Data-efficient Analytics for Optimal Human-Cyber-Physical Systems

Ming Jin

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2017-228
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-228.html

December 15, 2017

Data-efficient Analytics for Optimal Human-Cyber-Physical Systems

by

Ming Jin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

and the Designated Emphasis

in

Communication, Computation and Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Costas J. Spanos, Chair
Professor Pieter Abbeel
Professor Stefano Schiavon
Professor Alexandra von Meier

Fall 2017

Data-efficient Analytics for Optimal Human-Cyber-Physical Systems

Abstract

Data-efficient Analytics for Optimal Human-Cyber-Physical Systems

by

Ming Jin

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences
and the Designated Emphasis in
Communication, Computation and Statistics

University of California, Berkeley

Professor Costas J. Spanos, Chair

The goal of this research is to enable optimal human-cyber-physical systems (h-CPS) by data-efficient analytics. The capacities of societal-scale infrastructures such as smart buildings and power grids are rapidly increasing, becoming physical systems capable of cyber computation that can deliver human-centric services while enhancing efficiency and resilience. Because people are central to h-CPS, the first part of this thesis is dedicated to learning about the human factors, including both human behaviors and preferences. To address the central challenge of data scarcity, we propose physics-inspired sensing by proxy and a framework of “weak supervision” to leverage high-level heuristics from domain knowledge. To infer human preferences, our key insight is to learn a functional abstraction that can rationalize people’s behaviors. Drawing on this insight, we develop an inverse game theory framework that determines people’s utility functions by observing how they interact with one another in a social game to conserve energy. We further propose deep Bayesian inverse reinforcement learning, which simultaneously learns a motivator representation to expand the capacity of modeling complex rewards and rationalizes an agent’s sequence of actions to infer its long-term goals.

Enabled by this contextual awareness of the human, cyber, and physical states, we introduce methods to analyze and enhance system-level efficiency and resilience. We propose an energy retail model that enables distributed energy resource utilization and that exploits demand-side flexibility. The synergy that naturally emerges from integrated optimization of thermal and electrical energy provision substantially improves efficiency and economy. While data empowers the aforementioned h-CPS learning and control, malicious attacks can pose major security threats. The cyber resilience of power system state estimation is analyzed. The envisioning process naturally leads to a power grid resilience metric to guide “grid hardening.” While the methods introduced in the thesis can be applied to many h-CPS systems, this thesis focuses primarily on the implications for smart buildings and smart grid.

Contents

Acknowledgments

I could not ask for a better advisor than Costas Spanos. His enthusiasm and patience, insight and vision on research, academia and life have been a perpetuating motivation throughout my Ph.D., and will continue to empower me in my future career and life. I am deeply grateful that Costas took a chance on me after my "elevator pitch" during his visit to the Hong Kong University of Science and Technology – I would not have been where I am today without his invaluable support and guidance.

I would like to express my gratitude to the other committee members – Pieter Abbeel, Stefano Schiavon, and Sascha von Meier – for bringing insightful and interdisciplinary perspectives on my research. Pieter Abbeel for introducing me to the fascinating world of robotic research. Stefano Schiavon for sharing his deep insights on human-centric design in smart buildings. Sascha von Meier for the fruitful and constructive discussions on power system research.

Working at the Lawrence Berkeley National Lab has been an amazing experience, and I would like to thank my colleagues and collaborators, Wei Feng, Chris Marnay, Bo Shen, Nan Zhou, Jiang Lin, Lynn Price, Michel Foure, and Tianzhen Hong, for offering the unique opportunity to gain perspectives on conducting impactful research to solve societal-scale energy problems.

Many of my recent work has been in close collaboration with Javad Lavaei. Javad has given me invaluable advice and shared his vast knowledge on optimization and control theory. It is truly empowering to collaborate with Javad on high-impact research problems.

I am very grateful for the faculty who helped shape my research during my time at Berkeley: Alex Bayen for his vital support and guidance, and productive collaboration on the work "sensing by proxy" that led to my first best paper award. Peter Bartlett for his insightful discussions on statistical learning theory. Claire Tomlin and Kameshwar Poolla for the advice and constructive discussions. Murat Arcak, Shankar Sastry, and Laurent El Ghaoui for introducing me to the world of control and optimization. Sergey Levine for his inspiring class on deep reinforcement learning. Martin Wainwright, Bin Yu, Jitendra Malik, and Michael Jordan for the introduction to machine learning, computer vision, and theoretical statistics, all of which underpin the various parts of my research.

I have also been so fortunate to have many other collaborators and mentors outside Berkeley: Lin Zhang for sharing his profound knowledge in mobile sensing and information theory, and for his valuable support and guidance throughout my Ph.D. Kalle Johansson for the thoughtful discussions and productive collaboration on cyber security. Jinyue Yan for his world-class insights on energy transformation and passion for promoting young scholars to conduct clean technology research and entrepreneurial endeavor. Lillian Ratliff for the fruitful collaboration on game theory and her unique perspectives on academia and research. Nikos Bekiaris-Liberis and Andreas Damianou for the in-depth discussions on control theory and deep Bayesian network that have led to exciting research. Sanjib Kumar Panda, Johanna Mathieu, and Therese Peffer for flying to Sydney in support of our first international workshop on smart building and smart grid and sharing their valuable insights.

I would also like to thank the faculty and teachers at Berkeley who have tremendously influenced my teaching philosophy: Ronald Fearing, Anant Sahai, and Michel Maharbiz for being amazing mentors and teaching me how to design a class and transform student learning into a fun and inspiring experience. Tsu-Jae King Liu for the great class on teaching techniques for engineering classes. Linda von Hoene and Sabrina Soracco for the superb opportunity during the summer of 2017 to think and discuss about my roles and aspirations in academia and higher education more broadly.

I would also like to thank the people at CREST, SinBerBEST, EECS, IEOR, and LBNL for the fun and productive exchange of ideas over the years: Kevin Weekly, Ioannis Konstantakopoulos, Han Zou, Yuxun Zhou, Ruoxi Jia, Zhaoyi Kang, Jae Yeon Baek, Shichao Liu, Weixi Gu, Hao Jiang, Li Dan, Jason Poon, Hari Prasanna Das, Lucas Spangher, Toby Cheung, Chris Hsu, Chris Soyza, Ping Liu, Yu Zhang, Wei Qi, Angela Liu, Chao Ding, Jing Ge, Daniel Gerber, Dai Wang, Salar Fattahi, Richard Zhang.

Many thanks to special people of CREST, EECS, ITS, and LBNL, who made everything run smoothly: Yovana Gomez, Judy Huang, Shirley Salanio, Patrick Hernan, Yulia Golubovskaya, Helen Bassham, and Sammi Leung.

Going back many years, I am very grateful for the tremendous support, constructive advice and generous mentorship from faculty at HKUST, Penn and Princeton during my undergraduate research and study – Hoi Sing Kwok, Tony Smith, George Jie Yuan, Weichuan Yu, Levent Yobas, Ross Murch, Mansun Chan, Zexiang Li, Zhiyong Fan, Ling Shi, Oscar Au, Shiyou Xu, and Nan Yao. Thank you all for believing in me and sharing your valuable experiences and outlooks.

I shall be forever grateful for my parents for their unconditional support, understanding, love and inspirations. You have always encouraged me to pursue my dream and aspirations, be honest, independent, persevering, responsible and compassionate. Thank you for nurturing me and being my best role models. Last but not least, I want to thank my wife and colleague, Ruoxi, for making every day special. I am fortunate to collaborate with such a wonderful lady in research and life, and to seek meaning in each other's love and support. I would like to dedicate this thesis to my family.

Chapter 1

Introduction

In the coming century, societies collectively face enormous challenges, including massive urbanization, aging and unreliable infrastructure, natural resource depletion and cyber space security.[1] None of these challenges can be met without finding ways to overcome the social and economic barriers to change. Progress in technologies such as the Internet of Things (IoT), machine learning and advanced computation has caused the world to become more inclusive, more connected, and more productive. However, to meet the challenges, our societal-scale systems must become not only more technologically advanced, but also more sustainable, safer, and form healthier and happier places for people. Consider, for example, buildings, which consume about 40% of total U.S. energy consumption and provide working and living spaces for people.[2] There is a huge potential for solutions to improve the comfort, health and productivity of people inside these buildings. From environmental sensors that monitor indoor environmental quality to wearables that track people's activities, diverse sources of information can be fused and filtered *to tailor the indoor environment for optimal worker performance* (for example, by creating individualized "environmental bubbles", by shifting the color "temperature" of office LED lights, and by dynamically controlling ventilation and air conditioning systems in response to both CO_2 levels, temperature, and levels of airborne particulate matter). Ultimately, the goal is to *transform human resources* (for example, by revealing communication patterns, engaging workers in organic collaboration, and by configuring work areas to increase knowledge transfer and prevent knowledge silos).

> *A human-cyber-physical system (h-CPS) is a physical system with a "cyber brain" that engages humans in myriad aspects from system operation to service delivery.*

Driven by the fundamental push for energy efficiency, resilience, and human-centric values, this thesis focuses on using data analytics to achieve optimal h-CPS performance.

Integrating and coordinating the human, cyber and physical components in an h-CPS is the key to enabling cross-layer design, testing, certification, operation, maintenance, ren-

[1] "NAE Grand Challenges for Engineering", National Academy of Engineering, 2008

[2] "How much energy is consumed in U.S. residential and commercial buildings?", The U. S. Energy Information Administration, 2017

ovation and upgradability (Fig. 1.1). Each layer, nevertheless, has its own functions and characteristics. How do we model and learn about each layer, in particular the human factor, which can be stochastic, complex and not directly observable? Learning in h-CPS is often constrained by data availability, how to learn efficiently in a low data regime? And how to leverage the learned knowledge to improve the functionality of the overall system? A typical h-CPS can be viewed as a system of systems (e.g., a building is a system that includes lighting, heating, ventilation, and air conditioning subsystems, and an energy grid is a system that includes natural gas, electricity and thermal energy subsystems). How to exploit the synergy from a system-level integration point of view to improve efficiency? Last but not least, with the increasing reliance on data and data analytics, h-CPSs have become vulnerable to malicious cyberattacks. How to evaluate the vulnerability of the data analytics and enhance cyber resilience?

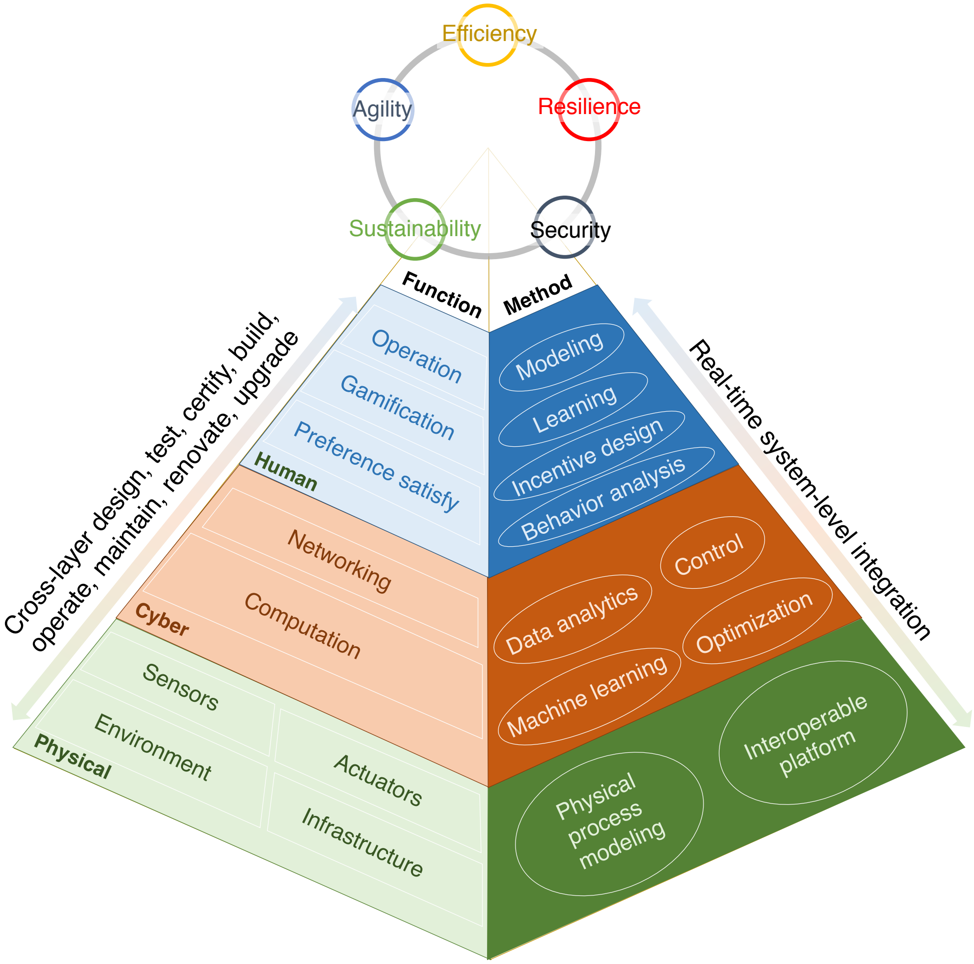


Figure 1.1: A human-cyber-physical system organically engages human factors with cyber-physical infrastructure, creating a cross-layer design and operation to improve overall efficiency, resilience, agility, security and sustainability.

1.1 Thesis approach

> *This thesis bridges ideas from machine learning, control theory and optimization from two distinctive aspects: a focus on human factor learning and human-centric operation, and the development of data-efficient algorithms for h-CPSs.*

The human factor

In societal-scale systems like buildings and power grids, *the human factor* refers to the relevant aspects of people as "users" who receive services, "agents" and "operators" who influence or operate the system, and "sensors" that monitor the context (Fig. 1.2). Due to the central role that people occupy, investigating the human factor can potentially provide in-depth knowledge to identify the root causes of inefficiency, support the tuning of system controls, and assess the performance of h-CPS. We adopt a "*human-centric design*"—a holistic approach to infer the needs of people, optimize their experiences, and enable the system to respond to their feedback.

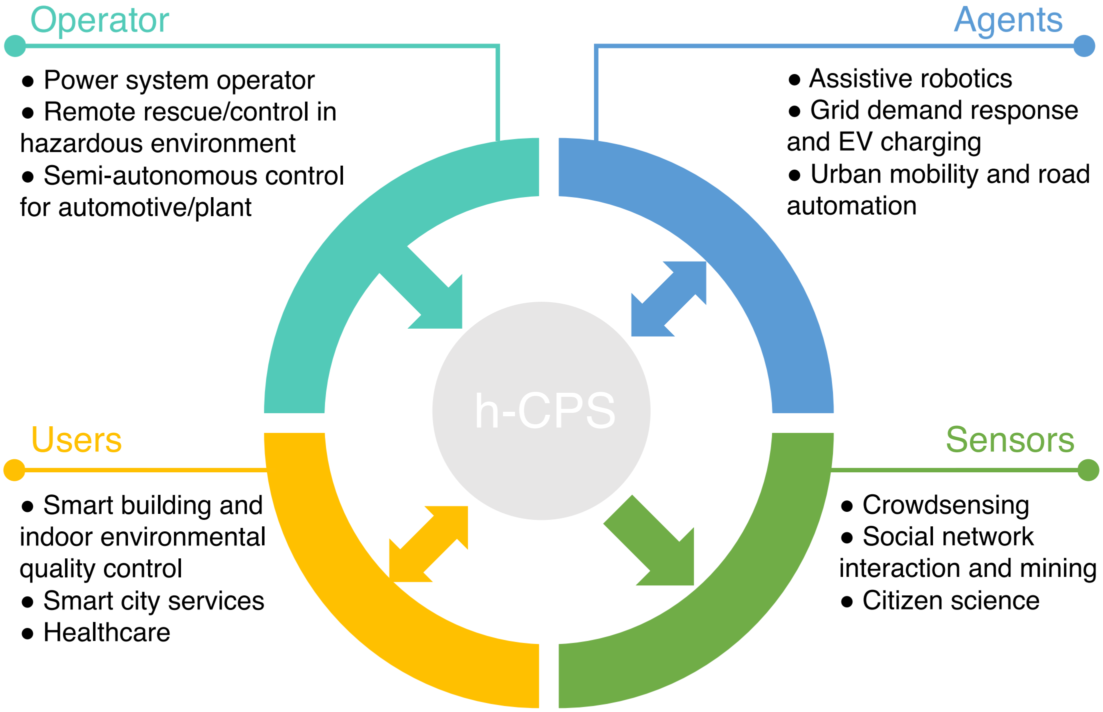

Operator
● Power system operator
● Remote rescue/control in hazardous environment
● Semi-autonomous control for automotive/plant
Agents
● Assistive robotics
● Grid demand response and EV charging
● Urban mobility and road automation
h-CPS
Users
● Smart building and indoor environmental quality control
● Smart city services
● Healthcare
Sensors
● Crowdsensing
● Social network interaction and mining
● Citizen science

Figure 1.2: People are central to h-CPS, and play diverse roles in its operation.

The human factor is remarkably distinct from the cyber and physical components of the system: people are social beings, privacy-sensitive, risk-averse, and non-stationary. Their behaviors can be patterned, yet their individual differences remain unpredictable. With respect to these special characteristics, the learning approach needs to be tailored accordingly:

- *Sensing* needs to be simple and non-intrusive to minimize both collection effort and interference. This idea motivated the proposal of the sensing by proxy paradigm (Chap. 2);
- *Learning* needs to be indirect, data-efficient and *in-vivo* to reduce bias and cost. This led to the development of learning under weak supervision (Chap. 3), inverse game theory (Chap. 4) and deep Bayesian inverse reinforcement learning (Chap. 5);
- *Control* needs to be soft and incentivized to encourage participation. This underlies the efficiency via gamification concept (Chap. 4) and the economic incentive design for optimal energy retail and dispatch (Chap. 6).

The overall thesis approach to understand, model, infer and influence the human factor employs inter-disciplinary tools that combine cyber, physical and physiological measurements with user engagement and behavior (surveys, interactions, simulated performance activities) to perform commensurate analysis and control development.

The paradigm shift of data-efficient analytics

Machine learning has risen in part because of big data collection in areas like computer vision, machine translation, search and advertising. However, in many domains, particularly h-CPS, data is significantly limited due to availability, cost, and privacy and security concerns. There is a broad agreement that new techniques are needed that are capable of working with less data; thus, we are witnessing the emergence of many lines of research in estimation and prediction [11], [184], [193], computer vision [208], [235], and reinforcement learning [55], [65], [115], [224].

From the perspective of data analytics in h-CPS, *this thesis adopts an approach to tackle the data efficiency issue at key steps in the analytics pipeline*, from experimental design, data collection and analysis to model learning, evaluation and online adaptation (Fig. 1.3):

- In *an optimal experimental design*, the goal is to apply domain knowledge to determine what data to collect, and where and when to collect it, as well as to determine the duration of the experiment. For instance, we proposed a method to determine the duration of an experiment for observing (just) enough samples to test a hypothesis with a certain level of confidence [106];
- During *data collection*, instead of relying on a fixed sensor network, we proposed an "automated mobile sensing" strategy based on a mobile robot that actively takes samples to infer an event [104];

- For *data analysis*, we pursued several approaches to alleviate the need for additional data, such as incorporating explicit domain knowledge to design effective features [35], [76], [90], [99], and to make robust estimations via bootstrapping [132];

- The main part of the thesis is dedicated to discussing *data-efficient model learning* such as sensing by proxy, which uses inspirations from physics (Chap. 2), weak supervision, which leverages high-level heuristic rules (Chap. 3), and a deep Bayesian network, which achieves simultaneous representation learning and inverse reinforcement learning (Chap. 5);

- *Model adaptation* is often needed to generalize knowledge across domains (e.g., we employed transfer learning to use an existing dataset in new but similar scenarios for occupancy sensing [99]), and to adapt to changing environments (e.g., we designed a non-parametric algorithm that exploits both historical data and online samples for indoor positioning [249]);

- During *evaluation*, methods like multi-modal fusion and prediction pooling can be used to leverage heterogeneous data sources and improve accuracy [93], [94], [107], [109].

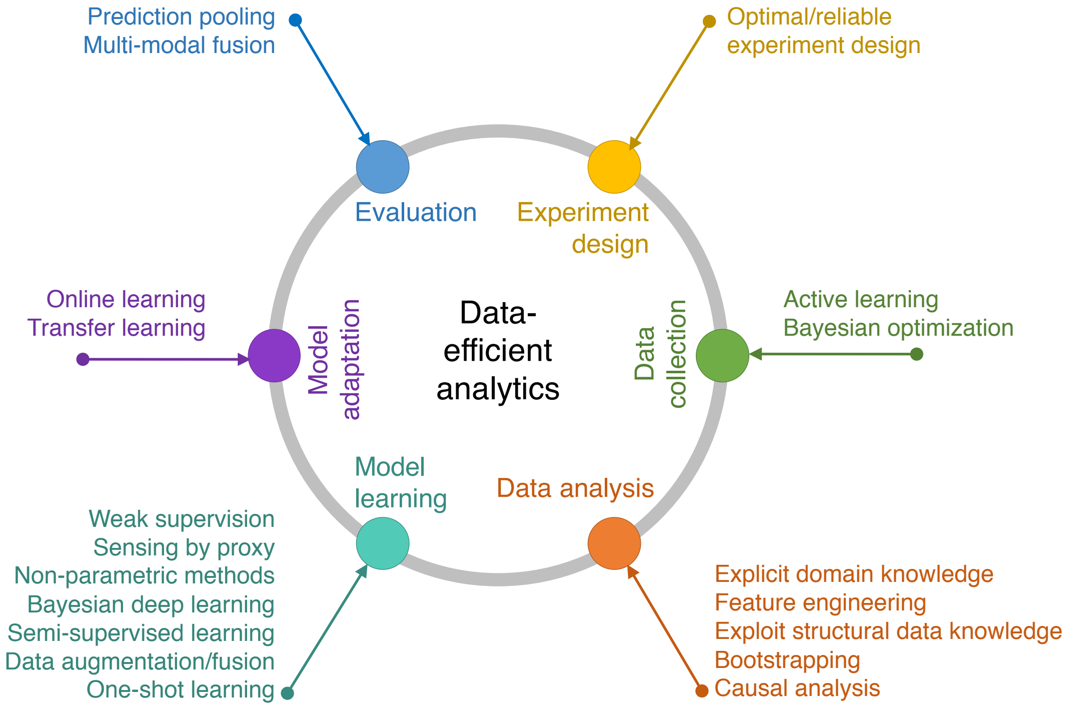


Figure 1.3: Data efficiency can be enhanced throughout the analytics pipeline, from experimental design, data collection and analysis to model learning, adaptation and evaluation.

Overall, the expectations are that a paradigm shift in data-efficient analytics will enable a control system to adapt to changing circumstances, detect regime changes with agility, draw actionable insights from limited amounts of information, and extend the applications of data-driven analytics to more vital services in h-CPS.

1.2 Contributions

> The goal of this thesis is to develop data-efficient analytics to improve the optimal performance of human-cyber-physical systems.

Human factor determination by proxy sensing:

Data analytics research for learning about the human factor is in its relative infancy. It is a challenging task, not just because the accuracy and reliability of sensing depends on sensor types, locations, data fusion and processing, and so on but also becuase human perceptions of the sensing system (e.g., that it is non-intrusive and non-interruptive) are critically important for practical implementation. One key contribution of my thesis is that it investigates a range of "proxies" for human factor determination, from environmental parameters (e.g., CO_2, magnetic field) [75], [94], [95] and in-vivo feedback (e.g., control actions) [96], to personal devices (e.g., smartphones) [90], [249] and bio-markers (e.g., skin temperature and heart rate) [109], as illustrated in Fig. 1.4.

> *By exploring the wide spectrum of proxies, we can extend the possibilities of human factor sensing, thus enabling a flexible and customized trade-off among cost, accuracy, availability, information granularity and privacy.*

To derive actionable information from proxy measurements, we have designed data-analytic algorithms that overcome some inherent drawbacks of proxy sensing, such as delayed response to, and indirect/time-changing relations with human factors. Consider occupancy determination (i.e., people counting in a room) in the context of smart buildings as an example. Indoor CO_2 is shown to be a good proxy for human presence because people naturally exhale CO_2; however, previous determination methods using CO_2 are slow in response to occupancy changes because CO_2 takes time to accumulate or dissipate. Sensing by proxy, as proposed in [95], is a sensing paradigm based on constitutive models that takes sensor delays into account and abstracts the underlying physical dynamics to make fast and reliable inferences (Chap. 2). In addition, works like the non-parametric algorithm, which can overcome the drift in "WiFi fingerprint" signals for robust and persistent indoor positioning [249], and sensor fusion frameworks based on wearable devices, which can overcome the limitations of individual sensors for human activity recognition [109], share a common theme of developing algorithmic solutions to address practical limitations, thus enabling accurate and reliable proxy sensing.

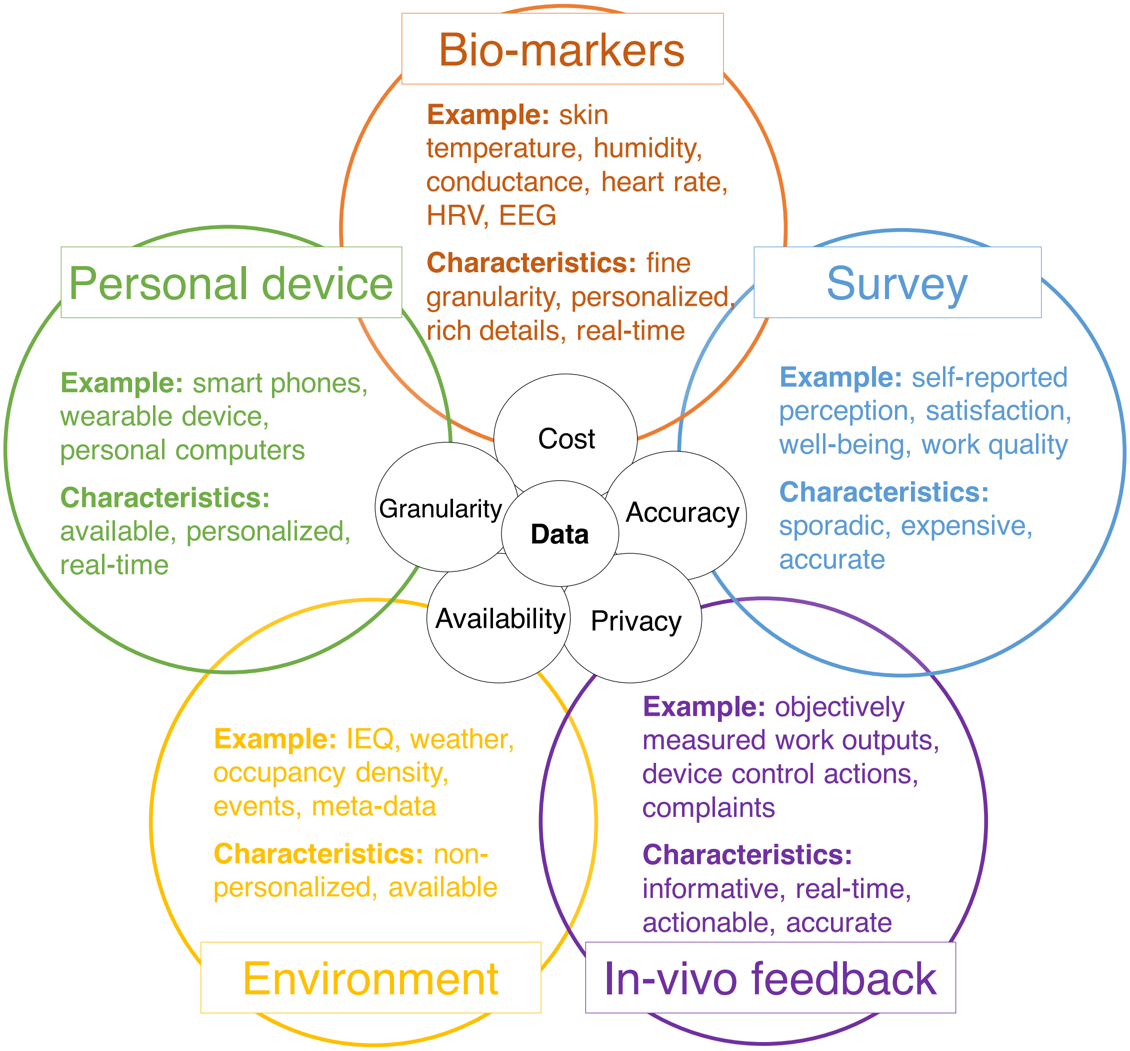

Bio-markers
Example: skin temperature, humidity, conductance, heart rate, HRV, EEG
Characteristics: fine granularity, personalized, rich details, real-time
Personal device
Example: smart phones, wearable device, personal computers
Characteristics: available, personalized, real-time
Survey
Example: self-reported perception, satisfaction, well-being, work quality
Characteristics: sporadic, expensive, accurate
Cost
Granularity
Data
Accuracy
Availability
Privacy
Example: IEQ, weather, occupancy density, events, meta-data
Characteristics: non-personalized, available
Environment
Example: objectively measured work outputs, device control actions, complaints
Characteristics: informative, real-time, actionable, accurate
In-vivo feedback

Figure 1.4: Human factors can be revealed through multiple channels and data sources, that achieve trade-offs among cost, accuracy, granularity, availability and privacy.

Label-free learning under weak supervision:

Previous learning methods for human factors predominantly involve supervised learning, which requires the collection of a sufficiently large *labeled* dataset for representation of common scenarios and generalization to unknown situations. This poses a practical challenge, particularly in the case of human factor determination, because people may be reluctant to label their data due to privacy concerns or simply be too busy to do so. To overcome this bottleneck, we have developed algorithms that can be trained with high-level heuristic rules rather than low-level labels, a form known as "weak supervision."

By incorporating domain knowledge in the form of high-level heuristics, we can

> *alleviate the need to collect low-level labels from people, thus enabling scalable human factor learning.*

Learning under weak supervision is related to—but clearly distinguished from—unsupervised or semi-supervised learning, particularly in the aspect that it leverages higher level heuristics as guidance. For example, the multi-view iterative training proposed in [100] uses heuristics to initialize noisy labels for an unlabeled dataset, iteratively refines the "weak" labels, and rules out absurd predictions to drive down the label noise until a termination condition is met or the learning converges. The high-level rules often stem from domain knowledge, and because people's behavior is fundamentally patterned and structured, such heuristics abound in numerous applications, from indoor positioning and occupancy detection to thermal comfort monitoring and activity recognition, thus significantly expand the possibilities of learning while requiring minimal user effort.

Inverse game theory and incentive design:

Previous approaches to designing and operating h-CPS treat people either as siloed agents or simply neglect the human factors. However, such neglect leads to inefficient operation and unexplored potential, because people are an integral part of the h-CPS nexus, and social interactions are central to human behaviors (see Fig. 1.1). We introduce the idea of gamification to h-CPS operation by wrapping services (e.g., lighting and temperature control in a building) as a game among people that explores and exploits the social dimensions (i.e., peer-pressure, collaboration, risk aversion and reward probability distortion) of human decision making. We also developed an inverse game theory framework to infer the strategies encoded in people's utility functions, further facilitating the design of incentives to nudge people in the desired direction.

> *By gamifying the parts of h-CPS operation that closely involve people and by learning their strategies in a game context, we can tap into the human factor potential—and further—tailor the incentives to enhance the system's efficiency and resilience.*

We lay the foundation for computational gamification by formulating the hierarchical control architecture as a Stackelberg game, where the participants are modeled as non-cooperative utility maximizers, and the leader can issue incentives to encourage desired behaviors [190]. Our contributions to inverse game theory include the parametric utility learning framework [190], and its robust version that employs constrained feasible generalized least squares estimation enhanced by bootstrapping to improve the forecasting capability [132]. We also extended the framework to model players with multiple modes of strategies using a probabilistic interpretation that combines multiple utility functions. This approach allows us to capture the fact that players' utility functions are not static but depend on their current states [134]. Motivated by observations during the test period, we further analyzed situation in which people naturally form "coalitions" during a game. This analysis led to proposing utility learning frameworks for the coalition game, which is shown to effectively capture the community dynamics within players [133]. Social games in the h-CPS context can be fun and also substantially improve energy efficiency, as our game for lighting controls in a smart office demonstrates [135], [190]. Furthermore, the utility functions learned during the game about users can be leveraged for socialized automation even after the game period.

Preference inference from behavior demonstrations:

When attempting to automate a physical system that matches people's desires, the traditional approach is to first solicit people's preferences and then encode them in the system's control logic. However, this approach is ineffective because there are myriad potential situations to enumerate, and people's survey responses contain implicit biases. A learning system that considers in-vivo feedback is clearly preferable because it can directly observe how people interact with the system and infer their preferences from their actual behaviors in specific contexts. This concept is commonly known as inverse reinforcement learning. However, previous algorithms are either limited in the representational power of complex preference functions or they require substantial amounts of demonstrations. Our proposal leverages a deep Bayesian network capable of representing highly complex functions while retaining trainability.

> *By simultaneously learning a good representation of the context and a preference function that rationalizes the behaviors, we can automate the system to match people's preferences after observing only limited demonstrations.*

Specifically, we model the agent in a Markov decision process using a reward function (i.e., preference) that depends on the current state. The agent takes a sequence of actions to maximize its total reward. Given a set of demonstrations that consist of state-action pairs, as well as the underlying dynamics of state transitions, the task of inverse reinforcement learning is to infer the reward function. Our contribution is to employ a deep Gaussian process to model the reward [96]. Because training the deep Gaussian process involves maximizing the likelihood function, which is intractable in its exact integral form, we propose an innovative variational training method that results in a tractable lower bound. The latent layers in the deep Gaussian process can capture complex feature dynamics and learn an effective representation, while the Bayesian training procedure acts to regularize the process to prevent overfitting and improves the generalization performance from limited demonstrations.

Leveraging flexibility and synergy from system integration:

Existing methods have treated subsystem control and optimization as "siloes" where people are treated as consumers with rigid demands. For example, consider energy systems. We envision transformations of both electrical and thermal energy into an integrated energy supply that optimizes its diverse resources from renewables, natural gas and power grid and leverages local demand flexibility.

> *By breaking the traditional "silos" in services and integrating subsystems during design and operation, we can leverage the flexibility and synergy that naturally emerge from optimization to improve overall system efficiency.*

This methodology has been employed for both smart building design [92] and microgrid operation [97], [98]. For smart building technology investment, we propose a platform-based design approach that abstracts building subsystems (e.g., lighting, HVAC, security, etc.) into several layers (functional design, module design, and implementation design) to facilitate a holistic design space exploration [92]. For microgrid operation, we model an integrated energy system (i.e., electrical and thermal) where diverse energy resources can be interconnected and coordinated to exploit the synergistic potential. Building owners are modeled as either utility maximizers who derive satisfaction from energy use [97] or as responsive consumers whose energy use portfolio consists of both critical demand and curtailable demand that is sensitive to price changes [98]. By formulating the dispatch problem as a mixed integer program, synergy and flexibility are naturally encoded in the optimized strategy, which results in improved efficiency and economy.

Cyber resilience analysis of h-CPS data analytics:

Data lie at the core of h-CPS and are used ubiquitously for estimation, optimization and control. Yet existing data analytics are seldom designed to be resilient to adversarial injections, rendering h-CPSs vulnerable to potential cyberattacks. Consider power grid state estimation, a key procedure conducted on a regular basis to filter and fuse various measurements collected from grid sensors to estimate the complex voltages at system buses. At present, bad data detection can filter invalid data due to sensor faults, but the filters can be evaded by systematically injected adversarial noise. Such attacks on data integrity may lead to system failure and financial loss; thus, they are worthy of analysis.

> *By envisioning the possible ways in which the integrity of data analytics can be attacked, we can evaluate a system's vulnerability and design effective countermeasures to enhance the resilience of h-CPSs to attacks.*

Specifically, we investigate the cyber resilience of power system AC-based state estimation [102], [103]. It was commonly held that due to the nonlinear physics and nonconvex formulation, it was difficult (if not impossible) to attack such systems using sparse injections without being detected. However, our analysis indicated that under convex relaxation by semidefinite programming, a sparse and stealthy attack can indeed be formed in polynomial time that can evade the current bad data detector. We thus propose to solve such programs to evaluate the cyber resilience of existing power systems against potential data integrity attacks. Our methodological contributions also include the design of a rank-1 penalty matrix and the derivation of performance bounds for the convexified problem.

1.3 Thesis overview

Learning about the human factor

People are an integral part of h-CPSs. Consider buildings as an example, where people spend about 90% of their time in various activities. People assume multiple roles in buildings, from "users" who simply enjoy building services to "operators" who control building lighting or temperature, to active "agents" who work and collaborate with colleagues, and "sensors" who provide feedback about current events. Human-centric values such as comfort, productivity, health and well-being are high-priorities for building managers.

> The goal is to efficiently and reliably learn about the human factor to improve human-centric values in the design and operation of h-CPSs.

Specifically, we examine the human factors in the context of a smart building, where people are engaged in different activities, and meanwhile relate to/interact with the building environment, system and other occupants, as illustrated in Fig. 1.5. A common theme that threads the aspects of learning is the process of *deriving actionable intelligence from limited data and resources (e.g., time, labor and cost).*

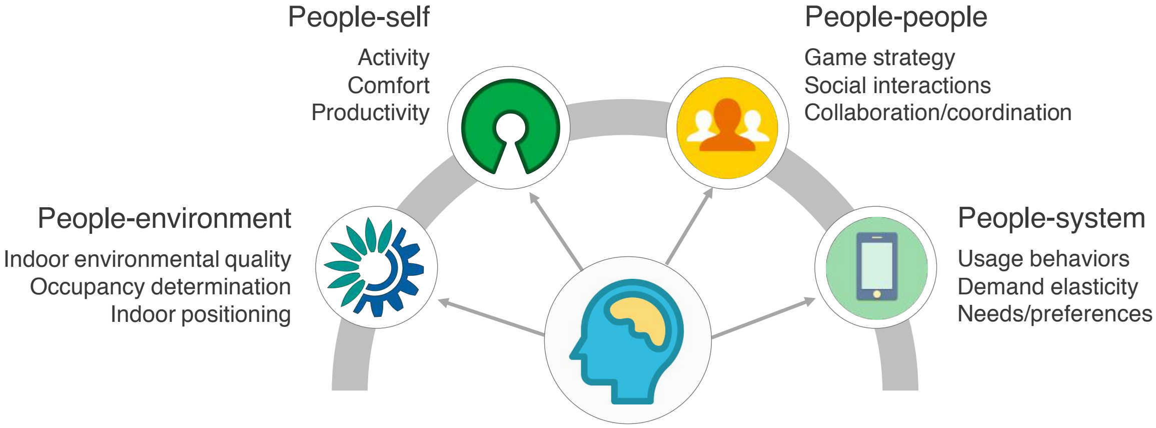


Figure 1.5: The multiple dimensions of human factors, delineated by people's interactions with the environment, the system, other people and themselves.

People in a built environment: The convergence of ubiquitous sensing and information technology enhances the awareness of the environment and the activities of people in the buildings. Traditional approaches rely on static sensor networks deployed at preselected locations, which often require tedious setup, calibration and maintenance.

> *We innovate sensing technology, making it more portable and agile.*

Along this line of research, our idea of a Building-in-Briefcase (BiB) is that "everything is in the briefcase, including the sensors and the router, and you can take the briefcase wherever you go" [228]. The BiB system is "trivially" easy to deploy in almost any building environment, making it easier than ever to monitor building conditions.[3] Indoor environmental quality is critical to people's comfort, health and well-being, but its measurements often require expensive sensors that are not scalable to cover a large area. Our proposed "automated mobile sensing" paradigm solves this issue by leveraging the mobility of a navigation-enabled sensor-rich robot to actively survey the indoor space [110]. The platform is agile to the dynamic changing environment due to its simultaneous localization and automatic mapping capability, which drives down the sensing infrastructure cost and frees users from laborious sensor calibration efforts. This line of work is based on collaboration with Kevin Weekly, Shichao Liu, Stefano Schiavon, Alexandre M. Bayen, and Costas J. Spanos.

Managed by the building operating system, data can be collected, stored and processed for building operation and fault diagnosis. Going beyond controls based on low-level data alone (such as temperature and illuminance), sensor measurements can be pooled and fused to reveal high-level human factors.

> *We enhanced the contextual awareness of buildings by sensing where people are, what they are doing, and to some extent, how they are feeling.*

Along this line of research, we investigated non-intrusive methods to determine room occupancy, people's indoor positions, activities and thermal comfort using wireless sensors (e.g., CO_2, smart meters) and mobile sensors (e.g., smart phones, wearables). For *occupancy determination*, we proposed sensing by proxy, a system that uses an ordinary differential equation coupled with a partial differential equation to sense indoor CO_2 concentration. This approach a has faster response rate and higher reliability than previously used machine learning models, and could be used to improve the efficiency of demand-controlled ventilation systems currently in use [94], [95].[4] We also explored the use of smart meter data to detect home occupancy without any initial information from a home owner [99], [100]. Using a weak supervision approach, the proposed algorithm can tease out detailed power consumption characteristics when a home is occupied; subsequently, it can determine when someone is home—even when that person's patterns are outside the norm.[5] The methods used for this aspect are discussed in Chaps. 2 and 3, and are based on joint work with Kevin Weekly, Nikos Bekiaris-Liberis, Ruoxi Jia, Alexandre M. Bayen, and Costas J. Spanos.

The WiFi signals from smartphones can be utilized to sense people's *indoor positions*; however, such signals often suffer from signal drift and are unstable. To address these issues, we developed a nonparametric method to adapt to the signal dynamics online [249], optimize the locations of WiFi routers [101], and investigated incorporating the floorplan to regularize the estimation [93], all of which have been demonstrated to improve the positioning

[3] "Brains for buildings, packaged in a smart briefcase", Berkeley Engineering magazine, Oct, 2017

[4] 'CO$_2$ sensor occupancy detection", CO2Meter.com, Feb, 2017

[5] "What does your smart meter know about you?", IEEE Spectrum, Jun, 2017

accuracy. In addition, we proposed a new system based on sound that is essentially a form of echolocation. The system can identify different rooms based on a relatively small dataset gathered in advance [90].[6] Regarding the determination of people's *indoor activities and thermal comfort*, we pioneered the work of tapping into the physiological signals of human body such as skin temperature and heart rate collected by wearable devices, and demonstrated prototypes that can make reliable inferences about indoor activities (e.g., sitting, running, climbing the stairs) and thermal comfort (i.e., individual responses to the current temperature conditions) [109]. While this part of our work, in collaboration with Han Zou, Ruoxi Jia, Shichao Liu, Stefano Schiavon, Lihua Xie and Costas J. Spanos, is not included in detail in this thesis, it is closely related to the methods introduced in Chaps. 2 and 3.

Utility learning in a social game: People are social beings who naturally interact with one another to connect, compete or collaborate. When immersed in an entertaining, real-world game, people tend to adopt a strategy that arises from their preferences and social inclinations and leads to certain actions. Economists have long studied and applied "nudges" such as default effect and distorted perception of reward probability to tap into social dimensions for marketing and policy making purposes. Utility functions are abstractions of individual preferences that have been widely used to rationalize people's behaviors.

> *In a social game designed to improve energy efficiency, we learned people's utility functions as a means to identify their strategies, predict their actions, and gently "nudge" them toward a desired outcome.*

A social game platform consisting of an intelligent lighting system and an online web portal was set up in the Center for Research in Energy Systems Transformation at UC Berkeley, where a group of about 20 people participated in a game to reduce lighting usage with a probability of winning a monetary reward. Chap. 4 of this thesis is dedicated to describing the theoretical framework of inverse game theory that rationalizes the observations we made during the game. We also introduce a hierarchical control structure based on a Stackelberg game and an incentive design scheme to advance building manager's goal [190]. This part of the thesis involved joint work with Ioannis C. Konstantakopoulos, Lillian J. Ratliff, S. Shankar Sastry, and Costas J. Spanos.

Preferences revealed from interactions: People interact with the system on a regular basis, such as adjusting the indoor lighting and temperature setpoint, or playing music that matches their mood while making a Doppio Espresso. Our key insight in Chap. 5 is to exploit the unique data resulting from personalized interactions with the system to gain insight into individuals' likes and dislikes, without explicitly querying them for such information.

[6] "An indoor positioning system based on echolocation", MIT Technology Review, Jul, 2014

> *We rationalized an agent's actions by modeling its preferences using a deep Bayesian network trained with inverse reinforcement learning.*

In our approach, the world is modeled as a Markov decision process, where the next state depends on the current state and action, and an agent takes a sequence of actions to collect rewards. Reinforcement learning solves for an optimal policy that maximizes the reward and assumes that the reward is given. Inverse reinforcement learning solves the opposite problem—reward specification, given the set of demonstrations. The advantage of using a deep Bayesian network to model the reward is that it can handle complexity through its representation power, while remaining sample-efficient due to the Bayesian regularization. Together with Andreas Damianou, Pieter Abbeel, and Costas J. Spanos, we introduced the deep Gaussian process for inverse reinforcement learning (DGP-IRL) algorithm and proposed a novel training method based on variational inequality [96].

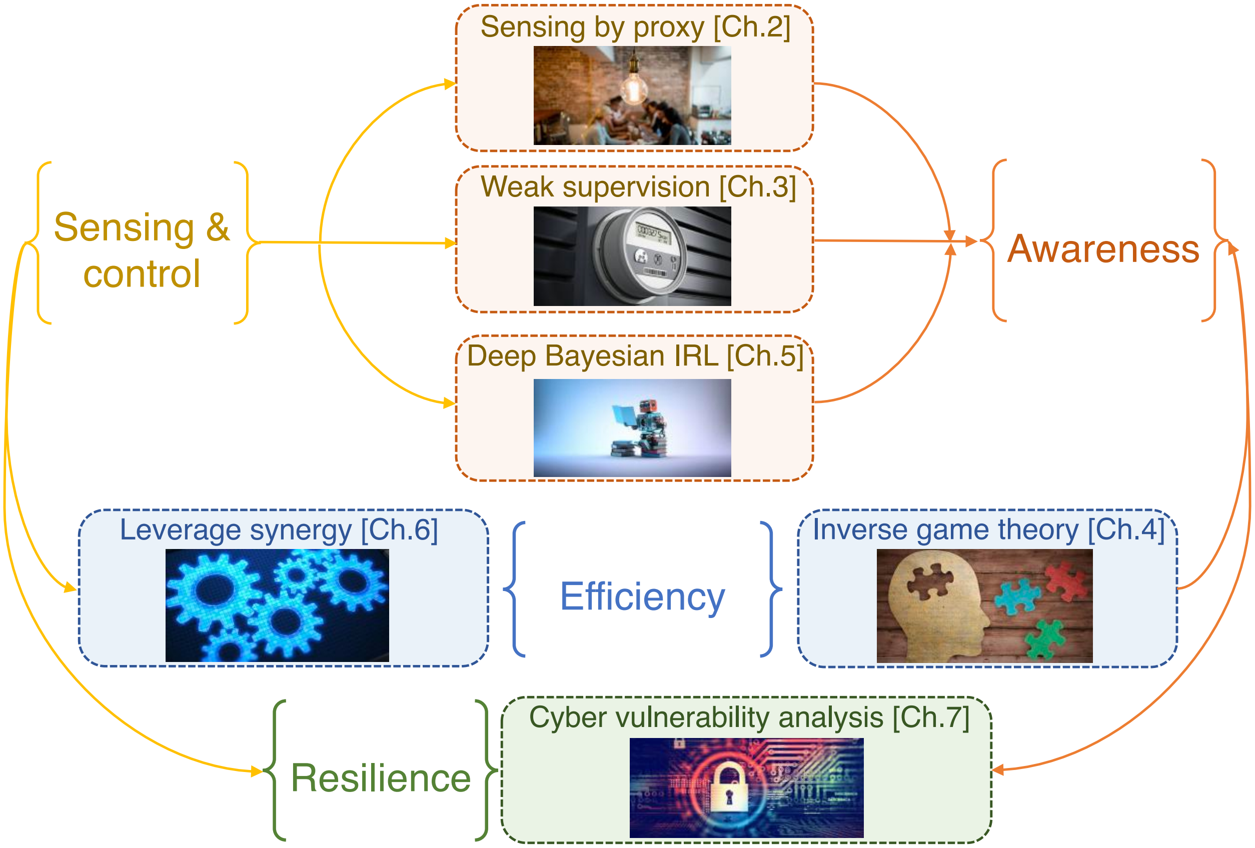


Figure 1.6: Thesis overview: chapters are organized with respect to the four key h-CPS modules: sensing and control, awareness, efficiency, and resilience.

System-level efficiency and resilience

Advancing from a traditional cyber-physical system view to one that respects the human factor in its design and operation, the first part of the thesis focuses on learning about the human factor and developing high sample-efficiency methods that enhance contextual awareness from ubiquitous sensing. This naturally leads to a quest for approaches that further enhance the efficiency and resilience, two important metrics of h-CPS optimal performance.

> Our key insight is to address the efficiency and resilience of h-CPS from a holistic perspective that weaves the cyber, physical and human factors into a single tapestry.

Thus, the second part of the thesis centers on the analysis and optimization of system-level efficiency and resilience, completing the four pillars of h-CPS performance, namely sensing and control, awareness, efficiency and resilience. The logical connections of the thesis chapters are illustrated in Fig. 1.6. Overall, we identify the four pillars as general modules in h-CPS operation.

Efficiency via flexibility and synergy: Efficiency stems from the use of the best available resources. In a typical h-CPS like a smart grid, efficiency means the ability to consume renewable energy when it is abundant, the ability to switch between different fuel sources such as coal and natural gas, and the ability to store surplus energy for use during peak hours—an operation called economic dispatch or unit commitment. However, renewable energy is stochastic and intermittent, and the system is limited by its capacity. In Chap. 6, we identify strategies to tap into the potentials of human factors through economic incentives, namely to price energy for the retail market, and of the synergy that stems from amalgamating the electricity market with the thermal energy market [97], [98], [105].

> *In a foreseeable energy retail market, we exploit the flexibility of end-user demand and the synergy from thermal-electrical coordination to promote renewable energy integration and system economy.*

For example, our strategy capitalizes on the so-called “spark spread,” which is the difference between the prices of natural gas and electricity. When the electricity price from the grid peaks, the district energy system operator can rely on its combined heat and power plant, which burns natural gas to produce electricity. Waste heat can be recycled to provide heating capacity or channeled to an absorption chiller to provide cooling. Our district system model is capable of exploring such interconnected energy flows among distributed energy resources and central plants; synergistic dispatch arises naturally from the optimization. Based on the economic theory of demand elasticity, we also optimize retail rates to achieve mutual benefits for both the customers and the retailer. This portion of the thesis contributes to the toolset that I developed while I was working at the Lawrence Berkeley National Laboratory with Wei Feng, Chris Marnay, Ping Liu and Costas J. Spanos.

Vulnerability of data analytics against cyberattacks: h-CPS operation has become increasingly reliant on data analytics results. Up to this point, the thesis has discussed approaches to enhance contextual awareness and system efficiency under an implicit assumption of trustworthy data. However, incoming data might be insecure at its source, during transmission, or at the point of analysis. If data integrity were to be compromised by an adversary, the vulnerability of the data analytics currently in use is unknown.

> *We analyzed the vulnerability of power grid state estimation against potential cyberattack as a means to evaluate cyber resilience of the grid.*

First, we showed that to conduct such an attack with limited numbers of sensor modifications and without being detected, the adversary needs to solve an optimization problem that is non-linear, non-convex and discrete, implying the computational barrier that naturally exists for state estimation based on an AC model. But it is far from the truth that such problem is tenable; we showed that an innovative relaxation based on semidefinite programming was able to provide a near-global stealthy and sparse solution to the original attack problem. For any grid topology and sensing infrastructure, the solution (e.g., the feasibility of the problem, or the number of sensors that need to be altered) can therefore serve as a realistic metric for power grid cyber resilience [102], [103]. The materials presented in Chap. 7 are based on joint work with Javad Lavaei and Karl H. Johansson.

Part I

Learning about the human factor

Chapter 2

Sensing by proxy

Human behaviors are often not directly observable. Nevertheless, the influences of humans on the environment can often be characterized and measured. Thus, by obtaining ambient information (such as those listed in Fig. 1.4), it is possible to non-intrusively discern the hidden factors without interfering with people's daily routines. To this end, we propose *sensing by proxy*, a sensing paradigm based on constitutive models that capture the physical processes of human influences on environmental parameters in order to infer the human factors.

2.1 System model with distributed sensor delays

Sensing by proxy employs a state observer applicable to multiple-input, multiple-output (MIMO), linear time-invariant (LTI) systems, where the sensor output channels have distributed delays. We consider the following system:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{K}\mathbf{u}(t) \tag{2.1}$$

$$y_i(t) = \int_0^{D_i} \mathbf{Q}_i(\sigma)\mathbf{x}(t-\sigma)d\sigma, \text{ for } i = 1, ..., m \tag{2.2}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the state at time $t > 0$, $\mathbf{u}(t) \in \mathbb{R}^p$ is the input, $y_i(t)$ is the i-th sensor output (out of m in total) with delay $D_i > 0$, and $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{K} \in \mathbb{R}^{n\times p}$ and $\mathbf{Q}_i \in \mathbb{R}^{1\times n}$ are coefficient matrices.

Such a system can be viewed as the dual of a predictor-based controller for a system whose input has a delayed effect on the state, a problem that has been studied in population dynamics [10] and liquid mono-propellant rocket motors [237] to name just a few. In this realm of research, delay compensation for linear [10], [137], [237] and nonlinear systems [120], [157] has been achieved using predictor-based techniques. Sensing by proxy employs an observer equivalent to the predictor feedback design for the case of distributed sensor delays, and the exponential convergence of the estimation error is guaranteed (see Theorem 2.1). The main difference of our approach and prior work is the use of infinite-dimensional forwarding-backstepping transformation of the infinite-dimensional actuator states, because the traditional backstepping method is inapplicable for both single-input systems with distributed input delays and for multi-input systems with different delays. This enables us to apply an ordinary differential equation (ODE) coupled with a partial differential equation (PDE) system to capture the temporal and spatial dynamics and effectively regularizes the inference output.

Sensing by proxy is clearly different from the discriminative models in ML widely used to infer human behaviours, which assume that measurements are independent and identically distributed—and perhaps more markedly, that a sufficiently large dataset of *labeled data* is available for model training [138], [140], [147], [242]. Compiling such labeled datasets is often impractical because they require substantial human labor and cost to record the ground truth data, whose collection processes inevitably raise privacy concerns [88]. To reduce the data-labeling costs, sensing by proxy is derived from physical models that capture the human influence through succinct representations (i.e., the parameters can be identified with a small amount of data), and are thus more accurate and reliable for inference.

2.2 Sensing by proxy methodology

Sensing by proxy is a physics-inspired inference method capable of real-time detection. The core is *an observer-based detector for a MIMO, LTI system with distributed sensor delays.*

Proxy design and modeling

A proxy is a phenomenon that reveals about human factors to some extent. To steamline the presentation, we will focus on environmental parameters, such as temperature, humidity, and CO_2, but the framework is applicable (can be extended) to other phenomena. We model the dynamics of the proxy parameters using a MIMO, LTI system with distributed delays in the sensing channels.

The source term, $\mathbf{x}(t) \in \mathbb{R}^m$, comprises m proxy measurements at time t in their respective units. The state $\mathbf{x}(t)$ is the output of an LTI system whose input, $\mathbf{v}(t) \in \mathbb{R}^m$ represents the (unknown) humans' effect. The relation is characterized by the following ODE:

$$\dot{\mathbf{x}}(t) = -\mathbf{A}\mathbf{x}(t) + \mathbf{v}(t), \tag{2.3}$$

where the matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$ characterizes the inertia of proxy phenomena. The human factors change at discrete events, and remain relatively constant between two adjacent events. This is represented by the form of a piece-wise constant signal,

$$\dot{\mathbf{v}}(t) = \mathbf{0}, \tag{2.4}$$

which is congruent with our experimental observation that the response of the proxy parameters due to changes of the human's presence has some similarities with the step response of a low-pass filter.

The ODE is coupled with a PDE that models the evolution of the proxy variables:

$$\mathbf{u}_t(s,t) = -\mathbf{B}\mathbf{u}_s(s,t) + \mathbf{B}_X\mathbf{x}(t) \tag{2.5}$$
$$\mathbf{u}(0,t) = \mathbf{U}_0(t) \tag{2.6}$$
$$\mathbf{u}(1,t) = \mathbf{U}_1(t) \tag{2.7}$$

where $\mathbf{u}(s,t) \in \mathbb{R}^m$ denotes the proxy at time $t \geq 0$ and for $0 \leq s \leq 1$, $\mathbf{u}_t(s,t), \mathbf{u}_s(s,t)$ are standard notations for partial derivatives with respect to t and s, and $\mathbf{U}_0(t), \mathbf{U}_1(t) \in \mathbb{R}^m$ are the sensor measurements at time t. For a physical process, parameter matrix $\mathbf{B} = \text{diag}(b_1, \ldots, b_m)$ represents the speed of convection, and $\mathbf{B}_X \in \mathbb{R}^{m\times m}$ is the rate of dispersion.

Proxy inference

The latent human factors $\mathbf{v}(t)$ can be inferred by proxy measurements $\mathbf{U}_0(t)$ and $\mathbf{U}_1(t)$. For compactness, define $\mathbf{z}(t) = \begin{bmatrix}\mathbf{x}(t)\\ \mathbf{v}(t)\end{bmatrix} \in \mathbb{R}^{2m}$, and rewrite (2.3) and (2.4) as:

$$\dot{\mathbf{z}}(t) = \bar{\mathbf{A}}\mathbf{z}(t), \tag{2.8}$$

where $\bar{\mathbf{A}} = \begin{bmatrix}-\mathbf{A} & \mathbf{I}_{m\times m}\\ \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m}\end{bmatrix}$ with $\mathbf{A}$ from (2.3). Similarly, (2.5) can be recasted as:

$$\mathbf{u}_t(s,t) = -\mathbf{B}\mathbf{u}_s(s,t) + \mathbf{B}_Z\mathbf{z}(t), \tag{2.9}$$

where $\mathbf{B}_Z = \begin{bmatrix}\mathbf{B}_X & \mathbf{0}_{m\times m}\end{bmatrix}$ is the augmented matrix of $\mathbf{B}_X$. Consider the following observer:

$$\hat{\mathbf{u}}_t(s,t) = -\mathbf{B}\hat{\mathbf{u}}_s(s,t) + \mathbf{B}_Z\hat{\mathbf{z}}(t) + \mathbf{r}(s)\mathbf{L}\left(\mathbf{U}_1(t) - \hat{\mathbf{u}}(1,t)\right) \tag{2.10}$$
$$\hat{\mathbf{u}}(0,t) = \mathbf{U}_0(t) \tag{2.11}$$
$$\dot{\hat{\mathbf{z}}}(t) = \bar{\mathbf{A}}\hat{\mathbf{z}}(t) + \mathbf{L}\big(\mathbf{U}_1(t) - \hat{\mathbf{u}}(1,t)\big) \tag{2.12}$$

where $\mathbf{r}(s) \in \mathbb{R}^{m\times 2m}$, $\mathbf{L} \in \mathbf{R}^{2m\times m}$ are yet to be determined, and $\mathbf{U}_0(t)$, $\mathbf{U}_1(t)$ are the measurements of environmental parameters at two separate locations. We use $[\hat{\mathbf{B}}_X]_{i,:}$ to denote the i-th row of $\hat{B}_X$, and the hat notation to indicate estimated quanity. The following result guarantees the exponential convergence rate of human factor estimation error [17] [94]:

Theorem 2.1. *Consider the system* (2.10)–(2.12), *where*

$$\mathbf{r}(s) = \begin{bmatrix}\mathbf{r}_1(s) & \cdots & \mathbf{r}_m(s)\end{bmatrix}^\top \tag{2.13}$$
$$\mathbf{r}_i(s) = \left(\mathbf{C}_i - \int_0^{(1-s)/b_i} [\mathbf{B}_Z]_{i,:}e^{-\bar{\mathbf{A}}y}dy\right) e^{\bar{\mathbf{A}}(1-s)/b_i} \tag{2.14}$$
$$\mathbf{C}_i = \int_0^{1/b_i} [\mathbf{B}_Z]_{i,:}e^{-\bar{\mathbf{A}}\sigma}d\sigma, \quad i = 1, ..., m \tag{2.15}$$

Let the pair $(\bar{\mathbf{A}}, \bar{\mathbf{C}})$ *be observable, where* $\bar{\mathbf{C}} = \begin{bmatrix} \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_m \end{bmatrix} \in \mathbb{R}^{m \times 2m}$*, and choose* $\mathbf{L}$ *such that the matrix* $\bar{\mathbf{A}} - \mathbf{L}\bar{\mathbf{C}}$ *is Hurwitz. Then, for any* $\mathbf{z}(0) \in \mathbb{R}^{2m}$*,* $u_i(s,t)$*,* $\hat{u}_i(s,t) \in L^2(0,1)$*,* $i = 1, ..., m$*, where* u_i *is the* i*-th component of* $\mathbf{u}$*, there exist positive constants* λ *and* κ *such that the following holds for all* $t \geq 0$

$$\Omega(t) \leq \kappa \Omega(0) e^{-\lambda t}, \tag{2.16}$$

where

$$\Omega(t) = \int_0^1 \|\mathbf{u}(s,t) - \hat{\mathbf{u}}(s,t)\|^2 ds + \|\mathbf{z}(t) - \hat{\mathbf{z}}(t)\|^2. \tag{2.17}$$

Proof. See Appendix A.1. □

2.3 Application: occupancy detection in buildings

Intelligent buildings are conscious about both its occupants and environments in order to optimize user comfort and energy efficiency. Occupancy information (i.e., how many people in each room) can be used in several value-added ways:

- *To monitor occupant-specific energy usage*: total energy consumption can be assigned to individuals using occupancy data [32], and information such as individual energy efficiency and entropy can be used to classify occupants consumption behaviors [77];
- *To improve occupant behavior modeling*: study how the connectivity and interaction among people can be used to improve energy saving [33], or use the data to validate existing or future occupancy models for different types of buildings [84], [91];
- *To drive real-time building automatic controls and adaptive services*, such as demand-controlled ventilation, and "geo-fencing" [5], [48], [61], [183], [201].

Occupancy can be detected in various ways, such as passive infrared (PIR) [4], [5], camera [28], [48], [61], [85], sound [90], [213], [220], pressure sensors [170], electricity meters [100], [130], and environmental measurements like particulate matters (PM2.5), CO_2, temperature, and humidity [24], [50], [227], [139], and even wireless traffics [21], [155], [201], [250]. They can be broadly categorized based on the information granularity (as listed in Table 2.1):

- Level 1 and 2 – presence (whether the room is occupied or not) and count (how many people are inside). Almost all the sensors can provide presence information, but several methods fail to provide counts, such as PIR, which can only sense the motion.
- Level 3 – activity (what the occupants are engaged in, e.g., having a meeting or working on a computer). This can be provided by, for example, ambient sound or PC usage.

- Level 4 and 5 – identity and tracking (of each occupant), such as using camera, sound and WiFi signal for indoor localization.

Based on this taxonomy, we provide a simple, working definition of "non-intrusiveness":

> *Occupancy sensors are non-intrusive if they can provide information up to the activity level (Level 3) but not the identity or tracking levels (Level 4 and 5).*

Table 2.1: Survey of occupancy sensing methods and their capability of providing different levels of information granularity.

	Ref.	Presence	Count	Activity	Identity	Tracking
CO_2	[24], [25], [50], [139], [161]	✓	✓	✗	✗	✗
PIR	[4], [5]	✓	✗	✗	✗	✗
Power meter	[100], [130]	✓	✗	✓	✗	✗
Pressure	[170]	✓	✓	✗	✓	✗
Camera	[28], [48], [61], [85]	✓	✓	✓	✓	✓
Sound	[90], [213], [220]	✓	✓	✓	✓	✓
Wireless signal	[21],[155], [201], [250]	✓	✓	✗	✓	✓
PC usage	[215]	✓	✓	✓	✓	✗

Sensing by proxy for occupancy detection

Among all the "non-intrusive" sensing parameters, indoor CO_2 represents a good proxy for occupancy, as humans naturally exhale CO_2 and are the main source of its indoor variations. CO_2 concentration is a good indicator for indoor air quality, which has been found to influence productivity [66], [70]. CO_2 sensors have also been integrated in some commercial sensors and heating, ventilation, and air conditioning (HVAC) systems. However, CO_2 based detection might be influenced by people's individual characteristics like gender and physiques, or opening/closing of windows. And previous methods have slow response rate to occupancy changes (e.g., when the air feels "stuffy", the occupants might be there for at least some time) [24], [25], [50], [139], [161]. Sensing by proxy represents a new category of methods [1],

> The proposed "Sensing by proxy" model is more accurate than previously used machine learning models, and could be used to improve the efficiency of Demand-Controlled Ventilation systems (DCV) currently in use.

