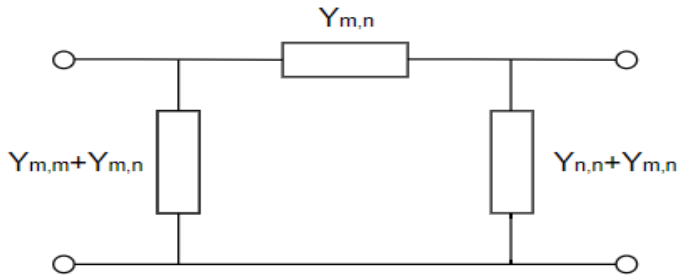

Figure 4.1: Y-equivalent circuit for a two-bus network.

provided in section 4.5, while in Section 4.6 we introduce the robust counterpart of our scheme. Finally, the proposed framework is numerically assessed in Section 4.7 and we draw our conclusions in Section 4.8.

4.2 System Model

Consider a power network composed of N buses denoted by the set $\mathscr{N} \in \{1, \ldots, N\}$. The interconnection between these buses is modeled by graph $\mathbf{G}$. Accordingly, the pair of buses $m, n \in \mathscr{N}$ are interconnected if $(m, n) \in \mathbf{G}$. From the Kirchhoff's law, the complex current injections at the buses, i.e., $\mathbf{i} = [I_1, \ldots, I_N]^T$, satisfies:

$$\mathbf{i} = \mathbf{Y}\mathbf{v} \tag{4.1}$$

with $\mathbf{v} = [V_1, \ldots, V_N]^T$ standing for the complex bus voltages that define the state of the system, and $\mathbf{Y} \in \mathbb{C}^{N \times N}$ being the nodal admittance matrix[1] with entries given by:

$$Y_{m,n} = \begin{cases} -y_{m,n} & \text{if} \quad (m,n) \in \mathbf{G}, \\ \overline{y}_n + \sum_{l \in \mathscr{N}} (y_{n,l} + \overline{y}_{n,l}) & \text{if} \quad m = n, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

Variables $y_{m,n}$, $\overline{y}_{m,n}$ stand for the admittance and shunt admittance in the line connecting buses m and n, respectively; and $\overline{y}_n$ stands for the shunt admittance at the nth bus, Fig. 4.1.

Typically, measurements are categorized into legacy and synchronized measurements. The former group relates the state vector $\mathbf{v}$ through *nonlinear* equations. In this category, we have the following types:

- Power injections: The power injection at bus n reads:

$$S_n = P_n + \mathrm{j}Q_n = V_n I_n^* \tag{4.3}$$

[1]For simplicity, we have ignored transformer taps in the formulation.

where S_n stands for the complex power at bus n, P_n for the active power injection, Q_n for the reactive power injection, V_n for the complex voltage and I_n^* for the conjugate of the complex current.

Let us define here the matrix $\mathbf{M}_n \in \mathbb{R}^{2N\times 2N}$ as follows:

$$\mathbf{M}_{P,n} = \sum_{j=1}^{N} \left(\mathbf{e}_n \mathbf{e}_j^T\right) \otimes \mathbf{B}_P(Y_{n,j}) \tag{4.4}$$

with $\mathbf{e}_n \in \mathbb{R}^N$ which takes 1 at its n^{th} position and 0 elsewhere. In addition, define:

$$\mathbf{B}_P(Y_{n,j}) = \begin{bmatrix} \Re\{Y_{n,j}\} & -\Im\{Y_{n,j}\} \\ \Im\{Y_{n,j}\} & \Re\{Y_{n,j}\} \end{bmatrix}. \tag{4.5}$$

Bearing this in mind, the active power injection can be expressed as:

$$P_n = \mathbf{x}^T \mathbf{M}_{P,n} \mathbf{x} \tag{4.6}$$

where $\mathbf{x}$ stands for the Cartesian representation of the system state, that is $\mathbf{x} = [\Re\{V_1\}, \Im\{V_1\}, \ldots, \Re\{V_N\}, \Im\{V_N\}]^T \in \mathbb{R}^{2N}$.

Similarly, the reactive power injection at the nth bus reads

$$Q_n = \mathbf{x}^T \mathbf{M}_{Q,n} \mathbf{x} \tag{4.7}$$

where

$$\mathbf{M}_{Q,n} = \sum_{j=1}^{N} \left(\mathbf{e}_n \mathbf{e}_j^T\right) \otimes \mathbf{B}_Q(Y_{n,j}) \tag{4.8}$$

with

$$\mathbf{B}_Q(Y_{n,j}) = \begin{bmatrix} -\Im\{Y_{n,j}\} & -\Re\{Y_{n,j}\} \\ \Re\{Y_{n,j}\} & -\Im\{Y_{n,j}\} \end{bmatrix}. \tag{4.9}$$

- Power branch measurements: The complex power flowing from bus m to bus n is expressed as:

$$\begin{aligned} S_{m,n} &= P_{m,n} + \mathrm{j} Q_{m,n} = V_m I_{m,n}^* \\ &= V_m \left(V_m^* - V_n^*\right) y_{m,n}^* + |V_m|^2 \overline{y}_m^* \end{aligned} \tag{4.10}$$

The active and reactive power flows from bus m to bus n can be expressed as follows:

$$\begin{aligned} P_{m,n} &= \mathbf{x}^T \mathbf{M}_{P,mn} \mathbf{x} \\ Q_{m,n} &= \mathbf{x}^T \mathbf{M}_{Q,mn} \mathbf{x} \end{aligned} \tag{4.11}$$

where

$$\begin{aligned}\mathbf{M}_{P,mn} &= \left(\mathbf{e}_m\mathbf{e}_m^T\right) \otimes \left(\mathbf{B}_P(y_{m,n}) + \mathbf{B}_P(\overline{y}_m)\right) \\ &\quad - \left(\mathbf{e}_m\mathbf{e}_n^T\right) \otimes \mathbf{B}_P(y_{m,n}) \qquad (4.12)\\ \mathbf{M}_{Q,mn} &= \left(\mathbf{e}_m\mathbf{e}_m^T\right) \otimes \left(\mathbf{B}_Q(y_{m,n}) + \mathbf{B}_Q(\overline{y}_m)\right) \\ &\quad - \left(\mathbf{e}_m\mathbf{e}_n^T\right) \otimes \mathbf{B}_Q(y_{m,n}) \qquad (4.13)\end{aligned}$$

- Voltage measurements $\{|V_n|^2\}$: In this case, we have that:

$$|V_n|^2 = \mathbf{x}^T\mathbf{M}_{V,n}\mathbf{x} \qquad (4.14)$$

with

$$\mathbf{M}_{V,n} = \mathbf{e}_n\mathbf{e}_m^T \otimes \mathbf{I}_2 \qquad (4.15)$$

with $\mathbf{I}_N$ standing for the $N \times N$ identity matrix.

To summarize, exploiting the Cartesian representation of the system state, legacy measurements can be conveniently written in the following unified form:

$$z_l = \mathbf{x}^T\mathbf{M}_l\mathbf{x} + w_l \quad l = 1, \ldots, L \qquad (4.16)$$

with z_l standing for each legacy measurement for $l = 1, \ldots, L$, $\mathbf{M}_l \in \mathbb{R}^{2N\times 2N}$ being defined according to corresponding type of measurement, and w_l standing for the noise.

As for the synchronized measurements, which are taken by the so-called PMUs, the following linear model results:

$$\mathbf{r} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon} \qquad (4.17)$$

with $\mathbf{A} \in \mathbb{R}^{2K\times 2N}$ and $\boldsymbol{\varepsilon} \in \mathbb{R}^{2K}$ standing for the measurement matrix and the noise, respectively. In both models, the noise, w_l for $l = 1, \ldots, L$ and $\boldsymbol{\varepsilon} \in \mathbb{R}^{2K}$, has been assumed to be zero-mean and independent (also between the real and imaginary parts of the measured values from PMUs) according to the literature [25], [76].

4.3 Multi-area State Estimation

Consider now that the power network is divided into M distinct geographical areas, and that each area may be operated by a different utility. Define the state vector of area i by $\mathbf{x}_i = [x_{i,1}, x_{i,2} \ldots, x_{i,2N_i}]^T \in \mathbb{R}^{2N_i}$ with N_i denoting its number of buses. Due to the physical interconnection between areas, bus voltages at the frontier edges, called in the sequel border state variables, are also incorporated into the local state vector of each neighboring area. To be precise, if the pair $x_{i,l}$ and $x_{j,m}$ are border state variables between adjacent areas i and j, the condition $x_{i,l} = x_{j,m}$ must be satisfied. To generalize

this, define $\mathscr{N}_i$ as set of adjacent areas to area i and $\mathscr{S}_{i,j}$ as the set of border variables between areas i and j, with cardinalities given by $|\mathscr{N}_i|$ and $|\mathscr{S}_{i,j}|$. Then, the system state has to satisfy $x_{i,l} = x_{j,m}$ for $j \in \mathscr{N}_i$ and $(l,m) \in \mathscr{S}_{i,j}$. Bearing all the above in mind, the multi-area SE optimization problem is defined as follows:

$$\min_{\{\mathbf{x}_i\}} \quad \sum_{i=1}^{M} \left(\frac{\alpha}{2} \left\| \mathbf{r}_i - \mathbf{A}_i \mathbf{x}_i \right\|_2^2 + \frac{\beta}{2} \sum_{l=1}^{L_i} \left(z_{i,l} - \mathbf{x}_i^T \mathbf{M}_{i,l} \mathbf{x}_i \right)^2 \right) \tag{4.18a}$$

$$\text{s. t.} \quad \mathbf{e}_l^{(i)T} \mathbf{x}_i = \mathbf{e}_m^{(j)T} \mathbf{x}_j, \quad j \in \mathscr{N}_i, (l,m) \in \mathscr{S}_{i,j}, \quad \forall i \tag{4.18b}$$

$$\mathbf{x}_i \in \mathscr{X}_i, \quad \forall i \tag{4.18c}$$

where $\mathbf{r}_i \in \mathbb{R}^{2K_i}$ and $\{z_{i,l}\}_{l=1}^{L_i}$ correspond to the synchronized measurements and legacy measurements at area i, and the definitions of matrices $\mathbf{A}_i \in \mathbb{R}^{2K_i \times 2N_i}$ and $\mathbf{M}_{i,l} \in \mathbb{R}^{2N_i \times 2N_i}$ follow from the rationale of Section 4.2. Column vector $\mathbf{e}_l^{(i)}$ equals 1 at the lth position and zero elsewhere. With some abuse of notation, we have also used the superscript (i) in $\mathbf{e}_l^{(i)}$ to denote that the vector size is that of $\mathbf{x}_i$. Besides, we have included (4.18c) that constraints our solution to lie in the feasible sets $\{\mathscr{X}_i\}$ which are assumed to be convex and compact. Finally, we have introduced the positive scalars $\alpha \in [0,1]$ and $\beta \in [0,1]$ where $\beta = 1 - \alpha$ to weight the contribution of the two type of measurements.

4.4 Successive Convex Approximation Approach

Note that the optimization problem in (4.18) is non-convex (due to the fourth-order terms) and, thus, global optimality of the solution cannot be guaranteed, in general. Hence, with the aim of finding a local minimum, we resort to the novel iterative optimization algorithm proposed in [71]. To that end, first we rewrite the cost function in our problem as $U(\mathbf{x}) = \sum_{i=1}^{M} U_i(\mathbf{x}_i)$ with

$$U_i(\mathbf{x}_i) = \frac{\alpha}{2} \left\| \mathbf{r}_i - \mathbf{A}_i \mathbf{x}_i \right\|_2^2 + \frac{\beta}{2} \sum_{l=1}^{L_i} \left(z_{i,l} - \mathbf{x}_i^T \mathbf{M}_{i,l} \mathbf{x}_i \right)^2 \tag{4.19}$$

According to [71], we need to approximate $U_i(\mathbf{x}_i)$ by a strongly-convex function. To do so, we simply linearize the (fourth-order) non-convex terms around a feasible point $\tilde{\mathbf{x}}_i$ and add a proximal term, that is,

$$\begin{aligned} \tilde{U}_i(\mathbf{x}_i, \tilde{\mathbf{x}}_i) = \frac{\alpha}{2} \left\| \mathbf{r}_i - \mathbf{A}_i \mathbf{x}_i \right\|_2^2 + \frac{\beta}{2} \Bigg(& \mathbf{g}_{\tilde{\mathbf{x}}_i}^T (\mathbf{x}_i - \tilde{\mathbf{x}}_i) \\ & + \frac{\rho}{2} (\mathbf{x}_i - \tilde{\mathbf{x}}_i)^T \tilde{\mathbf{H}}_i(\tilde{\mathbf{x}}_i)(\mathbf{x}_i - \tilde{\mathbf{x}}_i) \Bigg) \end{aligned} \tag{4.20}$$

where $\rho > 0$, and the gradient reads

$$\mathbf{g}_{\tilde{\mathbf{x}}_i} = -2 \sum_{l=1}^{L_i} \left(z_{i,l} - \tilde{\mathbf{x}}_i^T \mathbf{M}_{i,l} \tilde{\mathbf{x}}_i \right) (\mathbf{M}_{i,l} + \mathbf{M}_{i,l}^T) \tilde{\mathbf{x}}_i \tag{4.21}$$

with $\tilde{\mathbf{H}}_i(\tilde{\mathbf{x}}_i)$ standing for a positive definite matrix (or semidefinite positive if $\mathbf{A}_i^T\mathbf{A}_i$ is full rank[2]). One could simply let $\tilde{\mathbf{H}}_i(\tilde{\mathbf{x}}_i)$ be the identity matrix, i.e., $\tilde{\mathbf{H}}_i(\tilde{\mathbf{x}}_i) = \mathbf{I}$. However, since this does not exploit any second order information of (4.19), we propose to *approximate* the Hessian of the non-convex term in (4.19) by a positive definite matrix (referred to in the sequel as the A-Hessian). First, note that the Hessian of (4.19) reads:

$$\begin{aligned}\mathbf{H}_i(\tilde{\mathbf{x}}_i) = \beta \sum_{l=1}^{L_i} \big[&(\mathbf{M}_{i,l} + \mathbf{M}_{i,l}^T)\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T(\mathbf{M}_{i,l} + \mathbf{M}_{i,l}^T) \\ &-(\mathbf{M}_{i,l} + \mathbf{M}_{i,l}^T)\left(z_{i,l} - \tilde{\mathbf{x}}_i^T\mathbf{M}_{i,l}\tilde{\mathbf{x}}_i\right)\big]\end{aligned} \tag{4.22}$$

We observe that the first part in (4.22) is the sum of rank-1 matrices. Hence, by discarding the second term in the summation and adding a properly scaled identity matrix, we obtain our candidate A-Hessian matrix, namely:

$$\tilde{\mathbf{H}}_i(\tilde{\mathbf{x}}_i) = \beta \sum_{l=1}^{L_i} (\mathbf{M}_{i,l} + \mathbf{M}_{i,l}^T)\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T(\mathbf{M}_{i,l} + \mathbf{M}_{i,l}^T) + \delta\mathbf{I} \tag{4.23}$$

where δ is a small positive scalar and $\mathbf{I} \in \mathbb{R}^{2N\times 2N}$ stands for the identity matrix. In doing so, we preserve some second order information of (4.19) with an affordable computational complexity (see alternative strategy in Section 4.7).

Hence, the approximate optimization problem at iteration ν given the current iterate, $\mathbf{x}_i^{(\nu)}$, is as follows:

$$\begin{aligned}\min_{\{\mathbf{x}_i\}} \quad & \sum_{i=1}^{M} \tilde{U}_i(\mathbf{x}_i, \mathbf{x}_i^{(\nu)}) \\ \text{s.t.} \quad & \mathbf{e}_l^{(i)T}\mathbf{x}_i = \mathbf{e}_m^{(j)T}\mathbf{x}_j, \quad j \in \mathscr{N}_i, (l,m) \in \mathscr{S}_{i,j}, \forall i \\ & \mathbf{x}_i \in \mathscr{X}_i, \quad \forall i\end{aligned} \tag{4.24}$$

Let $\{\hat{\mathbf{x}}_i^{(\nu)}\}$ denote the unique optimal solution to (4.24). Bearing all the above in mind, the successive approximation method for state estimation (SCA-SE) is summarized in Algorithm 4.1. In short, the algorithm finds the solution of the convex approximation problem in (4.24), $\{\hat{\mathbf{x}}_i^{(\nu)}\}$ (given the last point $\mathbf{x}_i^{(\nu)}$), and computes the new iterate as a convex combination of $\{\hat{\mathbf{x}}_i^{(\nu)}\}$ and $\{\mathbf{x}_i^{(\nu)}\}$ until convergence.

Theorem 4.1. *Algorithm 4.1 converges to a stationary solution of* (4.18).

Proof. The optimization problem in (4.18) satisfies the following requirements:

1. $\sum_{i=1}^{M} U_i(\mathbf{x}_i)$ is continuously differentiable in $\{\mathscr{X}_i\}$: this clearly follows from the definition of each $U_i(\mathbf{x}_i)$.

[2]Which is not realistic since we should ensure full observability with PMU measurements only.

Algorithm 4.1 Successive Convex Approximation for SCA-SE

1: Initialize $\nu = 0$, $\gamma^{(\nu)} \in (0, 1]$, $\{\mathbf{x}_i^{(0)} \in \mathscr{X}_i\}$
2: Compute $\{\hat{\mathbf{x}}_i^{(\nu)}\}$ by solving (4.24) in a distributed (local) fashion (see Section 4.5)
3: $\{\mathbf{x}_i^{(\nu+1)} \leftarrow \mathbf{x}_i^{(\nu)} + \gamma^{(\nu)}(\hat{\mathbf{x}}_i^{(\nu)} - \mathbf{x}_i^{(\nu)})\}$
4: Compute $\tilde{\mathbf{H}}_i(\mathbf{x}_i^{(\nu+1)})$ as in (4.23) for each area.
5: If $\mathbf{x}_i^{(\nu+1)}$ is a stationary solution STOP
6: $\nu \leftarrow \nu + 1$ and go to step 2

2. $\nabla_{\mathbf{x}_i} U_i$ is Lipschitz continuous $\forall \mathbf{x}_i \in \mathscr{X}_i$. To show this, note that

$$\begin{aligned}\nabla_{\mathbf{x}_i} U_i = &-\alpha \mathbf{A}_i^T (\mathbf{r}_i - \mathbf{A}_i \mathbf{x}_i) \\ &- \beta \sum_{l=1}^{L_i} \left(z_{i,l} - \mathbf{x}_i^T \mathbf{M}_{i,l} \mathbf{x}_i\right) (\mathbf{M}_{i,l} + \mathbf{M}_{i,l}^T) \mathbf{x}_i.\end{aligned}$$

   Since this is a continuously differentiable function on $\mathscr{X}_i$, we have that $\nabla_{\mathbf{x}_i} U_i$ is Lipschitz continuous.

3. U_i is coercive on the feasible set. Since, by definition, $\mathscr{X}_i$ is compact each U_i is coercive and, thus, the sum of coercive functions is coercive.

Hence, from [71], convergence to a stationary solution is guaranteed. ■

4.5 Distributed Implementation via ADMM

In this section, we provide a distributed implementation of Step 2 in Algorithm 4.1. Due to its superior performance, we propose a decentralized solution based on the well-known ADMM[3] [72]. To that aim, first we decompose the optimization problem by introducing auxiliary consensus variables, $c_{ij,(l,m)}$, on the border state variables:

$$\min_{\{\mathbf{x}_i, c_{ij,(l,m)}\}} \quad \sum_{i=1}^{M} \tilde{U}_i(\mathbf{x}_i, \tilde{\mathbf{x}}_i) \tag{4.25a}$$

$$\text{s.t.} \quad \mathbf{e}_l^{(i)T} \mathbf{x}_i = c_{ij,(l,m)},\ j \in \mathscr{N}_i, (l,m) \in \mathscr{S}_{i,j}, \forall i \tag{4.25b}$$

$$\mathbf{x}_i \in \mathscr{X}_i, \tag{4.25c}$$

where, with some abuse of notation, we let $c_{ij,(l,m)}$ and $c_{ji,(m,l)}$ denote the same variable. Following the rationale of ADMM, we obtain the following sequence of primal updates:

$$\begin{aligned}\mathbf{x}_i^{(t+1)} = \underset{\mathbf{x}_i \in \mathscr{X}_i}{\operatorname{argmin}} \tilde{U}_i(\mathbf{x}_i, \tilde{\mathbf{x}}_i) + \sum_{j \in \mathscr{N}_i} \sum_{(l,m) \in \mathscr{S}_{i,j}} \Big[&\lambda_{i,j,(l,m)}^{(t)} \\ \cdot (\mathbf{e}_l^{(i)T} \mathbf{x}_i - c_{ij,(l,m)}) + \frac{\theta}{2} &(\mathbf{e}_l^{(i)T} \mathbf{x}_i - c_{ij,(l,m)})^2 \Big]\end{aligned} \tag{4.26}$$

[3] Note that other methods such as dual decomposition are also available.

$$\mathbf{x}_i^{(t+1)} = \left(\alpha \mathbf{A}_i^T \mathbf{A}_i + \theta \sum_{j \in \mathscr{N}_i} \sum_{(l,m) \in \mathscr{S}_{i,j}} \mathbf{e}_l^{(i)} \mathbf{e}_l^{(i)T} + (\beta\rho/2) \tilde{\mathbf{H}}_i \right)^{-1} \cdot \left((\beta\rho/2) \tilde{\mathbf{H}}_i \tilde{\mathbf{x}}_i - \beta \mathbf{g}_{\tilde{\mathbf{x}}_i} + \alpha \mathbf{A}_i^T \mathbf{r}_i + \sum_{j \in \mathscr{N}_i} \sum_{(l,m) \in \mathscr{S}_{i,j}} (\theta c_{ij,(l,m)}^{(t)} - \lambda_{i,j,(l,m)}^{(t)}) \mathbf{e}_l^{(i)} \right) \tag{4.29}$$

Algorithm 4.2 ADMM for computing Step 2 in Algorithm 4.1

for $t = 1, \ldots, t_{\mathsf{max}}$ **do**
  for $i = 1, \ldots, M$ **do**
    Compute $\{\mathbf{x}_i^{(t+1)}\}$ from (4.26)
    Broadcast your border variables to your neighbors
    Compute $\{c_{ij,(l,m)}^{(t+1)}\}$ from (4.27)
    Update the Lagrangian multipliers (4.28)
    Broadcast the Lagrangian multipliers associated to the border variables to your neighbors
  end for
  $t \leftarrow t + 1$
end for

$$c_{ij,(l,m)}^{(t+1)} = \frac{\mathbf{e}_l^{(i)T} \mathbf{x}_i^{(t+1)} + \mathbf{e}_m^{(j)T} \mathbf{x}_j^{(t+1)}}{2} + \frac{\lambda_{i,j,(l,m)}^{(t)} + \lambda_{j,i,(m,l)}^{(t)}}{2\theta} \tag{4.27}$$

with $\theta > 0$ standing for weight associated to the augmented constraint in the ADMM formulation [72]. Further, we have denoted by $\{\lambda_{j,i,(l,m)}\}$ the Lagrangian multipliers associated to the constraints of (4.25b). These dual variables are updated as follows:

$$\lambda_{i,j,(l,m)}^{(t+1)} = \lambda_{i,j,(l,m)}^{(t)} + \theta(\mathbf{e}_l^{(i)T} \mathbf{x}_i^{(t+1)} - c_{ij,(l,m)}^{(t+1)}) \tag{4.28}$$

$\forall i,\ j \in \mathscr{N}_i$ and $(l,m) \in \mathscr{S}_{i,j}$. When the constraint in $\mathbf{x}_i \in \mathscr{X}_i$ is trivially satisfied, as might occur when the set $\mathscr{X}_i$ is sufficiently large, the primal update in (4.26) accepts a closed form solution given in (4.29). This observation can be used to reduce the computational complexity required to obtain $\mathbf{x}_i^{(t+1)}$. This distributed method is summarized in Algorithm 4.2.

4.6 Robust State Estimation (RSCA-SE)

Here, we consider a scenario where few measurements may be subject to gross error (bad data) due to e.g., malfunctioning, malicious attacks, or communication failures. In order

to perform bad data detection, we reformulate the *centralized* SE problem in a LASSO framework:

$$\min_{\{\mathbf{x},\mathbf{o},\boldsymbol{\xi}\}} \left(\frac{\alpha}{2} \left\| \mathbf{r} - \mathbf{A}\mathbf{x} - \mathbf{o} \right\|_2^2 + \lambda_1 \left\| \mathbf{o} \right\|_1 \right. \\ \left. + \frac{\beta}{2} \sum_{l=1}^{L} \left(z_l - \mathbf{x}^T \mathbf{M}_l \mathbf{x} - \mathbf{e}_l^T \boldsymbol{\xi} \right)^2 + \lambda_2 \left\| \boldsymbol{\xi} \right\|_1 \right) \tag{4.30}$$

where vectors $\mathbf{o} \in \mathbb{R}^K$ and $\boldsymbol{\xi} \in \mathbb{R}^L$ account for the presence of bad data in the PMU and legacy measurements, respectively (i.e., with nonzero entries for corrupted measurements). Consequently, the quadratic and fourth-order terms in the objective function now account for the residual errors, whereas the norm-1 (regularization) terms $\|\mathbf{o}\|_1$ and $\|\boldsymbol{\xi}\|_1$ attempt to enforce sparsity in the solution. The λ_1 and λ_2 weights (along with α and β) control the model-fit vs. sparsity/robustness trade-off. And, further, they define an upper bound on the absolute value of the residuals (see Section 4.6.1 ahead).

To circumvent the non-convexity of the composite problem (4.30), we resort to the SCA-based non-convex optimization framework of [86]. Specifically, at iteration ν the algorithm attempts to solve the strongly-convex problem:

$$\left\{ \hat{\mathbf{x}}^{(\nu)}, \hat{\mathbf{o}}^{(\nu)}, \hat{\boldsymbol{\xi}}^{(\nu)} \right\} = \underset{\{\mathbf{x} \in \mathscr{X}\}}{\operatorname{argmin}} \; \tilde{U}(\mathbf{x}, \tilde{\mathbf{x}}^{(\nu)}, \mathbf{o}, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}^{(\nu)}). \tag{4.31}$$

In the above expression, the *approximate* objective function $\tilde{U}(\mathbf{x}, \tilde{\mathbf{x}}^{(\nu)}, \mathbf{o}, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}^{(\nu)})$ can be obtained by linearizing the non-convex term of the objective function in (4.30) at the solution for the current iterate, $\tilde{\mathbf{x}}^{(\nu)}, \tilde{\boldsymbol{\xi}}^{(\nu)}$, that is (see derivation in Appendix B):

$$\tilde{U}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{o}, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}},) = \frac{\alpha}{2} \left\| \mathbf{r} - \mathbf{A}\mathbf{x} - \mathbf{o} \right\|_2^2 + \lambda \left\| \mathbf{o} \right\|_1 \\ + \frac{\beta}{2} \left\{ \mathbf{g}_{\tilde{\mathbf{x}}}^T (\mathbf{x} - \tilde{\mathbf{x}}) + \mathbf{g}_{\tilde{\boldsymbol{\xi}}}^T \left(\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}} \right) \right. \\ \left. + \frac{\rho}{2} \left\| \mathbf{x} - \tilde{\mathbf{x}} \right\|_2^2 + \frac{\rho}{2} \left\| \boldsymbol{\xi} - \tilde{\boldsymbol{\xi}} \right\|_2^2 \right\} + \lambda \left\| \boldsymbol{\xi} \right\|_1 \tag{4.32}$$

where

$$\mathbf{g}_{\tilde{\mathbf{x}}} = -2 \sum_{l=1}^{L} \left(z_l - \tilde{\mathbf{x}}^T \mathbf{M}_l \tilde{\mathbf{x}} - \mathbf{e}_l^T \boldsymbol{\xi} \right) (\mathbf{M}_l + \mathbf{M}_l^T) \tilde{\mathbf{x}} \tag{4.33}$$

and

$$\mathbf{g}_{\tilde{\boldsymbol{\xi}}} = -2 \sum_{l=1}^{L} \mathbf{e}_l \left(z_l - \tilde{\mathbf{x}}^T \mathbf{M}_l \tilde{\mathbf{x}} - \mathbf{e}_l^T \tilde{\boldsymbol{\xi}} \right) \tag{4.34}$$

denote the gradients with respect to $\tilde{\mathbf{x}}$ and $\tilde{\boldsymbol{\xi}}$, respectively. The approximate problem (4.31) admits a closed form solution which is given by (4.41). The original problem can thus be efficiently solved in an iterative manner by letting

$$\hat{\mathbf{x}}^{(\nu+1)} = \underset{\mathbf{x} \in \mathscr{X}}{\operatorname{argmin}} \, \tilde{U}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{o}, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}), \tag{4.35}$$

followed by an update of the sparse vectors

$$\hat{\mathbf{o}}^{(\nu+1)} = (\mathbf{r} - \mathbf{A}^T\hat{\mathbf{x}}^{(\nu+1)})^+_{\lambda/\alpha} \tag{4.36}$$

and

$$\hat{\boldsymbol{\xi}}^{(\nu+1)} = \left[\tilde{\boldsymbol{\xi}} - (1/\rho)\mathbf{g}_{\tilde{\boldsymbol{\xi}}}\right]^+_{2\lambda/\rho\beta} \tag{4.37}$$

where $(\chi)^+_\lambda$ denotes the element-wise thresholding operator, namely,

$$S_\lambda(\chi) = \begin{cases} \chi + \lambda & \text{if } \chi < -\lambda; \\ 0 & \text{if } |\chi| \leq \lambda; \\ \chi - \lambda & \text{if } \chi - \lambda. \end{cases} \tag{4.38}$$

Hereinafter, the proposed scheme, which is summarized in Algorithm 4.3, will be referred to as Robust Successive Convex Approximation for State Estimation (RSCA-SE). Convergence to a stationary solution of the original problem can be guaranteed as long as the sequence of step sizes $\gamma^{(\nu)}$ is carefully chosen (see [71], [86] for details). The main features of RSCA-SE is that, it provides jointly state estimation and outliers suppression and its numerical behavior is mainly dominated from the weights λ_1, λ_2.

