Data-Driven and Robust Predictive Control and Optimization

With Applications to Building Energy Management

Yun Li

Data-Driven and Robust Predictive Control and Optimization

With Applications to Building Energy Management

Data-Driven and Robust Predictive Control and Optimization

With Applications to Building Energy Management

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus,
Prof. dr. ir. H. Bijl,
chair of the Board of Doctorates
to be defended publicly on
Wednesday, May 6, 2026 at 12:30 PM.

by

Yun Li

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. T. Keviczky,	Delft University of Technology, promotor
Dr. N. Yorke-Smith,	Delft University of Technology, promotor
Independent members:	
Prof. dr. ir. B.H.K. De Schutter,	Delft University of Technology
Prof. dr. L.C.M. Itard,	Delft University of Technology
Prof. dr. R. Scattolini,	Politecnico di Milano, Italy
Prof. dr. S. Baldi,	Southeast University, China
Dr. C.N. Jones,	École Polytechnique Fédérale de Lausanne, Switzerland

Keywords: Robust optimization, model predictive control, system level synthesis, adjustable uncertainty set, data-driven uncertainty set, building climate control, demand-side management, noise control, heat pump

Cover: Generated using ChatGPT by Y. Li

ISBN 978-94-6518-276-6

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

CONTENTS

SUMMARY

Buildings, as major global energy consumers, can help mitigate the impact of growing renewable energy in smart grids through demand-side management (DSM). Smart energy management of buildings requires advanced control schemes that can cope with economic objectives, environmental uncertainties, occupant comfort, physical constraints, and external communication signals. Robust optimization (RO) and model predictive control (MPC) provide systematic and effective frameworks for this purpose. Beyond building energy management, RO and MPC methods are also fundamental to a broad range of engineering applications, such as chemical process planning, transportation, robotics, etc. This thesis focuses on RO and MPC design for linear systems as well as their applications in building energy management. The research is organized into three topics:

- MPC designs for building energy management to enable energy-flexible DSM and improve environmental sustainability (Chapters 2–4).
- data-driven RO designs and algorithmic solutions for linear models to reduce conservatism and improve computational efficiency (Chapters 5 & 6).
- a distributionally robust MPC design for constrained linear systems to robustify control performance against additive model uncertainties and disturbances (Chapter 7).

Within the first topic, three key research works are carried out:

1. An important task in DSM is the so-called "reserve and provision" problem, which can be formulated as a robust optimal control (ROC) problem with adjustable uncertainty sets. Existing studies primarily address continuous uncertainties and thus fail to capture on/off-type device behaviour. We investigate the ROC with binary adjustable uncertainties for constrained linear systems. A new metric is introduced to quantify the scope of uncertainties, together with an ROC framework that effectively models the corresponding "reserve and provision" problem. To address the computational challenges and balance control optimality, affine control policies are adopted. By exploiting strong duality and big-M techniques, the ROC problem is reformulated as a mixed-integer linear program (MILP) that can be efficiently solved. Numerical case studies on smart building DSM demonstrate the effectiveness of the proposed approach.
2. Given its substantial share in household electricity consumption and widespread deployment in modern buildings, the heat pump and thermal energy storage (HPTES) system offers significant potential for providing demand response services. The complex and nonlinear dynamics of the HPTES system preclude the

application of existing "reserve and provision" methods. We develop a practical two-step DSM strategy compatible with the EU future-proof flexibility interface to exploit the energy flexibility of HPTES systems. In the first step, a mixed-integer economic MPC formulation is used to quantitatively assess the system's flexibility potential. In the second step, this flexibility is utilized to respond to demand response requests. Both numerical simulations and experimental results confirm the feasibility and effectiveness of the proposed strategy.

3. The noise generated by heat pumps (HPs) remains an acoustic market barrier to their widespread adoption. To mitigate this issue, an MPC-based solution is proposed to reduce HP acoustic nuisance from a control-oriented perspective. A piecewise linear approximation is employed to model HP noise patterns, enabling adaptability to various operating conditions. Two cost-function formulations for noise mitigation are investigated. Numerical results using a high-fidelity building simulator demonstrate that the proposed MPC design can effectively reduce HP noise without significantly increasing energy consumption.

Under the second topic, two main research studies are conducted:

1. The conservatism and computational tractability of an RO model are critically influenced by how the uncertainty set is represented. With the availability of abundant data and advances in machine learning (ML) tools, we investigate the construction of compact and computationally tractable uncertainty sets using ML techniques. By combining density-based spatial clustering of applications with noise (DBSCAN), Gaussian mixture models (GMM), and principal component analysis (PCA), we propose an algorithm that generates data-driven uncertainty sets that are both compact and adaptive to irregular data distributions. Detailed parameter tuning guidelines and a Python-based toolkit are also provided. Furthermore, a computationally efficient solution for linear two-stage RO models with the proposed uncertainty set is derived. The effectiveness of the proposed approach is extensively validated on both synthetic and real-world datasets, including a building climate control case study.

2. Representing uncertainty sets as unions of multiple basic subsets has been shown to improve compactness and flexibility in data-driven RO. Building on this structure, two separate research problems are addressed: (i) the computational challenge for solving the corresponding RO problem when applying the uncertainty set structure in predictive control problems, and (ii) the development of a new RO formulation that mitigates the conservatism of RO models by exploiting the multi-subset structure. For the first problem, a monolithic mixed-integer representation is proposed to uniformly capture the union of multiple subsets over the prediction horizon, enabling the computation of the worst-case scenario via a single MILP problem. For the second, a new formulation integrating the RO and distributionally robust optimization (DRO) is proposed to mitigate the conservatism of conventional RO formulation while avoiding the high computational burden of the standard DRO models. For both research problems, algorithmic solutions

based on column-and-constraint generation (CCG) are proposed to solve the corresponding RO problems. Simulation results from three case studies validate the efficiency and effectiveness of the proposed methods.

Under the last topic, a distributionally robust model predictive control design for constrained uncertain linear systems is investigated. To improve the expressiveness of control policies and mitigate the associated computational burden in predictive control problems, a novel system level synthesis (SLS) parameterization considering affine output feedback control policies is proposed. The SLS formulation is extended to a distributionally robust setting to enhance the control performance against model mismatch and additive disturbances by optimizing the control performance against the worst-case disturbance distribution within a Wasserstein-metric-based ambiguity set. Theoretical analysis quantifies the effect of model errors on the distributional shift between predicted and actual closed-loop responses. A computationally efficient relaxation is derived to solve the proposed distributionally robust SLS design. The effectiveness of the proposed approach is numerically tested.

In summary, this thesis addresses the issues related to data-driven robust optimization and predictive control as well as their applications in energy management of smart buildings. The proposed frameworks enhance decision-making resilience and reduce conservatism for optimization and predictive control of linear systems, and enable energy-flexible and environmentally-sustainable operation of smart buildings.

SAMENVATTING

Gebouwen behoren tot de grootste energieverbruikers wereldwijd en kunnen via vraagsturing (demand-side management, DSM) bijdragen aan het opvangen van de groei van hernieuwbare energie in slimme elektriciteitsnetten. Slim energiebeheer in gebouwen vereist geavanceerde regelstrategieën die economische doelstellingen, omgevingsonzekerheden, bewonerscomfort, fysieke beperkingen en externe communicatiesignalen kunnen verwerken. Robuuste optimalisatie (RO) en modelgebaseerde voorspellende regeltechniek (model predictive control, MPC) bieden systematische en effectieve raamwerken voor dit doel. Naast gebouwenergiemanagement vormen RO en MPC methoden ook de basis voor uiteenlopende ingenieurstoepassingen, zoals chemische procesplanning, transport en robotica. Dit proefschrift richt zich op RO en MPC ontwerp voor lineaire systemen en hun toepassingen in gebouwenergiemanagement. Het onderzoek is georganiseerd in drie thema's:

- MPC ontwerpen voor gebouwenergiemanagement ter ondersteuning van energieflexibele DSM en verbetering van milieuduurzaamheid (Hoofdstukken 2–4).
- Data-gedreven RO ontwerpen en algoritmische oplossingen voor lineaire modellen met als doel het verminderen van conservatisme en het verbeteren van rekenkundige efficiëntie (Hoofdstukken 5 & 6).
- Een distributioneel robuust MPC ontwerp voor begrensde lineaire systemen om de regelprestaties te versterken bij additieve modelonzekerheden en verstoringen (Hoofdstuk 7).

Binnen het eerste thema worden drie kernstudies uitgevoerd:

1. Een belangrijke DSM taak is het zogenaamde "reserve en voorziening" probleem, dat kan worden geformuleerd als een robuust optimaal regelprobleem (ROC) met aanpasbare onzekerheidsverzamelingen. Bestaande studies richten zich vooral op continue onzekerheden en houden daardoor geen rekening met aan/uit-gedrag van apparaten. Wij onderzoeken daarom het ROC met binaire aanpasbare onzekerheden voor begrensde lineaire systemen. Hiervoor wordt een nieuwe maatstaf geïntroduceerd om de omvang van onzekerheden te kwantificeren en een ROC-raamwerk ontwikkeld dat het "reserve en voorziening" probleem accuraat modelleert. Om de rekenlast te beperken en de regeloptimaliteit te waarborgen, worden affine regelstrategieën toegepast. Door gebruik te maken van sterke dualiteit en big-M-technieken wordt het ROC geherformuleerd als een mixed-integer lineair programma (MILP). Numerieke casestudies in slimme gebouwen tonen de effectiviteit van de voorgestelde aanpak.

2. Warmtepompen in combinatie met thermische energieopslag (HPTES) vertegenwoordigen een aanzienlijk deel van het elektriciteitsverbruik in huishoudens en bieden grote kansen voor vraagrespons. Door hun complexe en niet-lineaire dynamiek zijn bestaande “reserve and provision” methoden echter niet toepasbaar. Daarom ontwikkelen wij een praktische tweestaps DSM strategie, compatibel met de EU future-proof flexibility interface, om de energie-flexibiliteit van HPTES systemen te benutten. In de eerste stap wordt een mixed-integer economische MPC formulering gebruikt om de flexibiliteit te kwantificeren. In de tweede stap wordt deze flexibiliteit veilig ingezet om aan vraagresponsverzoeken te voldoen. Zowel numerieke simulaties als experimentele resultaten bevestigen de haalbaarheid en effectiviteit van de voorgestelde methode.