[1] 'CO_2 Sensor Occupancy Detection", CO2Meter.com, Feb, 2017 [Accessed: 12/1/2017]

Sensing by proxy is based on a physical model that captures the dynamics of CO_2 concentration inside a typical room, where the fresh air is brought in at the supply vent and exhausted at the return vent on the ceilings (see Fig. 2.1a). As people breath, the warm air that contains CO_2 rises like a bubble to the ceilings through mixing and convection effects. A CO_2 sensor is placed at the exhaust vent to measure the occupancy effects. More specifically, as reduced from (2.3) – (2.6), an ordinary differential equation coupled with partial differential equation is used to capture the CO_2 dynamics:

$$\dot{x}(t) = -ax(t) + v(t) \tag{2.18}$$

$$\dot{v}(t) = 0 \tag{2.19}$$

$$u_t(s,t) = -bu_s(s,t) + b_X x(t) \tag{2.20}$$

$$u(0,t) = U_0(t) \tag{2.21}$$

$$u(1,t) = U_1(t) \tag{2.22}$$

where both $x(t)$ and $v(t)$ are scalars. As illustrated in Fig. 2.1b, the ODE (2.18) is used to model human's CO_2 production on the concentration in their local vicinity, and the effect rate is specified by the time constant, a (unit is 1/second). As delineated in the PDE, the highly concentrated air of CO_2 then diffuses to the environment at rate b_X (1/second), and is exhausted by the return vent through convection at rate b (1/second).

The corresponding observer can be obtained by the general result in Theorem 2.1:

$$\hat{u}_t(s,t) = -b\hat{u}_s(s,t) + \begin{bmatrix} b_X & 0 \end{bmatrix} \begin{bmatrix} \hat{x}(t) \\ \hat{v}(t) \end{bmatrix} + \mathbf{r}(s) \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \big(U_1(t) - \hat{u}(1,t)\big) \tag{2.23}$$

$$\hat{u}(0,t) = U_0(t) \tag{2.24}$$

$$\begin{bmatrix} \dot{\hat{x}}(t) \\ \dot{\hat{v}}(t) \end{bmatrix} = \begin{bmatrix} -a & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x}(t) \\ \hat{v}(t) \end{bmatrix} + \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \big(U_1(t) - \hat{u}(1,t)\big) \tag{2.25}$$

where $\mathbf{r}(s) = \begin{bmatrix} \pi_1(s) & \pi_2(s) \end{bmatrix}$, and

$$\pi_1(s) = \frac{b_X}{a}\big(e^{\frac{a}{b}s} - 1\big) \tag{2.26}$$

$$\pi_2(s) = \frac{b_X}{ba}s + \frac{b_X}{a^2}\big(1 - e^{\frac{a}{b}s}\big) \tag{2.27}$$

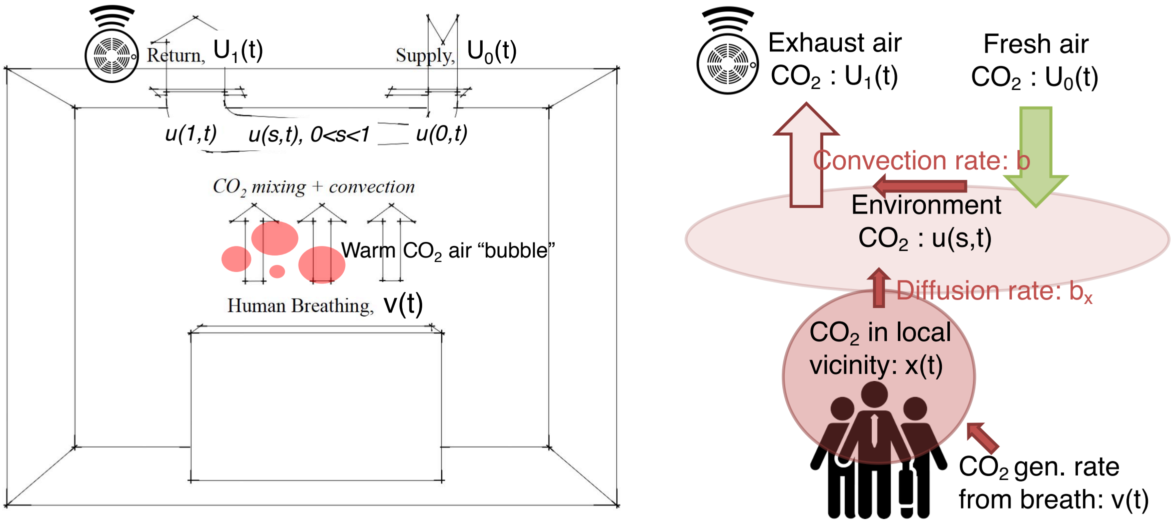
Corollary 2.1. *Consider the system (2.18)–(2.22) and the observer (2.23)–(2.27). Let $b_X \neq 0$ and choose L_1, L_2 such that the matrix $\bar{\mathbf{A}} - \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \mathbf{C}_1$ is Hurwitz, where $\bar{\mathbf{A}} = \begin{bmatrix} -a & 1 \\ 0 & 0 \end{bmatrix}$, and $\mathbf{C}_1 = \begin{bmatrix} \pi_1(1) & \pi_2(1) \end{bmatrix}$. Then for any $x(0)$, $\hat{x}(0)$, $v(0)$, $\hat{v}(0) \in \mathbb{R}$, there exists positive constant λ and κ such that the following holds for all $t \geq 0$,*

$$\Omega(t) \leq \kappa\Omega(0)e^{-\lambda t} \tag{2.28}$$

$$\Omega(t) = \int_0^1 \big(u(s,t) - \hat{u}(s,t)\big)^2 ds + \big(x(t) - \hat{x}(t)\big)^2 + \big(v(t) - \hat{v}(t)\big)^2 \tag{2.29}$$

Proof. See Appendix A.1. □

The corresponding occupancy detection algorithm is illustrated in Fig. 2.2. To implement Sensing by proxy, the CO_2 concentration is measured at the exhaust vent, and applied the observer (2.23) – (2.25) to obtain the CO_2 production rate $\hat{v}(t)$ (unit is ppm/second). This rate is then passed through a median filter for smoothing, and normalized by human breathing rate (0.183ppm/(second·person)) to obtain occupancy estimation.[2]



(a) Physical representation of a typical room. (b) Illustration of human CO_2 effect.

Figure 2.1: Physical illustrations of the model. Fresh air with CO_2 level $U_0(t) = 400$ ppm enters the room from the supply vent, and exits the room after convection and mixing with human breath $v(t)$. The condition of the air at the return vent $U_1(t)$ is measured.

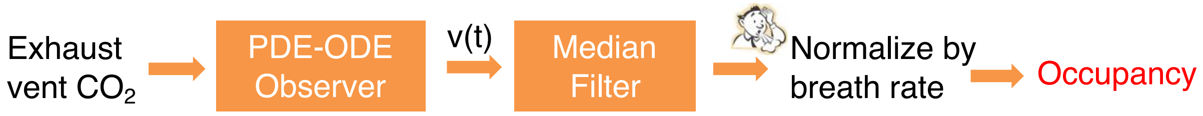


Figure 2.2: Sensing by proxy algorithm illustration.

Experimental evaluation

The testbed is set up in a conference room at UC Berkeley, where a single CO_2 sensor is placed at the return vent (Fig. 2.3). To conduct the experiment, we asked several people to stay in and out of the room following a predefined schedule, and measured the indoor CO_2

[2] The source code and data can be accessed at: https://github.com/jinming99/Sensing-by-proxy

concentration continuously. Fig. 2.4a shows how the CO_2 concentration evolved over time. We can notice that when the number of occupants was high, there was a high rate of increase and also a high level of concentration of CO_2. But, it took time for CO_2 to accumulate or deplete in space. Consequently, different occupancy levels can correspond to similar CO_2 concentration, and vice versa, which fundamentally limits most ML methods or rule-based methods that rely on CO_2 concentration level.

On the contrary, Sensing by proxy captured the indoor CO_2 dynamics, as shown in Fig. 2.4a. Given the occupancy (input $v(t)$), the PDE–ODE model (2.18) – (2.22) accurately predicted about the CO_2 concentration in the space (output $u(1, t)$, see the "simulated return" vent CO_2 in Fig. 2.4a). This consequently results in better estimation of occupancy, as shown in Fig. 2.4b. Sensing by proxy has a fairly fast response to occupancy changes, and the inference is quite close to the ground truth. This method is also robust to non-uniformity of physiques, as exhibited in the subject group. We also evaluated Sensing by proxy (SbP) against other ML methods, such as Bayes Net, Multi-layer Perceptron (MLP), and found that the performance is much better on average according to the root mean squared error metric. Fig. 2.5 compares the actual estimation by Sensing by proxy with Bayes Net, which is best among ML methods, where we plot the percentage of the estimated occupancy vs. the true occupancy. While Bayes Net sometimes made large errors, as confounded by the CO_2 levels (for example, when the room is full of people, it indicates that the room is vacant, so the ventilation is actually turned off), Sensing by proxy is reliable almost all the time, and the error is typically bounded by 1 person. This is important for efficient ventilation strategies to ensure occupant comfort.

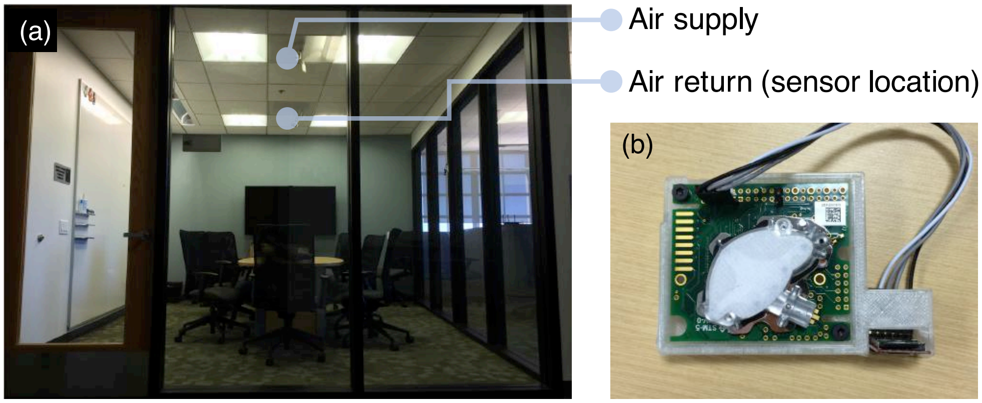
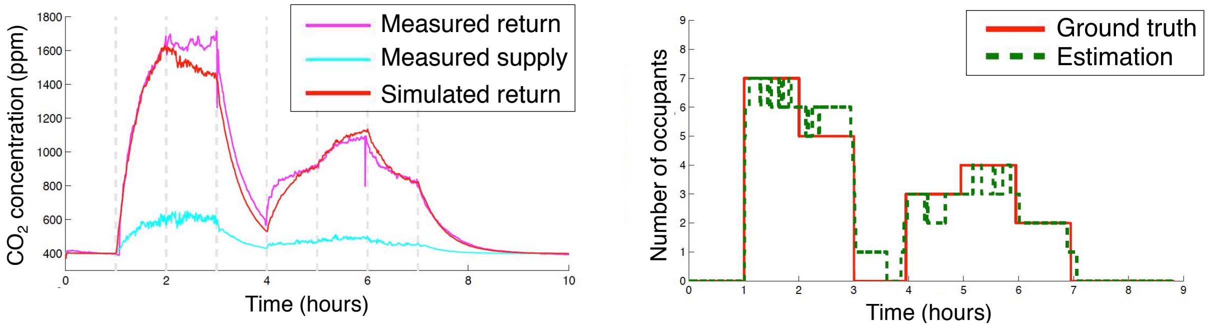


Figure 2.3: (a) The testbed is a conference room of size $14 \times 10 \times 9$ ft^3, equipped with a full ventilation system including an air return vent and air supply vent, as illustrated in Fig. 2.1a. (b) CO_2 sensor up close, which is placed at multiple locations (supply vent, return vent, and blackboard); however, for occupancy detection in real-time, we only need to measure the CO_2 level at the return vent.



(a) Given occupancy, simulate return vent CO$_2$. (b) Given CO$_2$ measurement, infer occupancy.

Figure 2.4: Simulation result and occupancy detection by SbP for Exp. C. Parameters: $a = 0.06$ sec^{-1}, $b = 2.5$ sec^{-1}, $b_X = 1.5$ sec^{-1}. The response time is less than few minutes.

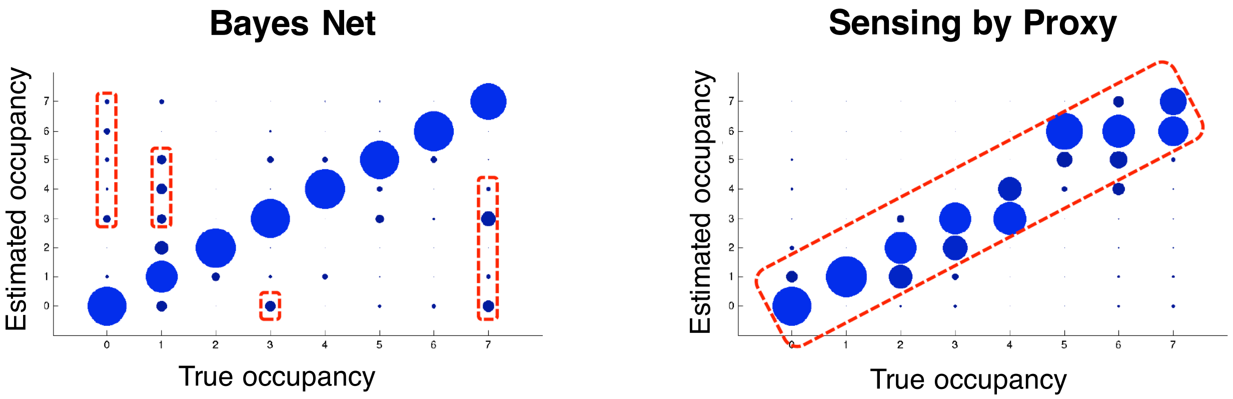


Figure 2.5: Visualization of confusion matrix for Bayes Net (left) and SbP (right), where the position of blue circles represents the true occupancy (x-axis) and estimated occupancy (y-axis), and the size indicates occurrence frequency. Bayes Net makes nonnegligible errors (red rectangle), whereas SbP performs reliably with errors bounded within the ±1 region.

Table 2.2: Comparison of root mean-squared error of estimation with other models in occupants experiments. Details for Exp. A, B, C and ML methods can be found in [94].

	Naïve Bayes	Bayes Net	MLP	RBF	Logistic	SMO	AdaBoost	SbP
Exp. A	1.3	1.2	1.0	1.1	1.1	1.2	1.6	**0.6**
Exp. B	0.7	0.7	0.6	0.7	0.6	0.6	0.7	**0.5**
Exp. C	1.7	1.5	1.6	1.6	1.5	1.6	2.4	**0.6**
Average	1.2	1.1	1.1	1.1	1.1	1.1	1.6	**0.6**

2.4 Chapter summary

The sensing by proxy paradigm described in this chapter is a latent variable determination method based on proxy measurements governed by constitutive models. Because physical phenomena often require time to take effect, we model the system with ordinary differential equations with sensing delays and show that it is equivalent to an ordinary differential equation coupled with a partial differential equation. A system observer is designed that uses multiple time series of sensor measurements as input and outputs an estimate of the latent state—which is the goal of inference.

We demonstrated an application for occupancy determination in a built environment. Occupancy determination is critical for enabling demand-controlled ventilation and lighting and to enhance the contextual awareness of building automation systems. The results showed that by monitoring the CO_2 concentration inside a room, sensing by proxy can reliably and quickly determine the number of people inside the room. The system outperformed typical machine learning algorithms, particularly with regard to its robustness in the presence of large errors; sensing by proxy managed to limit the magnitude of error to 1 person, while the comparison algorithms resulted in large errors that can directly lead to inefficient operations.

One key aspect in which sensing by proxy differs from other data-driven algorithms is its sample efficiency. Because sensing by proxy requires only a few model parameters, training the model is relatively easy and fast. Because the model is constitutive and physics-based, the parameters can be also estimated using simulation programs. Another promising approach to estimate the model is to employ methods of learning under weak supervision as discussed in Chap. 3. The idea behind this approach is to use high-level heuristic rules to initialize the occupancy data. For example, a rule might indicate high occupancy when the CO_2 level exceeds 3,000 ppm, or a rule might use intrinsic information such as meeting calendars to indicate the number of people in the conference room. Using such heuristics can substantially reduce the need to collect occupancy ground truth and simplify practical implementations.

Chapter 3

Learning under weak supervision

We switch gears to take a different approach to data-efficient analytics that can potentially reduce or even eliminate the need for labeled data by leveraging weak supervision heuristics/constraints (such as human behavioral patterns, physical rules and biological constraints). Contemporary supervised learning algorithms (such as deep learning) rely on various strategies to search for an *empirical risk minimizer (ERM)* [19]:

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} l(f(\mathbf{x}_i), y_i), \tag{3.1}$$

where $\mathcal{F}$ is a function space, $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is a loss function (such as the 0-1 loss $l_{01}(y, f(\mathbf{x})) = \mathbf{1}(y \neq f(\mathbf{x}))$ for classification), $\mathbf{x}_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$ are the input (a feature vector) and output (a label) of the i-th data point, which together form an n-point dataset $\mathcal{S} = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$. The *regret* characterizes how "poorly" the learner has performed compared to the performance of the best learner in the class:

$$\delta R_{l,\mathcal{F}}(f) = \underbrace{R_{l,\mathcal{F}}(f)}_{\text{risk of } f} - \underbrace{\inf_{f' \in \mathcal{F}} R_{l,\mathcal{F}}(f')}_{\text{risk of the best detector}}, \tag{3.2}$$

where $R_{l,\mathcal{F}}(f) = \mathbb{E}_{f \in \mathcal{F}}\left[l(y, f(\mathbf{x}))\right]$ is the risk of model f, which is empirically minimized in ERM (3.1). A tight upper bound on the regret suggests that the algorithm has performance guarantee when used in practice. While breakthroughs in AI thus far are highly dependent on the availability of massive datasets (such as IBM Watson winning at Jeopardy! (2011) after 8.6M documents were made accessible by Wikipedia and Project Gutenberg (2010) and Google's GoogLeNet (2014) after the release of 1.5M labeled images by ImageNet (2010)), there is an emerging consensus among ML researchers that acquiring labeled data is a major bottleneck to further advances. This chapter is dedicated to describing a new learning paradigm based on weak supervision. We also describe algorithms to enable learning under weak supervision.

3.1 Overview of weak supervision

The idea of learning under weak supervision is simple:

> Rather than soliciting low-level, accurate, but expensive labels from users or domain experts, one can employ higher-level, noisier—but cheaper (even free)—heuristics to initialize an unlabeled dataset.

In its general form a heuristic appears as conditional probability distribution $\mathbb{P}(\tilde{y}|t(\mathbf{x}))$, where $t(\cdot)$ is any transformation on the feature $\mathbf{x} \in \mathcal{X}$ (such as taking a subset or low-dimensional embedding). Throughout this section, we use $\tilde{y}$ to denote a weak label generated from heuristics, and y to refer to a ground-truth label. By accepting and recognizing the noise in the "weak labels", we can then modify the methods used to train the algorithm (such as the iterative training procedure in Sec. 3.2 or the surrogate loss approach in Sec. 3.3) to construct a more refined and customized model.

By and large, many lines of ML research have shared the fundamental goal of making effective use of scarce labeled data and abundant unlabeled data (see Fig. 3.1). A broad category of methods termed *unsupervised learning* aims at uncovering the hidden structures in data by learning compressed yet informative representations. Typical examples are principle component analysis [112], the EM algorithm for clustering [27], low-dimensional embedding [199] causal analysis [244], and, more recently, the auto-encoder [126] and generative adversarial nets [73]. Another broad category is *semi-supervised learning*, which leverages unlabeled data (as a regularizer during training, or to learn a compact representation) to improve the efficacy of labeled data (see [29] for a good overview). *Active learning* also recognizes the difficulty of obtaining high-quality labels; instead, it solicits only the most valuable labels to improve the model's decision-making capability [54]. *Transfer learning* tackles the data efficiency issue by transferring data or models across several datasets, or by jointly learning several tasks to regularize the model [26], [180]. Various methods have

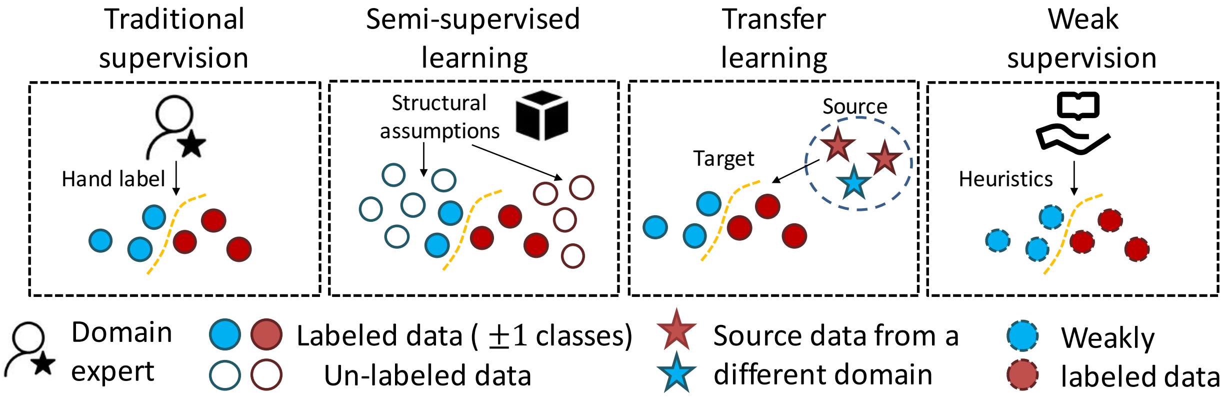


Figure 3.1: Overview of four lines of ML to tackle the data scarsity issue.

generalized supervised learning to allow for more label possibilities. In *multiple-instance learning*, a label is provided for a group of objects that holds for at least one object in the group [46]. Labels can be also expressed as ranks among candidates, such as the use of clickthrough data to improve the retrieval quality of search engines [111]. While the boundaries of these categories are blurry and hybrid approaches are possible (and often beneficial in practice), the main difference of learning under weak supervision is the *incorporation of high-level knowledge into the learning process*.

An emerging body of literature has employed weak supervision rules in learning. The use of explicit constraints on the output space to force it to have a particular meaning or semantics or to follow certain physical laws have been demonstrated to significantly reduce the need for labeled training data [208]. Data programming uses a labeling function to automatically process unlabeled data and learns a generative model to resolve the output of these labeling functions [193]. A probabilistic framework has been described to model the noise simultaneously during training [235]. In the following, we will describe both a multi-view iterative training (MIT) algorithm that refines noisy labels during each iteration until convergence (Sec. 3.2), and a surrogate loss approach that accounts for noisy distributions by modifying the loss function to be unbiased in expectation (Sec. 3.3).

Because human behaviors are fundamentally patterned and structured [114], learning under weak supervision has wide applications to h-CPSs. We will present a use case involving smart meter data analytics that automatically learns occupant behaviors from raw smart meter data streams.

3.2 Multi-view iterative training

We assume that a task can be partitioned into several "views", each carries some relevant information: $\mathcal{X} = \mathcal{X}_1 \bigtimes \cdots \bigtimes \mathcal{X}_k$, where $\mathcal{X}$ is the entire information space and $\mathcal{X}_j$ is the j-th view (out of k in total). For instance, to infer about the individual thermal acceptance (i.e., whether the user feels the thermal condition is acceptable or not [35]), one "view" can be the time of the day, one "view" can be the *set of* physiological signals collected from personal wearables, one "view" can be the activity picked up by smart phones, yet another "view" can be the control signals issued to the user's personal fans/heaters (in-vivo feedback). We recognize that *some views might be easier to support heuristics than others* (for example, when the personal heater is turned on, it is likely that the user is discontented with the current condition. And the person might regularly feel hot or cold during the early morning/late night.) Define the set of heuristics as $\mathcal{H}$, which consists of rules h_j for view $\mathcal{X}_j$ that can be in the form of a conditional probability $h_j(\tilde{y}, \mathbf{x}) = \mathbb{P}(\tilde{y}|t_j(\mathbf{x}))$ for $t_j(\mathbf{x}) \in \mathcal{X}_j$ that processes the relevant view for each data point. The combined heuristic is thus

$$h(\tilde{y}, \mathbf{x}) = \frac{\prod_{h_j \in \mathcal{H}} h_j(\tilde{y}, \mathbf{x})}{\sum_{\tilde{y}} \prod_{h_j \in \mathcal{H}} h_j(\tilde{y}, \mathbf{x})}. \tag{3.3}$$

The key idea of multiview-iterative training (MIT) is to *initialize the unlabeled data points probabilistically with heuristic rules, and iteratively refine the labels until convergence or some stopping signals* [99]. For each iteration, we resolve the conflicts by the majority rule and probabilistic update. To streamline the presentation, we focus on the case of binary classification (the label y is either $+1$ or -1), but the procedure is applicable to the general multi-class classification:

1. *Initialization:* Initialize training set labelsby the multi-view heuristics h in (3.3).

2. *Multiview training:* For rounds $t \in \{1, 2, ...\}$, train the classifier using all the views (either in a single classifier or multiple classifiers pooled by the majority rule), and determine the new "guesses" for the data, partitioned into $\hat{\mathcal{L}}_{-1} = \{(\mathbf{x}_i, \tilde{y}_i) | \tilde{y}_i = -1\}$ and $\hat{\mathcal{L}}_{+1} = \{(\mathbf{x}_i, \tilde{y}_i) | \tilde{y}_i = +1\}$ for data with weak labels of -1 and $+1$, respectively.

3. *Weak label updates:* Perform the following update:

$$\mathcal{L}_y^{t+1} = \{\mathcal{L}_y^t \cap \hat{\mathcal{L}}_y\} \cup \mathit{Sample}\{\mathcal{L}_y^t \Delta \hat{\mathcal{L}}_y; \alpha_y\}, \text{ for } y \in \{-1, +1\} \tag{3.4}$$

   where $\mathcal{L}_y^t = \{(\mathbf{x}_i, \tilde{y}_i) | \tilde{y}_i = y\}$ is the set of data whose weak labels are y in the current round t, $\mathcal{L}_y^t \Delta \hat{\mathcal{L}}_y$ is the symmetric difference set operation, and $\alpha_y \in (0, 1)$ is the sampling rate for label $y \in \{-1, +1\}$ that controls the probabilistic update.

4. *Stopping condition:* Stop the iteration whenever the labels do not change, or the condition (3.8) in Theorem 3.2 is satisfied.

The weak label update rule (3.4) keeps the weak labels that are "agreed upon" in two successive rounds $\{\mathcal{L}_y^t \cap \hat{\mathcal{L}}_y\}$, and resolves the conflicts by randomly sampling α_y portion of the weak labels from the "controversial" set $\{\mathcal{L}_y^t \Delta \hat{\mathcal{L}}_y\}$. The parameter α_y thus controls the trade-off between "learning speed" and "weak label noise". The stopping condition (3.8) is triggered when it seems unlikely that additional iterations can improve the accuracy, as proved in Theorem 3.2. MIT operates in an environment where the noise in the training set can not be ignored, which has been studied in the probably approximately correct (PAC) framework [221].

Theorem 3.1. *[221] If we draw a sequence of*

$$n \geq \frac{2}{\epsilon^2 (1 - 2\eta)^2} \log\left(\frac{2N}{\delta}\right) \tag{3.5}$$

samples from a distribution and find any hypothesis $\hat{f} \in \mathcal{F}$ that minimizes disagreement with the training labels, where ϵ is the hypothesis worst-case classification error rate, η is the upper bound on the training noise rate, N is the number of possible hypotheses in the class $\mathcal{F}$, and δ is the desired confidence level, then the following PAC property is satisfied:

$$\mathbb{P}\Big(d\big(\hat{f}, f^*\big) \geq \epsilon\Big) \leq \delta \tag{3.6}$$

where $d(\cdot,\cdot)$ is the sum over the probability of elements from the symmetric difference set labeled by hypothesis $\hat{f}$ and the optimal decision f^*.

The above theorem provides a high probability bound on the classification error ϵ that depends on the training noise rate η to be estimated. We partition the weak labeled set $\mathcal{L}^t_{-1} = L^t_{-1,\checkmark} \cup L^t_{-1,\times}$ by the correctly-labeled set $\mathcal{L}^t_{-1,\checkmark}$ (true negative) and incorrectly-labeled set $\mathcal{L}^t_{-1,\times}$ (false negative), and similarly we partition $\mathcal{L}^t_{+1} = \mathcal{L}^t_{+1,\checkmark} \cup \mathcal{L}^t_{+1,\times}$ by the correctly-labeled set $\mathcal{L}^t_{+1,\checkmark}$ (true positive) and incorrectly-labeled set $\mathcal{L}^t_{+1,\times}$ (false positive). Let $\mathcal{U}^t = \mathcal{U}^t_{-1} \cup \mathcal{U}^t_{+1}$ be a partition of the unlabeled dataset. Then the training noise rate η_t exhibited in the weakly labeled dataset $\mathcal{L}^t_{+1} \cup \mathcal{L}^t_{-1}$ is given by:

$$\eta_t = \frac{|\mathcal{L}^t_{-1,\times}| + |\mathcal{L}^t_{+1,\times}|}{|\mathcal{L}^t_{-1}| + |\mathcal{L}^t_{+1}|}, \tag{3.7}$$

where $|\mathcal{L}^t_{+1}|$ denote the set cardinality. Let ϵ_t be the error rate of the current classification model to be estimated. Then the training noise rate η_t and model classification error rate ϵ_t in the t-th iteration can be estimated as follows.

Lemma 3.1. *The training noise rate η_t and classification error rate ϵ_t can be estimated with the access to any two of the following (approximated) quantities:*

1. *The number of negative samples in the dataset $|\mathcal{L}^t_{-1,\checkmark}| + |\mathcal{L}^t_{+1,\times}| + |\mathcal{U}^t_{-1}|$*
2. *The number of negative samples in the labeled set $|\mathcal{L}^t_{-1,\checkmark}| + |\mathcal{L}^t_{+1,\times}|$*
3. *The number of positive samples in the dataset $|\mathcal{L}^t_{+1,\checkmark}| + |\mathcal{L}^t_{-1,\times}| + |\mathcal{U}^t_{+1}|$*
4. *The number of positive samples in the labeled set $|\mathcal{L}^t_{+1,\checkmark}| + |\mathcal{L}^t_{-1,\times}|$*
5. *The misclassification rate for the positive samples $|\mathcal{L}^t_{-1,\times}| / \left(|\mathcal{L}^t_{-1,\times}| + |\mathcal{L}^t_{+1,\checkmark}|\right)$*
6. *The misclassification rate for the negative samples $|\mathcal{L}^t_{+1,\times}| / \left(|\mathcal{L}^t_{+1,\times}| + |\mathcal{L}^t_{-1,\checkmark}|\right)$*

Proof. See Appendix A.2. □

Based on Lemma 3.1, we can estimate the *classification noise rate* η_t in the t-th round, which can be then used in the stopping rule to guarantee that the performance of the classification can only improve in each round before program termination.

Theorem 3.2. *The gap between the learned and optimal hypotheses in the PAC property (3.6) will decrease with high probability in each iteration with suitable sampling rates, α_{-1} and α_{+1}, whenever the following condition is satisfied:*

$$\left(|\mathcal{L}^{t+1}_{-1}| + |\mathcal{L}^{t+1}_{+1}|\right)(1 - 2\eta_{t+1})^2 > \left(|\mathcal{L}^t_{-1}| + |\mathcal{L}^t_{+1}|\right)(1 - 2\eta_t)^2 \tag{3.8}$$

where $\left(|\mathcal{L}^{t+1}_{-1}| + |\mathcal{L}^{t+1}_{+1}|\right)$ is the total number of weakly labeled samples in round $t+1$, and η_{t+1} is the (estimated) training noise rate.

Proof. See Appendix A.2. □

Theorem 3.2 suggests a *stopping indicator* as follows:

$$\mathbb{I}\left\{\left(|\mathcal{L}_{-1}^{t+1}|+|\mathcal{L}_{+1}^{t+1}|\right)(1-2\eta_{t+1})^2 \leq \left(|\mathcal{L}_{-1}^{t}|+|\mathcal{L}_{+1}^{t}|\right)(1-2\eta_{t})^2\right\} \tag{3.9}$$

which evaluates to 1 when the condition in (3.8) is violated. This can be evaluated during training time after each iteration. Frequent violations of (3.8) might be a strong indication to stop the algorithm and avoid potential deterioration, since it is no longer guaranteed that the weak labels are better in the next iteration. It is possible to apply this result to other iterative algorithms with weak supervision.

3.3 Learning with surrogate loss

The MIT algorithm proposed in Sec. 3.2 modifies the training procedure to be iterative in order to improve the high-level heuristics provided by weak supervision. Another approach is to modify the loss function directly to account for the presence of label noise. Let $\rho_{-1} = \mathbb{P}(\tilde{y} = +1|y = -1)$ and $\rho_{+1} = \mathbb{P}(\tilde{y} = -1|y = +1)$ denote the conditional label noise, where y is the truth label and $\tilde{y}$ is the weak label. We can design the loss function as a substitute for the original loss to "clean up" the noise in expectation. The procedure of learning with surrogate loss is described below (details can be found in [99]):

1. *Initialization:* Initialize training set labels by the multi-view heuristics h in (3.3).

2. *Cross-validation:* For each candidate $\theta \in \Theta^{CV}$ (the hyperparameter θ can be (ρ_{+1}, ρ_{-1}) in (3.10) or γ in (3.11)), obtain the empirical risk following the standard cross-validation procedure using the surrogate loss, then select the best candidate $\hat{\theta}$.

3. *Learning with surrogate loss:* Using the surrogate loss, train the model with labels initialized by h and obtain the predicted labels.

The initialization step is the same as MIT. The cross-validation step is used to select the best candidate in the set Θ^{CV}, since the conditional noise rates of weak labels ρ_{+1} and ρ_{-1} are unknown. Once the best hyperparameters are identified, the corresponding surrogate loss function can be minimized during training. We consider two surrogate loss functions – one depends on the conditional noise rates [196]:

$$\tilde{l}(f(\mathbf{x}), \tilde{y}) = \frac{(1-\rho_{-\tilde{y}})l(f(\mathbf{x}), \tilde{y}) - \rho_{\tilde{y}} l(f(\mathbf{x}), -\tilde{y})}{1-\rho_{+1}-\rho_{-1}} \tag{3.10}$$

where $\rho_{\tilde{y}}$ is the conditional noise rate, and $l(\cdot,\cdot)$ is the original loss function; and one depends on the average label noise [171]:

$$l_\gamma(f(\mathbf{x}), \tilde{y}) = (1-\gamma)\mathbb{I}\{\tilde{y} = +1\}l(f(\mathbf{x}), \tilde{y}) + \gamma\mathbb{I}\{\tilde{y} = -1\}l(-f(\mathbf{x}), \tilde{y}) \tag{3.11}$$

where γ is the weight chosen according to the conditional noise rates $\frac{1-\rho_{+1}+\rho_{-1}}{2}$, and $\mathbb{I}\{\}$ is the identity function which evaluates to 1 when the inside condition is satisfied. The surrogate loss functions are designed such that the procedure of seeking the empirical risk minimizer (3.1) under the weakly supervised distribution is as if we are working with the original loss function under the "clean" distribution, where the labels are the ground truth (the derivation of (3.10) is obtained in the Appendix A.2). The next theorem provides an upper bound on the regret (3.2) using the surrogate loss function (3.10).

Theorem 3.3 ([171])**.** *Let $l(t, y)$ be L-Lipschitz in t for every y, then with probability at least $1-\delta$, and $\hat{f} = \arg\min_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \tilde{l}(f(\mathbf{x}_i), y_i)$ be the ERM with the weakly-labeled data,*

$$\delta R_{l,\mathcal{F}}(\hat{f}) \leq 4L_\rho \mathcal{R}(\mathcal{F}) + 2\sqrt{\frac{\log(1/\delta)}{2n}} \tag{3.12}$$

where $\mathcal{R}(\mathcal{F}) = \mathbb{E}_{\mathbf{x}_i,\epsilon_i}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(\mathbf{x}_i)\right]$ is the Rademacher complexity of the function class $\mathcal{F}$ with ϵ_i as the i.i.d. Rademacher (symmetric Bernoulli) random variables [14], and $L_\rho \leq 2L/(1-\rho_{+1}-\rho_{-1})$ is the Lipschitz constant of $\tilde{l}$ given in (3.10).

Theorem 3.3 suggests that the upper bound on regret decreases as we have more samples, whose infimum depends on the Lipschitz constant and complexity of the function class only. As the proposed procedure acts like "exploration in the darkness", the result offers performance guarantee; nevertheless, the precondition is specified that neither $\rho_{\pm 1}$ is greater than 0.5. Indeed, the bound improves as the conditional noise rates reduce. We will refer to (3.10) as the *unbiased loss* (U. L.), since the expectation of the original loss $l(\cdot,\cdot)$ under the true label distribution is identical to that of the surrogate loss $\tilde{l}(\cdot,\cdot)$ under the weak label distribution [196]. The γ*-weighted label-dependent loss* is designed in a similar fashion, except that the risk $R_{l_\gamma,\mathcal{F}}(f)$ now is an affine transformation of the original risk $R_{l,\mathcal{F}}(f)$:

Lemma 3.2 ([171])**.** *There exists a constant B that is independent of f such that by choosing $\gamma = \frac{1-\rho_{+1}+\rho_{-1}}{2}$ and $A_\rho = \frac{1-\rho_{+1}-\rho_{-1}}{2}$, and for all functions $f \in \mathcal{F}$,*

$$R_{l_\gamma,\mathcal{F}}(f) = A_\rho R_{l,\mathcal{F}}(f) + B \tag{3.13}$$

Intuitively, the loss puts more weights on data with labels that have higher confidence of being correct. With the chosen γ, optimization with $\tilde{l}_\gamma$ is equivalent to that with the original loss due to the affine relation (3.13). The next theorem gives performance guarantee.

Theorem 3.4 ([171])**.** *Let L be the Lipschitz constant for $l(\cdot,\cdot)$ as before, and let*

$$\hat{f} = \arg\min_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \tilde{l}_\gamma(f(\mathbf{x}_i), y_i), \tag{3.14}$$

then with probability at least $1-\delta$,

$$\delta R_{\tilde{l}_\gamma,\tilde{D},\mathcal{F}}(\hat{f}) \leq 4L\mathcal{R}(\mathcal{F}) + 2\sqrt{\frac{\log(1/\delta)}{2n}} \tag{3.15}$$

where $\mathcal{R}(\mathcal{F})$ is the Rademacher complexity.

There are two keys for using the surrogate loss approach in the weak supervision framework. First, the initialization of the training set is based on a common occupancy schedule, which provides useful information for human behavior mining. Secondly, the surrogate loss is designed such that the expectation of the objective under the weak label distribution is equivalent to that of the original problem. Hence, the solution is unaffected by the noise introduced in the weak heuristic rules.

3.4 Application: smart meter data analytics

Utility companies had installed 65 million smart meters as of 2015, covering more than half of U.S. households, and the number was projected to reach 90 million by 2020 [216]. Smart meter data contain rich and useful information, and have been anlyzed to carry out tasks like load disaggregation [89], [119], [182], household activity recognition [31], [100], [108], user segmentation for demand response [9], [118], and electricity theft detection [43], to name just a few. Among all the applications, knowing whether the user is at home is useful not just for home automation and intrusion detection [159], but also for utilities, who can call or show up to perform necessary maintenance when knowing that the user is home, and not waste personnel time trying to reach the user (see Fig. 3.2 for an overview of the smart meter presence detection system).

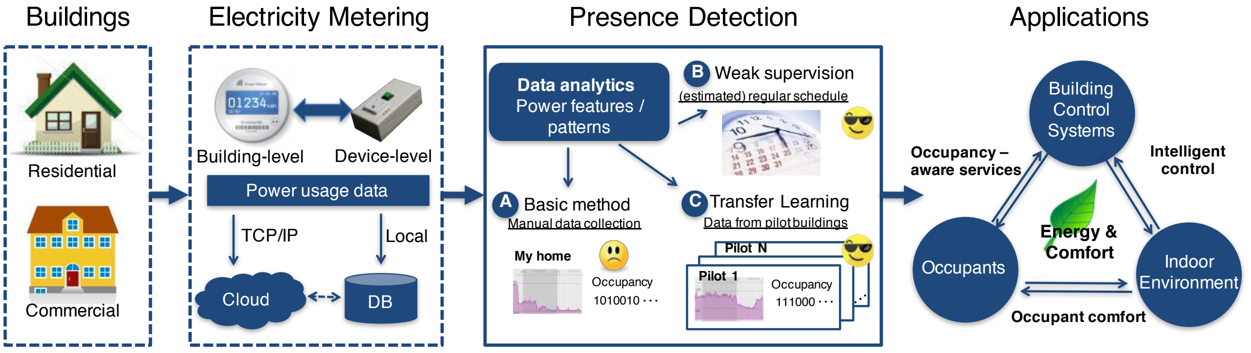
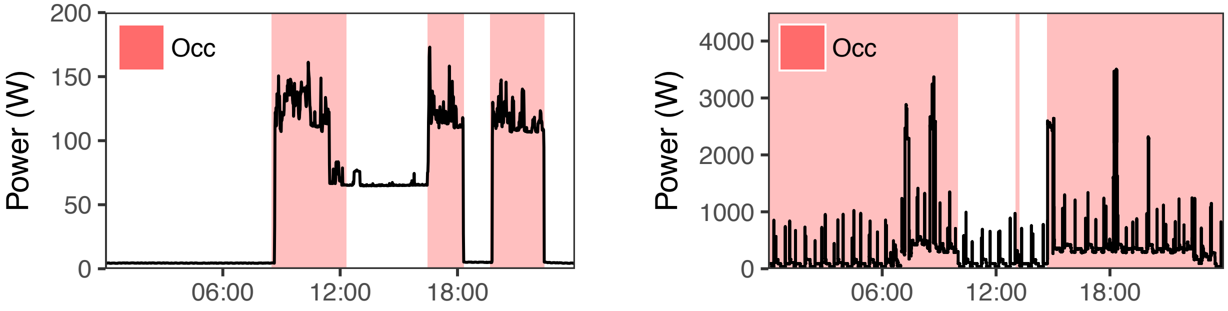


Figure 3.2: Smart meter data for household presence detection.

While previous works that employed power measurements for home occupancy detection assumed abundant labeled data from households [30], [50], [129], [130], [167], [238], collecting the occupancy data is laborious and difficult (if not impossible). A systematic approach that relaxes the requirement of occupancy data collection is desirable for practical reasons. The method “PresenceSense” was the first to infer home occupancy without collecting any labeled data from users [100], based on three main observations (see Fig. 3.3):

1. Energy consumption differs markedly when a building is occupied or vacant.

2. Office (residential) buildings are usually vacant (present) during non-business hours.



(a) Typical office cubicle.

(b) Typical household.

Figure 3.3: Daily power consumption of (a) an occupant in a commercial building, and (b) a household. The red color indicates user presence.

3. As noted by Nobel laureate Kahneman in his book "Thinking, Fast and Slow" [114], people's behaviors are patterned (the "slow" system) with spontaneous deviations (the "fast" system). These patterns are consistent over time.

The above observations imply a high-level weak supervision heuristic that assigns high probability of vacancy (occupancy) during non-business hours for commercial (residential) buildings, and high probability of occupancy (vacancy) during the rest of the day. This eliminates the need to collect presence data for an extended period of time to train the algorithms, as has been done in most previous studies [30], [50], [129], [130], [238]. As reported by an IEEE Spectrum article on our work,[1]

> In a recent paper, Jin and his colleagues demonstrated that machine learning systems can be trained to detect occupancy without any initial information from a home owner. "You just need a smart meter that listens over time," he says, "as well as the basic assumption that different types of buildings have different occupancy patterns, for example, commercial buildings are typically occupied during the day and not the night and homes are the opposite." Using this assumption, the machine learning algorithms were able to tease out more detailed characteristics about power consumption when a home is occupied; they then are able to tell when someone is home or not, even when that person's patterns are outside the norm.

Experimental evaluation

To conduct experiments of learning under weak supervision, we used the publicly available Electricity Consumption and Occupancy (ECO) dataset, which consists of fine-grained electricity and occupancy measurements for five Swiss households (id: r1, ..., r5) during summer

[1] "What Does Your Smart Meter Know About You?", IEEE Spectrum, Jun, 2017 [Accessed: 12/1/2017]

of 2012 [130]. We also employed the UMass Smart* home dataset [30] from two homes during the summer in western Massachusetts, and the personal cubicle (PC) dataset for 5 users in a campus building [100]. Aggregate power was collected by off-the-shelf digital electricity meters at a sampling rate of 1 Hz for ECO and Smart*. Occupancy information was entered manually by residents using the tablet mounted near the main entrance. Details about the households (number of occupants, types of devices, etc.) and data preprocessing techniques were described in [30], [100], [130].

Results for PC and ECO are reported in Figs. 3.4 and 3.5 for MIT, unbiased loss (U. L., (3.10)), and γ-weighted loss (3.11). As illustrated in Fig. 3.6, we implemented Naïve Bayes in each iteration of MIT, and terminated the iteration based on the stopping conditions. LibSVM was used to realize the γ-weighted loss with radial basis function (RBF) kernel. The logistic loss was selected for the unbiased loss, optimized by Nesterov's accelerated gradient method [174]. For comparison, we implemented simple threshold models are based on magnitudes (Mag/Th), changes in power magnitude (Chg/Th), and changes in percentage (Prc/Th), where the thresholds can be optimized over the training set as detailed in [100]. We also used the static schedule as the baseline, which indicates occupancy (vacancy) from 8am to 6pm for the PC (ECO), and vacancy (occupancy) for the rest of the day.

We employed the Matthews correlation coefficient (MCC), as suggested in [130] as a balanced measure of the prediction quality to overcome the difficulties in comparing different sizes of positive and negative instances:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP, TN, FP, FN are the numbers of true positive instances, true negative instances, false positive instances (estimation is occupancy when the ground truth vacancy,), and false negative instances (estimation is vacancywhen the ground truth is occupancy). The MCC returns a value between -1 and +1: a coefficient of +1 represents a perfect prediction, a coefficient of 0 represents random prediction, and a coefficient of -1 indicates total disagreement between prediction and observation. The true positve rate $(TP/(TP + FN))$ and true negative rate $(TN/(TN + FP))$ were also calculated. Generally, all methods deliver satisfactory performances compared to the static schedule and threshold models. However, due to the occupancy variability and device diversity, ECO is more challenging than PC. According to the true negative rate (TNR) and true positive rate (TPR), the γ-weighted loss seems to be a better absence detector, and unbiased loss excels at presence detection (Fig. 3.5). Overall, the proposed NL methods significantly outperformed the baseline models in the PC dataset, but only slightly exceeded the baseline models in the ECO dataset. The TNR/TPR rates report for ECO (Fig. 3.5) indicates that the weak supervision methods had relatively low TNR, implying that more mistakes were made when the households are vacant. This is mainly due to: (1) user behavior – sometimes the power consumption was high or has considerable fluctuations despite user absence, (2) the balance of dataset – users in some ECO households tended to be present in late mornings or early afternoons, as shown in the example traces (Figs. 3.3 and 3.7), resulting in more instances of occupancy than vacancy data [130]. While (1) illustrates the fundamental limitation in the proposed approach of using only power data to indicate user presence, (2) can be resolved by using existing methods such as putting a larger penalty on vacancy misclassification to improve the TNR performance [149].

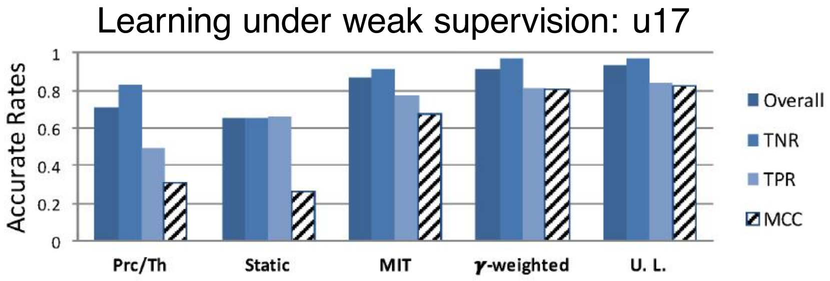


Figure 3.4: Results for user u17 in the PC dataset by 10-fold cross-validation, including baseline methods, MIT, γ-weighted, and unbiased losses.

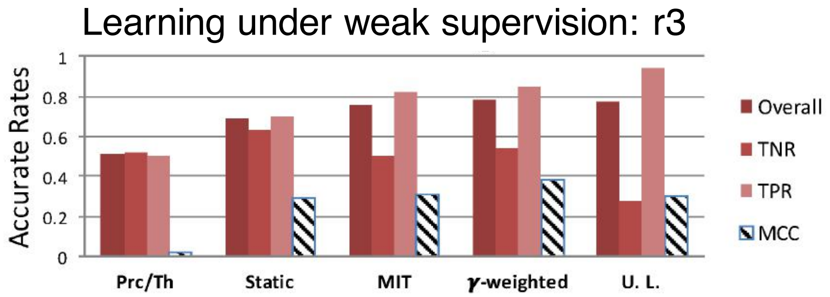


Figure 3.5: Results for household r3 in the ECO dataset by 10-fold cross-validation.

Examples of real-time occupancy detection are demonstrated in Fig. 3.7 for both commercial and residential buildings. The results suggest that power-based detectors mainly relied on power magnitudes, power transient features (short-term fluctuation captured by standard deviation) and power transition features (changes in power magnitudes in a sustained period of time). Compared to thresholding methods, the weakly supervised models were more effective in occupancy detection, especially for the ECO dataset where periodic power surges and diversified device usage patterns were common. In Fig. 3.8, we further evaluated the weak supervision model by plotting the estimated and true occupancy profile, which consists of the probability of the user being present at each hour throughout the day. Closely following the ground truth, the learned occupancy schedule significantly improved over the weak heuristics provided by for initialization.

Furthermore, we compared the best results from previous work that also used the ECO and Smart* datasets for evaluation [30], [129], [130]. For supervised learning, the methods

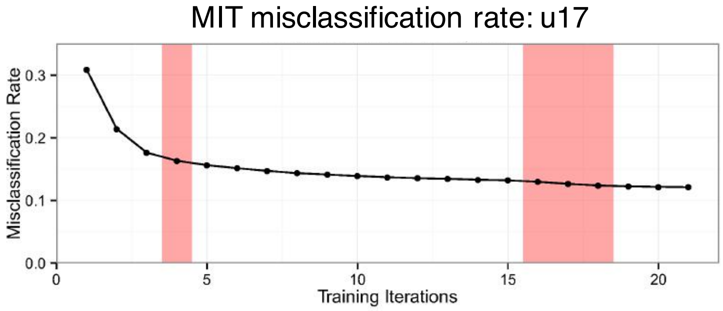
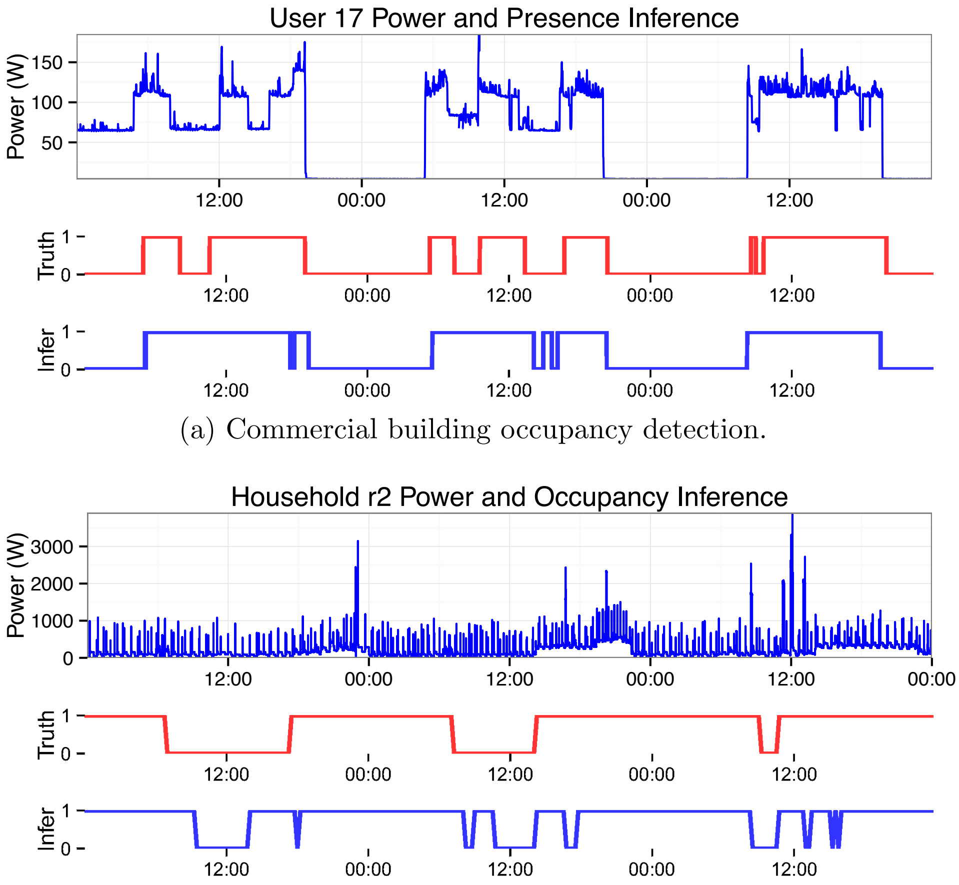


Figure 3.6: Misclassification rate during training iterations of MIT for u17 in PC. As the user can only observe the stopping conditions (red region), the user can terminate the training to avoid deterioration [100].

Table 3.1: Comparison of our results with prior art [30], [129], [130], showing the overall accuracy metric. The best performances in the weak supervision category are underlined.

	Supervised learning				Weak supervision		
	HMM[130]	SVM-PCA[129]	HMM-PCA[129]	NIOM[30]	MIT	γ-weighted	U. L.
r1	0.83	0.83	0.83		0.74	0.83	0.83
r2	0.82	0.92	0.90		0.74	0.75	0.77
r3	0.81	0.83	0.82		0.76	0.78	0.77
hA				0.79	0.81	0.78	0.84
hB				0.91	0.85	0.84	0.89

include hidden Markov model (HMM) [130], HMM-PCA, and SVM-PCA [129], which employed principal component analysis (PCA) for feature selections. For unsupervised learning, a threshold-based non-intrusive occupancy monitoring (NIOM) algorithm has been proposed, which assumed constant presence during nighttime, and then clustered the occupancy based on the deviation from nighttime power features [30]. One drawback with this approach is that the nighttime power usage used to set the thresholds may not be an accurate indicator for occupancy. We addressed this problem by two distinct ideas: MIT used a rough occupancy schedule to initialize the data, and then refined the labels by exploiting the power information; the surrogate loss methods recognized the noisy labels in initialized data, and customized loss functions to reduce the adverse effects. Table 3.1 indicates that MIT and surrogate loss methods outperformed NIOM for Home A (hA), but slightly underperformed for Home B (hB). Further, they remained consistently competitive for ECO dataset as compared to supervised learning methods [30], [129], [130].



(a) Commercial building occupancy detection.

(b) Residential building occupancy detection.

Figure 3.7: Examples of presence detection for (a) u17 and (b) r2 with MIT and γ-weighted loss, respectively. The power traces are shown on the top, whereas the bottom plots the true (red) and estimated (blue) occupancy, for comparisons.

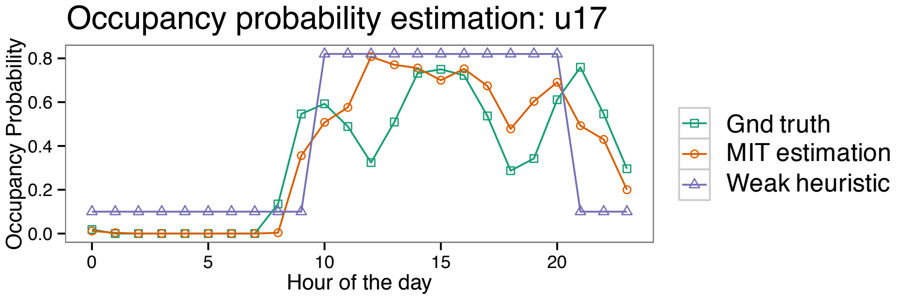


Figure 3.8: Occupancy schedules for NL include the shared profile (green), the learned one by MIT (blue), and the ground truth (red) for u17 in PC.

3.5 Chapter summary

Weak supervision is motivated by the fundamental need of learning a data-driven algorithm for h-CPS with limited data. It represents a new category of learning, where training labels are inherently noisy. While conventional supervised learning methods are misled by the training label inconsistencies, weak supervision methods resolve the conflicts algorithmically and have proved generalization properties. This chapter introduced two distinct methods of weak supervision: multi-view iterative training aims at refining the weak labels in each iteration while avoiding performance deterioration by checking for stopping conditions; surrogate loss method modifies the loss function to account for the training label noises that can be applied directly to existing data-driven algorithms.

To demonstrate the weak supervision paradigm, the proposed approaches were evaluated for occupancy detection with smart meter power data. By leveraging only high-level heuristics, accuracy rates of 74 to 89% for residential buildings and about 90% for offices can be obtained without accessing any labeled data. However, a main challenge for weak supervision is the reliable estimation of training noises. The key insight from multiview-iterative training (i.e., Lemma 3.1) is to estimate the noise *by examining the label consistencies between two iterations*. A Bayesian approach can be also adopted to utilize prior knowledge (e.g., [184]). While existing weak supervision methods are designed for classification problems, it will be meaningful to extend the framework to regression tasks. Since human behaviors are fundamentally patterned, a wide range of weak heuristics are available to enable innovative h-CPS applications.

Chapter 4

Gamification meets inverse game theory

The main goals of this chapter are to present the “inverse game theory,” which infers and rationalizes player utilities in a gamified context, and subsequently, to design and adapt incentives after learning the utilities to maximize system-level benefits. More specifically, we consider a non-cooperative game in which multiple players repeatedly make decisions with the aim of optimizing their individual utility functions without regard for the objectives of the other players [15]. We focus on the following question:

> Given the players’ equilibrium actions in a non-cooperative game, what are the utility functions that could motivate each player?

The problem of discerning people’s intentions given their actions has been under examination in multiple disciplines. In economics, the area of revealed preference studies the purchasing behavior of an agent over time to reveal more information about its utilities [222]. The identification literature in econometrics focuses on the rationalization problem in arbitrary games in which utilities are exponentially-sized [12]. In robotics, the problem of inverse reinforcement learning infers the hidden reward function of an agent in a long-horizon Markovian setting; however, these do not directly apply to a game that involves interactions among multiple agents (see Chap. 5 for more details). Last but not least, the problem of inverse optimization aims at recovering the objective function of an optimization program from its solution [6]. Differentiated from existing works, our study estimates utilities to rationalize players’ equilibrium behaviors based on a convex program that can be solved in polynomial time with limited data points. Moreover, we design a hierarchical control architecture in gamified services based on a Stackelberg game, which allows the game manager to induce desired behaviors by issuing incentive signals.

4.1 Overview of gamification

Gamification is defined as "the use of game design elements in non-game contexts" [44]. Gamification emerged as a promising method to support user engagement and enhance positive patterns in service uses and has been predicted by Gartner to have a significant impact on innovation, the design of employee performance, globalization of higher education, and emergence of customer engagement platforms [22]. Numerous contexts have reaped benefits as a result of positive, intrinsically motivating "gamified" experiences, including education and learning [42], [51], commerce [80], intra-organizational systems [63], engaging workplaces [194], sustainable consumption [79], and innovation/ideation [113], to name just a few.

Gamification typically consists of three main parts: implemented motivational affordances, resulting psychological outcomes, and further behavioral outcomes [80]. Specifically, the term "motivational affordances" means that motivation is afforded when the relations between game features and user abilities satisfy a user's needs when interacting with the service (examples of some commonly employed tactics are points, leaderboards and badges) [44]. While gamification can provide positive effects, the effects are highly dependent on the context in which the game is implemented and on its user community [80]. To effectively invoke desired psychological experiences and induce subsequent behavioral outcomes, the key is to understand and infer the users' psychological responses to design useful schemes and incentives, as illustrated in Fig. 4.1. In fact, one of the central and earliest questions in game theory is how to predict the behavior of an agent under certain incentives and game setup (for example, Nash equilibrium is used to determine the players' actions from their utilities). These predictions, in turn, may be used to inform and improve gamification design and to construct models of human behaviors.

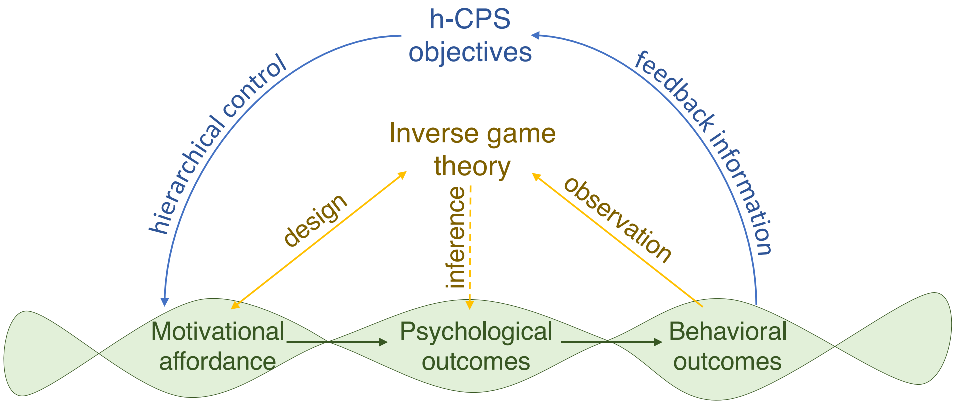


Figure 4.1: Gamification consists of three main parts: motivational affordance, psychological outcome, and behavioral outcome. The key idea of this chapter is to combine gamification with inverse game theory to learn about people's preferences in real context, and to enable customized incentive design to meet‘ overall h-CPS objectives.

4.2 Game-theoretic formulation

We model the interaction between the service provide (leader) and the users (followers) in a Stackelberg game [74]. In this model the followers are utility maximizers that play in a non-cooperative game, and the leader is also a utility maximizer with a utility that is dependent on the choices of the followers. The leader can influence the equilibrium of the game among the followers through the use of incentives which impact the utility and thereby the decisions of each follower.

A p-player game is described in terms of the strategy spaces and utility functions for each player. Consider a succinct game where the utility function of player i, denoted as f_i (which is usually exponentially-sized object), can be represented by a small number of parameters $\boldsymbol{\theta}_i \in \Theta$. We denote by $\mathcal{I} = \{1, \dots, p\}$ the index set for players. Let $\mathcal{X}_i \in \mathbb{R}^{m_i}$ denote the Euclidean strategy space of dimension m_i for player i and $\mathbf{x}_i \in \mathcal{X}_i$ denote its strategy vector. Define $m = \sum_i m_i$ and denote by $\mathcal{X} = \mathcal{X}_1 \times \cdots \mathcal{X}_p$ the joint strategy space and $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$ the joint strategy. Each player's strategy vector $\mathbf{x}_i$ is constrained to a convex set $\mathcal{C}_i \subset \mathcal{X}_i$. Let ℓ_i be the number of constraints on player i's problem and let $\ell = \sum_{i=1}^{p} \ell_i$. Denote $\mathcal{C} = \mathcal{C}_1 \times \cdots \times \mathcal{C}_p$ as the constraint set which we can explicitly characterize in terms of mappings $\mathbf{h}_i : \mathcal{X}_i \to \mathcal{C}_i$ where each component $h_{i,j}(\mathbf{x})$, $j = 1, \dots, \ell_i$ is a concave function of $\mathbf{x}_i$: $\mathcal{C}_i = \{x_i | \mathbf{h}_i(\mathbf{x}_i) \geq \mathbf{0}\}$. It is assumed that $\mathcal{C}_i$ is non-empty and bounded. Furthermore, the formulation can be extended for coalition games, where players form groups to jointly optimize their utilities when incentivized to do so (see [133] for further details).