Algorithm 4.3 Robust SCA for State Estimation (RSCA-SE)

1: Initialize $\nu = 0$, $\gamma^{(\nu)} \in (0, 1]$, $\{\tilde{\mathbf{x}}^{(\nu)}, \tilde{\boldsymbol{\xi}}^{(\nu)}, \tilde{\mathbf{o}}^{(\nu)} \in \mathscr{X}\}$
2: Compute $\{\hat{\mathbf{x}}^{(\nu)}, \hat{\boldsymbol{\xi}}^{(\nu)}, \hat{\mathbf{o}}^{(\nu)}\}$ by solving (4.31).
3: $\{\tilde{\mathbf{x}}^{(\nu+1)} \leftarrow \tilde{\mathbf{x}}^{(\nu)} + \gamma^{(\nu)}(\hat{\mathbf{x}}^{(\nu)} - \tilde{\mathbf{x}}^{(\nu)})\}$
4: $\{\tilde{\boldsymbol{\xi}}^{(\nu+1)} \leftarrow \tilde{\boldsymbol{\xi}}^{(\nu)} + \gamma^{(\nu)}(\hat{\boldsymbol{\xi}}^{(\nu)} - \tilde{\boldsymbol{\xi}}^{(\nu)})\}$
5: $\{\tilde{\mathbf{o}}^{(\nu+1)} \leftarrow \hat{\mathbf{o}}^{(\nu)}\}$
6: If $\{\tilde{\mathbf{x}}^{(\nu+1)}, \tilde{\mathbf{o}}^{(\nu+1)}, \tilde{\boldsymbol{\xi}}^{(\nu+1)}\}$ is a stationary solution Stop.
7: $\nu \leftarrow \nu + 1$ and go to step 2.

4.6.1 Upper bound on residual errors

The following proposition establishes an upper bound on the residual errors associated to linear/non-linear measurements:

Proposition 4.2. *Let $\check{\mathbf{x}}$, $\check{\mathbf{o}}$, $\check{\boldsymbol{\xi}}$ be the stationary solution given by Algorithm 4.3. Then, the residuals (after data cleansing) are bounded as follows:*

$$\left|\mathbf{r} - \mathbf{A}\check{\mathbf{x}} - \check{\mathbf{o}}\right|_\infty \leq \lambda_1/\alpha \tag{4.39}$$

$$\max_l \left|z_l - \check{\mathbf{x}}^T\mathbf{M}_l\check{\mathbf{x}} - \check{\boldsymbol{\xi}}\right| \leq \lambda_2/\beta \tag{4.40}$$

with $|\cdot|_\infty$ denoting the ℓ_∞ norm. In addition, if $[\check{\mathbf{o}}]_k \neq 0$ the residual associated to the k-th PMU measurement satisfies $\left|[\mathbf{r} - \mathbf{A}\check{\mathbf{x}} - \check{\mathbf{o}}]_k\right| = \lambda_1/\alpha$. Similarly, if $[\check{\boldsymbol{\xi}}]_l \neq 0$ the residual associated to the l-th legacy measurement satisfies $\left|z_l - \check{\mathbf{x}}^T\mathbf{M}_l\check{\mathbf{x}} - \check{\xi}_l\right| = \lambda_2/\beta$.

$$\mathbf{x}^{(t+1)} = \left(\alpha \mathbf{A}^T \mathbf{A} + \theta \sum_{j \in \mathcal{N}_i} \sum_{(l,m) \in \mathcal{S}_{i,j}} \mathbf{e}_l^{(i)} \mathbf{e}_l^{(i)T} + (\beta\rho/2)\mathbf{I} \right)^{-1} \cdot \left((\beta\rho/2)\tilde{\mathbf{x}} - \beta \mathbf{g}_{\tilde{\mathbf{x}}} + \alpha \mathbf{A}^T (\mathbf{r} - \mathbf{o}) + \sum_{j \in \mathcal{N}_i} \sum_{(l,m) \in \mathcal{S}_{i,j}} (\theta c_{ij,(l,m)}^{(t)} - \lambda_{i,j,(l,m)}^{(t)}) \mathbf{e}_l^{(i)} \right) \tag{4.41}$$

Proof. Refer to Appendix A. ■

In other words, the absolute value of the residual error associated to an arbitrary PMU (legacy) measurement is upper bounded by λ_1/α (λ_2/β). Further, the upper bound is attained for those measurements with bad data. Interestingly, from Proposition 4.2, Algorithm 4.3 can be regarded as an instantiation of the Huber estimator. Vectors $\mathbf{o}$ and $\boldsymbol{\xi}$ control the amount of residual error to be accounted in a ℓ_1-norm sense, whereas the amount of squared losses is upper bounded by λ_1 and λ_2 (model fit vs. robustness/sparsity trade-off).

4.7 Numerical Results

The proposed state estimation schemes have been numerically assessed by means of computer simulations, for different IEEE test cases and measurements sets. As in [29], power flow and injection measurements, voltage magnitude measurements and PMU measurements have been corrupted with independent and zero-mean Gaussian noise of standard deviation $\sigma = \{0.02, 0.02, 0.01, 0.002\}$ p.u., respectively. We have assumed that when a PMU is placed at a given node, it provides the voltage phasor of this bus and the related current phasors from the incident branches. In all test cases, observability is ensured according to [88, 89]. That means, the number of available measurements (legacy and PMUs) suffices in order to provide a full-rank invertible matrix in (4.29), which in turn means that the closed form solution of the SCA-SE is solvable. Where relevant, the Semi-Definite Relaxation-based SE method (SDR-SE) proposed in [29], and the Normal Equations-based method (NE-SE) from [81] have been used as benchmarks. In the first case, the SE problem is posed as a semidefinite programming (SDP) optimization model. In order to deal with the non-convexity of the SDP problem the authors exploit the semidefinite relaxation (SDR) technique, which amounts to drop the rank-1 constraint. However, the SDR-SE approach guarantees the near-optimality of the solution under a number of assumptions with respect to the system topology and the available measurement set. The second benchmark, NE-SE, is based on the classical WLS SE model. The latter is expanded into its Taylor series by computing the function's derivatives at

a single point and, after neglecting the higher order terms, the computation of the first order optimality condition leads into the normal equations. Again, this approach is not able to guarantee the convergence of the algorithm to a near-optimal solution, presenting sensitivity to the initialization point and the measurement set. Unless otherwise stated, computer simulation results are averaged over 500 realizations where the voltage magnitudes were Gaussian distributed, $V_m \sim \mathcal{N}(1, 0.01)$ p.u. and; angles were uniformly distributed over $[-0.5\pi, 0.5\pi]$ rad [29]. In all simulations, the SE is initialized[4] with flat start, i.e., $\hat{\mathbf{x}}^0 = [\mathbf{1^T 0^T}]^\mathbf{T}$ where $\mathbf{1} \in \mathbb{R}^\mathbf{N}$ is a vector with all ones and $\mathbf{0} \in \mathbb{R}^\mathbf{N}$ with all zeros. For power flow analysis we have used Matpower [90].

4.7.1 Centralized scenario

First, we present some results for a *centralized* scenario and the IEEE 30-bus test case, Fig. 4.2. The measurement set consists of 41 pairs of (re-)active power flows, 30 voltage magnitude measurements (placed according to the first test case in [29]) and no PMUs. To mention that the A-Hessian matrix has not been considered in this computer simulation. For the measurement noise and the number of realizations applies what it has been described above. Fig. 4.3 depicts the absolute estimation error in the voltage magnitudes and angles at all buses. Clearly, the proposed SCA-SE approach outperforms SDR-SE in terms of accuracy in the case of voltage magnitudes (top graph). At the same time, the accuracy of the SCA-SE and SDR-SE schemes is comparable for the angle estimates. On the contrary, the estimation error of NE-SE for both magnitude and angle estimates attains higher values.

Next, in Fig. 4.4, we analyze the impact of PMU measurements[5] in the resulting state estimation accuracy. To that aim, up to 4 PMUs are progressively introduced at a number of pre-defined locations, specifically at buses 10, 12, 27 and 15. The set of legacy measurements is the same as in the previous computer simulation. Results are given in terms of normalized estimation error, that is, $\frac{1}{2N}\|\mathbf{x} - \hat{\mathbf{x}}\|_2$, with $\mathbf{x}$ and $\hat{\mathbf{x}}$. The figure reveals, that, in many realizations NE-SE fails to converge (outliers in the top boxplot). On the contrary, the SDR-SE and SCA-SE schemes perform well in all realizations and cases. This follows from the fact that NE-SE convergence is largely affected by the initialization point (i.e., $\mathbf{x}^0$). For instance, in the case without PMUs, the mean value of the average error for NE-SE is 1.9×10^{-3}. In contrast, the same value for SCA-SE and SDR-SE is 7.8×10^{-4} and 9×10^{-4}, respectively. We also observe that, unsurprisingly, accuracy increases for the increasing number of PMUs; and, further, the accuracy of the

[4]To note here that the algorithm on any random time instant t_2 can be initialized with the estimated state from the previous estimation i.e., t_1, as in steady state conditions, $\|\mathbf{x}_{t_2} - \mathbf{x}_{t_1}\| \leq \|\mathbf{x}_{t_2} - \mathbf{x}_0\|$. This approach (hot start) can lead the algorithm to improved convergence rate.

[5]Where PMU and legacy measurements coexist, for the positive scalars α and β we have used the rule, $\beta = 1 - \alpha$ with $\alpha = 0.99$, in order to weight the contribution of each type of meters. However other rules can be followed.

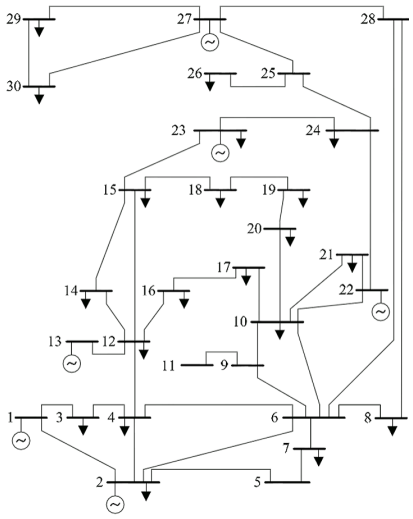

Figure 4.2: IEEE 30-bus test case [6].

proposed SCA-SE algorithm is very similar to that of SDR-SE. However the performance of SDR-SE degrades when the number of voltage magnitude measurements decreases. This drawback may arise in the SDR-SE method if voltage magnitude measurements are not available in all buses and the system does not have a tree topology, e.g. transmission system (see Proposition 1 in [29]).

To show this, in Fig. 4.5 is presented a scenario where the number of voltage magnitude measurements has been reduced from 30 to 15. We focus on a specific realization of the system state and then average the results over 100 realizations of the observation noise. In contrast with the previous scenario, PMUs are placed progressively at the buses 2, 10, 18, 27, 12, 28, 25, 6, 1 and 15. SCA-SE is able to converge efficiently to a stationary point. This, does not apply for SDR-SE, which has not converged (in all realizations). Moreover, as it was expected, for the proposed SCA-SE we observe a decreasing error when the number of PMU measurements increases. Important to mention that, the introduction of the approximate Hessian in the SCA-SE approach (bottom graph) has a dramatic impact in terms convergence rate: the total number of iterations needed to find a stationary solution decreases one order of magnitude (from few hundreds to 20-25). Hence, the exploitation of second order information information (embedded into the A-Hessian), even if approximate, is of an utmost importance. This holds in particular when there are few PMUs since, in this case, such 'high quality' measurements cannot be exploited to speed up convergence.

Complementarily, we investigate an alternative to the computation of the A-Hessian matrix in (4.23) which is based on the eigenvalue decomposition. Rather than just dropping the second terms of the Hessian (see equation (4.22)), we construct a positive-definite matrix (the EV-Hessian) by discarding the eigenvectors associated to its negative eigen-

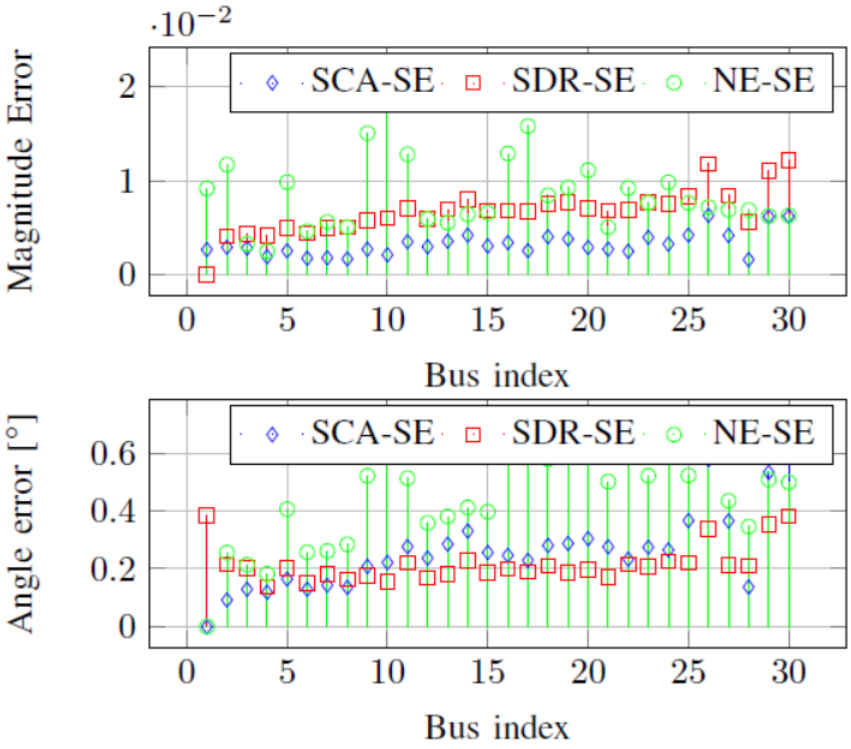

Figure 4.3: Absolute error for the magnitude (top) and angle (bottom) estimates in each bus, for SCA-SE, SDR-SE and NE-SE algorithms. IEEE 30-bus test case. The parameters has been set as $\gamma = 0.01$ and $\rho = 10$.

values (and, again, add a properly scaled identity matrix). The performance exhibited by SCA-SE with both the A- and EV-Hessian matrices is shown in Fig. 4.6. Results are averaged over 100 realizations for the IEEE 30-bus test system, where $V_m \sim \mathcal{N}(1, 0.0025)$ p.u. and $\theta_m \sim [-0.1\pi, 0.1\pi]$ for $m \in \{1, \ldots, N\}$. Further, the measurement set consists of 41 pairs of power flow measurements, 15 voltage magnitudes and no PMUs. Interestingly, the normalized error after convergence is identical in both cases. However, the number of iterations needed by the A-Hessian approach halves that of the EV-Hessian. This is due to the higher complexity of the EV-Hessian. Hence, the additional computational burden that the computation of the eigenvalue decomposition entails does not pay off, when it comes to preserving curvature information. The low-complexity A-Hessian matrix that we initially proposed is preferable.

4.7.2 Decentralized (multi-area) scenario

Figures 4.9 and 4.10 illustrate the convergence behavior of the decentralized version of SCA-SE in two scenarios for a multi-area setting of the IEEE 57-bus test case (partitioned in 4 areas, Fig. 4.7). In the first graph, Fig. 4.9, we present the convergence behavior

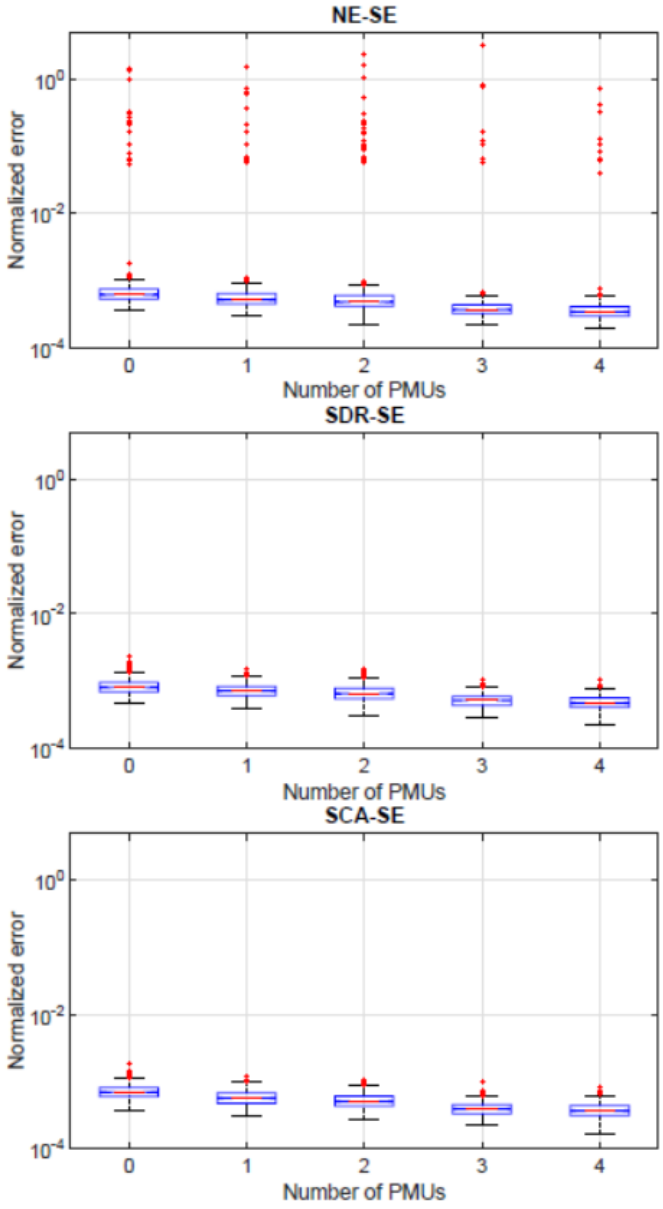

Figure 4.4: Normalized error vs. number of PMUs, for SCA-SE, SDR-SE and NE-SE algorithms. Parameter set: $\gamma = 0.01$, $\rho = 10$, $\alpha = 0.99$, $\beta = 0.01$.

of the algorithm in a case where the system is observable via only PMUs. Specifically, the measurement set consists of 17 PMUs according to the meter configuration presented in [91]. The graph illustrates that all areas converge in less than 20 iterations. The second scenario considers a case with a decreased number of PMUs aided by legacy measurements, that is, 80 pairs of power flows, 57 voltage magnitudes and 10 PMUs. Interestingly, Fig. 4.10 shows that the distributed SCA framework aided by the A-Hessian approaches the convergence rate of the idealistic convex case, i.e., 4.9. Notably,

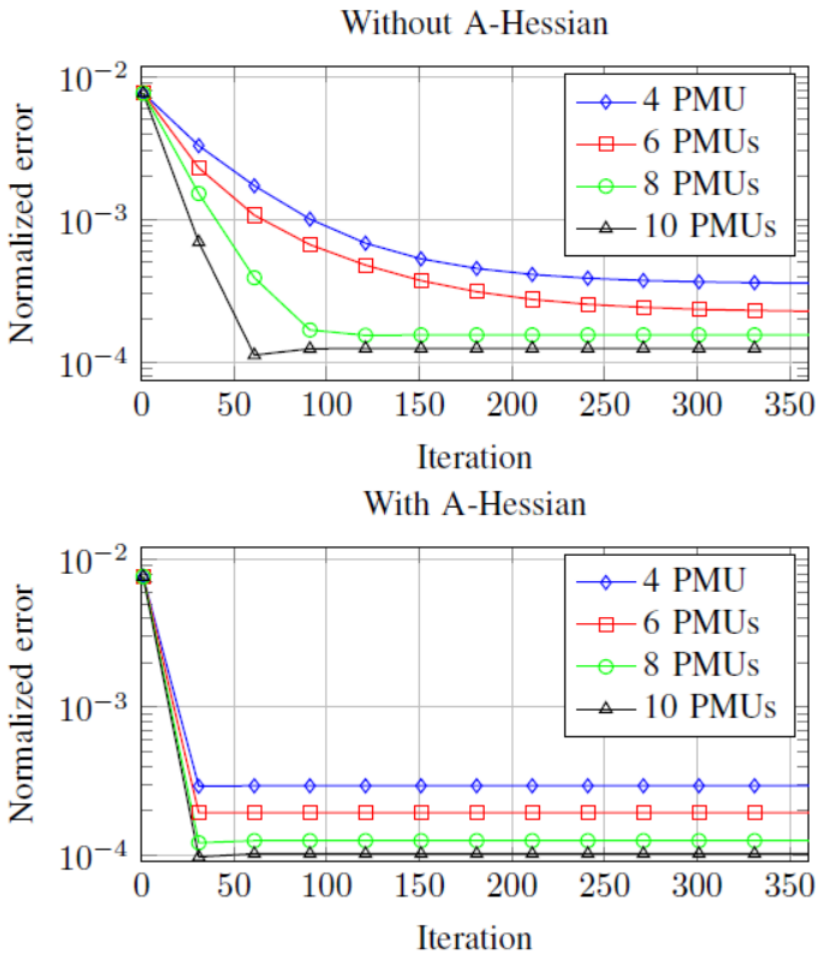

Figure 4.5: Normalized error vs. iteration number, with (bottom) and without (top) the approximate Hessian. IEEE 30-bus system. Parameter set: $\gamma = 0.01$, $\rho = 10$, $\alpha = 0.99$, $\beta = 0.01$ and $\delta = 10^{-4}$.

as the same graph depicts, the estimation error of decentralized SCA-SE (blue curve, $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_1^T, \hat{\mathbf{x}}_2^T, \hat{\mathbf{x}}_3^T, \hat{\mathbf{x}}_4^T]^T$), after the convergence to a stationary solution, is identical to that of its *centralized* counterpart (unlike other works [29] no additional penalty from the use of ADMM).

Next, the decentralized version of the SCA-SE algorithm is numerically assessed on a test case which includes a higher number of buses, namely the IEEE 118-bus test system partitioned in 9 areas, Fig. 4.8. By doing so, we aim to illustrate the efficiency of the proposed scheme for larger networks. The measurement set here includes 186 pairs of power flows, 118 voltage magnitudes and 32 PMUs. Again, the phasor meters are positioned according to the configuration proposed in [91]. Fig. 4.11 shows that each area finds a stationary solution within 25-30 iterations (Algorithm 4.1). Again, this is

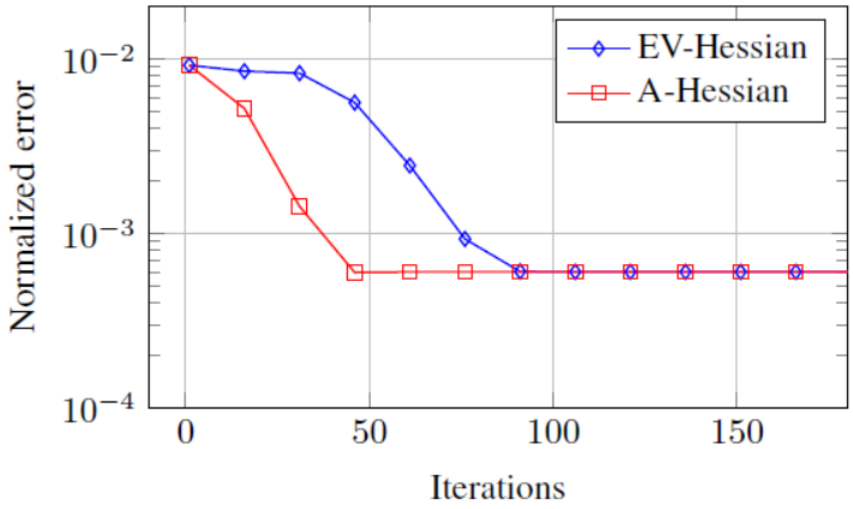

Figure 4.6: Normalized error vs. iteration number for the cases of EV-Hessian and A-Hessian.

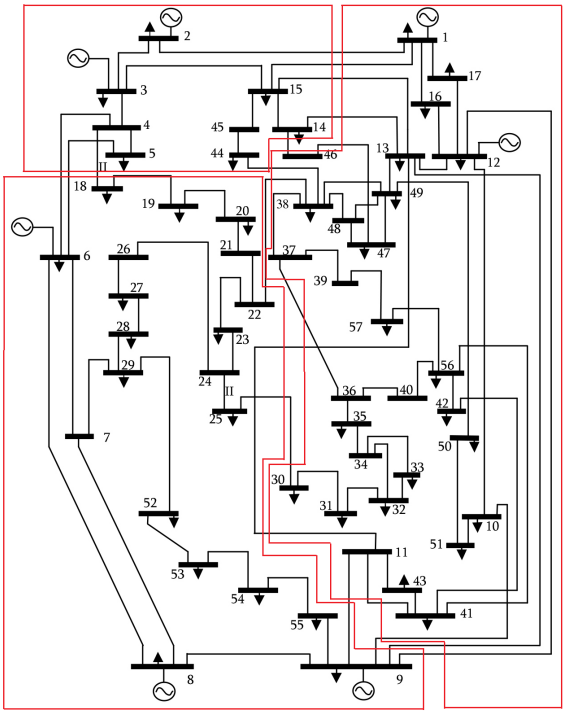

Figure 4.7: IEEE 57-bus test case partitioned in 4 areas [7].

thanks to the use of the A-Hessian matrix that, in such decentralized setting, needs to be computed on a per area basis (locally). For each SCA-SE iteration, 2-3 iterations of the ADMM scheme (Algorithm 4.2) suffice. In summary, for both scenarios, the decentralized

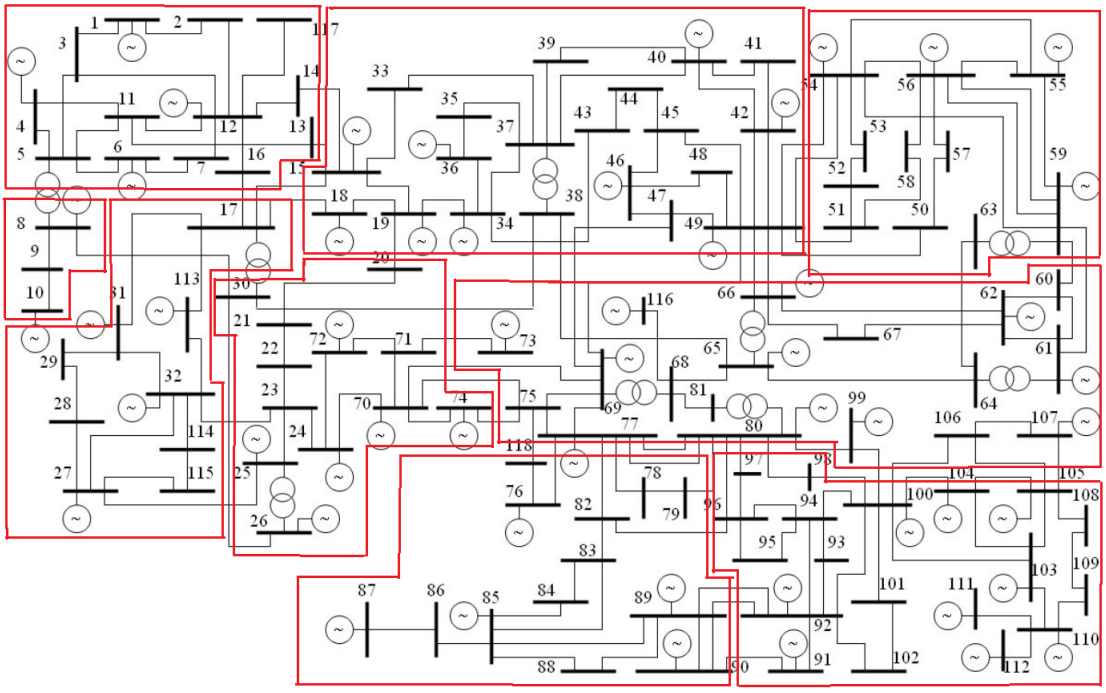

Figure 4.8: IEEE 118-bus test case partitioned in 9 areas [8].

version of the SCA-SE scheme exhibits a remarkable performance.