3. Geluidsoverlast door warmtepompen vormt een marktbarrière voor brede adoptie. Om dit te beperken wordt een MPC gebaseerde benadering ontwikkeld om de akoestische hinder vanuit regeltechnisch perspectief te reduceren. Met behulp van een stuklijnige benadering worden geluidspatronen van warmtepompen gemodelleerd, zodat deze zich kunnen aanpassen aan diverse bedrijfsconditie. Twee kostenfuncties voor geluidsmitigatie worden onderzocht. Numerieke resultaten met een gedetailleerde gebouwsimulator tonen aan dat de voorgestelde MPC-strategie het geluid van warmtepompen effectief kan verminderen zonder een significante toename van het energieverbruik.

Binnen het tweede thema worden twee hoofdonderzoeken uitgevoerd:

1. De conservativiteit en rekenbaarheid van een RO model worden sterk beïnvloed door de representatie van de onzekerheidsverzameling. Met de beschikbaarheid van grote hoeveelheden data en de vooruitgang in machine-learning technieken onderzoeken wij hoe compacte en rekenkundig efficiënte onzekerheidssets kunnen worden geconstrueerd. Door DBSCAN clustering, Gaussiaanse mengmodellen (GMM) en principal component analysis (PCA) te combineren, ontwikkelen wij een algoritme dat data-gedreven onzekerheidssets genereert die compact zijn en zich kunnen aanpassen aan onregelmatige dataverdelingen. Richtlijnen voor parameterafstemming en een Python-toolkit worden meegeleverd. Verder wordt een computationeel efficiënte oplossingsmethode afgeleid voor lineaire tweefasen robuuste optimalisatiemodellen met de voorgestelde onzekerheidsverzameling. De effectiviteit van de methode wordt gevalideerd op zowel synthetische als real-world datasets, waaronder een casestudy op gebouwklimaatregeling.

2. Het representeren van onzekerheidssets als een unie van meerdere deelverzamelingen verhoogt de compactheid en flexibiliteit in data-gedreven RO. Op basis van deze structuur worden twee onderzoeksvragen behandeld: (i) het oplossen van RO problemen in voorspellende regelcontext wanneer deze multi-subset structuur wordt toegepast, en (ii) de ontwikkeling van een nieuwe RO formulering die het conservatisme vermindert door de multi-subset structuur te benutten. Voor het eerste probleem wordt een monolithische mixed-integer representatie voorgesteld om de unie van subsets uniform te modelleren over de voorspellingshorizon, zodat het worst-case scenario met één MILP kan worden bepaald. Voor het

tweede probleem wordt een nieuwe formulering ontwikkeld die RO en distributioneel robuuste optimalisatie (DRO) combineert om conservatisme te verminderen zonder de hoge rekenlast van standaard DRO modellen. Voor beide onderzoeksvragen worden oplossingsmethoden gebaseerd op column-and-constraint generation (CCG) gepresenteerd. Drie casestudies tonen de efficiëntie en effectiviteit van de voorgestelde methoden.

In het derde thema wordt een distributioneel robuust MPC ontwerp voor begrensde lineaire systemen met onzekerheden onderzocht. Om de expressiviteit van regelstrategieën te vergroten en de rekenlast te beperken, wordt een nieuwe system level synthesis (SLS) parameterisatie ontwikkeld op basis van affine output-feedback regelaars. Deze SLS formulering wordt uitgebreid naar een distributioneel robuuste setting, waarbij de regelprestaties worden geoptimaliseerd tegen de slechtst mogelijke verstoringsverdeling binnen een op de Wasserstein-maat gebaseerde ambiguïteitsset. Een theoretische analyse kwantificeert het effect van modelafwijkingen op de distributieverschuiving tussen voorspelde en werkelijke gesloten-lusresponsen. Daarnaast wordt een rekenkundig efficiënte relaxatie afgeleid. De effectiviteit van de voorgestelde aanpak wordt numeriek aangetoond.

Samenvattend, behandelt dit proefschrift vraagstukken rond data-gedreven robuuste optimalisatie en voorspellende regeltechniek, evenals hun toepassingen in energiebeheer van slimme gebouwen. De voorgestelde raamwerken vergroten de veerkracht van besluitvorming, verminderen conservatisme in optimalisatie en voorspellende regeling van lineaire systemen, en faciliteren energie-flexibel en duurzaam gebouwbeheer.

NOMENCLATURE

ARX	auto regressive with exogeneous inputs
ASHP	air-source heat pump
AUS	adjustable uncertainty sets
BMS	building management system
CCG	column-and-constraint generation
CDF	cumulative density function
CE	certainty equivalence
COP	coefficient of performance
CPNP	chemical process network planning
CVaR	conditional value at risk
DBSCAN	density-based spatial clustering of applications with noise
DDU	decision-dependent uncertainties
DNN	deep neural network
DR	demand response
DR-SLS	distributionally robust system level synthesis
DRMPC	distributionally robust model predictive control
DRO	distributionally robust optimization
DSM	demand side management
EFI	energy flexibility interface
EU	European union
EVs	electric vehicles
GHG	greenhouse gas
GMM	Gaussian mixture model
HP	heat pump

HPTES	heat pump and thermal energy storage
HVAC	heating, ventilation, and air conditioning
I.I.D.	independent and identically distributed
KDE	kernel density estimation
LB	lower bound
LP	linear program
MAE	mean absolute error
MILP	mixed-integer linear programming
ML	machine learning
MPC	model predictive control
N-SLS	nominal system level synthesis
NPV	net present value
PCA	principal component analysis
RC	resistor–capacitor
RES	renewable energy sources
RMPC	robust model predictive control
RO	robust optimization
ROC	robust optimal control
SAA	sample average approximation
SD	strong duality
SLS	system level synthesis
SMPC	stochastic model predictive control
SO	stochastic optimization
SOCP	second-order cone program
SVC	support vector clustering
TES	thermal energy storage
UB	upper bound
V2H	vehicle-to-home

1

INTRODUCTION

This chapter presents a brief overview of the topics covered in this thesis: model predictive control design for building energy management, data-driven robust optimization, and distributionally robust model predictive control. For each topic, the open problems and associated challenges that motivate the research works of this thesis are discussed. The novelty and the main contributions of this thesis are summarized. Finally, the structure of this thesis is outlined.

1

1.1. BACKGROUND

MODERN systems – ranging from industrial processes and energy networks to building infrastructures – are becoming increasingly complex, interconnected, and data-intensive. Efficient and reliable operation of such systems requires advanced decision-making and control strategies capable of handling uncertainty, balancing performance, safety, and sustainability objectives. Uncertainties may arise from imperfect models, external disturbances, fluctuating resource provision and demand, or changing environmental conditions, and they often lead to performance degradation or constraint violations when not properly accounted for.

Among these systems, buildings represent one of the largest global consumers of energy and a major source of carbon emissions [1, 2]. On the one hand, as urbanization accelerates and decarbonization goals intensify, improving the operational efficiency of building energy systems has become both an economic necessity and an environmental imperative. On the other hand, the increased attention on building climate comfort motivates more intelligent schemes for indoor climate management. Advanced control and optimization strategies have been regarded as a crucial solution in achieving these objectives by enabling optimal and robust energy management that balances occupant comfort, energy cost, and sustainability objectives [3].

Furthermore, beyond the application of building energy management, designing optimization and control algorithms against uncertainties is itself of scientific importance, and is prevalently exploited in solving many practical engineering problems, including chemical process optimization, power system operation, supply chain management, and transportation planning [4–6]. Therefore, investigating the general principles and methodologies of optimization and control against uncertainties is not only beneficial in building energy management but also contributes to a wide range of engineering domains.

To ensure reliable decision-making against uncertainties, robust optimization (RO) has gained increased attention and has emerged as an essential tool. Notably, robust approaches for linear systems have gathered considerable attention and have been applied to solving real-world problems due to their theoretical tractability and computational efficiency. Rather than optimizing for a nominal or expected scenario, robust methods explicitly consider uncertainties within predefined sets and following unknown distributions, and seek decisions that guarantee feasibility and acceptable performance for all possible realizations [7, 8]. This worst-case formulation provides formal safety and reliability guarantees, which are crucial in safety-critical and performance-sensitive applications. For enhancing the resilience of decisions in a dynamic environment, the approaches developed in RO have been incorporated into predictive control design to develop robust control schemes [9–11].

However, it should be pointed out that robust optimization methods can be overly conservative when the uncertainty set is poorly characterized or when only limited information about the support of uncertainties is available, leading to suboptimal performance. The increasing availability of high-resolution sensing, monitoring, and operational data has enabled the rise of data-driven robust optimization and control [12, 13]. These approaches leverage data to learn system behaviour, identify patterns, or forecast future conditions, thereby reducing dependence on explicit physical models or expert

knowledge. Data-driven techniques have achieved notable success in various domains, including building energy scheduling, power system operation, etc [14–16].

While the above-mentioned research topics – smart energy management of buildings, robust optimization and predictive control and their data-driven variants – have been investigated in many recent publications, there are still some important research gaps to be bridged, and some meaningful and practical problems have been neglected in the existing literature. In the following subsections, the above topics will be further elaborated to present a detailed introduction for each topic.

1.1.1. MODEL PREDICTIVE CONTROL DESIGN FOR BUILDING ENERGY MANAGEMENT

Considering the fact that the energy sector is responsible for over 75% of the European Union's (EU) greenhouse gas (GHG) emissions, the EU has set ambitious energy and climate targets, including reducing net GHG emissions by at least 55% and increasing the share of renewable energy sources (RES) to at least 40% of the total energy mix by 2030 [17]. While the increasing integration of RES contributes to decarbonization and reduces dependence on fossil fuels, it also introduces new challenges to power system operation. The intermittent and unpredictable nature of RES generation causes fluctuations in power supply, potentially leading to supply–demand imbalances and a higher demand for reserve or backup capacity [18, 19]. Conventional generation-side management approaches, however, often struggle to cope with such rapid and uncertain variations, as traditional generation units typically have limited ramping capabilities.

To mitigate these issues and enhance grid stability, demand-side management (DSM) has been proposed as a complementary solution. DSM involves adjusting or shifting energy consumption patterns in response to grid conditions, thereby improving the power system flexibility and reliability. Unlike generation-side measures, DSM does not require costly infrastructure upgrades, making it an economically attractive approach for balancing supply and demand. A comprehensive survey of DSM strategies can be found in [20, 21].

Among various DSM resources, buildings play a crucial role, accounting for approximately 40% of global energy consumption, with heating, ventilation, and air-conditioning (HVAC) systems responsible for about half of this amount [22]. Importantly, HVAC loads are often shiftable or curtailable, providing significant potential for flexible energy usage. For example, buildings can preheat or precool thermal zones within acceptable comfort ranges before the peak hours to avoid energy usage during peak demand periods. Similarly, domestic hot water can also be preheated when the electricity price is low and stored in thermal tanks for later use. Given their substantial energy consumption and the growing deployment of automated building energy management systems, buildings represent a highly promising resource for DSM. Recent international initiatives, such as IEA EBC Annex 67 & 82, demonstrate the increasing interest in leveraging building energy flexibility to mitigate the supply variability introduced by high RES penetration. Detailed surveys on building-oriented DSM approaches are available in [23, 24].