In each round of the game, the player solves their individual optimization problem

$$\max_{\mathbf{x}_i \in \mathcal{C}_i} f_i(\mathbf{x}_i, \mathbf{x}_{-i}; \boldsymbol{\gamma}), \tag{4.1}$$

where $\mathbf{x}_{-i} = (\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_p)$ is the marginal strategy vector for all players excluding player i, and $\boldsymbol{\gamma} \in \Gamma$ is the game incentive signal (omitted in expressions if it does not cause confusions). To wit, the agents are non-cooperative players in a continuous game with convex constraints. We model their interaction using the Nash equilibrium concept:

Definition 4.1 (Nash Equilibrium)**.** *A point* $\mathbf{x}_i \in \mathcal{C}_i$ *is a* Nash equilibrium *for the* p*–player non-cooperative game* $(f_1, \dots, f_p)$ *on* $\mathcal{C}$ *if for each* $i \in \mathcal{I}$*,*

$$f_i(\mathbf{x}_i, \mathbf{x}_{-i}) \geq f_i(\mathbf{x}'_i, \mathbf{x}_{-i}), \quad \forall\, \mathbf{x}'_i \in \mathcal{C}_i. \tag{4.2}$$

It is well known that Nash equilibria exist for concave games [197, Thm. 1]. The definition can be relaxed as follows:

Definition 4.2 (ϵ–approximate Nash equilibrium)**.** *Given* $\epsilon > 0$*, a point* $\mathbf{x}_i \in \mathcal{C}_i$ *is a* ϵ*–approximate Nash equilibrium for the game* $(f_1, \dots, f_p)$ *if for each* $i \in \mathcal{I}$*,*

$$f_i(\mathbf{x}_i, \mathbf{x}_{-i}) \geq f_i(\mathbf{x}'_i, \mathbf{x}_{-i}) - \epsilon, \quad \forall\, \mathbf{x}'_i \in \mathcal{C}_i. \tag{4.3}$$

Define the Lagrangian of each player's optimization problem as follows:

$$L_i(\mathbf{x}_i, \mathbf{x}_{-i}, \boldsymbol{\mu}_i) = f_i(\mathbf{x}_i, \mathbf{x}_{-i}) + \sum_{j \in \mathcal{A}_i(\mathbf{x}_i)} \mu_{i,j} h_{i,j}(\mathbf{x}_i), \tag{4.4}$$

where $\mathcal{A}_i(\mathbf{x}_i)$ is the active constraint set at $\mathbf{x}_i$, and define

$$\boldsymbol{\omega}(\mathbf{x}, \boldsymbol{\mu}) = \left[\tfrac{\partial}{\partial \mathbf{x}_1} L_1(\mathbf{x}_1, \mathbf{x}_{-1}, \boldsymbol{\mu}_1)^\top \quad \cdots \quad \tfrac{\partial}{\partial \mathbf{x}_p} L_p(\mathbf{x}_p, \mathbf{x}_{-p}, \boldsymbol{\mu}_p)^\top\right]^\top \in \mathbb{R}^m. \tag{4.5}$$

We further define a differential Nash equilibrium as follows:

Definition 4.3 ([188])**.** *A point $\mathbf{x}^* \in \mathcal{C}$ is a differential Nash equilibrium for the game $(f_1, \ldots, f_p)$ if the following conditions are satisfied:*

- $\boldsymbol{\omega}(\mathbf{x}^*, \boldsymbol{\mu}^*) = 0,$
- $\mathbf{z}^\top \frac{\partial^2}{\partial \mathbf{x}_i^2} f_i(\mathbf{x}_i^*, \mathbf{x}_{-i}^*, \boldsymbol{\mu}_i^*)\mathbf{z} < 0$ *for all* $\mathbf{z} \neq \mathbf{0}$ *such that* $\frac{\partial}{\partial \mathbf{x}_i} h_{i,j}(\mathbf{x}_i^*)^\top \mathbf{z} = 0,$
- $\mu_{i,j} > 0$ *for* $j \in \mathcal{A}_i(\mathbf{x}_i^*)$ *and* $i \in \mathcal{I}$.

Proposition 1. *A differential Nash equilibrium of the p-person concave game $(f_1, \ldots, f_p)$ on $\mathcal{C}$ is a Nash equilibrium.*

Proof. See Appendix A.3. □

A sufficient condition guaranteeing that a Nash equilibrium $\mathbf{x}$ is isolated is that the Jacobian of $\boldsymbol{\omega}(\mathbf{x}, \boldsymbol{\mu})$ is invertible [197]. We refer to such points as being *non-degenerate*.

4.3 Reverse Stackelberg game – incentive design

A reverse Stackelberg game is a hierarchical control problem in which sequential decision making occurs; in particular, the leader announces an incentive signal to the followers, after which the followers determine their optimal strategies [74]. Both the leader and the followers wish to maximize their pay-off functions $f_L(\mathbf{x}, \boldsymbol{\gamma})$ and $\{f_1(\mathbf{x}, \boldsymbol{\gamma}), \ldots, f_n(\mathbf{x}, \boldsymbol{\gamma})\}$ respectively. We now consider each of the follower's utility functions to be a function of the incentives $\boldsymbol{\gamma} \in \Gamma$ chosen by the leader.

The basic approach to solving the reversed Stackelberg game is as follows. Let $\boldsymbol{\gamma}$ and $\mathbf{x}$ take values in Γ and $\mathcal{C}$, respectively and let $f_L, f_i : \mathcal{C} \times \Gamma \to \mathbb{R}$ for each $i \in \mathcal{I}$. We define the desired choice for the leader as

$$(\mathbf{x}^*, \boldsymbol{\gamma}^*) \in \arg\max_{\mathbf{x}, \boldsymbol{\gamma}} \{f_L(\mathbf{x}, \boldsymbol{\gamma}) | \boldsymbol{\gamma} \in \Gamma, \mathbf{x} \in \mathcal{C} \text{ is a differential Nash under } \boldsymbol{\gamma}\}. \tag{4.6}$$

By ensuring that the desired agent action $\mathbf{x}^*$ is a non-degenerate differential Nash equilibrium (i.e., structural stability), we can make the solution robust to measurement and environmental noise [189]. Further, it establishes that it is (locally) isolated – it is globally isolated if the followers' game is concave. This indicates that the solution of the reversed Stackelberg game is obtained by a bi-level optimization problem. To solve the inner level of the bi-level optimization problem, we replace the condition that the occupants play a Nash equilibrium with the dynamical system determined by the gradients of each player's utility with respect to their own choice variables, i.e.,

$$\dot{\mathbf{x}}_i = \frac{\partial}{\partial \mathbf{x}_i} f_i(\mathbf{x}_i, \mathbf{x}_{-i}, \boldsymbol{\gamma}),\ \mathbf{x}_i \in \mathcal{C}_i,\ \forall\ i \in \mathcal{I}. \tag{4.7}$$

It has been show that the solution obtained by a projected gradient descent method for computing stationary points of the dynamical system in (4.7) converges to Nash equilibria [67]. We can also add the constraint to the leader's optimization problem that at the stationary points of this dynamical system (i.e. the Nash equilibria), the matrix $-\frac{\partial^2}{\partial \mathbf{x}^2}\boldsymbol{\omega}$ is positive definite, thereby ensuring that each of the equilibria is non-degenerate and isolated.

Denote the set of non-degenerate stationary points of the dynamical system $\dot{\mathbf{x}}$ in (4.7) as $\mathrm{Stat}(\dot{\mathbf{x}})$. The leader then solves the following problem:

$$\begin{aligned} \max_{\boldsymbol{\gamma}\in\Gamma} \quad & f_L(\mathbf{x}, \boldsymbol{\gamma}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathrm{Stat}(\dot{\mathbf{x}}). \end{aligned} \tag{4.8}$$

The bi-level optimization can be solved by methods like trust-region method [153] and evolutionary approaches [36], [203].

4.4 Inverse game theory framework

Consider a succinct game where the i–th player's utility function is parameterized as follows:

$$f_i(\mathbf{x}_i, \mathbf{x}_{-i}) = \varphi_{i,0}(\mathbf{x}_i, \mathbf{x}_{-i}) + \textstyle\sum_{j=1}^{N_i} \varphi_{i,j}(\mathbf{x}_i, \mathbf{x}_{-i})\theta_{ij}, \tag{4.9}$$

where $\{\varphi_{i,j}\}_{j=0}^{N_i}$ is a set of non-constant, concave basis functions and $\boldsymbol{\theta}_i = [\theta_{i1}\ \cdots \theta_{iN_i}]^\top \in \Theta_i$ are the parameters, which are assumed unknown thus to be learned. Let n_i denote the number of data points for player i and define $n_d = \sum_{i=1}^{p} n_i$ be the total number of data points. We assume that each observation $\mathbf{x}^{(k)}$ corresponds to an ϵ–approximate Nash equilibrium where the superscript notation $(\cdot)^{(k)}$ indicates the k–th observation. We define residual functions capturing the amount of suboptimality of the observations $\mathbf{x}_i^{(k)}$ [122]. Indeed, let the residual of the stationarity and complementary conditions for player i's optimization problem be given by, respectively,

$$\mathbf{r}_{\mathrm{s},i}^{(k)}(\boldsymbol{\theta}_i, \boldsymbol{\mu}_i) = \frac{\partial}{\partial \mathbf{x}_i} f_i(\mathbf{x}^{(k)}) + \sum_{j=1}^{\ell_i} \mu_{i,j} \frac{\partial}{\partial \mathbf{x}_i} h_{i,j}(\mathbf{x}_i^{(k)}) \tag{4.10}$$

and

$$r_{\mathrm{c},i}^{j,(k)}(\boldsymbol{\mu}_i) = \mu_{i,j} h_{i,j}(\mathbf{x}_i^{(k)}),\ j \in \{1, \ldots, \ell_i\}, \tag{4.11}$$

where $\boldsymbol{\theta}_i \in \Theta$ is the utility function parameter and $\boldsymbol{\mu}_i = (\mu_{i,j})_{j=1}^{\ell_i}$ are Lagrange multipliers. Define

$$\mathbf{r}_{\mathrm{s}}^{(k)}(\boldsymbol{\theta}, \boldsymbol{\mu}) = [\mathbf{r}_{\mathrm{s},1}^{(k)}(\boldsymbol{\theta}_1, \boldsymbol{\mu}_1)^\top \ \cdots \ \mathbf{r}_{\mathrm{s},p}^{(k)}(\boldsymbol{\theta}_p, \boldsymbol{\mu}_p)^\top]^\top \in \mathbb{R}^m, \tag{4.12}$$

and

$$\mathbf{r}_{\mathrm{c}}^{(k)}(\boldsymbol{\mu}) = [\mathbf{r}_{\mathrm{c},1}^{(k)}(\boldsymbol{\mu}_1)^\top \ \cdots \ \mathbf{r}_{\mathrm{c},p}^{(k)}(\boldsymbol{\mu}_p)^\top]^\top \in \mathbb{R}^\ell, \tag{4.13}$$

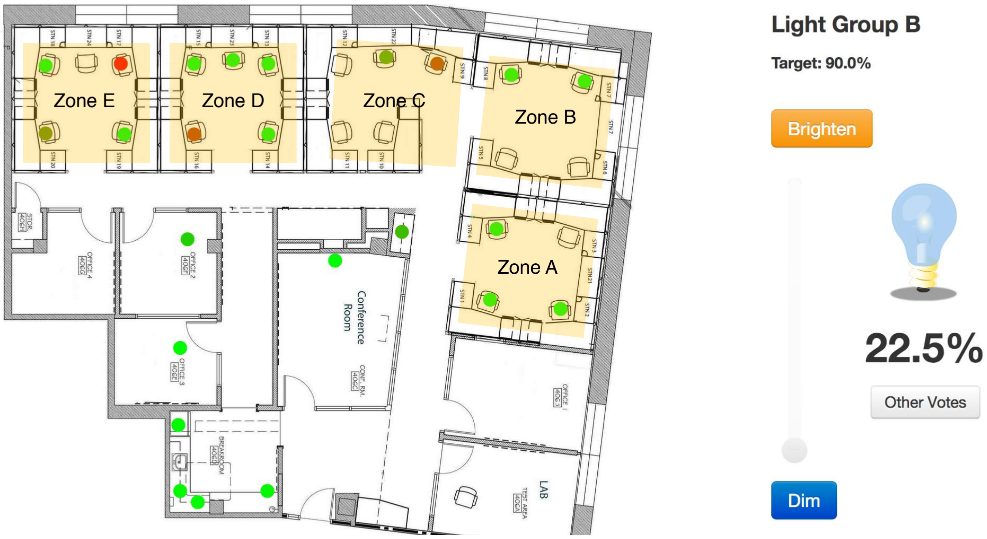
where $\boldsymbol{\theta} = (\boldsymbol{\theta}_i)_{i=1}^p$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_i)_{i=1}^p$ and $\mathbf{r}_{\mathrm{c},i}^{(k)}(\boldsymbol{\mu}_i) = [r_{\mathrm{c},i}^{1,(k)}(\boldsymbol{\mu}_i) \ \ldots \ r_{\mathrm{c},i}^{\ell_i,(k)}(\boldsymbol{\mu}_i)]^\top$. Given the observations of the agents' decisions, we solve the following convex optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\mu},\boldsymbol{\theta}} \quad & \textstyle\sum_{k=1}^{n_d} \chi(\mathbf{r}_s^{(k)}(\boldsymbol{\theta}, \boldsymbol{\mu}), \mathbf{r}_c^{(k)}(\boldsymbol{\mu})) \\ \text{s.t.} \quad & \boldsymbol{\theta}_i \in \Theta_i, \boldsymbol{\mu}_i \geq \mathbf{0}, \qquad \forall\, i \in \mathcal{I} \end{aligned} \tag{IGT}$$

where $\chi : \mathbb{R}^m \times \mathbb{R}^\ell \to \mathbb{R}_+$ is a nonnegative, convex penalty function satisfying $\chi(\mathbf{z}_1, \mathbf{z}_2) = 0$ if and only if $\mathbf{z}_1 = \mathbf{0}$ and $\mathbf{z}_2 = \mathbf{0}$ (i.e. any norm on $\mathbb{R}^m \times \mathbb{R}^\ell$), the inequality $\boldsymbol{\mu}_i \geq \mathbf{0}$ is element-wise and the Θ_i's are constraint sets for the parameters $\boldsymbol{\theta}_i$ that collect prior information about the utility functions f_i. For learning utilities in a game theoretic context, we would like to ensure that the observations are ϵ–approximate Nash equilibria for the estimated game; hence, we select Θ_i such that each player's parameterized utility function is concave. As indicated in [122], it is important to select each Θ_i such that it encodes enough prior information about each f_i so as to prevent trivial solutions; we ensure this by selecting the set of basis functions $\{\varphi_{i,j}\}_{j=1}^N$ for each player to be non-constant, concave functions and assuming $\varphi_{i,0} \not\equiv 0$ in our parameterization. As an example, if $\{\varphi_{i,j}\}_{j=0}^{N_i}$ are all concave, then $\Theta_i = \mathbb{R}_+^{N_i}$ ensures that f_i is concave. The optimization (IGT) can be converted to a standard regression problem, as detailed in [190]. Furthermore, robustness can be enhanced by assuming heteroskedasticity [69, Chap. 5] which also allows for inference of correlated errors in the resulting regression model. These correlated errors can then be used to determine the relationship between players' decision-making processes, as detailed in [132], [133].

4.5 Energy efficiency via gamification

We have designed *a social game for energy savings* where occupants in an office building vote according to their usage preferences of shared resources and are rewarded with points based on how *energy efficient* their strategy is in comparison with other occupants. Earning more points increases the likelihood of the occupant winning the weekly lottery. The prizes in the lottery consist of Amazon gift cards. We have installed a Lutron system in the office, which allows us to precisely control the lighting level within the lighting zones. We use it to set the default lighting level as well as to implement the average of the votes each time the occupants change their lighting preferences. There are twenty-two occupants in the office, which is divided into five lighting zones, each with four occupants (see Fig. 4.2a). We have developed an online platform in which the occupants can log in and participate in the game.



(a) Gamified lighting zones.

(b) Voting UI.

Figure 4.2: (a) The office at UC Berkeley campus where social game was carried out. The space has five lighting zones, each can be controlled separately. (b) User interface for occupants to view and vote for the lighting level.

In the platform, the occupants can cast their lighting dim level votes (e.g., Fig. 4.2b shows the user interface for an occupant to select their lighting preference), view point balances of all occupants, and observe the voting patterns of all occupants.

An occupant's vote x_i can change the lighting level in their zone as well as for neighboring zones. The lighting setting that is implemented in each zone is the average of all the votes weighted according to location proximity to that zone. In addition, there is a default lighting setting $d \in [0, 100]$ selected by the leader. An occupant can leave the lighting setting as the default after logging in or they can change it to some other value in the interval $[0, 100]$ depending on their preferences. There are three different states for an occupant in the lighting game. Each day when an occupant logs into the online platform after they enter the office, they are considered present for the remainder of the day. If they actively change their votes from the default to some other values, we consider them *active*. On the other hand, if they choose not to change their vote from the default setting, then they are considered *default* for the day. If they do not enter the office on a given day, then they are considered *absent*.[1] To reduce the complexity of computing the expectation for the joint distribution across player states for $p = 22$ players, we currently restrict the set of admissible incentive mechanisms to be a constant map $\boldsymbol{\gamma} = (d, \rho)$, where d is the default light level and $\rho \geq 0$ is

[1] We implemented a presence detection algorithm based on plugload power data [100].

the incentive point, such that the i-th player's utility is

$$f_i(x_i, x_{-i}, \rho) = -\underbrace{\left(x_i - \frac{1}{p}\sum_{j=1}^{p} x_j\right)^2}_{\text{Taguchi loss}} \underbrace{-\theta_i \rho \left(\frac{x_i}{100}\right)^2}_{\text{Desire to win}}, \tag{4.14}$$

where the Taguchi loss function is interpreted as modeling occupant dissatisfaction in such a way that it increases as variation increases from their desired lighting setting [212], and ρ is the total number of points distributed by the building manager that affects each occupant's desire to win. In addition, the nature of default setting $d \in [0, 100]$ is that it is an option provided to the followers; they must actively vote in order for this value not to be taken as their current vote when they are present in the office. In a sense, it is the outside option. Thus, the leader only selects the incentives (d, ρ), as shown in Fig. 4.3. For the social game, the leader's utility function is given as follows:

$$f_L(\mathbf{x}, \rho) = \mathbb{E}\left[\underbrace{K - g(\mathbf{x})}_{energy} - \underbrace{c_2 p(\rho)}_{effort} - \underbrace{c_1 \sum_{i=1}^{n} \beta_i f_i(x_i, x_{-i}, y)}_{benevolence}\right] \tag{4.15}$$

where K is is the maximum consumption of the Lutron lighting system in kilowatt-hours (kWh), $g(\mathbf{x})$ is the is the energy consumption in kWh at a given $\mathbf{x}$ (see Fig. 4.4a), $p(\cdot)$ is a cost-for-effort function on the points ρ and $c_1, c_2 \in \mathbb{R}_+$ are scaling factors for the last two terms describing how much utility and total points respectively the leader is willing to exchange for 1 kWh. The last term is the *benevolence* term where the β_i's are the *benevolence factors*. This term captures the fact that the leader cares about the followers' satisfaction which is related to their productivity level [7]. The expectation is taken with respect to the joint distribution defined by distributions across the player states *absent*, *active*, *default*. By drawing from this joint distribution, we simulate the game using the estimated utility functions. In Fig. 4.4b, we can see that our model captures most of the variation in the true votes. Since the prize in the lottery is currently a fixed monetary value delivered to the winner through an Amazon gift card, varying the points does not cost the leader anything explicitly. However, we model the cost of giving points by a function $p(\cdot)$ which captures the fact that after some critical value of ρ the points no longer seem as valuable to the followers. The followers' perceive the points that they receive as having some value towards winning the prize. The leader's goal is to choose ρ and d so they induce the followers to play the game and choose the desired lighting setting.

Currently we do not add individual rationality constraints to the leader's optimization problem which would ensure that the players' utilities are at least as much as what they would get by selecting the default value. The impact being that this constraint would ensure players are active. With respect to economics literature, the default lighting setting is similar to the outside option in contract theory. It is interesting to note that in the current situation the leader has control over the outside option.

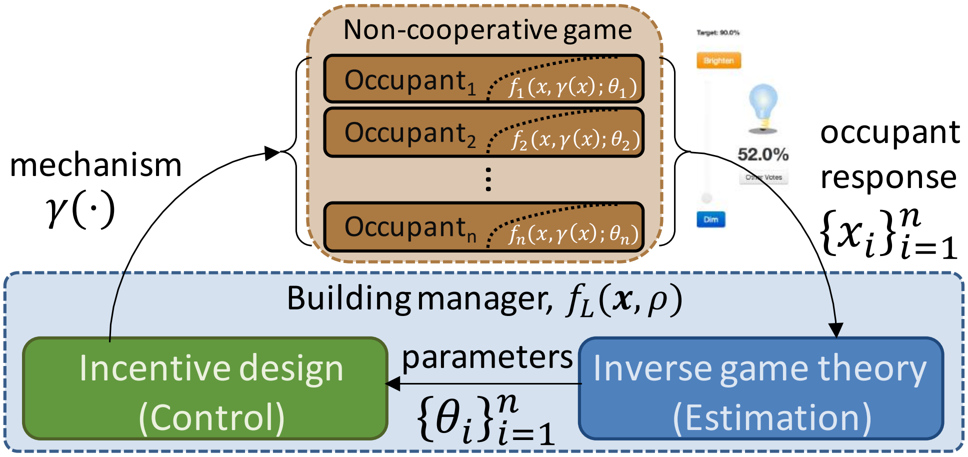
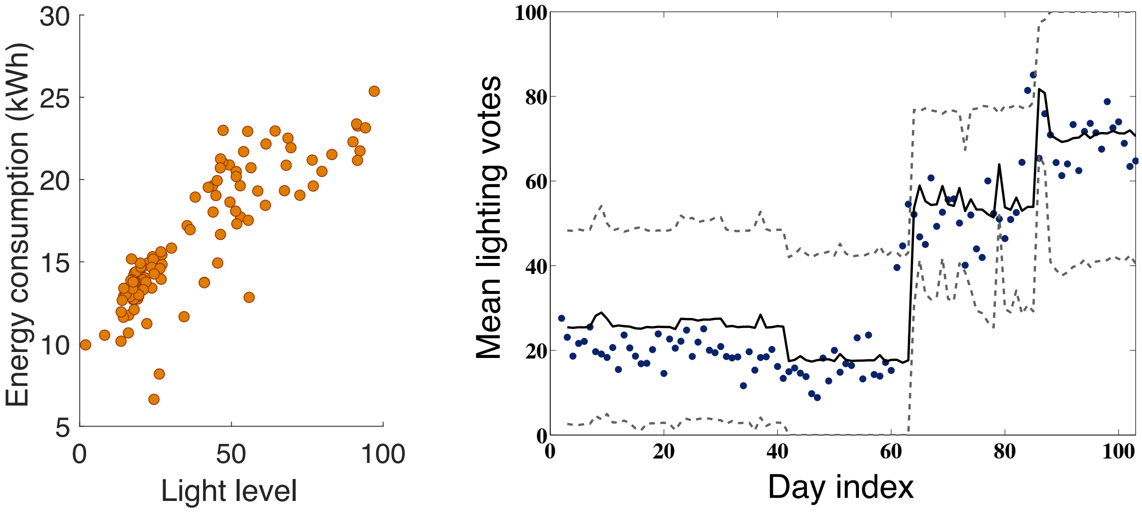


Figure 4.3: Inverse game theory and incentive design in a social game.

The particle swarm optimization (PSO) method was employed for the leader incentive design, which is a population based stochastic optimization technique in which the algorithm is initialized with a *population* of random solutions and searches for optima by updating *generations* [121]. The potential solutions are called *particles*. Each particle stores its coordinates in the problems space which are associated with the best solution achieved up to the current time. The best over all particles is also stored and at each iteration the algorithm updates the particles' velocities. For each particle in the PSO algorithm, we sample from the distribution across player states and compute Nash for the resulting game via simulation of the dynamical system (4.7). We compute the mean of the votes at the Nash equilibrium to get the lighting setting. We repeat this process and use the mean of the lighting settings over all the simulations to compute the leader's utility for each of the particles. To estimate the energy consumption function, we collected data for different lighting settings and created a piecewise affine map from the lighting dim level to energy consumption in kilowatt–hours (kWh) (see Fig. 4.4a). Using this map, we formulate a utility for the leader which takes the average lighting votes as the input and returns the difference between the maximum consumption in kWh, i.e. 25 kWh, and the piecewise affine map for energy saving of the lights.

Using the past data, i.e. data collected for default settings $\{10, 20, 60, 90\}$, and θ_i estimates for each occupant, we created a piecewise affine map for interpolating the parameters of the occupants utility functions for different default settings. Similarly, we interpolated the joint distribution across player states (*absent*, *active*, *default*) as a function of the default setting. This enabled the optimization of the leader's utility function, given in (4.15), over both the total points ρ and the default setting d.

In the implementation of the leader's optimization problem in this example we make the following choices for the parameters and scaling of the leader's utility function. For each particle in the PSO algorithm, we map each follower's true utility f_i to an interpolated utility



(a) Energy vs. light level.

(b) Nash equilibria prediction.

Figure 4.4: (a) Energy consumption data for the Lutron lighting system in kWh as a function of the lighting setting. (b) Prediction of lighting votes, showing the true mean of the lighting votes for each day over the duration of the experiment (blue dots), the predicted Nash equilibria with the estimated utilities (solid black line), and one standard deviation of the prediction (dashed black lines).

$\hat{f}_i$ taking a value in the range $[0, 100]$ by finding the global maximum and minimum of their utility under the current particle to determine an appropriate affine scaling of their original utility. We use $\hat{f}_i$ in place of f_i in the leader's utilty. We use $c_1 = 1/2$ to reflect the fact that the leader is willing to exchange 1 kWh savings for a utility value of 2 in the total sum of the followers' utilities $\sum_i \beta_i \hat{f}_i$ under the current particle value for $y = (d, \rho)$. Similarly, we use $c_2 = 1/500$ to indicate that the leader is willing to exchange 500 points in return for 1 kWh of savings.

Examining each of the occupant's estimated utility functions has given us a sense of which occupants are the most sensitive to changes in ρ and d. For instance, occupant id2 was quite inflexible to changes in the points ρ and appeared to care less about winning and more about his comfort level (see Fig. 4.5). This fact is also reflected in the very low parameter estimate for θ_2. It is also the case that occupant id2's behavior was largely affected by others' votes. In addition, occupants in the set $\mathcal{S}_c = \{2, 6, 8, 14, 20\}$ were the most active players in a probabilistic sense. As a result, we gave non-zero benevolence terms to players in this set. We refer to this set as the leader's *care-set*. For all $i \in \{1, \dots, 20\} \backslash \mathcal{S}_c$, we set $\beta_i = 0$. Further, we normalized $\sum_{j \in \mathcal{S}_c} \beta_j = 1$. Since occupant id2 exhibited particularly interesting behavior, we varied β_2, and let $\beta_j = (1 - \beta_2) \frac{1}{|\mathcal{S}_c|}$ for all $j \in \mathcal{S}_c$ and where $|\mathcal{S}_c|$ is the cardinality of $\mathcal{S}_c$. Table 4.1 contains the energy savings in dollars per day for the leader given the energy cost of the lights and how much of the occupants' utility that the leader is willing to exchange for

1 kWh.[2] The optimal incentives were computed by solving the leader's optimization problem via the PSO method where we simulated the game of the occupants via the dynamic system (4.7). Table 4.1 lists the leader's utility in dollars for previous values of (d, ρ) after the start of the social game, as well as the values after optimizing over (d, ρ) for some given benevolence factors $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$. We can see that by computing even the local optimal (d, ρ) of the leader's bi-level optimization problem, the leader has a much higher utility.

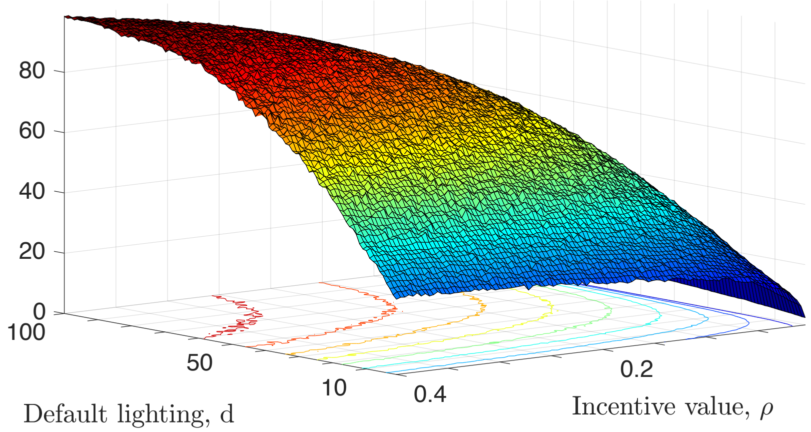


Figure 4.5: Utility of occupant id2 as a function of (d, ρ) at the mean Nash equilibrium after running 1000 simulations. Notice that for fixed values of d the utility value is near constant in ρ. Also, occupant 2 has very large utility when the default setting is around 70.

Table 4.1: Leader's utility in dollars for previously implemented $(d, \rho)$and benevolence factors $\boldsymbol{\beta} = (\beta_2, \sum_{j \in \mathcal{S}_c} \beta_j)$ where $\mathcal{S}_c = \{6, 8, 14, 20\}$. We also show the results for optimized leader incentives (d, ρ) by solving the leader's optimization problem (4.8) using PSO. The value is interpreted as the energy saved in dollars by the leader plus the utility as measured in dollars. We use a rate of \$0.12 per kWh as this is the approximate rate in California.

	Benevolence factors $\boldsymbol{\beta}$					
(d, ρ)	(0.9,0.1)	(0.75,0.25)	(0.6,0.4)	(0.45,0.55)	(0.3,0.7)	(0.2,0.8)
$(10, 7000)$	\$2.01	\$2.10	\$2.19	\$2.28	\$2.37	\$2.42
$(20, 7000)$	\$1.98	\$2.01	\$2.06	\$2.08	\$2.10	\$2.13
$(60, 7000)$	\$1.70	\$1.67	\$1.66	\$1.65	\$1.65	\$1.64
$(90, 7000)$	\$1.35	\$1.33	\$1.32	\$1.31	\$1.31	\$1.30
Optimized	**\$4.56**	**\$4.73**	**\$4.67**	**\$4.69**	**\$5.07**	**\$5.43**

We have not yet factored in the cost of the prize in the lottery, which was \$100 per week. The values we report in Table 4.1 represent per day savings on weekdays. More substantial savings can be achieved by scaling up the social game. For example, we are in the process

[2] The electricity cost is \$0.12 per kWh based on the California rate.

of implementing a social game in an entire building in Singapore with more than $1,000$ occupants. This social game will include options for the consumer to choose lighting settings, HVAC and personal cubicle plug-load consumption. In addition, we plan to implement a social game of this nature in Sutarja Dai Hall on the UC Berkeley campus. At this scale, with a lottery cost of $100 the building manager can potentially save a considerable amount.

4.6 Chapter summary

This chapter introduced a theoretical framework of inverse game theory and incentive design in a gamified environment. The key insight is to design gamification (i.e., motivational affordance, psychological outcome and behavioral change) as an enabler for simultaneous learning and influence of people's preferences and behaviors, which are revealed and reinforced in their interactions with others in the game. In a gamified context, utility functions can be used to capture social dimensions of people-people interaction and individual preferences for comfort and incentives; even "indifference" and "inaction" common to real-world game (due to lack of engagement or motivation) can be naturally encoded. By rationalizing people's behaviors using ϵ-approximate Nash equilibrium, the proposed inverse game theory can learn the parameters of people's utility functions with high data-efficiency. The estimated utilities are used in a reverse Stackelberg game to design the optimal incentives, which can be issued by the leader as a signal to "nudge" people in the *desired* directions, thereby "closing the loop".

We applied the theory to improve energy efficiency in buildings via gamification. The results are promising: just by introducing the social game without optimizing the incentive, we obtained about $2.00/day savings for a small office (∼1/20-th of the building). By optimizing the incentives, we achieved a 150% increase in the savings (∼$5/day). The scaled-up game has a potential for even more substantial savings. Furthermore, the game offered valuable data about people's preferences, which can be respected in automated building control after the game period. However, due to complexity and uncertainty of people's behaviors, it is meaningful to examine action uncertainty (e.g., the mixture model of utility function [134]), as well as different forms of game (e.g., coalition game [133], polymatrix game [72]). While this chapter considers actions within a short time span (i.e., Nash equilibrium), the next chapter discusses preference learning that spans multiple time scales to account for people's long-term planning capability.

Chapter 5

Deep Bayesian inverse reinforcement learning

Inverse reinforcement learning (IRL) aims at inferring the latent reward function that the agent subsumes by observing its demonstrations or trajectories in the task. It has been successfully applied to tackle practical challenges, e.g., navigation [1], [191], [247], and robotics [2], [131], [175]. As people's actions often involve long-term planning and their motivations depend on factors that cannot be known *a priori*, human behavior learning needs to span multiple time scales and account for complex reward. However, existing IRL methods have limited representation power due to the linearity assumption [2], [191], [211], [247].

The success of deep learning in a wide range of domains has drawn the community's attention to its structural advantages that can improve learning in complicated scenarios, e.g., [163] recently achieved a deep reinforcement learning (RL) breakthrough. Nevertheless, most deep models require massive data to be properly trained and can become impractical for human preference learning. A deep Gaussian process (deep GP) is a deep belief network comprising a hierarchy of latent variables with Gaussian process mappings between the layers. Analogous to how gradients are propagated through a standard neural network, deep GPs aim at propagating uncertainty through Bayesian learning of latent posteriors. This constitutes a useful property for approaches involving stochastic decision making and also guards against overfitting by allowing for noisy features. More importantly, it can not only learn abstract structures with *smaller* data sets, but also retain the non-parametric properties which has been demonstrated to be important for IRL. However, previous methodologies employed for approximate Bayesian learning of deep GPs fail when diverging from the simple case of fixed output data modeled through a Gaussian regression model [37], [156]. In particular, in the IRL setting, the reward (output) is only revealed through the demonstrations, which is guided by the policy given by the reinforcement learning. In light of this contemplation, we propose a deep Bayesian inverse reinforcement learning in this chapter. We introduce a non-standard variational approximation framework to extend previous inference schemes for deep GPs, which allows for approximate Bayesian inference to learn the complex reward functions.

5.1 Introduction of inverse reinforcement learning

The Markov Decision Process (MDP) is characterized by $\{\mathcal{S}, \mathcal{A}, \mathcal{T}, \gamma, \boldsymbol{r}\}$, which represents the state space, action space, transition model, discount factor, and reward function, respectively. Take robot navigation as an example. The goal is to travel to the goal spot while avoiding stairwells. The state describes the current location and heading. The robot can choose actions from going forward or backward, turning left or right. The transition model specifies $p(s_{t+1}|s_t, a_t)$, i.e., the probability of reaching the next state given the current state and action, which accounts for the kinematic dynamics. The reward is +1 if it achieves the goal, -1 if it ends up in the stairwell, and 0 otherwise. The discount factor, γ, is a positive number less than or equal to 1, e.g., 0.9, to discount the future rewards. The optimal policy is then given by maximizing the expected reward,

$$\pi^* = \arg\min_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t)|\pi\right], \tag{5.1}$$

where the expectation is taken over the stochastic state transitions and policy actions.

The IRL task is to find the reward function r^* such that the induced optimal policy is in alignment with the demonstrations, given $\{\mathcal{S}, \mathcal{A}, \mathcal{T}, \gamma\}$ and $\mathcal{M} = \{\zeta_1, ..., \zeta_H\}$, where $\zeta_h = \{(s_{h,1}, a_{h,1}), ..., (s_{h,T}, a_{h,T})\}$ is the demonstration trajectory consisting of state-action pairs. Under the linearity assumption, the feature representation of states forms the linear basis of reward, namely $r(s) = \mathbf{w}^\top \boldsymbol{\phi}(s)$, where $\boldsymbol{\phi}(s) : \mathcal{S} \mapsto \mathbb{R}^{m_0}$ is the m_0-dimensional mapping from the state to the feature vector. From this definition, the *expected reward* for policy π is given by

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t)|\pi\right] = \mathbf{w}^\top \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \boldsymbol{\phi}(s_t)|\pi\right] = \mathbf{w}^\top \boldsymbol{\mu}(\pi),$$

where $\boldsymbol{\mu}(\pi) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \boldsymbol{\phi}(s_t)|\pi\right]$ is the feature expectation for policy π. The reward parameter $\mathbf{w}^*$ is learned such that

$$\mathbf{w}^{*\top} \boldsymbol{\mu}(\pi^*) \geq \mathbf{w}^{*\top} \boldsymbol{\mu}(\pi), \forall \pi \tag{5.2}$$

a prevalent idea that appears in the maximum margin planning (MMP) [191] and feature expectation matching [211].

Motivated by the perspective of expected reward that parametrizes the policy class, the maximum entropy (MaxEnt) model considers a stochastic decision framework, where *the optimal policy randomly chooses the action* according to the associated reward [247]:

$$p(a|s) = \exp\{Q^*(s, a; \boldsymbol{r}) - V^*(s; \boldsymbol{r})\} \tag{5.3}$$

where $V(s; \boldsymbol{r}) = \log \sum_a \exp(Q(s, a; \boldsymbol{r}))$ follows the Bellman equation, $Q(s, a; \boldsymbol{r})$ and $V(s; \boldsymbol{r})$ are measures of how "desirable" the corresponding state s and state-action pair (s, a) are under rewards $\boldsymbol{r}$. In principle, for a given state s, the best action corresponds to the highest Q-value $Q(s, a; \boldsymbol{r})$, which represents the "optimality" of the action considering the accumulated rewards in the future. Assuming independence among state-action pairs from demonstrations, the likelihood of the demonstration is equal to the joint probability of taking a sequence of actions $a_{h,t}$ under states $s_{h,t}$, as indicated by the Bellman equation:

$$p(\mathcal{M}|\boldsymbol{r}) = \prod_{h=1}^{H}\prod_{t=1}^{T} p(a_{h,t}|s_{h,t}) = \exp\left(\sum_{h=1}^{H}\sum_{t=1}^{T}\left(Q(s_{h,t}, a_{h,t}; \boldsymbol{r}) - V(s_{h,t}; \boldsymbol{r})\right)\right). \tag{5.4}$$

Though directly optimizing the above criteria with respect to $\boldsymbol{r}$ is possible, it does not lead to generalized solutions transferrable in a new test case where no demonstrations are available; hence, we need a "model" of $\boldsymbol{r}$. MaxEnt assumes linear structure for rewards, while GPIRL uses GPs to relate the states to rewards [143].

5.2 Deep GP for inverse reinforcement learning

We first discuss the reward modeling through GP, proceed to incorporate the representation learning modules, and introduce a variational framework to train the model for IRL.

Gaussian Process reward modeling

We consider the setup of *discretizing the world* into n states. Let the observed state-action pairs (demonstrations) $\mathcal{M} = \{\zeta_1, \ldots, \zeta_h\}$ be generated by a set of m_0-dimensional state features $\mathbf{X} \in \mathbb{R}^{n,m_0}$ through the reward function r. Throughout this chapter we denote points (rows of $\mathbf{X}$) as $[\mathbf{X}]_{i,:} = \mathbf{x}_i$, features (columns of $\mathbf{X}$) as $[\mathbf{X}]_{:,m} = \mathbf{x}^m$ and single elements as $[\mathbf{X}]_{i,m} = x_i^m$.

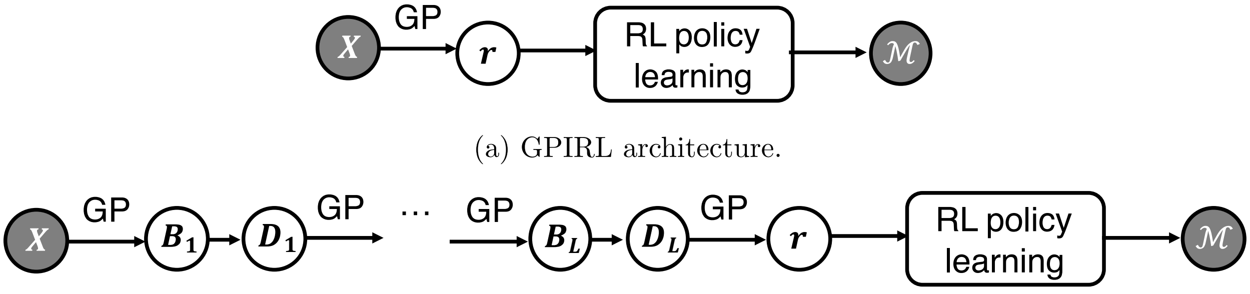
In this modeling framework, the reward function r plays the role of an unknown mapping, thus we wish to treat it as *latent* and keep it flexible and non-linear. Therefore, we model it with a zero-mean GP prior [143], [187]:

$$r \sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}),$$

where $k_{\boldsymbol{\theta}}$ is the covariance function, e.g., $k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_k^2 e^{-\frac{\xi}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top(\mathbf{x}_i - \mathbf{x}_j)}$ with parameters $\boldsymbol{\theta} = \{\sigma_k, \xi\}$. Given a finite amount of data, this induces the probability $\boldsymbol{r}|\mathbf{X}, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, K_{\mathbf{XX}})$, where $\boldsymbol{r} \triangleq r(\mathbf{X})$ is a vector of rewards evaluated for each row of $\mathbf{X}$ and the covariance matrix is obtained by $[K_{\mathbf{XX}}]_{i,j} = k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)$ for each entry. The GPIRL training objective is the likelihood function, which comes from integrating out the latent reward (see Fig. 5.1a):

$$p(\mathcal{M}|\mathbf{X}) = \int p(\mathcal{M}|\boldsymbol{r})p(\boldsymbol{r}|\mathbf{X}, \boldsymbol{\theta})\mathrm{d}\boldsymbol{r} \tag{5.5}$$

and the maximizing parameter is $\boldsymbol{\theta}$, which we drop from our expressions from now on. The above integral is intractable, because $p(\mathcal{M}|\boldsymbol{r})$ has the complicated expression of (5.4) (this



(a) GPIRL architecture.

(b) DGP-IRL architecture.

Figure 5.1: Comparison of the GPIRL and DGP-IRL architectures. GPIRL models the reward function as a Gaussian process, while DGP-IRL stacks latent spaces ($\mathbf{B}_l$ and $\mathbf{D}_l$) connected through GPs to form a deep GP representation.

is in contrast to the tranditional GP regression where $\mathcal{M}|\boldsymbol{r}$ is a Gaussian or other simple distribution). This can be alleviated using the approximation of [143]. We will describe this approximation in the next section, as it is also used by our approach. Notice that all latent function instantiations are linked through a joint multivariate Gaussian. Thus, prediction of the function value $r^* = r(\mathbf{x}^*)$ at a test input $\mathbf{x}^*$ is found through the conditional

$$r^*|\boldsymbol{r}, \mathbf{X}, \mathbf{x}^* \sim \mathcal{N}\left(K_{\mathbf{x}^*\mathbf{X}} K_{\mathbf{X}\mathbf{X}}^{-1} \boldsymbol{r}, k_{\mathbf{x}^*\mathbf{x}^*} - K_{\mathbf{x}^*\mathbf{X}} K_{\mathbf{X}\mathbf{X}}^{-1} K_{\mathbf{X}\mathbf{x}^*}\right)$$

As can be seen, the prediction $r(\mathbf{x}^*)$ is reliant on the effectiveness of feature representation: states with features close in Euclidean distance are assumed to be associated with similar rewards. This motivates our novel deep GP method which is obtained by considering additional layers to increase the feature expressiveness.

Incorporating the representation learning layers

The traditional model-based IRL approach is to learn the latent reward $\boldsymbol{r}$ that best explains the demonstrations $\mathcal{M}$. In this chapter we aim at additionally and simultaneously uncovering a highly descriptive state feature representation. To achieve this, we introduce a *latent* state feature representation $\mathbf{B} = \left[\mathbf{b}^1, ..., \mathbf{b}^{m_1}\right] \in \mathbb{R}^{n,m_1}$, where $\mathbf{B}$ constitutes the instantiations of an introduced function b which is learned as a non-linear GP transformation from $\mathbf{X}$. To account for noise we further introduce $\mathbf{D}$ as the noisy versions of $\mathbf{B}$, i.e., $d_i^m = b_i^m + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \lambda^{-1})$. Together, $\mathbf{B}$ and $\mathbf{D}$ form a hidden layer, and the layer can be repeated several times, as illustrated in Fig. 5.1b. To streamline the presentation, we will illustrate the method with a 2-layered deep GP, but the method can be applied to deeper structures.

Importantly, rather than performing two separate steps of learning (for the GPs on r and on b), we nest them into a single objective function, to maintain the flow of information during optimization. This results in a deep GP whose top layers perform representation learning and lower layers perform model-based IRL (Fig. 5.1b), called Deep Gaussian Process for

Inverse Reinforcement Learning (DGP-IRL). By using $\mathbf{x}^m$, $\mathbf{d}^m$, $\mathbf{b}^m$ to represent the m-th column of $\mathbf{X}$, $\mathbf{D}$, $\mathbf{B}$ respectively, the full generative model can be written as:

$$\begin{aligned} p(\mathcal{M}, \boldsymbol{r}, \mathbf{D}, \mathbf{B}|\mathbf{X}) &= \underbrace{p(\mathcal{M}|\boldsymbol{r})}_{\text{IRL}} \underbrace{p(\boldsymbol{r}|\mathbf{D})}_{\mathcal{GP}(\mathbf{0},k^r(\mathbf{d}_i,\mathbf{d}_j))} \underbrace{p(\mathbf{D}|\mathbf{B})}_{\text{Gaussian noise}} \underbrace{p(\mathbf{B}|\mathbf{X})}_{\mathcal{GP}(\mathbf{0},k^b(\mathbf{x}_i,\mathbf{x}_j))} \\ &= e^{\sum_{h=1}^{H}\sum_{t=1}^{T}\left(Q(s_{h,t},a_{h,t};\boldsymbol{r})-V(s_{h,t};\boldsymbol{r})\right)} \mathcal{N}(\boldsymbol{r}|\mathbf{0}, K_{\mathbf{DD}}) \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{d}^m|\mathbf{b}^m, \lambda^{-1}\mathbf{I})\mathcal{N}(\mathbf{b}^m|\mathbf{0}, K_{\mathbf{XX}}), \end{aligned} \tag{5.6}$$

where the IRL term $p(\mathcal{M}|\boldsymbol{r})$ takes the form of (5.4), $K_{\mathbf{XX}}$ and $K_{\mathbf{DD}}$ are the covariance matrices in each layer, constructed with covariance functions k^b and k^r, respectively. Compared to GPIRL, the proposed framework has substantial gain in flexibility by introducing the abstract representation of states in the hidden layers $\mathbf{B}$ and $\mathbf{D}$.

We can compress the statistical power of each generative layer into a set of auxiliary variables within a sparse GP framework [204]. Specifically, we introduce *inducing outputs and inputs*, denoted by $\boldsymbol{f} \in \mathbb{R}^{\alpha}$ and $\mathbf{Z} \in \mathbb{R}^{\alpha,m_1}$ respectively for the lower layer and by $\mathbf{V} \in \mathbb{R}^{\alpha,m_1}$ and $\mathbf{W} \in \mathbb{R}^{\alpha,m_0}$ for the top layer (as illustrated in Fig. 5.2). The inducing outputs and inputs are related with the same GP prior appearing in each layer. For example, $\boldsymbol{f}|\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, K_{\mathbf{ZZ}})$ with $K_{\mathbf{ZZ}} = k^r(\mathbf{Z}, \mathbf{Z})$. By relating the original and inducing variables through the conditional Gaussian distribution, the auxiliary variables are learned to be *sufficient statistics* of the GP. The augmented model, shown in Fig. 5.2, has the following full distribution:

$$\begin{aligned} &p(\mathcal{M}, \boldsymbol{r}, \boldsymbol{f}, \mathbf{B}, \mathbf{D}, \mathbf{V}|\mathbf{X}, \mathbf{Z}, \mathbf{W}) \\ &= p(\mathcal{M}|\boldsymbol{r})p(\boldsymbol{r}|\boldsymbol{f}, \mathbf{D}, \mathbf{Z})p(\boldsymbol{f}|\mathbf{Z})p(\mathbf{D}|\mathbf{B})p(\mathbf{B}|\mathbf{V},\mathbf{X},\mathbf{W}) \qquad (5.7) \\ &= p(\mathcal{M}|\boldsymbol{r})\mathcal{N}(\boldsymbol{r}|K_{\mathbf{DZ}}K_{\mathbf{ZZ}}^{-1}\boldsymbol{f}, \boldsymbol{\Sigma}_r)\mathcal{N}(\boldsymbol{f}|\mathbf{0}, K_{\mathbf{ZZ}}) \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{d}^m|\mathbf{b}^m, \lambda^{-1}\mathbf{I})\mathcal{N}(\mathbf{b}^m|K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\mathbf{v}^m, \boldsymbol{\Sigma}_B), \end{aligned}$$

where we adopt the fully independent training conditional (FITC) to preserve the exact variances in $\boldsymbol{\Sigma}_B = diag\left(K_{\mathbf{XX}} - K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}K_{\mathbf{WX}}\right)$, and the deterministic training conditional (DTC) in $\boldsymbol{\Sigma}_r = \mathbf{0}$ as in GPIRL to facilitate the integration of $\boldsymbol{r}$ in the training objective [185].

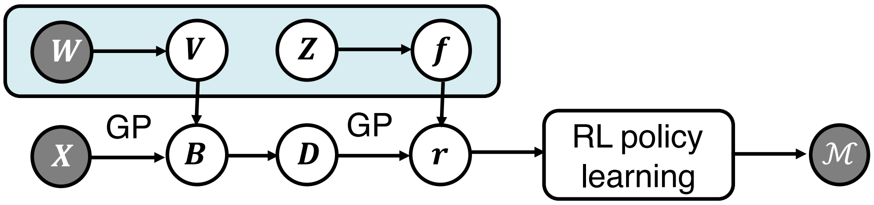


Figure 5.2: Illustration of DGP-IRL with the inducing outputs $\boldsymbol{f}, \mathbf{V}$ and inputs $\mathbf{Z}, \mathbf{W}$.

In the following, we will omit the inducing inputs $\mathbf{W}, \mathbf{Z}$ in the conditions, with the convention to treat them as model parameters [37], [117]. By selecting the number of inducing points $\alpha \ll n$ the complexity reduces from $\mathcal{O}(n^3)$ to $\mathcal{O}(n\alpha^2)$. While DGP-IRL resolves the case when the outputs have complex dependencies with the latent layers, the training of the model based on variational inference requires gradients for the parameters, as in backpropagation in deep neural network training, whose convergence can be improved by leveraging advancements in deep learning. Additionally, in DGP-IRL, the role of auxiliary variables goes further than just introducing *scalability*. Indeed,the auxiliary variables play a distinct role in our model by forming the base of a variational framework for Bayesian inference.

Variational inference and transfer learning

For model training, our task is to optimize the model evidence

$$p(\mathcal{M}|\mathbf{X}) = \int p(\mathcal{M}, \boldsymbol{f}, \boldsymbol{r}, \mathbf{V}, \mathbf{D}, \mathbf{B}|\mathbf{X})\mathrm{d}(\boldsymbol{f}, \boldsymbol{r}, \mathbf{V}, \mathbf{D}, \mathbf{B}). \tag{5.8}$$

However, this quantity is intractable. Firstly because the latent variables $\mathbf{D}$ appear nonlinearly in the inverse of covariance matrices. Secondly because the latent rewards $\boldsymbol{f}, \boldsymbol{r}$ relate to the observation $\mathcal{M}$ through the reinforcement learning layer; the choice of $\boldsymbol{\Sigma}_r = \mathbf{0}$ in (5.7) does not completely solve this problem because in DGP-IRL there is additional uncertainty propagated by the latent layers. This indicates that Laplace approximation is not practical, neither is the variational method employed for deep GP, where the output is related to the latent variable in a simple regression framework [37], [117].

To this end, we show that we can derive an analytic lower bound on the model evidence by constructing a variational framework using the following special form of variational distribution $\mathcal{Q} = q(\boldsymbol{f})q(\mathbf{D})q(\mathbf{B})q(\mathbf{V})$, where

$$q(\boldsymbol{f}) = \delta(\boldsymbol{f} - \tilde{\boldsymbol{f}}) \tag{5.9}$$

$$q(\mathbf{B}) = p(\mathbf{B}|\mathbf{V}, \mathbf{X}) \tag{5.10}$$

$$q(\mathbf{D}) = \prod_{m=1}^{m_1} \delta\left(\mathbf{d}^m - K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} \tilde{\mathbf{v}}^m\right) \tag{5.11}$$

$$q(\mathbf{V}) = \prod_{m=1}^{m_1} \mathcal{N}\left(\mathbf{v}^m | \tilde{\mathbf{v}}^m, \mathbf{G}^m\right) \tag{5.12}$$

where $\tilde{\boldsymbol{f}}$, $\tilde{\mathbf{v}}^m$, $\mathbf{G}^m$ are variational parameters. The delta distribution is equivalent to taking the mean of normal distributions for prediction, which is reasonable in the context of reinforcement learning. Also note that the delta distribution is *applied only in the bottom layer and not repeatedly*; therefore, representation learning is indeed being manifested in the latent layers. In addition, $q(\mathbf{B})$ matches the exact conditional $p(\mathbf{B}|\mathbf{V}, \mathbf{X})$ so that these two terms cancel in the fraction of (5.14) and the number of variational parameters is minimized, as in [219]. As for $q(\mathbf{D})$, it is chosen as delta distributions such that, in tandem with $\boldsymbol{\Sigma}_r = \mathbf{0}$, the IRL term $p(\mathcal{M}|\boldsymbol{r})p(\boldsymbol{r}|\boldsymbol{f}, \mathbf{D})$ in (5.13) becomes tractable and information can flow through the latent layers $\mathbf{B}, \mathbf{D}$. The variational marginal $q(\mathbf{V})$ is factorized across its dimensions with fully parameterized normal densities. Notice that $\tilde{\boldsymbol{f}}$ and $\tilde{\mathbf{v}}$ are the mean of the inducing outputs, corresponding to pseudo-inputs $\mathbf{Z}$ and $\mathbf{W}$, where $\mathbf{Z}$ (initialized with random numbers from uniform distributions [210]) can be learned to further maximize the marginaliked likelihood, and $\mathbf{W}$ is chosen as a subset of $\mathbf{X}$. Further, the variational means of $\mathbf{D}$ can be augmented with input data $\mathbf{X}$ to improve stability during training [57].

The variational lower bound, $\mathcal{L}$, follows from the Jensen's inequality, and can be derived analytically due to the choice of variational distribution $\mathcal{Q}$ (see Appendix A.4 for detailed derivation):

$$\log p(\mathcal{M}|\mathbf{X}) = \log \int p(\underbrace{\mathcal{M}|\boldsymbol{r})p(\boldsymbol{r}|\boldsymbol{f},\mathbf{D}}_{p(\mathcal{M}|\boldsymbol{r}=K_{\mathbf{DZ}}K_{\mathbf{ZZ}}^{-1}\boldsymbol{f}) \text{ by DTC: } \boldsymbol{\Sigma}_r=\mathbf{0}})p(\boldsymbol{f})p(\mathbf{D}|\mathbf{B})p(\mathbf{B}|\mathbf{V},\mathbf{X})p(\mathbf{V})\mathrm{d}(\boldsymbol{r},\boldsymbol{f},\mathbf{V},\mathbf{D},\mathbf{B}) \tag{5.13}$$

$$\geq \int q(\boldsymbol{f})q(\mathbf{D})p(\mathbf{B}|\mathbf{V},\mathbf{X})q(\mathbf{V}) \log \frac{p(\mathcal{M}|K_{\mathbf{DZ}}K_{\mathbf{ZZ}}^{-1}\boldsymbol{f})p(\boldsymbol{f})p(\mathbf{D}|\mathbf{B})p(\mathbf{V})}{q(\boldsymbol{f})q(\mathbf{D})q(\mathbf{V})} \tag{5.14}$$

$$= \mathcal{L}_{\mathcal{M}} + \mathcal{L}_{\mathcal{G}} - \mathcal{L}_{\mathrm{KL}} + \mathcal{L}_{\mathcal{B}} - \frac{nm_1}{2}\log(2\pi\lambda^{-1}) \tag{5.15}$$

with

$$\mathcal{L}_{\mathcal{M}} = \log p(\mathcal{M}|K_{\tilde{\mathbf{D}}\mathbf{Z}}K_{\mathbf{ZZ}}^{-1}\tilde{\boldsymbol{f}})) \tag{5.16}$$

$$\mathcal{L}_{\mathcal{G}} = \log p(\boldsymbol{f} = \tilde{\boldsymbol{f}}|\mathbf{Z}) = \log \mathcal{N}(\boldsymbol{f} = \tilde{\boldsymbol{f}}|\mathbf{0}, K_{\mathbf{ZZ}}) \tag{5.17}$$

$$\mathcal{L}_{KL} = \mathrm{KL}(q(\mathbf{V})||p(\mathbf{V}|\mathbf{W})) = \sum_{m=1}^{m_1} \mathrm{KL}(\mathcal{N}(\mathbf{v}^m|\tilde{\mathbf{v}}^m, \mathbf{G}^m)||\mathcal{N}(\mathbf{v}^m|\mathbf{0}, K_{\mathbf{WW}})) \tag{5.18}$$

$$\mathcal{L}_{\mathcal{B}} = -\frac{\lambda}{2}\sum_{m=1}^{m_1} \mathrm{Tr}(\boldsymbol{\Sigma}_{\mathbf{B}} + K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\mathbf{G}^m K_{\mathbf{WW}}^{-1}K_{\mathbf{WX}}) \tag{5.19}$$

where $|K_{\mathbf{WW}}|$ is the determinant of $K_{\mathbf{WW}}$, $\tilde{\mathbf{D}} = \left[\tilde{\mathbf{d}}^1, ..., \tilde{\mathbf{d}}^{m_1}\right]$ with $\tilde{\mathbf{d}}^m = K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\tilde{\mathbf{v}}^m$, $\mathcal{L}_{\mathcal{M}}$ is the term associated with RL, $\mathcal{L}_{\mathcal{G}}$ is the Gaussian prior on inducing outputs $\boldsymbol{f}$, and $\mathcal{L}_{KL}$ denotes the Kullback–Leibler (KL) divergence between the variational posterior $q(\mathbf{V})$ to the prior $p(\mathbf{V})$, acting as a regularization term. The lower bound $\mathcal{L}$ can be optimized with gradient-based method like backpropagation. In addition, we can find the optimal fixed-point equations for the variational distribution parameters $\tilde{\mathbf{v}}^m, \mathbf{G}^m$ for $q(\mathbf{V})$ using variational calculus, in order to raise the variational lower bound $\mathcal{L}$ further (refer to Appendix A.4 for this derivation). Notice that the approximate marginalization of all hidden spaces in (5.15) approximates a Bayesian training procedure, according to which model complexity is automatically balanced through the Bayesian Occam's razor principle. Optimizing the objective $\mathcal{L}$ turns the variational distribution $\mathcal{Q}$ into an approximation to the true model posterior.

The inducing points provide a succinct summary of the data by the property of FITC, which means only the inducing points are necessary for prediction [185]. Given a set of new states $\mathbf{X}^*$, DGP-IRL can infer the latent reward through the full Bayesian treatment:

$$p(\boldsymbol{r}^*|\mathbf{X}^*, \mathbf{X}) = \int p(\boldsymbol{r}^*|\boldsymbol{f}, \mathbf{D}^*)q(\boldsymbol{f})p(\mathbf{D}^*|\mathbf{B}^*)p(\mathbf{B}^*|\mathbf{V}, \mathbf{X}^*)q(\mathbf{V})\mathrm{d}(\boldsymbol{f}, \mathbf{B}^*, \mathbf{D}^*, \mathbf{V}) \tag{5.20}$$

Given that the above integral is computationally intensive to evaluate, a practical alternative adopted in our implementation is to use point estimates for latent variables; hence, the rewards are given by:

$$\boldsymbol{r}^* = K_{\mathbf{D}^*\mathbf{Z}}K_{\mathbf{ZZ}}^{-1}\tilde{\boldsymbol{f}}, \tag{5.21}$$

where $\mathbf{D}^* = [\mathbf{d}_*^1, ..., \mathbf{d}_*^{m_1}]$, with $\mathbf{d}_*^m = K_{\mathbf{X}^*\mathbf{W}}K_{\mathbf{WW}}^{-1}\tilde{\mathbf{v}}^m$. The above formulae suggest that instead of making inference based on $\mathbf{X}$ layer directly as in [143], DGP-IRL first estimates the latent representation of the states $\mathbf{D}^*$, then makes GP regression using the latent variables.

5.3 Experiments on benchmarks

For the experimental evaluation, we employ the *expected value difference* (EVD) as a metric of optimality,

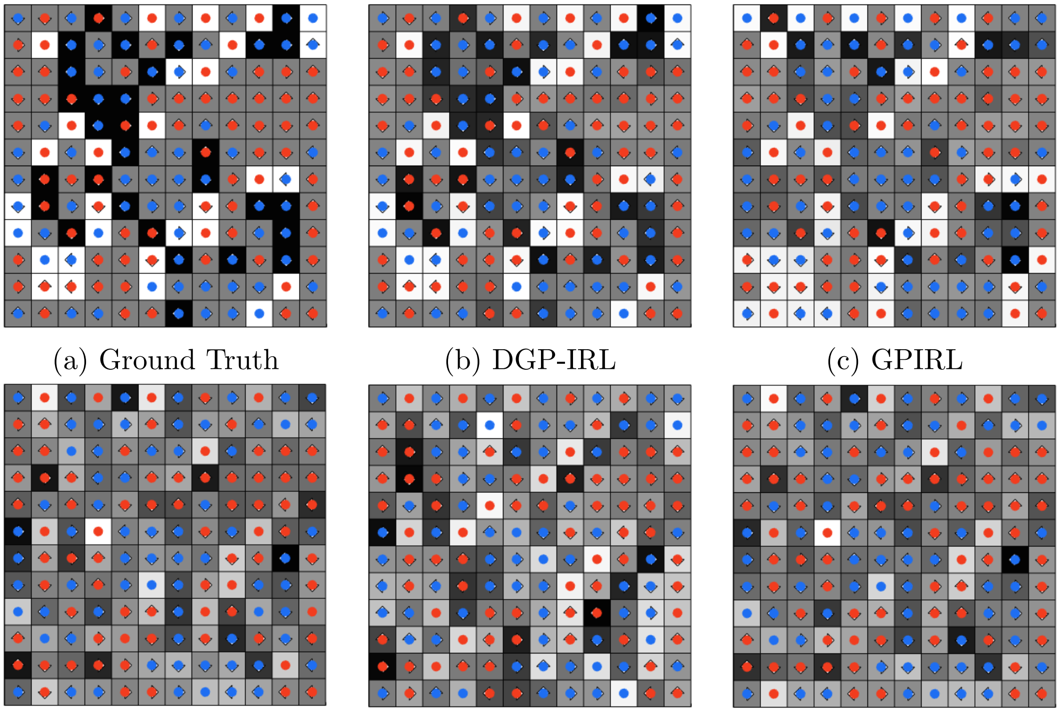
$$\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t r(s_t)|\pi^*\right] - \mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t r(s_t)|\hat{\pi}\right], \tag{5.22}$$

which is the difference between the expected reward earned under the optimal policy, π^*, given by the true rewards, and the policy derived from the IRL rewards, $\hat{\pi}$.[1]

Binary world (BW) is a benchmark whose reward depends on combinatorics of features [234]. More specifically, in a $N \times N$ gridworld where each block is randomly assigned with either a blue or red dot, the state is associated with the +1 reward if there are 4 blues in the 3×3 neighborhood, -1 if there are 5 blues in the neighborhood, and 0 otherwise (illustrated in Fig. 5.3a). The feature represents the color of the 9 dots in the neighborhood. The agent maximizes its expected discounted reward by following a policy which provides the probabilities of actions (moving up/down/left/right, or stay still) at each state, subject to a transition probability.

The objective of the experiment is to compare the performances of various IRL algorithms to recover the latent rewards given limited demonstrations. Candidates that have been evaluated include Learning to search (LEARCH) [192], MaxEnt [247], and MMP [191], which assume a linear reward function, and GPIRL [143], which is the state-of-the-art method for IRL. BW sets up a challenging scenario, where states that are maximally separated in feature space can have the same rewards, yet those that are close in euclidean distance may have opposite rewards. While linear models were limited by their capacity of representation, the results of GPIRL also deviated from the latent rewards as it could not generalize from training data with the convoluted features. DGP-IRL, nevertheless, was able to recover the ground truth with the highest fidelity, as shown in Fig. 5.3.