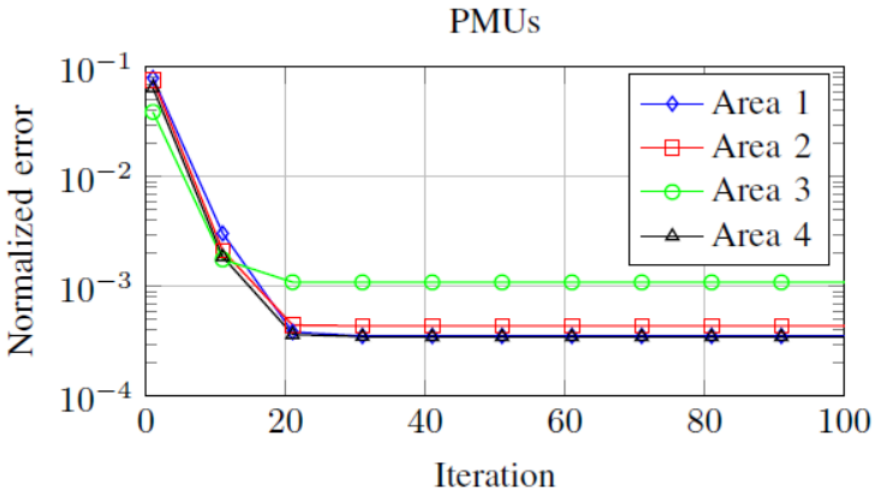

Figure 4.9: Normalized error vs. iteration number for the 4 areas of the IEEE 57-bus test case. Parameter set: $\rho = 30$, $\alpha = 1$ and $\beta = 0$.

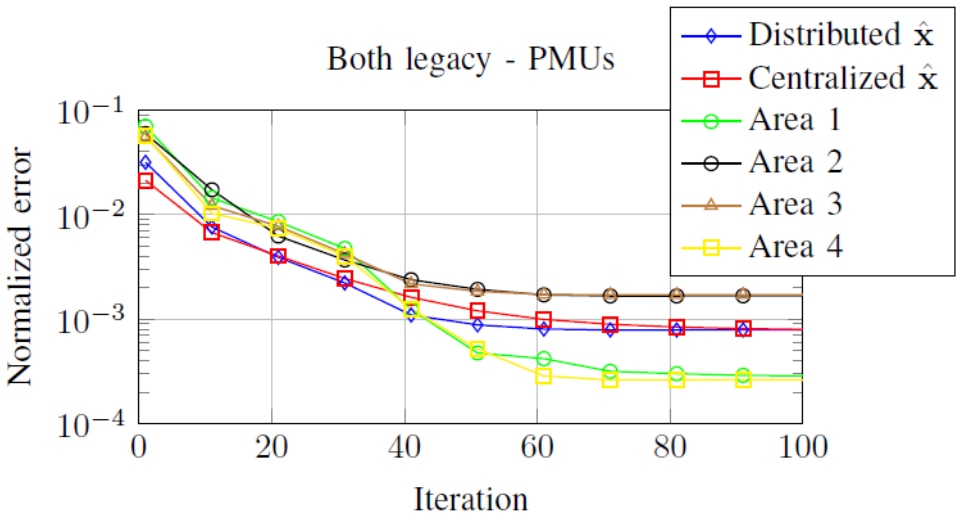

Figure 4.10: Normalized error vs. iteration number for the 4 areas, the aggregated decentralized solution and its centralized counterpart of the IEEE 57-bus test case. Parameter set: $\gamma = 0.01$, $\rho = 30$, $\alpha = 0.99$, $\beta = 0.01$ and $\delta = 10^{-4}$.

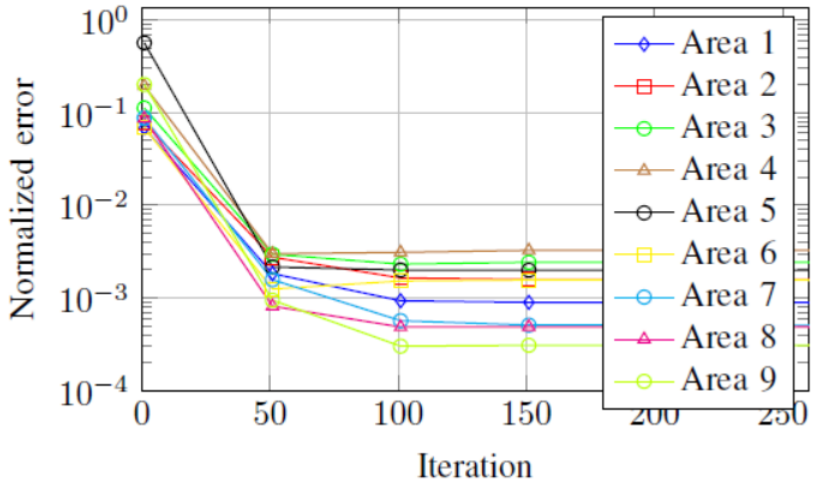

Figure 4.11: Normalized error vs. iteration number for the 9 areas of the IEEE 118-bus test case. Parameter set: $\gamma = 0.01$, $\rho = 30$, $\alpha = 0.99$, $\beta = 0.01$ and $\delta = 10^{-4}$.

4.7.3 Robust approach (RSCA-SE)

Here, we investigate the impact of bad data in the various state estimation schemes. Results are averaged over 100 realizations of system state, where the loads have been assumed $P_m \sim \mathcal{N}(P_m, 0.001)$, $Q_m \sim \mathcal{N}(Q_m, 0.001)$, and the additive noise. Bad data is generated by scaling the correct measurements by a 1.2 factor, as in [25]. As benchmarks, we have considered the *robust* versions of the classical NE-SE estimator (RNE-SE), on

the one hand; and the SDR-SE scheme [29] (RSDR-SE), on the other. RNE-SE leverages on the so-called Largest Normalized Residual Test (LNRT) to detect bad data. Such detection is performed *a posteriori*, that is, based on the analysis of the normalized residuals that the classical NE-SE yields. Measurements with too large residuals are then removed from the entire set and NE-SE is run again. Yet fairly simple, the RNE-SE approach is subject to detection failure. This may happen, for instance, in the presence of multiple and conforming bad data. Conforming bad data are considered the ones which appear in measurements with strongly correlated residuals and in turn, the residual errors appear consistent with each other [81, Ch.5]. As for SDR-SE, its robust version can be formulated by including an ℓ-1 regularization term in the original cost function (see [87] for details).

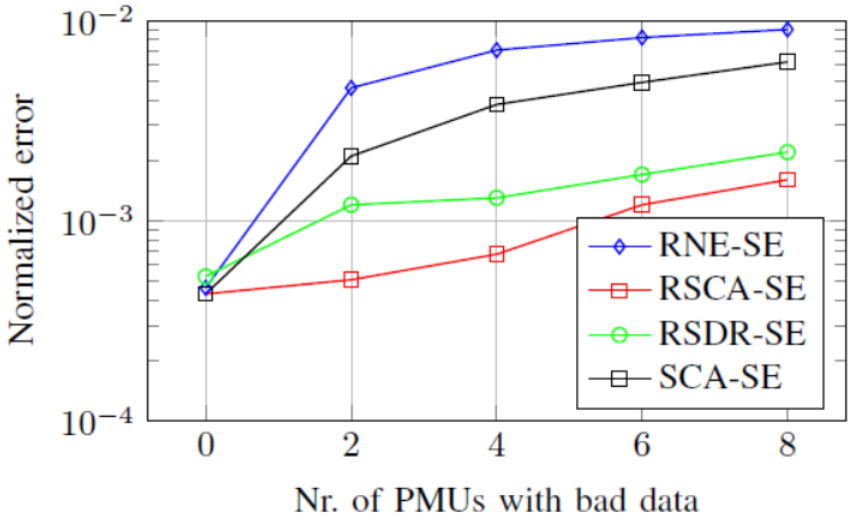

Figure 4.12: Normalized error vs. number of PMUs with bad data for the IEEE 57-bus test case. Parameter set: $\gamma = 0.01$, $\alpha = 0.99$, $\beta = 0.01$ and $\delta = 10^{-4}$.

Figure 4.12 refers to the scenario where bad data appear within the PMU measurements for the IEEE 57-bus test case. The measurement set includes 81 pairs of power flows, 57 voltage magnitudes and 17 PMUs. Further, bad data is progressively introduced in the PMUs at buses 2, 6, 10, 12, 19, 22, 46, 49. This subset of buses has been chosen as a worst-case scenario where the corrupted phasor measurements are spatially extended over the system. The graph reveals that the performance degradation experienced by the proposed RSCA-SE scheme (Section 4.6) is, by far, the lowest. In fact, with up to 2 corrupt PMU measurements the performance degradation with respect to the baseline case (clean measurements) is barely noticeable. On the contrary, RNE-SE is critically affected by gross errors. In fact, bad data goes undetected in all cases this resulting into a large normalized error. This is due to the fact that, PMU measurements

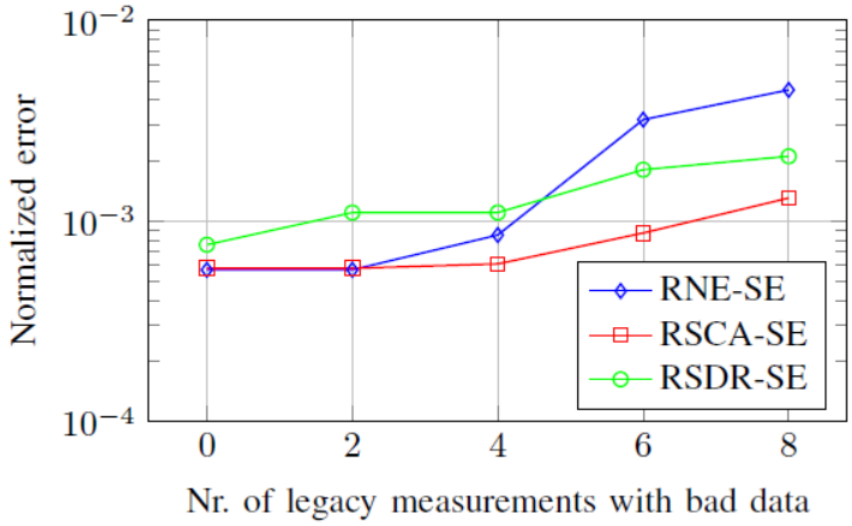

Figure 4.13: Normalized error vs. number of legacy measurements with bad data for the IEEE 30-bus test case. Parameter set: $\gamma = 0.01$, $\alpha = 0.99$, $\beta = 0.01$ and $\delta = 10^{-4}$.

are introduced in the traditional NE-SE with large weights as highly accurate measurements. The latter may lead the LNRT residual analysis, which involves the measurement weights, in misleading decisions [82]. Interestingly, the contrast with respect to the accuracy between the RSCA-SE and SCA-SE schemes is rather broad. This evidences the effectiveness of the followed approach introduced to render SCA-SE more robust to bad data (see Section 4.6). Notice also that, unlike in RNE-SE, the state estimation and bad data detection/removal in RSCA-SE is conducted in parallel. This, along with the more sophisticated detection method, results into an enhanced performance. As for the RSDR-SE, its gain with respect to the *plain* (i.e., non robust) SCA-SE scheme can be regarded as profitable, but at the same time moderate, compared to RSCA-SE.

Similar results are depicted in Fig. 4.13 in a scenario where bad data are contained in legacy measurements for the IEEE 30-bus test case. The measurement set consists of 41 pairs of power flows and 30 voltage magnitudes and bad data is progressively introduced in the voltage meters of buses 15 to 22. These specific voltage magnitude measurements has been chosen as corrupted in order to numerically assess the proposed scheme and benchmarks in a worst case scenario, which involves multiple and conforming bad data. Again, RSCA-SE is superior compared to the benchmarks. Interestingly, we observe that with up to 2 corrupted legacy measurements RNE-SE succeeded in detecting and extracting the bad data (equal estimation error with baseline case). However, as the number of bad data increases, RNE-SE presents a significant performance degradation due to the fact that LNRT was unable to detect the outliers.

Finally, we empirically assess the validity of Proposition 4.2, for both synchronized

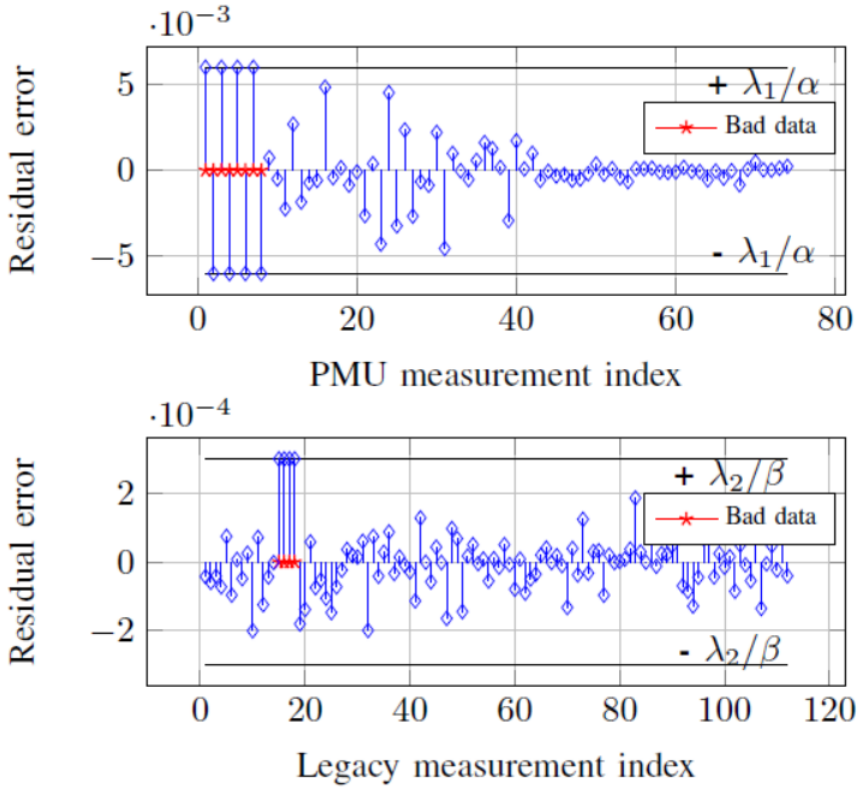

Figure 4.14: Residual error vs. measurement index: PMU - real and imaginary parts / IEEE 57-bus (top) and legacy / IEEE 30-bus (bottom). Parameter set: $\lambda_1/\alpha = 6 \cdot 10^{-3}$, $\lambda_2/\beta = 3 \cdot 10^{-4}$

and legacy measurements. For the former, bad data is contained in PMU measurements at buses 2,6, 10 and 12 for the case of IEEE 57-bus test system. For the latter, bad data is contained in voltage magnitude measurements at the buses 15 to 18 for the IEEE 30-bus test system. Indeed, in Fig. 4.14 we observe how, for all individual measurements, the absolute value of the residual error is upper bounded by λ_1/α (or λ_2/β). And, further, that such upper bound is reached when measurements effectively contain bad data (which is denoted by red crosses), as anticipated by Proposition 4.2.

4.8 Conclusions

In this chapter, we have considered a SE scenario with measurements available from both PMUs and legacy meters. The resulting SE model constitutes a non-convex optimization problem. In order to overcome this challenge, we have leveraged on a successive convex approximation approach (SCA-SE), where the non-convex objective function is approximated by a sequence of strongly convex ones, solving the problem in an iterative manner.

The proposed scheme has been numerically assessed over a variety of measurement set scenarios and different test systems, namely the IEEE 30-bus, 57-bus and 118-bus test systems. For comparison reasons, we have chosen as benchmarks an efficient SE scheme based on a semidefinite relaxation and the traditional normal equations based SE. The numerical results have shown that the SCA-SE scheme outperforms in terms of accuracy. This applies in all different scenarios of measurement redundancy and system state under consideration. In addition, we incorporate to the SCA-SE second order information from the original non-convex objective function via the presented A-Hessian matrix. Computer simulations have shown that this approach has concluded in an improved convergence rate of the proposed algorithm avoiding additional computational burden. Then, we have exploited the properties of the successive convex approximation framework and we have presented a distributed implementation of the SCA-SE. To do so, we have leveraged on the ADMM, an efficient framework suitable for multi-area settings. The computer simulations have revealed that, in contrast to other decentralized approaches, our scheme exhibits slight performance degradation with respect to the centralized version. Finally, in order to detect and clean measurements with bad data, the SCA-SE scheme has been reformulated as a LASSO-like estimation problem by promoting sparsity in the vector of corrupted measurements (RSCA-SE). This applies for both PMU and legacy observations. Additionally, we have derived an upper bound of the residual error and concluded that the resulting robust state estimator resembles the Huber's estimator. The RSCA-SE has been numerically assessed in different scenarios where PMU and legacy measurements have been corrupted with gross error. The numerical analysis have proved that the proposed scheme detects and cleanses bad data measurements more effectively than other methods from the literature, such as the LNRT.

4.9 Appendix A: Proof for Proposition I

The surrogate objective function for the robust state estimation optimization problem is:

$$\begin{aligned}\tilde{U}(\mathbf{x},\tilde{\mathbf{x}},\mathbf{o},\boldsymbol{\xi},\tilde{\boldsymbol{\xi}}) = \frac{\alpha}{2}\left\|\mathbf{r}-\mathbf{A}\mathbf{x}-\mathbf{o}\right\|_2^2 + \lambda\left\|\mathbf{o}\right\|_1 \\ + \frac{\beta}{2}\Big\{\mathbf{g}_{\tilde{\mathbf{x}}}^T(\mathbf{x}-\tilde{\mathbf{x}}) + \mathbf{g}_{\tilde{\boldsymbol{\xi}}}^T\left(\boldsymbol{\xi}-\tilde{\boldsymbol{\xi}}\right) \\ + \frac{\rho}{2}\left\|\mathbf{x}-\tilde{\mathbf{x}}\right\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{\xi}-\tilde{\boldsymbol{\xi}}\right\|_2^2\Big\} + \lambda\left\|\boldsymbol{\xi}\right\|_1\end{aligned} \tag{4.42}$$

We define the auxiliary variables $\boldsymbol{\mu}\in\mathbb{R}^K$, $\boldsymbol{\pi}\in\mathbb{R}^{2N}$, $\boldsymbol{\upsilon}\in\mathbb{R}^L$ as follows:

$$\boldsymbol{\mu} = \mathbf{r}-\mathbf{A}\mathbf{x}-\mathbf{o} \tag{4.43}$$

$$\boldsymbol{\pi} = \mathbf{x}-\tilde{\mathbf{x}} \tag{4.44}$$

$$\boldsymbol{\upsilon} = \boldsymbol{\xi}-\tilde{\boldsymbol{\xi}} \tag{4.45}$$

and accordingly, we define the Lagrangian function:

$$\begin{aligned}\mathscr{L} = \frac{\alpha}{2}\left\|\mu\right\|_2^2 + \lambda_1\left\|\mathbf{o}\right\|_1 \\ +\frac{\beta}{2}\Big\{\mathbf{g}_{\tilde{\mathbf{x}}}^T\boldsymbol{\pi} + \mathbf{g}_{\tilde{\boldsymbol{\xi}}}^T\boldsymbol{\upsilon} + \frac{\rho}{2}\left\|\boldsymbol{\pi}\right\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{\upsilon}\right\|_2^2\Big\} + \lambda_2\left\|\boldsymbol{\xi}\right\|_1 \\ +\boldsymbol{\psi}(\mathbf{r}-\mathbf{A}\mathbf{x}-\mathbf{o}-\boldsymbol{\mu}) + \boldsymbol{\tau}(\mathbf{x}-\tilde{\mathbf{x}}-\boldsymbol{\pi}) + \boldsymbol{\delta}(\boldsymbol{\xi}-\tilde{\boldsymbol{\xi}}-\boldsymbol{\upsilon})\end{aligned} \tag{4.46}$$

where $\boldsymbol{\psi}\in\mathbb{R}^K$, $\boldsymbol{\tau}\in\mathbb{R}^{2N}$, $\boldsymbol{\delta}\in\mathbb{R}^L$ are the dual variables. The partial derivatives with respect the primal variables read:

$$\frac{\partial\mathscr{L}}{\partial\boldsymbol{\mu}} \implies \alpha\boldsymbol{\mu}-\boldsymbol{\psi}^T=0 \implies \boldsymbol{\mu}=\boldsymbol{\psi}/\alpha \tag{4.47a}$$

$$\frac{\partial\mathscr{L}}{\partial\boldsymbol{\pi}} \implies \frac{\beta}{2}\mathbf{g}_{\tilde{\mathbf{x}}}+\frac{\beta\rho}{2}\boldsymbol{\pi}-\boldsymbol{\tau}=0 \implies \boldsymbol{\pi}=\frac{2}{\beta\rho}(\boldsymbol{\tau}-\frac{\beta}{2}\mathbf{g}_{\tilde{\mathbf{x}}}) \tag{4.47b}$$

$$\frac{\partial\mathscr{L}}{\partial\boldsymbol{\upsilon}} \implies \frac{\beta}{2}\mathbf{g}_{\tilde{\boldsymbol{\xi}}}+\frac{\beta\rho}{2}\boldsymbol{\upsilon}-\boldsymbol{\delta}=0 \implies \boldsymbol{\upsilon}=\frac{2}{\beta\rho}(\boldsymbol{\delta}-\frac{\beta}{2}\mathbf{g}_{\tilde{\boldsymbol{\xi}}}) \tag{4.47c}$$

$$\frac{\partial\mathscr{L}}{\partial\mathbf{x}} \implies -\mathbf{A}^T\boldsymbol{\psi}+\boldsymbol{\tau}=0 \tag{4.47d}$$

$$\frac{\partial\mathscr{L}}{\partial\mathbf{o}} \implies -\boldsymbol{\psi}+\lambda_1\partial\left\|\mathbf{o}\right\|_1 \tag{4.47e}$$

$$\frac{\partial\mathscr{L}}{\partial\boldsymbol{\xi}} = \boldsymbol{\delta}+\lambda_2\partial\left\|\boldsymbol{\xi}\right\|_1 \tag{4.47f}$$

Leveraging on the conditions (4.47) we can re-write and re-arrange (4.46) as:

$$\begin{aligned}\mathscr{L} = -\frac{1}{2\alpha}\left\|\psi\right\|_2^2 + \lambda_1\left\|\mathbf{o}\right\|_1 - \frac{\beta}{4\rho}\left\|\boldsymbol{\tau}-\frac{\beta}{2}\mathbf{g}_{\tilde{\mathbf{x}}}\right\|_2^2 \\ -\frac{\beta}{4\rho}\left\|\boldsymbol{\delta}-\frac{\beta}{2}\mathbf{g}_{\tilde{\boldsymbol{\xi}}}\right\|_2^2 + \lambda_2\left\|\boldsymbol{\xi}\right\|_1 + \boldsymbol{\psi}(\mathbf{r}-\mathbf{A}\mathbf{x}-\mathbf{o}-\boldsymbol{\mu}) \\ +\mu^T\mathbf{r}-\boldsymbol{\tau}^T\tilde{\mathbf{x}}+\boldsymbol{\delta}^T\tilde{\boldsymbol{\xi}}+\mathbf{x}^T(-\mathbf{A}^T\boldsymbol{\psi}+\boldsymbol{\tau})-\psi^T\mathbf{o}-\boldsymbol{\delta}^T\boldsymbol{\xi}\end{aligned} \tag{4.48}$$

Hence, the dual problem reads:

$$\begin{aligned} -\frac{1}{2\alpha}\left\|\psi\right\|_2^2 - \frac{\beta}{4\rho}\left\|\boldsymbol{\tau} - \frac{\beta}{2}\mathbf{g}_{\tilde{\mathbf{x}}}\right\|_2^2 - \frac{\beta}{4\rho}\left\|\boldsymbol{\delta} - \frac{\beta}{2}\mathbf{g}_{\tilde{\boldsymbol{\xi}}}\right\|_2^2 \\ +\psi^T\mathbf{r} - \boldsymbol{\tau}^T\tilde{\mathbf{x}} + \boldsymbol{\delta}^T\tilde{\boldsymbol{\xi}} \\ \text{s.t. } \mathbf{A}^T\boldsymbol{\psi} = \boldsymbol{\tau} \\ \left|\boldsymbol{\psi}\right|_\infty \leq \lambda_1 \\ \left|\boldsymbol{\delta}\right|_\infty \leq \lambda_2 \end{aligned} \tag{4.49}$$

Now, assume that the algorithm converges to a stationary solution $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\xi}}$. That means:

$$\mathbf{x} - \tilde{\mathbf{x}} = 0 \tag{4.50}$$

$$\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}} = 0 \tag{4.51}$$

Then, from (4.47a) we have:

$$\boldsymbol{\psi} = \alpha\boldsymbol{\mu} \tag{4.52}$$

and by definition (4.43):

$$\boldsymbol{\mu} = \mathbf{r} - \mathbf{A}\mathbf{x} - \mathbf{o} \tag{4.53}$$

so, again from (4.47a):

$$\left|\boldsymbol{\psi}\right|_\infty = \alpha\left|\boldsymbol{\mu}\right|_\infty = \alpha\left|\mathbf{r} - \mathbf{A}\mathbf{x} - \mathbf{o}\right|_\infty \leq \lambda_1 \tag{4.54}$$

and the residual is constrained by:

$$\left|\mathbf{r} - \mathbf{A}\mathbf{x} - \mathbf{o}\right|_\infty \leq \lambda_1/\alpha \tag{4.55}$$

Accordingly, for the part that refers to the legacy measurements, as $\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}} = 0$ from assumption (4.51), we have $\boldsymbol{\upsilon} = 0$ and from (4.47c):

$$\left|\boldsymbol{\delta}\right|_\infty = \left|\frac{\beta}{2}\mathbf{g}_{\tilde{\boldsymbol{\xi}}}\right|_\infty \leq \lambda_2 \tag{4.56}$$

where:

$$\mathbf{g}_{\tilde{\boldsymbol{\xi}}} = -2\sum_{l=1}^{L}\mathbf{e}_l\left(z_l - \tilde{\mathbf{x}}^T\mathbf{M}_l\tilde{\mathbf{x}} - \mathbf{e}_l^T\tilde{\boldsymbol{\xi}}\right) \tag{4.57}$$

4.10 Appendix B: SCA framework for the Robust SE

Theorem 4.3. *The successive convex approximation framework proposed in [86] for non-convex optimization problems in the form:*

$$\min_{\{\mathbf{w}\}} \quad V(\mathbf{w}) = G(\mathbf{w}) + F(\mathbf{w}) \tag{4.58}$$

where $G(\mathbf{w})$ is possibly a non-convex function and $F(\mathbf{w})$ is a non-smooth but convex one, is equally valid for the case of the optimization problem in (4.30)*:*

$$\min_{\{\mathbf{x},\boldsymbol{\xi}\}} \quad V(\mathbf{x},\boldsymbol{\xi}) = G(\mathbf{x},\boldsymbol{\xi}) + F(\boldsymbol{\xi}) \tag{4.59}$$

with

$$G(\mathbf{x},\boldsymbol{\xi}) = \frac{\beta}{2}\sum_{l=1}^{L}\left(z_l - \mathbf{x}^T\mathbf{M}_l\mathbf{x} - \mathbf{e}_l^T\boldsymbol{\xi}\right)^2$$

and

$$F(\boldsymbol{\xi}) = \lambda_\beta \left\|\boldsymbol{\xi}\right\|_1$$

Proof. First, we define the auxiliary vector:

$$\mathbf{w} = (\mathbf{x},\boldsymbol{\xi})^T \in \mathbb{R}^{(2N+L)\times 1} \tag{4.60}$$

with $\mathbf{x} \in \mathbb{R}^{2N}$ and $\boldsymbol{\xi} \in \mathbb{R}^L$, and the auxiliary matrices:

$$\boldsymbol{\Omega}_1 = [\mathbf{I}_{2N}, \mathbf{0}] \in \mathbb{R}^{2N\times(2N+L)} \tag{4.61}$$

where $\mathbf{I}_{2N} \in \mathbb{R}^{2N\times 2N}$ stands for the eye matrix and $\mathbf{0} \in \mathbb{R}^{2N\times L}$ is a matrix with all zeros, and:

$$\boldsymbol{\Omega}_2 = [\mathbf{0}, \mathbf{I}_L] \in \mathbb{R}^{L\times(2N+L)} \tag{4.62}$$

where, accordingly, $\mathbf{I}_L \in \mathbb{R}^{L\times L}$ stands for the eye matrix and $\mathbf{0} \in \mathbb{R}^{L\times 2N}$ is again, a matrix with all zeros.