Moreover, effective control of HVAC systems in buildings not only supports DSM and energy efficiency but also enhances occupant comfort and well-being. As people spend

over 80% of their time indoors, maintaining comfortable thermal conditions is essential for health, satisfaction, and productivity [25]. Hence, advanced building climate control and energy management serve both environmental and socio-economic objectives.

To fully exploit the energy flexibility of buildings, traditional rule-based control strategies are inadequate. These conventional methods are typically designed for static operating conditions and fail to account for dynamic factors such as weather forecasts, occupancy patterns, grid signals, and varying energy prices. In contrast, advanced control techniques, particularly Model Predictive Control (MPC), offer a systematic framework for optimizing building operation under uncertainty and constraints, balancing comfort and energy efficiency, and incorporating external signals for grid interaction [2, 22, 26].

While the energy-saving potential of MPC in building climate control has been extensively validated in the literature, several open challenges remain. In particular, integrating MPC-based building control into DSM frameworks raises additional issues related to uncertainty management, flexibility quantification, communication with grid operators, and the mitigation of side effects such as acoustic pollution of HVAC devices. Addressing these challenges is essential for enabling MPC to serve as a practical and scalable solution for future smart energy systems.

1.1.2. DATA-DRIVEN ROBUST OPTIMIZATION

Mathematical programming has been successfully applied across a wide range of domains, including chemical processes, finance, transportation systems, and energy systems [27–31]. In practical applications, the parameters involved in optimization models are often subject to various sources of uncertainty, such as process disturbances, measurement noise, and parameter estimation errors. To ensure performance guarantees and avoid performance degradation when applying the optimization models, uncertainties need to be properly taken into consideration. To address these uncertainties, two principal paradigms have been developed: stochastic optimization (SO) and RO.

In the stochastic setting, it is typically assumed that the probability distribution of the uncertainties is known, and the goal is to minimize the expected value of a cost function with respect to these uncertainties [32]. While this approach can yield less conservative decisions than the so-called worst-case robust approaches, analytical solutions exist only for limited classes of distributions and problem structures. Moreover, in many practical scenarios, the true probability distribution of uncertainties is unknown or difficult to estimate accurately. To overcome this limitation, scenario-based approximations of the stochastic formulation are often employed, where the uncertainty distribution is approximated by a finite set of uncertainty samples [33]. Although this relaxation broadens applicability, ensuring a high-confidence performance guarantee requires a sufficiently large number of samples, and accordingly a high computational burden, particularly in multi-stage or large-scale optimization problems.

The robust optimization paradigm represents an alternative and complementary approach. In the RO setting, uncertain parameters are assumed to lie within a prescribed uncertainty set, which captures all possible realizations [7, 34]. The objective function and constraints are then optimized and enforced against the worst-case realization within this set. Unlike the stochastic approach, robust optimization does not require

explicit probabilistic information about the uncertainties – only their support or range. Moreover, RO formulations are generally computationally efficient, and their computational cost does not depend on the number of uncertainty samples.

However, a key challenge in RO lies in the construction of the uncertainty set, as it directly determines the trade-off between robustness and conservatism [12]. Conventional choices, such as box, ellipsoidal, and budget uncertainty sets, are popular due to their analytical and computational convenience. Nonetheless, the construction and tuning of these sets for practical problems often rely on domain-specific knowledge, making their selection problem-dependent and sometimes excessively time-consuming. Besides, these conventional choices might fail to generate compact uncertainty sets for complex uncertainty distributions.

With the rapid advancement of sensing technologies, computational power, and wireless connectivity, vast amounts of data are now being generated across diverse sectors such as industrial processes, transportation networks, and the built environment. This abundance of data opens new opportunities for data-driven decision making, where optimization models can exploit data rather than rely solely on expert-defined parameters. Besides, a large variety of machine learning (ML) techniques are available to analyze data and extract useful information to facilitate decision-making effectively.

Recent progress in mathematical programming, coupled with advances in machine learning, has sparked growing interest in ML-based data-driven optimization to enhance model adaptability and to reduce the conservatism of the optimal decision. In particular, data-driven RO frameworks have emerged as a promising direction. In data-driven RO, the uncertainty sets are constructed directly from empirical data rather than predefined geometric assumptions. This allows the data itself to characterize the variability and dependence structure of uncertain parameters. By allowing the uncertainty set to be informed by data, data-driven RO methods can automatically capture latent patterns for constructing compact uncertainty sets to reduce unnecessary conservatism and lead to more informed decision-making in uncertain environments [12, 35, 36].

1.1.3. DISTRIBUTIONALLY ROBUST MODEL PREDICTIVE CONTROL

MPC has long been an active and influential research topic within the control community. Owing to its ability to handle multi-variable systems and explicit constraints, MPC has been successfully applied to a wide range of practical problems, including process control, automotive systems, power grids, and building energy management [37–40]. The core idea of MPC is to approximate an infinite-horizon optimal control problem by repeatedly solving in a receding horizon fashion, at each control step, a finite-horizon optimal control problem that predicts and optimizes the future behaviour of the system.

Designing an MPC scheme requires an accurate system model to predict the evolution of the system dynamics over the prediction horizon. In practice, however, the controlled system inevitably suffers from various sources of uncertainties, such as unmodeled dynamics, external disturbances, sensor noise, and parameter estimation errors. These uncertainties can significantly degrade the performance of MPC and even lead to constraint violations if not properly accounted for. To address this challenge, there are two common well-established design frameworks: robust MPC (RMPC) and stochastic MPC (SMPC), primarily differing in how they treat uncertainty [41, 42].

Analogously to Robust Optimization (RO), RMPC models uncertainties as elements within a predefined bounded uncertainty set and optimizes the worst-case performance over all possible realizations in this set. A notable advantage of RMPC is that it does not require explicit probabilistic information about uncertainties, making it suitable for systems where only deterministic bounds are available. However, because RMPC explicitly guards against the worst-case scenario when optimizing the objective function and ensuring constraint satisfaction, it often leads to overly conservative control actions. Moreover, it cannot effectively handle unbounded uncertainties, which might be the case in some practical applications.

In contrast, SMPC, similar in spirit to SO, incorporates the probabilistic characteristics of uncertainties and optimizes the expected system performance. By exploiting distributional information, SMPC can achieve less conservative results compared to RMPC. Nevertheless, the limitations of SO also carry over to SMPC: 1) accurate distributional knowledge of uncertainties is rarely available in practice; 2) analytical solutions exist only for specific systems and distributions; and 3) sample-based approaches require a large number of uncertainty realizations, leading to computational complexity that grows exponentially with the prediction horizon.

In real-world applications, the available information about uncertainties typically lies somewhere between these two extremes. Often, only partial or approximate statistical knowledge can be obtained, for instance, from limited historical data or expert judgment. Furthermore, even when the uncertainty distribution can be estimated accurately from historical data, distributional shifts may occur due to changes in the operating environment, invalidating previous models. These challenges have motivated the development of distributionally robust optimization (DRO), an emerging framework that aims to balance robustness and statistical fidelity [43, 44].

Within the DRO framework, the true distribution of uncertainties is assumed to belong to an ambiguity set – a family of plausible distributions characterized by available information (e.g., moments, support, or distance from empirical data). The goal is to optimize the worst-case expected performance over all distributions in this ambiguity set. Compared with RO, DRO can produce less conservative decisions if the ambiguity set is properly calibrated. Compared with SO, DRO offers robustness against distributional ambiguity. Owing to these advantages, DRO has attracted substantial attention in recent years, leading to a variety of distributionally robust MPC (DRMPC) formulations that enhance the robustness and adaptability of MPC in the presence of uncertain and data-dependent disturbances [45–51].

1.2. PROBLEM STATEMENT AND CHALLENGES

1.2.1. MODEL PREDICTIVE CONTROL DESIGN FOR BUILDING ENERGY MANAGEMENT

MPC has been widely recognized as a promising advanced control strategy for optimizing indoor climate operations, and numerous studies have demonstrated its effectiveness in balancing energy efficiency with thermal comfort [26, 40]. However, when integrated into DSM frameworks, several critical challenges remain unresolved. Most existing DSM strategies rely on price-based or incentive-based programs, where energy

consumers are expected to voluntarily adjust their consumption patterns in response to dynamic electricity prices or external incentives [52, 53]. While these schemes are easy to implement and require only unidirectional communication protocols, they fail to fully leverage the latent energy flexibility of buildings, that is, their capability to adapt consumption in a controlled and predictable manner.

To unlock this flexibility potential, more sophisticated DSM architectures with bi-directional communication between building management systems (BMS) and grid operators are required. This enables both parties to exchange information: grid operators or energy providers can send demand response (DR) requests, while buildings can report their available flexibility reserves. In this context, the EU S2 Standard has been introduced as a future-proof framework to standardize such communications, ensuring interoperability across diverse household devices, market contracts, and privacy requirements [54].

Within this bi-directional structure, the DSM task fits into the so-called "reserve and provision" problem: first, assessing and reserving flexibility capacity, and second, realizing it when a DR signal is issued. The first step, referred to as *flexibility assessment*, is of central importance – it quantifies how much and under what conditions a building can adjust its energy consumption [55, 56]. This practical problem aligns with a theoretical framework known as *robust optimization with decision-dependent uncertainties* (RO-DDU) in the operations research community [57], or *robust optimal control with adjustable uncertainty sets* (ROC-AUS) in the control community [58, 59].

Unlike conventional robust optimization and control problems, where uncertainties are exogenous and uncertainty sets are predefined and static, ROC-AUS considers adjustable uncertainty sets whose size and shape depend on decision variables. This feature naturally captures the flexibility interaction between grid requests and BMS capabilities: while the DR signal represents an external uncertainty for BMS, its admissible range is constrained by the flexibility capacity that the BMS commits to provide. Thus, the uncertainties are endogenous, and their admissible scopes are determined by decision makers themselves.

The introduction of decision-dependent uncertainties substantially increases computational and theoretical complexity. Directly extending existing RO methods often leads to nonconvex and computationally intractable formulations, even for linear systems with polyhedral uncertainty sets [58]. Moreover, current studies primarily focus on continuous uncertainties modeled by polyhedral or ellipsoidal sets. While the ROC-AUS with continuous uncertainties can cover a wide range of engineering problems, in many practical energy systems, uncertainties may be discrete or hybrid in nature. For example, on-off operational modes of devices such as heat pumps or chillers. In these cases, the existing approaches developed for the continuous uncertainties are no longer applicable. Consequently, it is of both theoretical importance and practical necessity to consider ROC-AUS problems with non-continuous uncertainties.