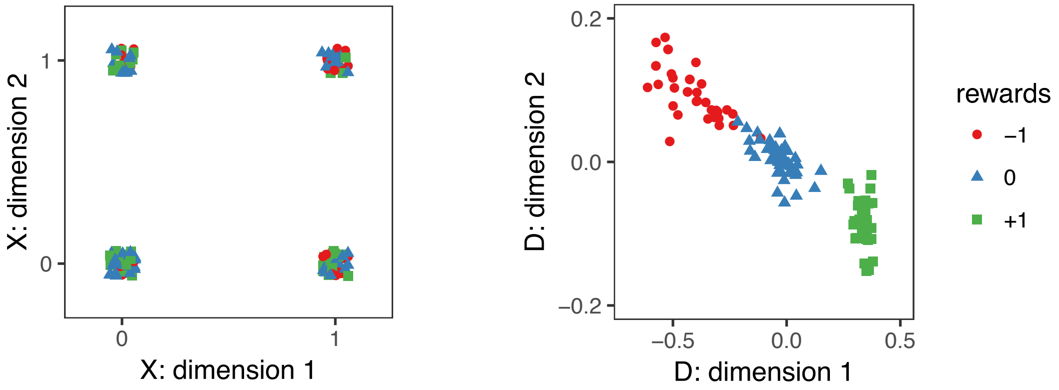
[1] The software implementation can be accessed at: https://github.com/jinming99/DGP-IRL.



(a) Ground Truth (b) DGP-IRL (c) GPIRL

(d) LEARCH (e) MaxEnt (f) MMP

Figure 5.3: BW benchmark evaluation with 128 demonstrated traces for DGP-IRL, GPIRL [143], LEARCH [192], MaxEnt [247], and MMP [191].



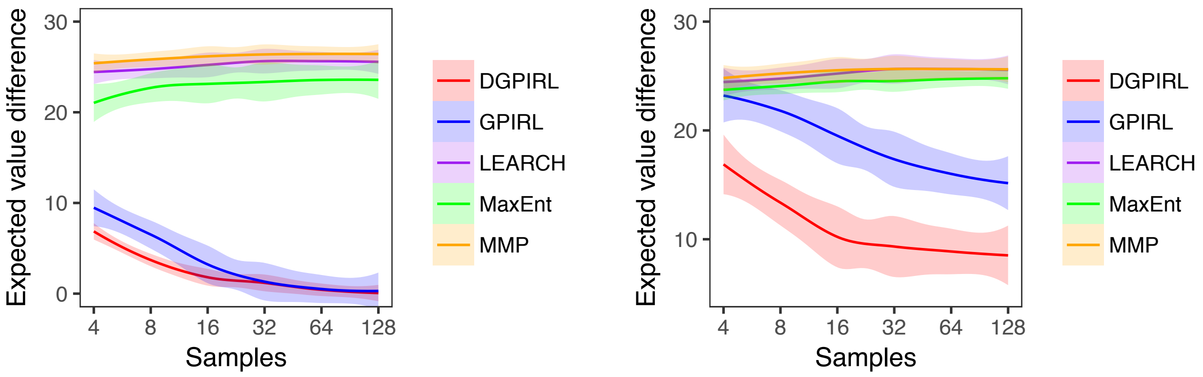
(a) Input space **X** and rewards. (b) Latent space **D** and rewards.

Figure 5.4: Visualization of points (features of states) along two arbitrary dimensions in the (a) input space **X** and (b) latent space **D** of DGP-IRL. The rewards are entangled in the input space **X** but separated in the latent space **D**.

By successively warping the original feature space through the latent layers, DGP-IRL could learn an abstract representation that revealed the reward structure. As illustrated in Fig. 5.4, though the points were mixed up in the input space, making it impossible to separate those with the same rewards based on the input features **X** alone, their positions in the latent space clearly formed clusters when viewed from latent features **D**, implying that DGP-IRL had remarkably uncovered the mechanism of reward generation by simply observing the traces of actions.

Additionally, the transferability test was carried out by examining EVD in a new world where no demonstrations were available, which required the ability of knowledge transfer from the previous learning scenario. As the features were interlinked not only with the reward but also with themselves in a very nonlinear way, this scenario was particularly challenging for linear models like LEARCH, MaxEnt and MMP. The advantage of simultaneous representation and inverse reinforcement learning was demonstrated in Fig. 5.5, where DGP-IRL outperformed GPIRL and other models in both the training and transfer cases, and the improvement was obvious as more data became accessible.

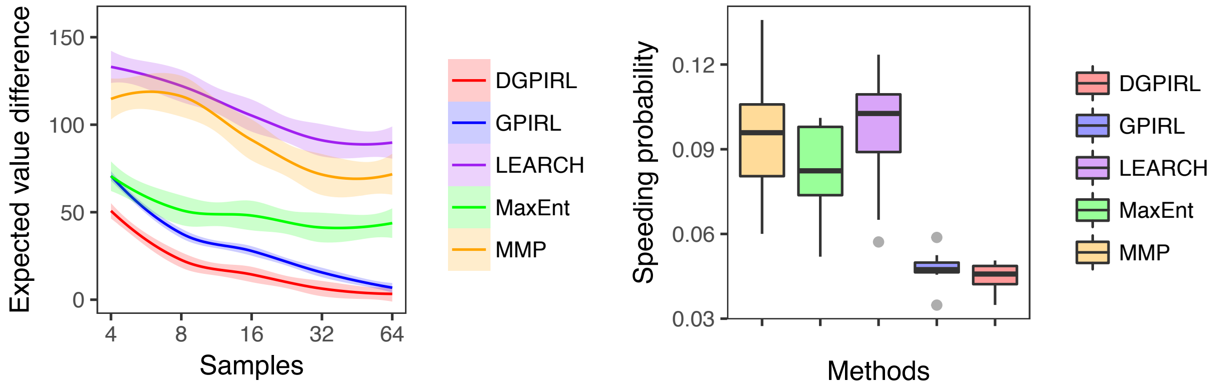
For another benchmark, highway driving behavior modeling is a concrete example to examine the capacity of IRL algorithms in learning the underlying motives from human demonstrations [142], [143]. In a three-lane highway, vehicles of specific class (civilian or police) and category (car or motorcycle) are positioned at random, driving at the same constant speed. The autonomous car can switch lanes and navigate at up to three times the traffic speed. The state is described by a continuous feature which consists of the closest distances to vehicles of each class and category in the same lane, together with the left, right, and any lane, both in the front and back of the robot car, in addition to the current speed and position.



(a) Training EVD.

(b) Transfer task EVD.

Figure 5.5: Plots of EVD in the training (a) and transfer (b) tests for the BW benchmark as the number of training samples varies. The shaded area indicates the standard deviation of EVD among independent experiments.



(a) Training EVD.

(b) Speeding risk.

Figure 5.6: Plots of EVD in the training (a) and the risk of speeding (with 64 demonstrations) (b) in the highway driving simulation benchmark, with three lanes and 32 car lengths.

The goal is to navigate the robot car as fast as possible, but to avoid speeding by checking that the speed is no greater than twice the current traffic when the police car is 2 car lengths nearby. As the reward is a nonlinear function determined by the current speed and distance to the police, linear models were outrun by GPIRL and DGP-IRL. Performance generally improved with more demonstrations, and DGP-IRL remained to yield the policy closest to the optimal in EVD, and with minimal risk of speeding, as illustrated in Fig. 5.6a and 5.6b, respectively.

5.4 Chapter summary

In this chapter we proposed deep Bayesian inverse reinforcement learning to infer an agent's intention from its behaviors. By extending the deep GP framework to the IRL domain, we enabled learning latent rewards with more complex structures from limited data. As training the network involves evaluating a likelihood that is intractable, we derived a variational lower bound using an innovative definition of the variational distributions. This methodological contribution enables Bayesian learning in our model and can be applied to other scenarios where the observation layer's dynamics cause similar intractabilities. We compared the proposed DGP-IRL with existing approaches in benchmark tests as well as highway driving tasks with human demonstrations, and verified its capability of handling complex reward with scarce data. This represents a new category of human preference learning in h-CPS with *in-vivo* observations.

Part II

System-level efficiency and resilience

Chapter 6

Enabling optimal energy retail in a microgrid

The conventional approach of h-CPS system operator to improve system efficiency and resilience is by optimizing physical control and operations (for example, the unit commitment task for the electricity grid). When the system is under unusual stress, this strategy tends to compromise economy to ensure reliability and safety (e.g., oversized equipment and excessive reserves). Enabled by the contextual awareness of human factors and system states in the first part of this thesis, this chapter employs “behavioral nudges” as a “nexus point” between end-users and system operations, and exploits *end-use demand flexibility* to further enhance efficiency and reliability. We focus on designing energy retail rates and dispatch generators in a microgrid, which is a group of interconnected loads and distributed energy resources within clearly defined electrical boundaries that acts as a single controllable entity with respect to the grid. We leverage the demand elasticity concept from economics to model the end-users’ responses to price signals. By modeling an integrated energy system that provides both thermal and electrical power, synergy naturally emerges from the optimization for generation dispatch and optimal rate design, enabling enhanced system efficiency and substantial savings for both the energy provider and end-use customers.

6.1 Microgrid and optimal energy retail

The transition from an economy that relies heavily on fossil fuels to one that is powered primarily by renewable energy has been accelerating in recent years, bolstered by mounting concerns over climate change and falling prices of solar and wind energy [178]. However, the penetration of volatile, distributed renewable resources can potentially destablize the grid. Furthermore, grid resilience and rapid self-recovery in the face of natural disasters and malicious attacks are extremely necessary features [172]. Driven by the evolution of technologies and markets, there is a fundamental push across the industry to update utility rate structures as the existing tariff becomes less and less efficient [60]. Meanwhile, the emergence

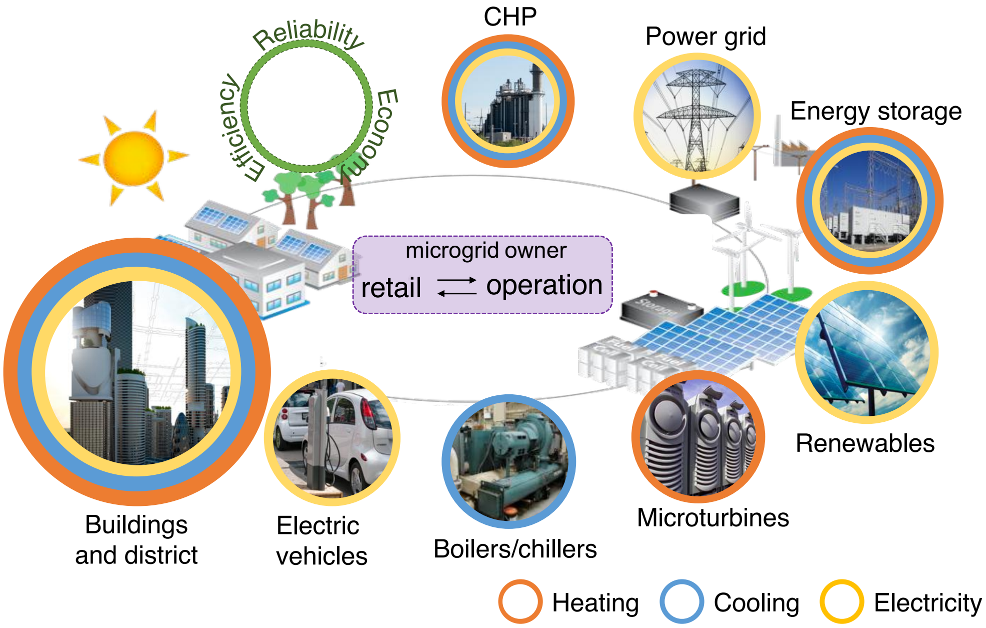


Figure 6.1: Schematic of MR-POD that jointly optimizes energy retail and dispatch by considering demand flexibility and generator synergy.

of electricity retail services enables customers to choose providers [168]. Increased competition exposes retailers to greater risks, while not necessarily reducing customers' bills [168]. Clearly, a systematic strategy for rate design and resource management is central to the ongoing transformation of the system. This paradigm shift can be further driven by demand response (DR) by the institution of time-differentiated retail pricing, e.g., time-of-use (TOU) and real-time pricing (RTP), which reflect fluctuating wholesale prices and explore end-user demand flexibility [116], [125], [168], [223]. Currently, RTP is most popular in the wholesale market, while being experimented on a few sites like the Illinois Power Company on the retail side. However, with the increasing penetration of internet-of-things (IoT) devices and occupancy-aware building controls, buildings' responsiveness can be significantly enhanced through automated services. Thus, study on DR at the retail level that finds its optimal pricing scheme and relationship with local distributed energy resources adoption and operation becomes increasingly important [71].

On the supply side, the division of the grid into productive sub-systems, microgrids (MGs), that integrate distributed generation (DG) and storage to serve local demand, has been proposed to increase manageability, energy efficiency, and resilience [141], [240]. The rapid development of integrated energy systems (IESs), which exploit the synergistic potential of thermal and electrical provision, is crucial for flexibility enhancement, carbon dioxide (CO_2) reduction, and renewable integration [150]. Nevertheless, in places like Jilin province in northeastern China, about 89% of the total wind power curtailment is caused by operating conventional CHP at full load to satisfy high heat demands and lack of curtailable supply [233]. The central task for a retailer with generation capacity is thus to design energy rates and operate the facility to gain profits and preserve system stability. Previous works on MG operation often treat it as a non-profit entity that does not price its energy output, which confines its application to campuses and other situations where the total bill is paid by the MG owner [16], [34], [47], [78], [82], [97], [176], [177], [179], [223]. As a result, DR in a MG is limited to contracts [176] or a mutual agreement where the MG operator has central control over DR-enabled loads [71], [97], [124]. This often requires the integration of advanced communication infrastructure and might raise security and privacy issues. Furthermore, the scope is predominantly within electricity provision, rather than exploiting synergies of IESs [40], [81], [124], [176]. We focus on future smart MG with time-differentiated pricing on the retail side and propose a **M**icrogrid **R**etailer **P**ricing and **O**peration strategy with **D**emand response, namely MR-POD, to capture the new opportunity (see Fig. 6.1).

MG modeling and dispatch. Previous work has been undertaken on modeling high-level system design for MGs to study their profitability and optimal technology selection [16], [49], [82], [154], [177]. The dispatch of MG has been attempted through a variety of approaches, including mixed integer linear programming (MILP) [41], dynamic programming [49], simulated annealing [223], particle swarm optimization [16], evolutionary algorithms [177] and game theoretic agent-based formulations [47]. An empirical comparison of LP, MILP, and non-linear programming (NLP), had been conducted, and the study concluded that MILP is the most appropriate model from the viewpoints of accuracy and runtime [179]. In comparison, our formulation of the dispatch problem as MIQP also *addresses the uncertainty in renewable generation and the flexibility in demand that facilitates DR.*

Demand response. Demand response (DR) is becoming a cost-effective balancing resource in power systems. According to the US Department of Energy, DR is “a tariff or program established to motivate changes in electric usage by end-use customers, in response to changes in the price of electricity over time, or to give incentive payments designed to induce lower electricity usage at times of high market prices or when grid reliability is jeopardized” [60]. There are mainly two groupings of DR programs: price-based DR and incentive-based DR, with the key difference that the former offers customers time-varying or localized prices, while the latter grants fixed or time-varying payments under specific contracts [62]. The efficacies of price-based DR have been empirically examined in several studies [62]. This work focuses on the design of rate signals for price-based DR, with a special focus on *retailers who own distributed generation system and can price the energy for profits and DR.*

Optimal rate design. Smart pricing plays a vital role in DR to increase system reliability, reduce generation costs, and lower consumers’ bills [166]. To determine the consumer response, price-elastic load models were proposed [165], [239], where the elasticity is often estimated using panel data [8], [64]. Methods based on mixed-integer stochastic programming

[81], [176] and noncooperative games [116] have been proposed to determine the optimal sale price of electricity and the electricity procurement policy of a retailer. However, these works only focus on electricity supply and profit maximization for the retailer without providing DR incentivization to enroll customers in the programs, which are often voluntary in practice. Differentiated from the previous studies, MR-POD is aimed at *providing guidance on optimal MG operation and pricing on a district level with integrated energy provision.* By leveraging the efficiency of energy coupling and demand flexibility, the model provides a cost-effective and grid-cooperative strategy in a competitive and uncertain market.

6.2 Integrated energy retail model

The MG dispatch and retailer pricing with DR problem is formulated within an optimization framework. Key components, including the MG generator and building loads, are shown in Fig. 6.2, where flows of cash, energy, and information within the MG are illustrated.

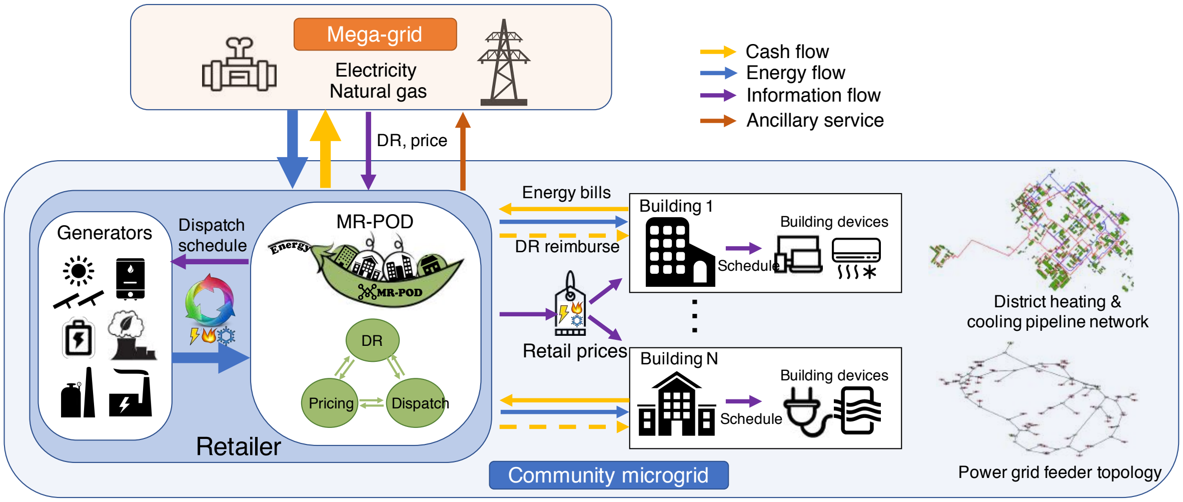


Figure 6.2: Overview of the retailer model, incorporating generator dispatch and energy retailing to serve a community. The microgrid can optionally connect to the utility grid for electricity procurement and participate in ancillary services like DR.

Problem formulation

The key problem that MR-POD solves is: "How should energy prices be set and the microgrid operated to maximize retailer profit while satisfying building demand and grid requirements?" Two prominent factors are involved:

- *Price elasticity* of loads for individual buildings under the DR scheme

- *Uncertainty and fluctuation* of energy demand, electricity and thermal tariffs, and weather conditions

Price elasticity refers to the change in energy demand in response to a change in product price [62], [128], which can be used by the retailer to estimate peak demand reduction potential, and provision of ancillary services to the grid. The uncertainty aspect, inherent for all planning problems, is addressed by forecasting, as discussed in Section 6.3.

The basic MR-POD problem is formulated as follows (see Fig. 6.3 for an illustration):

$$\max_{\{\mathbf{x}_t,\boldsymbol{p}_t\}_{t=1}^T} \sum_{t=1}^{T} f_t^{\text{Rev}}(\mathbf{d}_t,\boldsymbol{p}_t) - f_t^{\text{Ope}}(\mathbf{x}_t,\mathbf{z}_t,\boldsymbol{\xi}_t) - \lambda_{env} f_t^{\text{Env}}(\mathbf{x}_t,\mathbf{z}_t)$$
$$\text{s.t.}\quad \mathbf{x}_t \in \mathcal{X}_t(\mathbf{z}_t,\mathbf{d}_t,\boldsymbol{\xi}_t), \mathbf{d}_t \in \mathcal{D}_t(\boldsymbol{p}_t), \boldsymbol{p}_t \in \mathcal{P}_t, \quad \forall t = 1,...,T \qquad \text{(MR-POD)}$$

where $\mathbf{x}_t$ is the dispatch proposal at time t, which includes variables in three categories: *generation* from on-site power plants, *storage charging/discharging*, and *grid import/export*. The energy demand of buildings $\mathbf{d}_t$ is a function of retail prices and DR incentives $\boldsymbol{p}_t$ determined by MR-POD. The state variable $\mathbf{z}_t$ captures the state-of-charge (SOC) of the storage as governed by the previous state and any actions. The external quantities, e.g., solar irradiation Irr_t and electricity price c_t^{grid}, are summarized in $\boldsymbol{\xi}_t$.

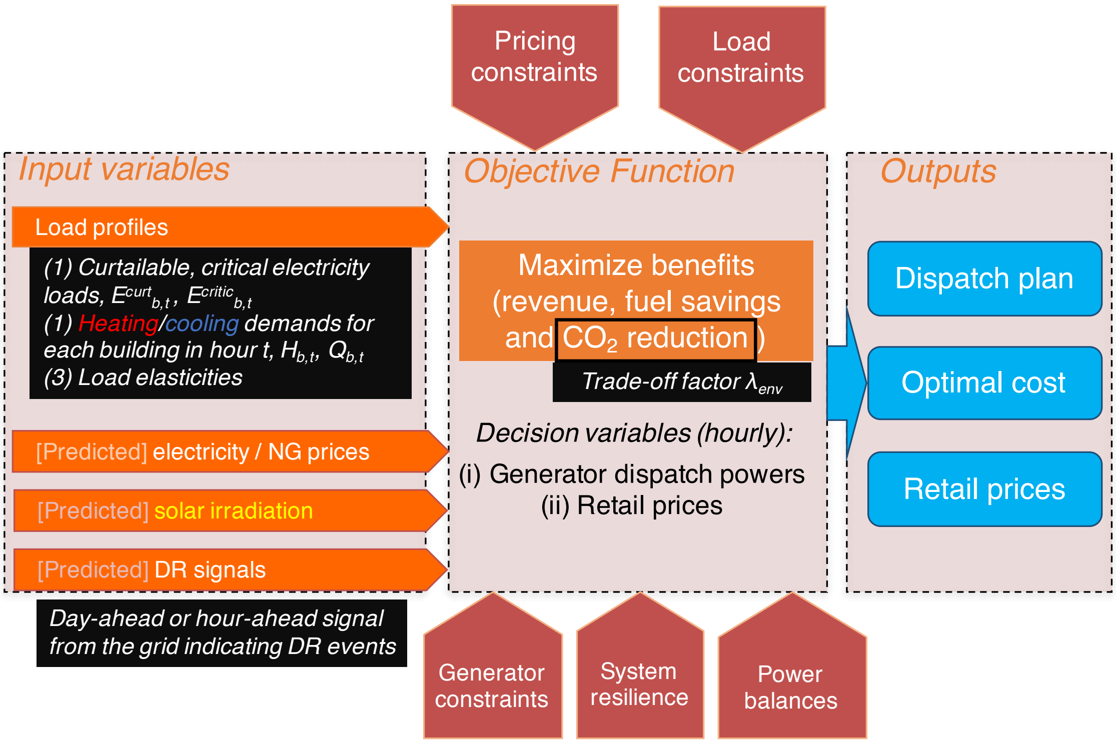

Pricing constraints
Load constraints
Input variables
Load profiles
(1) Curtailable, critical electricity loads, $E^{curt}_{b,t}$, $E^{critic}_{b,t}$
(1) Heating/cooling demands for each building in hour t, $H_{b,t}$, $Q_{b,t}$
(3) Load elasticities
[Predicted] electricity / NG prices
[Predicted] solar irradiation
[Predicted] DR signals
Day-ahead or hour-ahead signal from the grid indicating DR events
Objective Function
Maximize benefits (revenue, fuel savings and CO$_2$ reduction)
Trade-off factor λ_{env}
Decision variables (hourly):
(i) Generator dispatch powers
(ii) Retail prices
Outputs
Dispatch plan
Optimal cost
Retail prices
Generator constraints
System resilience
Power balances

Figure 6.3: Illustration of the optimization framework of MR-POD.

Objective function. Driven by both economic gains and environmental consciousness, the retailer tries to maximize its profit $f_t^{\text{Rev}}(\mathbf{d}_t, \boldsymbol{p}_t) - f_t^{\text{Ope}}(\mathbf{x}_t, \mathbf{z}_t, \boldsymbol{\xi}_t)$, and at the same time, minimize its environmental impact $f_t^{\text{Env}}(\mathbf{x}_t, \mathbf{z}_t)$. The revenue collected by selling energy to buildings $f_t^{\text{Rev}}(\mathbf{d}_t, \boldsymbol{p}_t)$ depends on retail prices $\boldsymbol{p}_t$ and building loads $\mathbf{d}_t$. This is a *quadratic function* of the retail prices $\boldsymbol{p}_t$, since the building demand $\mathbf{d}_t$ depends linearly on the price as in (6.6). The operational cost $f_t^{\text{Ope}}(\mathbf{x}_t, \mathbf{z}_t, \boldsymbol{\xi}_t)$ is the expenditure on fuel imports net of any revenue from energy sold back, in addition to maintenance expenses for facilities with on-site personnel. On top of the former commonly adopted economic incentives [40], [81], [97], [124], [176], the environmental impact $f_t^{\text{Env}}(\mathbf{x}_t, \mathbf{z}_t)$, measured by the amount of carbon dioxide (CO_2) emissions, is incorporated to encourage the use of renewable energy and natural gas in favor of grid electricity.[1] Through the parameter λ_{env} controlling trade-offs such as a carbon tax[2], MR-POD is able to offer guidance to balance the economic and environmental benefits for the retailer.

Constraints. There are two main groupings of constraints in MR-POD related to pricing and operation. The pricing constraints $\boldsymbol{p}_t \in \mathcal{P}_t$ ensure regulatory compliance, market competitiveness, and customer satisfaction. The operation constraints include (1) power balance between load and generation, $\mathbf{x}_t \in \mathcal{X}_t(\mathbf{z}_t, \mathbf{d}_t, \boldsymbol{\xi}_t)$, for *heating, cooling, and electricity*, (2) feasibility for dispatch variables $\mathbf{x}_t$ and storage states $\mathbf{z}_t$ delineated by the generation and storage technologies, e.g., CHP partial loads, PV output, and storage charge/discharge rate limits, (3) the building load identity $\mathbf{d}_t \in \mathcal{D}_t(\boldsymbol{p}_t)$, based on the price elasticity model, (4) system resilience requirements, as prescribed in either the cap on the total imported power from the grid [124], [164], [177], or the spinning reserve limits on the storage resources [164], [177], as well as (5) DR targets like peak load reduction, which can be achieved through energy price setting. Due to the involvement of integer variables like discrete on/off decisions for CHP and charging/discharging for storage, in addition to quadratic coupling between prices and building loads, the resulting problem requires MIQP.

MG energy pricing

The key to a sustainable pricing policy should align the incentives of the retailer, its customers, and its regulators, and ensure reliability, customer equity, and social welfare maximization [39], [168]. In the following, we introduce the guiding principles of day-ahead rate setting for electricity ($\boldsymbol{p}_t^E$), heating ($\boldsymbol{p}_t^H$), and cooling ($\boldsymbol{p}_t^C$) services (see Fig. 6.4 for an illustration).

Time-differentiated rate structure. While the DA prices of RTP can vary from hour to hour, TOU typically has three levels corresponding to off-, mid-, and on-peak hours, i.e.,

[1] Based on the statistics from the U.S. Energy Information Administration, electricity generated from coal (0.98kgCO_2/kWh) emits more carbon dioxide than that generated from natural gas (0.55kgCO_2/kWh). Since coal combustion accounts for 71% of CO_2 emissions of the grid electricity while natural gas only accounts for 28%, it is cleaner to generate electricity from natural gas than import from the grid in the U.S.

[2] For example, a carbon tax of \$0.026/kgCO_2 is levied in Denmark, while the tax is \$0.131/kgCO_2 in Sweden: "Where Carbon Is Taxed?", Carbon tax center, Jun, 2017 [Accessed: 12/1/2017].

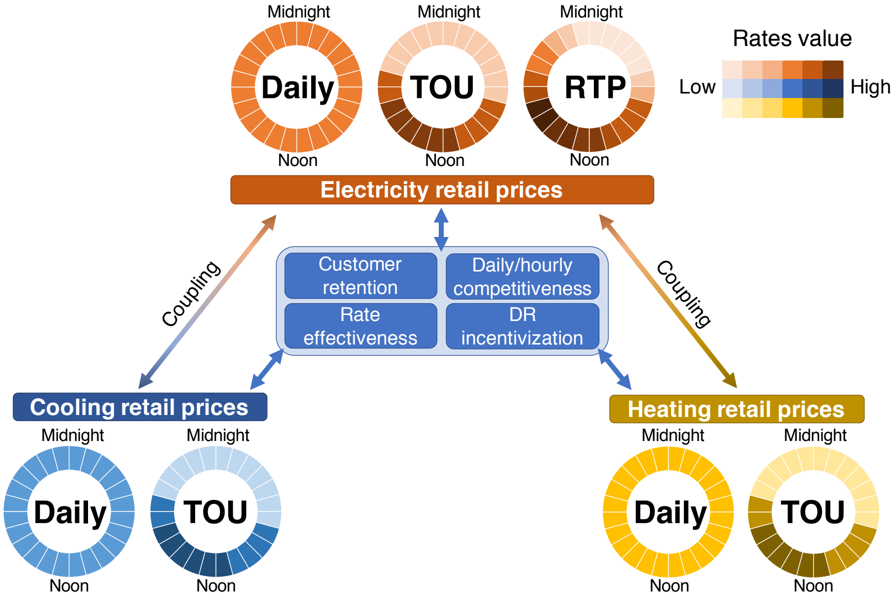


Figure 6.4: Overview of the pricing strategy, including the time-differentiated rates structure and energy price coupling. The strategy considers customer retention, price competitiveness, rate effectiveness, and DR incentivization for energy pricing.

$\boldsymbol{p}_{t_1}^E = \boldsymbol{p}_{t_2}^E$ if t_1, t_2 are in the same time group. To avoid "response fatigue" due to price variation [13], it is enforced that the hourly price change and the difference between average rates of off-, mid-, and on-peak are limited,

$$\left|\boldsymbol{p}_t^E - \boldsymbol{p}_{t+1}^E\right|, \left|\frac{1}{n_{\text{mid}}}\sum_{t\in\text{mid}} \boldsymbol{p}_t^E - \frac{1}{n_{\text{off}}}\sum_{t\in\text{off}} \boldsymbol{p}_t^E\right|, \left|\frac{1}{n_{\text{on}}}\sum_{t\in\text{on}} \boldsymbol{p}_t^E - \frac{1}{n_{\text{mid}}}\sum_{t\in\text{mid}} \boldsymbol{p}_t^E\right| \leq \delta^{\text{diff}}, \tag{6.1}$$

where $n_{\text{off}}, n_{\text{mid}}, n_{\text{on}}$ are the sizes of each group, and δ^{diff} is capped at 0.1\$/kWh for electricity.

Rate competitiveness. Both hourly and daily average limits are imposed on thermal and electricity rates:

$$r_t^{\min} \leq \boldsymbol{p}_t^E \leq r_t^{\max}, \quad r_{\text{avg}}^{\min} \leq \frac{1}{24}\sum_{t=1}^{24} \boldsymbol{p}_t^E \leq r_{\text{avg}}^{\max} \tag{6.2}$$

where typical values of $r_t^{\max}$ and $r_{\text{avg}}^{\min}$ are 0.3\$/kWh and 0.05\$/kWh, respectively, while $r_t^{\min}$ and $r_{\text{avg}}^{\max}$ can be chosen as the forecasted wholesale market tariff/the flat rate in the area to protect the retailer/customers. Further, to hedge consumers against high prices, the K-factor is introduced, K, as the upper bound on the ratio between the bills under the new rate $(\boldsymbol{p}^E, \boldsymbol{p}^H, \boldsymbol{p}^Q)$ and the flat rate $(p_{\text{flat}}^E, p_{\text{flat}}^H, p_{\text{flat}}^Q)$:

$$\sum_{t=1}^{24}\left(\boldsymbol{p}_t^E E_{t,b} + \boldsymbol{p}_t^H H_{t,b} + \boldsymbol{p}_t^Q Q_{t,b}\right) \leq K\sum_{t=1}^{24}\left(p_{\text{flat}}^E E_{t,b} + p_{\text{flat}}^H H_{t,b} + p_{\text{flat}}^Q Q_{t,b}\right), \forall b \in \{1, ..., B\} \tag{6.3}$$

where $E_{t,b}, H_{t,b}, Q_{t,b}$ are the electricity/heating/cooling loads of building b. Setting the K-factor less than one implies that the new rates will reduce customers' bills relative to the incumbent utility. But, this would often result in a loss of profits for the retailer. A K-factor slightly larger than one allows more flexibility, and can, interestingly, lead to a win-win situation in tandem with the DR mechanism.

Integrated energy coupling. The co-existence of multiple energy vectors in the system indicates the potential coupling between electricity and thermal loads. For a hypothetical consumer with a heat pump, competitive thermal rates could make it more cost-effective to purchase thermal energy from the retailer than self-generate(Fig. 6.4):

$$\boldsymbol{p}_t^H \leq \frac{\boldsymbol{p}_t^E}{\mathrm{COP}^H}, \quad \boldsymbol{p}_t^Q \leq \frac{\boldsymbol{p}_t^E}{\mathrm{COP}^Q} \tag{6.4}$$

where $\mathrm{COP}^H, \mathrm{COP}^Q$ are the coefficients of performance (COP) for heat pumps, which can be as high as 4 for some commercial brands.

DR effectiveness. It is desirable to shape the loads during DR events, such as for peak load reduction, which can be incorporated in rate optimization. The key is to differentiate the price elasticity of demands, as discussed in the following section.

It is worth mentioning that electricity rates often include the commodity costs, transmission/distribution infrastructure charges, and public purpose programs, such as energy efficiency and low-income subsidies, which can be either fixed or variable [168]. Demand charges are also sometimes applied on maximum demand over a certain time. This study focuses on variable operational costs that arise from generation and fuel imports, though it could be combined with other fixed charges in practice.

Energy demand and supply

The effectiveness of price setting depends on the price sensitivity of energy demands. The load profiles of buildings in a MG, such as in residential and commercial buildings, hospitals, and public services, can be characterized as critical or curtailable loads.

Critical load. For electricity usage in data centers and ICUs of hospitals, for example, it is of utmost importance that critical loads are satisfied, i.e.,

$$E_{t,b}^{\text{critic}} = E_{t,b}^{\text{critic}} \tag{6.5}$$

where $b \in \{1, ..., B\}$ for a building within the community, and t denotes an hourly time step.

Curtailable load. Apart from critical loads, demands like heating, cooling, ventilation, and lighting usually fall as the energy price increases. A consumer's sensitivity to price changes is measured by the coefficient of elasticity, ϵ, which indicates a ϵ% change in energy demands due to a 1% change in price. The curtailable load, therefore, is modeled as:

$$E_{t,b}^{\text{curt}} = E_{t,b}^{\text{curt,ref}} \left(1 + \epsilon_{t,b} \underbrace{\frac{p_t^{\text{E}} - p_t^{\text{E,ref}} + \beta_t^{\text{DR}}}{p_t^{\text{E,ref}}}}_{\text{\% change in price}} \right) \tag{6.6}$$

where $\epsilon_{t,b}$ is the elasticity coefficient for building b at time t, $E_{t,b}^{\mathrm{curt,ref}}$ is the curtailable load under the price $p_t^{\mathrm{E,ref}}$, which usually corresponds to historical data [64], [128], [165].

The elasticity coefficient $\epsilon_{t,b}$ is typically negative, indicating the reciprocal relationship between demand and price; its value depends on (1) time of the day: the load is usually more price responsive during on-peak than off-peak hours [60], (2) rate structures: it is found that loads under TOU rates are less elastic than those under RTP rates [64], and (3) planning horizon: the elasticity is usually greater in the long-run when customers can react to a price increase by purchasing more energy efficient appliances [20], [64]. For instance, the elasticity of electricity demands for residential buildings in the US ranges from -0.20 to -0.35 in the short-run, and -0.30 to -0.80 in the long-run [8]. Differentiated from more complex non-linear models based on logarithm or potential [239], the linear model simplifies the optimization and is also more accurate and reliable [239]. We focus our attention on *own-price elasticity*, which limits the influence of price on demands in the same time period, since it is sufficient for capturing how customers adjust their consumption to price changes [62].

As for the supply side, we consider an integrated energy system to satisfy the buildings' electric and thermal loads. By exploiting synergies and complementarities of various energy vectors, this approach can improve energy efficiency, reduce CO_2 emissions, and facilitate renewable integration [233]. Apart from CHP and conventional thermal generators like electric/natural gas/absorption chillers/boilers and heat pumps, renewable resources like solar thermal and photovoltaics (PV) are included in the retailer's facility to harness solar energy and reduce carbon footprints. Electric and thermal storage with dynamic charging/discharging behaviors are available to enable smooth operation and exploit time-shifting opportunities. Maintaining a minimum amount of stored energy, typically 5% of the total capacity, i.e., state-of-charge (SOC), is referred to here as the *spinning reserve requirement* [164], [177]. Modeling details can be found in [97].

6.3 Optimal rate design and operation strategy

This section introduces MG planning under uncertain market and weather conditions, as well as the DR incentivization scheme.

Planning under uncertainty

Using MR-POD for strategizing, the operator can optimize the energy dispatch and retailing in five critical steps, as illustrated in Fig. 6.5: Before the actual day of dispatch (*day 0*), data related to weather, energy demands, and MG status are acquired from installed sensors and meters (*step 1*); this is used to predict and estimate key quantities such as DR potentials, renewable energy, and electricity wholesale tariffs (*step 2*). Based on the prediction, MR-POD produces the optimal dispatch plan and retail rates (*step 3*), which are announced to generation facilities and consumers (*step 4*). The plan is executed on the actual day of

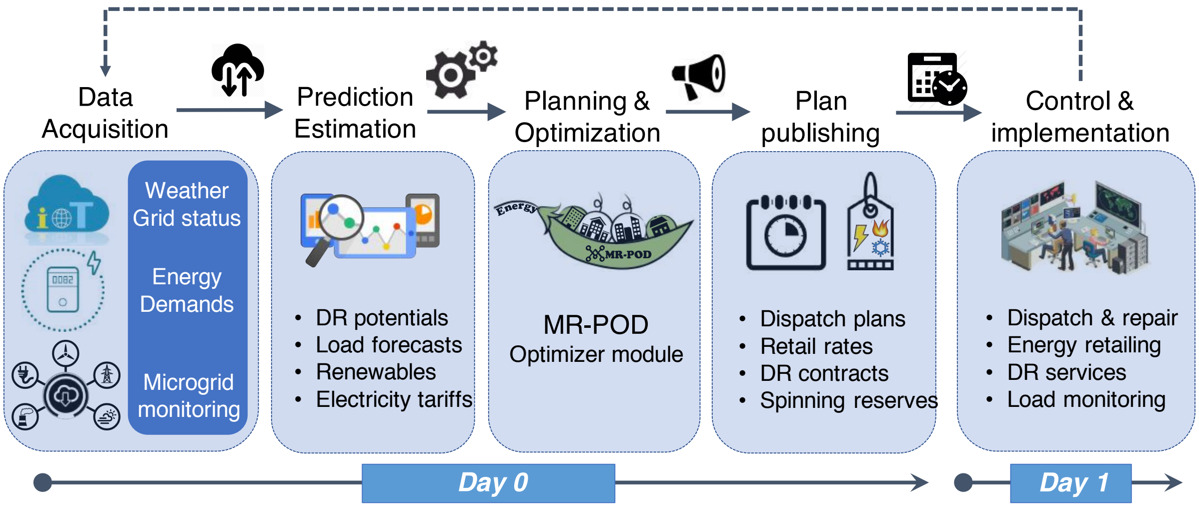


Figure 6.5: System overview of MR-POD, illustrating key components: data acquisition, estimation and prediction, planning and optimization, and control and actuation.

dispatch (*day 1*), and repaired to adjust to unaccounted for fluctuations in demand and renewable generation (*step 5*).

Prediction of uncertain variables. Methods for solar and load forecasting can be grouped into data-driven or model-based methods [45]. A comprehensive review of price prediction approaches has been recently conducted [230]. Specifically, we employ the "forecast combination" method based on ordinary least squares (OLS[c]), which combines M forecasts from a committee of predictors, $\hat{y}_{m,t}$, according to

$$y_t^{\text{OLS}} = c_{\text{OLS}} + \sum_{m=1}^{M} w_m \hat{y}_{m,t} \tag{6.7}$$

where constant c_{OLS} and weights $\{w_m\}_{m=1}^{M}$ are learned from past performance of the forecasts [230]. Its performance is shown to be superior among an array of candidates for solar and tariff prediction [97].

Generator dispatch and energy retailing. As with electricity market bidding, upon receiving predictions on *day 0*, the retailer performs MR-POD optimization to prepare a day-ahead dispatch plan and its energy retail rates and announces them to the generation facility and building owners. The original dispatch proposal is amended for actual execution on *day 1* by exploiting the cheapest sources/destinations of energy immediately available, e.g., storage (if any) or grid, to maintain the power balance.

Setting the DA retail rates is common practice, such as the DA RTP tariff used by the Illinois Power Company, pilots in California, Idaho, and New Jersey, and the three-level TOU pricing in Ontario, Canada. And it reaps several benefits [13]. First, the DA prices like RTP can best reflect the costs of energy procurement incurred by the retailer. Also, it can handle exceptional days, for instance, by declaring DA CPP when the forecasted loads are high. Importantly, it allows consumers sufficient time to schedule their consumption, while not being "fatigued" by hourly rate changes [123].

DR incentivization

Time-differentiated rates bring about changes in customers' energy consumption by differentiating prices during peak and off-peak hours. The targeted change patterns, or load shaping, are described by:

$$a^t_{\min} E^{\text{curt,ref}}_{t,b} \leq E^{\text{curt}}_{t,b} \leq a^t_{\max} E^{\text{curt,ref}}_{t,b} \tag{6.8}$$

where $a^t_{\min}$ and $a^t_{\max}$ are design parameters indicating the ranges of actual loads when the retail rates are in effect; for instance, normal load conditions typically correspond to $a^t_{\min} = 0.85$ and $a^t_{\max} = 1.1$ [53], while load reduction requires $a^t_{\min} < a^t_{\max} < 1$. Occasionally, in response to unusual events, the retailer can employ additional incentive/penalty terms, β^{DR}_t, in tandem with the regular retail rates to induce further changes in loads, as predicted by the curtailable load model (6.6).

While the success of DR relies on customer engagement, in practice, interest in switching to RTP rates wane due to a lack of financial incentives and increased exposure to market volatility [127]. One viable strategy is to motivate DR participation by offering guarantees of energy bill reduction. This can be achieved by dictating the "K factor" to be less than one when setting the rates (MR-POD); however, experiments show that this strategy often yields inefficient pricing, and even leads to a significant loss of profits for the retailer.

Our proposal (Fig. 6.6) allows an initial increase in customer energy payments, but later compensates the customers with performance-based dividends, which serve several purposes: 1) alignment of the financial interests of stakeholders; 2) incentivization of DR; 3) protection of customers, e.g., low-income families, by reducing their bills.

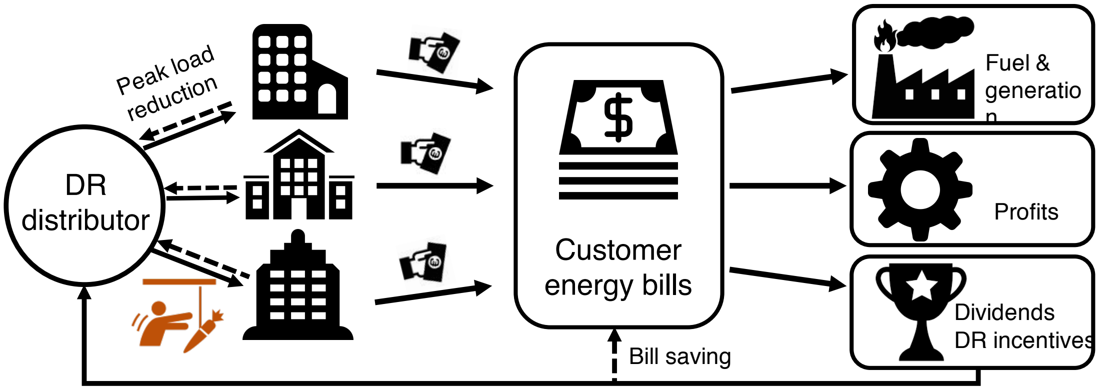


Figure 6.6: The mechanism of DR incentivization with performance-based dividends, which uses a portion of the retailer's profits as rewards to buildings based on their peak load reduction performance.

The performance-based dividends are calculated relative to a baseline, which is usually the flat rate pricing. First, assuming baseline loads, any increase in energy bills due to RTP is compensated. This ensures non-increasing bills for customers who opt for RTP over flat rates. Second, a share of retailer's total fuel cost savings is distributed among DR participants. The amount that each building receives is proportional to its contribution to total peak load reduction of the community, though it is possible to factor customer type and income levels into the distribution weights.[3] As ancillary services are usually scheduled by the ISO a day ahead and called upon as needed on short notice, the scheme is able to introduce added flexibility to MG load responses, thus improving services to the grid [127].

6.4 Scenario analysis and case study

This section studies the impact of optimal dispatch and pricing on system economy and reliability. First, the scenario without DR is examined with fixed retail rates. The DR option is enabled by jointly optimizing rates and dispatch.

Experimental setup

We first present the data for solar irradiation, building loads, and energy prices. We also specify six (6) campus scale MGs with different generators to serve three buildings with electricity, cooling and heating.

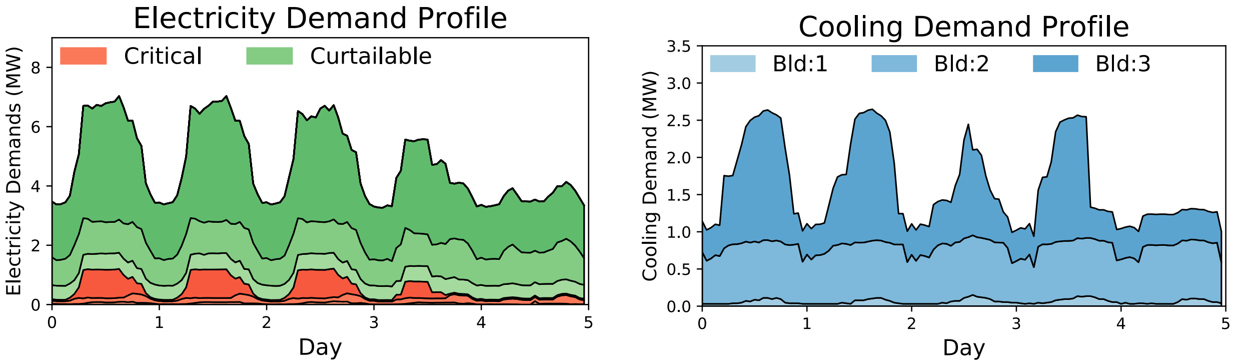


Figure 6.7: Electricity load profiles (left), which display the critical loads (red) and curtailable loads (green), for three buildings (different shadings). Cooling demands (right) for the buildings in a stacked plot.

[3] For instance, if the bill for a customer enrolled in RTP is \$92 but would be \$90 under the flat rate, then he would be compensated by \$2 to bring down the bill. If, in addition, the building contributes 50kW out of 1000kW of total peak load reduction of the community, and the cost savings of the retailer is \$200, then, with a sharing rates of 0.5, an additional \$5 rebate will apply, leading to a reduced bill of \$85.

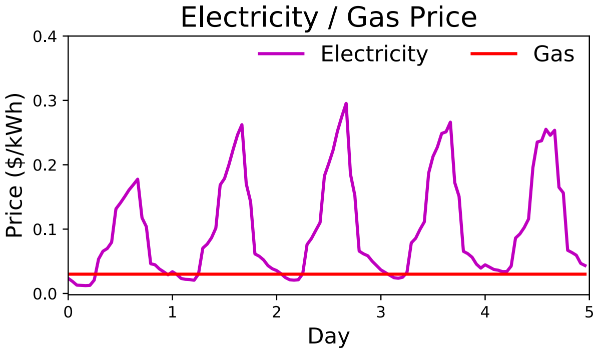


Figure 6.8: Electricity and natural gas tariffs, where the spark spread is mainly driven by the daily fluctuation of electricity prices. Data sources: see footnotes 7 and 8.

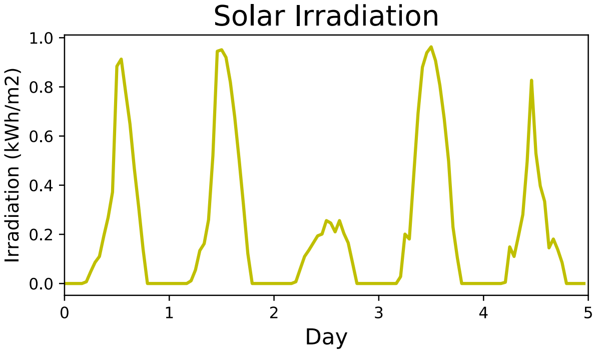


Figure 6.9: Solar irradiation measured by the GHI index (kWh/m^2) on several days of the study period, which clearly exhibits diurnal patterns.

Table 6.1: Building elasticity parameters for off-peak hours (12am-7am, 7pm-12am), mid-peak hours (7am-11am, 5pm-7pm), and on-peak hours (11am-5pm) in the summer period, where cooling loads are dominant.

	off-peak			mid-peak			on-peak		
	elec.	heat	cool	elec.	heat	cool	elec.	heat	cool
Bld:1	-.1	-.2	-.2	-.3	-.3	-.3	-.46	-.4	-.4
Bld:2	-.12	-.22	-.2	-.32	-.35	-.3	-.48	-.45	-.4
Bld:3	-.15	-.24	-.2	-.34	-.4	-.3	-.5	-.5	-.4

Dataset. The TMY3 dataset [231] is queried for the Global Horizontal Irradiance (GHI) index [4] in Oakland, California (Fig. 6.9) to determine PV outputs.

The load data is retrieved from the Open Energy Information (OpenEI) for a research facility (Bld:1) [5], a large hotel (Bld:2), and a commercial building (Bld:3) [6]. During the period of study, i.e., May, the thermal loads are predominantly for cooling (Fig. 6.7). The elasticity parameters (Table 6.1), prudently derived from [8], [53], [60], differentiated responses in off-/mid-/on-peak hours and building types.

The electricity spot price is accessed from the National Grid Online Database[7] and adapted to be similar to the California wholesale market and to reflect the time of use rates (Fig. 6.8). The natural gas price, which according to the U.S. Energy Information

[4] GHI, measured in 1 kWh/m^2, is the total amount of direct and diffuse solar radiation received on a horizontal surface during the 60-minute period.

[5] NREL RSF Measured Data 2011, accessed: 12/2017

[6] OpenEI Load Profiles, accessed: 12/2017

[7] National Grid Online Database, accessed: 12/2017

Administration experiences less fluctuations throughout the month, is assumed to be at a constant level of 0.03$/kWh.[8]

MG specification. We have prototyped six (6) MGs with different generation capacities (Table 6.2). MG1 is considered as the baseline, which imports electricity from the grid and provides heating and cooling energy by a NG boiler and an electric chiller. The aim of the rest of the prototypes is to study the effects of energy storage (MG2 vs. MG1), renewables (MG3 vs. MG1), CHP and absorption chiller (MG5 vs. MG4), and grid imports (MG6 vs. MG5) on operations.

The core MIQP programs (MR-POD) are built in Python and solved by Gurobi. The following experiments are performed on a MacBook with a 2.8 GHz Intel Core i7 CPU and 16 GB RAM memory.

Table 6.2: MG specifications. The storage capacities follow the format of heating storage/cooling storage/electric battery. Four discrete CHP plants are considered. The modeling and specifications of generator technologies can be found in [97]. For those MGs with grid imports, they can also function as islands.

	NG boiler	**Electric chiller**	**Storage**	**PV**	**Solar thermal**	**Absorption chiller**	**CHP**	**Grid import**
MG1	5MW	10MW						Yes
MG2	5MW	10MW	1/1/4MW					Yes
MG3	5MW	10MW		1.5MW	.75MW			Yes
MG4	5MW	10MW	1/1/4MW	1.5MW	.75MW			Yes
MG5	5MW	10MW	1/1/4MW	1.5MW	.75MW	10MW	1.5/2/3/4MW	Yes
MG6	5MW	10MW	1/1/4MW	1.5MW	.75MW	10MW	1.5/2/3/4MW	No

Energy dispatch and uncertainty effect

This section demonstrates the optimal energy dispatch planning of MR-POD while keeping retail prices fixed. Several observations can be made about the energy dispatch plan (Fig. 6.10) for MG4, which includes CHP, storage, and PVs: 1) the predicted spot price follows the trend of the true spot price[9]; as a result, 2) the battery takes advantage of its variation by charging during the night (off-peak) and discharging during the afternoon (on-peak); also, 3) CHP and the absorption chiller are dispatched for electricity and cooling generation to exploit the spark spread.

By comparison (Table 6.4, "Daily" columns), given the same revenue from customer bills, MG1–or the baseline–earns the least profit, whereas MG5 brings in the most profit, which

[8] U.S. Energy Information Administration, accessed: 12/2017

[9] To reduce uncertainty in the spot market and solar irradiation, an OLS forecast combination scheme based on an array of forecasters (Gaussian process, support vector regression, multi-layer perceptron, etc.) is employed, which use a month of data for training and to make day-ahead predictions.

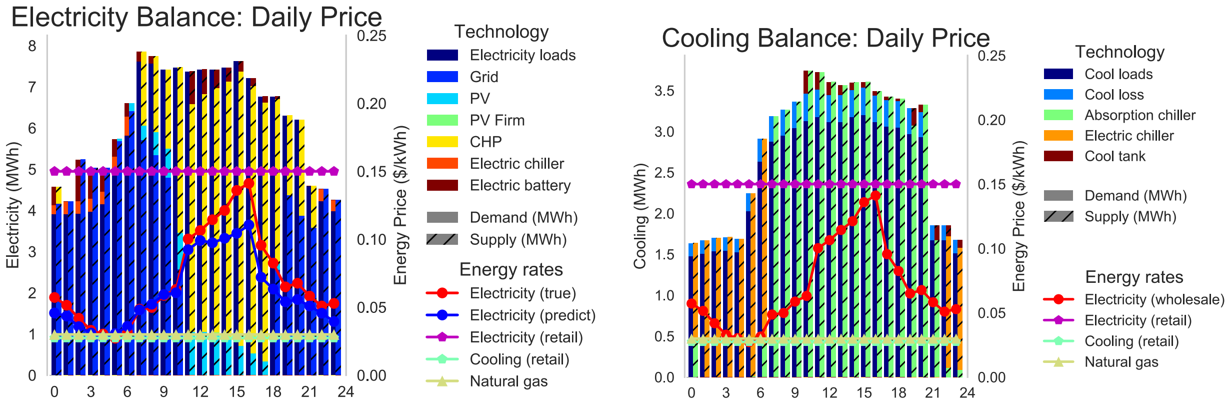


Figure 6.10: Electricity and cooling balances with daily flat rates. The graph also shows the forecasted and true wholesale price, as well as the natural gas rates. Since the experiment is conducted during the summer, the heat balance is not shown due to insignificant loads.

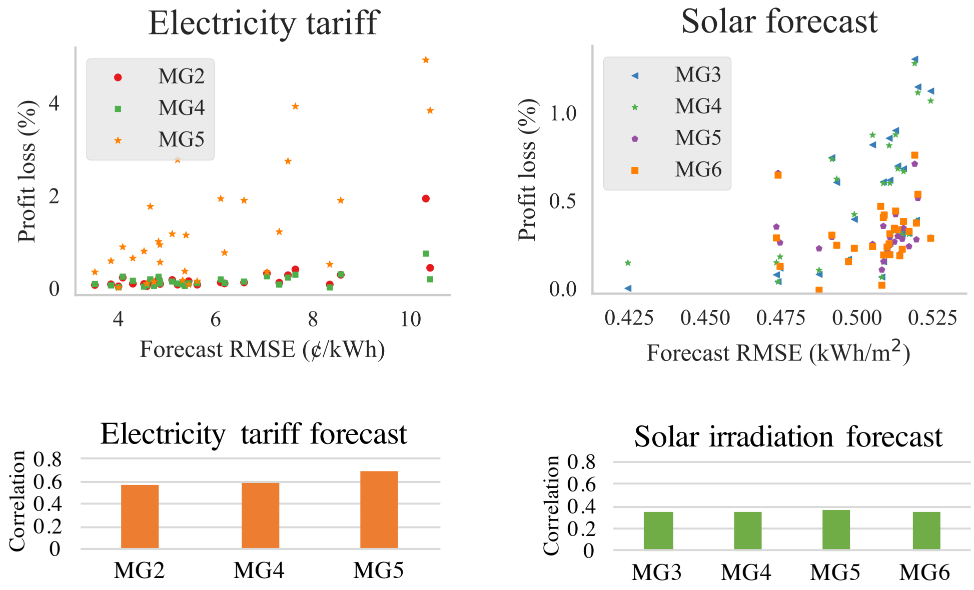


Figure 6.11: Top panel: scatter plots of the profit loss against electricity tariff (left) and solar (right) forecast error. The baseline is an oracle that uses true electricity tariff and solar irradiation for dispatch and pricing. Bottom panel: Pearson correlation between profit loss and forecast error. A positive number closer to 1 occurs when the two random variables follow similar trend.

exceeds MG6 that operates in "island-mode." This illustrates the energy cost reduction offered by storage, renewables and CHP.

The uncertainty effect of solar and electricity prices is studied by collating the daily profit loss with the forecasting error[10] (Fig. 6.11). We can see that there is a positive correlation between profit loss and forecasting error. Since the dispatch of CHP relies on accurate prediction of the spark spread, the effect of wholesale spot price forecasting error is more pronounced for MG5 than both MG2 and MG4. This result is in alignment with the findings from [41]; however, their studies incorporated the situation with only electricity loads and no distributed generation capacity, and the forecasting errors were simulated from a noise model rather than derived for state-of-the-art predictors.

Optimal retail pricing strategies

The central question in this section is: "How can the retailer strategize its operation and retailing to promote mutual benefits for its customers and the grid."

Firstly, we investigate the benefits of time-differentiated rate structures with elastic building loads (Table 6.1) and practical-oriented pricing constraints (Table 6.3). The DA electricity and thermal rates are evaluated over a month during the summer, as illustrated in Fig. 6.12 for MG4 and Fig. 6.13 for MG5 (which differ by the installation of CHP plants), where the monthly average and 90% confidence interval of the retail prices and true/predicted spot prices are shown. While the optimal RTP and TOU rates share similar trends, RTP exhibits more flexibility for accommodating hourly fluctuations in loads and spot prices. Prices are relatively stable over the month, which reduces customers' risks of exposure to the wholesale market volatility. One crucial difference between the rate profiles of MG1 to MG4 and that of MG5 and MG6 is focused on the peak hours (see Figs. 6.12 and 6.13). For MGs that rely on grid imports for electricity provision, the retail price *peaks along with the spot price* to reflect the increased cost of generation, while this increase in rates is absent for MGs that can use natural gas as an alternative source. Indeed, as is shown in the previous section, CHP is dispatched when the grid electricity is expensive (Fig. 6.10).

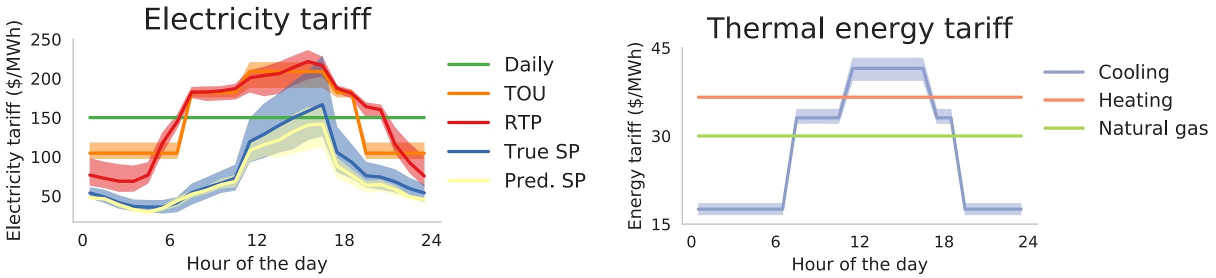


Figure 6.12: Optimized electricity (left) and thermal (right) retail rates under different pricing structures (Daily, TOU, RTP) for MG4. The shading indicates 90% confidence interval. Both the predicted and true wholesale electricity tariffs are shown.

[10] The forecasting error is measured by the root mean squared error (RMSE), given by $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$, with y_i and $\hat{y}_i$ denoting the true and predicted values at time $i \in \{1, ..., n\}$.

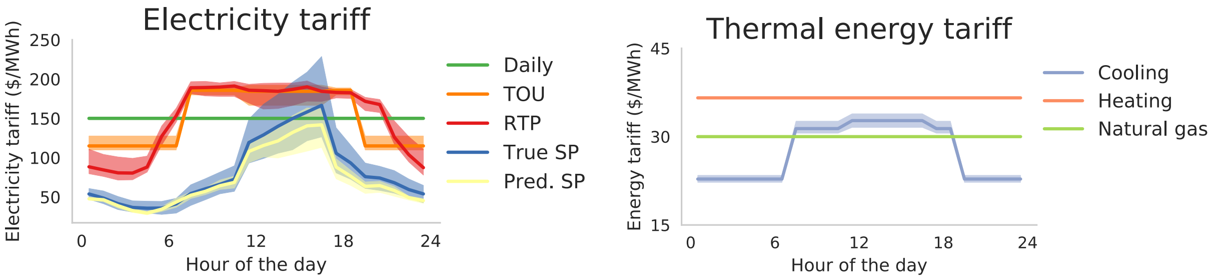


Figure 6.13: Optimized electricity (left) and thermal (right) retail rates under different pricing structures for MG5.

Table 6.3: Parameters of optimal rate design. Each parameter category is followed by the equation reference. For hourly rates limits, $\hat{y}_t^E$ is the predicted wholesale tariff at hour t. The unit for rates-related quantities is $/kWh.

	Electricity	**Thermal**
Hourly change cap (6.1)	$\delta_E^{\text{diff}} = 0.2$	$\delta_{H,Q}^{\text{diff}} = 0.1$
Hourly rates limits (6.2)	$r_t^{\max} = 0.3$, $r_t^{\min} = \min(0.05, \hat{y}_t^E)$	$r_t^{\max} = 0.05$, $r_t^{\min} = 0.01$
Daily rates limits (6.2)	$r_{\text{avg}}^{\max} = 0.15$, $r_{\text{avg}}^{\min} = 0.05$	$r_t^{\max} = 0.15/\text{COP}$, $r_t^{\min} = 0.01$
K factor (6.3)	$K = 1.2$ (unless otherwise specified)	
Energy coupling (6.4)	COP $= 3.0$ for both heating and cooling	
Reference rates (6.6)	$p_t^{\text{E,ref}} = 0.15$	$p_t^{\text{H,ref}} = 0.036$, $p_t^{\text{Q,ref}} = 0.028$
DR requirements (6.8)	$a_{\min}^t = 0.85$, $a_{\max}^t = 1.1$	
TOU groupings	off-peak: 7pm-7am, mid-peak: 7am-11am, 5pm-7pm, on-peak: 11am-5pm	
DR dividends	customer share 50% of retailer profits	

The economic and environmental impacts are assessed (Fig. 6.14 and Table 6.4), illustrating increased daily profits and reduced total energy and CO$_2$ emission.[11] To gain insights into the impact on MG-level efficiency, we study the measures of peak electricity usage, peak-to-valley distance and load factors, which indicate the average peak loads (11am – 5pm), the difference between peak loads and valley loads (7pm – 7am), and the ratio between the average loads and peak loads, respectively. The RTP scheme is shown to significantly bring down peak loads and peak-to-valley distance while raising the load factors, which lessens the burden of the MG to invest in peak capacity and improves resource management and system reliability. Above all, RTP is shown to improve the economics more significantly over the daily rates when CHP is not present, due to the substantial reduction in peak hour loads

[11] The profit is calculated as the revenue minus the fuel cost, e.g., electricity from the grid or natural gas, which also include the dividends for the buildings due to DR. The total energy consumption includes daily electricity and thermal energy demands. The CO$_2$ emission is estimated from the use of grid electricity (0.98kgCO$_2$/kWh) and natural gas (0.55kgCO$_2$/kWh).

Table 6.4: Scenario analysis result summary. The reported daily values for the cost of generation, profits, and CO_2 emissions are averaged over 30 days period. Compared to the baseline model that uses flat daily retail rates, the percentage differences are shown in the parenthesis. Graphical illustrations for other indicators, such as peak electricity and load factors, are shown in Fig. 6.14.