Leveraging on the auxiliary vector $\mathbf{w}$, the problem in (4.59) can be re-arranged as follows :

$$U(\mathbf{w}) = \frac{\beta}{2}\left\{\sum_{l=1}^{L}\left(z_l - (\boldsymbol{\Omega}_1\mathbf{w})^T\mathbf{M}_l(\boldsymbol{\Omega}_1\mathbf{w}) - \mathbf{e}_l^T(\boldsymbol{\Omega}_2\mathbf{w})\right)^2\right\} + \lambda_\beta\left\|\boldsymbol{\Omega}_2\mathbf{w}\right\|_1 \tag{4.63}$$

Following the rationale of [86] and, simply linearizing the function around a feasible point $\tilde{\mathbf{w}}$, we can produce a strongly convex function of (4.63), that is:

$$\tilde{U}(\mathbf{w},\tilde{\mathbf{w}}) = \frac{\beta}{2}\Bigg\{ -2\sum_{l=1}^{L}\Big(z_l - (\boldsymbol{\Omega}_1\tilde{\mathbf{w}})^T\mathbf{M}_l(\boldsymbol{\Omega}_1\tilde{\mathbf{w}}) - \mathbf{e}_l^T(\boldsymbol{\Omega}_2\tilde{\mathbf{w}})\Big)$$
$$\Big((\boldsymbol{\Omega}_1^T\mathbf{M}_l\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_1^T\mathbf{M}_l^T\boldsymbol{\Omega}_1)\tilde{\mathbf{w}}_i + \boldsymbol{\Omega}_2^T\mathbf{e}_l)\Big)^T\Big(\mathbf{w}-\tilde{\mathbf{w}}\Big) + \frac{\rho}{2}\left\|\mathbf{w}-\tilde{\mathbf{w}}\right\|_2^2\Bigg\} + \lambda_\beta\left\|\boldsymbol{\Omega}_2\mathbf{w}\right\|_1 \tag{4.64}$$

Now, in (4.64) we have:

$$\Big((\boldsymbol{\Omega}_1^T\mathbf{M}_l\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_1^T\mathbf{M}_l^T\boldsymbol{\Omega}_1)\tilde{\mathbf{w}} + \boldsymbol{\Omega}_2^T\mathbf{e}_l)\Big)^T\Big(\mathbf{w}-\tilde{\mathbf{w}}\Big) \tag{4.65}$$

where,

$$(\boldsymbol{\Omega}_1^T\mathbf{M}_l\mathbf{M}_1 + \boldsymbol{\Omega}_1^T\mathbf{M}_l^T\boldsymbol{\Omega}_1)\tilde{\mathbf{w}} = (\boldsymbol{\Omega}_1^T\mathbf{M}_l\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_1^T\mathbf{M}_l^T\boldsymbol{\Omega}_1)\cdot$$

$$\left[\begin{bmatrix}\tilde{\mathbf{x}}\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{0}\\ \tilde{\boldsymbol{\xi}}\end{bmatrix}\right] = (\boldsymbol{\Omega}_1^T\mathbf{M}_l\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_1^T\mathbf{M}_l^T\boldsymbol{\Omega}_1)\begin{bmatrix}\tilde{\mathbf{x}}\\ \mathbf{0}\end{bmatrix}$$

based on the fact that:

$$(\boldsymbol{\Omega}_1^T\mathbf{M}_l\boldsymbol{\Omega}_1)\cdot\left[\begin{bmatrix}\tilde{\mathbf{x}}\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{0}\\ \tilde{\boldsymbol{\xi}}\end{bmatrix}\right] = (\boldsymbol{\Omega}_1^T\mathbf{M}_l\boldsymbol{\Omega}_1)\begin{bmatrix}\tilde{\mathbf{x}}\\ \mathbf{0}\end{bmatrix}$$

Additionally, bearing in mind that:

$$(\mathbf{w}-\tilde{\mathbf{w}}) = \left[\begin{bmatrix}\mathbf{x}\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{0}\\ \boldsymbol{\xi}\end{bmatrix} - \begin{bmatrix}\tilde{\mathbf{x}}\\ \mathbf{0}\end{bmatrix} - \begin{bmatrix}\mathbf{0}\\ \tilde{\boldsymbol{\xi}}\end{bmatrix}\right] \tag{4.66}$$

we can re-arrange accordingly (4.65) and, we have:

$$\left((\boldsymbol{\Omega}_1^T\mathbf{M}_l\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_1^T\mathbf{M}_l^T\boldsymbol{\Omega}_1)\begin{bmatrix}\tilde{\mathbf{x}}\\ \mathbf{0}\end{bmatrix} + \boldsymbol{\Omega}_2^T\mathbf{e}_l\right)\cdot\left[\begin{bmatrix}\mathbf{x}\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{0}\\ \boldsymbol{\xi}\end{bmatrix} - \begin{bmatrix}\tilde{\mathbf{x}}\\ \mathbf{0}\end{bmatrix} - \begin{bmatrix}\mathbf{0}\\ \tilde{\boldsymbol{\xi}}\end{bmatrix}\right] =$$
$$(\boldsymbol{\Omega}_1^T\mathbf{M}_l\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_1^T\mathbf{M}_l^T\boldsymbol{\Omega}_1)\begin{bmatrix}\tilde{\mathbf{x}}\\ \mathbf{0}\end{bmatrix}\left[\begin{bmatrix}\mathbf{x}\\ \mathbf{0}\end{bmatrix} - \begin{bmatrix}\tilde{\mathbf{x}}\\ \mathbf{0}\end{bmatrix}\right] + (\boldsymbol{\Omega}_2^T\mathbf{e}_l)\left[\begin{bmatrix}\mathbf{0}\\ \boldsymbol{\xi}\end{bmatrix} - \begin{bmatrix}\mathbf{0}\\ \tilde{\boldsymbol{\xi}}\end{bmatrix}\right] \tag{4.67}$$

Going one step further in (4.67) and dropping the auxiliary matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, we have:

$$(\mathbf{M}_l + \mathbf{M}_l^T)\tilde{\mathbf{x}}(\mathbf{x}-\tilde{\mathbf{x}}) + (\mathbf{e}_l)(\boldsymbol{\xi}-\tilde{\boldsymbol{\xi}}) \tag{4.68}$$

Finally, substituting (4.68) in (4.64) and again dropping $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$, we have:

$$\tilde{U}(\mathbf{w},\tilde{\mathbf{w}}) = \frac{\beta}{2}\Bigg\{ -2\sum_{l=1}^{L}\left(z_l - \tilde{\mathbf{x}}^T\mathbf{M}_l\tilde{\mathbf{x}} - \mathbf{e}_l^T\tilde{\boldsymbol{\xi}}\right)\left((\mathbf{M}_l + \mathbf{M}_l^T)\tilde{\mathbf{x}}\right)^T(\mathbf{x}-\tilde{\mathbf{x}}) \\ -\sum_{l=1}^{L}\mathbf{e}_l\left(z_l - \tilde{\boldsymbol{\xi}}^T\mathbf{M}_l\tilde{\mathbf{x}} - \mathbf{e}_l^T\tilde{\mathbf{x}}\right)\left(\boldsymbol{\xi}-\tilde{\boldsymbol{\xi}}\right) + \frac{\rho}{2}\left\|\mathbf{x}-\tilde{\mathbf{x}}\right\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{\xi}-\tilde{\boldsymbol{\xi}}\right\|_2^2\Bigg\} + \lambda_\beta\left\|\boldsymbol{\xi}\right\|_1 \quad (4.69)$$

that is equal with the strongly convex function 4.32 in Section 4.6.

■

Chapter 5

A Regularized State Estimation Scheme for a Robust Monitoring of the Distribution Grid

In this chapter, we propose a regularized SE scheme for the DG. The ultimate goal is to track accurately the system state at a faster time scale according to the requirements of the new operational environment. The SE algorithm operates at two different time scales in which the set of available measurements are different. At the main time instants (every 15 minutes) the set of observations comprises SCADA measurements, pseudomeasurements and μPMUs. In this case, we resort to a *regularized* version of the normal equations-based SE (R-NESE). In the intermediate time instants, only a reduced number of μPMU measurements is available. To circumvent observability issues, we exploit the fact that the voltage drop in adjacent buses is limited and, on that basis, a *regularized* weighted total variation estimation (WTVSE) problem is formulated. Then, the impact of in-line voltage regulators (IVLRs) is explicitly taken into consideration and that, forces us to *decompose* and solve the original SE problem for a number of smaller regions (D-WTVSE). The latter can be iteratively solved by resorting to the ADMM. Complementarily, we also present a μPMU placement method (μPP) in order to improve the conditioning of the R-NESE problem. This problem can be posed as a MISDP and, thus, can be efficiently solved. The performance of the proposed scheme is numerically assessed on (mostly) a 95-bus distribution system for a number of realistic conditions of noise, load and photovoltaic generation profiles. A number of benchmarks are provided, as well.

5.1 Introduction

SE is a well-studied problem for the TG. In contrast, SE for the DG has attracted lots of research interest over the last decade. However, DG SE turns out to be very challenging [40], [92], as we discuss next.

To start with, the uniqueness of DG network topologies often renders the SE problem ill-conditioned [37]. Several factors, such as the radial topologies and the low reactance/resistance (X/R) ratio [2], degrade the efficiency of the conventional estimators operating at the transmission level. At the same time, these peculiarities raise the need for analytical modeling of the distribution network [93]. Several works on DG SE [43], [94] have proposed alternative SE schemes suitable for radial-shaped DGs. The underlying idea is to consider the branch currents as the state variables instead of bus voltage and angles. This modification can ensure increased simplicity of the non-linear functions that describe the relationship between the state variables and the AC load flows. From a different point of view, a number of schemes have leveraged on the aforementioned characteristics in order to built heuristic methods. A monitoring algorithm based on a new formulation of the AC radial load flows can be found in [95]. The proposed scheme relies on the voltage measurements of a crucial set of buses and the *neighboring* state variables are adjusted iteratively to fit the measured values. In the same context, a model order reduction method is performed on the DG in [49]. The algorithm selects a number of nodes and branches and reduces them to an equivalent one. Consequently, the SE is capable of providing an estimate relying on a reduced number of PMUs. One more recent heuristic approach is presented in [96]. The method relies on an ad-hoc regularized parameter in order to deal with possible numerical instability introduced by the DG specificities. Interestingly, the scheme is implemented with slight modifications to the classical weighted least squares (WLS) estimator.

Another challenge here is the limited availability of measurements. In this case, utilities have leveraged on data mining techniques on historical and smart meter data, constructing pseudomeasurement sets [46]. By using such pseudomeasurements in conjunction with actual SCADA readings, (slightly) redundant measurement sets can be constructed. Nonetheless, these observation sets consist of a large number of (pseudo) power injection measurements, that in turn affects significantly the numerical conditioning of the normal equations [81] and deteriorate the SE accuracy [97]. In [98], it has been shown that the correlation among these pseudomeasurements reduces substantially the estimation performance. Consequently, an advanced processing of the high uncertainty that such load injections exhibit becomes a necessity [99]. A prior statistical treatment in order to overcome the nonsynchronized nature of the smart meter data is proposed in [100]. Based on the observation that the short-time load variations follow a normal distribution, the authors present a modified WLS-based SE.

Further, non-technical restrictions also affect the precision of the pseudo loads and in turn, the accuracy of SE. In [101], the authors emphasize the privacy restrictions with respect to the data aggregation from smart meters. These restrictions affect significantly the accuracy of the pseudomeasurements, degrading the voltage profile accuracy. In order to overcome this barrier, two strategies are proposed. The first is based on the consideration of the correlations between the errors of each pair of measurements. To do so, instead of the commonly adopted diagonal covariance matrix, the authors propose a non-diagonal one, which accounts for such correlations. The second strategy leverages on the power loss estimation at the side of the LV system. Given that, the precision of the considered pseudo loads at the MV/LV substations will be improved.

Furthermore, the increasing penetration of DER along with the introduction of new components, such as electric vehicles and distributed energy storage systems, have transformed the DG into an active infrastructure [102]. Besides, the power production profile of DERs (e.g., photovoltaic plants) exhibits variability at a shorter time scale. This results into an increasing number of voltage fluctuations and reverse power flows, which in turn may result in: i) the disconnection of the renewable resources; and ii) financial losses for the utilities [103]. For that reason, the distribution system operators attempt to keep an appropriate (within operational constraints) voltage profile at the medium voltage (MV) level system leveraging on active elements, such as the on-load tap changers at the primary distribution substations and the ILVRs across the feeders [38]. However, under specific operational situations these control devices may fail [104]. Consequently, the interest in monitoring (and optimizing) DG's MV systems with an increased temporal granularity and the maximum reliability has become a necessity.

To that aim, the distribution system operators may resort to the recently introduced, low-cost, μPMUs [39]. These devices are exclusively designed to provide visibility for the MV feeders, where, unlike high voltage systems, the phase angle differences between two measured locations lie usually at the level of small fractions of a degree. Thus, on the one hand, μPMUs typically operate at faster timescale, as opposed to legacy SCADA and pseudomeasurements (millisecond vs. second or minute timescales). On the other, the precision of their (linear and synchronized) measurements is higher, compared to the classical PMUs deployed at the TG. Indeed, a low number of μPMUs is able to improve significantly the estimated voltage profile [105], but also to enhance other applications for the DG [106]. However, a number of challenges, mostly related with the optimal number and position of the phasor meters, have to be also addressed. In [107], the authors present a mathematical analysis on the impact of synchrophasor technology on the voltage profile estimation. The main sources of uncertainty introduced from the PMUs are analyzed and interestingly, the authors prove that the number and location of the meters is of a high importance. As a matter of fact, several studies in the literature deal with these issues, following different rationales. For instance, an incremental placement

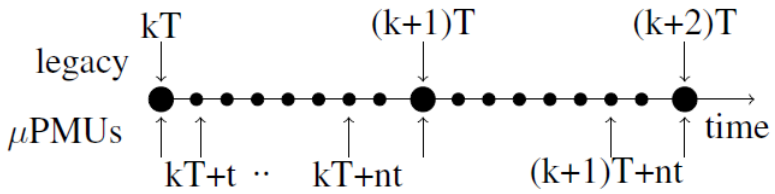

Figure 5.1: Legacy and μPMU measurements are available at time instants $\{kT, (k+1)T, \ldots\}$ and $\{kT, kT+t, kT+2t, \ldots\}$, respectively.

approach can be found in [108], [109]. In the latter, the authors present an integer linear programming model aimed to reach a predefined reliability level of system observability using a minimum number of phasor meters.

5.1.1 Contribution

In this chapter, we propose a regularized SE scheme for the DG robust monitoring, which operates in a two-time scale fashion. The ultimate goal is to accurately track the system state at a faster time scale with increased reliability, according to the needs of the new operational environment. In contrast with the previous schemes [44], [100] the presented algorithm is able to provide monitoring with an increased temporal granularity. Given that, it is capable of tracking short-term off-limit system conditions, which are mostly generated by voltage fluctuations and reverse power flows. This feature could be proven to be beneficial for the distribution system utilities in order to optimize the operation of the DERs and minimize their financial losses. At the same time, the scheme presents robustness against the high uncertainty of the available information. This overcomes the barrier of the restricted and low quality measurement sets, commonly available in the case of the DG. The main contributions of this chapter are as follows:

1) A state estimator that operates on the main time instants kT. The observations comprise SCADA measurements and pseudomeasurements, available every 15 minutes, Fig. 5.1. On top of these, a limited number of precise synchrophasors is added. This low redundant measurement set, characterized by a noise variation diversity, leads into an ill-conditioned optimization problem and affects significantly the SE accuracy [81]. In order to overcome this barrier and in line with [12], the SE is posed as a regularized non-linear least squares optimization problem. Resorting to the well-known Tikhonof regularization [5], the problem formulation refers to the *regularized* version of the normal equations-based SE (R-NESE). In addition, prior system information is introduced to the regularization term, that is, the last computed state estimate. Specifically, the proposed R-NESE presents the following properties:

- It improves the conditioning of the initial SE problem. For this reason, the quadratic

convergence characteristics of the normal equations are preserved, leading to a fast convergence rate compared to the commonly used normal equations as well as compared to other interesting approaches [110].

- It presents robustness against uncertainties raised by the specificities of the DG, e.g., low X/R ratio. More importantly, in contrast with other works [46], the algorithm overcomes the barriers of the low redundant measurement sets and the high error variance of pseudomeasurements, and it attains a high accuracy.

- The proposed scheme is practical for implementation since it just requires slight modifications with respect to the existing normal equations-based estimators.

2) A state estimation scheme that operates on the intermediate time instants kT+nt. In this case, we exclusively leverage on the positioned μPMUs at a subset of buses, Fig. 5.1. Since unavoidably the number of measurements is low, observability cannot be guaranteed. To circumvent that, we again resort to *regularization* techniques [5] leveraging on prior (expert) information on selected DG features. Specifically, the SE is formulated as a weighted *total variation* estimation (WTVSE) problem which limits variations in voltage estimates in adjacent buses. This stems from the fact that, in DGs, branch impedances (electrical lengths) and current flows tend to be lower than in TGs and so are voltage drops in adjacent buses [111]. On that basis, we propose a rule to define branch-specific weights for the regularizer. Further, possible zero power injection buses are exploited and introduced to the problem as constraints. Finally, the presence of ILVRs across the feeders has been taken into consideration. This allows us to re-formulate the problem as a constrained *decomposed* WTVSE (D-WTVSE). Then the D-WTVSE problem is effectively solved by finding an iterative solution based on the ADMM [72]. At a glance, the D-WTVSE scheme presents the following features:

- An efficient SE model that relies on a limited number of μPMUs and the zero power injection buses. The model counts for feeder structural differences (e.g., rural or urban), the intermittent nature of DER power production and active components installed within the feeder.

- The SE numerical treatment leverages on the ADMM which offers advantages over the ℓ_1-norm (WTV regularization) solution and model simplicity compared to traditional methods for ILVRs incorporation [111].

- Based on the superior properties of the ADMM, the algorithm presents satisfactory convergence rate and a low computational time is needed for full convergence. Given that, the implementation of D-WTVSE offers the capability for multiple consecutive estimates within the intermediate time instants.

3) An ad-hoc μPMU placement method (μPP). The presented model has been designed according to the demands of the proposed two-time scale SE scheme. To do

so, following a similar rationale with already developed algorithms for sensor scheduling purposes [112], we pose it as MISDP optimization problem. In particular, the objectives of the μPP problem are: (i) to ensure system observability (if needed) for the R-NESE scheme, while taking into consideration the existing measurements; (ii) to optimize the conditioning of the R-NESE problem [97] that, in turn, will result into more accurate state estimates and (iii) to allocate a sufficient subset of the predefined number of μPMUs in each DG region, according to the requirements imposed by D-WTVSE.

Comparing the μPP method with other works from the literature, a similar approach can be found in [113]. However, the authors emphasize on the placement of (non-synchronized) voltage and power flow meters. The same applies for [114], where a mixed integer linear programming model is presented. With respect to the synchrophasor technology, a work of similar nature can be found in [108]. The authors present a PMU placement method that improves the numerical stability of the SE problem. More specifically, the target of the proposed approach is to decrease the variances of the state estimates in order to improve the algorithmic accuracy, and to avoid the correlation among measurements aiming to optimize the bad data analysis. Other interesting works can be found in [115], [116], where the number of meters (to be placed) is optimized in relevance with financial criteria. In contrast, the presented method is able to allocate a predefined number of μPMUs according to possible budget availability and communication constraints [14]. Besides, based on the problem constraints, the operator is able to control the number of meters located to the different feeder areas. For instance, this can be exploited for increased monitoring in a specific subset of buses where industrial loads or high PV penetration are located.

The rest of the chapter is organized as follows. In the next section we provide the basic DG SE system model. In subsections 5.3.1 and 5.3.2 we present the algorithms for the SE schemes at the main (R-NESE) and intermediate time instants (D-WTVSE), respectively. The solution of D-WTVSE via ADMM is presented in Section 5.4, while Section 5.5 introduces the μPMU placement method. Finally, the proposed algorithms are numerically assessed in Section 5.6. To conclude, we present our conclusions in Section 5.7.

5.2 System Model

Consider a balanced[1] radial distribution grid (see Fig. 5.2) represented by the graph $\mathscr{G} = (\mathscr{V}, \mathscr{B})$, where $\mathscr{V}$ denotes the set of buses with cardinality $|\mathscr{V}| = N$; and $\mathscr{B}$ stands for the set of edges that describes their interconnections (branches), with cardinality $|\mathscr{B}| = N - 1$. The complex current injections at the buses, i.e., $\mathbf{i} = [I_1, \ldots, I_N]^T$, satisfy

[1] A radial distribution grid is considered balanced when it does not include un-transposed lines which are unbalanced because of single phase, two phase and three phase loads.

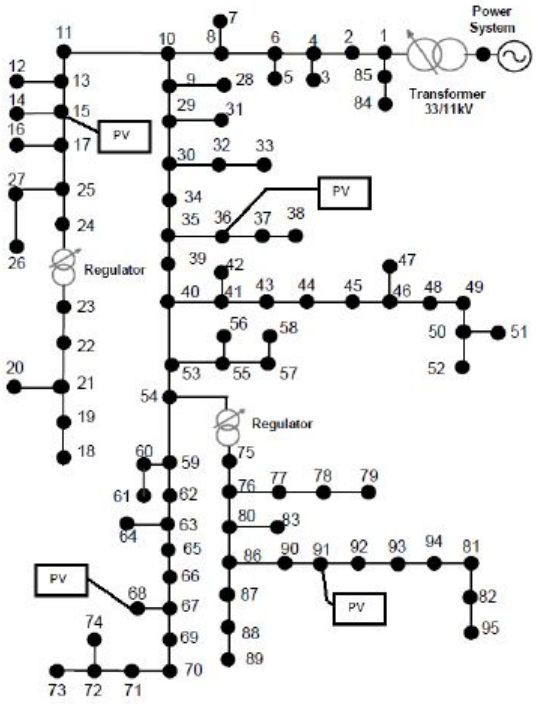

Figure 5.2: A 95-bus radial distribution grid comprising four photovoltaic (PV) generation plants (adapted from [9]).

$\mathbf{i} = \mathbf{Y}\mathbf{v}$, where $\mathbf{Y} \in \mathbb{C}^{N\times N}$ is the nodal admittance matrix [117] and $\mathbf{v} = [V_1, V_2, ..V_N]^T \in \mathbb{C}^N$ stands for the complex voltages at all the buses that determines the state of the DG. The ultimate goal of power system state estimator (SE) is to estimate $\mathbf{v}$ from a set of measurements at selected buses and branches.

Two types of measurements are considered available, namely, legacy and μPMU measurements. As depicted in Fig. 5.1, their availability is at two different time scales and, moreover, observations are of different nature. The first category comprises readings from the SCADA system as well as pseudo-measurements. Legacy measurements are typically obtained at time intervals of $T = 15$ minutes [48] and include: squared voltage magnitudes, $|V_m|^2$ for $m \in \mathscr{V}$; power flows at the branches, given by $S_{m,l} = P_{m,l}+jQ_{m,l} = V_m I^*_{m,l}$ for $(m,l) \in \mathscr{B}$; power injections, given by $S_m = P_m+jQ_m = V_m I^*_m$ for $m \in \mathscr{V}$ and squared branch current magnitudes, $|I_{m,l}|^2$ for $(m,l) \in \mathscr{B}$. In contrast, the sampling period of μPMUs $t = T/q$ is smaller, typically on the order of few milliseconds [39], and provide the DG operator with updated snapshots of complex bus voltages, $V_m = \Re\{V_m\} + j\Im\{V_m\}$ for $m \in \mathscr{V}$. Each of these complex bus voltages is exploited using its polar or Cartesian coordinates. Notice that legacy measurements are non-linear functions of the system state, whereas μPMU measurements turn out to be linear ones.

Let vector $\mathbf{x} = [\Re\{V_1\}, \Im\{V_1\}, \ldots, \Re\{V_N\}, \Im\{V_N\}]^T \in \mathbb{R}^{2N}$ denote the Cartesian

representation of the system state, and $\mathbf{z} = \left[\mathbf{r}^T, \mathbf{y}^T\right]^T \in \mathbb{R}^M$ stand for the corresponding measurement vector, where $\mathbf{r} \in \mathbb{R}^L$ and $\mathbf{y} \in \mathbb{R}^K$ are the vectors of legacy and μPMU measurements, respectively. It is worth noting that, typically, applies $M = L + K > 2N$ but $K < 2N$ due to deployment and operational costs. Finally, we can write:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e} \tag{5.1}$$

where $\mathbf{h}(\mathbf{x})$ denotes a non-linear function of $\mathbf{z}$ on $\mathbf{x}$ in compliance with the AC power flow model [117]; and $\mathbf{e} \in \mathbb{R}^M$ stands for zero-mean Gaussian noise with known covariance matrix $\mathbf{R_e}$.

5.3 A Two-time scale State Estimation Scheme

Bearing all the above in mind, we propose the following SE algorithm. At time instants $kT, (k+1)T, \ldots, (k+n)T$ for $k \in \mathbb{N}$, where *both* legacy and μPMU measurements are available, a SE algorithm based on the regularized normal equations is adopted (see Section 5.3.1 next). On the contrary, for the intermediate time instants between kT and $(k+1)T$ we adopt a state estimator formulated as a constrained decomposed weighted total variation regularization problem (subsection 5.3.2). Note that the latter relies only on the observations coming from a reduced number of μPMUs and, possibly, the available zero power injections.

5.3.1 Regularized Normal Equations-based State Estimator (R-NESE)

According to (5.1), the conventional state estimator is given by the solution to the following non-convex optimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\{\mathbf{x}\}} \frac{1}{2} \|\mathbf{z} - \mathbf{h}(\mathbf{x})\|_2^2 \tag{5.2}$$

where, in the above (and following) expression(s), the time index has been dropped for the ease of notation. To solve (5.2), one can resort to the so-called Normal Equations [81]:

$$\mathbf{G}^{(\nu)} \Delta\mathbf{x}^{(\nu)} = \mathbf{J}_{\mathbf{x}}^{(\nu)T} \mathbf{R}_{\mathbf{e}}^{-1} [\mathbf{z} - \mathbf{h}(\mathbf{x}^{(\nu)})] \tag{5.3}$$

where ν stands for the algorithm iteration index; matrix $\mathbf{J}_{\mathbf{x}}^{(\nu)} \in \mathbb{R}^{M \times 2N}$ denotes the Jacobian matrix of $\mathbf{h}(\mathbf{x})$ evaluated at ν^{th} iteration (i.e., the matrix of all first order partial derivatives of the non-linear equations vector $\mathbf{h}(\mathbf{x}^{(\nu)})$); $\mathbf{R_e} \in \mathbb{R}^{M \times M}$ is the *known* covariance matrix of the measurements; $\mathbf{G}^{(\nu)} = \mathbf{J}_{\mathbf{x}}^{(\nu)T} \mathbf{R}_{\mathbf{e}}^{-1} \mathbf{J}_{\mathbf{x}}^{(\nu)}$ stands for the gain matrix and; $\Delta\mathbf{x}^{(\nu)} = \mathbf{x}^{(\nu+1)} - \mathbf{x}^{(\nu)}$ captures the variation of the state variables over consecutive iterations.

In the DG SE, the accuracies of the various elements in the measurement vector $\mathbf{z}$ can be substantially different. This results in a near-singular gain matrix $\mathbf{G}$ [81] which, in turn, results into an ill-conditioned system of equations in (5.3).

For that reason, we adopt the following *regularized* non-linear least-squares optimization problem [118]:

$$\hat{\mathbf{x}} = \arg\min_{\{\mathbf{x}\}} \frac{1}{2}\big[(\mathbf{z}-\mathbf{h}(\mathbf{x}))^T\mathbf{R}_{\mathbf{e}}^{-1}(\mathbf{z}-\mathbf{h}(\mathbf{x})) + \tau\,\|\mathbf{x}-\mathbf{x}_p\|_2^2\big] \tag{5.4}$$

that refers to the Tikhonof regularization method [5], also known as ridge regression [119] in the field of statistics. In equation (5.4), τ is the parameter that controls the amount of regularization establishing the trade-off between the regularization term $\|\mathbf{x}-\mathbf{x}_p\|_2^2$ and data fidelity [120]. In addition, $\mathbf{x}_p$ is an application dependent vector, here, the estimated state in the previous time instant.

The Tikhonof regularization method is commonly applied to non-well posed optimization models [121]. These problems present numerical instability, namely, small perturbations to the input data results in large variances of the outputs. The regularization method exploits prior information in order to avoid this. More precisely, based on the assumption that the exact solution is a *smooth* function, the algorithm returns a result that approximates efficiently the real one. From a mathematical perspective, the underlying rationale is to extract the source of ill-conditioning from the respective system of equations. With respect to the SE problem (5.3), this entails the filtering of the contributions to the state estimate $\hat{\mathbf{x}}$ which correspond to the lower singular values of $\mathbf{G}$. The latter is linked with the value of the control parameter τ and consequently, an appropriate choice is needed [120].