Additionally, due to its theoretical and computational tractability, almost all existing ROC-AUS frameworks assume linearity or convexity in system models and constraints, which limits their applicability to real-world energy systems characterized by nonlinear and nonconvex dynamics as well as operational constraints. Consequently, there is also a pressing need to develop computationally tractable, theoretically sound, and practi-

cally implementable DSM methodologies for assessing and exploiting the energy flexibility of buildings that can accommodate the practical nonlinear system behaviors. Such developments are essential for better exploiting building energy flexibility and enabling scalable, intelligent DSM in future smart energy systems.

Heat pumps (HPs) have been widely deployed in modern buildings for space heating and cooling due to their high energy efficiency and ability to integrate with low-carbon energy systems. However, as their penetration continues to grow, especially in densely populated residential and commercial areas, a growing concern has emerged: noise. The acoustic emissions generated during HP operation are increasingly viewed as a market barrier that may hinder the widespread adoption of HPs [60]. Existing noise-mitigation solutions – such as sound-absorbing enclosures, acoustic barriers, or anti-vibration mountings – can reduce noise but often require intrusive, costly, and space-consuming retrofits, making them impractical for many existing installations.

Most existing MPC-based building climate control strategies focus primarily on energy savings. Incorporating HP noise considerations directly into MPC design offers a promising, control-oriented pathway to mitigate acoustic disturbances in a flexible and economical manner. Nevertheless, formulating an MPC scheme that is scalable across different HP systems and effective in reducing acoustic nuisance remains an underexplored and nontrivial challenge.

1.2.2. DATA-DRIVEN ROBUST OPTIMIZATION

With the availability of abundant historical data and the rapid development of machine learning techniques, ML-based methods have been increasingly explored to construct data-driven uncertainty sets for RO problems [61]. By uncovering latent structures in uncertainty data, these approaches can yield more compact and computationally efficient uncertainty sets than conventional options, such as box, ellipsoidal, or budget sets. Moreover, instead of relying on domain-specific expertise, data-driven uncertainty sets can be automatically generated and tailored to specific engineering applications.

Despite their potential to mitigate the conservatism of traditional RO approaches, several open challenges remain in utilizing ML techniques for uncertainty set construction.

First, there are fundamental differences between solving ML tasks and constructing uncertainty sets. ML models are typically evaluated by prediction or classification accuracy, whereas such metrics may not exist – or may even be irrelevant – when defining uncertainty sets. For instance, clustering algorithms (e.g., K-Means) have been adopted to generate data-driven uncertainty sets [36, 62]. In ML contexts, these algorithms aim to minimize classification or clustering errors. In contrast, when used for RO, the focus shifts to building uncertainty sets that are compact and computationally tractable, which is generally not directly reflected by the minimized prediction/classification errors. Besides, uncertainty data may lack any label information in practice.

Second, computational complexity poses a significant challenge. Many state-of-the-art ML methods employ nonlinear transformations – such as kernel functions or deep neural networks – to enhance representation capability and prediction accuracy. However, when incorporated into uncertainty set construction, these nonlinearities often introduce nonconvexity and computational intractability, limiting their practical applica-

bility in RO formulations.

Furthermore, the methodological gap between standard ML practices and the requirements of RO remains underexplored. Conventional ML tuning or validation criteria may not be appropriate for developing ML-enabled uncertainty sets. Therefore, systematic guidelines are needed for selecting, tuning, and validating data-driven approaches within the RO context – an aspect that is overlooked in existing studies.

Finally, unlike classical RO formulations that typically involve a single uncertainty set, many ML-based data-driven sets yield composite or hierarchical uncertainty sets composed of multiple subsets. For such cases, further research is required to design computationally-efficient solution algorithms and to exploit the structure of these sets to reduce conservatism and enhance robust optimality.

1.2.3. DISTRIBUTIONALLY ROBUST MODEL PREDICTIVE CONTROL

Properly handling uncertainties has long been a central challenge in MPC design. To bridge the complementary benefits of the conventional RMPC and SMPC designs, the DRO framework has been introduced into MPC, leading to a variety of DRMPC formulations, see the works in [46, 49, 63–65], and the references therein.

Existing DRMPC formulations primarily focus on mitigating the effects of additive disturbances while assuming perfect knowledge of system dynamics and full state measurements. Although these assumptions simplify the analysis and enable robustness guarantees to some extent, they are far from realistic for practical situations. Moreover, most DRMPC designs adopt open-loop control actions instead of closed-loop control policies. This open-loop setting offers theoretical tractability but sacrifices the benefits of closed-loop policies in reducing conservatism and improving performance.

In real-world systems, the assumptions of perfect modeling and full state observability rarely hold. Due to modelling complexity, system identification errors, unmeasured disturbances, and the trade-off between model accuracy and computational efficiency, only approximated system models can be obtained in general for controller design. Likewise, because of the physical inaccessibility of certain states and sensor or cost limitations, only partial state measurements are typically available. These practical constraints also fundamentally limit the practicability of the existing DRMPC formulations. Furthermore, closed-loop feedback control policies – as opposed to open-loop action sequences – offer more expressiveness and adaptability, allowing the controller to dynamically react to real-time disturbances and estimation errors. However, incorporating such feedback structures into DRMPC frameworks introduces additional challenges in both theoretical analysis and computational tractability.

Considering the above discussion, it is more practical and meaningful to develop DRMPC schemes that explicitly account for imperfect system models, partial state measurements, and closed-loop feedback policies. Nevertheless, addressing these factors introduces several challenges:

- Imperfect models cause biased predictions of future system evolution, degrading the reliability of constraint satisfaction and performance guarantees.

- Feedback control policies amplify the coupling between uncertainty realizations and decision variables, complicating the reformulation and tractability of the un-

derlying optimization problem.

- Partial state measurements further complicate the construction of closed-loop prediction and performance analysis.

Developing DRMPC designs that overcome these challenges is essential to enhance the robustness, adaptability, and real-world applicability of predictive control under uncertainty.

1.3. OUR CONTRIBUTIONS

Given the scientific challenges and open problems discussed in the above section, this thesis will investigate the associated problems. The main contributions of this thesis are briefly summarized as follows:

1. A novel robust optimal control formulation considering binary adjustable uncertainties, along with computationally efficient solution methods, is proposed to address the "reserve and provision" problem with on-off type control inputs. (Chapter 2)

2. Motivated by grid congestion issues in the Netherlands, a practical DSM framework and an associated energy-flexible MPC design for HPTES systems are developed to exploit the energy flexibility of the HPTES system, and are experimentally validated on a real HPTES installation. (Chapter 3)

3. An MPC-based building climate control approach is proposed to mitigate acoustic pollution induced by HP operation, representing the first control-oriented solution to alleviate HP noise issues in smart building climate control. (Chapter 4)

4. An ML-based method for constructing data-driven uncertainty sets in RO models is proposed, together with guidelines for algorithm tuning and a Python-based toolkit, enabling compact, adaptive, and computationally tractable uncertainty quantification. (Chapter 5)

5. For uncertainty sets composed of multiple subsets, innovative formulations and corresponding computational algorithms are developed to: 1) address computational challenges in applying the RO model to predictive control problems, and 2) reduce the conservatism of RO formulations by exploiting the structure of uncertainty sets. (Chapter 6)

6. A novel distributionally robust system-level synthesis (DR-SLS) formulation, along with a computationally tractable relaxation, is proposed for constrained linear systems to enhance the robustness of control policies against model mismatch and additive disturbances. (Chapter 7)

1.4. THESIS OUTLINE

The overall structure of this thesis is graphically illustrated in Figure 1.1. Chapters 2–7 present a collection of works that have been published or submitted to peer-reviewed

journals and conferences. Since each chapter addresses distinct research problems and design frameworks, separate mathematical notations are adopted accordingly.

As discussed in Section 1.2, the main contributions of this thesis can be categorized into three research themes: 1) MPC design for building energy management (Chapters 2–4) (Chapter 2 considers a methodological study, Chapters 3 & 4 contain different application studies); 2) Data-driven robust optimization (Chapters 5–6); and 3) DRMPC for uncertain linear systems (Chapter 7). Among these chapters, Chapters 2 and 5–7 deal with robust optimization and predictive control design against uncertainties. Based on the types of the considered uncertainties, these chapters can be classified as: i) uncertainties induced by adjustable additive disturbances (Chapter 2), ii) uncertainties induced by fixed additive disturbances (Chapters 5 & 6); and iii) uncertainties induced by both fixed additive model mismatch and disturbances (Chapter 7). Furthermore, Chapters 5, 6 & 7 all aim to mitigate the conservatism of RO designs, albeit from different perspectives. Chapter 5 focuses on developing compact and data-informed uncertainty sets while retaining the conventional RO formulation. Chapters 6 & 7 investigate novel DRO formulations, which integrate RO and SO designs, and develop computationally efficient solutions by exploiting problem structures.

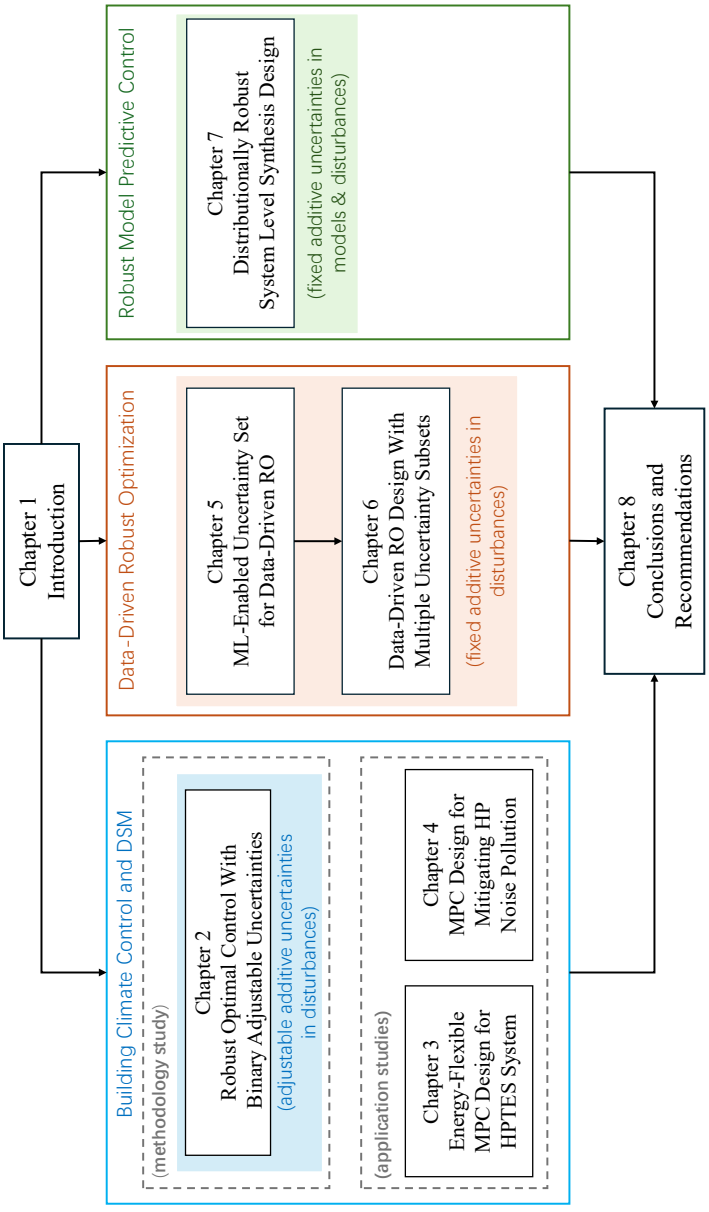


Figure 1.1: Thesis outline and structure.