	Cost of generation (k$)			Profits (k$)			CO_2 emissions (ton)		
	Daily	TOU	RTP	Daily	TOU	RTP	Daily	TOU	RTP
MG1	11.9	11.0(-7.6%)	10.9(-8.4%)	8.4	9.3(+10.3%)	9.4(+11.9%)	134	133(-0.7%)	129(-3.7%)
MG2	11.7	10.8(-7.7%)	10.7(-8.5%)	8.6	9.5(+10.5%)	9.6(+11.6%)	135	134(-0.7%)	130(-3.7%)
MG3	11.0	10.1(-8.2%)	10.0(-9.1%)	9.3	10.2(+9.7%)	10.3(+10.8%)	125	124(-0.8%)	120(-4.0%)
MG4	10.8	9.9(-8.3%)	9.8(-9.3%)	9.6	10.4(+8.3%)	10.6(+10.4%)	127	124(-2.4%)	121(-4.7%)
MG5	6.8	6.6(-2.9%)	6.5(-4.4%)	13.5	13.7(+1.5%)	13.9(+3.0%)	128	127(-0.8%)	123(-3.9%)
MG6	7.3	7.3(0%)	7.1(-2.7%)	13.0	13.1(+0.8%)	13.2(+1.5%)	133	133(0%)	128(-3.8%)

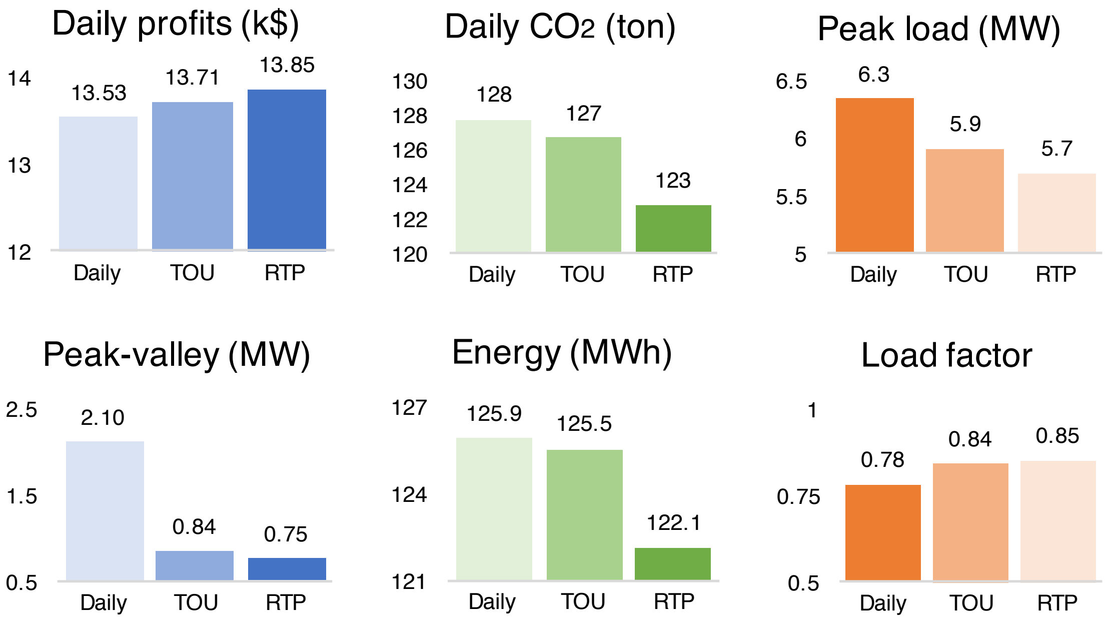


Figure 6.14: Comparison of different dynamic rate structures (Daily, TOU, RTP) for MGs, based on the economic (daily profits), environmental (CO_2 emission, total energy), and reliability (peak electricity, peak-valley distance, load factors) indicators.

that lowers the cost of generation (see Table 6.4 for MG1 – MG4).

Next, we evaluate the performance-based dividend strategy to promote customers' participation in RTP and demand response. Three rate settings (K factors 0.95, 1.0, 1.2) are considered. Fig. 6.15 illustrates the percentages of customer bill savings, energy production cost saving, and retailer profit increase for MG4 before (denoted as A) and after (B) the dividend. Customers achieve the most significant bill saving under the price setting with K-factor of 0.95; however, the conservative pricing does not induce peak load shedding in order to reduce the retailer generation cost, causing a considerable loss of profits. On the contrary, by allowing more flexibility in pricing (K factor of 1.2), the time-differentiated rates become more effective to reduce peak loads (Fig. 6.16), whose benefits can be shared among buildings (1 to 5% bill saving) and the retailer (3 to 6% profit increase) through the dividend mechanism. Since most RTP programs in the U.S. are voluntary [13], this offers economic incentives for enrollment. From the above results, customers with more elastic demands (Bld:2 and Bld:3) are more likely to save, since they tend to reduce more usage when the price is high. To assess the effects of energy load elasticity, four types of profiles are examined, namely, "very rigid", "rigid", "elastic", and "very elastic", which correspond to -100, -50, 0, 100% changes of elasticity parameters in Table 6.1 for all buildings.[12] There seems to be a positive correlation between the elasticity of customers and energy bill savings, retailer profit increase, and peak load reductions (Fig. 6.17), indicating the potential benefits of programs like openADR that aim at improving responsiveness to price through building automation [71], [160].

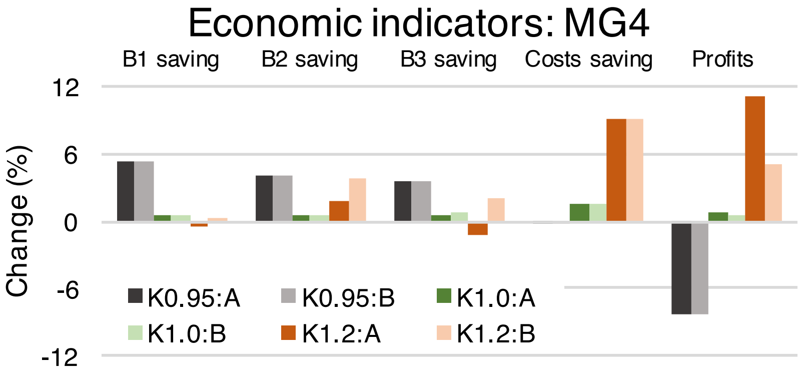


Figure 6.15: Economic indicators of building bill savings, cost savings, and profits increase (percentage) for different K factors (0.95, 1.0, 1.2). Scheme A and B represent the indicators before and after the performance-based dividends are rewarded to each building (Fig. 6.6). With K factor of 0.95, while buildings can enjoy substantial bill savings, the retailer incurs a profit loss of -8%. By introducing more flexibility in rate setting, e.g., K factors of 1.2, both consumer bill savings and retailer profits will improve after the performance-based dividends.

[12] For instance, the electricity elasticity for B1 during off-peak hours would be $-0.1 * (1 - 0.5) = -0.05$ for a "rigid" profile, and $-0.1 * (1 + 1) = -0.2$ for a "elastic" profile.

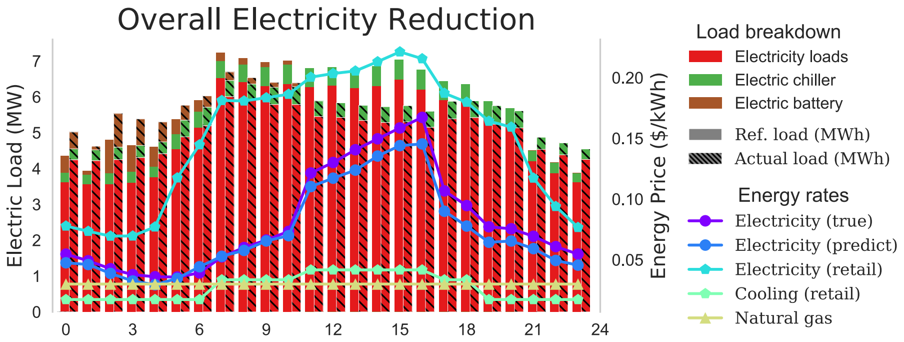


Figure 6.16: Overall electricity reduction with RTP rates. During peak hours, the original thermal and electricity loads are reduced (shaded bars) due to the high rates, while some of the loads are shifted to off-peak hours.

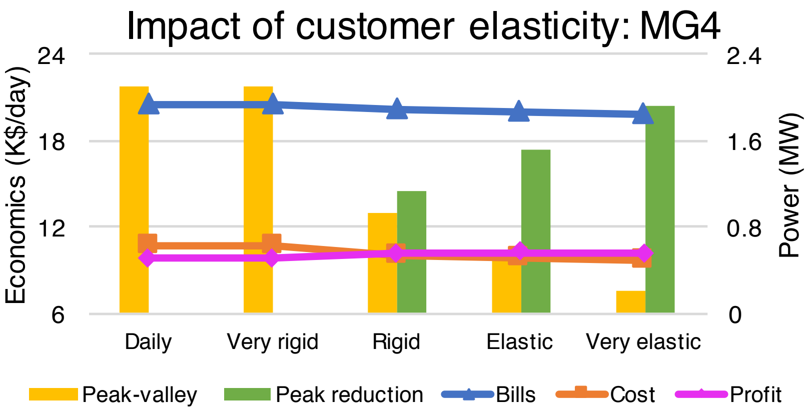


Figure 6.17: The economic and system indicators for four different customer profiles (elastic, baseline: elasticity in Table 6.1, very elastic: elasticity is 2 times the baseline, very rigid: elasticity is 0, rigid: elasticity is 50% of baseline). The performance of the system with elastic demands under daily rates is identical to that with very rigid consumers under RTP. Both indicators are improved with the customers being more elastic.

Microgrid case study

Due to the increasing penetration of renewables and heightened environmental awareness, it is crucial to ensure economic and environmental viability and system stability. This section demonstrates the capability of MR-POD in addressing the following issues:

- *Case 1:* Environmentally aware pricing and operation
- *Case 2:* Demand response for PV over-generation

The operation of a clean MG that aims to reduce the environmental impact, such as that of greenhouse gas emissions, is often pursued as a positive externality for society. According to a recent report by the World Bank, about 40 national jurisdictions worldwide put a price on carbon, a.k.a. carbon tax, which spans from less than 1\$/tCO$_2$e to 131\$/tCO$_2$e.[13] *Case 1* focuses on the design of environmentally aware pricing and operation strategies. More specifically, the cost of CO$_2$ emission can be considered by setting the λ_{Env} parameter in the optimization (MR-POD), which acts as a "virtual carbon tax." The tradeoff between profits and carbon dioxide emission is demonstrated for different MG infrastructures (Fig. 6.18), which illustrates the *Pareto frontier* in a multi-objective optimization.

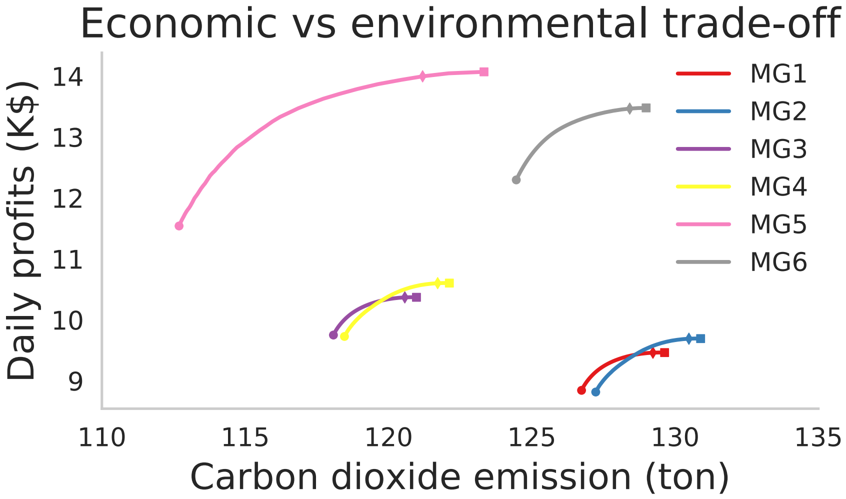


Figure 6.18: The trade-off between daily profits and CO$_2$ emission in MG operations and pricing. The square, diamond, and circle markers indicate λ_{env} being 0, 40, and 1000\$/tCO$_2$e. Clearly, MG5 is at the Pareto frontier, which can achieve more profits with less emissions due to the capability of fuel switching.

The results indicate that there is a limited range of trade-off for MGs with a single fuel source (MG1, 2, 3, 4, 6) that can only control through the price signal, as compared to MG5 that can also perform fuel switching. At a reasonable level of carbon taxes, or 40\$/tCO$_2$e, MG5 can substantially reduce CO$_2$ emissions while maintaining a high profit. As can be seen in Fig. 6.19, the use of an electric chiller and grid electricity is replaced by the absorption chiller and CHP at hours 11pm-2am, except during hours when the grid electricity price is relatively low to save generation cost. Indeed, the proportion of natural gas consumption significantly rises for environmentally aware operations during mid- and on-peak hours as the spark spread widens (Fig. 6.20).

By leveraging the natural gas fired, electricity powered devices, and renewable sources within a MG, it is possible to perform fuel switching as circumstances dictate. In particular,

[13] "State and trends of carbon pricing 2016", World Bank report, 2016 [Accessed: 12/1/2017]

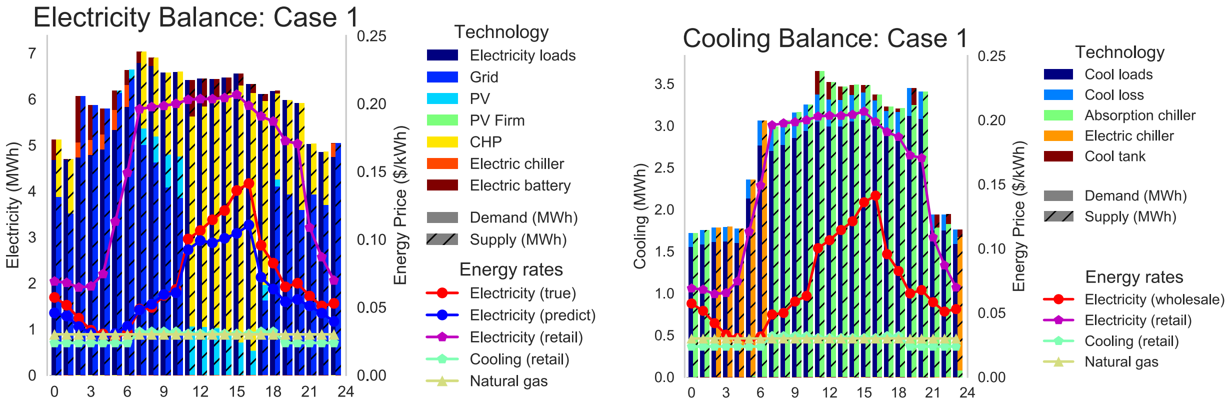


Figure 6.19: Electricity and cooling balances with a reasonable level of carbon taxes at 40\$/tCO$_2$e. For comparison, the plot is presented for the same day as in Fig. 6.10, which adopts a flat rate but does not include carbon tax equivalence in its operation.

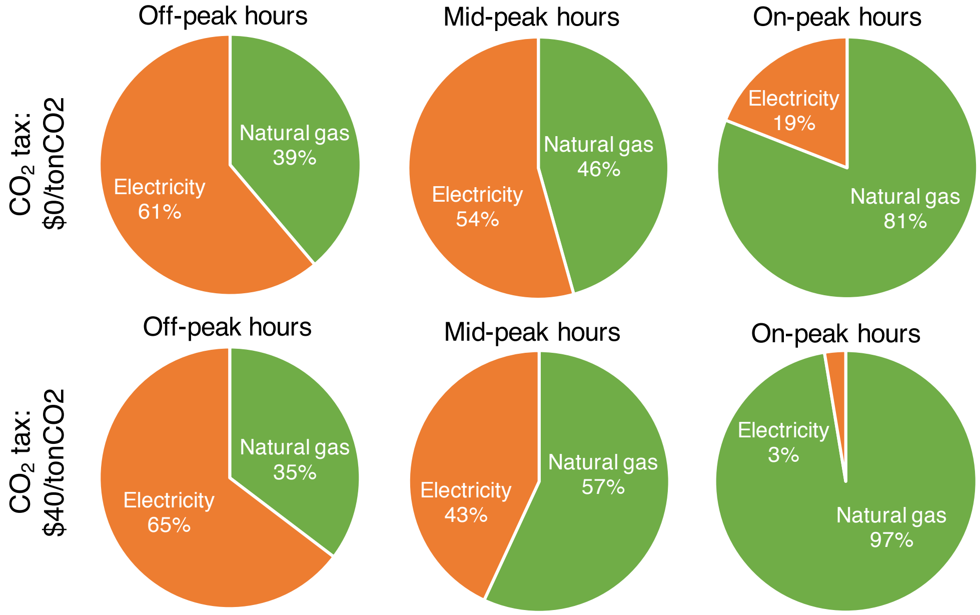


Figure 6.20: Fuel mixing during off-, mid-, on-peak hours for a schemes with λ_{env} being 0 and 40\$/tCO$_2$e. The latter results in more natural gas usage during mid- and on-peak hours for clean operations. However the usage of natural gas does not change significantly due to off-peak hours, due to the lower price of grid electricity.

Case 2 focuses on the problem of curtailed electric energy [144], when some of the renewable energy generation must be wasted to keep real-time power balance.

To simulate the case of PV over-generation, MG5 is assumed to have a high level of renewable generation (solar panels with 15MW rated capacity). Consequently, the problem often arises during a sunny day, when the supply of electricity far exceeds the demand. However, due to the prediction of the event, the retailer can promptly respond by lowering the electricity rates to encourage consumption, in addition to coordinating the charging of battery to shift the excessive generation to the night, which avoids the destabilization of the system and reduces customer bills (Fig. 6.21). In light of the upward tendency of renewable adoptions, this illustrates the added flexibility of MG enabled by optimal coordination and retail rates setting.

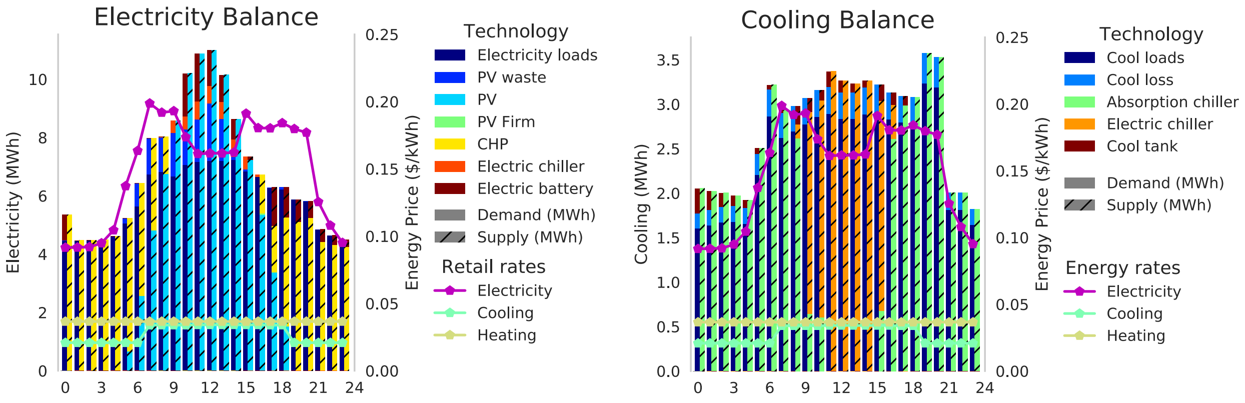


Figure 6.21: Electricity and cooling balances under RTP. When there is a PV surplus during the noon, the rates are set lower to encourage flexible consumption while the storage is charged, which reduces the amount of PV curtailment.

6.5 Chapter summary

With the increasing penetration of renewables and the advent of electric vehicles as mobile batteries, fundamental changes in utility rate structures and energy system operation are vital. In this chapter, an optimal strategy for energy dispatch and pricing was investigated, which was shown experimentally to promote energy efficiency and retailer profitability, bill savings for the customers, and demand response for the grid. The key insight is that “behavior nudges” like retail prices can be co-optimized with integrated energy dispatch to leverage demand flexibility (e.g., by incentivizing load curtailment during on-peak hours, the DR scheme ensures both reliability and economy) and system synergy (e.g., the most significant reduction in operational costs is brought by the CHP plant, which performs fuel switching by exploiting the spark spread when the electricity wholesale tariff is high). However, while data-driven planning is shown in this chapter to improve h-CPS efficiency, it also makes the system more vulnerable to potential attacks that could compromise security. We will examine such a scenario in power grid in the next chapter for cyber resilience analysis, which is equally important (if not more so) for h-CPS operation.

Chapter 7

Cyber resilience of power grid state estimation

Human-cyber-physical systems continually face variable operational conditions caused by both internal and external factors. In Chap. 6, we explored data-driven strategies to improve system efficiency in a dynamic and uncertain environment. This chapter focuses on the aspect of resilience—the capability to predict, absorb, and recover from disturbances. With the growing concerns about the effects of potential cyberattacks on critical infrastructures like power grid, we analyze the vulnerability of a key procedure known as power grid state estimation against potential cyberattacks on data integrity, also known as a false data injection attack (FDIA). A general form of FDIA can be formulated as an optimization problem whose objective is to find a stealthy and sparse data injection vector on the sensor measurements that cause the state estimate to be spurious and misleading. Due to the nonlinear AC measurement model and the cardinality constraint, the problem includes both continuous and discrete nonlinearities. To solve the FDIA problem efficiently, we investigate a novel convexification framework based on semidefinite programming (SDP) and prove that the attack can be stealthy and sparse.

7.1 Power grid resilience and state estimation

The convergence of automation and information technology has enhanced reliability, efficiency, and agility of the modern grid. Managed by supervisory control and data acquisition (SCADA) systems, a wealth of sensor data from transmission and distribution infrastructures are collected and filtered in order to facilitate a key procedure known as power system state estimation (SE), which is conducted on a regular basis (e.g., every few minutes), as shown in Fig. 7.1 [3], [214]. The outcome presents system operators with essential information about the real-time operating status to improve situational awareness, make economic decisions, and take contingency actions in response to potential threat that could engander the grid reliability [172].

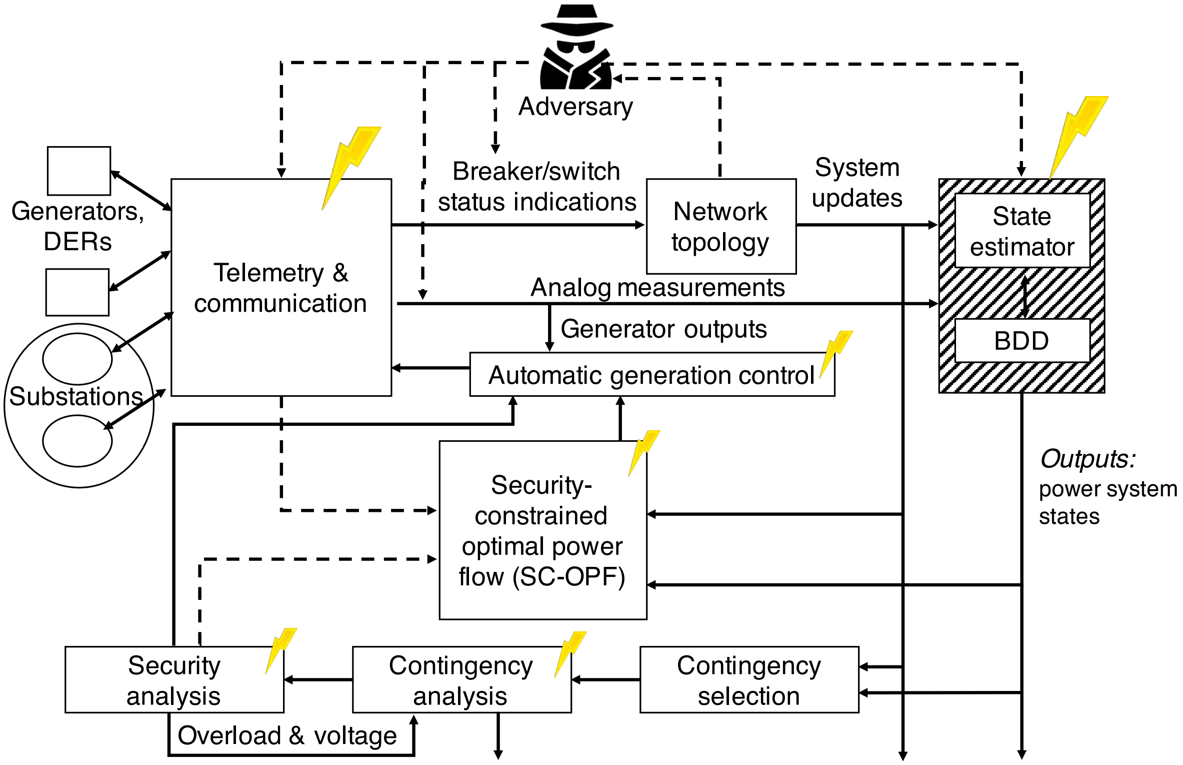


Figure 7.1: Illustration of power system operation and its vulnerability to cyberattacks (adapted from [158]). With unfettered access to the communication network and grid information system through cyber-intrusion, an adversary would be able to stage an attack on the system without any physical sabotage by simply injecting false data to the state estimator to impact the decision making for the system.

In smart grid where information is sent via remote terminal units (RTUs), maintaining the security of the communication network is imperative to guard against system intrusion and ensure operational integrity [214], [226], [229], [245]. However, traditional approaches such as security software, firewalls, and "air gaps", i.e., no connection between systems, are recognized as inadequate in the face of growing likelihood of breaches and cyber threat, such as the 2016 cyberattack on Ukraine's electricity infrastructure [23], [145]. In a recent report from the National Academies of Sciences, Engineering, and Medicine, titled "Enhancing the resilience of the nation's electricity system", the committee concluded that the United States' electric grid is vulnerable to a range of threats, among which terrorism and cyberattacks are most severe and could potentially cause long-term and widespread blackouts [172]. A process called "envisioning process" is recommended to improve the cyber security and resilience, which stresses the importance of "anticipating myriad ways in which the system might be disrupted and the many social, economic, and other consequences of such disruptions".

In this chapter, we analyze power grid vulnerability against cyberattack – more specifically, one critical class of threat known as false data injection attack, which attempts to stealthily modify data to introduce error into grid SE (Fig. 7.1) [148]. To stage an FDIA, the attacker needs to compromise power measurements by hacking the communication with SCADA. Previous works [38]. [136], [148], [181], [198], [205], [236], [241] have demonstrated that a stealthy FDIA is possible to evade bad data detection (BDD) by the control center, and can cause potential damages of load shedding [241], economic loss [145], [225], and even blackouts [146]. While these works have primarily studied a simplified power flow model, i.e., DC model [38], [83], [136], [148], [181], [198], [206], [236], [241], an FDIA based on a more accurate AC model is within the realm of possibility [3]. In a system where measurements are nonlinear functions of the state parameters, it is usually not easy to construct a state that evades BDD. Indeed, DC-based FDIA can be easily detected by AC-based BDD [186], [226]. On the other hand, the nonlinearity of equality power-flow constraints also makes the co-existence of multiple states and spurious solutions possible, which is a fundamental reason why an AC-based FDIA with sparse attacks is feasible and perhaps more detrimental than an DC-based FDIA. Once constructed, this new class of attacks could be hard to detect by existing methods. Thus, it is vital to understand its mechanism and devise protection/detection methods to thwart such attacks.

Adversarial FDIA. Potential adversarial FDIA strategies have been addressed in previous works on power system vulnerability analysis [86], [136], [148], [186], [241]. The negative impacts and possible defense mechanisms have also been studied [136], [146], [198], [241]. From a practitioner's point of view, there are mainly two categories, based on either DC or AC models [145], [225]. For DC-FDIA, an unobservability condition was derived and the attack was numerically shown to be sparse [136], [148], [241]. Distributed DC-FDIA with partial knowledge about the topology was considered in [181] ,[226]. The vulnerability was quantified by the minimum number of sensors needed to compromise in order to stage stealth FDIA [38], [136], [198]. This can be formulated as a minimum cardinality problem, where different algorithms have been proposed for efficient computation [83], [206]. As for the attack impact, FDIA has been studied on the electric market [236] and load redistribution [241] to show significant financial losses.

Only a few works have been published on AC-based FDIA, due to the recognized complexity of nonlinear systems [186], [214]. The paper [86] introduced a graph-based algorithm to identify a set of compromised sensors that suffices to construct an unobservable attack; however, this only offers an *upper bound* on the cardinality, rather than resource-constrained sparsity. The work [186] studied AC-based FDIA based on linearization around the target state under the assumption that SE is obtained by a specific algorithm, which could be too stringent in practice.

Contributions. Differentiated from prior literature, this study is the *first of its kind* to solve a general FDIA for the AC-based SE, with theoretical guarantees of sparsity and unobservability. Motivated by the theoretical challenges of continuous nonconvexity and discrete nonlinearity posed by AC-based FDIA, we propose a novel convexification framework using SDP, and prove conditions on stealth attack and performance bounds. This broadens the perspectives on power system security and vulnerability analysis. By investigating the least-effort strategy from the attacker's perspective, this study provides a realistic metric for the grid security based on the number of individual sensors required to thwart an FDIA. The results also motivate protection mechanisms for AC-based SE, such as the redesign of BDD [207].

Notations

Set notations. We use $\mathbb{R}$ and $\mathbf{C}$ as the sets of real and complex numbers, and $\mathbb{S}^n$ and $\mathbb{H}^n$ to represent the spaces of $n \times n$ real symmetric matrices and $n \times n$ complex Hermitian matrices, respectively. A set of indices $\{1, 2, ..., k\}$ is denoted by $[k]$. The set cardinality $\mathrm{Card}(\cdot)$ is the number of elements in a set. The support of a vector $\mathbf{x}$, denoted as $\mathrm{supp}\,(\mathbf{x})$, is the set of indices of the nonzero entries of $\mathbf{x}$. For a set $\mathcal{S} \subset \mathbb{R}^n$, we use $\mathcal{S}^c = \mathbb{R}^{n_m} \setminus \mathcal{S}$ to denote its complement. The notation $\mathrm{int}\,\Gamma$ is used to represent the interior of the set Γ.

Matrix notations. Vectors are shown by bold letters, and matrices are shown by bold and capital letters. The symbols $\mathbf{0}_n$, $\mathbf{1}_n$, $\mathbf{0}_{m\times n}$, $\mathbf{I}_{n\times n}$ denote the $n \times 1$ zero vector, $n \times 1$ one vector, $m \times n$ zero matrix, and $n \times n$ identity matrix, respectively. Let $[\mathbf{x}]_i$ denote the i-th element of the vector $\mathbf{x}$. For an $m \times n$ matrix $\mathbf{W}$, let $\mathbf{W}[\mathcal{X}, \mathcal{Y}]$ denote the submatrix of $\mathbf{W}$ whose rows are chosen from $\mathcal{X} \in [m]$ and whose columns are chosen from $\mathcal{Y} \in [n]$. The notation $\mathbf{W} \succeq 0$ indicates that $\mathbf{W}$ is Hermitian and positive semidefinite (PSD), and $\mathbf{W} \succ 0$ indicates that $\mathbf{W}$ is Hermitian and positive definite.

Operator notations. The symbols $(\cdot)^\top$ and $(\cdot)^*$ represent the transpose and conjugate transpose operators. We use $\Re(\cdot)$, $\Im(\cdot)$, $\mathrm{trace}\,(\cdot)$, and $\det(\cdot)$ to denote the real part, imaginary part, trace, and determinant of a scalar/matrix. The dot product is represented by $\mathbf{x}_1 \cdot \mathbf{x}_2 = \mathbf{x}_1^\top \mathbf{x}_2$, for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. The imaginary unit is denoted as $\mathbb{i}$. The notations $\angle x$ and $|x|$ indicate the angle and magnitude of a complex scalar; moreover, $\angle \mathbf{x}$ and $|\mathbf{x}|$ are defined based on the angles and magnitudes of all entries of the vector $\mathbf{x}$. For a convex function $g(\mathbf{x})$, we use $\partial g(\mathbf{x})$ to denote its subgradient. The notations $\|\mathbf{x}\|_0$, $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$ and $\|\mathbf{x}\|_\infty$ show the cardinality, 1-norm, 2-form and ∞-norm of $\mathbf{x}$.

Power system modeling

We model the electric grid as a graph $\mathcal{G} := \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N} := [n_b]$ and $\mathcal{L} := [n_l]$ represent its set of buses and branches. Denote the admittance of each branch $l \in \mathcal{L}$ that connects bus s and bus t as y_{st}. The mathematical framework of this work applies to more detailed models with shunt elements and transformers; but to streamline the presentation, these are not considered in the theoretical analysis of this paper. The grid topology is encoded in the bus admittance matrix $\mathbf{Y} \in \mathbb{C}^{n_b \times n_b}$, as well as the *from* and *to* branch admittance matrices $\mathbf{Y}_f \in \mathbb{C}^{n_l \times n_b}$ and $\mathbf{Y}_t \in \mathbb{C}^{n_l \times n_b}$, respectively (see [248, Ch. 3]).

The power system state is described by the bus voltage vector $\mathbf{v} = \left[v_1, ..., v_{n_b}\right]^\top \in \mathbb{C}^{n_b}$, where $v_k \in \mathbb{C}$ is the complex voltage at bus $k \in \mathcal{N}$ with magnitude $|v_k|$ and phase $\angle v_k$. Given the complex nodal vector, the nodal current injection can be written as $\mathbf{i} = \mathbf{Yv}$, and the branch currents at the from and to ends of all branches are given by $\mathbf{i}_f = \mathbf{Y}_f\mathbf{v}$ and $\mathbf{i}_t = \mathbf{Y}_t\mathbf{v}$, respectively. Define $\{\mathbf{e}_1, ..., \mathbf{e}_{n_b}\}$ and $\{\mathbf{d}_1, ..., \mathbf{d}_{n_l}\}$ as the sets of canonical vectors in $\mathbb{R}^{n_b}$ and $\mathbb{R}^{n_l}$, respectively. We can derive various types of power and voltage measurements as follows:

- *Voltage magnitude.* The voltage magnitude at bus k is given by $|v_k|^2 = \operatorname{trace}(\mathbf{E}_k\mathbf{vv}^*)$, where $\mathbf{E}_k := \mathbf{e}_k\mathbf{e}_k^\top$.
- *Nodal power injection.* The power injection at bus node k consists of real and reactive powers, $p_k + \mathbb{i}q_k$, where:

$$\begin{aligned} p_k &= \Re\left(i_k^* v_k\right) = \operatorname{trace}\left(\tfrac{1}{2}\left(\mathbf{Y}^*\mathbf{E}_k + \mathbf{E}_k\mathbf{Y}\right)\mathbf{vv}^*\right) \\ q_k &= \Im\left(i_k^* v_k\right) = \operatorname{trace}\left(\tfrac{1}{2\mathbb{i}}\left(\mathbf{Y}^*\mathbf{E}_k - \mathbf{E}_k\mathbf{Y}\right)\mathbf{vv}^*\right). \end{aligned}$$

- *Branch power flows.* Given a line $l \in \mathcal{L}$ from node s to node t, the real and reactive power flows in both directions are given by:

$$\begin{aligned} p_f^{(l)} &= \Re\left([\mathbf{i}_f]_l^* v_s\right) = \operatorname{trace}\left(\tfrac{1}{2}\left(\mathbf{Y}_f^*\mathbf{d}_l\mathbf{e}_s^\top + \mathbf{e}_s\mathbf{d}_l^\top\mathbf{Y}_f\right)\mathbf{vv}^*\right) \\ p_t^{(l)} &= \Re\left([\mathbf{i}_f]_l^* v_t\right) = \operatorname{trace}\left(\tfrac{1}{2}\left(\mathbf{Y}_f^*\mathbf{d}_l\mathbf{e}_t^\top + \mathbf{e}_t\mathbf{d}_l^\top\mathbf{Y}_f\right)\mathbf{vv}^*\right) \\ q_f^{(l)} &= \Im\left([\mathbf{i}_f]_l^* v_s\right) = \operatorname{trace}\left(\tfrac{1}{2\mathbb{i}}\left(\mathbf{Y}_f^*\mathbf{d}_l\mathbf{e}_s^\top - \mathbf{e}_s\mathbf{d}_l^\top\mathbf{Y}_f\right)\mathbf{vv}^*\right) \\ q_t^{(l)} &= \Im\left([\mathbf{i}_f]_l^* v_t\right) = \operatorname{trace}\left(\tfrac{1}{2\mathbb{i}}\left(\mathbf{Y}_f^*\mathbf{d}_l\mathbf{e}_t^\top - \mathbf{e}_t\mathbf{d}_l^\top\mathbf{Y}_f\right)\mathbf{vv}^*\right). \end{aligned}$$

Thus, each common measurement in power systems that belongs to one of the above *measurement types* can be written as:

$$f_i(\mathbf{v}) = \operatorname{trace}(\mathbf{M}_i\mathbf{vv}^*), \tag{7.1}$$

where $\mathbf{M}_i \in \mathbb{H}^{n_b}$ is the Hermitian measurement matrix for the i-th noiseless measurement (it is straightforward to include linear PMU measurements in our analysis as well).

AC-based state estimation

The SE problem aims at finding the unknown operating point of a power network, namely $\mathbf{v}$, based on a given set of measurements. During the operation, a set of measurements $\mathbf{v} \in \mathbb{R}^{n_m}$ are acquired:

$$\mathbf{v} = \mathbf{f}(\mathbf{v}) + \mathbf{e} + \mathbf{b}, \tag{7.2}$$

where $\mathbf{f} : \mathbb{C}^{n_b} \mapsto \mathbb{R}^{n_m}$ is the measurement function whose scalar elements are designated in (7.1), $\mathbf{e} \in \mathbb{R}^{n_m}$ denotes random noise, and $\mathbf{b} \in \mathbb{R}^{n_m}$ is the bad data error that accounts for sensor failure or adversarial injection. In the case of no bad data error, the common strategy for solving SE is to form the nonlinear weighted least squares problem:

$$\min_{\hat{\mathbf{v}}\in\mathcal{V}} \sum_{i=1}^{n_m} w_i(m_i - f_i(\hat{\mathbf{v}}))^2, \tag{7.3}$$

where $\mathcal{V}$ is the region of potential operating points, w_i is the inverse variance of sensor i, and $f_i(\hat{\mathbf{v}})$ is given in (7.1).

In the case that the sensor measurements are not corrupted by bad data and noise, i.e., $\mathbf{b} = \mathbf{e} = \mathbf{0}$, we describe a condition under which a state is "observable" based on the measurement types (matrices) $\mathcal{M} = \{\mathbf{M}_1, ..., \mathbf{M}_{n_m}\}$ [151]. First, we introduce some notations. Let $\mathcal{O}$ denote the set of all buses except the slack bus. The complex vector $\mathbf{v} \in \mathcal{C}^{n_b}$ can be represented by its real-valued counterpart:

$$\overline{\mathbf{v}} = \left[\Re\left(\mathbf{v}[\mathcal{N}]^\top\right) \quad \Im\left(\mathbf{v}[\mathcal{O}]^\top\right)\right]^\top \in \mathbb{R}^{2n_b-1}.$$

Accordingly, any $n \times n$ Hermitian matrix $\mathbf{M}$ can be characterized by a $(2n-1) \times (2n-1)$ real skew-symmetric matrix:

$$\overline{\mathbf{M}} = \begin{bmatrix} \Re\left(\mathbf{M}[\mathcal{N},\mathcal{N}]\right) & -\Im\left(\mathbf{M}[\mathcal{N},\mathcal{O}]\right) \\ \Im\left(\mathbf{M}[\mathcal{O},\mathcal{N}]\right) & \Re\left(\mathbf{M}[\mathcal{O},\mathcal{O}]\right) \end{bmatrix} \in \mathbb{R}^{(2n-1)\times(2n-1)}.$$

Based on (7.1) and the above notations, the vector-valued function $\mathbf{f}(\mathbf{v})$ maps the state to a set of noiseless measurements:

$$\mathbf{f}(\mathbf{v}) = \begin{bmatrix} \mathbf{v}^*\mathbf{M}_1\mathbf{v} \\ \vdots \\ \mathbf{v}^*\mathbf{M}_{n_m}\mathbf{v} \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{v}}^\top\overline{\mathbf{M}}_1\overline{\mathbf{v}} \\ \vdots \\ \overline{\mathbf{v}}^\top\overline{\mathbf{M}}_{n_m}\overline{\mathbf{v}} \end{bmatrix} \in \mathbb{R}^{n_m}, \tag{7.4}$$

whose Jacobian matrix is given by:

$$\mathbf{J}(\mathbf{v}) = 2\left[\overline{\mathbf{M}}_1\overline{\mathbf{v}} \quad \cdots \quad \overline{\mathbf{M}}_{n_m}\overline{\mathbf{v}}\right]. \tag{7.5}$$

Motivated by the inverse function theorem, which states that the inverse of the function $\mathbf{f}(\mathbf{v})$ exists locally if $\mathbf{J}(\mathbf{v})$ has full row rank, an "observability" definition is introduced below.

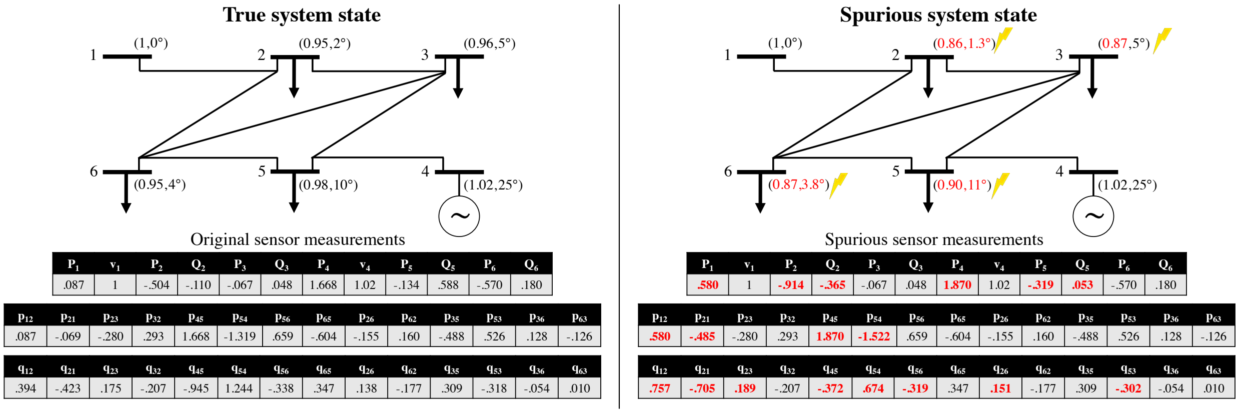
Definition 7.1 (Observability). *A state $\mathbf{v} \in \mathbf{C}^{n_b}$ is observable from a set of measurement types $\mathcal{M}$ if the Jacobian $\mathbf{J}(\mathbf{v})$ has full row rank. For a given set of measurement types $\mathcal{M}$, the observable set $\mathcal{V}(\mathcal{M})$ is the set of all observable states.*

In practice, the SE problem (7.3) can be solved efficiently using first-order methods such as the Gauss-Newton algorithm or a recent method based on SDP relaxation [151], [243]. Furthermore, as implied by the observability property and the Kantorovich theorem, if the state $\mathbf{v}$ is observable, then we can find it using the Gauss-Newton method by starting from any initial point sufficiently close to $\mathbf{v}$.

As captured by the bad data vector **b**, the sensor measurements might be corrupted by aberrant data. The common practice is to employ a BDD based on statistical hypothesis testing [214]. Under the null hypothesis that no bad injection exists, namely $b_i = 0$, the residual $(m_i - f_i(\hat{\mathbf{v}}))^2$ should follow the chi-squared distribution, where $\hat{\mathbf{v}}$ is the estimated state and the random error e_i is assumed to be normally distributed. A threshold value is set based on confidence levels to detect large residuals, whose corresponding data are discarded and a new iteration of SE starts. This procedure is able to sift out randomly occurring bad data; however, it can be ineffective to guard against systematically fabricated bad data, a type of cyberattack known as FDIA.

7.2 Vulnerability of AC-based state estimation

FDIA is a cyberattack on the data analytic process, where a malicious agent intentionally injects false data $\mathbf{b} \in \mathbb{R}^{n_m}$ into the n_m grid sensors to make system operators believe in an operating state, namely $\tilde{\mathbf{v}}$, other than the true state $\mathbf{v}$ [145], [226]. As an illustrative example (Fig. 7.2), the operator would be "tricked" if the attacker manages to tamper with certain power flow measurements to generate a fake state estimate of the system.



Original sensor measurements

P_1	v_1	P_2	Q_2	P_3	Q_3	P_4	v_4	P_5	Q_5	P_6	Q_6
.087	1	-.504	-.110	-.067	.048	1.668	1.02	-.134	.588	-.570	.180

p_{12}	p_{21}	p_{23}	p_{32}	p_{45}	p_{54}	p_{56}	p_{65}	p_{26}	p_{62}	p_{35}	p_{53}	p_{36}	p_{63}
.087	-.069	-.280	.293	1.668	-1.319	.659	-.604	-.155	.160	-.488	.526	.128	-.126

q_{12}	q_{21}	q_{23}	q_{32}	q_{45}	q_{54}	q_{56}	q_{65}	q_{26}	q_{62}	q_{35}	q_{53}	q_{36}	q_{63}
.394	-.423	.175	-.207	-.945	1.244	-.338	.347	.138	-.177	.309	-.318	-.054	.010

Spurious sensor measurements

P_1	v_1	P_2	Q_2	P_3	Q_3	P_4	v_4	P_5	Q_5	P_6	Q_6
.580	1	**-.914**	**-.365**	-.067	.048	**1.870**	1.02	**-.319**	**.053**	-.570	.180

p_{12}	p_{21}	p_{23}	p_{32}	p_{45}	p_{54}	p_{56}	p_{65}	p_{26}	p_{62}	p_{35}	p_{53}	p_{36}	p_{63}
.580	**-.485**	-.280	.293	**1.870**	**-1.522**	.659	-.604	-.155	.160	-.488	.526	.128	-.126

q_{12}	q_{21}	q_{23}	q_{32}	q_{45}	q_{54}	q_{56}	q_{65}	q_{26}	q_{62}	q_{35}	q_{53}	q_{36}	q_{63}
.757	**-.705**	**.189**	-.207	**-.372**	**.674**	**-.319**	.347	**.151**	-.177	.309	**-.302**	-.054	.010

Figure 7.2: An example of a 6-bus system, where the nodal voltage magnitudes and power injections as well as branch power flows are measured (p.u.). The attacker injects false data (red) to influence the bus state estimates (shown on the right side of each bus). The per unit bases for power and voltage are 100MW and 240KV, respectively. The line admittance values are identical to $1+1\mathbb{i}$. The FDIA injection is solved by (SDP-FDIA), with parameters shown in Table 7.1. Note that p_{ij} and q_{ij} show the active and reactive power flows over the line (i, j).

FDIA differs from randomly occurring bad data in its stealth operation to evade BDD. Existing works have investigated stealth conditions for FDIA on DC-based SE [136], [148]. The following definition of "stealth" is provided to include cases of both DC- and AC-based models.

Definition 7.2 (Stealth)**.** *An attack* $\mathbf{b}$ *is stealthy under state* $\mathbf{v}$ *if, in the absence of the measurement noise* $\mathbf{e}$*, there exists a nonzero vector* $\boldsymbol{c}$ *such that* $\mathbf{f}(\mathbf{v}) + \mathbf{b} = \mathbf{f}(\mathbf{v} + \boldsymbol{c})$*.*

The following lemma provides a sufficient condition for AC-based attacks to remain stealthy.

Lemma 7.1 (Sufficient condition for stealth attack)**.** *An attack* $\mathbf{b}$ *is stealthy if there exists a nonzero vector* $\boldsymbol{c}$ *such that* $\mathbf{M}_i\boldsymbol{c} = \mathbf{0}$ *for every* $i \in [n_m]$ *that is not in the support of* $\mathbf{b}$*.*

Proof. See Appendix A.5. □

Lemma 7.1 implies that an attack is unobservable if the state deviation $\boldsymbol{c}$ lies in the *null space* of the measurement matrices of those sensors the attacker does not tamper with. This is applicable to the situation discussed in [86] for a single bus attack. To better understand this, consider a vector $\boldsymbol{c}$ that has zeros everywhere except at location j. Since the j-th column of $\mathbf{M}_i$, denoted as $[\mathbf{M}_i]_{:j}$, is zero unless $\mathbf{M}_i$ corresponds to the measurement of a branch that connects to bus j, this delineates a "superset" of sensors needed to hack to guarantee a stealth attack.

An upper bound on the minimum number of compromised sensors can be derived for a multi-bus attack; however, the sufficient condition could be too stringent because the attacker only needs to satisfy $b_i = \operatorname{trace}(\mathbf{M}_i\boldsymbol{c}\boldsymbol{c}^*) + \operatorname{trace}(\mathbf{M}_i\boldsymbol{c}\mathbf{v}^*) + \operatorname{trace}(\mathbf{M}_i\mathbf{v}\boldsymbol{c}^*) = 0$ for all $i \notin \operatorname{supp}(\mathbf{b})$ to remain stealthy. For instance, consider the system in Fig. 7.2. Since the bus states are all under attack, the upper bound on the minimum number of sensors to infiltrate is 40, or all the measurements, according to [86] and Lemma 1. But due to the "clever" design, FDIA is conducted successfully by tampering with only 18 sensors, which is a sparser subset of the upper bound. It is also worthwhile to note that one can think of a strategy that offsets the phases of bus voltages at bus 2, 3, 5 and 6 by a constant. This will keep the real power flows the same as before and only change the reactive flows. However, even with this ad hoc strategy, the number of sensors to tamper with is 19. This indicates the efficiency of the demonstrated strategy. However, to find such an attack vector, a general strategy can be formulated as an optimization problem to maximize sabotage with limited resources and to evade detection:

$$\begin{aligned} \min_{\tilde{\mathbf{v}}\in\mathbb{C}^{n_b},\mathbf{b}\in\mathbb{R}^{n_m}} \quad & h(\tilde{\mathbf{v}}) \\ \text{s. t.} \quad & \mathbf{f}(\tilde{\mathbf{v}}) = \mathbf{v} + \mathbf{b} \\ & \|\mathbf{b}\|_0 \leq c \end{aligned} \tag{NC-FDIA}$$

where $\mathbf{f}(\cdot)$ is the AC-model measurement function (7.1), $\tilde{\mathbf{v}}$ is the spurious state, $h(\cdot)$ is an optimization criterion to be specified later, and c is a constant number. The constraints amount to the unobservability condition (Definition 7.2) and the sparsity requirement. The following assumption is made on the adversary attack capability:

Assumption 1. *The attacker can form a strategy after accessing the grid topology and the measurement vector* $\mathbf{v}$*.*

The above assumption depicts a powerful adversary and a completely adversarial scenario. Using the full set of measurements, the attacker can perform SE to estimate the true state $\mathbf{v}$, and tailor the attack to be stealthy. However, if this assumption is violated, the attacker risks being detected by the BDD [226]. The analysis provided in this paper is based on Assumption 1 because it helps understand the behavior of the system under the worst attack possible (using the full knowledge of the system) and simplifies the mathematical treatment.

Several objectives are possible for the attacker to fulfill various malicious goals, such as:

- *Target state attack:* $h(\tilde{\mathbf{v}}) = \|\tilde{\mathbf{v}} - \mathbf{v}_{tg}\|_2^2$ to misguide the operator towards $\mathbf{v}_{tg}$
- *Voltage collapse attack:* $h(\tilde{\mathbf{v}}) = \|\tilde{\mathbf{v}}\|_2^2$ to deceive the operator to believe in low voltages
- *State deviation attack:* $h(\tilde{\mathbf{v}}) = -\|\tilde{\mathbf{v}}-\mathbf{v}\|_2^2$ to yield the estimated state $\tilde{\mathbf{v}}$ to be maximally different from the true state $\mathbf{v}$

An FDIA attack can be formed by solving (NC-FDIA) with one of the above objectives; however, the problem is challenging due to: 1) a possibly nonconvex objective function, e.g., concave for the state deviation attack, 2) nonlinear equalities, and 3) cardinality constraints. The next section develops an efficient strategy to deal with these issues.

7.3 SDP convexification of the FDIA problem

Since the original attack problem (NC-FDIA) is nonconvex and difficult to tackle, we propose a convexification method based on SDP, which can be solved efficiently. Based on this framework, an "attackable region" of system states is characterized, where a strategy is guaranteed to exist and can be found efficiently. To streamline the presentation, we focus the analysis on the case of "target state attack," where $h(\tilde{\mathbf{v}}) = \|\tilde{\mathbf{v}} - \mathbf{v}_{tg}\|_2^2$ with $\mathbf{v}_{tg}$ chosen by the adversary *a priori*. The results hold for many other objective functions as well.

SDP convexification

By introducing an auxiliary variable $\mathbf{W} \in \mathbb{H}^{n_b}$ and the associated function $\bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) = \operatorname{trace}(\mathbf{W}) - \tilde{\mathbf{v}}^*\mathbf{v}_{tg} - \mathbf{v}_{tg}^*\tilde{\mathbf{v}}$, (NC-FDIA) can be reformulated as:

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}}\in\mathbb{C}^{n_b},\mathbf{b}\in\mathbb{R}^{n_m},\\ \mathbf{W}\in\mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) \\ \text{s. t.} \quad & \operatorname{trace}(\mathbf{M}_i\mathbf{W}) = m_i + b_i, \quad \forall i \in [n_m] \\ & \|\mathbf{b}\|_0 \leq c \\ & \mathbf{W} = \tilde{\mathbf{v}}\tilde{\mathbf{v}}^* \end{aligned} \tag{NC-FDIA-r}$$

A cardinality-included SDP relaxation of the above nonconvex problem can be obtained by replacing $\mathbf{W} = \tilde{\mathbf{v}}\tilde{\mathbf{v}}^*$ with a general PSD constraint:

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}} \in \mathbb{C}^{n_b}, \mathbf{b} \in \mathbb{R}^{n_m}, \\ \mathbf{W} \in \mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) \\ \text{s. t.} \quad & \operatorname{trace}(\mathbf{M}_i \mathbf{W}) = m_i + b_i, \quad \forall i \in [n_m] \\ & \|\mathbf{b}\|_0 \leq c \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0 \end{aligned} \tag{NC-FDIA-c}$$

To study the relationship between the nonconvex problem (NC-FDIA-r) and its cardinality-included relaxation (NC-FDIA-c), we define an augmented matrix:

$$\hat{\mathbf{Z}} = \begin{bmatrix} 1 & \hat{\mathbf{v}}^* \\ \hat{\mathbf{v}} & \hat{\mathbf{W}} \end{bmatrix}, \tag{7.6}$$

where $(\hat{\mathbf{v}}, \hat{\mathbf{W}})$ is a solution of (NC-FDIA-c). It is straightforward to verify that if rank$(\hat{\mathbf{Z}})$ is equal to 1, then we must have $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$. Thus, $(\hat{\mathbf{v}}, \hat{\mathbf{W}})$ is feasible for (NC-FDIA-r) and consequently optimal since the objective value of (NC-FDIA-c) is a lower bound for (NC-FDIA-r). In fact, by exploring the special features of the problem, we can derive a milder condition to guarantee the equivalence. This will be elaborated next.

Assumption 2a. *Given a solution $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ of (NC-FDIA-c), $\hat{\mathbf{v}}$ and $\mathbf{v}_{tg}$ point along the same "general direction" in the sense that:*

$$\hat{\mathbf{v}}^* \mathbf{v}_{tg} + \mathbf{v}_{tg}^* \hat{\mathbf{v}} > 0. \tag{7.7}$$

Note that the objective function of (NC-FDIA-c) helps with the satisfaction of Assumption 2a, since the objective aims at making $\hat{\mathbf{v}}$ and $\mathbf{v}_{tg}$ be as close as possible to each other.

Theorem 7.1. *The relaxation (NC-FDIA-c) recovers a solution of the nonconvex problem (NC-FDIA) and finds an optimal attack if it has a solution $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ satisfying Assumption 2a such that rank$(\hat{\mathbf{W}}) = 1$.*

Proof. See Appendix A.5. □

Theorem 7.1 ensures that if rank$(\hat{\mathbf{W}}) = 1$, then rank$(\hat{\mathbf{Z}}) = 1$ (even though it could theoretically be 2), in which case (NC-FDIA-c) is able to find an optimal attack. Nevertheless, the optimal solution of (NC-FDIA-c) is not guaranteed to be rank-1, and in addition the cardinality constraint $\|\mathbf{b}\|_0 \leq c$ in this optimization problem is intractable. We introduce a series of techniques to deal with each issue.

To enforce (NC-FDIA-c) to possess a rank-1 solution, we aim at penalizing the rank of its solution via a convex term. The literature of compressed sensing suggests using the nuclear norm penalty trace $(\mathbf{W})$ [52]. However, this penalty is not appropriate for power systems, since it penalizes the voltage magnitude at each bus and may yield impractical results. Instead, a more general penalty term in the form of trace $(\mathbf{M}_0\mathbf{W})$ will be used in this paper:

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}}\in\mathbb{C}^{n_b},\mathbf{b}\in\mathbb{R}^{n_m},\\ \mathbf{W}\in\mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}},\mathbf{W}) + \operatorname{trace}(\mathbf{M}_0\mathbf{W}) \\ \text{s. t.} \quad & \operatorname{trace}(\mathbf{M}_i\mathbf{W}) = m_i + b_i, \quad \forall i \in [n_m] \\ & \|\mathbf{b}\|_0 \leq c \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0, \end{aligned} \tag{NC-FDIA-p}$$

where $\mathbf{M}_0$ is to be designed. Similar to Lasso [218], we can replace the cardinality constraint in the above problem with an l_1-norm penalty added to the objective function to induce sparsity, which leads to the convex program:

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}}\in\mathbb{C}^{n_b},\mathbf{b}\in\mathbb{R}^{n_m},\\ \mathbf{W}\in\mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}},\mathbf{W}) + \operatorname{trace}(\mathbf{M}_0\mathbf{W}) + \alpha\|\mathbf{b}\|_1 \\ \text{s. t.} \quad & \operatorname{trace}(\mathbf{M}_i\mathbf{W}) = m_i + b_i, \ \forall i \in [n_m] \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0 \end{aligned} \tag{SDP-FDIA}$$

where α is a constant regularization parameter. After this convexification, (SDP-FDIA) is thus an SDP (after reformulating the l_1-norm term in a linear way), which can be solved efficiently using standard numerical solvers (e.g., SeDuMi and SDPT3) [232]. On the other hand, we recognize that by including penalty terms for rank and sparsity, we inevitably introduce bias to the optimization problem. Thus, the result obtained by (SDP-FDIA) should be described as "near-optimal," in comparison to a global minimum of (NC-FDIA). This is an artifact that arises from the computational complexity of the problem, and can be only remedied by a careful selection of the penalty coefficients.

Assumption 2b. *Given a solution* $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ *of* (SDP-FDIA), $\hat{\mathbf{v}}$ *and* $\mathbf{v}_{tg}$ *have the same general direction in the sense of* (7.7).

Lemma 7.2 (Stealth attack). *Let* $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ *be a solution of* (SDP-FDIA) *satisfying Assumption 2b. The attack* $\hat{\mathbf{b}}$ *is stealthy if rank*$(\hat{\mathbf{W}}) = 1$.

Proof. See Appendix A.5. □

Attackable region

In this section, we first introduce and characterize the set of voltages that the attacker can achieve by solving (SDP-FDIA) for the malicious data injection. Then, we analyze the sabotage scale under the studied FDIA. Throughout this section, let $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ denote an optimal solution of (SDP-FDIA). Given any stealth attack $\mathbf{b}$, we define an optimization problem based on (SDP-FDIA) to minimize over $(\mathbf{v}, \mathbf{W})$ with a fixed $\mathbf{b}$, and denote its optimal objective value as $g(\mathbf{b})$:

$$\begin{aligned} g(\mathbf{b}) = \min_{\substack{\tilde{\mathbf{v}} \in \mathbb{C}^{n_b}, \\ \mathbf{W} \in \mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) + \operatorname{trace}(\mathbf{M}_0 \mathbf{W}) \\ \text{s. t.} \quad & \operatorname{trace}(\mathbf{M}_i \mathbf{W}) = m_i + b_i, \quad \forall i \in [n_m] \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0 \end{aligned} \tag{FDIA-SE}$$

In the following, we will use $g(\mathbf{b})$ as a proxy for the sabotage scale.[1] Now, we define an "attackable" state below.

Definition 7.3 (Attackable state)**.** *A state $\mathbf{v}_{at}$ is attackable if $(\mathbf{v}_{at}, \mathbf{W} = \mathbf{v}_{at}\mathbf{v}_{at}^*)$ is the unique and optimal solution of* (FDIA-SE) *for some stealth attack vector $\mathbf{b} \in \mathbb{R}^m$.*

Definition 7.4 (Attackable region)**.** *The attackable region $\mathcal{A}(\mathcal{M}, \rho)$ for a given set of measurement types $\mathcal{M}$ is the set of states $\mathbf{v}_{at}$ that is attackable for some $\mathbf{M}_0$ with bounded norm $\|\mathbf{M}_0\|_2 \leq \rho$.*

In other words, for any state $\mathbf{v}_{at} \in \mathcal{A}(\mathcal{M}, \rho)$ in the attackable region, there exists a stealth attack $\mathbf{b}$ such that $(\mathbf{v}_{at}, \mathbf{W} = \mathbf{v}_{at}\mathbf{v}_{at}^*, \mathbf{b})$ is a feasible solution of (SDP-FDIA) and that $(\mathbf{v}_{at}, \mathbf{W} = \mathbf{v}_{at}\mathbf{v}_{at}^*)$ is optimal if we fix the attack $\mathbf{b}$. The size of $\mathcal{A}(\mathcal{M}, \rho)$ also depends on ρ; more specifically, we have $\mathcal{A}(\mathcal{M}, \rho_1) \subseteq \mathcal{A}(\mathcal{M}, \rho_2)$ for $\rho_1 \leq \rho_2$. In what follows, we will characterize the attackable region.

Theorem 7.2. *If $\mathcal{A}(\mathcal{M}, \rho)$ is non-empty for some $\rho > 0$, the intersection of the attackable region and the observable set, i.e., $\mathcal{A}(\mathcal{M}, \rho) \cap \mathcal{V}(\mathcal{M})$, is an open set.*

Proof. See Appendix A.5. □

For some special cases, we can have a more explicit characterization of the attackable region, as explained later.

Theorem 7.3. *Consider the "target state attack" with $\bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) = \mathit{trace}(\mathbf{W}) - \tilde{\mathbf{v}}^*\mathbf{v}_{tg} - \mathbf{v}_{tg}^*\tilde{\mathbf{v}}$, where $\mathbf{v}_{tg} \in \mathcal{V}(\mathcal{M})$ is chosen to be observable. Then, $\mathbf{v}_{tg} \in \mathcal{A}(\mathcal{M}, \rho)$ for some $\rho > 0$, i.e., $\mathbf{v}_{tg}$ is attackable.*

Proof. See Appendix A.5. □

[1] For an optimal solution of (SDP-FDIA), the term $\operatorname{trace}\left(\mathbf{M}_0 \hat{\mathbf{W}}\right)$ can be bounded within limited ranges; as a result, $g(\mathbf{b})$ acts as a "proxy" for $\bar{h}(\hat{\mathbf{v}}, \hat{\mathbf{W}})$.

Note that the proof of Theorem 7.3 allows computing ρ explicitly. Define a set of voltages $\mathcal{R}(\mathbf{Y}) \subset \mathbf{C}^{n_b}$ such that $\mathbf{v} \in \mathcal{R}(\mathbf{Y})$ if and only if, for each line $l \in \mathcal{L}$ that connect nodes s and t, we have:

$$-\pi \leq \angle v_s - \angle v_t - \angle y_{st} \leq 0 \tag{7.8}$$

$$0 \leq \angle v_s - \angle v_t + \angle y_{st} \leq \pi \tag{7.9}$$

where y_{st} is the branch admittance between buses s and t. Since real-world transmission systems feature low resistance-to-reactance ratios, the angle of each line admittance y_{st} is close to $-\pi/2$ [3], and thus a realistic vector $\mathbf{v}$ would belong to $\mathcal{R}(\mathbf{Y})$ under normal conditions where the voltage phase difference along each line is relatively small. The following result gives an explicit form for a region that is attackable, in the case where the set of measurement types includes only the branch power flows and nodal voltage magnitudes, but not the nodal bus injections. Henceforth, we will refer to this case as the "special case" (compared to the "general case" where nodal bus injections can also be included in the measurements).