Next, we approximate the non-linear score function $f(\mathbf{x})$ in problem (5.4) by expanding it into its Taylor series around $\mathbf{x}^{(\nu)}$ and neglecting the higher order terms, which yields:

$$\tilde{f}(\mathbf{x}) = f(\mathbf{x}^{(\nu)}) + \nabla f(\mathbf{x}^{(\nu)})(\mathbf{x}-\mathbf{x}^{(\nu)}) + \frac{1}{2}\nabla^2 f(\mathbf{x}^{(\nu)})(\mathbf{x}-\mathbf{x}^{(\nu)})^2 \tag{5.5}$$

Hence, the first order optimality condition for (5.5) can be computed as:

$$\nabla\tilde{f}(\mathbf{x}) = \nabla f(\mathbf{x}^{(\nu)}) + \nabla^2 f(\mathbf{x}^{(\nu)})(\mathbf{x}-\mathbf{x}^{(\nu)}) = 0 \tag{5.6}$$

This last expression, in turn, can be re-written as the iterative Gauss-Newton method:

$$\mathbf{x}^{(\nu+1)} = \mathbf{x}^{(\nu)} - [\nabla^2 f(\mathbf{x}^{(\nu)})]^{-1}\nabla f(\mathbf{x}^{(\nu)}) \tag{5.7}$$

with:

$$\nabla f(\mathbf{x}^{(\nu)}) = -\mathbf{J}_{\mathbf{x}}^{(\nu)T}\mathbf{R}_{\mathbf{e}}^{-1}(\mathbf{z}-\mathbf{h}(\mathbf{x}^{(\nu)})) + \tau(\mathbf{x}^{(\nu)}-\mathbf{x}_p) \tag{5.8}$$

and

$$\nabla^2 f(\mathbf{x}^{(\nu)}) = \mathbf{J}_{\mathbf{x}}^{(\nu)T}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{J}_{\mathbf{x}}^{(\nu)} + \tau\boldsymbol{I} \tag{5.9}$$

Algorithm 5.1 Regularized NESE state estimation at an arbitrary time instant kT

1: Initialize $\nu = 0$, $\mathbf{x}^{(0)}$, $\mathbf{x}_p$, $\mathbf{J}_{\mathbf{x}}^{(0)}$, $\mathbf{G}^{(0)}$, τ.
2: **repeat**
3: $\quad \nu = \nu + 1$
4: $\quad$ Compute $\tau^{(\nu)}$.
5: $\quad$ Compute $\Delta\hat{\mathbf{x}}_{kT}^{(\nu)}$ with (5.10).
6: $\quad$ Update $\hat{\mathbf{x}}_{kT}^{\nu} = \Delta\hat{\mathbf{x}}_{kT}^{(\nu)} + \hat{\mathbf{x}}_{kT}^{\nu-1}$.
7: **until** $\Delta\hat{\mathbf{x}}_{kT}^{(\nu)} < \varepsilon$

where the last two expressions follow from equation (5.4). From all the above, equation (5.7) can be re-written as:

$$\begin{aligned} \left(\mathbf{J}_{\mathbf{x}}^{(\nu)T}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{J}_{\mathbf{x}}^{(\nu)} + \tau^{(\nu)}\boldsymbol{I}\right)\Delta\mathbf{x}^{(\nu)} = \\ \mathbf{J}_{\mathbf{x}}^{(\nu)T}\mathbf{R}_{\mathbf{e}}^{-1}[\mathbf{z} - \mathbf{h}(\mathbf{x}^{(\nu)})] - \tau^{(\nu)}(\mathbf{x}^{(\nu)} - \mathbf{x}_p) \end{aligned} \tag{5.10}$$

that is, as a regularized version of the normal equations, see (5.3), where the parameter $\tau^{(\nu)}$ has to be re-computed in each iteration ν (see Section 5.6).

Algorithm 5.1 summarizes the corresponding iterative procedure.

5.3.2 Constrained Weighted Total Variation State Estimator (WTVSE)

The following linear model relates the set of measurements provided by the μPMUs (i.e., $\mathbf{y}$) at an arbitrary time instant $kT + nt$ (bus voltage phasors in Cartesian coordinates) with the system state $\mathbf{x}$ (again, the time index has been omitted for the ease of notation):

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{5.11}$$

where $\mathbf{H} \in \mathbb{R}^{K\times 2N}$ stands for the corresponding measurement matrix, and $\mathbf{w}$ denotes zero-mean Gaussian noise with known covariance matrix $\mathbf{R}_{\mathbf{w}}$. In comparison with the *full* measurement vector $\mathbf{z}$, the number of measurements provided by the μPMUs is typically lower than the dimension of the system state vector, i.e, $K \leq 2N$. This possibly renders the SE problem unsolvable since the system of equations (5.11) is underdetermined. This can be alleviated by introducing a regularization term accounting for the fact that, in DGs, the voltage difference between adjacent buses is small.

From all the above, the state estimation problem can be posed as a *constrained* weighted total variation regularization problem (WTVSE) [122]:

$$\begin{aligned} \hat{\mathbf{x}} = \arg\min_{\{\mathbf{x}\}} \frac{1}{2}\left\|\mathbf{y} - \mathbf{H}\mathbf{x}\right\|_2^2 + \lambda\left\|\mathbf{W}\mathbf{L}\mathbf{x}\right\|_1 \\ s.t. \quad \tilde{\mathbf{Y}}\mathbf{x} = \mathbf{0} \end{aligned} \tag{5.12}$$

where, in the score function above, the term $\frac{1}{2}\left\|\mathbf{y} - \mathbf{H}\mathbf{x}\right\|_2^2$ encourages *fidelity* in the solution to the vector of observations $\mathbf{y}$; $U(\mathbf{x}) = \left\|\mathbf{W}\mathbf{L}\mathbf{x}\right\|_1$ is the *regularization* term,

which introduces a prior model of the state vector $\mathbf{x}$; and λ denotes the regularization parameter, which controls the tradeoff between data fidelity and prior knowledge.

Matrix $\mathbf{L} \in \mathbb{R}(2N-2) \times 2N$ is in charge of generating the vector with the voltage differences for each pair of adjacent buses m and l. Since real and imaginary parts are stacked in the same vector, it can be expressed as

$$\mathbf{L} = \mathbf{\Lambda} \odot \mathbf{I}_{2\times 2} \tag{5.13}$$

where, assuming that the elements in the i-th row of matrix $\mathbf{\Lambda} \in \mathbb{R}N-1 \times N$ are associated to the (m,l) branch in the DG, they read

$$[\mathbf{\Lambda}]_{i,j} = \begin{cases} 1, & j = m; \\ -1, & j = l; \\ 0, & \text{otherwise} \end{cases} \tag{5.14}$$

and, further, the operator $\odot$ denotes Kronecker product, and $\mathbf{I}_{2\times 2}$ stands for the 2×2 identity matrix. The diagonal weighting matrix $\mathbf{W} \in \mathbb{R}(2N-2) \times (2N-2)$ in the regularization term of problem (5.12) above can, in turn, be expressed as

$$\mathbf{W} = \mathbf{\Omega} \odot \mathbf{I}_{2\times 2} \tag{5.15}$$

where (diagonal) matrix $\mathbf{\Omega} \in \mathbb{R}(N-1) \times (N-1)$ collects the weights for each branch. Taking into consideration that the voltage drop along the (m,l) branch is proportional to the branch impedance $Z_{m,l} = R_{m,l} + jX_{m,l}$, that is, $V_m - V_l = Z_{ml}I_{ml}$ (see Fig. 5.3), a sensible choice for the aforementioned weights is

$$\omega_{i,i} = 1 - \frac{|Z|_{m,l}}{|Z|_{\text{total}}} \tag{5.16}$$

where we have assumed again that the i-th row of matrix $\mathbf{\Lambda}$ is associated to the (m,l) branch, and Z_{total} denotes the total impedance for the feeder where the (m,l) branch lies. By doing so, the larger the branch impedance, the smaller the weight, this allowing for larger variations in the corresponding buses. This allows to design a regularization term of (5.12) which properly reflects the specific characteristics (i.e., electrical lengths) of the various sections in medium-voltage feeders[2].

Finally, the equality constraint $\tilde{\mathbf{Y}}\mathbf{x} = \mathbf{0}$ in problem (5.12) accounts for the d zero current (power) injection buses. For instance, for the $m \in \mathscr{V}$ zero current injection bus, we have

$$Y_{mm}V_m + \sum_{l \in \mathscr{N}} Y_{ml}V_l = 0 \tag{5.17}$$

[2]These may be substantially different in e.g., rural or urban areas.

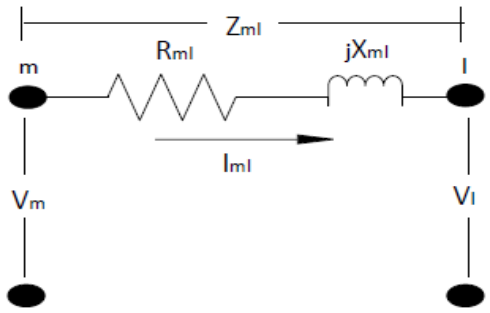

Figure 5.3: Voltage drop along the (m, l) branch.

where Y_{mm} stands for the sum of the admittances directly connected to bus m, $\mathcal{N}$ is the set of buses connected to the bus m, Y_{ml} stands for the admittance that connect the buses m, l, and V_m is the complex-valued voltage at bus m. By (i) splitting equation (5.17) into its real and imaginary parts; (ii) stacking the resulting two parts in consecutive equations of the system; (iii) leveraging on the Cartesian representation of the (complex) bus voltages in the system state vector $\mathbf{x}$; and (iv) re-arranging terms, the individual elements in matrix $\tilde{\mathbf{Y}}$ follow.

5.3.2.1 Inclusion of In-Line Voltage regulators

Next, the WTVSE problem (5.12) can be re-formulated to account for active elements installed in the DG. Specifically, we consider the presence of the so-called ILVRs, which automatically provide voltage regulation if the respective limits have been exceeded (e.g., 0.95 - 1.05 per-unit). However, we have to take into account that (in contrast to the case of substation transformers), the voltage base (used for the per-unit calculations) in both areas on each side of ILVRs are identical. Hence, the ILVRs (when active) will *break down* the feeders in areas of different per-unit voltage levels, leading the total variation regularization term in (5.12) to possible inefficiency. In order to deal with this specificity of WTVSE, for the numerical treatment of the ILVRs we will follow a decomposed formulation based on their location. Nevertheless, this offers the advantage of increased model simplicity compared to the commonly implemented methods, which are based on the augmented state vector and admittance matrix [111]. As an illustrative example, Fig. 5.4 presents the 95-bus UK DG. The system is divided into three areas according to the location of the ILVRs and in each case, the buses connected to the ILVR branch, i.e., {23-24} and {54-75}, become the border nodes of the respective adjacent regions.

Now, consider a DG feeder that includes P ILVRs. Lets assume that each regulator divides the system in two different areas, namely r and s, of different voltage levels (Fig. 5.5) and the total number of areas is A. Accordingly, we define $\mathscr{A}_r$ as the set of adjacent

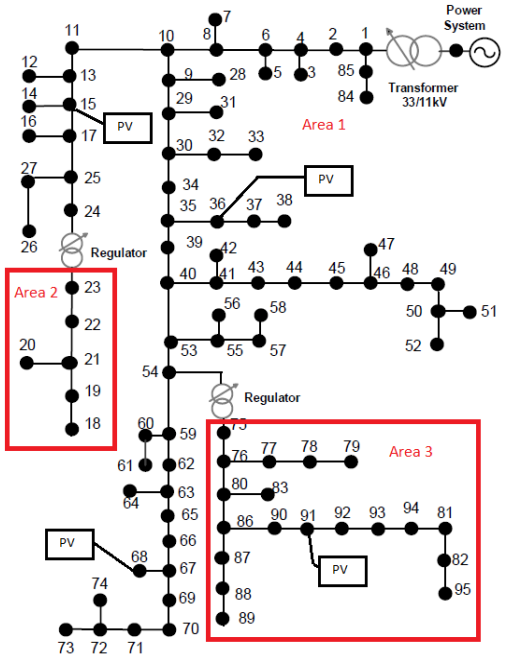

Figure 5.4: Decomposition of the 95-bus UK DG according to the location of the ILVRs.

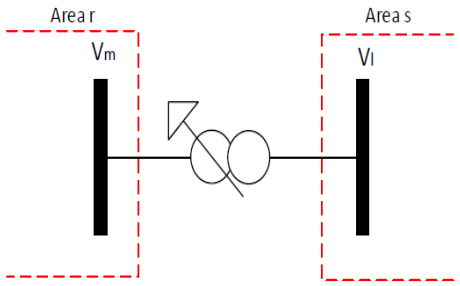

Figure 5.5: WTVSE decomposition according to the ILVR.

areas to area r and $\mathscr{S}_{rs}$ as the set of adjacent buses m and l between areas r and s. Then, the constraints $a_{rs}V_{mR} = V_{lR}$ and $a_{rs}V_{mI} = V_{lI}$ have to be satisfied, where a_{rs} stands for the tap position[3] indicator of the corresponding ILVR, usually displayed in the distribution management system.

Taking all the above into consideration, problem (5.12) can be re-formulated as a decomposed WTVSE (D-WTVSE) optimization problem based on the positions of the ILVRs (Fig. 4):

$$\begin{aligned} \arg\min_{\{\mathbf{x}_r\}} \sum_{r=1}^{A} \frac{1}{2}\left\|\mathbf{y}_r - \mathbf{H}_r\mathbf{x}_r\right\|_2^2 + \lambda\left\|\mathbf{W}_r\mathbf{L}_r\mathbf{x}_r\right\|_1 \\ s.t. \quad \tilde{\mathbf{Y}}_r\mathbf{x}_r = \mathbf{0}_r \\ \begin{bmatrix} a_{rs} & 0 & -1 & 0 \\ 0 & a_{rs} & 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{e}_{mR}^{(r)T}\mathbf{x}_r \\ \mathbf{e}_{mI}^{(r)T}\mathbf{x}_r \\ \mathbf{e}_{lR}^{(s)T}\mathbf{x}_s \\ \mathbf{e}_{lI}^{(s)T}\mathbf{x}_s \end{bmatrix} = \mathbf{0} \quad s \in \mathscr{A}_r,\ (m,l) \in \mathscr{S}_{rs},\ \forall r \end{aligned} \tag{5.18}$$

where $\mathbf{x}_r \in \mathbb{R}^{2N_r}$, $\mathbf{y}_r \in \mathbb{R}^{K_r}$, $\mathbf{H}_r \in \mathbb{R}^{K_r \times 2N_r}$ refer to area r and follow the rationale explained above. The same applies for $\mathbf{W}_r \in \mathbb{R}^{(2N_r-2)\times(2N_r-2)}$, $\mathbf{L}_r \in \mathbb{R}^{(2N_r-2)\times 2N_r}$ and $\tilde{\mathbf{Y}}_r$. The vector $\mathbf{e}$ is the singleton vector with all elements zero except one with 1 as its entry. For instance, $\mathbf{e}_{mR}^{(r)}$ ($\mathbf{e}_{mI}^{(r)}$) stands for a vector with 1 at the position $2m-1$ ($2m$) that refers to the real (imaginary) part of the voltage at the bus m. In addition, the superscript r denotes that the vector size is that of $\mathbf{x}_r$. Finally, as already mentioned, a_{rs} refers to the real value ratio of the voltage at the two sides of the ILVR (buses m and l).

5.4 Solving the D-WTVSE Problem via ADMM

In order to solve the D-WTVSE problem (5.18), we resort to the ADMM [72]. Among other candidates, ADMM presents computational advantages over the non-smooth ℓ_1-norm treatment and adapts well to decomposed optimization problems. To that aim, however, problem (5.18) must be re-written as:

$$\begin{aligned} \arg\min_{\{\mathbf{x}_r\},\{\boldsymbol{c_{rs}}\},\{\boldsymbol{\theta}_r\}} \sum_{r=1}^{A} \frac{1}{2}\left\|\mathbf{y}_r - \mathbf{H}_r\mathbf{x}_r\right\|_2^2 + \lambda\left\|\boldsymbol{\theta}_r\right\|_1 \\ s.t. \quad \begin{bmatrix} \mathbf{W}_r\mathbf{L}_r \\ \tilde{\mathbf{Y}}_r \\ \mathbf{M}_{rs} \end{bmatrix} \mathbf{x}_r - \begin{bmatrix} \boldsymbol{\theta}_r \\ \mathbf{0}_r \\ \boldsymbol{c_{rs}} \end{bmatrix} = \mathbf{0}_r \end{aligned} \tag{5.19}$$

[3]The tap position refers to the number of winding turns at the primary and secondary side of the ILVR (amount of voltage regulation provided) which determines the actual ratio (e.g. 1.05) between the respective primary and secondary side voltage.

where $\mathbf{M}_{rs} \in \mathbb{R}^{2\times 2N_r}$ for all $s \in \mathscr{A}_r$ is defined as:

$$[\mathbf{M}_{rs}]_{i,j} = \begin{cases} a_{rs}, & i = 1, \quad j = 2m-1; \\ a_{rs}, & i = 2, \quad j = 2m; \\ 0, & \text{otherwise} \end{cases} \tag{5.20}$$

and $\boldsymbol{c_{rs}} = [c_{rsml\{R\}}; c_{rsml\{I\}}]$ for all $s \in \mathscr{A}_r$ stands for an introduced auxiliary vector per each pair of connected areas via ILVR, in order to fully decompose the problem. Moreover, $\boldsymbol{\theta}_r$ plays the role of an auxiliary variable and, hence, it introduces an additional constraint. Accordingly, the *augmented* Lagrangian reads:

$$\begin{aligned} \mathscr{L}\big(\mathbf{x}_r, \boldsymbol{c_{rs}}, \boldsymbol{\theta}_r, \boldsymbol{\mu}_r, \boldsymbol{\pi}_r, \boldsymbol{\kappa_{rs}}\big) = \sum_{r=1}^{A} \Bigg[& \frac{1}{2}\left\|\mathbf{y}_r - \mathbf{H}_r\mathbf{x}_r\right\|_2^2 + \lambda\left\|\boldsymbol{\theta}_r\right\|_1 \\ & + \boldsymbol{\mu}_r^T(\mathbf{W}_r\mathbf{L}_r\mathbf{x}_r - \boldsymbol{\theta}_r) + \frac{c_1}{2}\left\|\mathbf{W}_r\mathbf{L}_r\mathbf{x}_r - \boldsymbol{\theta}_r\right\|_2^2 + \boldsymbol{\pi}_r^T(\tilde{\mathbf{Y}}_r\mathbf{x}_r) \\ & + \frac{c_2}{2}\left\|\tilde{\mathbf{Y}}_r\mathbf{x}_r\right\|_2^2 + \sum_{s\in\mathscr{A}_r}\frac{c_3}{2}\left\|\mathbf{M}_{rs} - \boldsymbol{c_{rs}}\right\|_2^2 + \boldsymbol{\kappa}_{\boldsymbol{rs}}^T(\mathbf{M}_{rs} - \boldsymbol{c_{rs}})\Bigg] \end{aligned} \tag{5.21}$$

with $\boldsymbol{\mu}_r$, $\boldsymbol{\pi}_r$ and $\boldsymbol{\kappa_{rs}}$ standing for the vectors of Lagrange multipliers, and c_1, c_2, c_3 denoting pre-defined positive constants. The ADMM iterates for the updates of the primal and dual variables yield:

$$\mathbf{x}_r^{\nu+1} = \arg\min_{\mathbf{x}_r} \mathscr{L}\big(\mathbf{x}_r, \boldsymbol{c}_{\boldsymbol{rs}}^{\nu}, \boldsymbol{\theta}_r^{\nu}, \boldsymbol{\mu}_r^{\nu}, \boldsymbol{\pi}_r^{\nu}, \boldsymbol{\kappa}_{\boldsymbol{rs}}^{\nu}\big) \tag{5.22}$$

$$\boldsymbol{c}_{\boldsymbol{rs}}^{\nu+1} = \arg\min_{\boldsymbol{c_{rs}}} \mathscr{L}\big(\mathbf{x}_r^{\nu+1}, \boldsymbol{c_{rs}}, \boldsymbol{\theta}_r^{\nu}, \boldsymbol{\mu}_r^{\nu}, \boldsymbol{\pi}_r^{\nu}, \boldsymbol{\kappa}_{\boldsymbol{rs}}^{\nu}\big) \tag{5.23}$$

$$\boldsymbol{\theta}_r^{\nu+1} = \arg\min_{\boldsymbol{\theta}_r} \mathscr{L}\big(\mathbf{x}_r^{\nu+1}, \boldsymbol{c}_{\boldsymbol{rs}}^{\nu+1}, \boldsymbol{\theta}_r, \boldsymbol{\mu}_r^{\nu}, \boldsymbol{\pi}_r^{\nu}, \boldsymbol{\kappa}_{\boldsymbol{rs}}^{\nu}\big) \tag{5.24}$$

$$\boldsymbol{\mu}_r^{\nu+1} = \boldsymbol{\mu}_r^{\nu} + c_1\big(\mathbf{W}_r\mathbf{L}_r\mathbf{x}_r^{\nu+1} - \boldsymbol{\theta}_r^{\nu+1}\big) \tag{5.25}$$

$$\boldsymbol{\pi}_r^{\nu+1} = \boldsymbol{\pi}_r^{\nu} + c_2\big(\tilde{\mathbf{Y}}_r\mathbf{x}_r^{\nu+1}\big). \tag{5.26}$$

$$\boldsymbol{\kappa}_{\boldsymbol{rs}}^{\nu+1} = \boldsymbol{\kappa}_{\boldsymbol{rs}}^{\nu} + c_3\big(\mathbf{M}_{rs} - \boldsymbol{c}_{\boldsymbol{rs}}^{\nu+1}\big). \tag{5.27}$$

One can easily prove that the closed form of the sequential updates of the primal variables at iteration ν can be written as

$$\begin{aligned}\mathbf{x}_r^{\nu+1} = \Bigg(&\mathbf{H}_r^T\mathbf{H}_r + c_1\mathbf{L}_r^T\mathbf{W}_r^T\mathbf{W}_r\mathbf{L}_r + c_2\tilde{\mathbf{Y}}_r{}^T\tilde{\mathbf{Y}}_r \\ &+ \sum_{s\in\mathscr{A}_r}\mathbf{M}_{rs}^T\mathbf{M}_{rs}\Bigg)^{-1} \\ \Bigg(&\mathbf{H}_r^T\mathbf{y}_r - \mathbf{L}_r\mathbf{W}_r^T\boldsymbol{\mu}_r^\nu + c_1\mathbf{L}_r^T\mathbf{W}_r^T\boldsymbol{\theta}_r{}^\nu - \tilde{\mathbf{Y}}_r\boldsymbol{\pi}_r^\nu \\ &+ \sum_{s\in\mathscr{A}_r}\left(\mathbf{M}_{rs}^T\boldsymbol{c_{rs}} - \boldsymbol{\kappa_{rs}}\mathbf{M}_{rs}^T\right)\Bigg)\end{aligned} \tag{5.28}$$

$$\boldsymbol{c_{rs}}^{(\nu+1)} = \frac{\mathbf{M}_{rs}\mathbf{x}_r^{(\nu+1)} + \mathbf{M}_{sr}\mathbf{x}_s^{(\nu+1)}}{2} + \frac{\boldsymbol{\kappa}_{\boldsymbol{rs}}^{(\nu)} + \boldsymbol{\kappa}_{\boldsymbol{sr}}^{(\nu)}}{2c_3} \tag{5.29}$$

$$\boldsymbol{\theta}_r{}^{\nu+1} = \mathscr{S}_{\lambda/c_1}\left(\mathbf{W}_r\mathbf{L}_r\mathbf{x}_r^{\nu+1} + c_1^{-1}\boldsymbol{\mu}_r^\nu\right) \tag{5.30}$$

with $S_a(\chi)$ standing for the soft thresholding operator, that is,

$$S_a(\chi) = \begin{cases}\chi + a & \text{if } \chi < -a; \\ 0 & \text{if } |\chi| \leq a; \\ \chi - a & \text{if } \chi > a.\end{cases} \tag{5.31}$$

5.5 μPMU Placement (μPP)

As already mentioned (Sections 5.1 and 5.3), in the case of DG SE the system observability can be achieved using a large number of power injection measurements. These values are either pseudomeasurements with a noise standard deviation σ on the order of 10^{-1} (i.e. high) [101], or virtual measurements, such as zero power injections, with σ on the order of 10^{-5} (i.e. low). Unfortunately, as discussed earlier, disparity in the quality of measurements results into an ill-conditioned gain matrix $\mathbf{G}$ [81] that, in turn, affects the SE accuracy [97].

The degree to which $\mathbf{G}$ is ill-conditioned is given by the condition number κ, that is:

$$\kappa = \|\mathbf{G}\| \left\|\mathbf{G}^{-1}\right\| \tag{5.32}$$

or equivalently:

$$\kappa = \lambda_{max}(\mathbf{G})/\lambda_{min}(\mathbf{G}) \tag{5.33}$$

where λ_{max} and λ_{min} stand for the maximum and minimum eigenvalue of $\mathbf{G}$, respectively.

Bearing all the above in mind, the purpose of this section is to propose a μPMU placement method that enhances robustness to the R-NESE estimator. Specifically, the

target is, first to construct a full-rank $\mathbf{G}$ and, more importantly, to optimize on the condition number κ.

First, let us consider a number of o μPMUs for installation and a binary decision vector $\mathbf{q} = (q_1, ..., q_N)^T$, where each element q_i for $i = 1...N$ denotes the presence (1) or absence (0) of a μPMU at bus i. With respect to the decision vector $\mathbf{q}$, the gain matrix $\mathbf{G}$ can be written as follows [65]:

$$\mathbf{G}(\mathbf{q}) = \left[\mathbf{G}_0 + \sum_{i=1}^{N} q_i \mathbf{G}_i\right] \tag{5.34}$$

where $\mathbf{G}_i = \mathbf{J}_i^T \mathbf{R}_i^{-1} \mathbf{J}_i$ is the gain matrix resulting for the inclusion of each candidate μPMU, for $i = 1...N$ and $\mathbf{G}_0 = \mathbf{J}_0^T \mathbf{R}_0^{-1} \mathbf{J}_0$ is the initial gain matrix generated from legacy measurements only. Both $\mathbf{G}_0$ and $\mathbf{G}_i$ for $i = 1...N$ are evaluated at $\mathbf{x}_0$, that is, the initial value of the state vector[4] in (5.3). The power system is said to be numerically observable when $\mathbf{G}(\mathbf{q}) \succ \mathbf{0}$.

Based on all the above, the μPMU placement problem (μPP) can be posed as the following MISDP optimization program:

$$\min_{\mathbf{q},\, \xi,\, \omega} \quad \xi - \omega \tag{5.35a}$$

$$s.t. \quad \omega \boldsymbol{I} \preceq \mathbf{G}(\mathbf{q}) \preceq \xi \boldsymbol{I} \tag{5.35b}$$

$$\omega_0 \leq \omega \leq \xi \leq \xi_0 \tag{5.35c}$$

$$\sum_{i=1}^{|\mathscr{N}|} q_i = o \tag{5.35d}$$

$$\forall r \in \mathscr{A}, \quad \forall i \in \mathscr{N}_r \quad \sum_{i=1}^{|\mathscr{N}_r|} q_i = o_r \quad \mathscr{N}_r \subseteq \mathscr{N} \tag{5.35e}$$

$$\forall i \in \mathscr{N} \setminus \{1\} \quad (q_i + q_{i-1}) \leq 1 \tag{5.35f}$$

$$\mathbf{q} \in \{0,1\}^N \tag{5.35g}$$

The first constraint in the μPP problem, (5.35b), can be interpreted as two separate linear matrix inequalities. The first, reads $\mathbf{G}(\mathbf{q}) - \omega \boldsymbol{I} \succeq \mathbf{0}$ and guarantees an invertible positive definite $\mathbf{G}(\mathbf{q})$ where ω stands for an auxiliary (real valued) variable [5]. The role of ω is to introduce a constraint on the minimum eigenvalue of $\mathbf{G}(\mathbf{q})$ and satisfies $\omega > \omega_0$ where $\omega_0 >> 0$ denotes an appropriately chosen constant[5]. The second linear matrix inequality, $\xi \boldsymbol{I} - \mathbf{G}(\mathbf{q}) \succeq \mathbf{0}$, follows the same rationale where the auxiliary variable ξ introduces a constraint on the maximum eigenvalue of $\mathbf{G}(\mathbf{q})$. The ultimate goal is to

[4]Typically, $\mathbf{x}_0$ is set as a flat voltage profile, i.e., all voltage magnitudes are set to 1 and all angles to 0.

[5]The constants ω_0 and ξ_0 can be chosen based on the λ_{min} and λ_{max} of the matrices $\mathbf{G}_0$ and $\mathbf{G}_i$ for $i = 1...N$ [112].

minimize the difference $\xi - \omega$ with respect to the binary decision vector $\mathbf{q}$ in equation (5.35g), that, in turn, minimizes the condition number κ.

The rest of the constraints are related with the number and the location of the μPMUs. Constraint (5.35d) refers to the total constant number o of μPMUs to be placed based on different criteria, such as economic budget and communication feasibility study [14]. Constraint (5.35e) aims to set a requirement on the minimum number of meters o_r to be placed at the buses of each area r (see subsection 5.3.2.1). The last constraint, (5.35f), is in charge of preventing the placement of two meters in adjacent buses (based on the fact that a phasor meter at a bus makes also observable its adjacent branches).

5.6 Numerical Results

The performance of the presented regularized SE scheme has been numerically assessed on the 95-bus UK distribution grid, Fig. 5.2 [9]. This system has a nominal voltage of 11 kV and the installed active/reactive load is 3.83 MW/0.95 MVar, respectively. As Figure 5.2 illustrates, we have considered four PV plants as DER. Their nominal capacity is 0.8, 0.8, 0.5 and 0.4 MW and they are connected to buses 15, 36, 68 and 91, respectively. Concerning observations, we have considered a set of legacy measurements with redundancy $L/N \approx 1.1$. This includes: (i) 5 voltage measurements located at the substation and the DER buses; (ii) 3 pairs of branch power flows $P_{i,j}$, $Q_{i,j}$ located at the substation; (iii) 4 squared current magnitudes at the feeders $|I_{i,j}|^2$; and (iv) 90 pairs of power injections P_i,Q_i, out of which 40 are pseudomeasurements, and 50 are zero power injections (virtual measurements). Following the rationale of similar works [123], each type of the measurement subsets is corrupted with independent and identical distributed (i.i.d.) Gaussian noise with standard deviations (σ) 10^{-3}, $2 \cdot 10^{-3}$, $4 \cdot 10^{-3}$, $3 \cdot 10^{-2}$, and $2 \cdot 10^{-6}$ per-unit (p.u.) [117], respectively. In addition, 12 μPMU measurements are incorporated, located according to the μPP method proposed in Section 5.5, with $\sigma = 10^{-4}$ p.u. and $2 \cdot 10^{-4}$ rad for voltage magnitude and angle, respectively [39]. Unless otherwise stated, computer simulation results are averaged over $N_r = 5000$ realizations of the noise, load and PV generation profiles. The loads have been assumed to be Gaussian-distributed with $P_m \sim \mathscr{N}(P_m, 0.06)$, $Q_m \sim \mathscr{N}(Q_m, 0.06)$ for $m = 1, 2, ..N \in \mathscr{V}$. For the power flow analysis, we have used MATPOWER 5.1 [90]. Results have been verified through CVX [124]. For the MISDP μPP problem solution, we have used the optimization software YALMIP[6] [125] and specifically the solver SEDUMI.