2

Robust Optimal Control for Linear Systems With Binary Adjustable Uncertainties

Robust Optimal Control (ROC) with adjustable uncertainties has proven to be effective in addressing critical challenges within modern energy networks, especially the reserve and provision problem. However, existing research on ROC with adjustable uncertainties has predominantly focused on the scenario of uncertainties modeled as continuous variables. This assumption limits its applicability in many practical settings, such as the on-off operation of HVAC devices and power transmission line switching, where uncertainties are inherently discrete. In this chapter, we explore ROC with binary adjustable uncertainties, where the uncertainties are modeled by binary decision variables, marking the first investigation of its kind. To tackle this new challenge, firstly we introduce a metric designed to quantitatively measure the extent of binary adjustable uncertainties. Then, to balance computational tractability and adaptability, we restrict control policies to be affine functions with respect to uncertainties, and propose a general design framework for ROC with binary adjustable uncertainties. To address the inherent computational demands of the original ROC problem, especially in large-scale applications, we employ strong duality (SD) and big-M-based reformulations to create a scalable and computationally efficient Mixed-Integer Linear Programming (MILP) formulation. Numerical simulations are conducted to showcase the performance of our proposed approach, demonstrating its applicability and effectiveness in handling binary adjustable uncertainties within the context of modern energy networks.

This chapter is based on the paper: Li, Y., Yorke-Smith, N., & Keviczky, T. (2024, June). Robust Optimal Control With Binary Adjustable Uncertainties. In 2024 European Control Conference (ECC) (pp. 3721-3727). IEEE.

2.1. INTRODUCTION

In recent years, the so-called reserve and provision problem has gained increasing attention owing to its ability to characterize the demand side management issues in modern energy networks. As an efficient approach for modeling the reserve and provision problem, a new optimal control framework called *robust optimal control* (ROC) *with adjustable uncertainties* was proposed [55, 57, 58, 66–68]. Unlike the conventional robust optimal control problem, where the uncertainty sets are fixed and are determined by exogenous factors, ROC with adjustable uncertainties considers the case where the scope of uncertainties is adjustable and is determined by decision makers.

By allowing the decision maker to actively decide the admissible scope of uncertainty, the corresponding robust optimal control framework can be more versatile to include more practical cases. However, this versatility comes with more computational challenges. It is worth pointing out that existing research about ROC with adjustable uncertainties all focus on the case of continuous adjustable uncertainties. In the continuous case, uncertainty sets are typically represented as polyhedra or norm balls. As continuous variables yield an infinite number of uncertain scenarios within a given set, this leads to a semi-infinite optimization problem in the ROC formulation with adjustable uncertainties. To create a computationally tractable reformulation, strong duality theory is employed, reducing the need to satisfy universal constraints for all uncertain scenarios and instead focusing only on the worst-case scenario [55, 57, 58, 66–68].

Despite the success of ROC with continuous adjustable uncertainties in supporting modern energy networks, it should be noted that it fails to deal with the case where uncertainties can be binary. For example, in building climate control, some heating, ventilation and air conditioning (HVAC) devices only operate with two modes: on and off. If such devices are considered for providing demand response services, the existing ROC with continuous adjustable uncertainties fails to characterize this scenario. In addition, the existing design framework for ROC with continuous adjustable uncertainties cannot be directly generalized to deal with the binary case. On the one hand, polyhedra or norm balls used in the continuous case to describe the scope of uncertainties is not suitable for binary uncertainties. On the other hand, duality-based reformulation, which is built upon the assumption of continuous uncertainties, utilized in the continuous case for finding the worst uncertainty scenario cannot be directly applied to deal with the binary case. Motivated by this fact, in this chapter, we investigate the problem of ROC with binary adjustable uncertainties. Our main contributions can be summarized as follows:

- The problem of ROC with binary adjustable uncertainties is investigated for the first time. By introducing a new metric to evaluate the scope of binary uncertainties, we propose a general design formulation for ROC with binary adjustable uncertainties.

- Due to the high computational demand and poor scalability of the orignal formulation of ROC with binary adjustable uncertainties, we provide a computationally efficient and scalable alternative together with detailed derivation procedures. In addition, the probabilistic robustness of the optimal solution for our proposed approach is analyzed via Markov inequality.

- We consider a practical case of reserve and provision problem: utilizing the energy flexibility of buildings to provide demand response services and show how our proposed approach can be utilized to explore the energy flexibility potential. Simulation results are provided to demonstrate the effectiveness of the proposed approach.

The remaining parts of this chapter are organized as follows. Section 2.2 introduces the control problem and proposes a novel formulation of the ROC with binary adjustable uncertainties. For this original formulation, Section 2.3 gives a computationally efficient alternative with detailed derivation procedures and also analyzes the probabilistic robustness of the optimal solution. Simulation results are presented in Section 2.4, and the conclusion follows in Section 2.5.

Notation: Bold-face letters are used to denote the stacked time sequences of related signals; the operator $|\cdot|$ for a given set of indices denotes the cardinality of the set; subscript i of a given vector/matrix represents the i-th element/row of the corresponding vector/matrix; $\mathbb{N}_{[0,a]}$ denotes the set of nonnegative integers less than equal to a; max or min for a given vector implies row-wise maximization/minimization. $\sum_k x_k$ denotes the summation of all terms x_k, $\mathbf{1}_N$ denotes the N-dimensional all 1 vector.

2.2. PROBLEM FORMULATION

In this section, we will introduce the problem of robust optimal control with binary adjustable uncertainties. The considered system is

$$x(t+1) = Ax(t) + Br(t) + Du(t) + Ev(t) \tag{2.1}$$

where the $x(t) \in \mathbb{R}^n$ are system states, $r(t) \in \mathbb{B}^m$ are reference control inputs, $u(t) \in \mathbb{R}^p$ and $v(t) \in \mathbb{B}^q$ are continuous and binary recourse control variables, respectively. For system (2.1), we assume that the following linear state and input constraints should be satisfied

$$G_x x(t) \leq g_x, \tag{2.2a}$$
$$G_r r(t) + G_u u(t) + G_v v(t) \leq g_r, \tag{2.2b}$$

where (G_x, G_r, G_u, G_v) and (g_x, g_r) are matrices and vectors with approximate dimensions, respectively.

In our design framework, the reference control input is assumed to be predetermined but suffers from adjustable uncertainties, which is equivalent to the case that the system is subject to additive uncertainties.

Within the context of the reserve and provision problem, system (2.1) can model the dynamics of several utilities in energy systems, such as building thermal dynamics or energy storage devices. For example, for building climate control, system (2.1) can model the thermal dynamics of buildings that draw power from both main power grid $r(t)$, which is usually determined according to the day-ahead electricity market, and local renewable energy sources (RES) ($u(t)$ and $v(t)$), which can be flexibly scheduled in real time.

Based on (2.1), the predicted state evolution over N time steps is

$$\mathbf{x} = F_x x(0) + F_r \mathbf{r} + F_u \mathbf{u} + F_v \mathbf{v}, \tag{2.3}$$

where $\mathbf{x} = [x(1)^{\mathrm{T}}, \cdots, x(N)^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n \cdot N}$, $\mathbf{r} = [r(0)^{\mathrm{T}}, \cdots, r(N-1)^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{B}^{m \cdot N}$, $\mathbf{u} = [u(0)^{\mathrm{T}}, \cdots, u(N-1)^{\mathrm{T}}] \in \mathbb{R}^{p \cdot N}$, and $\mathbf{v} = [v(0)^{\mathrm{T}}, \cdots, v(N-1)^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{B}^{q \cdot N}$, and x^0 is the initial state vector. The detailed format of the matrixes F_x, F_r, F_u and F_v can be found in [69].

Correspondingly, constraints (2.2) within the prediction horizon can be compactly denoted as

$$\mathbb{X}^N := \{\mathbf{x} | G_{\mathbf{x}} \mathbf{x} \leq g_{\mathbf{x}}\}, \tag{2.4a}$$

$$\mathbb{U}^N := \{(\mathbf{u}, \mathbf{v}) | G_{\mathbf{r}} \mathbf{r} + G_{\mathbf{u}} \mathbf{u} + G_{\mathbf{v}} \mathbf{v} \leq g_{\mathbf{r}}\}, \tag{2.4b}$$

where $G_{\mathbf{x}} = \mathrm{diag}(G_x, \cdots, G_x)$, $G_{\mathbf{r}} = \mathrm{diag}(G_r, \cdots, G_r)$, $G_{\mathbf{u}} = \mathrm{diag}(G_u, \cdots, G_u)$, $G_{\mathbf{v}} = \mathrm{diag}(G_v, \cdots, G_v)$, $g_{\mathbf{x}} = \mathbf{1}_N \otimes g_x$, and $g_{\mathbf{r}} = \mathbf{1}_N \otimes g_r$.

In the reserve and provision framework, the reference control signal suffers from unknown external manipulations to provide services for external entities, and the design objective is to optimally quantify how much uncertainty (also called flexibility) can be reserved while guaranteeing system constraints. For the case of continuous adjustable uncertainties, the scope of flexibility is denoted as adjustable polyhedrons or norm balls, and the volume of uncertainty sets are adopted as metrics to measure the scope of uncertainties [58]. However, in the case of binary uncertainty, this definition of uncertainty set is not applicable because polyhedron/norm balls cannot accurately characterize uncertain binary variables, and their volumes also cannot truly reflect the scope of uncertainties.