Theorem 7.4. *Let $\mathcal{V}(\mathcal{M}) \subset \mathbf{C}^{n_b}$ denote the set of observable states for a given set of measurement types $\mathcal{M}$ including the branch power flows and nodal voltage magnitudes, but not the nodal bus injections. Then, we have $\mathcal{V}(\mathcal{M}) \cap \mathcal{R}(\mathbf{Y}) \subseteq \mathcal{A}(\mathcal{M}, \rho)$ for some $\rho > 0$.*

Proof. See Appendix A.5. □

The attackable region is an important concept that characterizes the outcome of solving (SDP-FDIA), meaning that if a state is in the attackable region, then it is a candidate attack strategy as well as the unique solution of (FDIA-SE) for some stealth attack. However, this does not imply that no stealth attack exists for a state $\tilde{\mathbf{v}}$ that is not in the attackable region; in fact, we can always construct a stealth data injection $\mathbf{b} = \mathbf{f}(\tilde{\mathbf{v}}) - \mathbf{v}$, where $\mathbf{v}$ is the true state. For example, if the measurement set $\mathcal{M}$ is so small that a part of the grid remains unobservable (see Definition 7.1), then (FDIA-SE) does not have a unique solution for any stealth attack $\mathbf{b}$. In that case, the attack-targeted state $\tilde{\mathbf{v}}$ does not belong to $\mathcal{A}(\mathcal{M}, \rho)$. In light of Theorem 7.2, if a state $\mathbf{v}_{at}$ is attackable, then any state in its small neighborhood is also attackable. Since we do not know the outcome of (SDP-FDIA) a priori, it is helpful to design a particular rank penalty matrix $\mathbf{M}_0$; indeed, as shown in Theorem 7.3, this can guarantee that a desired observable state is attackable. Further, Theorem 7.4 indicates that any observable state is attackable over a set of branch power flow measurements. In fact, we will give an explicit formula for $\mathbf{M}_0$ in this case (see the proof of Theorem 7.4 in Appendix A.5) such that the solution to (SDP-FDIA) is unique and in the form of $(\hat{\mathbf{v}}, \mathbf{W} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$.

Performance bounds for (SDP-FDIA)

The main objective of this section is to compare the solution of (SDP-FDIA) to an "oracle attack" to be defined later, and provide guarantees for stealthy solutions (Lemma 7.2). First, we focus on the properties of the sabotage scale $g(\mathbf{b})$ defined in (FDIA-SE).

Lemma 7.3. *$g(\mathbf{b})$ is convex and sub-differentiable.*

Proof. See Appendix A.5. □

To proceed, we consider an "oracle attack" that is able to solve (NC-FDIA-p).

Definition 7.5 (Oracle attack)**.** *The oracle attack $\mathbf{b}^\star \in \mathbb{R}^{n_m}$ is a global minimum of the nonconvex program (NC-FDIA-p). Define $\mathcal{B} \subseteq \mathbb{R}^{n_m}$ as the set of all vectors in $\mathbb{R}^{n_m}$ with the same support as $\mathbf{b}^\star$.*

Let $\boldsymbol{\Delta}_\mathcal{B} = \arg\min_{\boldsymbol{\Delta}_t \in \mathcal{B}} \|\boldsymbol{\Delta} - \boldsymbol{\Delta}_t\|_2^2$ be the projection of a vector $\boldsymbol{\Delta}$ onto the set $\mathcal{B}$. The deviation of the solution of (SDP-FDIA) from the oracle, namely $\hat{\boldsymbol{\Delta}} = \hat{\mathbf{b}} - \mathbf{b}^\star$, belongs to a cone.

Lemma 7.4. *For every $\alpha \geq 2\|\partial g(\mathbf{b}^\star)\|_\infty$, the error $\hat{\boldsymbol{\Delta}} = \hat{\mathbf{b}} - \mathbf{b}^\star$ belongs to the cone $C(\mathcal{B}, \mathcal{B}^c; \mathbf{b}^\star) = \{\boldsymbol{\Delta} \in \mathbb{R}^{n_m} | \|\boldsymbol{\Delta}_{\mathcal{B}^c}\|_1 \leq 3\|\boldsymbol{\Delta}_\mathcal{B}\|_1\}$.*

Proof. See Appendix A.5. □

For a general set of measurements that might include an arbitrary set of voltage magnitudes, nodal injections, and branch power flows, the following theorem provides performance bounds and a condition for stealthy attack using (SDP-FDIA).

Theorem 7.5. *Consider (SDP-FDIA) for a "target state attack" with $\bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) = \mathrm{trace}(\mathbf{W}) - \tilde{\mathbf{v}}^*\mathbf{v}_{tg} - \mathbf{v}_{tg}^*\tilde{\mathbf{v}}$, where $\mathbf{v}_{tg} \in \mathcal{V}(\mathcal{M})$ is chosen to be observable. Let $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ denote an optimal solution of (SDP-FDIA) for an arbitrary α greater than or equal to $2\|\partial g(\mathbf{b}^\star)\|_\infty$. The difference between the sabotage scale of the solved attack and the oracle attack satisfies the inequalities:*

$$-2\alpha\|\hat{\boldsymbol{\Delta}}_\mathcal{B}\|_1 \leq g(\hat{\mathbf{b}}) - g(\mathbf{b}^\star) \leq \alpha\left(\|\hat{\boldsymbol{\Delta}}_\mathcal{B}\|_1 - \|\hat{\boldsymbol{\Delta}}_{\mathcal{B}^c}\|_1\right),$$

where $\hat{\boldsymbol{\Delta}} = \hat{\mathbf{b}} - \mathbf{b}^\star$ is the difference with the oracle $\mathbf{b}^\star$.

Proof. See Appendix A.5. □

According to Theorem 7.5, there is a trade-off between attack sparsity and outcome in the sense that a tighter bound can be achieved with more entries outside the oracle sparse set $\mathcal{B}$. However, this also means that the attacker needs to tamper with more sensors. Moreover, the matrix $\mathbf{M}_0$ in (SDP-FDIA) can be constructed systematically using the Gram-Schmidt process (as detailed in the proof of Theorem 7.3).

7.4 Experiments on IEEE standard systems

This section numerically studies the vulnerability of power system AC-based SE under FDIA. More specifically, the objective is to validate whether the solution of (SDP-FDIA) is sparse and stealthy.

We first study the 30-bus system provided in MATPOWER [248] (Fig. 7.3). The states of this system are randomly initialized with magnitudes close to 1 and small phases. We consider a comprehensive measurement portfolio, which includes nodal voltage magnitudes, power injections, and branch real/reactive power flows. To streamline the presentation, we will focus on the target state attack, i.e., $h(\tilde{\mathbf{v}}) = \|\tilde{\mathbf{v}} - \mathbf{v}_{tg}\|_2^2$, where the entries of the target $\mathbf{v}_{tg}$ have been deliberately chosen to have low magnitudes (around 0.9), and phases identical to their counterparts in the true state. This would often trigger misguided contingency response, in an attempt to recover from the voltage sag [23]. Throughout the experiments, we assume that the sensor noise has a standard deviation of 1% of the measurement value.

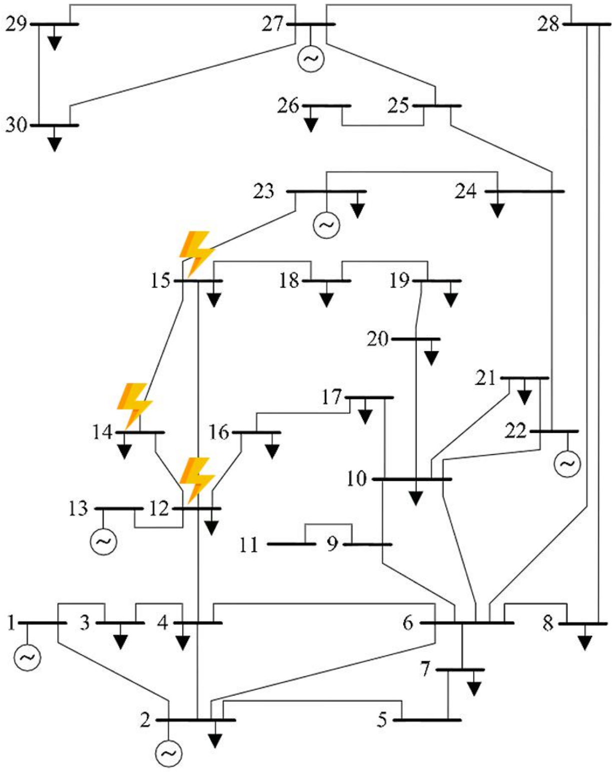
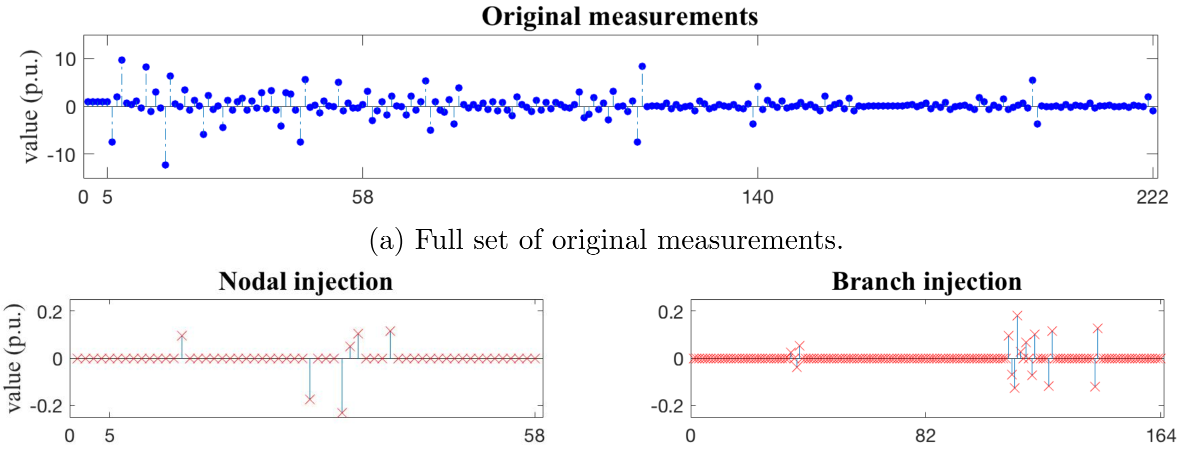


Figure 7.3: The IEEE 30-bus test case [248].

An FDIA injection is obtained in Fig. 7.4 by solving (SDP-FDIA) with parameters listed in Table 7.1. There are 222 measurements in total, which are organized in Fig. 7.4a by voltage magnitudes (indices 1–5), nodal real and reactive power injections (indices 5–58),



(a) Full set of original measurements.

(b) FDIA on nodal power and voltages.

(c) FDIA on branch power flows.

Figure 7.4: There are 222 measurements in total, which are organized in Figure (a) by voltage magnitudes (indices 1–5), nodal real and reactive power injections (indices 5–58), branch real power flows (indices 58–140), and branch reactive power flows (indices 140–222). The FDIA injections for nodal measurements are shown in Figure (b), where indices 1–5 and 5–58 correspond to voltage magnitudes and bus injections, respectively. The FDIA injections for branch measurements are provided in Figure (c), where indices 1–82 and 82–164 correspond to real power flows and reactive power flows, respctively.

branch real power flows (indices 58–140), and branch reactive power flows (indices 140–222). The FDIA injections for nodal measurements and branch measurements are also shown in Fig. 7.4. It can be observed that the injection values are relatively sparse, especially for real power flows over branches (indices 1–82 in Fig. 7.4c). This is due to the fact that they depend mainly on the phase differences between buses, but the target voltages have identical phases as the true state. The geographic locations of the attacked sensors include the locations of buses under attack (buses 12, 14 and 15) and the locations of the adjacent power lines, as confined within the superset used to calculate the upper bound [86]. In addition, the spurious measurements against the original values are depicted in Fig. 7.5. Given the presence of innate sensor noise, it is difficult to identify the attack on the raw measurement values by observation. In other words, the attack is "hidden" among the sensor noises.

Assume that the FDIA visualized in Fig. 7.4 is successfully implemented by the adversary on the set of measurements, and then the system operator solves the SE problem using the Gauss-Newton algorithm implemented in MATPOWER (note that the attack is SE-algorithm-agnostic). The obtained spurious states are plotted against the true states for the voltage magnitudes and phases in Fig. 7.6. Even though the system operates in a normal state with magnitudes in the prescribed interval $[0.98, 1.02]$, FDIA "tricks" the operator to believe in a potential voltage sag where some of the voltage magnitudes are outside of

Table 7.1: Simulation experiments, lists of the regularization parameters α and ϵ, the rank of $\hat{\mathbf{Z}}$, and the cardinality of $\hat{\mathbf{b}}$, as well as the upper bound given by [86].

system	α	ϵ	rank($\hat{\mathbf{Z}}$)	Card($\hat{\mathbf{b}}$)	upper bound	buses attacked*	pass BDD
6-bus†	.4	1/6	1	18	40	[2,3,5,6]	Yes
14-bus	.2	1/14	1	16	46	[2,3,4]	Yes
30-bus	1.16	1/30	1	21	54	[12,14,15]	Yes
39-bus	1.82	1/39	1	18	36	[26,28,29]	Yes
57-bus	0.5	1/57	1	30	92	[6,7,8]	Yes

* The attacked bus numbers are identical to the MATPOWER description.

† The 6-bus system is described in Fig. 7.2.

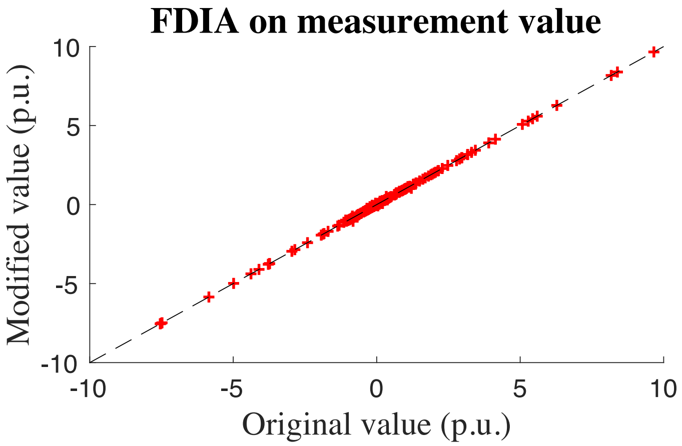


Figure 7.5: This plot shows the spurious values against the original values for all the measurements. The identity relation $y = x$ is illustrated by the dotted line. It can be observed that, given the presence of innate sensor noise, the spurious values are almost identical to the original measurements.

the above interval (green area in Fig. 7.6). Consequently, the operator may take harmful contingency actions. It is worthwhile to note that since the phases of the designed target states $\mathbf{v}_{tg}$ are identical to those of the true states by design, the spurious states estimated by the operator change insignificantly in phases, as shown in the right plot of Fig. 7.6.

To examine the effect of the regularization parameter α on the solution sparsity, we have run ten independent experiments with random sensor noise values and plotted the cardinality of $\hat{\mathbf{b}}$ with respect to α, as shown in Fig. 7.7. While the absence of $\|\cdot\|_1$ penalty (i.e., $\alpha = 0$) results in a dense solution, as α increases, the attack $\hat{\mathbf{x}}^a$ becomes significantly sparser compared to the upper bound provided by [86]. However, as α continuously increase, since the attack becomes sparser, its effect on SE reduces. This fact is reflected in the performance

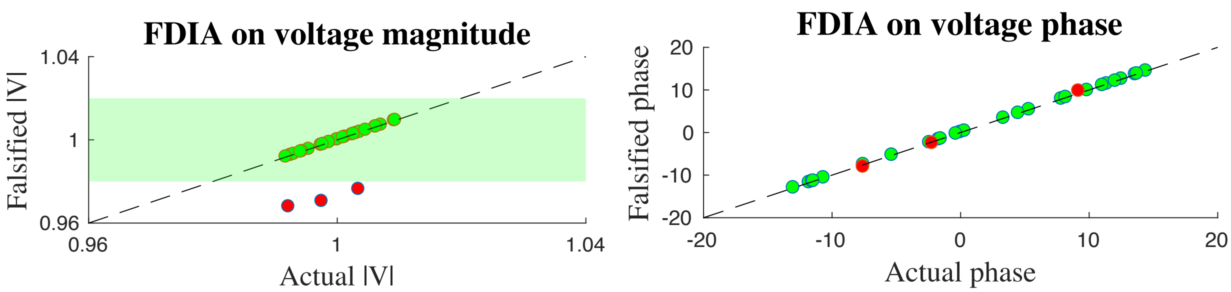


Figure 7.6: These plots depict spurious state estimation against true state for voltage magnitude (left) and voltage phases (right). In both plots, the dotted line indicates the $y = x$ relationship. For the magnitude plot, the green region specifies the normal operating interval $[098, 1.02]$. Observe that some spurious voltage magnitudes fall out of this prescribed operating region, while all of the spurious states have almost the same phases as their counterparts in the true states, due to the specifications by the FDIA target voltage vector.

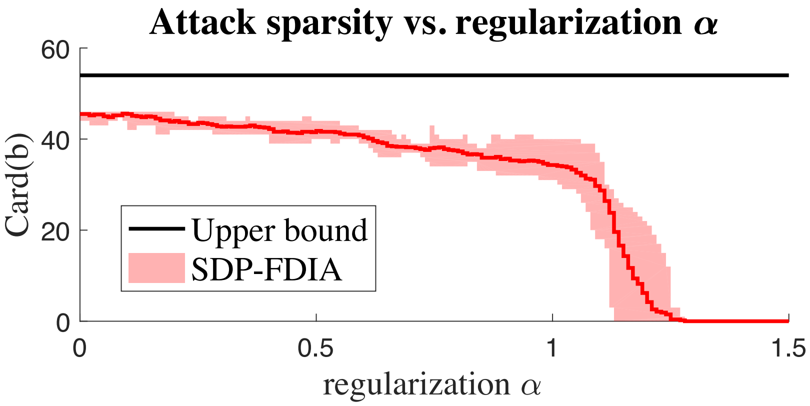


Figure 7.7: This plot shows the cardinality of the solution $\hat{\mathbf{b}}$ with respect to α. The upper bound is derived according to [86]. Ten independent experiments were performed to obtain the mean (red line) and min/max (shaded region).

bounds in Theorem 7.5.

As for the choice of $\mathbf{M}_0$, we set $\mathbf{M}_0 = -\mathbf{I} + \epsilon \mathbf{v}_{tg}\mathbf{v}_{tg}^* + \mathbf{L}_0$, for a matrix $\mathbf{L}_0$ that satisfies the following properties: 1) $\mathbf{L}_0 \succeq 0$, 2) 0 is a simple eigenvalue of $\mathbf{L}_0$, 3) the vector $\mathbf{v}_{tg}$ belongs to the null space of $\mathbf{L}_0$ (outlined in the proof of Theorem 7.3). The matrix $\mathbf{L}_0$ is obtained via the standard Gram-Schmidt procedure by starting with the target $\mathbf{v}_{tg}$. For the choice of ϵ, the proof of Theorem 7.3 (Appendix A.5) provides a guideline to use the equation $\epsilon = \frac{1}{\mathbf{v}_{tg}^*\hat{\mathbf{v}}}$; while $\hat{\mathbf{v}}$ cannot be known *a priori*, it is desirable to be close to $\mathbf{v}_{tg}^*$. Therefore, for the 30-bus system, a value of ϵ that leads to a rank-1 solution is close to 1/30≈0.033. In addition, the algorithm has been tested on several other power systems, with parameters listed in

Table 7.1. According to the results, the constructed FDIA attack can always evade BDD detection with ϵ close to $1/n_b$. Indeed, the measurement residuals are all on the order of 0.001, which are much lower than the BDD detection threshold. As for the sparsity, we have found that the cardinality Card($\hat{\mathbf{b}}$) is lower than the upper bound by [86] at the obtained scale of attack.

As the analysis shows, by having access to the sensor measurements, the adversary can solve (SDP-FDIA) to obtain a sparse attack vector. To thwart FDIA, a set of security sensors may need to be placed at locations under potential attack as indicated by $\hat{\mathbf{b}}$ of (SDP-FDIA). For any power system, the cardinality of a potential FDIA stealth attack can be used to indicate the vulnerability of the system against potential cyber threat [198].

7.5 Chapter summary

This chapter analyzed the vulnerability of power system AC-based state estimation against a critical class of cyberattacks known as false data injection attack. Since constructing an FDIA against AC-based state estimation requires solving a highly nonconvex problem, it is often believed that such attacks could be easily detected. However, this study showed that a near-globally optimal stealth attack can be found efficiently for a general scenario through a novel convexification framework based on SDP, where the measurement set could include nodal voltage magnitudes, real and reactive power injections at buses, and power flows over branches. We further analyzed the "attackable region" and derives performance bounds for a given set of measurement types and grid topology, where an attacker can plan an attack in polynomial time with limited resources.

The key insight from this chapter extends the spectrum of h-CPS data analytics in the following aspect: to learn about people, we want to collect as little data as possible to minimize sensor and labor costs (Part I of this thesis); to learn about system state, sensor redundancy can guard against sensor faults and ensure estimation integrity; and from a grid protection point of view, the results of this chapter can be used to design a security metric for the current practice against cyberattacks, redesign the bad data detection scheme, and inform proposals of grid hardening. Above all, the proposed convexification method and its associated theoretical analysis can be applied to other large-scale nonconvex problems in power systems and beyond.

Chapter 8

Conclusion and future directions

By mutually strengthening and converging technological advancements in sensing, learning, control and optimization, future human-cyber-physical systems will embody effective measures to address pressing issues such as global warming, environmental pollution, poverty, aging populations, and the fuel and food shortages collectively faced by human societies. ***Productivity*** will be enhanced by the proliferation of robots and automation, complementing and augmenting human forces to further improve ***efficiency*** and ***cost-effectiveness*** (e.g., see Fig. 8.1a for a manufacturing system example). The paradigm shift towards ***agile operation*** vis-à-vis lean production meets the needs of less predictable environments when volume is low and variability is high (e.g., customized services in smart buildings, as illustrated in Fig. 8.1c). Similarly, ***flexibility*** enhanced by smart architecture supports fast reconfiguration and response to changes (e.g., Fig. 8.1b for real-time pricing and automated responses in a smart grid). Along with the trend of interdependence among critical infrastructures, ***decentralization*** is key to improving system robustness and efficiency (e.g., distributed energy resources and the emergence of micro- and nano-grids). ***Safety*** and ***resilience*** have also become critical concerns as systems shift towards AI and data-driven automation (e.g., 8.1d for safety-oriented design in roadways and autonomous vehicles). Above all, because h-CPSs are fundamentally designed to serve people, ***human-centric*** values such as comfort, health and well-being will be continually promoted and optimized.

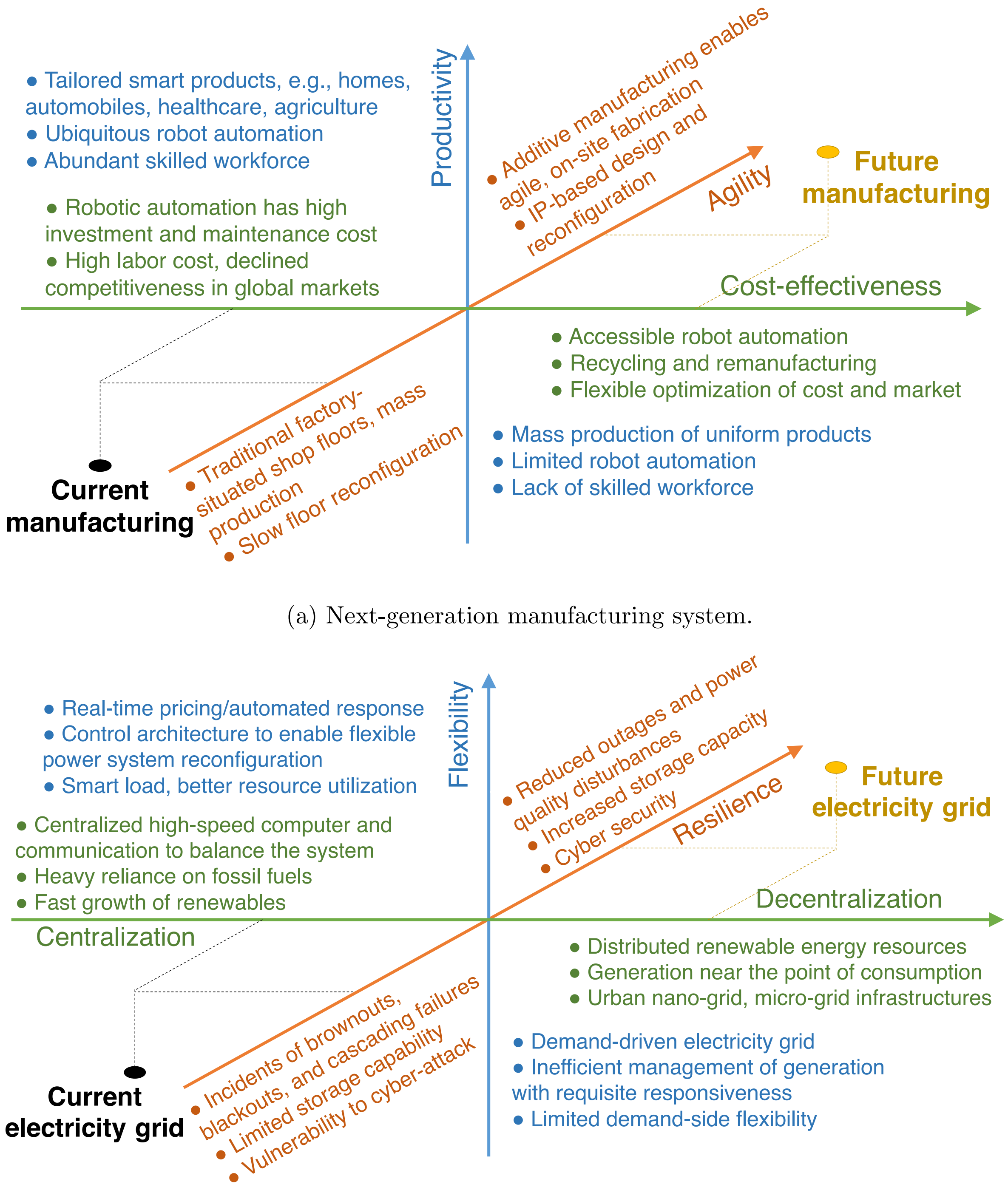
This thesis comprises a key step towards envisioned future optimal human-cyber-physical systems. Because people are central to an h-CPS, the first part of this thesis was dedicated to learning about human factors, including human behaviors and preferences. The goal is to deliver human-centric services and to facilitate interactive and cooperative controls while keeping humans in the loop. However, a main challenge in the learning task is the lack of labeled data due to cost and privacy concerns. To address this issue, we explored the physics-inspired sensing by proxy approach in Chap. 2, which determines human occupancy by measuring its impact on an indoor environment modeled by constitutive equations. On one hand, the resulting algorithm alleviates the need for ground truth collection and responds more quickly than existing methods to changes in occupancy. On the other hand, even without preexisting labels, data-driven algorithms can be employed under the framework of “weak supervision” discussed in Chap. 3. The core idea is to leverage high-level heuristics from domain knowledge to create noisy “weak labels,” and algorithmically resolve the inconsistencies by iterative refinement of the labels or by redesign of the loss function to account for the noise. When inferring human intentions and preferences, data is tricky to obtain due to survey biases and people’s internal inconsistencies. In this regard, our key insight is to abstract people’s preferences as a function that can rationalize their behaviors. Drawing on this insight, we developed an inverse game theory framework that determines people’s utility functions by observing how they interact with one another in a social game to conserve energy (Chap. 4). Because people’s actions often involve long-term planning and their motivations depend on factors that cannot be known *a priori*, learning should span multiple time scales and account for complex rewards. Along this aspect, we explored deep Bayesian inverse reinforcement learning, which simultaneously learns the motivator representation to expand the capacity of modeling complex rewards and rationalizes an agent’s sequence of actions to infer its long-term goals (Chap. 5). While the methods in Chap. 4 and Chap. 5 work in different settings (i.e., a gamified environment and long-term planning, respectively), both have been shown to have high data efficiency, which can enable wider applications in h-CPSs.

Enabled by the context awareness of the human, cyber-, and physical- components, the second part of this thesis explored methods to analyze and enhance system-level efficiency and resilience. Chap. 6 explored the next-generation energy retail model to enable distributed resource energy utilization and to exploit demand-side flexibility. The synergy that naturally emerges from integrated optimization of both thermal and electrical energy provision is able to substantially improve efficiency and reduce generation costs. While data empowers h-CPS learning and control to gain context awareness and to enhance efficiency, malicious attacks on data integrity can pose major security threats. Chap. 7 discussed the cyber resilience of power system state estimation, a key procedure for power grid operation. Although an adversarial attack on the more accurate AC-model state estimation is non-convex, an approach based on semidefinite programming relaxation produces a near-global optimal attack. The envisioning process naturally leads to a resilience metric for power grids, and can inform upgrades of bad data detection schemes to enhance cyber resilience. In the following, we discuss some key challenges in h-CPS data analytics as well as opportunities and future directions to address these challenges.

8.1 Challenges and opportunities in h-CPS

Human-centric learning and control

Applications in h-CPS that consider user behaviors (e.g., a smart home assistant that learns a user’s moods and activities) and preferences (e.g., utility companies that learn a user’s response to economic incentives) and that are aimed at improving human-centric values (e.g., comfort, well-being, health and productivity) are in increasing demand. Yet meeting

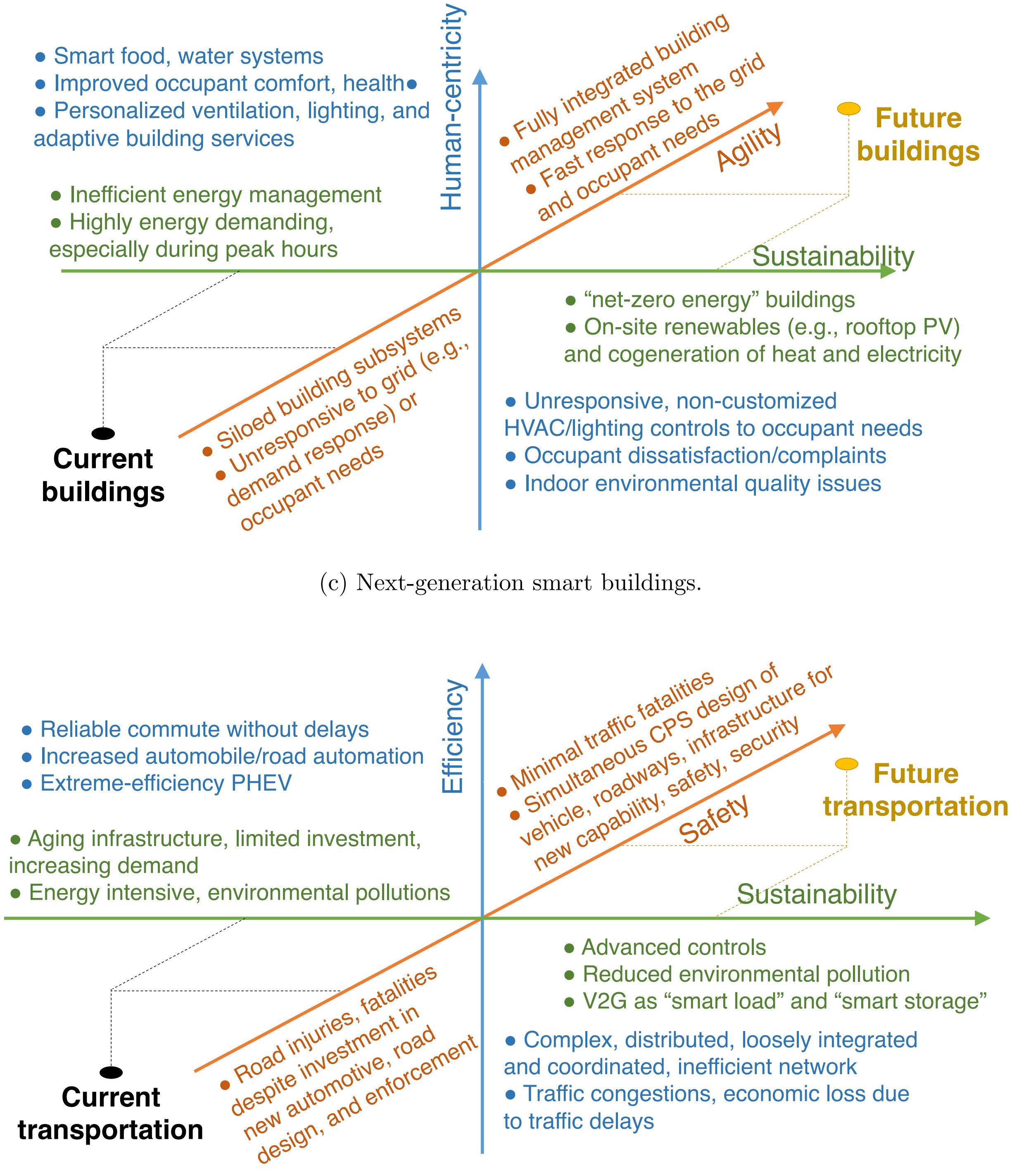


Productivity
• Tailored smart products, e.g., homes, automobiles, healthcare, agriculture
• Ubiquitous robot automation
• Abundant skilled workforce
• Robotic automation has high investment and maintenance cost
• High labor cost, declined competitiveness in global markets
• Additive manufacturing enables agile, on-site fabrication
• IP-based design and reconfiguration
Agility
Future manufacturing
Cost-effectiveness
• Accessible robot automation
• Recycling and remanufacturing
• Flexible optimization of cost and market
• Mass production of uniform products
• Limited robot automation
• Lack of skilled workforce
Current manufacturing
• Traditional factory-situated shop floors, mass production
• Slow floor reconfiguration

(a) Next-generation manufacturing system.

Flexibility
• Real-time pricing/automated response
• Control architecture to enable flexible power system reconfiguration
• Smart load, better resource utilization
• Centralized high-speed computer and communication to balance the system
• Heavy reliance on fossil fuels
• Fast growth of renewables
Centralization
• Reduced outages and power quality disturbances
• Increased storage capacity
• Cyber security
Resilience
Future electricity grid
Decentralization
• Distributed renewable energy resources
• Generation near the point of consumption
• Urban nano-grid, micro-grid infrastructures
• Demand-driven electricity grid
• Inefficient management of generation with requisite responsiveness
• Limited demand-side flexibility
Current electricity grid
• Incidents of brownouts, blackouts, and cascading failures
• Limited storage capability
• Vulnerability to cyber-attack

(b) Next-generation electricity grid.


● Smart food, water systems
● Improved occupant comfort, health●
● Personalized ventilation, lighting, and adaptive building services
Human-centricity
● Fully integrated building management system
● Fast response to the grid and occupant needs
Agility
Future buildings
● Inefficient energy management
● Highly energy demanding, especially during peak hours
Sustainability
● "net-zero energy" buildings
● On-site renewables (e.g., rooftop PV) and cogeneration of heat and electricity
● Siloed building subsystems
● Unresponsive to grid (e.g., demand response) or occupant needs
Current buildings
● Unresponsive, non-customized HVAC/lighting controls to occupant needs
● Occupant dissatisfaction/complaints
● Indoor environmental quality issues

(c) Next-generation smart buildings.

● Reliable commute without delays
● Increased automobile/road automation
● Extreme-efficiency PHEV
Efficiency
● Minimal traffic fatalities
● Simultaneous CPS design of vehicle, roadways, infrastructure for new capability, safety, security
Safety
Future transportation
● Aging infrastructure, limited investment, increasing demand
● Energy intensive, environmental pollutions
Sustainability
● Advanced controls
● Reduced environmental pollution
● V2G as "smart load" and "smart storage"
● Road injuries, fatalities despite investment in new automotive, road design, and enforcement
Current transportation
● Complex, distributed, loosely integrated and coordinated, inefficient network
● Traffic congestions, economic loss due to traffic delays

(d) Next-generation transportation system.

Figure 8.1: Societal-scale human-cyber-physical systems (e.g., smart building, power grid, manufacturing and transportation) are under transformation to enhance efficiency, cost-effectiveness, productivity, agility, flexibility, safety, resilience, and human-centric values.

this demand often requires collecting a vast quantity of sensitive personal data and making decisions that might affect people on an individual level.

> *The challenge is to infer and respond to human intentions and preferences while ensuring privacy, fairness and accountability.*

In this thesis, we have taken a step toward human-centric learning and control through the focus on data-efficient analytics as described in Part I. Future directions include understanding and mitigating privacy concerns for sensitive information (e.g., using notions of differential privacy [58], pan-privacy [59], local privacy [56] and free privacy [87]) and unintended discrimination in decision making [169]. To hold AI-enabled applications accountable, the algorithms need to embrace openness, transparency and interpretability, and they must work or interact effectively with people while observing regulations and standards.

Learning in large-scale nonconvex optimization

Many control and planning problems in h-CPSs involve large numbers of continuous and discrete variables and constraints and must be solved regularly within a limited time budget. For instance, the optimal power flow problem is a fundamental problem with thousands of variables and constraints that needs to be solved every 5 to 15 minutes to balance the supply and demand. However, due to the nonlinearity of the physics that relate the bus voltages to complex powers, the problem is highly nonlinear and nonconvex and quickly becomes formidable (even with the best state-of-the-art solvers), especially when *N-1 contingency planning* is considered, which adds millions of constraints to an already difficult problem.

> *The challenge is to solve highly nonlinear problems at much larger scale in a much shorter time horizon with higher accuracy.*

One important direction to pursue is lossless convexification, which transforms an original nonconvex problem into a convex formulation with provable optimality guarantees, as discussed in Chap. 7 for the semidefinite programming relaxation technique. We can also leverage learning techniques, both by simplifying the original problems (e.g., active constraint screening, linearization) and guiding the search for the optimal solution (e.g., learned initialization, adaptive step size selection).

Adaptable and safe operation in dynamic environments

Because the real world changes continually—often rapidly and unexpectedly—h-CPS operation must adapt to dynamic environments in safe and reliable ways. For example, in a power grid, when the environment changes because of either internal conditions (e.g., generator trips or line breaks, sensor failures) or external conditions (e.g., inclement weather, malicious sabotage), the grid must coordinate and respond with respect to the particular change. Similarly, a smart building should quickly respond to unexpected events (e.g., an emergency or a power outage) by learning and adapting in real time and by cooperating with human operators.

> *The challenge is to make decisions with limited computational resources (e.g., on-board chips) and within short time spans (e.g., seconds or microseconds) while ensuring the safety and reliability of the subsequent actions.*

To tackle this challenge, algorithms must support continual or life-long learning that constantly evolves in response to and interacts with the environment [162], [202], [217]. It should be able to generalize to multiple tasks, even previously unseen tasks, by efficiently transferring and utilizing knowledge from already learned skills (e.g., model-agnostic meta-learning [65], fast reinforcement learning via slow reinforcement learning [55], and learning to reinforcement learning [224]). Above all, such algorithms must be complex enough to address complex situations (e.g., deep reinforcement learning [163]) yet simple enough to be implemented on low-cost hardware and robust enough to provide theoretical safety guarantees.

Interactive and cooperative operation keeping humans in the loop

Depending on the task complexity and the requirements for safety, reliability and accountability, h-CPS applications range from being human-operated to semi-autonomous to fully-autonomous. For example, while the power system state estimation has been automated, in the face of a potential cyberattack (as discussed in Chap. 7), the system should warn a human operator of the possibility of intrusion and provide information regarding the intrusion pattern so the human operator can assist with mitigation. Similarly, an automated system should cooperate with human operators during emergencies to help steer people away from the danger zone as quickly as possible.

> *The challenge is to seamlessly and interactively include humans in the control loop, to augment and/or leverage human decision making capabilities.*

To tackle this challenge, an AI system should be able to infer and respect human intention and preferences, and it should occasionally explain its decisions to its human counterparts to seek approval or guidance. The inverse game theory (Chap. 4) and deep Bayesian inverse reinforcement learning (Chap. 5) are steps toward more accurate preference inference. Achieving explainable actions requires AI systems to go beyond black-box predictions and decisions, reaching a level that can identify the features of inputs most responsible for particular decisions, support interactive analysis and answer counterfactual questions [209]. This capability would dramatically increase the usability of AI in h-CPS operation and control.

Resilience of interdependent infrastructures

There is a trend to integrate critical infrastructures such as energy, water, agriculture, transportation and communication systems to achieve higher efficiencies. Even within an energy system, the integration of natural gas, electricity and thermal energy networks can dramatically reduce energy generation costs, as discussed in Chap. 6.

> *The challenge is to ensure the resilience of critical infrastructures as they become increasingly interdependent and reliant on shared data and services.*

To enhance resilience, a systematic approach can be adopted that involves preparing the system for possible stresses or attacks (e.g., resilience by design and the envisioning process described in Chap. 7), relying on resources to ameliorate the consequences of an event after it has occurred (e.g., a mitigation strategy), recovering as quickly as possible after the event is over (e.g., black-start mechanism), and remaining alert to insights and lessons for future events (e.g., hindsight evaluation) [68].

Robust and secure decision-making

Applications in h-CPSs involve components that interact through complex, coupled physical environments. For example, decisions for a power grid must be made at multiple time scales that account for stochastic behavior due to renewable energy resources, variable demand and unplanned outages. In addition, modern analytics fuse not only information from trustworthy central sources, but also data from untrusted crowd-sourced third-parties.

> *The challenge is to ensure the robustness and security of the decision-making process in the face of uncertainty, faulty data and malicious attacks.*

To tackle this issue, the first step is to understand the scenarios in which the decision making is not robust or not secure (e.g., the envisioning of a potential cyberattack on the state estimator in Chap. 7). Drawing from the insights gained in the envisioning process, one can design systems that track data provenance, and combine both hardware (e.g., device fingerprints) and software (e.g., bad data detection with contextual information) algorithms to enhance the reliability of the results. Similarly, the system can rely on interactive and cooperative decision making with humans to leverage the complementary capability.

8.2 Closing thoughts

The methods presented in this thesis attempt to combine sensing, learning, control, optimization, game theory and robotics to empower data-efficient analytics in h-CPSs. As discussed in this chapter, many challenges remain to be addressed and ample opportunities remain to be undertaken. To achieve the envisioned optimal h-CPS and provide a significant societal impact, these challenges and opportunities will require strong collaboration among various communities, disciplines and stakeholders, a thorough understanding of the social, economic and regulatory barriers and implications, and a sustained effort in education, research and entrepreneurship.

Appendix A

Main proofs and derivations

A.1 Chapter 2: Sensing by proxy

Theorem 2.1. *Consider the system (2.10)–(2.12), where*

$$\mathbf{r}(s) = \begin{bmatrix}\mathbf{r}_1(s) & \cdots & \mathbf{r}_m(s)\end{bmatrix}^\top \tag{A.1}$$

$$\mathbf{r}_i(s) = \left(\mathbf{C}_i - \int_0^{(1-s)/b_i} [\mathbf{B}_Z]_{i,:} e^{-\bar{\mathbf{A}}y} dy\right) e^{\bar{\mathbf{A}}(1-s)/b_i} \tag{A.2}$$

$$\mathbf{C}_i = \int_0^{1/b_i} [\mathbf{B}_Z]_{i,:} e^{-\bar{\mathbf{A}}\sigma} d\sigma, \quad i = 1, ..., m \tag{A.3}$$

Let the pair $(\bar{\mathbf{A}}, \bar{\mathbf{C}})$ *be observable, where* $\bar{\mathbf{C}} = \begin{bmatrix}\mathbf{C}_1 \\ \vdots \\ \mathbf{C}_m\end{bmatrix} \in \mathbb{R}^{m\times 2m}$*, and choose* $\mathbf{L}$ *such that the matrix* $\bar{\mathbf{A}} - \mathbf{L}\bar{\mathbf{C}}$ *is Hurwitz. Then, for any* $\mathbf{z}(0) \in \mathbb{R}^{2m}$*,* $u_i(s,t)$*,* $\hat{u}_i(s,t) \in L^2(0,1)$*,* $i = 1, ..., m$*, where* u_i *is the i-th component of* $\mathbf{u}$*, there exist positive constants* λ *and* κ *such that the following holds for all* $t \geq 0$

$$\Omega(t) \leq \kappa\Omega(0)e^{-\lambda t}, \tag{A.4}$$

where

$$\Omega(t) = \int_0^1 \|\mathbf{u}(s,t) - \hat{\mathbf{u}}(s,t)\|^2 ds + \|\mathbf{z}(t) - \hat{\mathbf{z}}(t)\|^2. \tag{A.5}$$

Proof. Consider the following MIMO, LTI system with distributed sensor delays:

$$\dot{\mathbf{z}}(t) = \bar{\mathbf{A}}\mathbf{z}(t) \tag{A.6}$$

$$y_i(t) = \int_0^{D_i} \mathbf{Q}_i \mathbf{z}(t-\sigma) d\sigma, \text{ for } i = 1, ..., m, \tag{A.7}$$

where $\mathbf{z}(t) = \begin{bmatrix}\mathbf{x}(t)\\ \mathbf{v}(t)\end{bmatrix} \in \mathbb{R}^{2m}$, $\bar{\mathbf{A}} = \begin{bmatrix}-\mathbf{A} & \mathbf{I}_{m\times m}\\ \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m}\end{bmatrix}$, $\mathbf{Q}_i = [\mathbf{B}_Z]_{i,:} \in \mathbb{R}^{1\times 2m}$, $y_i(t) \in \mathbb{R}$, and $D_i \in \mathbb{R}_+$ is a delay. In the following we use $\omega^{(i)}$ to denote the i-th entry of the vector $\boldsymbol{\omega}$. System (A.6) and (A.7) can be written equivalently as

$$\dot{\mathbf{z}}(t) = \bar{\mathbf{A}}\mathbf{z}(t) \tag{A.8}$$

$$y_i(t) = \omega^{(i)}(0,t), \text{ for } i = 1, ..., m, \tag{A.9}$$

where

$$\omega_t^{(i)}(l,t) = \omega_l^{(i)}(l,t) + \mathbf{Q}_i\mathbf{z}(t) \tag{A.10}$$

$$\omega_t^{(i)}(D_i,t) = 0, \text{ for } i = 1, ..., m. \tag{A.11}$$

One can see this by noting that the solution to (A.10) and (A.11) is

$$\omega_t^{(i)}(l,t) = \int_l^{D_i} \mathbf{Q}_i\mathbf{z}(t+l-\sigma)d\sigma.$$

We show next that system (2.3)–(2.7) can be written in the form of system (A.8)–(A.11), and hence, one can apply the results of [17, Thm. 2]. Define the spatial variable $l = \frac{1-s}{b_i}$, $D_i = 1/b_i$, $u^{(i)}(1-b_il,t) = \omega(l,t)$, we can write system (2.3)–(2.7) as

$$\dot{\mathbf{z}}(t) = \bar{\mathbf{A}}\mathbf{z}(t) \tag{A.12}$$

$$\omega_t^{(i)}(l,t) = \omega_l^{(i)}(l,t) + \mathbf{Q}_i\mathbf{z}(t) \tag{A.13}$$

$$\omega_t^{(i)}(D_i,t) = U_0^{(i)}(t) \tag{A.14}$$

$$\omega_t^{(i)}(0,t) = U_1^{(i)}(t), \text{ for } i = 1, ..., m. \tag{A.15}$$

System (A.12)–(A.15) is of the form (A.8), (A.10) and (A.11) with the difference of the nonhomogeneous boundary condition at $l = 0$ and D_i. However, the result in [17] applies with the trivial modification to account for the additional measured inputs. The observer (2.10)–(2.12) can be written in the $\hat{\boldsymbol{\omega}}$ variable as:

$$\hat{\omega}_t^{(i)}(l,t) = \hat{\omega}_l^{(i)}(l,t) + \mathbf{Q}_i\hat{\mathbf{z}}(t) + [\mathbf{r}(1-lb_i)\mathbf{L}\left(\mathbf{U}_1(t) - \hat{\boldsymbol{\omega}}(0,t)\right)]_i \tag{A.16}$$

$$\hat{\omega}^{(i)}(D_i,t) = U_0^{(i)}(t), \text{ for } i = 1, ..., m, \tag{A.17}$$

$$\dot{\hat{\mathbf{z}}}(t) = \bar{\mathbf{A}}\hat{\mathbf{z}}(t) + \mathbf{L}\big(\mathbf{U}_1(t) - \hat{\boldsymbol{\omega}}(0,t)\big). \tag{A.18}$$

The stability proof of [17, Thm. 2] is based on the dynamics of the observer errors $\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}$ and $\mathbf{z} - \hat{\mathbf{z}}$. Combining (A.12)–(A.15) with (A.16)–(A.18), and let $\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega} - \hat{\boldsymbol{\omega}}$ denote the observer error, we obtain that

$$\tilde{\omega}_t^{(i)}(l,t) = \tilde{\omega}_l^{(i)}(l,t) + \mathbf{Q}_i\tilde{\mathbf{z}}(t) - [\mathbf{r}(1-lb_i)\mathbf{L}\tilde{\boldsymbol{\omega}}(0,t)]_i \tag{A.19}$$

$$\tilde{\omega}^{(i)}(D_i,t) = 0, \text{ for } i = 1, ..., m, \tag{A.20}$$

$$\dot{\tilde{\mathbf{z}}}(t) = \bar{\mathbf{A}}\tilde{\mathbf{z}}(t) - \mathbf{L}\tilde{\boldsymbol{\omega}}(0,t), \tag{A.21}$$

which is the same error system as [17]. Since the pair $(\bar{\mathbf{A}}, \bar{\mathbf{C}})$ is observable, one can choose $\mathbf{L}$ such that the matrix $\bar{\mathbf{A}} - \mathbf{L}\bar{\mathbf{C}}$ is Hurwitz. One can apply [17, Thm. 2] to show that the observer (2.10)–(2.12) is stable. □

Corollary 2.1. *Consider the system* (2.18)*–*(2.22) *and the observer* (2.23)*–*(2.27)*. Let* $b_X \neq 0$ *and choose* L_1*,* L_2 *such that the matrix* $\bar{\mathbf{A}} - \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \mathbf{C}_1$ *is Hurwitz, where* $\bar{\mathbf{A}} = \begin{bmatrix} -a & 1 \\ 0 & 0 \end{bmatrix}$*, and* $\mathbf{C}_1 = \begin{bmatrix} \pi_1(1) & \pi_2(1) \end{bmatrix}$*. Then for any* $x(0)$*,* $\hat{x}(0)$*,* $v(0)$*,* $\hat{v}(0) \in \mathbb{R}$*, there exists positive constant* λ *and* κ *such that the following holds for all* $t \geq 0$*,*

$$\Omega(t) \leq \kappa \Omega(0) e^{-\lambda t} \tag{A.22}$$

$$\Omega(t) = \int_0^1 \big(u(s,t) - \hat{u}(s,t)\big)^2 ds + \big(x(t) - \hat{x}(t)\big)^2 + \big(v(t) - \hat{v}(t)\big)^2 \tag{A.23}$$

Proof. Recap that $\mathbf{r}(s) = \begin{bmatrix} \pi_1(s) & \pi_2(s) \end{bmatrix}$, $\pi_1(s) = \frac{b_X}{a}\big(e^{\frac{a}{b}s} - 1\big)$, and $\pi_2(s) = \frac{b_X}{ba}s + \frac{b_X}{a^2}\big(1 - e^{\frac{a}{b}s}\big)$. We show that the observability condition of the pair $\left(\bar{\mathbf{A}}, \int_0^{1/b_1} \begin{bmatrix} b_X & 0 \end{bmatrix} e^{-\bar{\mathbf{A}}\sigma} d\sigma\right)$ in Theorem 2.1 with only one sensor is equivalent to the observability condition of the pair $(\bar{\mathbf{A}}, \mathbf{C}_1)$ in this corollary. This follows by

$$\begin{aligned} \int_0^{1/b_1} \begin{bmatrix} b_X & 0 \end{bmatrix} e^{-\bar{\mathbf{A}}\sigma} d\sigma &= \frac{b_X}{a} \int_0^{1/b} \begin{bmatrix} a e^{a\sigma} & 1 - e^{a\sigma} \end{bmatrix} d\sigma \\ &= \frac{b_X}{a} \begin{bmatrix} e^{a/b} - 1 & \frac{1}{b} + \frac{1}{a}(1 - e^{a/b}) \end{bmatrix} = \mathbf{C}_1 \end{aligned}$$

To show that $(\bar{\mathbf{A}}, \mathbf{C}_1)$ is observable, note that the determinant of the observability matrix $\mathcal{O}$ is $\det(\mathcal{O}) = \pi_1(1)(\pi_1(1) + a\pi_2(1))$. It follows that $\det(\mathcal{O}) \neq 0$ whenever $b_X \neq 0$. The rest of the proof follows by the proof of Theorem 2.1. □

A.2 Chapter 3: Learning under weak supervision

Multi-view iterative training

Lemma 3.1. *The training noise rate* η_t *and classification error rate* ϵ_t *can be estimated with the access to any two of the following (approximated) quantities:*

1. *The number of negative samples in the dataset* $|\mathcal{L}^t_{-1,\checkmark}| + |\mathcal{L}^t_{+1,\times}| + |\mathcal{U}^t_{-1}|$
2. *The number of negative samples in the labeled set* $|\mathcal{L}^t_{-1,\checkmark}| + |\mathcal{L}^t_{+1,\times}|$
3. *The number of positive samples in the dataset* $|\mathcal{L}^t_{+1,\checkmark}| + |\mathcal{L}^t_{-1,\times}| + |\mathcal{U}^t_{+1}|$
4. *The number of positive samples in the labeled set* $|\mathcal{L}^t_{+1,\checkmark}| + |\mathcal{L}^t_{-1,\times}|$

5. *The misclassification rate for the positive samples* $|\mathcal{L}^t_{-1,\times}| / \left(|\mathcal{L}^t_{-1,\times}| + |\mathcal{L}^t_{+1,\checkmark}|\right)$

6. *The misclassification rate for the negative samples* $|\mathcal{L}^t_{+1,\times}| / \left(|\mathcal{L}^t_{+1,\times}| + |\mathcal{L}^t_{-1,\checkmark}|\right)$

Proof. According to the update rule:

$$\mathcal{L}^{t+1}_y = \{\mathcal{L}^t_y \cap \hat{\mathcal{L}}_y\} \cup Sample\{\mathcal{L}^t_y \Delta \hat{\mathcal{L}}_y; \alpha_y\}, \tag{A.24}$$

the expected number of elements in the labeled set of the next iteration depends on the current iteration as follow:

$$|\mathcal{L}^{t+1}_{-1,\checkmark}| = |\mathcal{L}^t_{-1,\checkmark}|(1-\epsilon_t) + (|\mathcal{L}^t_{+1,\times}| + |\mathcal{U}^t_{-1}|)(1-\epsilon_t)\alpha_{-1} \tag{A.25}$$

$$|\mathcal{L}^{t+1}_{-1,\times}| = |\mathcal{L}^t_{-1,\times}|\epsilon_t + (|\mathcal{L}^t_{+1,\checkmark}| + |\mathcal{U}^t_{+1}|)\epsilon_t\alpha_{-1} \tag{A.26}$$

$$|\mathcal{L}^{t+1}_{+1,\checkmark}| = |\mathcal{L}^t_{+1,\checkmark}|(1-\epsilon_t) + (|\mathcal{L}^t_{-1,\times}| + |\mathcal{U}^t_{+1}|)(1-\epsilon_t)\alpha_{+1} \tag{A.27}$$

$$|\mathcal{L}^{t+1}_{+1,\times}| = |\mathcal{L}^t_{+1,\times}|\epsilon_t + (|\mathcal{L}^t_{-1,\checkmark}| + |\mathcal{U}^t_{-1}|)\epsilon_t\alpha_{+1} \tag{A.28}$$

Since we can observe the number of samples in $|\mathcal{L}^t_{-1}| = |\mathcal{L}^t_{-1,\checkmark}| + |\mathcal{L}^t_{-1,\times}|$, $|\mathcal{L}^t_{+1}| = |\mathcal{L}^t_{+1,\checkmark}| + |\mathcal{L}^t_{+1,\times}|$, and $|\mathcal{U}^t| = |\mathcal{U}^t_{-1}| + |\mathcal{U}^t_{+1}|$, and also for those in round $t+1$, we can sum the pairs of (A.25, A.26), also (A.27, A.28). Together with two of the quantities proposed in Lemma 3.1, we can solve the system of equations for the estimation of ϵ_t and η_t. □

As a general remark, the system of equations to be solved in Lemma 3.1 is non-linear, which makes it computationally costly to solve. Since the problem is defined for $0 \le \epsilon_t \le 1$, we can perform a line search of ϵ_t. Given the value of ϵ_t, the system becomes linear and is very easy to solve by taking the inverse, or constrained quadratic programming. Then the optimal ϵ_t that corresponds to the solution that best fits the remaining single equation should be chosen.

Theorem 3.2. *The gap between the learned and optimal hypotheses in the PAC property (3.6) will decrease with high probability in each iteration with suitable sampling rates, α_{-1} and α_{+1}, whenever the following condition is satisfied:*

$$\left(|\mathcal{L}^{t+1}_{-1}| + |\mathcal{L}^{t+1}_{+1}|\right)(1-2\eta_{t+1})^2 > \left(|\mathcal{L}^t_{-1}| + |\mathcal{L}^t_{+1}|\right)(1-2\eta_t)^2 \tag{A.29}$$

where $\left(|\mathcal{L}^{t+1}_{-1}| + |\mathcal{L}^{t+1}_{+1}|\right)$ is the total number of weakly labeled samples in round $t+1$, and η_{t+1} is the (estimated) training noise rate.

Proof. (Sketch) Let $c = 2\mu \log\left(\frac{2N}{\delta}\right)$ where μ is chosen to make the equality holds in the PAC property (3.5), then we have $n_t = \frac{c}{\epsilon_t^2(1-2\eta_t)}$, where $n_t = |\mathcal{L}^{t+1}_{-1}| + |\mathcal{L}^{t+1}_{+1}|$ is the number of samples in the labeled set. We introduce u_t as follows for the simplicity of computation:

$$u_t = \frac{c}{\epsilon_t^2} = n_t(1-2\eta_t)^2 \tag{A.30}$$

Since u_t is proportional to $1/\epsilon_t^2$, we have $\epsilon_{t+1} < \epsilon_t$ satisfied as long as $u_{t+1} > u_t$, thus the claim is proved. □

Derivation of the surrogate loss

The supervised learning is described by $(l, \mathcal{F}, e_n)$, where $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is the *loss function* to penalize misdetection, $\mathcal{F}$ is the class of classifiers, $e_n : \mathcal{D} \to (\mathcal{X}, \mathcal{Y})^n$ is the repetitive experiments performed to acquire the dataset, $S = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\} \sim e_n(\mathcal{D})$, and $\mathcal{D}$ is the data distribution.

To study the mechanism of weak label initialization, we introduce the corruption process $\mathcal{T} : \mathcal{O} \to \tilde{\mathcal{O}}$ as a Markov kernel, which corrupts the outcome $\mathcal{O}$ of the experiments to be $\tilde{\mathcal{O}}$, i.e., $\tilde{e}_n = \mathcal{T}(e_n)$. Each Markov kernel is associated with a linear mapping, $\mathcal{T} : (\mathbb{R}^{\mathcal{O}})^* \to (\mathbb{R}^{\tilde{\mathcal{O}}})^*$, where $(\mathbb{R}^{\mathcal{O}})^*$ is the dual space of $(\mathbb{R}^{\mathcal{O}})$ for linear functionals. Weakly supervised learning is characterized by $(l, \mathcal{F}, \tilde{e}_n)$, as compared to the supervised learning.

Definition A.1 (Reconstructible Markov kernel)**.** *The Markov kernel $\mathcal{T} : \mathcal{O} \to \tilde{\mathcal{O}}$ is reconstructible if there exists a linear mapping $\mathcal{Q} : (\mathbb{R}^{\tilde{\mathcal{O}}})^* \to (\mathbb{R}^{\mathcal{O}})^*$, such that $\mathcal{Q}\mathcal{T} = 1$, where $\mathcal{Q}$ is known as the* reconstruction*.*

An immediate consequence of the reconstructible property is that we have:

$$\langle \mathcal{D}, l(\cdot, f(\cdot)) \rangle = \langle \mathcal{Q}\mathcal{T}(\mathcal{D}), l(\cdot, f(\cdot)) \rangle = \langle \mathcal{D}, \mathcal{Q}^*(l(\cdot, f(\cdot))) \rangle \tag{A.31}$$

where $\mathcal{D}$ is the original data distribution, $\mathcal{T}(\mathcal{D})$ is the corrupted distribution, $\mathcal{Q}^*(l(\cdot, f(\cdot)))$ is the corruption corrected loss function, and $\langle \mathcal{D}, l(\cdot, f(\cdot)) \rangle = \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}} l(y, f(\mathbf{x}))$ is the expectation under the distribution $\mathcal{D}$. The above property implies that working with the corrupted data with $\mathcal{Q}^*(l(\cdot, f(\cdot)))$ is equivalent to using the clean data with the original loss function $l(\cdot, f(\cdot))$ associated with learner $f \in \mathcal{F}$.

Since we are starting with noisy labels estimated by the occupancy schedules, the corruption process is characterized by $\rho_{+1} = \mathbb{P}(\tilde{y} = -1 | y = +1)$ and $\rho_{-1} = \mathbb{P}(\tilde{y} = +1 | y = -1)$; therefore, we can specify the Markov kernel $\mathcal{T}$ and $\mathcal{Q}^*$ as:

$$\mathcal{T} = \begin{pmatrix} 1 - \rho_{-1} & \rho_{+1} \\ \rho_{-1} & 1 - \rho_{+1} \end{pmatrix}, \tag{A.32}$$

$$\mathcal{Q}^* = \frac{1}{1 - \rho_{-1} - \rho_{+1}} \begin{pmatrix} 1 - \rho_{+1} & -\rho_{-1} \\ -\rho_{+1} & 1 - \rho_{-1} \end{pmatrix} \tag{A.33}$$

where $\mathcal{Q}^*$ is the conjugate transpose of $\mathcal{Q}$, and it can be verified as the *reconstruction* of $\mathcal{T}$, i.e., $\mathcal{Q}\mathcal{T} = 1$. With elementary calculations, the surrogate loss (3.10) in the main text follows.