[6]The binary decision vector $\mathbf{q}$ renders the μPP problem computationally intensive. However, the applied solvers, using a branch and bound framework, provide, if not the global optimal, an efficient local optimal solution for small and medium size applications (as in the case of DG feeders). Another approach is to solve the convex counterpart of the μPP problem where the binary constraint $\mathbf{q} \in \{0, 1\}$ is relaxed to $\mathbf{q} \in [0, 1]$. In that case, the decision vector will consist of real-valued elements; for that reason a rounding procedure and a local search should follow. Computer simulations reveal that both

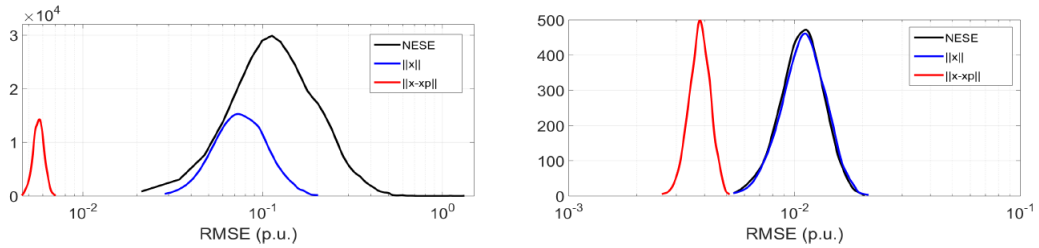

Figure 5.6: PDF of the RMSE without (left) and with (right) μPMUs. ($\sigma_1 = 0.03$ for pseudomeasurements)

5.6.1 R-NESE

First of all, we focus on the performance of the R-NESE scheme (5.10). To recall, it operates at time instants kT, when *both* legacy and μPMU measurements are available. The performance of the proposed scheme has been numerically assessed in a variety of scenarios. First, we have considered two cases with different error level in the pseudomeasurements, i.e., $\sigma_1 = 0.03$ and $\sigma_2 = 0.15$ p.u.. For each case, we present two scenarios, namely, with and without μPMUs. For the scheme with μPMUs, the measurement configuration with 12 meters proposed in subsection 5.6.3 has been incorporated. Further, in order to demonstrate explicitly the advantages provided by the regularization term in (5.4), the following benchmarks have been used: (i) the NESE scheme solving problem (5.3) and (ii) the R-NESE scheme solving problem (5.10) with $\mathbf{x}_p = 0$ (i.e. without information on the last state estimate). Finally, in all realizations $\mathbf{x}_p$ refers to the last state estimate obtained with the (D)-WTVSE scheme at the time instant $kT + nt$ (see Fig. 5.1 and subsection 5.6.2) with an additional error of $\sigma = 0.01$ p.u. in order to capture further uncertainty.

Figures 5.6 and 5.7 illustrate the probability density functions (PDFs), which have been estimated from the histograms of the realizations, and the RMSE (p.u.) averaged over the 5000 realizations, respectively. With respect to the PDFs, by definition the area below the curves is equal to 1. Clearly, NESE reaches high values of error because of the high percentage of pseudomeasurements ($100/205 = 48.8\%$) among the low redundant measurement set. With R-NESE and $\mathbf{x}_p = 0$, the accuracy has been drastically increased, however, the gain is moderate compared to R-NESE with $\mathbf{x}_p$ equal to the last (D)-WTVSE estimate. The proposed approach attains the highest accuracy, i.e., 5.7×10^{-3} p.u. (Table 5.1), which is one and two orders of magnitude lower compared to the benchmarks. In contrast with other works [46], the scheme presents robustness against the high uncertainty of the pseudo-load injections. The simulations reveal that for all buses the absolute error for the voltage magnitude stays below 1×10^{-2} p.u.. In-

approaches yield similar solutions.

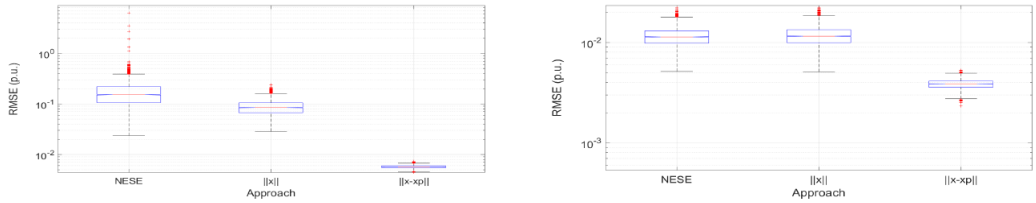

Figure 5.7: RMSE for the 95-bus DG and $\sigma_1 = 0.03$ p.u. without (left) and with (right) μPMUs.

Table 5.1: RMSE and condition number κ of $\mathbf{G}$ for each approach, without (top) and with (bottom) μPMUs. ($\sigma_1 = 0.03$ for pseudomeasurements)

Approach	RMSE (p.u.)	κ
NESE	1.84×10^{-1}	8.8×10^{12}
R-NESE with $\tau \|\mathbf{x}\|_2^2$	8.86×10^{-2}	3.62×10^{10}
R-NESE with $\tau \|\mathbf{x} - \mathbf{x}_p\|_2^2$	5.7×10^{-3}	4.41×10^{6}

Approach	RMSE (p.u.)	κ
NESE	1.15×10^{-2}	6×10^{7}
R-NESE with $\tau \|\mathbf{x}\|_2^2$	1.18×10^{-2}	5.9×10^{7}
R-NESE with $\tau \|\mathbf{x} - \mathbf{x}_p\|_2^2$	3.9×10^{-3}	4.1×10^{6}

terestingly, we also observe a significant improvement on the condition number κ of the gain matrix $\mathbf{G}$, which is related with the accuracy and convergence rate of the algorithm. Based on this improvement and the low computational complexity of the regularized normal equations (see Appendix C), a relatively low convergence time is needed for the proposed scheme. Concerning the second scenario (Figs. 5.6 and 5.7, right), where μPMUs have been included in the measurement set, the proposed approach, R-NESE, presents again a superior numerical behavior. However, with regard to these graphs, two further observations can be made: (i) as expected, synchrophasor measurements improve the accuracy for all approaches (Table 5.1); and (ii) interestingly, R-NESE with $\mathbf{x}_p = 0$ exhibits a slight performance degradation compared to NESE. The latter can be explained by the SVD analysis of the Tikhonof solution (5.10) [119]. More precisely, according to the chosen value of τ, specific contributions to $\hat{\mathbf{x}}$ that correspond to the lower eigenvalues of $\mathbf{G}$ will be filtered out. However, for this specific scenario, as the conditioning of $\mathbf{G}$ has been improved (due to the μPMUs), the optimal value of τ (see below) dominates further eigenvalues that compose profitable information.

Additionally, Fig. 5.8 shows the RMSE for *each* state variable in polar form averaged over the realizations. As shown on the left graph, the R-NESE presents high accuracy for the angle estimates. The maximum RMSE is below 10^{-2} rad. At the same time, the RMSE of the majority of the bus phase angles lies within the interval of $3\times10^{-3}-5\times10^{-3}$

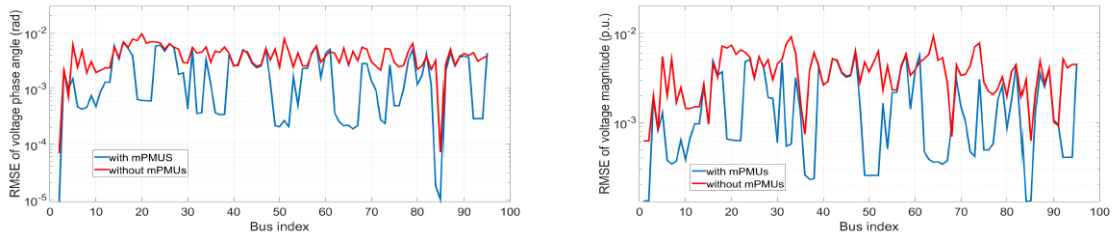

Figure 5.8: Average RMSE of R-NESE for each state variable with $\sigma_1 = 0.03$ p.u., with and without μPMUs. The left graph illustrates the voltage phase angles and the right graph the voltage magnitudes.

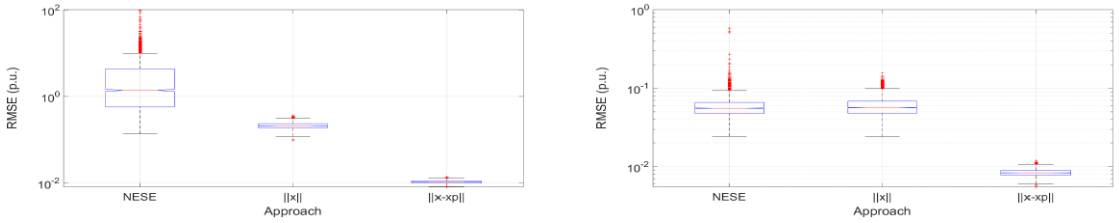

Figure 5.9: RMSE for the 95-bus DG and $\sigma_2 = 0.15$ p.u. without (left) and with (right) μPMUs.

rad. In contrast, for the second scenario where μPMUs have been incorporated, the majority of voltage angles RMSE stays below 10^{-3} rad and maximum error reads 6×10^{-3} rad. Important to note that, in the buses (and the respective neighboring ones) where the phasor meters have been positioned (e.g. 21, 50, 73, 84) the RMSE is substantially lower. The same behavior is observed for both scenarios in the case of voltage magnitudes.

Next, Fig. 5.9 shows the results for the case where an increased error level is considered for the pseudomeasurements, i.e., $\sigma_2 = 0.15$ p.u.. First, one observes an accuracy degradation for all approaches, Table 5.2. It is worth mentioning that, for the two benchmarks, the RMSE reaches high values (e.g., 3.28 p.u. for NESE). As it was expected, the high error deviation measurements lead to an ill-conditioned matrix $\mathbf{G}$ ($\kappa = 5.39\times10^{16}$), that in turn, leads many of the realizations to converge to a local minimum or not to converge at all. On the contrary, R-NESE exhibits a remarkable performance with RMSE equal to 1.1×10^{-2} p.u. without μPMUs and 8.4×10^{-3} p.u. with μPMUs.

Again, Fig. 5.10 presents the RMSE for each state variable. As expected, RMSE values are higher now. For instance, the majority of the RMSE for the bus phase angles lies within the interval $4\times10^{-3}-8\times10^{-3}$ rad. For the case with μPMUs this error interval is $1\times10^{-3}-5\times10^{-3}$ rad. However, in both angles and voltage magnitudes the R-NESE scheme, still, attains high accuracy despite the higher error of the pseudomeasurements.

Table 5.2: RMSE and condition number κ of $\mathbf{G}$ for each approach, without (top) and with (bottom) μPMUs, $\sigma_2 = 0.15$ for pseudomeasurements, subsection 5.6.1.

Approach	RMSE (p.u.)	κ
NESE	3.28	5.39×10^{16}
R-NESE with $\tau \|\mathbf{x}\|_2^2$	0.21	4×10^{9}
R-NESE with $\tau \|\mathbf{x} - \mathbf{x}_p\|_2^2$	1.1×10^{-2}	4.42×10^{6}

Approach	RMSE (p.u.)	κ
NESE	8.08×10^{-2}	9.7×10^{9}
R-NESE with $\tau \|\mathbf{x}\|_2^2$	5.96×10^{-2}	4.65×10^{8}
R-NESE with $\tau \|\mathbf{x} - \mathbf{x}_p\|_2^2$	8.4×10^{-3}	4.4×10^{6}

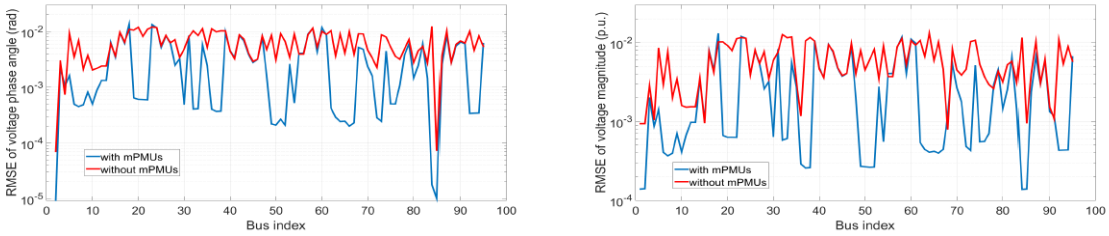

Figure 5.10: Average RMSE of R-NESE for each state variable with $\sigma_2 = 0.15$ p.u., with and without μPMUs. The left graph illustrates the voltage phase angles and the right graph the voltage magnitudes.

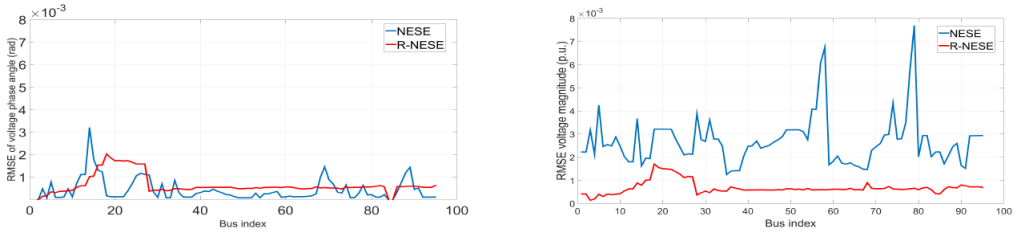

Figure 5.11: Average RMSE of each voltage phase angle θ_i (left) and voltage magnitude V_i (right) for $i = 1...N$, adopting an alternative approach on measurement accuracy for R-NESE and NESE.

Besides, we provide a comparison between the proposed R-NESE and NESE schemes for a scenario where different accuracies are used for each mesurement subset. Specifically, for each type of measurement we have considered the following accuracy: (i) 1% for voltage measurements, (ii) 3% for branch power flows and current magnitudes, (iii) 50% and 0.02% for pseudomeasurements and zero power injections, respectively, and (iv) 1% for the magnitude and 0.01 rad for the angles in the case of μPMUs. The standard deviation σ has been assumed as equal to a third of the accuracy value in each case. Figure 5.11 illustrates the RMSE for each state element $\hat{x}_i$ averaged, as before, over 5000 realizations, considering the same measurement set. The left graph reveals that NESE reaches a maximum RMSE value of 3×10^{-3} rad, presenting also additional error peaks in a number of buses. In contrast, the maximum RMSE for R-NESE is 2×10^{-3} rad, while the rest of the estimated angles present low error. In addition, with regard to the voltage magnitudes, R-NESE outperforms significantly in terms of accuracy. On the one hand, NESE reaches an RMSE at the level of $3 - 4 \times 10^{-3}$ with a maximum value of 8×10^{-3} p.u. On the other, R-NESE presents a maximum RMSE of 1.5×10^{-3} p.u. while the majority of the estimated voltage magnitudes error stays below 5×10^{-4} p.u. To conclude, independently of the adopted rationale on the measurement noise, the computer simulations have shown the same trend. That is, the proposed R-NESE attains high accuracy even in scenarios with low redundant measurement sets which include a large number of high uncertainty pseudomeasurements.

Moreover, we provide an additional comparison between the proposed R-NESE and the NESE scheme. With respect to the NE-SE scheme, instead of the commonly used flat start, i.e., $V_i = 1$ p.u. and $\theta = 0$ rad for $i = 1...N$, as initial condition is exploited the last available estimated state, i.e. $\hat{\mathbf{x}}_0 = \mathbf{x}_p$. This can be considered as a *hot start* for NE-SE. The computer simulation scenario considers the 95-bus DG and the measurement set already described, including μPMUs. Figure 5.12 shows the RMSE of each state variable $\hat{x}_i$ for both cases. As the graph illustrates, the R-NESE outperforms in terms of accuracy. More specifically, the left graph shows that the RMSE for all voltage phase angles stays

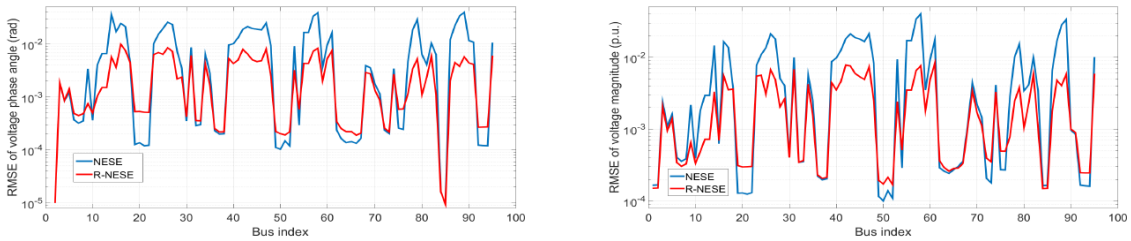

Figure 5.12: Average RMSE for each voltage phase angle θ_i (left) and voltage magnitude V_i (right) for $i = 1...N$ with μPMUs and $\sigma = 0.03$ p.u. for R-NESE and NESE.

below 1×10^{-2} rad. In contrast, in the case of NESE the maximum RMSE reaches the value of 5×10^{-2} rad and, a large number of estimated angles reach a similar level of error. The same applies for the voltage magnitudes (right graph) where the results are quite similar. Based on the above numerical behavior of the compared algorithms, the following conclusion can be drawn. Although the NESE algorithm starts iterations from a point close to the optimal one (i.e., $\mathbf{x}_p$), the bad conditioning of the optimization problem leads with high probability the algorithm to a local minima.

Finally, we discuss on how to determine the control parameter τ, which establishes the trade-off between regularization and data fidelity. Different methods exist in the literature, such as the generalized cross validation and the L-curve method [121]. The latter appears to provide robustness with regard to the control parameter estimation and consequently, we have adopted this approach. In Fig. 5.13 we provide the L-curve for an instance (specific iteration) of R-NESE (with μPMUs and $\sigma_1 = 0.03$ p.u.). This method constitutes a graphical tool depicting the error of the regularization term $\|\mathbf{x} - \mathbf{x}_p\|_2^2$ versus the error of the residual term $(\mathbf{z} - \mathbf{h}(\mathbf{x}))^T \mathbf{R}_\mathbf{e}^{-1} (\mathbf{z} - \mathbf{h}(\mathbf{x}))$. In particular, in every iteration of Algorithm 1, the two terms are computed for a number of different τ^ν values. As Fig. 5.13 illustrates, these values correspond to different points of the graph and the common method is to choose the one that falls within the maximum curvature, e.g., 10^1.

5.6.2 D-WTVSE

Here, we investigate the numerical behavior of the D-WTVSE scheme, which operates at the intermediate time instants, that is, between kT and $(k+1)T$. To recall, all simulations have been averaged over 5000 realizations of different noise and load injections. First, the impact of the control parameter λ on the algorithmic performance of WTVSE is illustrated. For this computer simulation the impact of ILVRs has not been taken into consideration. Fig. 5.14 presents the convergence and the RMSE for a number of different parameter values within the range $10^{-7} - 10^{-1}$. The graph shows that lower values are related with slower convergence rate, as λ controls the amount of regularization. For

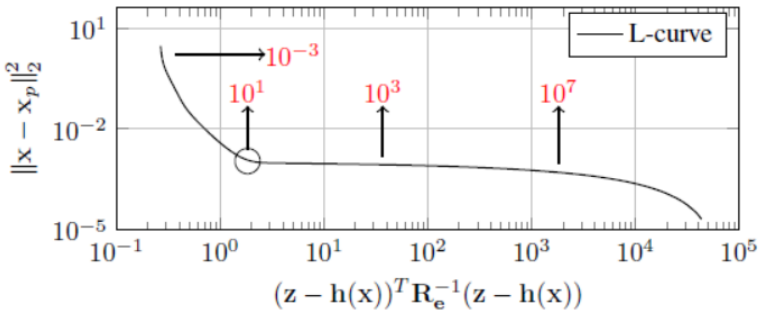

Figure 5.13: The L-curve that depicts the error of the regularization term (y-axis) versus the error of the corresponding residual term (x-axis) for R-NESE with $\tau^{(\nu)} = 10^{-3} - 10^{8}$.

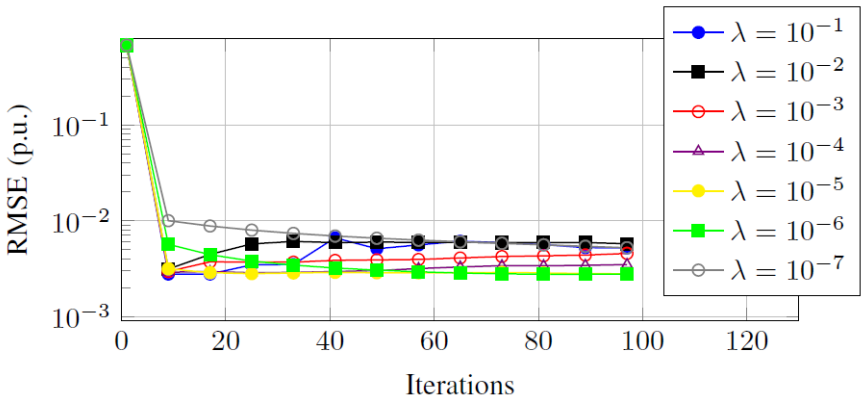

Figure 5.14: RMSE vs. iterations for the WTVSE ($c_1 = 5 \times 10^{-4}$, $c_2 = 10^{-5}$, $c_3 = 10^{-4}$).

instance, for $\lambda = 10^{-6}$ the algorithm attains the highest accuracy. However, more than 50 iterations are needed to converge. In contrast, for the case of $\lambda = 10^{-4}$ the scheme attains the same accuracy within 15-20 iterations. Exhaustive simulations with the trial-end-error method have shown that a parameter value within the range of $10^{-5} - 10^{-4}$ balances the trade-off between accuracy and convergence speed.

Further, Figs. 5.15 and 5.16 depict the convergence behavior and accuracy for the D-WTVSE optimization problem solved via ADMM. Here, just $K = 24$ measurements (voltage real and imaginary parts) are available from the μPMU configuration proposed in subsection 5.6.3 (Table 5.6). In the examined scenario, we have considered the presence of two ILVRs, providing 5% voltage regulation and positioned at the branches $\{23 - 24\}$ and $\{54 - 75\}$ (Fig. 5.4). Figure 5.15 indicates that, for this scenario, the program has converged for all areas within maximum 10-15 iterations reaching at the same time high accuracy. Moreover, the fact that the positive constants c_1, c_2, c_3 have been pre-defined improves significantly the convergence time of D-WTVSE (see Appendix C). Clearly, as

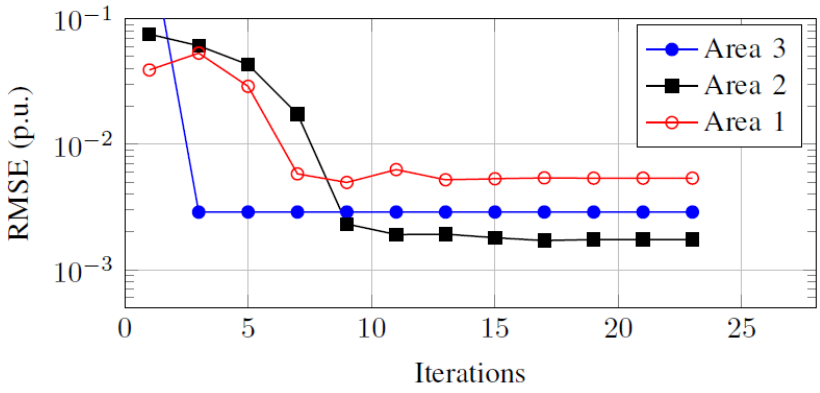

Figure 5.15: RMSE vs. iterations for the D-WTVSE scenario ($\lambda = 10^{-4}$, $c_1 = 5 \times 10^{-4}$, $c_2 = 1 \times 10^{-5}$, $c_3 = 1 \times 10^{-4}$).

the next figure shows (Fig. 5.16), the estimated magnitudes and angles approach with high precision the actual values, presenting a slight variance in a small number of buses. In both cases of ILVRs, $\{23 - 24\}$ and $\{54 - 75\}$, the estimated voltages exhibit high accuracy. However, with proper adjustment of the positive constants c_1 and c_3 (5.21), a trade-off can be established between ILVR-bus variable consensus and voltage variance limitation.

In addition, for the same scenario the proposed D-WTVSE is compared with WTVSE. To do so, WTVSE has been modified in order to incorporate the ILVR models based on the augmented admittance matrix approach [111]. Further, the regularization term $\|\mathbf{WLx}\|_1$ had to be modified accordingly. Interestingly, the proposed D-WTVSE presents higher accuracy, i.e. 4.8×10^{-3} p.u., compared to the WTVSE which reaches the value of 5.6×10^{-3} p.u.. In addition to that, D-WTVSE is able to converge in 0.0303 s when at the same time WTVSE needs 0.0395 s. The superiority of the proposed scheme is also illustrated in Fig. 5.17, which shows the error of voltage magnitude for both cases. Specifically, the augmented WTVSE attains slightly higher accuracy at the corresponding ILVR buses. However, this is not the case for the rest of the nodal estimated voltages. In general, the above analysis reveals that, by exploiting the proposed D-WTVSE scheme instead of the augmented WTVSE, performance degradation is avoided.

5.6.2.1 D-WTVSE with DER and Load Diversity

Finally, the efficiency of the regularization term $\|\mathbf{WLx}\|_1$, which renders the D-WTVSE problem solvable, is numerically assessed for three different cases:

- Case 1: A baseline case, where we consider the nominal data of the 95-bus UK DG [9].

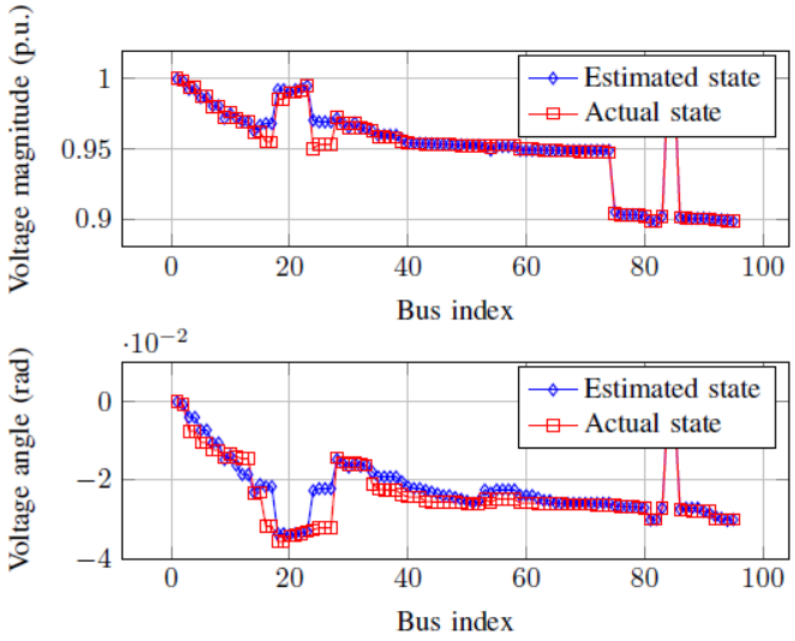

Figure 5.16: Actual and estimate voltage (magnitude and angle) for each bus with the D-WTVSE ($\lambda = 10^{-4}$, $c_1 = 5 \times 10^{-4}$, $c_2 = 1 \times 10^{-5}$, $c_3 = 1 \times 10^{-4}$).

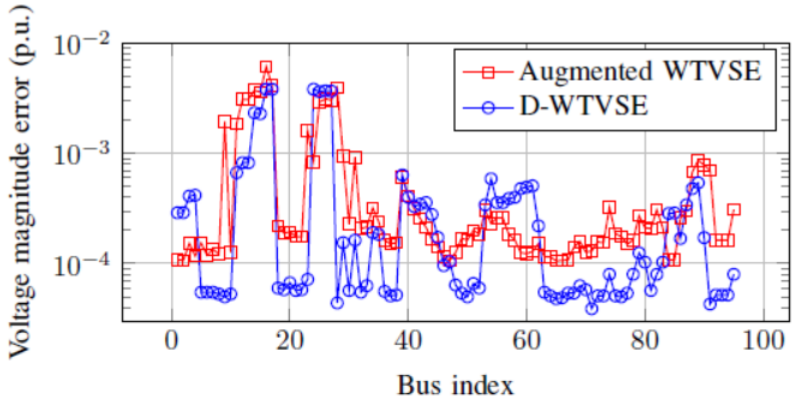

Figure 5.17: Absolute voltage magnitude error for each bus for the D-WTVSE and the augmented WTVSE.

Table 5.3: Addtional Active (MW) and Reactive (MVar) loads.