In the following, we introduce a new metric to deal with binary adjustable uncertainties. The indices $k \in \mathbb{N}_{[0,mN-1]}$ for $\mathbf{r}_k$ within the prediction horizon are partitioned into two disjunctive groups $\mathscr{C}$ and $\mathscr{U}$. For all $k \in \mathscr{U}$, the corresponding reference inputs $\mathbf{r}_k$ are allowed to be adjusted flexibly. For all $k \in \mathscr{C}$, the reference inputs $\mathbf{r}_k$ are not influenced by uncertainties and are fixed to their nominal values $\bar{\mathbf{r}}_k$.

However, due to system constraints, not all reference inputs $\mathbf{r}_k (k \in \mathscr{U})$ can be flexibly adjusted without violating (2.2). Here, we define $\mathscr{S} \subseteq \mathscr{U}$ as the index set in which the corresponding $\mathbf{r}_k$ $(k \in \mathscr{S})$ is adjusted from $\bar{\mathbf{r}}_k$ to a new value, denoted as $\tilde{\mathbf{r}}_k$ with $\tilde{\mathbf{r}}_k \neq \bar{\mathbf{r}}_k$, without violating system constraints. Based on the above definition, the values of $\mathbf{r}_k$ within the prediction horizon can be represented as

$$\mathbf{r}_k = \begin{cases} \tilde{\mathbf{r}}_k, & k \in \mathscr{S} \\ \bar{\mathbf{r}}_k, & k \in \mathscr{C} \cup \mathscr{U} \setminus \mathscr{S} \end{cases} \tag{2.5}$$

A schematic diagram of flexible input signals with $m = 1$ is given in Fig. 2.1. Similarly to the case of continuous adjustable uncertainty, $\bar{\mathbf{r}}_k = 0$ and $\tilde{\mathbf{r}}_k = 1$ means up-ward flexibility, and $\bar{\mathbf{r}}_k = 1$ and $\tilde{\mathbf{r}}_k = 0$ means down-ward flexibility.

In contrast to the continuous case, where the volume of the uncertainty set is adopted to measure the scope of reserved flexibility, according to the definition (2.5), a new variable Γ is defined as the largest cardinality of $\mathscr{S}$ that can be freely selected among $\mathscr{U}$, to measure the scope of the reserved uncertainty/flexibility. An admissible Γ implies

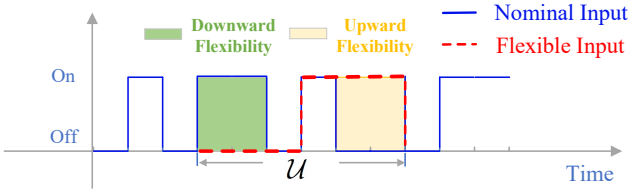


Figure 2.1: Schemetic diagram of flexible reference input signal $\mathbf{r}$.

that there exist feasible recourse inputs $(\mathbf{u},\mathbf{v})$ to ensure system constraint satisfaction for any possible reference input adjustment as long as $|\mathscr{S}| \leq \Gamma$. Clearly, a larger value of Γ implies that the controlled system can reserve more flexibility for providing services.

For the problem of ROC with adjustable uncertainties, a general control objective is to balance the minimization of operational cost and the maximization of the scope of reserved flexibility, while guaranteeing constraint satisfaction under all possible scenarios of uncertainties/flexibilities [58, 66]. Accordingly, the design objective for our considered problem takes the form

$$\min_{\mathbf{u},\mathbf{v},\Gamma} \left\{ \max_{\mathscr{S}} \{J(\mathbf{x},\mathbf{u},\mathbf{v})\} - \lambda\Gamma \right\} \tag{2.6a}$$

$$\text{s.t. } \mathbf{x} = F_x x^0 + F_r \mathbf{r} + F_u \mathbf{u} + F_v \mathbf{v} \tag{2.6b}$$

$$\mathbf{x} \in \mathbb{X}^N,\ (\mathbf{u},\mathbf{v}) \in \mathbb{U}^N, \tag{2.6c}$$

$$\mathbf{r}_k = \bar{\mathbf{r}}_k,\ \forall k \in \mathscr{C} \cup \mathscr{U} \setminus \mathscr{S}, \tag{2.6d}$$

$$\mathbf{r}_k = \tilde{\mathbf{r}}_k,\ \forall k \in \mathscr{S}, \tag{2.6e}$$

$$\forall \mathscr{S} : |\mathscr{S}| \leq \Gamma, \tag{2.6f}$$

where $J(\mathbf{x},\mathbf{u},\mathbf{v}) := \sum_{k=0}^{N} l_k(\mathbf{x},\mathbf{u},\mathbf{v})$ is the summation of linear stage cost functions $l_k(\mathbf{x},\mathbf{u},\mathbf{v})$, $\lambda > 0$ is a user-selected weighting parameter.

The objective function (2.6a) aims at minimizing the worst-case operational cost given Γ, namely $\max_{\mathscr{S}} J(\mathbf{x},\mathbf{u},\mathbf{v})$, w.r.t. the input uncertainties determined by $\mathscr{S}$ and in the meanwhile maximizing the scope of flexibility Γ. If the worst-case operational cost function $\max_{\mathscr{S}} J(\mathbf{x},\mathbf{u},\mathbf{v})$ is removed, optimization problem (2.6) is to find out the largest feasible Γ. Namely, the largest cardinality of $\mathscr{S}$ that can be freely selected without violating system constraints. For example, in Fig. 2.1, four inputs among $\{\mathbf{r}_k : \forall k \in \mathscr{U}\}$ are adjusted, if the optimal Γ^* computed in (2.6) satisfies $\Gamma^* \geq 4$, we are still capable of ensuring system constraints for this scenario of flexible inputs.

To ensure the feasibility of (2.6), it is assumed that for the predefined reference input $\bar{\mathbf{r}}$, there exists at least one input sequence $(\mathbf{u},\mathbf{v})$ such that the state and input constraints (2.6c) are satisfied.

Remark 2.1. We refer to the optimal control problem (2.6), similarly to its continuous counterparts: ROC with continuous adjustable uncertainties [58, 66], as *ROC with binary adjustable uncertainties*. The main difference between ROC with adjustable uncertainties and conventional ROC problem lies in the uncertainty set. For conventional ROC problems, the uncertainty sets are predetermined and fixed. But for the ROC with

adjustable uncertainties, the shape/size of uncertainty sets are not fixed but will be optimally determined by decision variables, such as Γ in (2.6).

Remark 2.2. We highlight that the problem formulation proposed in (2.6) can be applied to several important issues in modern energy systems. For example, similarly to [55, 67], formulation (2.6) can model the reserve and provision problem of buildings with on-off types of HVAC devices for providing demand response services. Besides, as shown in [70], it is also possible to extend the formulation (2.6) to analyze and design resilient control strategies for power grid networks to mitigate faults or cyberattacks.

2.3. ROBUST OPTIMAL CONTROL DESIGN AND ROBUSTNESS ANALYSIS

This section is devoted to solving the problem of ROC with binary adjustable uncertainties formulated in (2.6), and to analyzing the robustness of the derived solution.

2.3.1. ROBUST OPTIMAL CONTROL DESIGN

Considering the system dynamics (2.1), state and input constraints (2.4), and the linear property of the objective function, the optimization problem in (2.6) can be rewritten in the following compacted form

$$\min_{\theta,\Gamma,\mathbf{u},\mathbf{v}} \ \theta - \lambda\Gamma \tag{2.7a}$$

$$\text{s.t. } \mathbf{Or} + \mathbf{Pu} + \mathbf{Qv} \le \mathbf{h}, \tag{2.7b}$$

$$\mathbf{r}_k = \tilde{\mathbf{r}}_k, \ \forall k \in \mathscr{C} \cup \mathscr{U} \setminus \mathscr{S}, \tag{2.7c}$$

$$\mathbf{r}_k = \tilde{\mathbf{r}}_k, \ \forall k \in \mathscr{S}, \tag{2.7d}$$

$$\forall \mathscr{S} : \ |\mathscr{S}| \le \Gamma, \tag{2.7e}$$

where matrixes $\mathbf{O}$, $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{h}$ are constructed according to system dynamics, constraints and problem data, θ is introduced for the epigraph reformulation of the worst-case operational cost in (2.6) as robust constraints $J(\mathbf{x},\mathbf{u},\mathbf{v}) \le \theta \ \forall \mathscr{S} : |\mathscr{S}| \le \Gamma$, which are compactly included in (2.7b). Similar reformulation can be found in [58].

Unlike the continuous case, in which the original formulation of ROC with adjustable uncertainties is semi-infinite and is computationally intractable, the optimization problem (2.7) can be solved directly without further reformulation since it only entails a finite number of uncertainties $\tilde{\mathbf{r}}_k$ and $\mathscr{S}$ given a fixed Γ. By trying out all possible Γ and considering all corresponding uncertain scenarios, the optimal solution can be found by exhaustive search. However, it should be noted that given a fixed Γ, the number of constraints that need to be considered in (2.7) is proportional to $\sum_{i=0}^{\Gamma} \binom{|\mathscr{U}|}{i} \cdot N$, which can result in a large scale MILP that is computationally demanding. In the following, we will propose a novel reformulation for (2.7) such that the resulting problem is scalable and computationally efficient.

For the recourse control inputs $\mathbf{u}$ and $\mathbf{v}$, we will design closed-loop control policies instead of open-loop control actions since it has been demonstrated in many works that closed-loop control policies can tolerate more uncertainties than open-loop control actions [8, 71, 72]. Because considering arbitrary control policies can make the problem

computationally intractable, we consider the following affine control policies to balance robustness, optimality, and computational effort for our proposed scheme

$$\mathbf{u}_k = \sum_{i\in\mathscr{U}} M_{ki}\mathbf{r}_i + \eta_k, \tag{2.8a}$$

$$\mathbf{v}_k = \sum_{i\in\mathscr{U}} L_{ki}\mathbf{r}_i + \epsilon_k, \tag{2.8b}$$

where $M_{ki}\in\mathbb{R}$, $\eta_k\in\mathbb{R}$, $L_{ki}\in\mathbb{Z}$ and $\epsilon_k\in\mathbb{Z}$ are control policy parameters to be optimized. Since the exact value of the flexible inputs $\tilde{\mathbf{r}}_k (k\in\mathscr{U})$ is not determined when solving (2.7), the control policy (2.8) will adapt inputs $\mathbf{u}_k$ and $\mathbf{v}_k$ once $\mathbf{r}_i (i\in\mathscr{U})$ are revealed, which can increase the robustness of the controlled system so that more flexibility can be reserved. Note that binary restriction for $\mathbf{v}_k$ can be imposed by a linear constraint $0\le\mathbf{v}_k\le 1$ since the control policy (2.8b) will only generate integer outputs when $L_{ki}\in\mathbb{Z}$ and $\epsilon_k\in\mathbb{Z}$.