A.3 Chapter 4: Gamification meets inverse game theory

Proposition 1. *A differential Nash equilibrium of the p-person concave game $(f_1, \ldots, f_p)$ on $\mathcal{C}$ is a Nash equilibrium.*

Proof. Suppose the assumption holds that the constraints for each player do not depend on other players' choice variables. We can fix $\mathbf{x}^*_{-i}$ and apply Proposition 3.3.2 [18] to the i-th player's optimization problem

$$\max_{\mathbf{x}_i \in \mathcal{C}_i} f_i(\mathbf{x}_i, \mathbf{x}^*_{-i}; \boldsymbol{\gamma}). \tag{A.34}$$

Since each f_i is concave and each $\mathcal{C}$ is a convex set, $\mathbf{x}^*_i$ is a global optimum of the i-th player's optimization problem under the conditions of differential Nash equilibrium. Since this is true for each of the $i \in \mathcal{I}$ players, $\mathbf{x}^*$ is a Nash equilibrium. □

A.4 Chapter 5: Deep Bayesian inverse reinforcement learning

Details of the DGP-IRL model

DGP-IRL extends the deep GP framework to the IRL domain. DGP-IRL learns an abstract representation that reveals the reward structure by warping the original feature space through the latent layers, $\mathbf{D}, \mathbf{B}$. For a set of observed trajectories $\mathcal{M}$, our objective is to optimize the corresponding marginalized log-likelihood given the states in the world $\mathbf{X}$:

$$\log p(\mathcal{M}|\mathbf{X}) = \log \int p(\mathcal{M}|\boldsymbol{r}) p(\boldsymbol{r}|\mathbf{D}) p(\mathbf{D}|\mathbf{B}) p(\mathbf{B}|\mathbf{X}) d(\boldsymbol{r}, \mathbf{D}, \mathbf{B}) \tag{A.35}$$

where the integration is with respect to the latent layers, including the reward vector $\boldsymbol{r}$. As introduced in the main text, $\mathbf{d}^m \in \mathbb{R}^n$ is the m-th column of the latent layer $\mathbf{D} = \begin{bmatrix}\mathbf{d}^1 & \cdots & \mathbf{d}^{m_1}\end{bmatrix}$, and similarly for $\mathbf{B} = \begin{bmatrix}\mathbf{b}^1 & \cdots & \mathbf{b}^{m_1}\end{bmatrix}$:

$$p(\mathcal{M}|\boldsymbol{r}) = \sum_{i=1}^{h} \sum_{t=1}^{T} \big(Q(s_{i,t}, a_{i,t}; \boldsymbol{r}) - V(s_{i,t}; \boldsymbol{r})\big) \tag{A.36}$$

$$p(\boldsymbol{r}|\mathbf{D}) = \mathcal{N}(\boldsymbol{r}|\mathbf{0}, K_{\mathbf{DD}}) \tag{A.37}$$

$$p(\mathbf{D}|\mathbf{B}) = \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{d}^m|\mathbf{b}^m, \lambda^{-1}\mathbf{I}) \tag{A.38}$$

$$p(\mathbf{B}|\mathbf{X}) = \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{b}^m|\mathbf{0}, K_{\mathbf{XX}}) \tag{A.39}$$

where $p(\mathcal{M}|\boldsymbol{r})$ represents the reinforcement learning term, given by:

$$\log p(\mathcal{M}|\boldsymbol{r}) = \sum_{i} \sum_{t} (Q(s_{i,t}, a_{i,t}; \boldsymbol{r}) - V(s_{i,t}; \boldsymbol{r})) \tag{A.40}$$

$$= \sum_{t} \sum_{t} \left(\boldsymbol{r}_{s_{i,t}, a_{i,t}} - V(s_{i,t}; \boldsymbol{r}) + \sum_{s'} \gamma \mathcal{T}_{s'}^{s_{i,t}, a_{i,t}} V(s'; \boldsymbol{r}) \right) \tag{A.41}$$

The Q-value $Q(s_{i,t}, a_{i,t}; \boldsymbol{r})$ used above is a measure of how desirable is the corresponding state-action pair $(s_{i,t}, a_{i,t})$ under rewards $\boldsymbol{r}$ for all the world states, and is defined by:

$$Q(s_{i,t}, a_{i,t}; \boldsymbol{r}) = \boldsymbol{r}_{s_{i,t},a_{i,t}} + \sum_{s'} \gamma \mathcal{T}_{s'}^{s_{i,t},a_{i,t}} V(s'; \boldsymbol{r})$$

where $\boldsymbol{r}_{s_{i,t},a_{i,t}} = r(s_{i,t}, a_{i,t}) \in \mathbb{R}$ is the reward for $(s_{i,t}, a_{i,t})$, γ is the discount factor, $\mathcal{T}_{s'}^{s_{i,t},a_{i,t}} = p(s'|s_{i,t}, a_{i,t})$ is the transition probability by the transition model, and $V(s_{i,t}; \boldsymbol{r})$ is the value associated with state $s_{i,t}$, obtained by the modified Bellman backup operator:

$$V(s_{i,t}; \boldsymbol{r}) = \log \sum_{a \in \mathcal{A}} \exp \left(\boldsymbol{r}_{s_{i,t},a_{i,t}} + \sum_{s'} \gamma \mathcal{T}_{s'}^{s_{i,t},a} V(s'; \boldsymbol{r}) \right)$$

where we apply a **soft-max function** $V(s_{i,t}; \boldsymbol{r}) = \log \sum_{a \in \mathcal{A}} \exp \left(Q(s_{i,t}, a; \boldsymbol{r})\right)$ for the Q-values with all possible actions $a \in \mathcal{A}$. The value function $V(s; \boldsymbol{r})$ for state s can be obtained by repeatedly applying the above Bellman backup operator. For simplicity of notations, we use $V(s_{i,t}; \boldsymbol{r}), Q(s_{i,t}, a_{i,t}; \boldsymbol{r})$ to denote the solution after Bellman backup operators, unlike some literature that uses $V^*(s_{i,t}; \boldsymbol{r}), Q^*(s_{i,t}, a_{i,t}; \boldsymbol{r})$ to denote the difference. Detailed derivation of the above relationships can be found in [246].

Variational lower bound for DGP-IRL

It is intractable to perform the integration as in (A.35) for the marginal log-likelihood. In addition to $p(\mathcal{M}|\boldsymbol{r})$, which involves the latent variable $\boldsymbol{r}$ in a way which requires Q-value iterations, the term $p(\boldsymbol{r}|\mathbf{D}) = \mathcal{N}(\boldsymbol{r}|\mathbf{0}, K_{\mathbf{DD}})$ has a nonlinear dependency on $\mathbf{D}$ in the kernel matrix. To tackle this issue, we introduce inducing outputs $\boldsymbol{f}, \mathbf{V}$ and their corresponding inputs $\mathbf{Z}, \mathbf{W}$, as shown in Fig. 5.2. The resulting model follows the main text:

$$p(\mathcal{M}|\boldsymbol{r}) = \sum_{i=1}^{h} \sum_{t=1}^{T} \big(Q(s_{i,t}, a_{i,t}; \boldsymbol{r}) - V(s_{i,t}; \boldsymbol{r})\big) \tag{A.42}$$

$$p(\boldsymbol{r}|\boldsymbol{f}, \mathbf{D}, \mathbf{Z}) = \mathcal{N}(\boldsymbol{r}|K_{\mathbf{DZ}} K_{\mathbf{ZZ}}^{-1} \boldsymbol{f}, \mathbf{0}) \tag{A.43}$$

$$p(\boldsymbol{f}|\mathbf{Z}) = \mathcal{N}(\boldsymbol{f}|\mathbf{0}, K_{\mathbf{ZZ}}) \tag{A.44}$$

$$p(\mathbf{D}|\mathbf{B}) = \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{d}^m|\mathbf{b}^m, \lambda^{-1}\mathbf{I}) \tag{A.45}$$

$$p(\mathbf{B}|\mathbf{V}, \mathbf{X}, \mathbf{W}) = \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{b}^m|K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} \mathbf{v}^m, \boldsymbol{\Sigma}_B) \tag{A.46}$$

We also design the variation distribution as illustrated in the main text:

$$\begin{aligned}\mathcal{Q} &= q(\boldsymbol{f})q(\mathbf{D})p(\mathbf{B}|\mathbf{V},\mathbf{X})q(\mathbf{V}), \text{ with :}\\ q(\boldsymbol{f}) &= \delta(\boldsymbol{f}-\tilde{\boldsymbol{f}})\\ q(\mathbf{D}) &= \prod_{m=1}^{m_1}\delta\left(\mathbf{d}^m - K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\tilde{\mathbf{v}}^m\right)\\ q(\mathbf{V}) &= \prod_{m=1}^{m_1}\mathcal{N}\left(\mathbf{v}^m|\tilde{\mathbf{v}}^m,\mathbf{G}^m\right),\end{aligned}$$

where the variational distribution $\mathcal{Q}$ is to not be confused with the notation for Q-values, Q. Using the above distribution $\mathcal{Q}$, we can derive the variational lower bound as follows:

$$\log p(\mathcal{M}|\mathbf{X},\mathbf{Z},\mathbf{W}) = \log\int p(\mathcal{M},\boldsymbol{r},\boldsymbol{f},\mathbf{V},\mathbf{D},\mathbf{B}|\mathbf{Z},\mathbf{W},\mathbf{X})d(\boldsymbol{r},\boldsymbol{f},\mathbf{V},\mathbf{D},\mathbf{B}) \tag{A.47}$$

$$= \log\int \underbrace{p(\mathcal{M}|\boldsymbol{r})p(\boldsymbol{r}|\boldsymbol{f},\mathbf{D},\mathbf{Z})}_{p(\mathcal{M}|K_{\mathbf{DZ}}K_{\mathbf{ZZ}}^{-1}\boldsymbol{f})}p(\boldsymbol{f}|\mathbf{Z})p(\mathbf{D}|\mathbf{B})p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})p(\mathbf{V}|\mathbf{W})d(\boldsymbol{r},\boldsymbol{f},\mathbf{V},\mathbf{D},\mathbf{B}) \tag{A.48}$$

$$\geq \int q(\boldsymbol{f})q(\mathbf{D})p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})q(\mathbf{V})\log\frac{p(\mathcal{M}|K_{\mathbf{DZ}}K_{\mathbf{ZZ}}^{-1}\boldsymbol{f})p(\boldsymbol{f}|\mathbf{Z})p(\mathbf{D}|\mathbf{B})p(\mathbf{V}|\mathbf{W})}{q(\boldsymbol{f})q(\mathbf{D})q(\mathbf{V})} \tag{A.49}$$

$$\begin{aligned}&= \log p(\mathcal{M}|K_{\tilde{\mathbf{D}}\mathbf{Z}}K_{\mathbf{ZZ}}^{-1}\tilde{\boldsymbol{f}}) + \log p(\boldsymbol{f}=\tilde{\boldsymbol{f}}|\mathbf{Z})\\ &\quad + \int q(\mathbf{V})q(\mathbf{D})p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})\log\frac{p(\mathbf{D}|\mathbf{B})p(\mathbf{V}|\mathbf{W})}{q(\mathbf{V})}d(\mathbf{D},\mathbf{B},\mathbf{V}).\end{aligned} \tag{A.50}$$

In the above derivation, the combination of $p(\mathcal{M}|\boldsymbol{r})p(\boldsymbol{r}|\boldsymbol{f},\mathbf{D},\mathbf{Z})$ in (5.13) uses the deterministic training conditional (DTC) assumption [185], i.e., $p(\boldsymbol{r}|\boldsymbol{f},\mathbf{D},\mathbf{Z}) = \delta(\boldsymbol{r} - K_{\mathbf{DZ}}K_{\mathbf{ZZ}}^{-1}\boldsymbol{f})$, (5.14) applies Jensen's inequality with the variational distribution Q, (A.50) is a direct consequence of the choice of Q, and $\tilde{\mathbf{D}} = \begin{bmatrix}\tilde{\mathbf{d}}^1 & \cdots & \tilde{\mathbf{d}}^{m_1}\end{bmatrix}$, with $\tilde{\mathbf{d}}^m = K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\tilde{\mathbf{v}}^m$.

Utility 1 (Gaussian identities)**.** *If the marginal and conditional Gaussian distributions for $\boldsymbol{f}$ and $\mathbf{v}$ are in the form:*

$$\begin{aligned}p(\boldsymbol{f}|\mathbf{v}) &= \mathcal{N}(\boldsymbol{f}|\mathbf{M}\mathbf{v}+\mathbf{m},\boldsymbol{\Sigma}_{\boldsymbol{f}})\\ p(\mathbf{v}) &= \mathcal{N}(\mathbf{v}|\boldsymbol{\mu}_{\mathbf{v}},\boldsymbol{\Sigma}_{\mathbf{v}})\end{aligned}$$

Then the marginal distribution of $\boldsymbol{f}$ is:

$$p(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f}|\mathbf{M}\boldsymbol{\mu}_{\mathbf{v}}+\mathbf{m},\boldsymbol{\Sigma}_{\boldsymbol{f}}+\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{v}}\mathbf{M}^{\top}) \tag{A.51}$$

Using the Gaussian identities, the derivation of $\int q(\mathbf{V})p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})d\mathbf{V}$ is as follows:

$$\begin{aligned}\int q(\mathbf{V})p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})d\mathbf{V} &= \int \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{v}^m|\tilde{\mathbf{v}}^m, \mathbf{G}^m)\mathcal{N}(\mathbf{b}^m|K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\mathbf{v}^m, \mathbf{\Sigma}_{\mathbf{B}})d\mathbf{V} \\ &= \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{b}^m|\underbrace{K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\tilde{\mathbf{v}}^m}_{\tilde{\mathbf{b}}^m}, \underbrace{\mathbf{\Sigma}_{\mathbf{B}} + K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\mathbf{G}^m K_{\mathbf{WW}}^{-1}K_{\mathbf{WX}}}_{\tilde{\mathbf{\Sigma}}_{\mathbf{B}}^m})\end{aligned}$$

Therefore, we can obtained a closed form integration for the last term in (A.50) as follows:

$$\begin{aligned}&\int q(\mathbf{V})q(\mathbf{D})p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})\log p(\mathbf{D}|\mathbf{B})d(\mathbf{D},\mathbf{B},\mathbf{V}) \\ &= \int \left(\int q(\mathbf{V})p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})d\mathbf{V}\right) q(\mathbf{D})\log p(\mathbf{D}|\mathbf{B})d(\mathbf{D},\mathbf{B}) \\ &= \int \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{b}^m|\tilde{\mathbf{b}}^m, \tilde{\mathbf{\Sigma}}_{\mathbf{B}}^m)\log \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{d}^m = \tilde{\mathbf{d}}^m|\mathbf{b}^m, \lambda^{-1}\mathbf{I})d\mathbf{B} \\ &= \int \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{b}^m|\tilde{\mathbf{b}}^m, \tilde{\mathbf{\Sigma}}_{\mathbf{B}}^m)\log \prod_{m=1}^{m_1} \left((2\pi)^{-n/2}|\lambda^{-1}\mathbf{I}|^{-1/2}e^{-\frac{\lambda}{2}(\tilde{\mathbf{d}}^m-\mathbf{b}^m)^\top(\tilde{\mathbf{d}}^m-\mathbf{b}^m)}\right)d\mathbf{B} \\ &= \int \prod_{m=1}^{m_1} \mathcal{N}(\mathbf{b}^m|\tilde{\mathbf{b}}^m, \tilde{\mathbf{\Sigma}}_{\mathbf{B}}^m)\left(-\frac{nm_1}{2}\log(2\pi\lambda^{-1}) - \frac{\lambda}{2}\sum_{m=1}^{m_1}(\tilde{\mathbf{d}}^m-\mathbf{b}^m)^\top(\tilde{\mathbf{d}}^m-\mathbf{b}^m)\right)d\mathbf{B} \\ &= -\frac{nm_1}{2}\log(2\pi\lambda^{-1}) - \frac{\lambda}{2}\sum_{m=1}^{m_1}\left(Tr(\tilde{\mathbf{\Sigma}}_{\mathbf{B}}^m) + (\tilde{\mathbf{d}}^m-\tilde{\mathbf{b}}^m)^\top(\tilde{\mathbf{d}}^m-\tilde{\mathbf{b}}^m)\right)\end{aligned}$$

where $\tilde{\mathbf{\Sigma}}_B^m = \mathbf{\Sigma}_B + K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\mathbf{G}^m K_{\mathbf{WW}}^{-1}K_{\mathbf{WX}}$, $\tilde{\mathbf{b}}^m = K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\tilde{\mathbf{v}}^m$, and $\tilde{\mathbf{d}}^m = K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\tilde{\mathbf{v}}^m$, according to the variational distribution Q.

We now express the variational lower bound of the log likelihood as follow:

$$\mathcal{L} = \mathcal{L}_M + \mathcal{L}_G - \mathcal{L}_{KL} + \mathcal{L}_B - \frac{nm_1}{2}\log(2\pi\lambda^{-1}) \tag{A.52}$$

where

$$\mathcal{L}_M = \log p(\mathcal{M}|K_{\tilde{\mathbf{D}}\mathbf{Z}}K_{\mathbf{ZZ}}^{-1}\tilde{\boldsymbol{f}}) \tag{A.53}$$

$$\mathcal{L}_G = \log p(\boldsymbol{f} = \tilde{\boldsymbol{f}}|\mathbf{Z}) = \log \mathcal{N}(\boldsymbol{f} = \tilde{\boldsymbol{f}}|0, K_{\mathbf{ZZ}}) \tag{A.54}$$

$$= -\frac{1}{2}\tilde{\boldsymbol{f}}^\top K_{\mathbf{ZZ}}^{-1}\tilde{\boldsymbol{f}} - \frac{n_{inducing}}{2}\log(2\pi) - \frac{1}{2}\log|K_{\mathbf{ZZ}}| \tag{A.55}$$

$$\mathcal{L}_{KL} = KL(q(\mathbf{V})||p(\mathbf{V}|\mathbf{W})) = \sum_{m=1}^{m_1} KL(\mathcal{N}(\mathbf{v}^m|\tilde{\mathbf{v}}^m, \mathbf{G}^m)||\mathcal{N}(\mathbf{v}^m|0, K_{\mathbf{WW}})) \tag{A.56}$$

$$= \sum_{m=1}^{m_1}\frac{1}{2}\left(Tr(K_{\mathbf{WW}}^{-1}(\mathbf{G}^m + \tilde{\mathbf{v}}^m\tilde{\mathbf{v}}^{m\top}) - n_{inducing} + \log\left(\frac{|K_{\mathbf{WW}}|}{|\mathbf{G}^m|}\right)\right) \tag{A.57}$$

$$\mathcal{L}_B = -\frac{\lambda}{2}\sum_{m=1}^{m_1} Tr(\boldsymbol{\Sigma}_{\mathbf{B}} + K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\mathbf{G}^m K_{\mathbf{WW}}^{-1}K_{\mathbf{WX}}) \tag{A.58}$$

which is also described in the main paper. The learning of the model involves optimizing over the variational parameters, including $\tilde{\boldsymbol{f}}, \tilde{\mathbf{v}}^m, \mathbf{G}^m$, inducing inputs $\mathbf{Z}$, as well as hyperparameters for the kernel functions, which is performed through backpropagation based on the gradients of the variational lower bound (A.52) with respect to these parameters.

Optimizing the variational distribution $q(\mathbf{V})$

As can be seen, the variational lower bound (A.52) depends on the parameters of the variational distribution $q(\mathbf{V}) = \prod_{m=1}^{m_1}\mathcal{N}(\mathbf{v}^m|\tilde{\mathbf{v}}^m, \mathbf{G}^m)$, which can be optimized to improve the lower bound further. For the last term in (A.50), we have

$$\begin{aligned}
&\int q(\mathbf{V})q(\mathbf{D})p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})\log\frac{p(\mathbf{D}|\mathbf{B})p(\mathbf{V}|\mathbf{W})}{q(\mathbf{V})}d(\mathbf{D},\mathbf{B},\mathbf{V})\\
&= \int q(\mathbf{V})\left(\int q(\mathbf{D})p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})\log\frac{p(\mathbf{D}|\mathbf{B})p(\mathbf{V}|\mathbf{W})}{q(\mathbf{V})}d(\mathbf{D},\mathbf{B})\right)d\mathbf{V}\\
&= \int q(\mathbf{V})\left(\int p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})\log\frac{p(\mathbf{D}=\tilde{\mathbf{D}}|\mathbf{B})p(\mathbf{V}|\mathbf{W})}{q(\mathbf{V})}d\mathbf{B}\right)d\mathbf{B}\\
&= \int q(\mathbf{V})\log\frac{e^{\langle\log p(\mathbf{D}=\tilde{\mathbf{D}}|\mathbf{B})\rangle_{p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})}}p(\mathbf{V}|\mathbf{W})}{q(\mathbf{V})}d\mathbf{V}
\end{aligned}$$

where we have $\tilde{\mathbf{D}} = \begin{bmatrix}\tilde{\mathbf{d}}^1 & \cdots & \tilde{\mathbf{d}}^{m_1}\end{bmatrix}$, with $\tilde{\mathbf{d}}^m = K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\tilde{\mathbf{e}}^m$, and $\tilde{\mathbf{e}}^m$ for $m = 1, ..., m_1$ are variational parameters to optimize. To maximize the above quantity, we can reverse the Jensen's inequality to obtin the condition that:

$$\log q(\mathbf{V}) = C + \langle\log p(\mathbf{D}=\tilde{\mathbf{D}}|\mathbf{B})\rangle_{p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})} + \log p(\mathbf{V}|\mathbf{W})$$

where C denotes a constant. Now for the term $\langle \log p(\mathbf{D} = \tilde{\mathbf{D}}|\mathbf{B})\rangle_{p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})}$, we have:

$$\langle \log p(\mathbf{D} = \tilde{\mathbf{D}}|B)\rangle_{p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})} = \sum_{m=1}^{m_1} \langle \log \mathcal{N}(\mathbf{d}^m = \tilde{\mathbf{d}}^m|\mathbf{b}^m, \lambda^{-1}I)\rangle_{p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X})}$$
$$= C + \sum_{m=1}^{m_1} \left\langle -\frac{\lambda}{2} Tr\left(\tilde{\mathbf{d}}^m \tilde{\mathbf{d}}^{m\top} + \mathbf{b}^m \mathbf{b}^{m\top} - 2\tilde{\mathbf{d}}^m \mathbf{b}^{m\top}\right)\right\rangle_{\mathcal{N}(\mathbf{b}^m|K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\mathbf{v}^m, \mathbf{\Sigma_B})}$$
$$= C + \sum_{m=1}^{m_1} \left(-\frac{\lambda}{2} Tr\left(\tilde{\mathbf{d}}^m \tilde{\mathbf{d}}^{m\top} + \mathbf{\Sigma_B} + \mathbf{v}^{m\top} K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} \mathbf{v}^m - 2\mathbf{v}^{m\top} K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} \tilde{\mathbf{d}}^m\right)\right)$$

Therefore, we have:

$$\log q(\mathbf{v}^m) = C - \frac{1}{2}\Big(\lambda \mathbf{v}^{m\top} K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} \mathbf{v}^m - 2\lambda \mathbf{v}^{m\top} K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} \tilde{\mathbf{d}}^m + \mathbf{v}^{m\top} K_{\mathbf{WW}}^{-1} \mathbf{v}^m\Big)$$

Therefore by completing the squares we have $q(\mathbf{v}^m) = \mathcal{N}(\mathbf{v}^m|\tilde{\mathbf{v}}_*^m, \mathbf{\Sigma}_{\mathbf{v}*}^m)$:

$$\mathbf{\Sigma}_{\mathbf{v}*}^m = (K_{\mathbf{WW}}^{-1} + \lambda K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1})^{-1} = \lambda^{-1} K_{\mathbf{WW}} (\lambda^{-1} K_{\mathbf{WW}} + K_{\mathbf{WX}} K_{\mathbf{XW}})^{-1} K_{\mathbf{WW}}$$
$$\tilde{\mathbf{v}}_*^m = \lambda \mathbf{\Sigma}_{\mathbf{v}*}^m K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} \tilde{\mathbf{d}}^m = K_{\mathbf{WW}} \underbrace{(\lambda^{-1} K_{\mathbf{WW}} + K_{\mathbf{WX}} K_{\mathbf{XW}})^{-1}}_{\mathbf{\Gamma}} K_{\mathbf{WX}} \tilde{\mathbf{d}}^m$$

With the above optimized variational parameters for $q(\mathbf{v}^m)$, we first obtain:

$$\int q(\mathbf{v}^m) \langle \log p(\mathbf{d}^m = \tilde{\mathbf{d}}^m|\mathbf{b}^m)\rangle_{p(\mathbf{b}^m|\mathbf{v}^m,\mathbf{W},\mathbf{X})} d\mathbf{v}^m = -\frac{n}{2}\log(2\pi\lambda^{-1}) -$$
$$\frac{\lambda}{2} Tr\Big(\tilde{\mathbf{d}}^m \tilde{\mathbf{d}}^{m\top} + \mathbf{\Sigma_B} + K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} (\mathbf{\Sigma}_{\mathbf{v}*}^m + \tilde{\mathbf{v}}_*^m \tilde{\mathbf{v}}_*^{m\top}) - 2\tilde{\mathbf{v}}_*^{m\top} K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} \tilde{\mathbf{d}}^m\Big)$$

Next, we calculate $\int q(\mathbf{v}^m) \log p(\mathbf{v}^m|\mathbf{W}) d\mathbf{v}^m$:

$$\int q(\mathbf{v}^m) \log p(\mathbf{v}^m|\mathbf{W}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|K_{\mathbf{WW}}| - \frac{1}{2} Tr(K_{\mathbf{WW}}^{-1}(\mathbf{\Sigma}_{\mathbf{v}*}^m + \tilde{\mathbf{v}}_*^m \tilde{\mathbf{v}}_*^{m\top}))$$

Finally we have:

$$H(q(\mathbf{v}^m)) = q(\mathbf{v}^m) \log \frac{1}{q(\mathbf{v}^m)} = \frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\mathbf{\Sigma}_{\mathbf{v}*}^m| \tag{A.59}$$

Summarizing, we have:

$$\int q(\mathbf{V}) q(\mathbf{D}) p(\mathbf{B}|\mathbf{V},\mathbf{W},\mathbf{X}) \log \frac{p(\mathbf{D}|\mathbf{B}) p(\mathbf{V}|\mathbf{W})}{q(\mathbf{V})} d(\mathbf{D},\mathbf{B},\mathbf{V})$$
$$\leq \sum_{m=1}^{m_1} \Big[-\frac{n}{2}\log(2\pi\lambda^{-1}) - \frac{1}{2}\log|K_{\mathbf{WW}}| - \frac{1}{2} Tr(K_{\mathbf{WW}}^{-1}(\mathbf{\Sigma}_{\mathbf{v}*}^m + \tilde{\mathbf{v}}_*^m \tilde{\mathbf{v}}_*^{m\top})) + \frac{1}{2}\log|\mathbf{\Sigma}_{\mathbf{v}*}^m|$$
$$- \frac{\lambda}{2} Tr\Big(\tilde{\mathbf{d}}^m \tilde{\mathbf{d}}^{m\top} + \mathbf{\Sigma_B} + K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} (\mathbf{\Sigma}_{\mathbf{v}*}^m + \tilde{\mathbf{v}}_*^m \tilde{\mathbf{v}}_*^{m\top}) - 2\tilde{\mathbf{v}}_*^{m\top} K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} \tilde{\mathbf{d}}^m\Big)\Big]$$

We now express the variational lower bound of the log likelihood as follow:

$$\mathcal{L} = \mathcal{L}_M + \mathcal{L}_G + \mathcal{L}_{DBV} \tag{A.60}$$

where

$$\mathcal{L}_M = \log p(\mathcal{M}|K_{\tilde{\mathbf{D}}\mathbf{Z}}K_{\mathbf{Z}\mathbf{Z}}^{-1}\tilde{\boldsymbol{f}}) \tag{A.61}$$

$$\mathcal{L}_G = \log p(u = \tilde{u}|Z) = \log \mathcal{N}(u = \tilde{u}|0, K_{ZZ}) \tag{A.62}$$

$$= -\frac{1}{2}\tilde{u}^\top K_{ZZ}^{-1}\tilde{u} - \frac{K}{2}\log(2\pi) - \frac{1}{2}\log|K_{ZZ}| \tag{A.63}$$

$$\begin{aligned}\mathcal{L}_{DBV} = \sum_{m=1}^{m_1}\Big[& -\frac{n}{2}\log(2\pi\lambda^{-1}) - \frac{1}{2}\log|K_{\mathbf{WW}}| - \frac{1}{2}Tr(K_{\mathbf{WW}}^{-1}(\boldsymbol{\Sigma}_{\mathbf{v}*}^m + \tilde{\mathbf{v}}_*^m\tilde{\mathbf{v}}_*^{m\top})) + \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mathbf{v}*}^m| \\ & -\frac{\lambda}{2}Tr\left(\tilde{\mathbf{d}}^m\tilde{\mathbf{d}}^{m\top} + \boldsymbol{\Sigma}_{\mathbf{B}} + K_{\mathbf{WW}}^{-1}K_{\mathbf{WX}}K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}(\boldsymbol{\Sigma}_{\mathbf{v}*}^m + \tilde{\mathbf{v}}_*^m\tilde{\mathbf{v}}_*^{m\top}) - 2\tilde{\mathbf{v}}_*^{m\top}K_{\mathbf{WW}}^{-1}K_{\mathbf{WX}}\tilde{\mathbf{d}}^m\right)\Big]\end{aligned} \tag{A.64}$$

where $\tilde{\mathbf{d}}^m = K_{\mathbf{XW}}K_{\mathbf{WW}}^{-1}\tilde{\mathbf{e}}^m$, $\boldsymbol{\Gamma} = (\lambda^{-1}K_{\mathbf{WW}} + K_{\mathbf{WX}}K_{\mathbf{XW}})^{-1}$,$\Sigma_{\mathbf{v}*}^m = \lambda^{-1}K_{\mathbf{WW}}\boldsymbol{\Gamma}K_{\mathbf{WW}}$. The parameters we need to learn in this case include the variational parameters $\tilde{\boldsymbol{f}}$, and $\tilde{\mathbf{e}}^m$ for $m = 1, ..., m_1$, inducing inputs $\mathbf{Z}$, as well as hyperparameters for kernel functions.

Parameters learning by derivatives

In this section, we will obtain the derivatives of the marginal log likelihood $\mathcal{L}$ in (A.60) with respect to the variational parameters $\tilde{\boldsymbol{f}}$, $\tilde{\mathbf{e}}^m$ and inducing inputs $\mathbf{Z}$. The derivative of the reinforcement learning term, $p(\mathcal{M}|\boldsymbol{r})$ in (A.41), with respect to the reward $\boldsymbol{r}$, is given by:

$$\frac{\partial}{\partial \boldsymbol{r}}\log p(\mathcal{M}|\boldsymbol{r}) = \sum_i\sum_t\left(\frac{\partial}{\partial \boldsymbol{r}}\boldsymbol{r}_{s_{i,t},a_{i,t}} - \frac{\partial}{\partial \boldsymbol{r}}V_{s_{i,t}}^{\boldsymbol{r}} + \sum_{s'}\gamma\mathcal{T}_{s'}^{s_{i,t},a_{i,t}}\frac{\partial}{\partial \boldsymbol{r}}V_{s'}^{\boldsymbol{r}}\right) \tag{A.65}$$

The first term, $\sum_i\sum_t\frac{\partial}{\partial \boldsymbol{r}}\boldsymbol{r}_{s_{i,t},a_{i,t}}$, is simply a vector that counts the number of state-action pairs in the demonstrations $\hat{\mu}$, whose entry corresponding to (s, a) is given by: $\hat{\mu}_{s,a} = \sum_i\sum_t 1_{s_{i,t}=s\wedge a_{i,t}=a}$. The second term involves the derivative of the value function at state s with respect to rewards, as indicated in [246], equal to the expected visitation count of each state-action pair when starting from state s and following the optimal stochastic policy, i.e., $\frac{\partial}{\partial \boldsymbol{r}}V_s^{\boldsymbol{r}} = E[\mu|s]$, where μ is a vector with each entry $\mu_{s,a}$ corresponding to the expected visitation count for (s, a). Therefore, (A.65) can be written as:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{r}}\log p(\mathcal{M}|\boldsymbol{r}) &= \hat{\mu} - \sum_i\sum_t E[\mu|s_{i,t}] + \sum_i\sum_t\sum_{s'}\gamma\mathcal{T}_{s'}^{s_{i,t},a_{i,t}}E[\mu|s_{i,t}] \\ &= \hat{\mu} - \sum_s \hat{\nu}_s E[\mu|s]\end{aligned}$$

where $\hat{\nu}_s = \sum_a \hat{\mu}_{s,a} - \sum_i \sum_t \gamma \mathcal{T}_{s'}^{s_{i,t}, a_{i,t}}$. The term $\sum_s \hat{\nu}_s E[\mu|s]$ can be computed efficiently by a simple iterative algorithm described in [246], which we do not recount here. Note that the above derivation follows from [143]. For the variational parameters $\tilde{\boldsymbol{f}}$, we need to consider only two terms that involve it, i.e., $\mathcal{L}_{\mathcal{M}}, \mathcal{L}_G$:

$$\begin{aligned}
\frac{\partial \mathcal{L}_M}{\partial \tilde{\boldsymbol{f}}} &= \frac{\partial \boldsymbol{r}}{\partial \tilde{\boldsymbol{f}}} \frac{\partial \mathcal{L}_{\mathcal{M}}}{\partial \boldsymbol{r}} = K_{\tilde{\mathbf{D}}\mathbf{Z}} K_{\mathbf{ZZ}}^{-1} \frac{\partial \log p(\mathcal{M}|\boldsymbol{r})}{\partial \boldsymbol{r}} \\
\frac{\partial \mathcal{L}_G}{\partial \tilde{\boldsymbol{f}}} &= -K_{\mathbf{ZZ}}^{-1} \tilde{\boldsymbol{f}}
\end{aligned}$$

where $\boldsymbol{r} = K_{\tilde{\mathbf{D}}\mathbf{Z}} K_{\mathbf{ZZ}}^{-1} \tilde{\boldsymbol{f}}$ is the reward vector that we use for reinforcement learning.

For the variational parameters $\tilde{\mathbf{e}}^m$, let $\tilde{\mathbf{D}} = \left[K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} \tilde{\mathbf{e}}^1, ..., K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} \tilde{\mathbf{e}}^{m_1}\right] \in \mathbb{R}^{n \times m_1}$, and $\mathbf{E} = [\tilde{\mathbf{e}}^1, ..., \tilde{\mathbf{e}}^{m_1}] \in \mathbb{R}^{K \times m_1}$:

$$\frac{\partial \mathcal{L}_{\mathcal{M}}}{\partial \mathbf{E}} = \frac{\partial \tilde{\mathbf{D}}}{\partial \mathbf{E}} \frac{\partial K_{\tilde{\mathbf{D}}\mathbf{Z}}}{\partial \tilde{\mathbf{D}}} \frac{\partial \boldsymbol{r}}{\partial K_{\tilde{\mathbf{D}}\mathbf{Z}}} \frac{\partial \mathcal{L}_{\mathcal{M}}}{\partial \boldsymbol{r}}$$

In addition, by applying matrix derivatives,

$$\begin{aligned}
\frac{\partial \mathcal{L}_{DBV}}{\partial \mathbf{e}^m} = -\frac{\lambda}{2} \Big(& 2K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} + 2K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} K_{\mathbf{XW}} \mathbf{\Gamma} K_{\mathbf{WX}} K_{\mathbf{XW}} \Gamma K_{\mathbf{WX}} K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} \\
& -4K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} K_{\mathbf{XW}} \mathbf{\Gamma} K_{\mathbf{WX}} K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} \Big) \mathbf{e}^m - K_{\mathbf{WW}}^{-1} K_{\mathbf{WX}} K_{\mathbf{XW}} \mathbf{\Gamma} K_{\mathbf{WW}} \mathbf{\Gamma} K_{\mathbf{WX}} K_{\mathbf{XW}} K_{\mathbf{WW}}^{-1} \mathbf{e}^m.
\end{aligned}$$

The gradients are provided to minFunc [200], which calls a quasi-Newton strategy, where limited-memory BFGS updates with Shanno-Phua scaling are used in computing the step direction, and a bracketing line-search for a point satisfying the strong Wolfe conditions is used to compute the step direction.

A.5 Chapter 7: Cyber resilience of power grid state estimation

Lemma 7.1 (Sufficient condition for stealth attack)**.** *An attack* $\mathbf{b}$ *is stealthy if there exists a nonzero vector* $\boldsymbol{c}$ *such that* $\mathbf{M}_i \boldsymbol{c} = \mathbf{0}$ *for every* $i \in [n_m]$ *that is not in the support of* $\mathbf{b}$*.*

Proof. Since $f_i(\mathbf{v}) = \operatorname{trace}(\mathbf{M}_i \mathbf{v}\mathbf{v}^*)$, we have

$$f_i(\mathbf{v} + \boldsymbol{c}) = \operatorname{trace}(\mathbf{M}_i (\mathbf{v} + \boldsymbol{c})(\mathbf{v} + \boldsymbol{c})^*) = f_i(\mathbf{v}),$$

for every $i \in [n_m]$ that is not in the support of $\mathbf{b}$. □

Proof of Theorem 7.1 and Lemma 7.2

Theorem 7.1. *The relaxation* (NC-FDIA-c) *recovers a solution of the nonconvex problem* (NC-FDIA) *and finds an optimal attack if it has a solution* $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ *satisfying Assumption 2a such that rank*$(\hat{\mathbf{W}}) = 1$.

Proof. First, we prove that the equation $\text{rank}(\hat{\mathbf{W}}) = 1$ implies that $\hat{\mathbf{W}} = a^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*$, for some a such that $|a| \geq 1$. Since $\begin{bmatrix} 1 & \hat{\mathbf{v}}^* \\ \hat{\mathbf{v}} & \hat{\mathbf{W}} \end{bmatrix} \succeq 0$, by Schur complement, we have $\hat{\mathbf{W}} \succeq 0$, and $\hat{\mathbf{W}} - \hat{\mathbf{v}}\hat{\mathbf{v}}^* \succeq 0$. Due to $\text{rank}(\hat{\mathbf{W}}) = 1$, we can express $\hat{\mathbf{W}} = \mathbf{w}\mathbf{w}^*$. Since $\mathbf{w}\mathbf{w}^* - \hat{\mathbf{v}}\hat{\mathbf{v}}^* \succeq 0$, one can write $\mathbf{w} = a\hat{\mathbf{v}}$, where $|a| \geq 1$ (otherwise, there exists a vector $\boldsymbol{\nu} \in \mathbb{C}^{n_b}$ such that $\boldsymbol{\nu}^*\mathbf{w} = 0$, but $\boldsymbol{\nu}^*\hat{\mathbf{v}} \neq 0$ and $\boldsymbol{\nu}^*\left(\mathbf{w}\mathbf{w}^* - \hat{\mathbf{v}}\hat{\mathbf{v}}^*\right)\boldsymbol{\nu} = -|\boldsymbol{\nu}^*\hat{\mathbf{v}}|^2 < 0$, which violates the PSD condition).

Now, we show by contradiction that the equation $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$ holds at optimality. Assume that $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ is an optimal solution of (NC-FDIA-c) and that $\hat{a} > 1$ (the case $\hat{a} < -1$ is similar). It is obvious that $(\hat{a}\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ is also feasible. This gives rise to the relation:

$$\begin{aligned}\bar{h}(\hat{\mathbf{v}}, \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*) &= \text{trace}\left(\hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) - (\tilde{\mathbf{v}}^*\mathbf{v}_{tg} + \mathbf{v}_{tg}^*\tilde{\mathbf{v}}) \\ &> \text{trace}\left(\hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) - \hat{a}(\tilde{\mathbf{v}}^*\mathbf{v}_{tg} + \mathbf{v}_{tg}^*\tilde{\mathbf{v}}) \\ &= \bar{h}(\hat{a}\hat{\mathbf{v}}, \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*),\end{aligned}$$

where the inequality follows from Assumption 2a. This contradicts the optimality of $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$. Therefore, we must have $\hat{a} = 1$, implying that $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$.

Recall that (NC-FDIA-c) provides a lower bound for (NC-FDIA-r), which is a reformulation of (NC-FDIA). Therefore, since $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ is feasible for (NC-FDIA-r), it is optimal for (NC-FDIA). □

Lemma 7.2 (Stealth attack)**.** *Let* $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ *be a solution of* (SDP-FDIA) *satisfying Assumption 2b. The attack* $\hat{\mathbf{b}}$ *is stealthy if rank*$(\hat{\mathbf{W}}) = 1$.

Proof. Let $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ denote an optimal solution of (SDP-FDIA). If $\text{rank}(\hat{\mathbf{W}}) = 1$, then using a similar reasoning as in the proof for Theorem 7.1, we have $\hat{\mathbf{W}} = a^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*$ for every $|a| \geq 1$ due to the PSD constraint. Now, we show by contradiction that the relation $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$ holds at optimality. Let $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ be an optimal solution of (SDP-FDIA), and $\hat{a} > 1$ (the case $\hat{a} < -1$ is similar). It is obvious that $(\hat{a}\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ is also feasible. For a fixed $\hat{\mathbf{b}}$, this gives rise to the relation:

$$\begin{aligned}&\bar{h}(\hat{\mathbf{v}}, \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*) + \hat{a}^2\text{trace}(\mathbf{M}_0\hat{\mathbf{v}}\hat{\mathbf{v}}^*) \\ &= \text{trace}\left(\hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) - (\tilde{\mathbf{v}}^*\mathbf{v}_{tg} + \mathbf{v}_{tg}^*\tilde{\mathbf{v}}) + \hat{a}^2\text{trace}(\mathbf{M}_0\hat{\mathbf{v}}\hat{\mathbf{v}}^*) \\ &> \text{trace}\left(\hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) - \hat{a}(\tilde{\mathbf{v}}^*\mathbf{v}_{tg} + \mathbf{v}_{tg}^*\tilde{\mathbf{v}}) + \hat{a}^2\text{trace}(\mathbf{M}_0\hat{\mathbf{v}}\hat{\mathbf{v}}^*) \\ &= \bar{h}(\hat{a}\hat{\mathbf{v}}, \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*) + \hat{a}^2\text{trace}(\mathbf{M}_0\hat{\mathbf{v}}\hat{\mathbf{v}}^*)\end{aligned}$$

where the inequality follows from Assumption 2b. This contradicts the optimality of $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$. Therefore, we must have $\hat{a} = 1$, implying that $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$. Moreover, since

$$f_i(\hat{\mathbf{v}}) = \mathrm{trace}\left(\mathbf{M}_i\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) = \mathrm{trace}\left(\mathbf{M}_i\hat{\mathbf{W}}\right) = m_i + \hat{b}_i = f_i(\mathbf{v}) + \hat{b}_i,\ \forall i \in [n_m],$$

the stealth condition is satisfied, implying that $\hat{\mathbf{b}}$ is stealthy. □

Proof of Theorems 7.2 and 7.3

In the case of $\bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) = \mathrm{trace}\left(\mathbf{W}\right) - \tilde{\mathbf{v}}^*\mathbf{v}_{tg} - \mathbf{v}_{tg}^*\tilde{\mathbf{v}}$, the dual of (FDIA-SE) can be written as

$$\begin{aligned} \min_{\boldsymbol{\xi}\in\mathbb{R}^{n_m}, q_0\in\mathbb{R}} \quad & \boldsymbol{\xi}\cdot(\mathbf{v}+\mathbf{b}) \\ \text{s. t.} \quad & \begin{bmatrix} q_0 & -\mathbf{v}_{tg}^* \\ -\mathbf{v}_{tg} & \mathbf{I}+\mathbf{M}_0+\sum_i \xi_i\mathbf{M}_i \end{bmatrix} \succeq 0, \end{aligned}$$

where $\boldsymbol{\xi}$ is the vector of dual variables. The complementary slackness condition is given by:

$$\begin{bmatrix} q_0 & -\mathbf{v}_{tg}^* \\ -\mathbf{v}_{tg} & \mathbf{I}+\mathbf{M}_0+\sum_i \xi_i\mathbf{M}_i \end{bmatrix} \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} = \begin{bmatrix} q_0 - \tilde{\mathbf{v}}_{tg}^*\tilde{\mathbf{v}} & q_0\tilde{\mathbf{v}}^* - \tilde{\mathbf{v}}_{tg}^*\mathbf{W} \\ -\tilde{\mathbf{v}}_{tg} + \mathbf{Q}_0\tilde{\mathbf{v}} & -\tilde{\mathbf{v}}_{tg}\tilde{\mathbf{v}}^* + \mathbf{Q}_0\mathbf{W} \end{bmatrix} = \mathbf{0}. \tag{A.66}$$

Let $(\hat{\mathbf{v}}, \hat{\mathbf{W}})$ be an optimal solution of (FDIA-SE) and $(\hat{q}_0, \hat{\boldsymbol{\xi}})$ be a dual optimal solution. It follows from the above equation that $\hat{q}_0 = \tilde{\mathbf{v}}_{tg}^*\hat{\tilde{\mathbf{v}}}$. By defining

$$\mathbf{Q}_0 = \mathbf{I} + \mathbf{M}_0 + \sum_i \xi_i\mathbf{M}_i, \tag{A.67}$$

$$\mathbf{L}_0 = -\frac{1}{\hat{q}_0}\mathbf{v}_{tg}\mathbf{v}_{tg}^* + \mathbf{I} + \mathbf{M}_0, \tag{A.68}$$

$$\mathbf{H}(\boldsymbol{\xi}) = \mathbf{L}_0 + \sum_i \xi_i\mathbf{M}_i, \tag{A.69}$$

and using the Schur's complement, the dual problem can be reformulated as

$$\begin{aligned} \min_{\boldsymbol{\xi}\in\mathbb{R}^{n_m}} \quad & \boldsymbol{\xi}\cdot(\mathbf{v}+\mathbf{b}) \\ \text{s. t.} \quad & \mathbf{H}(\boldsymbol{\xi}) = \mathbf{L}_0 + \sum_i \xi_i\mathbf{M}_i \succeq 0. \end{aligned} \tag{FDIA-SE-d}$$

The following lemma proves strong duality between (FDIA-SE) and its dual formulation.

Lemma A.1. *Suppose that there exists a vector $\mathbf{v} \in \mathcal{V}(\mathcal{M})$ that is feasible for (FDIA-SE). Then, strong duality holds between (FDIA-SE) and its dual formulation (FDIA-SE-d).*

Proof. To prove the lemma, it suffices to find a strictly feasible point for the dual problem. Since there exists a vector $\mathbf{v} \in \mathcal{V}(\mathcal{M})$ that is feasible for (FDIA-SE), we have $\overline{\mathbf{v}}^\top\mathbf{J}(\mathbf{v}) \neq \mathbf{0}$

due to the full row-rank property of $\mathbf{J}(\mathbf{v})$. Therefore, there exists an index $i \in [n_m]$ such that $\mathbf{v}^*\mathbf{M}_i\mathbf{v} \neq 0$. Let $\{\mathbf{d}_1, ..., \mathbf{d}_{n_m}\}$ denote the standard basis vectors in $\mathbb{R}^{n_m}$. Then, we can select $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi} + \delta \times \mathbf{d}_i$ for any feasible dual vector $\boldsymbol{\xi}$, where $\delta \in \mathbb{R}$ is a nonzero number with an arbitrarily small absolute value such that $\delta \times \mathbf{v}^*\mathbf{M}_i\mathbf{v} > 0$. Therefore, one can write:

$$\mathbf{H}(\hat{\boldsymbol{\xi}}) = \mathbf{L}_0 + \sum_i \hat{\xi}_i\mathbf{M}_i = \mathbf{H}(\boldsymbol{\xi}) + c\mathbf{M}_i \succ 0 \tag{A.70}$$

if c is sufficiently small. Hence, $\hat{\boldsymbol{\xi}}$ is a strictly feasible dual point and, by Slater's condition, strong duality holds. □

Definition A.2. *Define $\Omega(\mathbf{L}_0, \mathbf{v})$ as a set of dual variables such that*

$$\mathbf{J}(\mathbf{v})\boldsymbol{\xi} = -2\overline{\mathbf{L}_0}\overline{\mathbf{v}}, \tag{A.71}$$

for every $\boldsymbol{\xi} \in \Omega(\mathbf{L}_0, \mathbf{v})$, where $\mathbf{J}(\mathbf{v}) \in \mathbb{R}^{(2n_b-1)\times n_m}$ is the Jacobian matrix in (7.5).

Since $\mathbf{H}(\boldsymbol{\xi}) = \mathbf{L}_0 + \sum_i \xi_i\mathbf{M}_i$, we have

$$\overline{\mathbf{H}(\boldsymbol{\xi})}\overline{\mathbf{v}} = \overline{\mathbf{L}_0}\overline{\mathbf{v}} + \sum_i \xi_i\overline{\mathbf{M}_i}\overline{\mathbf{v}} = \overline{\mathbf{L}_0}\overline{\mathbf{v}} + \tfrac{1}{2}\mathbf{J}(\mathbf{v})\boldsymbol{\xi} = \mathbf{0},$$

for all $\boldsymbol{\xi} \in \Omega(\mathbf{L}_0, \mathbf{v})$, which indicates that $\mathbf{v}$ lies in the null space of $\mathbf{H}(\boldsymbol{\xi}) \in \mathbb{S}^{n_b}$ for every $\boldsymbol{\xi} \in \Omega(\mathbf{L}_0, \mathbf{v})$.

Lemma A.2. *For every $\mathbf{v} \in \mathcal{V}(\mathcal{M})$ and $n_m \geq 2n_b - 1$, there is a vector $\boldsymbol{\xi} \in \mathbb{R}^{n_m}$ such that (A.71) is satisfied. Therefore, $\Omega(\mathbf{L}_0, \mathbf{v})$ is nonempty for every observable state vector $\mathbf{v}$.*

Proof. Since $\mathbf{v} \in \mathcal{V}(\mathcal{M})$ is observable, $\mathbf{J}(\mathbf{v})$ has full row rank. This implies that, for every $\mathbf{L}_0$, as long as the number of columns of $\mathbf{J}(\mathbf{v})$, namely n_m, is greater than or equal to the number of rows, namely $2n_b - 1$, there is a vector $\boldsymbol{\xi}$ satisfying (A.71). □

Theorem 7.2. *If $\mathcal{A}(\mathcal{M}, \rho)$ is non-empty for some $\rho > 0$, the intersection of the attackable region and the observable set, i.e., $\mathcal{A}(\mathcal{M}, \rho) \cap \mathcal{V}(\mathcal{M})$, is an open set.*

Proof. Define $\kappa(\mathbf{H}(\boldsymbol{\xi}))$ as the sum of the two smallest eigenvalues of the Hermitian matrix $\mathbf{H}(\boldsymbol{\xi}) \in \mathbb{S}^{n_b}$. It can be shown that the intersection of the attackable region and observable set, i.e., $\mathcal{A}(\mathcal{M}, \rho) \cap \mathcal{V}(\mathcal{M})$, can be represented as

$$\{\mathbf{v} \in \mathcal{V}(\mathcal{M}) | \kappa(\mathbf{H}(\boldsymbol{\xi})) > 0, \boldsymbol{\xi} \in \Omega(\mathbf{L}_0, \mathbf{v})\}.$$

The proof is similar to the argument made in Theorem 3 of [151]. Now, consider a vector $\mathbf{v}$ in $\{\mathbf{v} \in \mathcal{V}(\mathcal{M}) | \kappa(\mathbf{H}(\boldsymbol{\xi})) > 0, \boldsymbol{\xi} \in \Omega(\mathbf{L}_0, \mathbf{v})\}$, and let δ denote the second smallest eigenvalue of $\mathbf{H}(\mathcal{M}, \boldsymbol{\xi})$. Due to the continuity of the mapping from a state $\mathbf{v}$ to a set $\Omega(\mathbf{L}_0, \mathbf{v})$, there exists a neighborhood $\mathcal{T} \in \mathbf{C}^{n_b}$ such that there exists a $\boldsymbol{\xi}_t \in \Omega(\mathbf{L}_0, \mathbf{v}_t)$ with the following property:

$$\|\mathbf{H}(\boldsymbol{\xi}) - \mathbf{H}(\boldsymbol{\xi}_t)\|_F < \sqrt{\delta} \tag{A.72}$$

for every $\mathbf{v}_t \in \mathcal{V}(\mathcal{M}) \cap \mathcal{T}$ (note that $\|.\|_F$ represents the Frobenius norm). Using an eigenvalue perturbation argument (Lemma 5 in [152]), it can be concluded that $\mathbf{H}(\boldsymbol{\xi}_t) \succeq 0$ and $\text{rank}(\mathbf{H}(\boldsymbol{\xi}_t)) = n_b - 1$, which imply that $\kappa(\mathbf{H}(\boldsymbol{\xi}_t)) > 0$ and $\mathbf{v}_t \in \{\mathbf{v} \in \mathcal{V}(\mathcal{M}) | \kappa(\mathbf{H}(\boldsymbol{\xi})) > 0, \boldsymbol{\xi} \in \Omega(\mathbf{L}_0, \mathbf{v})\}$. Hence, $\mathcal{A}(\mathcal{M}, \rho) \cap \mathcal{V}(\mathcal{M})$ is an open set. □

Theorem 7.3. *Consider the "target state attack" with $\bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) = \mathit{trace}\,(\mathbf{W}) - \tilde{\mathbf{v}}^* \mathbf{v}_{tg} - \mathbf{v}_{tg}^* \tilde{\mathbf{v}}$, where $\mathbf{v}_{tg} \in \mathcal{V}(\mathcal{M})$ is chosen to be observable. Then, $\mathbf{v}_{tg} \in \mathcal{A}(\mathcal{M}, \rho)$ for some $\rho > 0$, i.e., $\mathbf{v}_{tg}$ is attackable.*

Proof. Let $\mathbf{M}_0$ be chosen as $\mathbf{M}_0 = -\mathbf{I} + \epsilon \mathbf{v}_{tg} \mathbf{v}_{tg}^* + \mathbf{L}_0$, for some $\epsilon > 0$ and a matrix $\mathbf{L}_0$ satisfying the following properties: 1) $\mathbf{L}_0 \succeq 0$, 2) 0 is a simple eigenvalue of $\mathbf{L}_0$, 3) the vector $\mathbf{v}_{tg}$ belongs to the null space of $\mathbf{L}_0$. Let $\rho = \|\mathbf{M}_0\|_2$ defined above. Note that $\boldsymbol{\xi} = \mathbf{0}$ is a feasible dual point since $\mathbf{H}(\mathbf{0}) = \mathbf{L}_0 \succeq 0$. Moreover, because of the equation $\mathbf{H}(\mathbf{0}) \mathbf{v}_{tg} = \mathbf{L}_0 \mathbf{v}_{tg} = \mathbf{0}$, we have $\mathbf{0} \in \Omega(\mathbf{L}_0, \mathbf{v}_{tg})$. Since 0 is a simple eigenvalue of $\mathbf{L}_0$, it holds that $\kappa(\mathbf{H}(\mathbf{0})) = \kappa(\mathbf{L}_0) > 0$. Therefore, it can be concluded that $\mathbf{v}_{tg} \in \{\mathbf{v} \in \mathcal{V}(\mathcal{M}) | \kappa(\mathbf{H}(\boldsymbol{\xi})) > 0, \boldsymbol{\xi} \in \Omega(\mathbf{L}_0, \mathbf{v})\}$. By the proof of Theorem 7.2, it follows that $\mathbf{v}_{tg}$ is attackable. □

Proof of Lemma 7.3

Lemma 7.3. *$g(\mathbf{b})$ is convex and sub-differentiable.*

For any two attacks $\mathbf{b}_1$ and $\mathbf{b}_2$, let the optimal states be denoted as $(\hat{\mathbf{v}}^{(1)}, \hat{\mathbf{W}}^{(1)})$ and $(\hat{\mathbf{v}}^{(2)}, \hat{\mathbf{W}}^{(2)})$. For every number $\lambda \in [0, 1]$, the point $(\lambda \hat{\mathbf{v}} + (1-\lambda)\hat{\mathbf{v}}^{(2)}, \lambda \hat{\mathbf{W}} + (1-\lambda)\hat{\mathbf{W}}^{(2)})$ is a feasible solution for the attack $\lambda \mathbf{b}_1 + (1-\lambda)\mathbf{b}_2$:

$$g(\lambda \mathbf{b}_1 + (1-\lambda)\mathbf{b}_2) \leq \lambda g(\mathbf{b}_1) + (1-\lambda) g(\mathbf{b}_2),$$

which proves the convexity. In what follows, in addition to proving the continuity of $g(\mathbf{b})$, we will derive a bound on the subgradient of $g(\mathbf{b})$, which is used in Theorem 7.5. The method is an extension of [232] to the primal formulation. In particular, our analysis is a type of parametric programming, which characterizes the change of the solution with respect to small perturbations of the parameters (see [232, Ch. 4]). Consider a disturbance γ to the vector $\mathbf{b} \in \mathbb{R}^{n_m}$ in (FDIA-SE) along the direction $\underline{\mathbf{b}}$. The primal problem changes as

$$\begin{aligned} \min_{\tilde{\mathbf{v}}, \mathbf{W}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) + \text{trace}\,(\mathbf{M}_0 \mathbf{W}) \\ \text{s. t.} \quad & \text{trace}\,(\mathbf{M}_i \mathbf{W}) = m_i + b_i + \gamma \underline{b}_i \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0 \end{aligned} \tag{P$_\gamma$}$$

and its dual formulation is given by:

$$\begin{aligned} \min_{\boldsymbol{\xi}, q_0} \quad & \boldsymbol{\xi} \cdot (\mathbf{v} + \mathbf{b} + \gamma \underline{\mathbf{b}}) \\ \text{s. t.} \quad & \begin{bmatrix} q_0 & -\mathbf{v}_{tg}^* \\ -\mathbf{v}_{tg} & \mathbf{I} + \mathbf{M}_0 + \sum_i \xi_i \mathbf{M}_i \end{bmatrix} \succeq 0 \end{aligned} \tag{D$_\gamma$}$$

Let Γ be the set of all vectors γ for which (D_γ) has a bounded solution and is strictly feasible. Assume that $0 \in \Gamma$. It is straightforward to verify that Γ is a closed (and possibly unbounded) interval. Due to duality, (P_γ) is feasible and has a bounded solution for every $\gamma \in \Gamma$, and the duality gap is zero.

Let $\mathcal{F}_\gamma$ denote the feasible set of (D_γ), $\mathbf{b}(\gamma) = \mathbf{v} + \mathbf{b} + \gamma\underline{\mathbf{b}}$, and $\boldsymbol{\xi}(\gamma) \in \{\boldsymbol{\xi}_\gamma : \boldsymbol{\xi}_\gamma = \arg\min\{\boldsymbol{\xi}_\gamma \cdot \mathbf{b}(\gamma), \boldsymbol{\xi}_\gamma \in \mathcal{F}_\gamma\}$. Moreover, let $\phi(\gamma; \mathbf{b}, \underline{\mathbf{b}}) = \boldsymbol{\xi}(\gamma) \cdot \mathbf{b}(\gamma)$ be the optimal value function. Obviously, we have $\phi(0; \mathbf{b}, \underline{\mathbf{b}}) = g(\mathbf{b})$ by the Slater's condition, and $\phi(\gamma; \mathbf{b}, \underline{\mathbf{b}})$ is *concave* in γ. We will use the shorthand notation $\phi(\gamma)$ henceforth.

Next, we derive the subdifferential of $\phi(\gamma)$, which is equivalent to $\partial g(\mathbf{b})$ when $\gamma = 0$ and $\underline{\mathbf{b}}$ is one of the canonical basis in $\mathbb{R}^{nm}$. For any $\gamma \in \operatorname{int}\Gamma$, choose $d\gamma$ small enough such that the point $\boldsymbol{\xi}(\gamma + d\gamma)$ lies in a compact set. Let $\boldsymbol{\xi}^+(\gamma)$ and $\boldsymbol{\xi}^-(\gamma)$ denote the limit as $d\gamma \to +0$ and -0, respectively.

Lemma A.3. *The equations*

$$\lim_{d\gamma \to +0} \frac{\mathbf{b}(\gamma) \cdot (\boldsymbol{\xi}(\gamma + d\gamma) - \boldsymbol{\xi}^+(\gamma))}{d\gamma} = 0$$

$$\lim_{d\gamma \to -0} \frac{\mathbf{b}(\gamma) \cdot (\boldsymbol{\xi}(\gamma + d\gamma) - \boldsymbol{\xi}^-(\gamma))}{d\gamma} = 0$$

hold for every $\gamma \in$ int Γ.

Proof. It is straightforward to verfiy that $\boldsymbol{\xi}^+(\gamma)$ is an optimal solution of (D_γ). Assume that

$$\lim_{d\gamma \to +0} \frac{\mathbf{b}(\gamma) \cdot (\boldsymbol{\xi}(\gamma + d\gamma) - \boldsymbol{\xi}^+(\gamma))}{d\gamma} \geq \epsilon > 0.$$

There exists a sequence $\{d\gamma_k\} \to +0$ such that

$$\begin{aligned}
&\mathbf{b}(\gamma + d\gamma_k) \cdot \boldsymbol{\xi}(\gamma + d\gamma_k) \\
&\geq \mathbf{b}(\gamma + d\gamma_k) \cdot \boldsymbol{\xi}^+(\gamma) + \epsilon d\gamma_k + d\gamma_k \underline{\mathbf{b}} \cdot (\boldsymbol{\xi}(\gamma + d\gamma_k) - \boldsymbol{\xi}^+(\gamma)) o(d\gamma_k) \\
&> \mathbf{b}(\gamma + d\gamma_k) \cdot \boldsymbol{\xi}^+(\gamma)
\end{aligned}$$

if $d\gamma_k$ is sufficiently small. This contradicts the optimality of $\boldsymbol{\xi}(\gamma + d\gamma_k)$ for $(\mathrm{D}_{\gamma + d\gamma_k})$. Similarly, assume that

$$\lim_{d\gamma \to +0} \frac{\mathbf{b}(\gamma) \cdot (\boldsymbol{\xi}(\gamma + d\gamma) - \boldsymbol{\xi}^+(\gamma))}{d\gamma} \leq \epsilon < 0$$

Then, there exists $\{d\gamma_k\} \to +0$ such that

$$\begin{aligned}
\mathbf{b}(\gamma) \cdot \boldsymbol{\xi}(\gamma + d\gamma_k) &\leq \mathbf{b}(\gamma) \cdot \boldsymbol{\xi}^+(\gamma) + \epsilon d\gamma_k + o(d\gamma_k) \\
&< \mathbf{b}(\gamma) \cdot \boldsymbol{\xi}^+(\gamma),
\end{aligned}$$

which contradicts that $\boldsymbol{\xi}^+(\gamma)$ is optimal for (D_γ). A similar argument can be made in the case where $d\gamma \to -0$. □

We now derive the directional derivative of $\phi(\gamma)$.

Lemma A.4. *The equations*

$$\lim_{d\gamma \to +0} \frac{\phi(\gamma + d\gamma) - \phi(\gamma)}{d\gamma} = \boldsymbol{\xi}^{+}(\gamma) \cdot \underline{\mathbf{b}}$$

$$\lim_{d\gamma \to -0} \frac{\phi(\gamma + d\gamma) - \phi(\gamma)}{d\gamma} = \boldsymbol{\xi}^{-}(\gamma) \cdot \underline{\mathbf{b}}$$

holds for every $\gamma \in int\, \Gamma$.

Proof. Since $\mathbf{b}(\gamma) \cdot \boldsymbol{\xi}(\gamma) = \mathbf{b}(\gamma) \cdot \boldsymbol{\xi}^{+}(\gamma) = \mathbf{b}(\gamma) \cdot \boldsymbol{\xi}^{-}(\gamma)$, one can write:

$$\begin{aligned}
&\lim_{d\gamma \to +0} \frac{\phi(\gamma + d\gamma) - \phi(\gamma)}{d\gamma} \\
&= \lim_{d\gamma \to +0} \frac{\boldsymbol{\xi}(\gamma + d\gamma) \cdot \mathbf{b}(\gamma + d\gamma) - \boldsymbol{\xi}(\gamma) \cdot \mathbf{b}(\gamma)}{d\gamma} \\
&= \lim_{d\gamma \to +0} \boldsymbol{\xi}(\gamma + d\gamma) \cdot \underline{\mathbf{b}} + \frac{\mathbf{b}(\gamma) \cdot (\boldsymbol{\xi}(\gamma + d\gamma) - \boldsymbol{\xi}(\gamma))}{d\gamma} \\
&= \boldsymbol{\xi}^{+}(\gamma) \cdot \underline{\mathbf{b}}
\end{aligned}$$

according to Lemma A.3. The proof for the case $d\gamma \to -0$ is similar. □

Notice that $\phi(\gamma)$ is continuously differentiable at γ if and only if $\underline{\mathbf{b}} \cdot \boldsymbol{\xi}^{+}(\gamma) = \underline{\mathbf{b}} \cdot \boldsymbol{\xi}^{-}(\gamma)$, which occurs either when (D_{γ}) has a unique solution or any feasible direction of the optimal face is orthogonal to $\underline{\mathbf{b}}$. To wrap up this section, we state the following lemma to bound the subdifferential $\partial g(\mathbf{b})$.

Lemma A.5. *Let $[\boldsymbol{\xi}^{+}(0)]_i$ and $[\boldsymbol{\xi}^{-}(0)]_i$ denote the i-th entry of $\boldsymbol{\xi}(d\gamma)$ as $d\gamma \to +0$ and -0 along the direction of the i-th canonical basis in $\mathbb{R}^{n_m}$. For every attack $\mathbf{b}$, assume that $0 \in \Gamma$. The subdifferential of $g(\mathbf{b})$ is bounded element-wise as*

$$[\boldsymbol{\xi}^{+}(0)]_i \leq [\partial g(\mathbf{b})]_i \leq [\boldsymbol{\xi}^{-}(0)]_i, \quad \forall i \in [n_m]$$

Proof. The proof follows from the strong duality between (P_{γ}) and (D_{γ}) at $\gamma = 0$, the concavity of $\phi(\gamma)$, and Theorem 24.1 in [195] on the monotonicity of subdifferential. □

To summarize, we have shown that $g(\mathbf{b})$ is continuous and convex (Lemma 7.3) with subdifferential depending on the dual solution (Lemma A.5). These results are useful for proving Theorem 7.5.