Bus	P (MW)	Q (MVar)
12	0.6	0.05
27	0.7	0.05
46	0.9	0.1
59	0.7	0.08
27	0.65	0.05
46	0.95	0.15

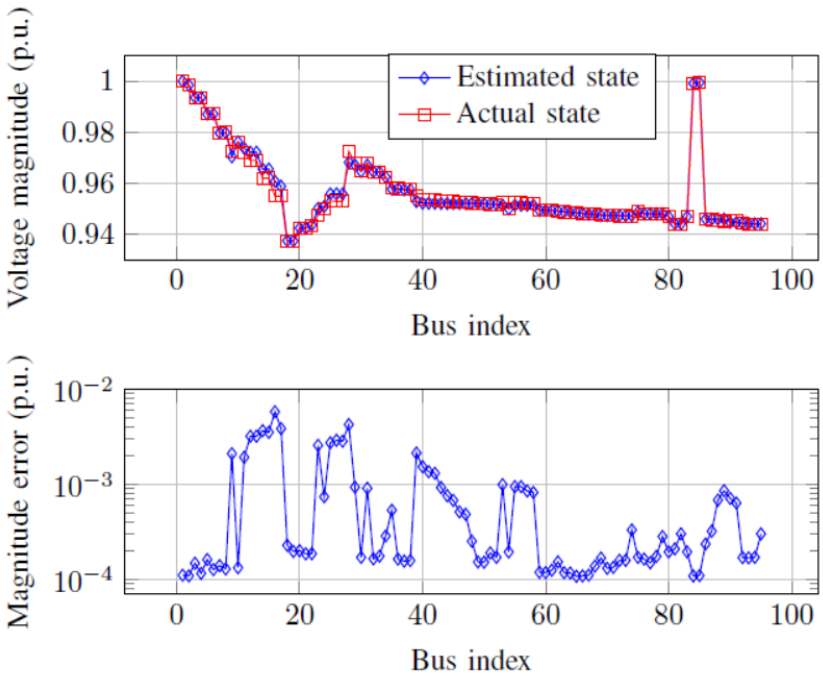

Figure 5.18: Actual vs estimated voltage magnitude (top) and absolute voltage magnitude error (bottom) for scenario (i) ($\lambda = 10^{-4}$, $c_1 = 5 \times 10^{-4}$, $c_2 = 1 \times 10^{-5}$, $c_3 = 1 \times 10^{-4}$).

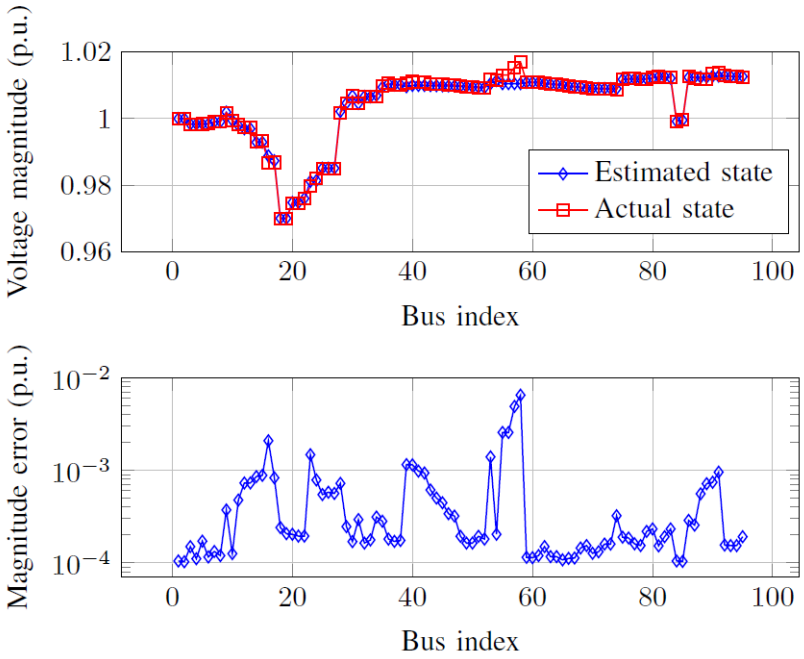

Figure 5.19: Actual vs estimated voltage magnitude (top) and absolute voltage magnitude error (bottom) for scenario (ii) ($\lambda = 10^{-4}$, $c_1 = 5 \times 10^{-4}$, $c_2 = 1 \times 10^{-5}$, $c_3 = 1 \times 10^{-4}$).

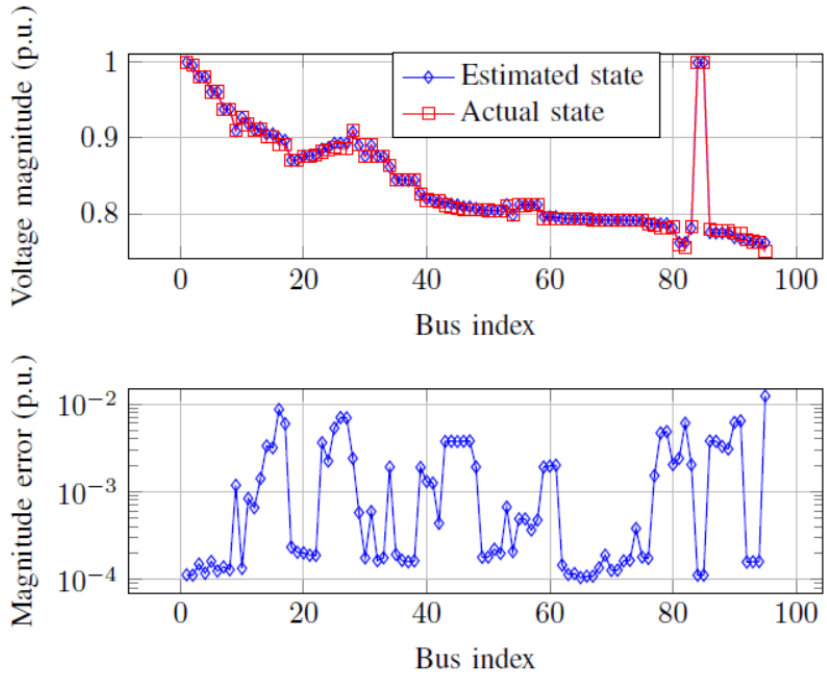

Figure 5.20: Actual vs estimated voltage magnitude (top) and absolute voltage magnitude error (bottom) for scenario (iii) ($\lambda = 10^{-4}$, $c_1 = 5 \times 10^{-4}$, $c_2 = 1 \times 10^{-5}$, $c_3 = 1 \times 10^{-4}$).

- Case 2: A scenario with PV generation (i.e., PV connected to the buses 15, 36, 68 and 91 with nominal capacity is 0.8, 0.8, 0.5 and 0.4 MW, respectively) added to the load profile.

- Case 3: A scenario without PV generation and with a number of additional high loads at the buses listed in Table 5.3. This last case can be regarded as a worst case scenario where adjacent buses with low/high loads (e.g., small/large number of households) absorb low/high currents.

Fig. 5.18 depicts the estimated and actual voltage magnitude profile (top) and the absolute voltage magnitude error (bottom) for Case 1. Here, the estimated voltage profile is really close to the actual one. The bottom graph shows that all absolute errors are below the value of 5.5×10^{-3} p.u.

In Case 2, Fig. 5.19, the error stays below 1×10^{-3} p.u. for the majority of voltage estimates. As expected, the high PV penetration has resulted in a smooth voltage profile (i.e., similar voltage magnitude in the majority of the buses) that, in turn, leads the total variation regularization term to superior performance.

Last, Fig. 5.20 shows the results for Case 3. The diversity between loads has resulted in a non-smooth real voltage profile. Accordingly, at the bottom graph we detect that the highest deviations with respect to the actual voltage magnitudes are located to the buses where the high loads have been positioned, e.g., buses 27, 46 and 59. Still, in this worst-case scenario, D-WTVSE is capable of providing a satisfactory state estimate where the absolute error of all the estimate voltages remain below 1×10^{-2} p.u.. More importantly, computer simulation results indicate that, the exploitation of the zero power injections and the weights introduced in the branches [12] significantly enhance the accuracy of D-WTVSE.

5.6.3 Numerical Assessment of μPP Algorithm

In this subsection, we evaluate the μPP method presented in Section 5.5. To do so, we consider three different MV DG systems, namely, a 33-bus, a 56-bus and the 95-bus UK DG that has been already considered for the SE computer simulations. In addition, in order to measure the impact of different μPMU configurations, the following metrics have been adopted: (i) the condition number κ of the gain matrix $\mathbf{G}_{sol}$ after the convergence of the SE problem (5.3); (ii) the $log(\kappa)$ and (iii) the average RMSE of the SE algorithm (5.3).

First, the proposed algorithm is evaluated considering the 33-bus DG. For this case, we have assumed a singular gain matrix $\mathbf{G}_0 \in \mathbb{R}^{66\times66}$ (i.e., non-observable system) and a limited number of 4/6/8 μPMUs to be placed. The table depicts the results for $\mathbf{G}_0$, two random μPMU placement configurations for comparison reasons ($\mathbf{G}_1$, $\mathbf{G}_2$) and the solution $\mathbf{G}_{sol}$ according to Section 5.5. Results (Table 5.4) indicate that the proposed

Gain matrix	Number of μPMUs	Buses	κ	$\log(\kappa)$	RMSE (p.u.)
$\mathcal{G}_0$	0	–	Inf	Inf	–
$\mathcal{G}_1$	4	$1, 9, 19, 29$	2.9×10^7	17.2	4.8×10^{-2}
$\mathcal{G}_2$	4	$2, 10, 19, 31$	2.4×10^7	17	4.4×10^{-2}
$\mathcal{G}_{sol}$	4	$2, 14, 26, 32$	2.3×10^6	14.8	2.9×10^{-2}

Gain matrix	Number of μPMUs	Buses	κ	$\log(\kappa)$	RMSE (p.u.)
$\mathcal{G}_0$	0	–	Inf	Inf	–
$\mathcal{G}_1$	6	$1, 6, 11, 19, 25, 33$	1.5×10^7	16.5	3.3×10^{-2}
$\mathcal{G}_2$	6	$1, 7, 12, 19, 26, 31$	6.3×10^6	15.7	2.5×10^{-2}
$\mathcal{G}_{sol}$	6	$2, 7, 11, 16, 24, 32$	1.3×10^6	14.1	1.9×10^{-2}

Gain matrix	Number of μPMUs	Buses	κ	$\log(\kappa)$	RMSE (p.u.)
$\mathcal{G}_0$	0	–	Inf	Inf	–
$\mathcal{G}_1$	8	$2, 5, 9, 14, 19, 22, 28, 33$	2.4×10^6	14.7	1.6×10^{-2}
$\mathcal{G}_2$	8	$2, 6, 11, 16, 18, 22, 26, 32$	1.2×10^6	14	1.1×10^{-2}
$\mathcal{G}_{sol}$	8	$1, 3, 11, 16, 20, 24, 26, 32$	3.5×10^5	12.8	7.3×10^{-3}

Table 5.4: Impact assessment of the μPP algorithm presented in Section 5.5 for the 33-bus test system and different number of μPMUs.

method has managed to provide a full-rank gain matrix. More importantly, the provided solution exhibits the best condition number (κ) and consequently, the highest SE accuracy. For all cases of different μPMU numbers and compared to $\mathbf{G}_1$ and $\mathbf{G}_2$, $\mathbf{G}_{sol}$ presents a condition number κ one order of magnitude below. For instance, comparing $\mathbf{G}_{sol}$ and $\mathbf{G}_1$ for the scenario of 6 μPMUs, Table 5.4 shows that the former presents an improved κ, i.e., 1.3×10^6 compared to 1.5×10^7, that corresponds to a decreased estimation error, i.e., 1.9×10^{-2} p.u. compared to 3.3×10^{-2} p.u.. A similar improvement of the numerical accuracy can been also observed for the other cases of μPMU numbers.

Further, we investigate the μPP algorithmic performance on the 56-bus DG, which includes a rural feeder [126]. In this case, the system includes a number of short, as well as long branches, which deteriorates the conditioning of $\mathbf{G}$ [81]. Table 5.5 indicates that the addition of a moderate number of synchrophasor meters, improves significantly the condition number κ of $\mathbf{G}_0$ (1.2×10^{10}). In particular, in the case of random placements $\mathbf{G}_1$ and $\mathbf{G}_2$, κ has been improved by two orders of magnitude. In contrast, the proposed method has reduced the condition number by three orders of magnitude, i.e., 3.6×10^7 compared to 1.2×10^{10}, something that translates into a lower estimation error, i.e., from 9.5×10^{-2} p.u. to 1.6×10^{-2} p.u..

Finally, for completeness, in Table 5.6 is presented the optimal solution for the 95-bus DG, which has been considered as the μPMU configuration in subsections 5.6.2 and 5.6.4. Again, the μPP method provides a gain matrix $\mathbf{G}_{sol}$ with an improved κ compared to $\mathbf{G}_0$, $\mathbf{G}_1$ and $\mathbf{G}_2$ (three and one orders of magnitude lower, respectively). Therefore, this results into a decreased estimation error, e.g., from 1.9×10^{-1} p.u. to 9×10^{-3} p.u..

Gain matrix	Number of μPMUs	Buses	κ	$\log(\kappa)$	RMSE (p.u.)
$\mathcal{G}_0$	0	–	1.2×10^{10}	23.2	9.5×10^{-2}
$\mathcal{G}_1$	8	5, 12, 20, 25, 34, 42, 50, 54	2.9×10^{8}	19.5	1.9×10^{-2}
$\mathcal{G}_2$	8	1, 10, 19, 26, 34, 43, 47, 56	9.4×10^{7}	18.4	1.8×10^{-2}
$\mathcal{G}_{sol}$	8	6, 13, 21, 25, 28, 39, 46, 50	3.6×10^{7}	17.4	1.6×10^{-2}

Table 5.5: Impact assessment of the μPP algorithm presented in Section 5.5 for the 56-bus test system.

Gain matrix	Number of μPMUs	Buses	κ	$\log(\kappa)$	RMSE (p.u.)
$\mathcal{G}_0$	0	–	3.4×10^{11}	26.56	1.9×10^{-1}
$\mathcal{G}_1$	12	2, 11, 18, 27, 33, 42, 54, 60, 68, 78, 86, 94	7.6×10^{8}	20.45	2×10^{-2}
$\mathcal{G}_2$	12	4, 10, 16, 27, 33, 42, 53, 60, 68, 78, 87, 95	5.2×10^{8}	20.1	$1,8 \times 10^{-2}$
$\mathcal{G}_{sol}$	12	1, 8, 21, 32, 37, 50, 64, 66, 73, 75, 84, 93	5.1×10^{7}	17.75	9×10^{-3}

Table 5.6: Impact assessment of the μPP algorithm presented in Section 5.5 for the 95-bus test system.

5.6.4 Full scheme: R-NESE and D-WTVSE

Finally, we evaluate the ability of the proposed SE scheme to track short time-scale voltage variations introduced from the stochastic behavior of the PV plants. The time-sequence diagram in Fig. 5.21 illustrates the operation of the two-time scale SE scheme based on the proposed R-NESE and D-WTVSE algorithms. The adopted sample of the *aggregated* power generation profile under consideration for the set of 4 PV plants is shown in Fig. 5.22. The profile refers to the PV generation on June 27, 2018, between 08:20 and 08:50 am at a MV feeder. The active power measurements have a resolution of 6 seconds. Further, for each active/reactive load-bus of the 95-bus UK DG a load pattern which refers to the same day and time period has been adopted, Fig. 5.23.

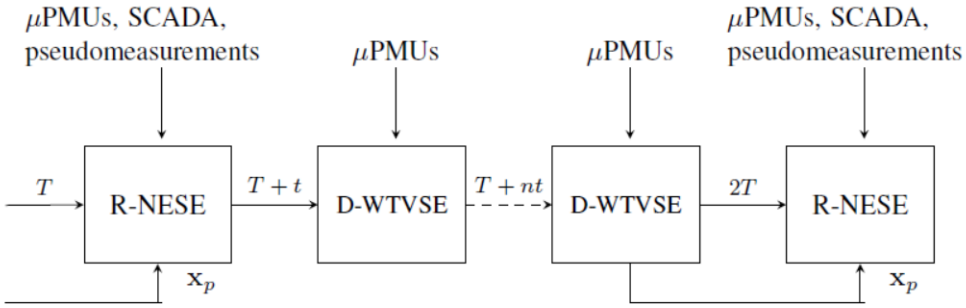

Figure 5.21: Time-sequence diagram of the proposed SE scheme based on the R-NESE and D-WTVSE algorithms.

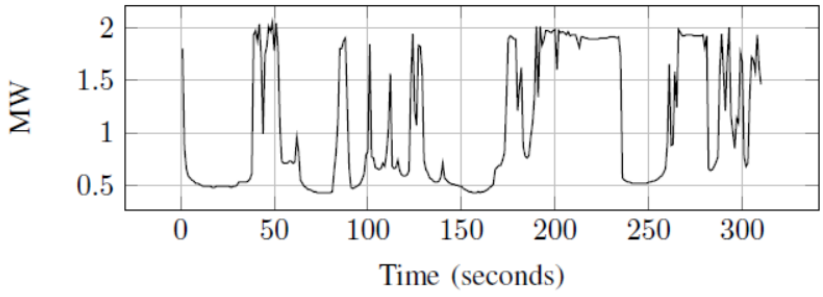

Figure 5.22: Aggregated power generation profile for the four PV plants. The data refer to June 27, 2018, between 08:20 and 08:50 am.

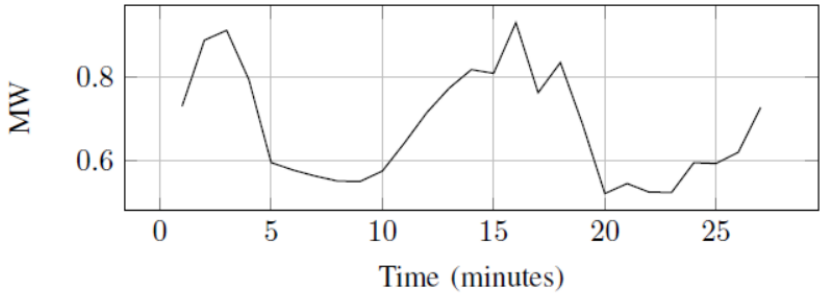

Figure 5.23: Active power load pattern for bus-19 of the 95-bus UK DG. The data refer to June 27, 2018, between 08:20 and 08:50 am.

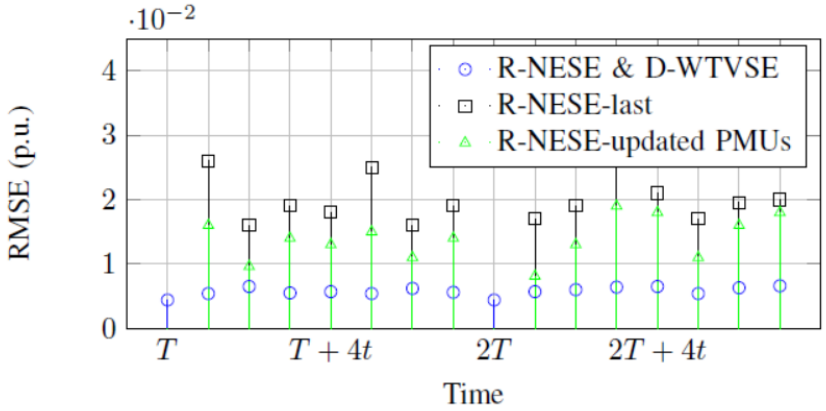

Figure 5.24: RMSE of the estimated state vector vs time (R-NESE parameters: $\tau^{(4)} = 1.5 \times 10^{1}$, D-WTVSE parameters: $\lambda = 10^{-4}$, $c_1 = 5 \times 10^{-4}$, $c_2 = 1 \times 10^{-5}$, $c_3 = 1 \times 10^{-4}$).

The performance of the proposed SE scheme (along with that of two benchmarks) is shown in Fig. 5.24 next. Specifically, the 'R-NESE & D-WTVSE' curve presents the estimation error associated to a SE running (i) the R-NESE scheme with both legacy and μPMU measurements at kT time instants (we let $T = 15$); and (ii) the proposed D-WTVSE scheme at the intermediate ones (we let $q = T/t = 8$, for illustrative purposes). The 'R-NESE-last' curve provides a benchmark for the case where no SE scheme is used at intermediate time instants. Hence, we compute the estimation error between the actual system state at time $kT + nt$ and the latest available R-NESE estimate computed at kT (on the basis of both legacy and μPMU measurements). The 'R-NESE-updated PMUs' curve depicts the achievable error at *intermediate* time instants with the outdated legacy measurements collected at kT (to preserve observability) and the updated μPMU readings gathered at $kT + nt$. Besides, a set of 12 μPMUs has been incorporated to the legacy measurement set. According to the configuration proposed from the μPP algorithm, the meters have been placed at buses 1, 8, 21, 32, 37, 50, 64, 66, 73, 75, 84 and 93.

Fig. 5.24 indicates that the estimation error for 'R-NESE-last' is the highest. This follows from the fact that, the last available state estimate is outdated. At the same time the system voltage profile has been differentiated according to the PV power variations. Indeed, the larger the deviation of the power generated by the PV plants at $kT + nt$ (see Fig. 5.22), the larger the estimation error (e.g., at $T + 4t$). Instead, by replacing the outdated μPMU measurements by updated ones ('R-NESE-updated PMUs' curve), performance significantly improves. This proves the large impact that μPMUs have on the computation of the system state. However, the proposed 'R-NESE & D-WTVSE' scheme clearly outperforms both benchmarks. As the graph shows, the proposed scheme reaches high accuracy in all time instants independently of the rapid variations of the PV power generation. Besides, the graph shows that the main time instants solution, i.e., R-NESE, presents lower estimation error compared to D-WTVSE. The latter comes from the exploitation of the regularization term incorporating prior information. As a general remark, computer simulations show that the proposed SE scheme can constitute a reliable monitoring tool, capable of tracking short-time voltage variations at the MV system.

5.7 Conclusions

In this chapter, we have proposed a *regularized* SE scheme for the robust monitoring of the distribution grid. The algorithm consists of two different time-scale parts; a robust state estimator that operates on the regular time instants (every 15 minutes) and a regularized SE scheme for the intermediate time period. The final goal was to accurately track the system state at a faster time scale with increased reliability, according to the needs of the new operational environment. With respect to the main time instants, the estimator

has been formulated as the regularized version of the normal equations-based SE solution (R-NESE). The computer simulations (on the 95-bus UK DG) have shown that R-NESE presents robustness against the high uncertainty raised by the low redundancy measurement sets and the pseudomeasurements. Moreover, the enhanced regularization term led the algorithm to increased accuracy and improved convergence rate compared with other approaches. For the intermediate time instants, we have presented a *decomposed* weighted total variation state estimation (D-WTVSE) model which is solved by means of the ADMM. The problem has been decomposed according to the ILVR locations across the feeder, exploiting further information from the tap changer indicators. As expected, the scheme presented a slightly worse performance compared to the main time instants estimator. Nonetheless, the extended simulation results have proven that D-WTVSE is able to provide an accurate state estimate under a diversity of scenarios. Further, the low computational complexity of the algorithm provides the opportunity for multiple state estimates within the intermediate time interval. Complementarily, we have presented a μPMU placement (μPP) method, which has been posed as a MISDP model. We have observed that the proposed approach is able to provide a full-rank gain matrix while taking into consideration the existing measurements,. More importantly, the presented solution improved significantly the conditioning of the SE problem and the estimation accuracy. Additional constraints related with the number and location of the μPMUs were also satisfied.

5.8 Appendix C: Computational Complexity Analysis

5.8.1 R-NESE

Let us first consider the computational complexity of the regularized normal equations (R-NESE) algorithm (5.10), presented in subsection 5.3.1. To do so, we will leverage on $\mathscr{O}(n)$ notation, that from now on corresponds to per iteration complexity. In addition, for the ease of notation, let us assume n state variables and k measurements. The main computational burden in:

$$\Delta \mathbf{x}^{(\nu)} = \left(\mathbf{J}_{\mathbf{x}}^{(\nu)T}\mathbf{R}_{\mathbf{e}}^{-1}\mathbf{J}_{\mathbf{x}}^{(\nu)} + \tau^{(\nu)}\boldsymbol{I}\right)^{-1}$$
$$\mathbf{J}_{\mathbf{x}}^{(\nu)T}\mathbf{R}_{\mathbf{e}}^{-1}[\mathbf{z} - \mathbf{h}(\mathbf{x}^{(\nu)})] - \tau^{(\nu)}(\mathbf{x}^{(\nu)} - \mathbf{x}_p) \tag{5.36}$$

entails to a number of matrix multiplications and the matrix inversion. The former can be formed at a cost of $\mathscr{O}(kn^2)$. On the other hand, the matrix inversion usually takes place with a Cholesky factorization at a cost of $\mathscr{O}(n^3)$. Taking into account that in distribution grid SE $n \approx k$, the overall computational complexity of R-NESE scales at $\mathscr{O}(n^3)$. Nevertheless, exploiting the inherent sparsity of $\mathbf{G}^{(\nu)}$ this cost can be reduced to $\mathscr{O}(n^2)$ [127], [128].

With regard to the convergence rate of the algorithm, the normal equations-based schemes commonly present quadratic convergence characteristics [129]. However, this is strongly dependent on the condition number κ of the gain matrix $\mathbf{G}^{(\nu)}$ in each iteration. According to the literature [130], in the case of a well-conditioned system of equations a few number of iterations are sufficient for the full convergence of the algorithm.

5.8.2 D-WTVSE

The computational complexity analysis for the D-WTVSE algorithm follows the same rationale. First, consider the ADMM-based solution presented in subsection 5.4. One can observe that the update step of $\mathbf{x}^{\nu+1}$ (28) dominates the computational burden of each ADMM iteration. More specifically, the computation of:

$$\mathbf{x}_r^{\nu+1} = \left(\mathbf{H}_r^T\mathbf{H}_r + c_1\mathbf{L}_r^T\mathbf{W}_r^T\mathbf{W}_r\mathbf{L}_r + c_2\tilde{\mathbf{Y}}_r{}^T\tilde{\mathbf{Y}}_r + \sum_{s\in\mathscr{A}_r}\mathbf{M}_{rs}^T\mathbf{M}_{rs}\right)^{-1}$$
$$\left(\mathbf{H}_r^T\mathbf{y}_r - \mathbf{L}_r\mathbf{W}_r^T\boldsymbol{\mu}_r^{\nu} + c_1\mathbf{L}_r^T\mathbf{W}_r^T\boldsymbol{\theta}_r{}^{\nu} - \tilde{\mathbf{Y}}_r\boldsymbol{\pi}_r^{\nu} + \sum_{s\in\mathscr{A}_r}\left(\mathbf{M}_{rs}^T\boldsymbol{c_{rs}} - \boldsymbol{\kappa_{rs}}\mathbf{M}_{rs}^T\right)\right) \tag{5.37}$$

consists of a number of matrix multiplications and mainly a matrix inversion. As all matrices involved in the computations are characterized from sparsity, again the overall computational complexity has a final cost of $\mathscr{O}(n^2)$. Important to note here that, in contrast with R-NESE where the gain matrix $\mathbf{G}^{(\nu)}$ has to be re-computed in each iteration, the implementation of D-WTVSE can exploit one additional advantage. In the case

where the control parameters c_1, c_2, c_3 have constant values during the ADMM iterations, the factorization of the matrix inversion can be cached in order to reduce significantly the computational complexity [72].

Concerning the convergence rate of the algorithm, a number of previous works have proved that ADMM presents a linear convergence behavior [131]. That means, the number of iterations needed for a full convergence grows linearly with the size of the problem. However, in the medium voltage SE case, where the number of the state variables can be characterized as moderate, a few decades of iterations are needed. Again, the convergence behavior of this method strongly depends on the conditioning of the problem for the update of $\mathbf{x}^{\nu+1}$. Given that, an improved condition number of the first part in (34) will lead to a decreased number of iterations.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this dissertation, we have focused on the modernization of smart grid monitoring techniques, according to the needs of the new operational environment. First, we have studied the SE problem for the transmission grid. In this case, we have taken into consideration the increased penetration of renewable energy sources and the energy market evolution. These recent concepts have raised the need for robust multiarea SE, according to which, different operation utilities will be able to interact in order to attain an optimal state estimate for networks extended over large geographical areas. Then, we have turned our attention to the distribution grid. In contrast with the high voltage system, here, the SE functionality faces a number of additional challenges, such as the restricted measurement infrastructure and the rapid variations of the system state imposed by the distributed energy resources. Below are summarized our main results and conclusions.

In Chapter 4, we have addressed the problem of transmission grid SE. We have considered a realistic scenario, where measurements are collected from both SCADA system and PMUs. This hybrid SE problem has been formulated as a non-convex optimization problem. Thus, we have leveraged on a successive convex approximation (SCA) framework. The latter, entails the approximation of the original objective function by a sequence of strongly convex ones, leading to an iterative solution (SCA-SE). For the numerical assessment of SCA-SE we have used other solutions from the literature, namely, the classical normal equations-based estimator and an interesting approach which exploits the semidefinite relaxation technique. Computer simulations have shown that, the proposed algorithm outperforms in terms of accuracy. This holds for a broad number of scenarios with respect to the available measurement set and the voltage profiles under consideration. In addition, we have improved the convergence rate of SCA-SE, exploiting second order information from the original objective function in an affordable computational cost. Then, we have proposed the distributed implementation of our algorithm

based on the ADMM. The numerical assessment over a number of different IEEE test cases has shown that, it is feasible to attain a similar accuracy with the centralized version of SCA-SE. Nevertheless, for this multiarea implementation, a restricted exchange of data is needed. Thus, privacy between different system operators is preserved and computational cost stays low. Finally, we have presented the robust counterpart of our scheme (RSCA-SE). Specifically, the SCA-SE model has been re-formulated as a LASSO-like estimation problem by promoting sparsity in the vector of corrupted measurements. This applies for legacy and PMU measurements. In addition to that, we have derived an upper bound of the residual error and concluded that the resulting robust state estimator resembles the Huber's estimator. Here, we have used as benchmarks the classical largest normalized residual test and a suitably re-formulated semidefinite relaxation-based SE scheme. RSCA-SE has proved to detect and cleanse corrupt (bad data) measurements more effectively in a number of different scenarios.