Theorem 2.1. Considering the system dynamics (2.1), the binary adjustable input uncertainties (2.5) and the control policies (2.8), the robust optimal control problem in (2.7) can be reformulated as the following MILP problem

$$\min_{\substack{\theta,\delta_k,\mu_{ij},\phi_{ij},y_{ij}\beta_{ij},\\ M_{kj},L_{kj},\eta_k,\epsilon_k}} \quad \theta-\lambda\sum_{k\in\mathscr{U}}\delta_k \tag{2.9a}$$

$$\text{s.t.} \sum_{j\in\mathscr{U}}\mu_{ij}+\sum_{j\in\mathscr{U}}\beta_{ij}+\phi_i+\sum_k\mathbf{P}_{ik}\eta_k+\sum_k\mathbf{Q}_{ik}\epsilon_k+\sum_{j\in\mathscr{C}}\mathbf{O}_{ij}\bar{\mathbf{r}}_j\le\mathbf{h}_i, \tag{2.9b}$$

$$\mu_{ij}+\pi_i\ge\frac{1}{2}\Big(\mathbf{O}_{ij}+\sum_k\mathbf{P}_{ik}M_{kj}+\sum_k\mathbf{Q}_{ik}L_{kj}+y_{ij}\Big)+\Big(\mathbf{O}_{ij}+\sum_k\mathbf{P}_{ik}M_{kj}+\sum_k\mathbf{Q}_{ik}L_{kj}\Big)\bar{\mathbf{r}}_j, \tag{2.9c}$$

$$-y_{ij}\le\mathbf{O}_{ij}+\sum_k\mathbf{P}_{ik}M_{kj}+\sum_k\mathbf{Q}_{ik}L_{kj}\le y_{ij}, \tag{2.9d}$$

$$\phi_i\ge\sum_{j\in\mathscr{U}}\Big(\mathbf{O}_{ij}+\sum_k\mathbf{P}_{ik}M_{kj}+\sum_k\mathbf{Q}_{ik}L_{kj}\Big)\cdot\bar{\mathbf{r}}_j, \tag{2.9e}$$

$$0\le\beta_{ij}\le M\delta_j, \tag{2.9f}$$

$$0\le\pi_i-\beta_{ij}\le M(1-\delta_j), \tag{2.9g}$$

$$\delta_j\in\mathbb{B},\ \mu_{ij}\ge 0,\ \pi_i\ge 0,\ y_{ij}\ge 0, \tag{2.9h}$$

$$i\in\mathscr{I},\ j\in\mathscr{U}, \tag{2.9i}$$

where $\mathscr{I}$ is the set of row indices of constraints (2.7b), and the optimal solution δ_k^* satisfies $\sum_{j\in\mathscr{U}}\delta_j^*=\Gamma^*$.

Proof: The proof is presented in the Appendix 2.A.1.

Remark 2.3. For the original formulation in (2.7), given a fixed Γ, in order to guarantee constraint satisfaction robustly, the number of constraints considered is proportional to $\sum_{i=0}^{\Gamma}\binom{|\mathscr{U}|}{i}\cdot N$. In contrast, the number of constraints in the reformulated optimization problem in (2.9) is only proportional to $|\mathscr{U}|\cdot N$, which is more favourable when $\sum_{i=0}^{\Gamma}\binom{|\mathscr{U}|}{i}\gg|\mathscr{U}|$. In addition, the constraint (2.22a) has nonconvex term $\Gamma\pi_i$ since

$\Gamma \in \mathbb{N}_{[0,|\mathscr{U}|]}$ is a decision variable. While this nonconvexity in (2.22a) can be dealt with by several solvers, e.g. `Gurobi`, with the big-M based reformulation in (2.23), the final optimization problem (2.9), which is an MILP, can be solved more efficiently by more off-the-shelf solvers, such as `Gurobi`, `CPLEX`, and `GLPK`.

2.3.2. ROBUSTNESS ANALYSIS OF THE OPTIMAL SOLUTION

The obtained optimal control strategy, namely $(M_{kj}^*, \eta_k^*, L_{kj}^*, \epsilon_k^*)$, is able to guarantee constraint satisfaction if no more than Γ of the flexible reference inputs $\mathbf{r}_j$ $(j \in \mathscr{U})$ are changed. In practice, however, it is possible that in some cases more than Γ number of reference inputs $\mathbf{r}_k$ are subjected to uncertainties, e.g., in case of device faults. As a result, it is of interest and importance to analyze the robustness of the derived solution for such cases. In the following, we will investigate the possibility of constraint violation with the derived optimal solution when more than Γ number of $\mathbf{r}_j$ $(j \in \mathscr{U})$ might be changed.

In our analysis, $\mathbf{r}_j$ $(j \in \mathscr{U})$ are regarded as random variables (r.v.). Consequently, the left-hand side of constraint (2.7b) are also r.v. For the i-th row of the constraint (2.7b), its probability of violation is

$$\mathbb{P}_{\text{vio}} = \Pr\big(\mathbf{O}_i^{\mathrm{T}}\mathbf{r} + \sum_{j\in\mathscr{U}} \mathbf{r}_j \sum_k \mathbf{P}_{ik} M_{kj}^* + \sum_{j\in\mathscr{U}} \mathbf{r}_j \sum_k \mathbf{Q}_{ik} L_{kj}^* + \sum_k \mathbf{P}_{ik}\eta_k^* + \sum_k \mathbf{Q}_{ik}\epsilon_k^* \geq \mathbf{h}_i\big). \tag{2.10}$$

For notational brevity, we define

$$a_{ij} := \mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik} M_{kj}^* + \sum_k \mathbf{Q}_{ik} L_{kj}^*, \tag{2.11a}$$

$$b_i := \mathbf{h}_i - \sum_{j\in\mathscr{C}} \mathbf{O}_{ij}\mathbf{r}_j - \sum_k \mathbf{P}_{ik}\eta_k^* - \sum_k \mathbf{Q}_{ik}\epsilon_k^*. \tag{2.11b}$$

Proposition 2.1. Assuming that $\mathbf{r}_j \in \mathbb{B}$ $(j \in \mathscr{U})$ are independent random variables, then the probability of violation for the i-th constraint (2.10) satisfies

$$\mathbb{P}_{\text{vio}} \leq \frac{\prod_{j\in\mathscr{U}} \mathbb{E}\big(\exp(a_{ij}\mathbf{r}_j)\big)}{\exp(b_i)}. \tag{2.12}$$

Proof: Based on the definitions of a_{ij} and b_j in (2.11), the probability of violation (2.10) can be rewritten as

$$\mathbb{P}_{\text{vio}} = \Pr\Big[\sum_{j\in\mathscr{U}} a_{ij}\mathbf{r}_j \geq b_i\Big] \tag{2.13a}$$

$$= \Pr\Big\{\exp\big(\sum_{j\in\mathscr{U}} a_{ij}\mathbf{r}_j\big) \geq \exp(b_i)\Big\}. \tag{2.13b}$$

Then, it follows from the Markov inequality [73] and the independence property of $\mathbf{r}_j$ that

$$\mathbb{P}_{\text{vio}} \leq \frac{\prod_{j\in\mathscr{U}} \mathbb{E}\big(\exp(a_{ij}\mathbf{r}_j)\big)}{\exp(b_i)}. \tag{2.14}$$

This completes the proof. □

It should be noted that the bound of the probability of violation in (2.12) relies on system information in a_{ij} and b_i. In the following, we give a probability bound that does not rely on system information and is easy to compute.

Proposition 2.2. Assuming that $\mathbf{r}_j$ $(j \in \mathscr{U})$ are independent random variables with $\Pr(\mathbf{r}_j \neq \bar{\mathbf{r}}_j) = \epsilon_j$, Then, we have $\mathbb{P}_{\text{vio}} \leq \frac{\sum_{j\in\mathscr{U}} \epsilon_j}{\Gamma}$.

Proof: Define random variables z_j as the indicator functions of the events $\mathbf{r}_j \neq \bar{\mathbf{r}}_j$ $(j \in \mathscr{U})$. Then, it is readily concluded that $\mathbb{P}_{\text{vio}} \leq \Pr(\sum_{j\in\mathscr{U}} z_j \geq \Gamma)$. By applying Markov inequality and the independence of $\mathbf{r}_j$, it yields

$$\mathbb{P}_{\text{vio}} \leq \frac{\mathbb{E}(\sum_{j\in\mathscr{U}} z_j)}{\Gamma} \tag{2.15a}$$

$$= \frac{\sum_{j\in\mathscr{U}} \mathbb{E}(z_j)}{\Gamma} = \frac{\sum_{j\in\mathscr{U}} \epsilon_j}{\Gamma}. \tag{2.15b}$$

This completes the proof. □

While the bound of probability violation in Proposition 2.2 is independent of system information and is easy to compute, it is very conservative when Γ is small or $\sum_{j\in\mathscr{U}} \epsilon_j$ is large.

2.4. SIMULATION RESULTS

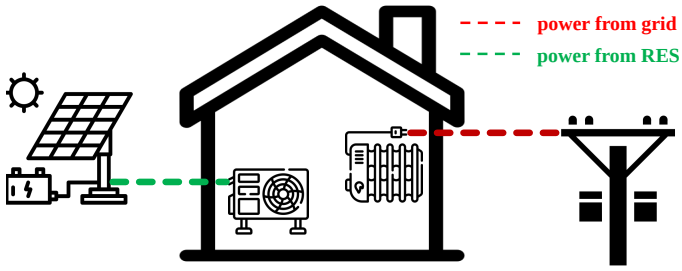


Figure 2.2: Diagram of energy consumption options.

The effectiveness of our proposed approach is illustrated via a numerical case study about the reserve and provision problem for a smart building system. A schematic diagram of the considered building system is depicted in Fig. 2.2. In our simulation, it is assumed that the thermal energy of the building is provided by three heat pumps (HP) that draw electricity from two sectors: power grid and local renewable energy sources (a solar panel). The HP powered by the grid has only two operation modes: on and off, which is denoted as $\mathbf{r}$, and the nominal operation status $\bar{\mathbf{r}}$ is predetermined according to day-ahead electricity price. For the remaining two HPs drawing power from local RES, their control signals $(\mathbf{u}, \mathbf{v})$ are available for adjusting in real-time. One HP has only on-off operational modes, and the other can operate continuously, e.g., PWM based operation. A solar panel is equipped to provide RES.

As ubiquitous storage devices of energy flexibility, buildings are utilized to provide demand response (DR) services to the power grid. During demand response period, the

building management system is asked to adjust its electricity consumption, namely the operational pattern of $\mathbf{r}$, according to the requests of the grid operator.