Proofs of Theorems 7.4 and 7.5

Theorem 7.4. *Let $\mathcal{V}(\mathcal{M}) \subset \mathbf{C}^{n_b}$ denote the set of observable states for a given set of measurement types $\mathcal{M}$ including the branch power flows and nodal voltage magnitudes, but not the nodal bus injections. Then, we have $\mathcal{V}(\mathcal{M}) \cap \mathcal{R}(\mathbf{Y}) \subseteq \mathcal{A}(\mathcal{M}, \rho)$ for some $\rho > 0$.*

Proof. For every $\hat{\mathbf{v}} \in \mathcal{V}(\mathcal{M}) \cap \mathcal{R}(\mathbf{Y})$, we show that by choosing $\mathbf{b} = \boldsymbol{f}(\hat{\mathbf{v}}) - \mathbf{v}$ where $f_i(\hat{\mathbf{v}})$ is given in (7.1), the unique optimal solution of (FDIA-SE) is given by $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*)$, hence $\hat{\mathbf{v}} \in \mathcal{A}(\mathcal{M}, \rho)$ is attackable for some ρ defined below. Let $\mathbf{M}_0$ in (SDP-FDIA) be given by the formula:

$$\mathbf{M}_0 = -\mathbf{I} + \epsilon \mathbf{v}_{tg}\mathbf{v}_{tg}^* + \sum_{l \in \mathcal{L}} \tilde{\mathbf{M}}_{pf}^{(l)} + \sum_{l \in \mathcal{L}} \tilde{\mathbf{M}}_{pt}^{(l)}, \tag{A.73}$$

where $\epsilon > 0$ is a constant parameter, and $\tilde{\mathbf{M}}_{pf}^{(l)}$ and $\tilde{\mathbf{M}}_{pt}^{(l)}$ are arbitrary matrices in $\mathbb{H}^{n_b}$. For every $(s,t) \in [n_b] \times [n_b]$, assume that the (s,t) entries of $\tilde{\mathbf{M}}_{pf}^{(l)}$ and $\tilde{\mathbf{M}}_{pt}^{(l)}$ are equal to zero if $(s,t) \notin \mathcal{L}$ and otherwise satisfy the following inequalities:

$$-\pi \le \angle y_{st} - \angle \tilde{M}_{pf,st}^{(l)} \le 0 \tag{A.74}$$

$$\pi \le \angle y_{st} + \angle \tilde{M}_{pt,st}^{(l)} \le 2\pi. \tag{A.75}$$

Choose $\rho = \mathbf{M}_0$ defined in (A.73) accordingly.

Let $\boldsymbol{\xi} \in \mathbb{R}^{n_m}$ and $\mathbf{Q} = \begin{bmatrix} q_0 & \boldsymbol{q}^* \\ \boldsymbol{q} & \mathbf{Q}_0 \end{bmatrix} \in \mathbb{H}^{n_b+1}$ be the dual variables. By the KKT conditions for optimality, we have: a) the stationarity conditions: $\boldsymbol{q} = -\mathbf{v}_{tg}$ and $\mathbf{Q}_0 = \mathbf{I} + \mathbf{M}_0 + \sum_i \xi_i \mathbf{M}_i$, b) the dual feasibility condition: $\mathbf{Q} \succeq 0$, and c) the complementary slackness condition: $\mathbf{Q}\begin{bmatrix} 1 & \mathbf{v}^* \\ \mathbf{v} & \mathbf{W} \end{bmatrix} = \mathbf{0}$. Let $\mathbf{H}(\boldsymbol{\xi}) = -\frac{1}{q_0}\mathbf{v}_{tg}\mathbf{v}_{tg}^* + \mathbf{Q}_0$ and $q_0 = \mathbf{v}_{tg}^*\mathbf{v}$. Based on a) and c), we have $\mathbf{H}(\boldsymbol{\xi})\mathbf{W} = \mathbf{0}$. Due to b) and Schur complement, it is required that $\mathbf{H}(\boldsymbol{\xi}) \succeq 0$.

By Slater's condition, strong duality holds if one can construct a strictly feasible dual solution $\hat{\boldsymbol{\xi}}$, which is optimal if KKT conditions are satisfied. The rank-1 condition for $\mathbf{W}$ follows if we can further show that $\text{rank}(\mathbf{H}(\hat{\boldsymbol{\xi}})) = n_b - 1$ (since together with $\mathbf{H}(\hat{\boldsymbol{\xi}})\mathbf{W} = \mathbf{0}$, it implies that $\mathbf{W}$ lies in the null space of $\mathbf{H}(\hat{\boldsymbol{\xi}})$, which is at most rank 1).

For the three types of measurements considered in this paper, the measurement matrices are: 1) $\mathbf{M}_i = \mathbf{E}_i$ for every $i \in \mathcal{N}$ (associated with voltage magnitudes), 2) $\mathcal{M}_{i+n_b} = \mathbf{Y}_{pf}^{(l)}$ for every $i \in \mathcal{L}$ (associated with real power flow *from* the bus), and 3) $\mathcal{M}_{i+n_b+n_l} = \mathbf{Y}_{pt}^{(l)}$ for every $i \in \mathcal{L}$ (associated with real power flow *to* the bus). By denoting $\hat{\boldsymbol{\xi}} = \sum_{l \in \mathcal{L}} \hat{\boldsymbol{\xi}}_{pf}^{(l)} + \sum_{l \in \mathcal{L}} \hat{\boldsymbol{\xi}}_{pt}^{(l)}$, we can write

$$\mathbf{H}(\hat{\boldsymbol{\xi}}) = \sum_{l \in \mathcal{L}} \mathbf{H}_{pf}^{(l)}(\hat{\boldsymbol{\xi}}_{pf}^{(l)}) + \sum_{l \in \mathcal{L}} \mathbf{H}_{pt}^{(l)}(\hat{\boldsymbol{\xi}}_{pt}^{(l)}),$$

where

$$\mathbf{H}_{pf}^{(l)}(\hat{\boldsymbol{\xi}}_{pf}^{(l)}) = \tilde{\mathbf{M}}_{pf}^{(l)} + \hat{\xi}_{pf,s}^{(l)}\mathbf{E}_s + \hat{\xi}_{pf,t}^{(l)}\mathbf{E}_t + \hat{\xi}_{pf,l+n_b}^{(l)}\mathbf{Y}_{pf}^{(l)}$$

$$\mathbf{H}_{pt}^{(l)}(\hat{\boldsymbol{\xi}}_{pt}^{(l)}) = \tilde{\mathbf{M}}_{pt}^{(l)} + \hat{\xi}_{pt,s}^{(l)}\mathbf{E}_s + \hat{\xi}_{pt,t}^{(l)}\mathbf{E}_t + \hat{\xi}_{pt,l+n_l+n_b}^{(l)}\mathbf{Y}_{pt}^{(l)}$$

and $\sum_{l\in\mathcal{L}}\tilde{\mathbf{M}}_{pf}^{(l)}+\sum_{l\in\mathcal{L}}\tilde{\mathbf{M}}_{pt}^{(l)}=\mathbf{I}+\mathbf{M}_0-\frac{1}{q_0}\mathbf{v}_{tg}\mathbf{v}_{tg}^*$. Define $\hat{\boldsymbol{\xi}}_{pf}^{(l)}$ in such a way that

$$\begin{aligned}\hat{\xi}_{pf,l+n_b}^{(l)}&=-\frac{2\Im\left(\hat{v}_s\hat{v}_t^*\tilde{M}_{pf,st}^{(l)*}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}^*\right)},\hat{\xi}_{pf,t}^{(l)}=\frac{|\hat{v}_s|^2\Im\left(\tilde{M}_{pf,st}^{(l)*}y_{st}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}^*\right)}\\ \hat{\xi}_{pf,s}^{(l)}&=\frac{|\hat{v}_t|^2}{|\hat{v}_s|^2}\hat{\xi}_{pf,t}^{(l)}+\Re(y_{st})\hat{\xi}_{pf,l+n_b}^{(l)}\end{aligned}\tag{A.76}$$

and $\hat{\boldsymbol{\xi}}_{pt}^{(l)}$ such that

$$\begin{aligned}\hat{\xi}_{pt,l+n_b+n_l}^{(l)}&=-\frac{2\Im\left(\hat{v}_s\hat{v}_t^*\tilde{M}_{pt,st}^{(l)*}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}\right)},\hat{\xi}_{pt,t}^{(l)}=-\frac{|v_s|^2\Im\left(\tilde{M}_{pt,st}^{(l)}y_{st}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}\right)}\\ \hat{\xi}_{pt,s}^{(l)}&=\frac{|\hat{v}_t|^2}{|\hat{v}_s|^2}\hat{\xi}_{pt,t}^{(l)}+\Re(y_{st})\hat{\xi}_{pt,l+n_b+n_l}^{(l)}\end{aligned}\tag{A.77}$$

where $\hat{\mathbf{v}}$ is an optimal solution of the primal problem (FDIA-SE). It can be verified that $\mathbf{H}_{pf}^{(l)}\hat{\mathbf{v}}=\mathbf{0}$, $\mathbf{H}_{pt}^{(l)}\hat{\mathbf{v}}=\mathbf{0}$, $\mathbf{H}_{pf}^{(l)}\succeq 0$ and $\mathbf{H}_{pt}^{(l)}\succeq 0$, as long as:

$$-\pi\le\angle\hat{v}_s-\angle\hat{v}_t-\angle y_{st}\le 0\tag{A.78}$$

$$0\le\angle\hat{v}_s-\angle\hat{v}_t+\angle y_{st}\le\pi\tag{A.79}$$

$$-\pi\le\angle y_{st}-\angle\tilde{M}_{pf,st}^{(l)}\le 0\tag{A.80}$$

$$\pi\le\angle y_{st}+\angle\tilde{M}_{pt,st}^{(l)}\le 2\pi.\tag{A.81}$$

The inequalities (A.78) and (A.79) are satisfied since $\hat{\mathbf{v}}\in\mathcal{R}(\mathbf{Y})$. The inequalities (A.80) and (A.81) require that $\tilde{M}_{pf,st}^{(l)}$ and $\tilde{M}_{pt,st}^{(l)}$ to lie in the second or third quadrants of the complex plane, which is satisfied by the design in (A.74) and (A.75).

Our next goal is to show that $\operatorname{rank}(\mathbf{H}(\hat{\boldsymbol{\xi}}))=n_b-1$, or equivalently, $\dim(\operatorname{null}(\mathbf{H}(\hat{\boldsymbol{\xi}})))=1$. For every $\mathbf{x}\in\operatorname{null}(\mathbf{H}(\hat{\boldsymbol{\xi}}))$, since $\mathbf{H}_{pf}^{(l)}\succeq 0$ and $\mathbf{H}_{pt}^{(l)}\succeq 0$, we have $\mathbf{H}_{pf}^{(l)}\mathbf{x}=\mathbf{H}_{pt}^{(l)}\mathbf{x}=\mathbf{0}$. By the construction of (A.76) and (A.77), for every line l with the endpoints s and t, it holds that $\frac{x_s}{\hat{v}_s}=\frac{x_t}{\hat{v}_t}$. This reasoning can be applied to another line $l':(t,a)$ to obtain $\frac{x_t}{\hat{v}_t}=\frac{x_a}{\hat{v}_a}$. By repeating the argument over a connected spanning graph of the network, one can obtain:

$$\frac{x_s}{\hat{v}_s}=\frac{x_t}{\hat{v}_t}=\frac{x_a}{\hat{v}_a}=\cdots=c\tag{A.82}$$

which indicates that $\mathbf{x}=\gamma\hat{\mathbf{v}}$. As a result, $\dim(\operatorname{null}(\mathbf{H}(\hat{\boldsymbol{\xi}})))=1$ and $\operatorname{rank}(\mathbf{H}(\hat{\boldsymbol{\xi}}))=n_b-1$. By the complementary slackness condition, it can be concluded that $\operatorname{rank}(\hat{\mathbf{W}})=1$. By Lemma 7.2, we have $\hat{\mathbf{W}}=\hat{\mathbf{v}}\hat{\mathbf{v}}^*$. We also know that $\mathbf{b}$ is stealthy since $\operatorname{trace}(\hat{\mathbf{v}}^*\mathbf{M}_i\hat{\mathbf{v}})=m_i+b_i$, $\forall i\in[n_m]$ by choice. □

Theorem 7.5. *Consider (SDP-FDIA) for a "target state attack" with* $\bar{h}(\tilde{\mathbf{v}},\mathbf{W})=\operatorname{trace}(\mathbf{W})-\tilde{\mathbf{v}}^*\mathbf{v}_{tg}-\mathbf{v}_{tg}^*\tilde{\mathbf{v}}$, *where* $\mathbf{v}_{tg}\in\mathcal{V}(\mathcal{M})$ *is chosen to be observable. Let* $(\hat{\mathbf{v}},\hat{\mathbf{W}},\hat{\mathbf{b}})$ *denote an optimal solution of (SDP-FDIA) for an arbitrary* α *greater than or equal to* $2\|\partial g(\mathbf{b}^\star)\|_\infty$*. The* *difference between the sabotage scale of the solved attack and the oracle attack satisfies the inequalities:*

$$-2\alpha\|\hat{\mathbf{\Delta}}_{\mathcal{B}}\|_1 \leq g(\hat{\mathbf{b}}) - g(\mathbf{b}^\star) \leq \alpha\left(\|\hat{\mathbf{\Delta}}_{\mathcal{B}}\|_1 - \|\hat{\mathbf{\Delta}}_{\mathcal{B}^c}\|_1\right),$$

where $\hat{\mathbf{\Delta}} = \hat{\mathbf{b}} - \mathbf{b}^\star$ *is the difference with the oracle* $\mathbf{b}^\star$*.*

Proof. In what follows, we will derive performance bounds for $\hat{\mathbf{x}}$ compared to $\mathbf{b}^\star$. By the definition of $g(\mathbf{b})$ in (FDIA-SE), we can rewrite (SDP-FDIA) only in terms of $\mathbf{b}$ as

$$\max_{\mathbf{b}} \quad g(\mathbf{b}) + \alpha\|\mathbf{b}\|_1 \tag{P4}$$

Define $r(\mathbf{\Delta}) = g(\mathbf{b}^\star + \mathbf{\Delta}) - g(\mathbf{b}^\star) + \alpha(\|\mathbf{b}^\star + \mathbf{\Delta}\|_1 - \|\mathbf{b}^\star\|_1)$ and $\hat{\mathbf{\Delta}} = \hat{\mathbf{b}} - \mathbf{b}^\star$. The separability of the l_1-norm yields that

$$\begin{aligned}\|\mathbf{b}^\star + \hat{\mathbf{\Delta}}\|_1 &\geq \|\mathbf{b}^\star_{\mathcal{B}} + \hat{\mathbf{\Delta}}_{\mathcal{B}^c}\|_1 - \|\mathbf{b}^\star_{\mathcal{B}^c} + \hat{\mathbf{\Delta}}_{\mathcal{B}}\|_1\\ &= \|\mathbf{b}^\star_{\mathcal{B}}\|_1 + \|\hat{\mathbf{\Delta}}_{\mathcal{B}^c}\|_1 - \|\hat{\mathbf{\Delta}}_{\mathcal{B}}\|_1\\ &= \|\mathbf{b}^\star\|_1 + \|\hat{\mathbf{\Delta}}_{\mathcal{B}^c}\|_1 - \|\hat{\mathbf{\Delta}}_{\mathcal{B}}\|_1.\end{aligned}$$

Together with $r(\hat{\mathbf{\Delta}}) \leq 0$ that results from the optimality of $\hat{\mathbf{b}}$, we have proved the upper bound. For the lower bound, one can write:

$$\begin{aligned}g(\hat{\mathbf{b}}) - g(\mathbf{b}^\star) &\geq \langle \partial g(\mathbf{b}^\star), \hat{\mathbf{\Delta}}\rangle \geq -|\langle \partial g(\mathbf{b}^\star), \hat{\mathbf{\Delta}}\rangle| && \text{(A.83)}\\ &\geq -\|\partial g(\mathbf{b}^\star)\|_\infty \|\hat{\mathbf{\Delta}}\|_1 && \text{(A.84)}\\ &\geq -\frac{\alpha}{2}\left(\|\hat{\mathbf{\Delta}}_{\mathcal{B}}\|_1 + \|\hat{\mathbf{\Delta}}_{\mathcal{B}^c}\|_1\right) && \text{(A.85)}\\ &\geq -2\alpha\|\hat{\mathbf{\Delta}}_{\mathcal{B}}\|_1 && \text{(A.86)}\end{aligned}$$

where (A.83) is due to the convexity of $g(\mathbf{b})$ (Lemma 7.3), (A.84) is by Hölder's inequality, (A.85) is due to the assumption of α, and (A.86) is due to Lemma 7.4 (see [173, Lem. 1]). □

List of Figures

List of Tables

Bibliography

[1] Pieter Abbeel, Dmitri Dolgov, Andrew Y Ng, and Sebastian Thrun. “Apprenticeship learning for motion planning with application to parking lot navigation”. In: *Proc. of the IEEE International Conference on Intelligent Robots and Systems.* 2008, pp. 1083–1090.

[2] Pieter Abbeel and Andrew Y Ng. “Apprenticeship learning via inverse reinforcement learning”. In: *Proc. of the International Conference on Machine learning.* ACM. 2004, p. 1.

[3] Ali Abur and Antonio Gomez Exposito. *Power system state estimation: theory and implementation.* CRC press, 2004.

[4] Anna A Adamopoulou, Athanasios M Tryferidis, and Dimitrios K Tzovaras. “A context-aware method for building occupancy prediction”. In: *Energy and Buildings* 110 (2016), pp. 229–244.

[5] Yuvraj Agarwal et al. “Occupancy-driven energy management for smart building automation”. In: *Proc. of the ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building.* ACM. 2010, pp. 1–6.

[6] Ravindra K Ahuja and James B Orlin. “Inverse optimization”. In: *Operations Research* 49.5 (2001), pp. 771–783.

[7] Yousef Al Horr et al. “Occupant productivity and office indoor environment quality: A review of the literature”. In: *Building and Environment* 105 (2016), pp. 369–389.

[8] Anna Alberini and Massimo Filippini. “Response of residential electricity demand to price: The effect of measurement error”. In: *Energy Economics* 33.5 (2011), pp. 889–895.

[9] Adrian Albert and Ram Rajagopal. “Smart meter driven segmentation: What your consumption says about you”. In: *IEEE Transactions on Power Systems* 28.4 (2013), pp. 4019–4030.

[10] Zvi Artstein. “Linear systems with delayed controls: A reduction”. In: *IEEE Transactions on Automatic Control* 27.4 (1982), pp. 869–879.

[11] Stephen H Bach, Bryan He, Alexander Ratner, and Christopher Ré. “Learning the structure of generative models without labeled data”. In: *arXiv preprint arXiv:1703.00854* (2017).

[12] Patrick Bajari, Han Hong, and Stephen P Ryan. "Identification and estimation of a discrete game of complete information". In: *Econometrica* 78.5 (2010), pp. 1529–1568.

[13] Galen Barbose, Charles Goldman, and Bernie Neenan. "A survey of utility experience with real time pricing". In: *Lawrence Berkeley National Laboratory* (2004).

[14] Peter L Bartlett and Shahar Mendelson. "Rademacher and Gaussian complexities: Risk bounds and structural results". In: *The Journal of Machine Learning Research* 3 (2003), pp. 463–482.

[15] Tamer Başar and Geert Jan Olsder. *Dynamic noncooperative game theory*. SIAM, 1998.

[16] Aliasghar Baziar and Abdollah Kavousi-Fard. "Considering uncertainty in the optimal energy management of renewable micro-grids including storage devices". In: *Renewable Energy* 59 (2013), pp. 158–166.

[17] Nikolaos Bekiaris-Liberis and Miroslav Krstic. "Lyapunov stability of linear predictor feedback for distributed input delays". In: *IEEE Transactions on Automatic Control* 56.3 (2011), pp. 655–660.

[18] Dimitri P Bertsekas. *Nonlinear programming*. Athena Scientific, 1999.

[19] Christopher M Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[20] Severin Borenstein. "The long-run efficiency of real-time electricity pricing". In: *The Energy Journal* (2005), pp. 93–116.

[21] Azzedine Boukerche, Horacio ABF Oliveira, Eduardo F Nakamura, and Antonio AF Loureiro. "Localization systems for wireless sensor networks". In: *IEEE wireless Communications* 14.6 (2007).

[22] Brian Burke. "Gamification 2020: What is the future of gamification". In: *Gartner, Inc., Nov* 5 (2012).

[23] Sharon Burke and Emily Schneider. "Enemy number one for the electric grid: mother nature". In: *SAIS Review of International Affairs* 35.1 (2015), pp. 73–86.

[24] Davide Calì, Peter Matthes, Kristian Huchtemann, Rita Streblow, and Dirk Müller. "CO2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings". In: *Building and Environment* 86 (2015), pp. 39–49.

[25] Luis M Candanedo and Véronique Feldheim. "Accurate occupancy detection of an office room from light, temperature, humidity and CO 2 measurements using statistical learning models". In: *Energy and Buildings* 112 (2016), pp. 28–39.

[26] Rich Caruana. "Multitask learning". In: *Learning to learn*. Springer, 1998, pp. 95–133.

[27] Gilles Celeux and Gérard Govaert. "A classification EM algorithm for clustering and two stochastic versions". In: *Computational statistics & Data analysis* 14.3 (1992), pp. 315–332.

[28] Feng-Min Chang, Feng-Li Lian, and Chih-Chung Chou. "Integration of Modified Inverse Observation Model and Multiple Hypothesis Tracking for Detecting and Tracking Humans". In: *IEEE Transactions on Automation Science and Engineering* 13.1 (2016), pp. 160–170.

[29] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. "Semi-supervised learning". In: *IEEE Transactions on Neural Networks* 20.3 (2009), pp. 542–542.

[30] Dong Chen, Sean Barker, Adarsh Subbaswamy, David Irwin, and Prashant Shenoy. "Non-Intrusive Occupancy Monitoring using Smart Meters". In: *Proc. of the ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM. 2013, pp. 1–8.

[31] Feng Chen et al. "Activity analysis based on low sample rate smart meters". In: *Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data mining*. ACM. 2011, pp. 240–248.

[32] Jiayu Chen and Changbum Ahn. "Assessing occupants' energy load variation through existing wireless network infrastructure in commercial and educational buildings". In: *Energy and Buildings* 82 (2014), pp. 540–549.

[33] Jiayu Chen, John E Taylor, and Hsi-Hsien Wei. "Modeling building occupant network energy consumption decision-making: The interplay between network structure and conservation". In: *Energy and Buildings* 47 (2012), pp. 515–524.

[34] Yen-Haw Chen, Su-Ying Lu, Yung-Ruei Chang, Ta-Tung Lee, and Ming-Che Hu. "Economic analysis and optimal energy management models for microgrid systems: A case study in Taiwan". In: *Applied Energy* 103 (2013), pp. 145–154.

[35] Toby C.T. Cheung, Stefano Schiavon, Elliott T. Gall, Ming Jin, and William W Nazaroff. "Longitudinal assessment of thermal and perceived air quality acceptability in relation to temperature, humidity, and CO2 exposure in Singapore". In: *Building and Environment* 115 (2017), pp. 80–90.

[36] Benoît Colson, Patrice Marcotte, and Gilles Savard. "An overview of bilevel optimization". In: *Annals of operations research* 153.1 (2007), pp. 235–256.

[37] Andreas Damianou and Neil Lawrence. "Deep Gaussian Processes". In: *Proc. of the International Conference on Artificial Intelligence and Statistics*. 2013, pp. 207–215.

[38] Gyorgy Dan and Henrik Sandberg. "Stealth attacks and protection schemes for state estimators in power systems". In: *IEEE International Conference on Smart Grid Communications*. 2010, pp. 214–219.

[39] Naïm R Darghouth, Ryan H Wiser, Galen Barbose, and Andrew D Mills. "Net metering and market feedback loops: Exploring the impact of retail rate design on distributed PV deployment". In: *Applied Energy* 162 (2016), pp. 713–722.

[40] Cedric De Jonghe, Benjamin F Hobbs, and Ronnie Belmans. "Optimal generation mix with short-term demand response and wind penetration". In: *IEEE Transactions on Power Systems* 27.2 (2012), pp. 830–839.

[41] Erik Delarue, Pieterjan Van Den Bosch, and William D'haeseleer. "Effect of the accuracy of price forecasting on profit in a Price Based Unit Commitment". In: *Electric power systems research* 80.10 (2010), pp. 1306–1313.

[42] Paul Denny. "The effect of virtual achievements on student engagement". In: *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems.* 2013, pp. 763–772.

[43] Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, and Vijay Devabhaktuni. "Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft". In: *Energy Policy* 39.2 (2011), pp. 1007–1015.

[44] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. "From game design elements to gamefulness: defining gamification". In: *Proc. of the ACM International Academic MindTrek Conference: Envisioning Future Media Environments.* 2011, pp. 9–15.

[45] Maimouna Diagne, Mathieu David, Philippe Lauret, John Boland, and Nicolas Schmutz. "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids". In: *Renewable and Sustainable Energy Reviews* 27 (2013), pp. 65–76.

[46] Thomas G Dietterich, Richard H Lathrop, and Tomás Lozano-Pérez. "Solving the multiple instance problem with axis-parallel rectangles". In: *Artificial intelligence* 89.1 (1997), pp. 31–71.

[47] Aris L Dimeas and Nikos D Hatziargyriou. "Operation of a multiagent system for microgrid control". In: *IEEE Transactions on Power Systems* 20.3 (2005), pp. 1447–1455.

[48] Giovanni Diraco, Alessandro Leone, and Pietro Siciliano. "People occupancy detection and profiling with 3D depth sensors for building energy management". In: *Energy and Buildings* 92 (2015), pp. 246–266.

[49] Guzmán Díaz and Blanca Moreno. "Valuation under uncertain energy prices and load demands of micro-CHP plants supplemented by optimally switched thermal energy storage". In: *Applied Energy* 177 (2016), pp. 553–569.

[50] Bing Dong et al. "An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network". In: *Energy and Buildings* 42.7 (2010), pp. 1038–1046.

[51] Tao Dong et al. "Discovery-based games for learning software". In: *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems.* 2012, pp. 2083–2086.

[52] David L Donoho. "Compressed sensing". In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306.

[53] Meysam Doostizadeh and Hassan Ghasemi. "A day-ahead electricity pricing model based on smart metering and demand-side management". In: *Energy* 46.1 (2012), pp. 221–230.

[54] Gregory Druck, Burr Settles, and Andrew McCallum. "Active learning by labeling features". In: *Proc. of the ACL Conference on Empirical Methods in Natural Language Processing.* 2009, pp. 81–90.

[55] Yan Duan et al. "RL^2: Fast Reinforcement Learning via Slow Reinforcement Learning". In: *arXiv preprint arXiv:1611.02779* (2016).

[56] John C Duchi, Michael I Jordan, and Martin J Wainwright. "Local privacy and statistical minimax rates". In: *IEEE Annual Symposium on Foundations of Computer Science.* IEEE. 2013, pp. 429–438.

[57] David Duvenaud, Oren Rippel, Ryan P. Adams, and Zoubin Ghahramani. "Avoiding pathologies in very deep networks". In: *Proc. of the International Conference on Artificial Intelligence and Statistics.* 2014, pp. 202–210.

[58] Cynthia Dwork. "Differential privacy: A survey of results". In: *Proc. of the International Conference on Theory and Applications of Models of Computation.* Springer. 2008, pp. 1–19.

[59] Cynthia Dwork, Moni Naor, Toniann Pitassi, Guy N Rothblum, and Sergey Yekhanin. "Pan-private streaming algorithms". In: *Proceedings of ICS.* 2010.

[60] US Department of Energy. *Benefits of demand response in electricity markets and recommendations for achieving them.* The National Academies Press, 2006.

[61] Varick L Erickson, Miguel Á Carreira-Perpiñán, and Alberto E Cerpa. "Occupancy modeling and prediction for building energy management". In: *ACM Transactions on Sensor Networks* 10.3 (2014), p. 42.

[62] Pedro Faria and Zita Vale. "Demand response in electrical energy supply: An optimal real time pricing approach". In: *Energy* 36.8 (2011), pp. 5374–5384.

[63] Rosta Farzan et al. "Results from deploying a participation incentive mechanism within the enterprise". In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems.* ACM. 2008, pp. 563–572.

[64] Massimo Filippini. "Short-and long-run time-of-use price elasticities in Swiss residential electricity demand". In: *Energy policy* 39.10 (2011), pp. 5811–5817.

[65] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks". In: *arXiv preprint arXiv:1703.03400* (2017).

[66] William J Fisk, Usha Satish, Mark J Mendell, Toshifumi Hotchi, and Douglas Sullivan. "Is CO2 Indoor Pollutant?" In: *ASHRAE Journal* 55.3 (2013), p. 84.

[67] Sjur D Flåm. "Solving non-cooperative games by continuous subgradient projection methods". In: *System Modelling and Optimization.* Springer, 1990, pp. 115–123.

[68] Stephen E Flynn. "America the resilient-defying terrorism and mitigating natural disasters". In: *Foreign Aff.* 87 (2008), p. 2.

[69] David A. Freedman. “Statistical models: theory and practice”. In: Cambridge University Press, 2009.

[70] Monika Frontczak et al. “Quantitative relationships between occupant satisfaction and satisfaction aspects of indoor environmental quality and building design”. In: *Indoor air* 22.2 (2012), pp. 119–131.

[71] Girish Ghatikar, Salman Mashayekh, Michael Stadler, Rongxin Yin, and Zhenhua Liu. “Distributed energy systems integration and demand optimization for autonomous operations and electric grid transactions”. In: *Applied Energy* 167 (2016), pp. 432–448.

[72] Asish Ghoshal and Jean Honorio. “Learning Sparse Potential Games in Polynomial Time and Sample Complexity”. In: *arXiv preprint arXiv:1706.05648* (2017).

[73] Ian Goodfellow et al. “Generative adversarial nets”. In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[74] N. Groot, B. De Schutter, and H. Hellendoorn. “Reverse Stackelberg games, Part I: Basic framework”. In: *IEEE International Conference on Control Applications,* 2012, pp. 421–426.

[75] Weixi Gu, Ming Jin, Zimu Zhou, Costas J Spanos, and Lin Zhang. “MetroEye: Smart Tracking Your Metro Trips Underground”. In: *Proc. of the International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM. 2016, pp. 84–93.

[76] Weixi Gu et al. “Measuring fine-grained metro interchange time via smartphones”. In: *Transportation Research Part C: Emerging Technologies* 81 (2017), pp. 153–171.

[77] Rimas Gulbinas, Ardalan Khosrowpour, and John Taylor. “Segmentation and classification of commercial building occupants by energy-use efficiency and predictability”. In: *IEEE Transactions on Smart Grid* 6.3 (2015), pp. 1414–1424.

[78] Li Guo, Nan Wang, Hai Lu, Xialin Li, and Chengshan Wang. “Multi-objective optimal planning of the stand-alone microgrid system based on different benefit subjects”. In: *Energy* 116 (2016), pp. 353–363.

[79] Anton Gustafsson, Cecilia Katzeff, and Magnus Bang. “Evaluation of a pervasive game for domestic energy engagement among teenagers”. In: *Computers in Entertainment* 7.4 (2009), p. 54.

[80] Juho Hamari, Jonna Koivisto, and Harri Sarsa. “Does gamification work?–a literature review of empirical studies on gamification”. In: *IEEE Hawaii International Conference on System Sciences*. 2014, pp. 3025–3034.

[81] AR Hatami, H Seifi, and MK Sheikh-El-Eslami. “Optimal selling price and energy procurement strategies for a retailer in an electricity market”. In: *Electric Power Systems Research* 79.1 (2009), pp. 246–254.

[82] AD Hawkes and MA Leach. "Modelling high level system design and unit commitment for a microgrid". In: *Applied energy* 86.7 (2009), pp. 1253–1265.

[83] Julien M Hendrickx, Karl Henrik Johansson, Raphael M Jungers, Henrik Sandberg, and Kin Cheong Sou. "Efficient computations of a security index for false data attacks in power networks". In: *IEEE Transactions on Automatic Control* 59.12 (2014), pp. 3194–3208.

[84] Tianzhen Hong, Hongsan Sun, Yixing Chen, Sarah C Taylor-Lange, and Da Yan. "An occupant behavior modeling tool for co-simulation". In: *Energy and Buildings* 117 (2016), pp. 272–281.

[85] Ching-Chun Huang and Sheng-Jyh Wang. "A bayesian hierarchical framework for multitarget labeling and correspondence with ghost suppression over multicamera surveillance system". In: *IEEE Transactions on Automation Science and Engineering* 9.1 (2012), pp. 16–30.

[86] Gabriela Hug and Joseph Andrew Giampapa. "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks". In: *IEEE Transactions on Smart Grid* 3.3 (2012), pp. 1362–1370.

[87] Ruoxi Jia, Roy Dong, Prashanth Ganesh, Shankar Sastry, and Costas Spanos. "Towards a Theory of Free-Lunch Privacy in Cyber-Physical Systems". In: *Proc. of the IEEE Annual Allerton Conference on Communication, Control, and Computing.* 2017.

[88] Ruoxi Jia, Roy Dong, S Shankar Sastry, and Costas J Spanos. "Privacy-enhanced architecture for occupancy-based HVAC control". In: *Proc. of the ACM International Conference on Cyber-Physical Systems.* 2017, pp. 177–186.

[89] Ruoxi Jia, Yang Gao, and Costas J Spanos. "A fully unsupervised non-intrusive load monitoring framework". In: *Proc. of the IEEE International Conference on Smart Grid Communications.* 2015, pp. 872–878.

[90] Ruoxi Jia, Ming Jin, Zilong Chen, and Costas J Spanos. "SoundLoc: Accurate room-level indoor localization using acoustic signatures". In: *IEEE International Conference on Automation Science and Engineering.* 2015, pp. 186–193.

[91] Ruoxi Jia and Costas Spanos. "Occupancy modelling in shared spaces of buildings: a queueing approach". In: *Journal of Building Performance Simulation* 10.4 (2017), pp. 406–421.

[92] Ruoxi Jia et al. "Design automation for smart buildings". In: *Proceedings of IEEE (submitted)* (2017).

[93] Ruoxi Jia et al. "MapSentinel: Can the Knowledge of Space Use Improve Indoor Tracking Further?" In: *Sensors* 16.4 (2016), p. 472.

[94] Ming Jin, Nikolaos Bekiaris-Liberis, Kevin Weekly, Costas J Spanos, and Alexandre M Bayen. "Occupancy Detection via Environmental Sensing". In: *IEEE Transactions on Automation Science and Engineering* PP.99 (2017), pp. 1–13.

[95] Ming Jin, Nikos Bekiaris-Liberis, Kevin Weekly, Costas Spanos, and Alexandre M. Bayen. "Sensing by proxy: Occupancy detection based on indoor CO2 concentration". In: *Proc. of the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. 2015, pp. 1–10.

[96] Ming Jin, Andreas Damianou, Pieter Abbeel, and Costas Spanos. "Inverse Reinforcement Learning via Deep Gaussian Process". In: *The Conference on Uncertainty in Artificial Intelligence*. 2017.

[97] Ming Jin, Wei Feng, Ping Liu, Chris Marnay, and Costas Spanos. "MOD-DR: Microgrid optimal dispatch with demand response". In: *Applied Energy* 187 (2017), pp. 758–776.

[98] Ming Jin, Wei Feng, Chris Marnay, and Costas Spanos. "Microgrid to enable optimal distributed energy retail and end-user demand response". In: *Applied Energy* 210 (2018), pp. 1321–1335.

[99] Ming Jin, R. Jia, and C. Spanos. "Virtual Occupancy Sensing: Using Smart Meters to Indicate Your Presence". In: *IEEE Transactions on Mobile Computing* 16.11 (2017), pp. 3264–3277.

[100] Ming Jin, Ruoxi Jia, Zhaoyi Kang, Ioannis C. Konstantakopoulos, and Costas J. Spanos. "PresenceSense: Zero-training Algorithm for Individual Presence Detection Based on Power Monitoring". In: *Proc. of the ACM Conference on Embedded Systems for Energy-Efficient Buildings*. 2014, pp. 1–10.

[101] Ming Jin, Ruoxi Jia, and Costas Spanos. "APEC: Auto Planner for Efficient Configuration of Indoor Positioning Systems". In: *Proc. of the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. 2015, pp. 100–107.

[102] Ming Jin, Javad Lavaei, and Karl Johansson. "A Semidefinite Programming Relaxation under False Data Injection Attacks against Power Grid AC State Estimation". In: *Proc. of the IEEE Annual Allerton Conference on Communication, Control, and Computing*. 2017.

[103] Ming Jin, Javad Lavaei, and Karl Johansson. "Power grid AC-based state estimation: vulnerability analysis against cyber attack". In: (2017). URL: http://www.ieor.berkeley.edu/~lavaei/FDIA_AC_2017.pdf.

[104] Ming Jin, Shichao Liu, Stefano Schiavon, and Costas Spanos. "Automated mobile sensing: Towards high-granularity agile indoor environmental quality monitoring". In: *Building and Environment* 127 (2018), pp. 268–276.

[105] Ming Jin, Chris Marnay, and Wei Feng. "Distributed energy resource integration by dispatch and retail optimization". In: *IEEE Power & Energy Society Innovative Smart Grid Technologies*. 2017. URL: http://www.jinming.tech/papers/der_isgt.pdf.

[106] Ming Jin, Lillian J Ratliff, Ioannis Konstantakopoulos, Costas Spanos, and Shankar Sastry. "REST: a reliable estimation of stopping time algorithm for social game experiments". In: *Proc. of the ACM/IEEE International Conference on Cyber-Physical Systems*. ACM. 2015, pp. 90–99.

[107] Ming Jin and Costas J Spanos. "BRIEF: Bayesian regression of infinite expert forecasters for single and multiple time series prediction". In: *Proc. of the IEEE Conference on Decision and Control*. IEEE. 2015, pp. 78–83.

[108] Ming Jin, Lin Zhang, and Costas Spanos. "Power Prediction through Energy Consumption Pattern Recognition for Smart Buildings". In: *IEEE International Conference on Automation Science and Engineering*. 2015, pp. 419–424.

[109] Ming Jin et al. "Environmental sensing by wearable device for indoor activity and location estimation". In: *Proc. of the Annual Conference of the IEEE Industrial Electronics Society*. 2014, pp. 5369–5375.

[110] Ming Jin et al. "Indoor environmental quality monitoring by autonomous mobile sensing". In: *Proc. of the ACM Conference on Embedded Systems for Energy-Efficient Buildings*. 2017.

[111] Thorsten Joachims. "Optimizing search engines using clickthrough data". In: *Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data mining*. ACM. 2002, pp. 133–142.

[112] Ian T Jolliffe. "Principal Component Analysis and Factor Analysis". In: *Principal component analysis*. Springer, 1986, pp. 115–128.

[113] JH Jung, Christoph Schneider, and Joseph Valacich. "Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments". In: *Management science* 56.4 (2010), pp. 724–742.

[114] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[115] Sanket Kamthe and Marc Peter Deisenroth. "Data-Efficient Reinforcement Learning with Probabilistic Model Predictive Control". In: *arXiv preprint arXiv:1706.06491* (2017).

[116] Farhad Kamyab, Mohammadhadi Amini, Siamak Sheykhha, Mehrdad Hasanpour, and Mohammad Majid Jalali. "Demand response program in smart grid using supply function bidding mechanism". In: *IEEE Transactions on Smart Grid* 7.3 (2016), pp. 1277–1284.

[117] Melih Kandemir. "Asymmetric Transfer Learning with Deep Gaussian Processes". In: *Proc. of the International Conference on Machine Learning*. 2015, pp. 730–738.

[118] Zhaoyi Kang, Ming Jin, and Costas J Spanos. "Modeling of end-use energy profile: An appliance-data-driven stochastic approach". In: *Proc. of the Annual Conference of the IEEE Industrial Electronics Society*. 2014, pp. 5382–5388.

[119] Zhaoyi Kang, Yuxun Zhou, Lin Zhang, and Costas J Spanos. "Virtual power sensing based on a multiple-hypothesis sequential test". In: *Proc. of the IEEE International Conference on Smart Grid Communications*. 2013, pp. 785–790.

[120] Iasson Karafyllis. "Finite-time global stabilization by means of time-varying distributed delay feedback". In: *SIAM Journal on Control and Optimization* 45.1 (2006), pp. 320–342.

[121] James Kennedy. "Particle swarm optimization". In: *Encyclopedia of machine learning*. Springer, 2011, pp. 760–766.

[122] Arezou Keshavarz, Yang Wang, and Stephen Boyd. "Imputing a convex objective function". In: *IEEE Intern. Symp. on Intelligent Control*. 2011, pp. 613–619.

[123] Jin-Ho Kim and Anastasia Shcherbakova. "Common failures of demand response". In: *Energy* 36.2 (2011), pp. 873–880.

[124] Seung-Jun Kim and Georgios Giannakis. "Scalable and robust demand response with mixed-integer constraints". In: *IEEE Transactions on Smart Grid* 4.4 (2013), pp. 2089–2099.

[125] Youngjin Kim and Leslie K Norford. "Optimal use of thermal energy storage resources in commercial buildings through price-based demand response considering distribution network operation". In: *Applied Energy* 193 (2017), pp. 308–324.

[126] Diederik P Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *Proc. of the International Conference on Learning Representations*. 2013.

[127] Michael CW Kintner-Meyer, Charles Goldman, Osman Sezgen, and Donna Pratt. *Dividends with demand response*. Tech. rep. Pacific Northwest National Laboratory, Richland, WA, 2003.

[128] Daniel S Kirschen and Goran Strbac. *Fundamentals of power system economics*. John Wiley & Sons, 2004.

[129] Wilhelm Kleiminger, Christian Beckel, and Silvia Santini. "Household occupancy monitoring using electricity meters". In: *Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. 2015, pp. 975–986.

[130] Wilhelm Kleiminger, Christian Beckel, Thorsten Staake, and Silvia Santini. "Occupancy detection from electricity consumption data". In: *Proc. of the ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM. 2013, pp. 1–8.

[131] J Zico Kolter, Pieter Abbeel, and Andrew Y Ng. "Hierarchical apprenticeship learning with application to quadruped locomotion". In: *Advances in Neural Information Processing Systems*. 2007, pp. 769–776.

[132] I. C. Konstantakopoulos, L. J. Ratliff, M. Jin, S. S. Sastry, and C. J. Spanos. "A Robust Utility Learning Framework via Inverse Optimization". In: *IEEE Transactions on Control Systems Technology* PP.99 (2017), pp. 1–17.

[133] Ioannis C Konstantakopoulos, Lillian J Ratliff, Ming Jin, and Costas J Spanos. "Leveraging correlations in utility learning". In: *Proc. of the IEEE American Control Conference*. 2017, pp. 5249–5256.

[134] Ioannis C Konstantakopoulos, Lillian J Ratliff, Ming Jin, Costas J Spanos, and S Shankar Sastry. "Inverse modeling of non-cooperative agents via mixture of utilities". In: *Proc. of the IEEE Conference on Decision and Control*. 2016, pp. 6327–6334.

[135] Ioannis C Konstantakopoulos, Lillian J Ratliff, Ming Jin, Costas Spanos, and S Shankar Sastry. "Smart building energy efficiency via social game: A robust utility learning framework for closing–the–loop". In: *IEEE International Workshop on Science of Smart City Operations and Platforms Engineering*. 2016, pp. 1–6.

[136] Oliver Kosut, Liyan Jia, Robert J Thomas, and Lang Tong. "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures". In: *IEEE International Conference on Smart Grid Communications*. 2010, pp. 220–225.

[137] W Kwon and A Pearson. "Feedback stabilization of linear systems with delayed control". In: *IEEE Transactions on Automatic control* 25.2 (1980), pp. 266–269.

[138] Timilehin Labeodan, Wim Zeiler, Gert Boxem, and Yang Zhao. "Occupancy measurement in commercial office buildings for demand-driven control applications – A survey and detection system evaluation". In: *Energy and Buildings* 93 (2015), pp. 303–314.

[139] Khee Poh Lam et al. "Occupancy detection through an extensive environmental sensor network in an open-plan office building". In: *IBPSA Building Simulation* (2009), pp. 1452–1459.

[140] Oscar D Lara and Miguel A Labrador. "A survey on human activity recognition using wearable sensors." In: *IEEE Communications Surveys and Tutorials* 15.3 (2013), pp. 1192–1209.

[141] Robert H Lasseter. "Microgrids". In: *IEEE Power Engineering Society Winter Meeting*. Vol. 1. 2002, pp. 305–308.

[142] Sergey Levine, Zoran Popovic, and Vladlen Koltun. "Feature construction for inverse reinforcement learning". In: *Advances in Neural Information Processing Systems*. 2010, pp. 1342–1350.

[143] Sergey Levine, Zoran Popovic, and Vladlen Koltun. "Nonlinear inverse reinforcement learning with Gaussian processes". In: *Advances in Neural Information Processing Systems*. 2011, pp. 19–27.

[144] Canbing Li et al. "Comprehensive review of renewable energy curtailment and avoidance: a specific example in China". In: *Renewable and Sustainable Energy Reviews* 41 (2015), pp. 1067–1079.

[145] Gaoqi Liang, Junhua Zhao, Fengji Luo, Steven Weller, and Zhao Yang Dong. "A review of false data injection attacks against modern power systems". In: *IEEE Transactions on Smart Grid* 8.4 (2016), pp. 1630–1638.

[146] Jingwen Liang, Oliver Kosut, and Lalitha Sankar. "Cyber attacks on AC state estimation: Unobservability and physical consequences". In: *IEEE PES General Meeting—Conference & Exposition*. 2014, pp. 1–5.

[147] Hui Liu, Houshang Darabi, Pat Banerjee, and Jing Liu. "Survey of wireless indoor positioning techniques and systems". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C* 37.6 (2007), pp. 1067–1080.

[148] Yao Liu, Peng Ning, and Michael K Reiter. "False data injection attacks against state estimation in electric power grids". In: *ACM Transactions on Information and System Security* 14.1 (2011), p. 13.

[149] Victoria López, Alberto Fernández, Jose G Moreno-Torres, and Francisco Herrera. "Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics". In: *Expert Systems with Applications* 39.7 (2012), pp. 6585–6608.

[150] Henrik Lund and Ebbe Münster. "Integrated energy systems and local energy markets". In: *Energy Policy* 34.10 (2006), pp. 1152–1160.

[151] Ramtin Madani, Javad Lavaei, and Ross Baldick. "Convexification of power flow equations for power systems in presence of noisy measurements". In: (2016). URL: http://www.ieor.berkeley.edu/~lavaei/SE_J_2016.pdf.

[152] Ramtin Madani, Javad Lavaei, and Ross Baldick. "Convexification of power flow problem over arbitrary networks". In: *IEEE Conference on Decision and Control*. 2015, pp. 1–8.

[153] Patrice Marcotte, Gilles Savard, and DL Zhu. "A trust region algorithm for nonlinear bilevel programming". In: *Operations research letters* 29.4 (2001), pp. 171–179.

[154] Chris Marnay et al. "Optimal technology selection and operation of commercial-building microgrids". In: *IEEE Transactions on Power Systems* 23.3 (2008), pp. 975–982.

[155] Claudio Martani, David Lee, Prudence Robinson, Rex Britter, and Carlo Ratti. "ENERNET: Studying the dynamic relationship between building occupancy and energy consumption". In: *Energy and Buildings* 47 (2012), pp. 584–591.

[156] César Lincoln C Mattos et al. "Recurrent Gaussian Processes". In: *Proc. of the International Conference on Learning Representations* (2016).

[157] Frédéric Mazenc and P-A Bliman. "Backstepping design for time-delay nonlinear systems". In: *IEEE Transactions on Automatic Control* 51.1 (2006), pp. 149–154.

[158] James McCalley. *Lecture notes EE 553: Steady-state analysis - Power system operation and control.* http://home.engineering.iastate.edu/~jdm/ee553/SE1.pdf. Accessed: 2017-9-21.

[159] Eoghan McKenna, Ian Richardson, and Murray Thomson. “Smart meter data: Balancing consumer privacy concerns with legitimate applications”. In: *Energy Policy* 41 (2012), pp. 807–814.

[160] Charles McParland. “OpenADR open source toolkit: Developing open source software for the smart grid”. In: *IEEE Power and Energy Society General Meeting.* 2011, pp. 1–7.

[161] Sean Meyn et al. “A sensor-utility-network method for estimation of occupancy in buildings”. In: *Proc. of the IEEE Conference on Decision and Control.* IEEE. 2009, pp. 1494–1500.

[162] Tom M Mitchell et al. “Never Ending Learning.” In: *AAAI.* 2015, pp. 2302–2310.

[163] Volodymyr Mnih et al. “Human-level control through deep reinforcement learning”. In: *Nature* 518.7540 (2015), pp. 529–533.

[164] Amjad Anvari Moghaddam, Alireza Seifi, and Taher Niknam. “Multi-operation management of a typical micro-grids using Particle Swarm Optimization: A comparative study”. In: *Renewable and Sustainable Energy Reviews* 16.2 (2012), pp. 1268–1281.

[165] M Parsa Moghaddam, A Abdollahi, and M Rashidinejad. “Flexible demand response programs modeling in competitive electricity markets”. In: *Applied Energy* 88.9 (2011), pp. 3257–3269.

[166] Amir-Hamed Mohsenian-Rad and Alberto Leon-Garcia. “Optimal residential load control with price prediction in real-time electricity pricing environments”. In: *IEEE transactions on Smart Grid* 1.2 (2010), pp. 120–133.

[167] Andrés Molina-Markham, Prashant Shenoy, Kevin Fu, Emmanuel Cecchet, and David Irwin. “Private memoirs of a smart meter”. In: *Proc. of the ACM workshop on embedded sensing systems for energy-efficiency in building.* ACM. 2010, pp. 61–66.

[168] Mathew J Morey and Laurence D Kirsch. “Retail Choice in Electricity: What Have We Learned in 20 Years?” In: *Washington, DC: Christensen Associates Energy Consulting LLC for Electric Markets Research Foundation* (2016).

[169] Cecilia Munoz, Megan Smith, and D Patil. “Big data: A report on algorithmic systems, opportunity, and civil rights”. In: *Executive Office of the President. The White House* (2016).

[170] Anindya Nag and Subhas Chandra Mukhopadhyay. “Occupancy Detection at Smart Home Using Real-Time Dynamic Thresholding of Flexiforce Sensor”. In: *IEEE Sensors Journal* 15.8 (2015), pp. 4457–4463.

[171] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. "Learning with noisy labels". In: *Advances in neural information processing systems*. 2013, pp. 1196–1204.

[172] National Academies of Sciences, Engineering, and Medicine. *Enhancing the Resilience of the Nation's Electricity System*. Washington, DC: The National Academies Press, 2017.

[173] Sahand N Negahban, Pradeep Ravikumar, Martin J Wainwright, and Bin Yu. "A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizers". In: *Statistical Science* 27.4 (2012), pp. 538–557.

[174] Yu Nesterov. "Gradient methods for minimizing composite functions". In: *Mathematical Programming* 140.1 (2013), pp. 125–161.

[175] Andrew Y Ng, Stuart J Russell, et al. "Algorithms for inverse reinforcement learning." In: *Proc. of the International Conference on Machine Learning*. 2000, pp. 663–670.

[176] Duong Tung Nguyen and Long Bao Le. "Risk-constrained profit maximization for microgrid aggregators with demand response". In: *IEEE Transactions on Smart Grid* 6.1 (2015), pp. 135–146.

[177] Taher Niknam, Rasoul Azizipanah-Abarghooee, and Mohammad Rasoul Narimani. "An efficient scenario-based stochastic programming framework for multi-objective optimal micro-grid operation". In: *Applied Energy* 99 (2012), pp. 455–470.

[178] Barack Obama. "The irreversible momentum of clean energy". In: *Science* 355.6321 (2017), pp. 126–129.

[179] Torben Ommen, Wiebke Brix Markussen, and Brian Elmegaard. "Comparison of linear, mixed integer and non-linear programming methods in energy system dispatch modelling". In: *Energy* 74 (2014), pp. 109–118.

[180] Sinno Jialin Pan and Qiang Yang. "A survey on transfer learning". In: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (2010), pp. 1345–1359.

[181] Fabio Pasqualetti, Ruggero Carli, and Francesco Bullo. "A distributed method for state estimation and false data detection in power networks". In: *IEEE International Conference on Smart Grid Communications*. 2011, pp. 469–474.

[182] Krystian X Perez et al. "Nonintrusive disaggregation of residential air-conditioning loads from sub-hourly smart meter data". In: *Energy and Buildings* 81 (2014), pp. 316–325.

[183] Anna Laura Pisello, Michael Bobker, and Franco Cotana. "A building energy efficiency optimization method by evaluating the effective thermal zones occupancy". In: *Energies* 5.12 (2012), pp. 5257–5278.

[184] Emmanouil Antonios Platanios, Avinava Dubey, and Tom Mitchell. "Estimating accuracy from unlabeled data: A bayesian approach". In: *Proc. of the International Conference on Machine Learning*. 2016, pp. 1416–1425.

[185] Joaquin Quiñonero-Candela and Carl Edward Rasmussen. “A unifying view of sparse approximate Gaussian process regression”. In: *The Journal of Machine Learning Research* 6 (2005), pp. 1939–1959.

[186] Md Ashfaqur Rahman and Hamed Mohsenian-Rad. “False data injection attacks against nonlinear state estimation in smart power grids”. In: *IEEE Power and Energy Society General Meeting*. 2013, pp. 1–5.

[187] Carl Edward Rasmussen and Christopher KI Williams. *Gaussian processes for machine learning*. Vol. 1. MIT press Cambridge, 2006.

[188] Lillian J. Ratliff, Samuel A. Burden, and S. Shankar Sastry. “Characterization and Computation of Local Nash Equilibria in Continuous Games”. In: *Proc. of the IEEE Annual Allerton Conference on Communication, Control, and Computing*. 2013, pp. 917–924.

[189] Lillian J. Ratliff, Samuel A. Burden, and S. Shankar Sastry. “Generictiy and Structural Stability of Non–Degenerate Differential Nash Equilibria”. In: *Proc. of the IEEE American Controls Conference*. 2014, pp. 3990–3995.

[190] Lillian J Ratliff, Ming Jin, Ioannis C Konstantakopoulos, Costas Spanos, and S Shankar Sastry. “Social game for building energy efficiency: Incentive design”. In: *Proc. of the IEEE Annual Allerton Conference on Communication, Control, and Computing*. 2014, pp. 1011–1018.

[191] Nathan D Ratliff, J Andrew Bagnell, and Martin A Zinkevich. “Maximum margin planning”. In: *Proc. of the International Conference on Machine Learning*. 2006, pp. 729–736.

[192] Nathan D Ratliff, David Silver, and J Andrew Bagnell. “Learning to search: Functional gradient techniques for imitation learning”. In: *Autonomous Robots* 27.1 (2009), pp. 25–53.

[193] Alexander J Ratner, Christopher M De Sa, Sen Wu, Daniel Selsam, and Christopher Ré. “Data programming: Creating large training sets, quickly”. In: *Advances in Neural Information Processing Systems*. 2016, pp. 3567–3575.

[194] Byron Reeves and J Leighton Read. *Total engagement: How games and virtual worlds are changing the way people work and businesses compete*. Harvard Business Press, 2009.

[195] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton University Press, 2015.

[196] Brendan van Rooyen and Robert C Williamson. “Learning in the Presence of Corruption”. In: *arXiv preprint arXiv:1504.00091* (2015).

[197] J. B. Rosen. “Existence and Uniqueness of Equilibrium Points for Concave N-Person Games”. In: *Econometrica* 33.3 (1965), p. 520.

[198] Henrik Sandberg, André Teixeira, and Karl H Johansson. "On security indices for state estimators in power networks". In: *First Workshop on Secure Control Systems, Stockholm*. 2010.

[199] Lawrence K Saul and Sam T Roweis. "Think globally, fit locally: unsupervised learning of low dimensional manifolds". In: *Journal of machine learning research* 4.Jun (2003), pp. 119–155.

[200] M Schmidt. "minFunc: unconstrained differentiable multivariate optimization in Matlab". In: (2012). URL: https://www.cs.ubc.ca/~schmidtm/Software/minFunc.html.

[201] James Scott et al. "PreHeat: controlling home heating using occupancy prediction". In: *Proc. of the International Conference on Ubiquitous Computing*. ACM. 2011, pp. 281–290.

[202] Daniel L Silver, Qiang Yang, and Lianghao Li. "Lifelong Machine Learning Systems: Beyond Learning Algorithms." In: *AAAI Spring Symposium: Lifelong Machine Learning*. Vol. 13. 2013, p. 05.

[203] Ankur Sinha, Pekka Malo, and Kalyanmoy Deb. "Evolutionary bilevel optimization: An introduction and recent advances". In: *Recent Advances in Evolutionary Multi-objective Optimization*. 2017, pp. 71–103.

[204] Edward Snelson and Zoubin Ghahramani. "Sparse Gaussian processes using pseudo-inputs". In: *Advances in neural information processing systems*. 2005, pp. 1257–1264.

[205] Saleh Soltan, Mihalis Yannakakis, and Gil Zussman. "Joint cyber and physical attacks on power grids: Graph theoretical approaches for information recovery". In: *ACM SIGMETRICS Performance Evaluation Review*. Vol. 43. 1. ACM. 2015, pp. 361–374.

[206] Kin Cheong Sou, Henrik Sandberg, and Karl Henrik Johansson. "On the exact solution to a smart grid cyber-security analysis problem". In: *IEEE Transactions on Smart Grid* 4.2 (2013), pp. 856–865.

[207] Siddharth Sridhar, Adam Hahn, and Manimaran Govindarasu. "Cyber–physical system security for the electric power grid". In: *Proceedings of the IEEE* 100.1 (2012), pp. 210–224.

[208] Russell Stewart and Stefano Ermon. "Label-Free Supervision of Neural Networks with Physics and Domain Knowledge." In: *Proc. of the AAAI Conference on Artificial Intelligence*. 2017, pp. 2576–2582.

[209] Ion Stoica et al. *A Berkeley View of Systems Challenges for AI*. Tech. rep. EECS Department, University of California, Berkeley, 2017.

[210] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. "On the importance of initialization and momentum in deep learning". In: *Proc. of the International Conference on Machine Learning*. 2013, pp. 1139–1147.

[211] Umar Syed and Robert E Schapire. "A game-theoretic approach to apprenticeship learning". In: *Advances in neural information processing systems*. Vol. 20. 20. 2007, pp. 1–8.

[212] Genichi Taguchi, Elsayed A Elsayed, and Thomas C Hsiang. *Quality engineering in production systems*. McGraw-Hill College, 1989.

[213] Stephen P Tarzia, Peter A Dinda, Robert P Dick, and Gokhan Memik. "Indoor localization without infrastructure using the acoustic background spectrum". In: *Proc. of the International Conference on Mobile systems, applications, and services*. 2011, pp. 155–168.

[214] Andre Teixeira, Kin Cheong Sou, Henrik Sandberg, and Karl Henrik Johansson. "Secure control systems: A quantitative risk management approach". In: *IEEE Control Systems* 35.1 (2015), pp. 24–45.

[215] Lakshmi V Thanayankizil, Sunil Kumar Ghai, Dipanjan Chakraborty, and Deva P Seetharam. "Softgreen: Towards energy management of green office buildings with soft sensors". In: *Proc. of the IEEE International Conference on Communication Systems and Networks*. 2012, pp. 1–6.

[216] The Edison Foundation Institute for Electric Innovation. "Electric Company Smart Meter Deployments: Foundation for A Smart Grid". In: *IEI Report* (2016).

[217] Sebastian Thrun. "Lifelong learning algorithms". In: *Learning to learn* 8 (1998), pp. 181–209.

[218] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), pp. 267–288.

[219] Michalis K Titsias and Neil D Lawrence. "Bayesian Gaussian process latent variable model". In: *Proc. of the International Conference on Artificial Intelligence and Statistics*. 2010, pp. 844–851.

[220] Sebastian Uziel et al. "Networked embedded acoustic processing system for smart building applications". In: *Proc. of the IEEE Conference on Design and Architectures for Signal and Image Processing*. 2013, pp. 349–350.

[221] Leslie G Valiant. "A theory of the learnable". In: *Communications of the ACM* 27.11 (1984), pp. 1134–1142.

[222] Hal R Varian. "Revealed preference". In: *Samuelsonian Economics and the Twenty-first Century* (2006), pp. 99–115.

[223] Rosemarie Velik and Pascal Nicolay. "Grid-price-dependent energy management in microgrids using a modified simulated annealing triple-optimizer". In: *Applied Energy* 130 (2014), pp. 384–395.

[224] Jane X Wang et al. "Learning to reinforcement learn". In: *arXiv preprint arXiv:1611.05763* (2016).

[225] Jingxuan Wang, Lucas CK Hui, SM Yiu, Eric Ke Wang, and Junbin Fang. "A survey on cyber attacks against nonlinear state estimation in power systems of ubiquitous cities". In: *Pervasive and Mobile Computing* (2017).

[226] Wenye Wang and Zhuo Lu. "Cyber security in the Smart Grid: Survey and challenges". In: *Computer Networks* 57.5 (2013), pp. 1344–1371.

[227] Kevin Weekly, Nikolaos Bekiaris-Liberis, Ming Jin, and Alexandre M Bayen. "Modeling and estimation of the humans' effect on the CO2 dynamics inside a conference room". In: *IEEE Transactions of Control Systems Technology* 23.5 (2015), pp. 1770–1781.

[228] Kevin Weekly et al. "Building-in-Briefcase (BiB)". In: *arXiv preprint arXiv:1409.1660* (2014).

[229] Dong Wei, Yan Lu, Mohsen Jafari, Paul M Skare, and Kenneth Rohde. "Protecting smart grid automation systems against cyberattacks". In: *IEEE Transactions on Smart Grid* 2.4 (2011), pp. 782–795.

[230] Rafał Weron. "Electricity price forecasting: A review of the state-of-the-art with a look into the future". In: *International Journal of Forecasting* 30.4 (2014), pp. 1030–1081.

[231] Stephen Wilcox and William Marion. *Users manual for TMY3 data sets.* 2008.

[232] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of semidefinite programming: theory, algorithms, and applications.* Vol. 27. Springer Science & Business Media, 2012.

[233] Jianzhnog Wu et al. "Integrated Energy Systems". In: *Applied Energy* 167 (2016), pp. 155–157.

[234] Markus Wulfmeier, Peter Ondruska, and Ingmar Posner. "Maximum entropy deep inverse reinforcement learning". In: *arXiv preprint arXiv:1507.04888* (2015).

[235] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang. "Learning from massive noisy labeled data for image classification". In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition.* 2015, pp. 2691–2699.

[236] Le Xie, Yilin Mo, and Bruno Sinopoli. "False data injection attacks in electricity markets". In: *IEEE International Conference on Smart Grid Communications.* 2010, pp. 226–231.

[237] Lihua Xie, Emilia Fridman, and Uri Shaked. "Robust control of distributed delay systems with application to combustion control". In: *IEEE Transactions on Automatic Control* 46.12 (2001), pp. 1930–1935.

[238] Longqi Yang, Kevin Ting, and Mani B Srivastava. "Inferring occupancy from opportunistically available sensor data". In: *IEEE International Conference on Pervasive Computing and Communications.* IEEE. 2014, pp. 60–68.

[239] Shaghayegh Yousefi, Mohsen Parsa Moghaddam, and Vahid Johari Majd. “Optimal real time pricing in an agent-based retail market using a comprehensive demand response model”. In: *Energy* 36.9 (2011), pp. 5716–5727.

[240] Jiancheng Yu et al. “Review of microgrid development in the United States and China and lessons learned for China”. In: *Renewable Energy Integration with Mini/Microgrids*. 2017. URL: http://www.jinming.tech/papers/REM2017.pdf.

[241] Yanling Yuan, Zuyi Li, and Kui Ren. “Modeling load redistribution attacks in power systems”. In: *IEEE Transactions on Smart Grid* 2.2 (2011), pp. 382–390.

[242] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. “A survey of affect recognition methods: Audio, visual, and spontaneous expressions”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.1 (2009), pp. 39–58.

[243] Y. Zhang, R. Madani, and J. Lavaei. “Conic Relaxations for Power System State Estimation with Line Measurements”. In: *IEEE Transactions on Control of Network Systems* PP.99 (2017), pp. 1–1.

[244] Yuxun Zhou and Costas J Spanos. “Causal meets submodular: Subset selection with directed information”. In: *Advances in Neural Information Processing Systems*. 2016, pp. 2649–2657.

[245] Yuxun Zhou et al. “Abnormal event detection with high resolution micro-PMU data”. In: *Proc. of the IEEE Power Systems Computation Conference*. 2016, pp. 1–7.

[246] Brian D Ziebart, J Andrew Bagnell, and Anind K Dey. “Modeling interaction via the principle of maximum causal entropy”. In: *Proc. of the International Conference on Machine Learning*. 2010, pp. 1247–1254.

[247] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. “Maximum Entropy Inverse Reinforcement Learning.” In: *Proc. of the AAAI Conference on Artificial Intelligence*. 2008, pp. 1433–1438.

[248] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. “MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education”. In: *IEEE Transactions on Power Systems* 26.1 (2011), pp. 12–19.

[249] Han Zou, Ming Jin, Hao Jiang, Lihua Xie, and Costas J Spanos. “WinIPS: WiFi-based Non-intrusive Indoor Positioning System with Online Radio Map Construction and Adaptation”. In: *IEEE Transactions on Wireless Communications* 16.12 (2017), pp. 8118–8130.

[250] Han Zou et al. “WinLight: A WiFi-based occupancy-driven lighting control system for smart building”. In: *Energy and Buildings* 158 (2018), pp. 924–938.