In Chapter 5 we have studied the SE problem for the robust monitoring of the distribution grid. Specifically, we have proposed a regularized SE scheme which consists of two different time-scale parts; a robust state estimator that operates on the regular time instants and, a regularized SE scheme for the intermediate time period. Our final goal has been to estimate accurately the system state at a faster time scale, tracking possible rapid variations over the voltage profile. With regard to the main time instants, we have formulated the estimator as a regularized non-linear least squares optimization problem, which refers to the regularized version of the normal equations-based SE solution (R-NESE). The computer simulations, mainly using the realistic 95-bus UK DG, have revealed that R-NESE has overcomed all barriers posed by the restricted and low quality measurement set. Regardless of the high error variation imposed by the pseudomeasurements, the adopted regularization technique enhanced with the last state estimate $\mathbf{x}_p$, leads the algorithm to an improved estimation accuracy. For the intermediate time instants, we have presented a weighted total variation state estimation model (WTVSE), which exploits a small number of available μPMUs and zero power injections. First, we have proposed a rule to compute the weights as a function of branch and feeder impedances. Based on that, the model accounts for distribution systems of different characteristics. Then, we have taken into consideration the presence of active elements across the feeders, namely, the In-Line Volatge Regulators (ILVRs). This has forced us to decompose the WTVSE problem (D-WTVSE) according to the ILVR locations, exploiting additional information from the tap changer indicators. For the solution of D-WTVSE we have leveraged on ADMM. The extended computer simulations have proven that D-WTVSE is capable of tracking rapid voltage fluctuations with high accuracy. This holds for a broad number of scenarios with respect to DER and load diversity. Nonetheless, the low computational complexity of the algorithm provides the opportunity for multiple consecutive state estimates within the intermediate time interval. Besides, we have presented

a μPMU placement (μPP) method in order to optimize the conditioning of the R-NESE scheme. The problem has been posed as a MISDP model and, thus, it has been efficiently solved. We have observed that the proposed placement method, taking into consideration the existing (legacy) measurements, is able to provide a full-rank gain matrix $\mathbf{G}$. More importantly, the μPP method decreases significanlty the condition number of $\mathbf{G}$, improving the accuracy of the R-NESE scheme.

6.2 Future Work

In this section, we outline and briefly describe several open issues and possible extensions resulting from the research work conducted in this PhD dissertation.

- **Derivation of decentralized versions of the proposed robust SE schemes.** In Chapter 4, we proposed a robust version of the SCA-SE scheme capable of coping with bad data in legacy and PMU measurements (i.e, RSCA-SE) by jointly conducting state estimation *and* bad data detection. The resulting scheme, however, was *centralized*. As such, this SE scheme is not suitable for multi-area/multi-operator scenarios which are very relevant for utilities. Hence, the derivation of *decentralized* versions of the RSCA-SE would be a natural extension of the work conducted in this PhD thesis. However, this raises a number of challenges. First, the corrupted measurements in the tie-lines connecting the different network areas are particularly detrimental and significantly degrade the performance of the (distributed) SE scheme in neighboring areas. Consequently, particular attention should be paid to the (re-) definition of strategies (e.g., based on sparsity principles) guaranteeing an almost sure detection of corrupted measurements in tie lines. Besides, due to the presence of bad data the problem formulation is expected to be highly non-convex. An appropriate convex approximation framework should thus be adopted, allowing to solve the problem with an affordable computational complexity and acceptable latency.

- **Analysis of the impact of communication network impairments on SE methods.** Throughout this PhD dissertation, we have assumed a seamless transmission of all legacy and PMU measurements from the collection points to the SCADA system/processing units. Nonetheless, data transmission in real-world scenarios must be accomplished via actual (wireless) networks which are typically subject to impairments/constraints in terms of latency, packet losses, or available bandwidth/throughput. This is particularly relevant in the case of PMU measurements. On the one hand, such measurements must be accurately synchronized (this encompassing the collection, transmission and processing stages). On the other, a widespread deployment of PMUs in transmission networks results into the generation of much larger volumes of data, given their higher sampling frequency, the

higher precision of such measurements, and the increasing number of PMU devices. An interesting research line here would be the analysis of the impact of such communication impairments and constraints on SE algorithms. This holds in particular for *decentralized* SE strategies where a critical volume of data needs to be exchanged between energy control centers. To do so, realistic traffic models for communication systems need to be introduced into the validation process. This could lead to a re-formulation of SE algorithms exploiting appropriate delay-tolerant solutions such as the asynchronous version of ADMM.

- **Further validation of SE algorithms in real-time simulation environments** The R-NESE/D-WTVSE scheme proposed in Chapter 5 is capable of tracking rapid variations of the voltage profile in medium-voltage feeders. The SE algorithms were developed in MATLAB and, for their validation, we leveraged on the power flows generated (simulated) in MATPOWER. This approach entails a number of simplifying assumptions (e.g., neglected power network elements, single-phase representation) with respect to the operation of real-world power systems. However, in order to successfully deploy such novel SE strategies in control centers, their performance must also be validated in real-time environments. This can be accomplished with the adoption of real-time simulation tools such as the Real Time Digital Simulator (RTDS) software. RTDS is typically used to test hardware equipment (e.g., relays, controllers) by exploiting hardware-in-the-loop configurations. To do that, a first challenge would be to encode and seamlessly integrate the SE software in a closed-loop simulation mode. The final goal, however, would be to validate such novel SE methods under more realistic operating conditions, such as fully detailed power networks models, rapid load and DER production changes and communications delays with respect to the measurements.

- **Optimal PMU placement strategies for realistic communication networks.** As discussed in Chapter 5, the adoption and use of PMUs pursues a two-fold goal: (i) to reinforce/guarantee system observability; and (ii) to increase SE accuracy. To achieve this, both the number of PMUs to be deployed and their locations must be precisely determined. The PMU placement strategies discussed in this PhD dissertation, however, completely ignore the latency and/or reliability (packet drop rate) constraints associated to the underlying communication networks. This naïve approach could e.g., lead to a situation where a system that, from the problem solution, should be observable, it is not in practice because part of the measurements were dropped or reached the SE functionality with excessive latency. Thus, a promising research line for future work in this area would be the reformulation of the PMU placement problem by taking such additional communication-related constraints into consideration.

- **Advanced methods for the generation of pseudomeasurement datasets in distribution grids.** The robust NESE scheme operating at the main time instants presented in Chapter 5 strongly relies on the availability of pseudomeasurements (e.g., pseudo-load injections). In general, pseudomeasurements exhibit a large error variance which, in turn, has a negative impact on SE accuracy. This motivates the need to construct more accurate pseudomeasurement datasets. Most of the works from the literature resort to data mining techniques leveraging on historical load data. In contrast, an interesting direction would be to exploit more recent concepts, such as dictionary learning (DL) techniques for graphs. DL strategies attempt to find sparse signal representations that capture prominent characteristics in a given data. And, further, they exploit the structure of the underlying topology where the data lies. This topology can provide further information, such as smoothness between the values of the data. In a similar context, the specificities of the distribution grid (e.g., tree topologies, smooth voltage variation between adjacent buses) could be leveraged in a DL framework to estimate the load injections in buses where communication infrastructure is absent.

Bibliography

[1] A. Gómez-Expósito, A. de la Villa Jaén, C. Gómez-Quiles, P. Rousseaux, and T. V. Cutsem, "A taxonomy of multi-area state estimation methods," *Electric Power Systems Research*, vol. 81, no. 4, pp. 1060 – 1069, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378779610002841

[2] D. Della Giustina, M. Pau, P. A. Pegoraro, F. Ponci, and S. Sulis, "Electrical distribution system state estimation: measurement issues and challenges," *IEEE Instrumentation Measurement Magazine*, vol. 17, no. 6, pp. 36–42, December 2014.

[3] N. T. Le, "Opportunistic hybrid network coding (ohnc) method and qos metrics modeling for smart grid wireless neighborhood area networks," Ph.D. dissertation, 07 2016.

[4] B. LAB. (2014) Data driven approach for monitoring and control of distribution system assets in microgrids using micro-pmu technology. [Online]. Available: https://building-microgrid.lbl.gov/projects/data-driven-approach-monitoring-and

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[6] (2015) Power flow test systems. [Online]. Available: https://al-roomi.org/power-flow/30-bus-system

[7] (2015) Power flow test systems. [Online]. Available: https://al-roomi.org/power-flow/57-bus-system

[8] (2015) Power flow test systems. [Online]. Available: https://al-roomi.org/power-flow/118-bus-system

[9] Y. P. Agalgaonkar, *Control and Operation of Power Distribution System for Optimal Accommodation of PV Generation*. Thesis of Doctor of Philosophy: Imperial College, Department of Electrical and Electronic Engineering, 2014.

[10] J. Matamoros, A. Tsitsimelis, M. Gregori, and C. Antón-Haro, "Multiarea state estimation with legacy and synchronized measurements," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[11] M. Cosovic, A. Tsitsimelis, D. Vukobratovic, J. Matamoros, and C. Anton-Haro, "5g mobile cellular networks: Enabling distributed state estimation for smart grids," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 62–69, Oct 2017.

[12] A. Tsitsimelis, J. Matamoros, and C. Antón-Haro, "An admm-based regularized state estimation scheme for the distribution grid," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2017, pp. 900–905.

[13] A. Tsitsimelis and C. Antón-Haro, "A regularized state estimation scheme for a robust monitoring of the distribution grid," *International Journal of Electrical Power & Energy Systems*, vol. 117, p. 105621, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S014206151831490X

[14] A. Tsitsimelis, C. Kalalas, J. Alonso-Zarate, and C. Antón-Haro, "On the impact of lte rach reliability on state estimation in wide-area monitoring systems," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2018, pp. 1–6.

[15] A. G. Phadke, "The wide world of wide-area measurement," *IEEE Power and Energy Magazine*, vol. 6, no. 5, pp. 52–65, September 2008.

[16] F. C. Schweppe and J. Wildes, "Power system static-state estimation, part i: Exact model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 120–125, Jan 1970.

[17] A. Garcia, A. Monticelli, and P. Abreu, "Fast decoupled state estimation and bad data processing," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-98, no. 5, pp. 1645–1652, Sept 1979.

[18] A. Gomez-Exposito, C. Gomez-Quiles, and A. d. l. V. Jaen, "Bilinear power system state estimation," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 493–501, Feb 2012.

[19] M. Gol and A. Abur, "A hybrid state estimator for systems with limited number of pmus," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1511–1517, May 2015.

[20] C. Bruno, C. Candia, L. Franchi, G. Giannuzzi, M. Pozzi, R. Zaottini, L. Franchi, G. Giannuzzi, and M. Zaramella, "Possibility of enhancing classical weighted least squares state estimation with linear pmu measurements," in *PowerTech, 2009 IEEE Bucharest*, June 2009, pp. 1–6.

[21] N. M. Manousakis, G. N. Korres, J. N. Aliprantis, G. P. Vavourakis, and G. C. J. Makrinas, "A two-stage state estimator for power systems with pmu and scada measurements," in *PowerTech (POWERTECH), 2013 IEEE Grenoble*, June 2013, pp. 1–6.

[22] R. Baltensperger, A. Loosli, H. Sauvain, M. Zima, G. Andersson, and R. Nuqui, "An implementation of two-stage hybrid state estimation with limited number of pmu," in *Developments in Power System Protection (DPSP 2010). Managing the Change, 10th IET International Conference on*, March 2010, pp. 1–5.

[23] A. Gomez-Exposito, A. Abur, A. de la Villa Jaen, and C. Gomez-Quiles, "A multilevel state estimation paradigm for smart grids," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 952–976, June 2011.

[24] G. N. Korres, "A distributed multiarea state estimation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 73–84, Feb 2011.

[25] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1617–1626, May 2013.

[26] Y. Guo, L. Tong, W. Wu, H. Sun, and B. Zhang, "Hierarchical multi-area state estimation via sensitivity function exchanges," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–12, 2016.

[27] S. Jayakumar Geetha, S. Chakrabarti, and K. Rajawat, "Hierarchical parallel dynamic estimator of states for interconnected power system," *IET Generation, Transmission Distribution*, vol. 12, no. 10, pp. 2299–2306, 2018.

[28] L. Xie, D. H. Choi, S. Kar, and H. V. Poor, "Fully distributed state estimation for wide-area monitoring systems," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1154–1169, Sept 2012.

[29] H. Zhu and G. B. Giannakis, "Power system nonlinear state estimation using distributed semidefinite programming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1039–1050, Dec 2014.

[30] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

[31] W. Zheng, W. Wu, A. Gomez-Exposito, B. Zhang, and Y. Guo, "Distributed robust bilinear state estimation for power systems with nonlinear measurements," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[32] A. Minot, Y. M. Lu, and N. Li, "A distributed gauss-newton method for power system state estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3804–3815, Sept 2016.

[33] M. Cosovic and D. Vukobratovic, "Distributed gauss–newton method for state estimation using belief propagation," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 648–658, Jan 2019.

[34] P. Chavali and A. Nehorai, "Distributed power system state estimation using factor graphs," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2864–2876, June 2015.

[35] M. Rostami and S. Lotfifard, "Distributed dynamic state estimation of power systems," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3395–3404, Aug 2018.

[36] V. Donde, X. Feng, I. Segerqvist, and M. Callavik, "Distributed state estimation of hybrid ac/hvdc grids by network decomposition," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 974–981, March 2016.

[37] C. Muscas, M. Pau, P. A. Pegoraro, and S. Sulis, "Smart electric energy measurements in power distribution grids," *IEEE Instrumentation Measurement Magazine*, vol. 18, no. 1, pp. 17–21, February 2015.

[38] Y. P. Agalgaonkar, B. C. Pal, and R. A. Jabr, "Distribution voltage control considering the impact of pv generation on tap changers and autonomous regulators," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 182–192, Jan 2014.

[39] A. von Meier, E. Stewart, A. McEachern, M. Andersen, and L. Mehrmanesh, "Precision micro-synchrophasors for distribution systems: A summary of applications," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2926–2936, Nov 2017.

[40] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2312–2322, March 2019.

[41] A. G. Exposito, A. de la Villa Jaen, and J. L. R. Izaga, "An alternative state estimation formulation for radial distribution networks," in *Power Tech, 2007 IEEE Lausanne*, July 2007, pp. 396–400.

[42] M. E. Baran and A. W. Kelley, "State estimation for real-time monitoring of distribution systems," *IEEE Transactions on Power Systems*, vol. 9, no. 3, pp. 1601–1609, Aug 1994.

[43] Whei-Min Lin and Jen-Hao Teng, "State estimation for distribution systems with zero-injection constraints," *IEEE Transactions on Power Systems*, vol. 11, no. 1, pp. 518–524, Feb 1996.

[44] W. Wu, Y. Ju, B. Zhang, and H. Sun, "A distribution system state estimator accommodating large number of ampere measurements," *International Journal of Electrical Power & Energy Systems*, vol. 43, pp. 839–848, 12 2012.

[45] M. Pau, P. A. Pegoraro, and S. Sulis, "Efficient branch-current-based distribution system state estimation including synchronized measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 9, pp. 2419–2429, Sep. 2013.

[46] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, "Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1888–1896, Nov 2012.

[47] J. Wu, Y. He, and N. Jenkins, "A robust state estimator for medium voltage distribution networks," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1008–1016, May 2013.

[48] A. Gómez-Expósito, C. Gómez-Quiles, and I. Džafić, "State estimation in two time scales for smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 421–430, Jan 2015.

[49] D. Carta, A. Benigni, and C. Muscas, "Model order reduction for pmu-based state estimation in distribution grids," *IEEE Systems Journal*, vol. 12, no. 3, pp. 2711–2720, Sep. 2018.

[50] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Enhancing observability in distribution grids using smart meter data," *IEEE Transactions on Smart Grid*, pp. 1–1, 2017.

[51] Y. Hu, A. Kuh, T. Yang, and A. Kavcic, "A belief propagation based power distribution system state estimator," *IEEE Computational Intelligence Magazine*, vol. 6, no. 3, pp. 36–46, Aug 2011.

[52] C. Klauber and H. Zhu, "Distribution system state estimation using semidefinite programming," in *North American Power Symposium (NAPS), 2015*, Oct 2015, pp. 1–6.

[53] A. von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES*, Feb 2014, pp. 1–5.

[54] M. B. Do Coutto Filho and J. C. Stacchini de Souza, "Forecasting-aided state estimation—part i: Panorama," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1667–1677, Nov 2009.

[55] S. Deshmukh, B. Natarajan, and A. Pahwa, "State estimation and voltage/var control in distribution network with intermittent measurements," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 200–209, Jan 2014.

[56] J. S. Thorp, A. Abur, M. Begovic, J. Giri, and R. Avila-Rosales, "Gaining a wider perspective," *IEEE Power and Energy Magazine*, vol. 6, no. 5, pp. 43–51, September 2008.

[57] B. Gou, "Generalized integer linear programming formulation for optimal pmu placement," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1099–1104, Aug 2008.

[58] B. Milosevic and M. Begovic, "Nondominated sorting genetic algorithm for optimal phasor measurement placement," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 69–75, Feb 2003.

[59] T. L. Baldwin, L. Mili, M. B. Boisen, and R. Adapa, "Power system observability with minimal phasor measurement placement," *IEEE Transactions on Power Systems*, vol. 8, no. 2, pp. 707–715, May 1993.

[60] M. Hajian, A. M. Ranjbar, T. Amraee, and B. Mozafari, "Optimal placement of pmus to maintain network observability using a modified bpso algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 1, pp. 28 – 34, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142061510001432

[61] S. P. Singh and S. P. Singh, "Optimal cost wide area measurement system incorporating communication infrastructure," *IET Generation, Transmission Distribution*, vol. 11, no. 11, pp. 2814–2821, 2017.

[62] M. Korkali and A. Abur, "Impact of network sparsity on strategic placement of phasor measurement units with fixed channel capacity," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, May 2010, pp. 3445–3448.

[63] V. Kekatos, G. B. Giannakis, and B. Wollenberg, "Optimal placement of phasor measurement units via convex relaxation," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1521–1530, Aug 2012.

[64] N. M. Manousakis and G. N. Korres, "Optimal pmu placement for numerical observability considering fixed channel capacity: A semidefinite programming approach," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3328–3329, July 2016.

[65] N. M. Manousakis and G. N. Korres, "An advanced measurement placement method for power system observability using semidefinite programming," *IEEE Systems Journal*, vol. 12, no. 3, pp. 2601–2609, Sep. 2018.

[66] R. Arghandeh, M. Gahr, A. von Meier, G. Cavraro, M. Ruh, and G. Andersson, "Topology detection in microgrids with micro-synchrophasors," in *2015 IEEE Power Energy Society General Meeting*, July 2015, pp. 1–5.

[67] A. L. Liao, E. M. Stewart, and E. C. Kara, "Micro-synchrophasor data for diagnosis of transmission and distribution level events," in *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T D)*, May 2016, pp. 1–5.

[68] L. Schenato, G. Barchi, D. Macii, R. Arghandeh, K. Poolla, and A. V. Meier, "Bayesian linear state estimation using smart meters and pmus measurements in distribution grids," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2014, pp. 572–577.

[69] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[70] M. Razaviyayn, "Successive convex approximation: Analysis and applications," 2014.

[71] G. Scutari, F. Facchinei, and L. Lampariello, "Parallel and distributed methods for constrained nonconvex optimization—part i: Theory," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 1929–1944, April 2017.

[72] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011. [Online]. Available: http://dx.doi.org/10.1561/2200000016

[73] D. Kirschen and G. Strbac, *Fundamentals of Power System Economics*. United Kingdom: John Wiley Sons Ltd, 2004.

[74] J. De La Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized phasor measurement applications in power systems," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 20–27, June 2010.

[75] L. Zhao and A. Abur, "Multi area state estimation using synchronized phasor measurements," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 611–617, May 2005.

[76] A. Gomez-Exposito, A. Abur, P. Rousseaux, A. De, L. Villa, and C. Gomez-Quiles, "On the use of pmus in power system state estimation," 01 2011.

[77] G. Giannakis, V. Kekatos, N. Gatsis, S.-J. Kim, H. Zhu, and B. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," vol. 30, no. 5, pp. 107–128, Sep. 2013.

[78] M. Cosovic and D. Vukobratovic, "Distributed gauss-newton method for ac state estimation: A belief propagation approach," in *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2016, pp. 643–649.

[79] D. H. Choi and L. Xie, "Fully distributed bad data processing for wide area state estimation," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, Oct 2011, pp. 546–551.

[80] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, Feb 2000.

[81] A. Abur and A. G. Exposito, *Power System State Estimation Theory and Implementation*. New York, NY, USA: Marcel Dekker, 2004.

[82] L. Zhang and A. Abur, "State estimator tuning for pmu measurements," in *North American Power Symposium (NAPS), 2011*, Aug 2011, pp. 1–4.

[83] ——, "Impact of tuning on bad data detection of pmu measurements," in *Innovative Smart Grid Technologies - Asia (ISGT Asia), 2012 IEEE*, May 2012, pp. 1–5.

[84] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J. S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 641–656, Feb 2014.

[85] J. Kaleva, A. Tölli, and M. Juntti, "Decentralized sum rate maximization with qos constraints for interfering broadcast channel via successive convex approximation," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2788–2802, June 2016.

[86] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1874–1889, April 2015.

[87] H. Zhu and G. B. Giannakis, "Robust power system state estimation for the nonlinear ac flow model," in *North American Power Symposium (NAPS), 2012*, Sept 2012, pp. 1–6.

[88] Bei Gou and A. Abur, "An improved measurement placement algorithm for network observability," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 819–824, Nov 2001.

[89] B. Xu and A. Abur, "Observability analysis and measurement placement for systems with pmus," in *Power Systems Conference and Exposition, 2004. IEEE PES*, Oct. 2004, pp. 943–946 vol.2.

[90] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb 2011.

[91] N. M. Manousakis and G. N. Korres, "A weighted least squares algorithm for optimal pmu placement," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3499–3500, Aug 2013.

[92] A. Primadianto and C. Lu, "A review on distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875–3883, Sep. 2017.

[93] A. Ranković, B. M. Maksimović, and A. T. Sarić, "A three-phase state estimation in active distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 54, pp. 154 – 162, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142061513002986

[94] M. E. Baran and A. W. Kelley, "A branch-current-based state estimation method for distribution systems," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 483–491, Feb 1995.

[95] J.-H. Menke, J. Hegemann, S. Gehler, and M. Braun, "Heuristic monitoring method for sparsely measured distribution grids," *International Journal of Electrical Power & Energy Systems*, vol. 95, pp. 146–155, 02 2018.

[96] H. G. Abood, V. Sreeram, and Y. Mishra, "A new algorithm for improving the numerical stability of power system state estimation," *IEEJ Transactions on Electrical and Electronic Engineering*, 10 2018.

[97] R. Ebrahimian and R. Baldick, "State estimator condition number analysis," *IEEE Transactions on Power Systems*, vol. 16, no. 2, pp. 273–279, May 2001.

[98] C. Muscas, M. Pau, P. A. Pegoraro, and S. Sulis, "Effects of measurements and pseudomeasurements correlation in distribution system state estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 12, pp. 2813–2823, Dec 2014.

[99] H. Brom, "Three-phase state estimation in the medium-voltage network with aggregated smart meter data," *International Journal of Electrical Power & Energy Systems*, vol. 98, 02 2018.

[100] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, "Distribution system state estimation based on nonsynchronized smart meters," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2919–2928, Nov 2015.

[101] Q. Chen, D. Kaleshi, Z. Fan, and S. Armour, "Impact of smart metering data aggregation on distribution system state estimation," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1426–1437, Aug 2016.

[102] G. T. Heydt, "The next generation of power distribution systems," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 225–235, Dec 2010.

[103] M. Hasheminamin, V. G. Agelidis, V. Salehi, R. Teodorescu, and B. Hredzak, "Index-based assessment of voltage rise and reverse power flow phenomena in a distribution feeder under high pv penetration," *IEEE Journal of Photovoltaics*, vol. 5, no. 4, pp. 1158–1168, July 2015.

[104] R. A. Walling, R. Saint, R. C. Dugan, J. Burke, and L. A. Kojovic, "Summary of distributed resources impact on power delivery systems," *IEEE Transactions on Power Delivery*, vol. 23, no. 3, pp. 1636–1644, July 2008.

[105] R. S. Silva, F. M. Laburu, and M. C. de Almeida, "On the use of μpmus for state estimation in distribution systems," in *2017 IEEE Power Energy Society General Meeting*, July 2017, pp. 1–5.

[106] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5152–5162, Sep. 2018.

[107] C. Muscas, M. Pau, P. A. Pegoraro, and S. Sulis, "Uncertainty of voltage profile in pmu-based distribution system state estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 5, pp. 988–998, May 2016.

[108] H. G. Abood, V. Sreeram, and Y. Mishra, "Incremental placement of pmus for enhancing state estimation accuracy," in *2016 Australasian Universities Power Engineering Conference (AUPEC)*, Sep. 2016, pp. 1–6.

[109] Y. Wang, C. Wang, W. Li, J. Li, and F. Lin, "Reliability-based incremental pmu placement," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2744–2752, Nov 2014.

[110] Y. Yao, X. Liu, D. Zhao, and Z. Li, "Distribution system state estimation: A semidefinite programming approach," *IEEE Transactions on Smart Grid*, vol. 10, pp. 4369–4378, 2019.

[111] H. Ahmadi, J. R. Martı, and A. von Meier, "A linear power flow formulation for three-phase distribution systems," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 5012–5021, Nov 2016.

[112] C. Rusu, J. Thompson, and N. M. Robertson, "Sensor scheduling with time, energy, and communication constraints," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 528–539, Jan 2018.

[113] T. C. Xygkis and G. N. Korres, "Optimized measurement allocation for power distribution systems using mixed integer sdp," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 11, pp. 2967–2976, Nov 2017.

[114] X. Chen, J. Lin, C. Wan, Y. Song, S. You, Y. Zong, W. Guo, and Y. Li, "Optimal meter placement for distribution network state estimation: A circuit representation based milp approach," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4357–4370, Nov 2016.

[115] J. Liu, J. Tang, F. Ponci, A. Monti, C. Muscas, and P. A. Pegoraro, "Trade-offs in pmu deployment for state estimation in active distribution grids," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 915–924, June 2012.

[116] J. Liu, F. Ponci, A. Monti, C. Muscas, P. A. Pegoraro, and S. Sulis, "Optimal meter placement for robust measurement systems in active distribution grids," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 5, pp. 1096–1105, May 2014.

[117] J. J. Grainger and W. Stevenson, *Power System Analysis*. MacGraw-Hill, 1994.

[118] S. Gurol, "Solving regularized nonlinear least-squares problem in dual space with application to variational data assimilation," June 2013. [Online]. Available: http://oatao.univ-toulouse.fr/9700/

[119] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[120] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, Dec. 1992. [Online]. Available: http://dx.doi.org/10.1137/1034115

[121] P. C. Hansen, *The L-Curve and Its Use in the Numerical Treatment of Inverse Problems*, 01 2001, vol. 4, pp. 119–142.

[122] Q. Yuan, L. Zhang, and H. Shen, "Multiframe super-resolution employing a spatially weighted total variation model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 3, pp. 379–392, March 2012.

[123] C. Gomez-Quiles, A. Gomez-Exposito, and A. de la Villa Jaen, "State estimation for smart distribution substations," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 986–995, June 2012.

[124] M. Grant and B. SP, "Cvx: Matlab software for disciplined convex programming," 01 2014.

[125] J. Lofberg, "Yalmip : a toolbox for modeling and optimization in matlab," in *2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)*, Sep. 2004, pp. 284–289.

[126] M. Farivar, R. Neal, C. Clarke, and S. Low, "Optimal inverter var control in distribution systems with high pv penetration," in *2012 IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–7.

[127] F. L. Alvarado, "Computational complexity in power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. 95, no. 4, pp. 1028–1037, July 1976.

[128] M. K. Enns, W. F. Tinney, and F. L. Alvarado, "Sparse matrix inverse factors (power systems)," *IEEE Transactions on Power Systems*, vol. 5, no. 2, pp. 466–473, May 1990.

[129] Björck, *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics, 1996. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611971484

[130] C. Gómez-Quiles, H. A. Gil, A. de la Villa Jaén, and A. Gómez-Expósito, "Equality-constrained bilinear state estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 902–910, May 2013.

[131] J. Eckstein and W. Yao, "Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives," *Pacific Journal of Optimization*, vol. 11, pp. 619–644, 10 2015.