In our simulation, the sampling period is selected as 30 min and the prediction horizon is set as $N = 48$, i.e., one day. The starting time instant of flexibility assessment period $\mathscr{U}$ is selected as 22 (11am). A solar panel with a maximal 2000W electricity output is included to provide RES. The initial indoor temperature is selected as $x(0) = 21°\text{C}$, and the indoor temperature bound is $[20°\text{C}, 24°\text{C}]$.

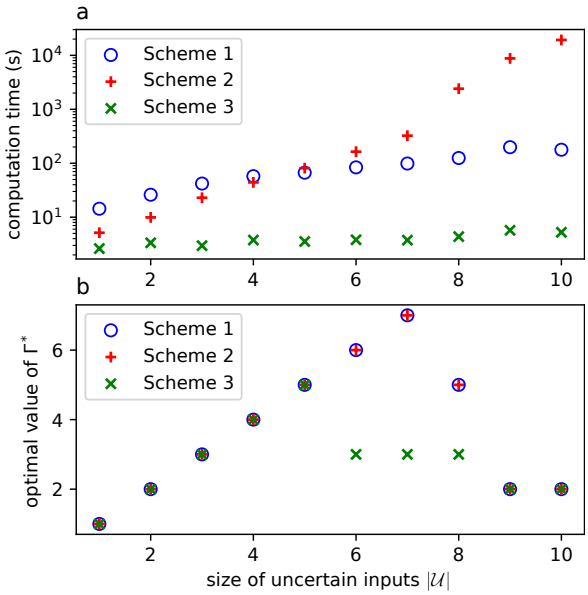


Figure 2.3: Computation time and optimal Γ^* for all schemes.

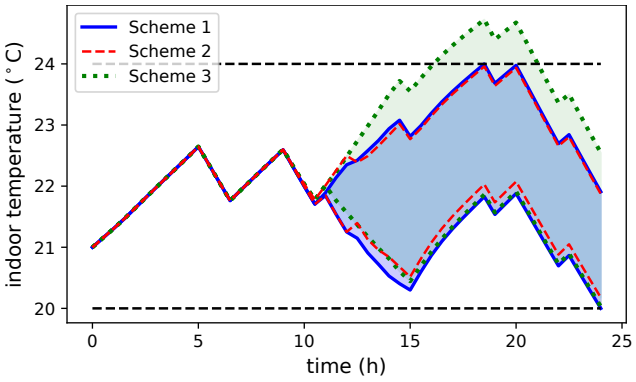


Figure 2.4: Indoor temperature envelopes for all schemes.

The objective for the optimization problem (2.7) is selected to maximize Γ. Namely,

we want to assess the largest flexibility potential of the considered building system within the period of $\mathscr{U}$. Three schemes are adopted in our simulation:

- Scheme 1: our proposed formulation (2.9).
- Scheme 2: original formulation (2.7) solved via exhaustive search.
- Scheme 3: our proposed formulation (2.9) but only considering open-loop control actions ($M_{ki} = 0$, $L_{ki} = 0$ in (2.8)).

All simulations are carried out using Gurobi 9.5.1 [74] on an Intel Xeon W-2223 CPU at 3.60GHz with 16G RAM. Optimization problems are modeled via the Python package Pyomo [75].

Fig. 2.3 gives the computation time and the optimal Γ^*, i.e., the largest number of flexible switches of $\mathbf{r}_j$ ($j \in \mathscr{U}$), for all schemes with an increasing size of $|\mathscr{U}|$. Clearly, it can be observed from Fig. 2.3a that the computation time of Scheme 2 (exhaustive search) increases exponentially with $|\mathscr{U}|$. On the contrary, our proposed approach is much more computationally efficient, and its computation time increases linearly with $|\mathscr{U}|$. Compared with Scheme 1, Scheme 3, which only considers open-loop control actions instead of closed-loop control policies as in Scheme 1, is computationally less demanding due to its smaller number of decision variables. Obviously, the downside of Scheme 3 is its suboptimality, as shown in Fig. 2.3b. In addition, we can also conclude from Fig. 2.3b that, for this case study, the affine control policy (2.8) restricted in our proposed approach (Scheme 1) does not compromise the optimality of the solution and can find the optimal Γ^* as with Scheme 2.

Fig. 2.4 depicts the indoor temperature envelopes with all 3 schemes under all possible uncertain scenarios of $\mathbf{r}$ given $|\mathscr{U}| = 8$ and $\Gamma = 5$. According to the computation results in Fig. 2.3b, it it shown that Schemes 1 and 2 are both able to tolerate all possible uncertainties but Scheme 3 can only ensure robust constraint satisfaction with $\Gamma = 3$. Clearly, as can be seen from Fig. 2.4, Scheme 3 is unable to robustly guarantee indoor comfort constraints, which validates the conclusion that open-loop control actions are less robust than closed-loop control policies.

2.5. CONCLUSIONS

This chapter investigated a class of ROC problems with binary adjustable uncertainties, that offers a framework for addressing the reserve and provision problem for energy systems with on-off type devices. In contrast to the existing methods for ROC with continuous adjustable uncertainties, a novel metric is introduced to measure the extent of binary uncertainties, and a general design framework for ROC with binary adjustable uncertainties is formulated. Subsequently, we reformulated and relaxed the ROC formulation by introducing auxiliary variables and applying strong duality, transforming it into a MILP problem amenable to efficient numerical solvers. This innovative approach empowers us to quantitively evaluate the flexibility potential inherent in modern energy systems with on-off type equipment. The efficacy of our proposed methodology is demonstrated through numerical experiments.

Future extensions include considering nonlinear constraints and nonaffine control policies, as well as hybrid types of adjustable uncertainties.

2.A. APPENDIX

2.A.1. PROOF OF THEOREM 2.1

Proof: Given a fixed Γ, system constraint (2.7b) has to be satisfied for all possible $\tilde{\mathbf{r}}_k$ and $\mathscr{S}$ such that $k \in \mathscr{S}$ and $|\mathscr{S}| \leq \Gamma$. To ensure constraint satisfaction for all possible uncertainties, which entails the order of $\sum_{i=0}^{\Gamma} \binom{|\mathscr{U}|}{i}$ uncertainty scenarios, we alternatively design a control scheme to guarantee constraint satisfaction under the worst case scenario of uncertainty, which can be represented as

$$\max_{\mathscr{S}} \left\{ \mathbf{O}\mathbf{r} + \mathbf{P}\mathbf{u} + \mathbf{Q}\mathbf{v} : |\mathscr{S}| \leq \Gamma, \mathbf{r}_j \text{ satisfies (2.5)} \right\} \leq \mathbf{h}, \tag{2.16}$$

where the maximization and inequality in (2.16) is computed and hold row-wise, respectively. The i-th row of constraints (2.16), combining the control policies (2.8) and the reference input pattern (2.5) can be expressed as

$$\max_{\mathscr{S}:|\mathscr{S}|\leq\Gamma} \left\{ \mathbf{O}_i^{\mathrm{T}}\mathbf{r} + \sum_k \mathbf{P}_{ik} \Big(\sum_{j\in\mathscr{U}} M_{kj}\mathbf{r}_j + \eta_k \Big) + \sum_k \mathbf{Q}_{ik} \Big(\sum_{j\in\mathscr{U}} L_{kj}\mathbf{r}_j + \epsilon_k \Big) \right\} \leq \mathbf{h}_i,$$

which can be further reformulated as

$$\begin{aligned} \max_{\mathscr{S}:|\mathscr{S}|\leq\Gamma} \Bigg\{ & \sum_{j\in\mathscr{S}} \Big(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} \Big) \tilde{\mathbf{r}}_j \\ & + \sum_{j\in\mathscr{U}\setminus\mathscr{S}} \Big(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} \Big) \bar{\mathbf{r}}_j \Bigg\} + \sum_k \mathbf{P}_{ik}\eta_k + \sum_k \mathbf{Q}_{ik}\epsilon_k + \sum_{j\in\mathscr{C}} \mathbf{O}_{ij}\bar{\mathbf{r}}_j \leq \mathbf{h}_i. \end{aligned} \tag{2.17}$$

As explained in [76], it can be verified that

$$\begin{aligned} \Big(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} \Big) \tilde{\mathbf{r}}_j \leq \frac{1}{2} \Bigg(& \mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} + \\ & \Big| \mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} \Big| \Bigg). \end{aligned} \tag{2.18}$$

Accordingly, by following a similar line as in [77] and introducing additional decision variables $z_j \in \mathbb{R}$ ($j \in \mathscr{U}$), the maximization problem in (2.17) is relaxed as

$$\begin{aligned} \max_{z_j} \sum_{j\in\mathscr{U}} \frac{z_j}{2} \cdot & \Bigg(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} + \Big| \mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} \Big| \Bigg) + && \text{(2.19a)} \\ & (1 - z_j) \cdot \Big(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} \Big) \bar{\mathbf{r}}_j \\ \text{s.t. } & 0 \leq z_j \leq 1, \quad \sum_{j\in\mathscr{U}} z_j \leq \Gamma. && \text{(2.19b)} \end{aligned}$$

Since (2.19) is a linear program (LP) w.r.t. z_j, according to strong duality theory of LP, see [78], the objective value of (2.19) coincides with that of the following dual problem:

$$\min_{\mu_{ij},\pi_i,\phi_i} \sum_{j\in\mathscr{U}} \mu_{ij} + \Gamma\pi_i + \phi_i \tag{2.20a}$$

$$\text{s.t. } \mu_{ij} + \pi_i \geq \frac{1}{2}\left(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} + \left|\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj}\right|\right) - \left(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj}\right)\bar{\mathbf{r}}_j, \tag{2.20b}$$

$$\phi_i \geq \sum_{j\in\mathscr{U}} \left(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj}\right)\bar{\mathbf{r}}_j, \tag{2.20c}$$

$$\mu_{ij} \geq 0,\ \pi_i \geq 0,\ \forall j \in \mathscr{U}, \tag{2.20d}$$

where μ_{ij} and π_i are Lagrange multipliers.

Optimization problem (2.20) is nonlinear because it contains the absolute function, which will incur increased computational effort. By introducing auxiliary decision variables $y_{ij} \geq 0$, optimization problem (2.20) can be further relaxed as the following linear optimization problem

$$\min_{\mu_{ij},\pi_i,\phi_i} \sum_{j\in\mathscr{U}} \mu_{ij} + \Gamma\pi_i + \phi_i \tag{2.21a}$$

$$\text{s.t. } \mu_{ij} + \pi_i \geq \frac{\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} + y_{ij}}{2} - \left(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj}\right)\cdot\bar{\mathbf{r}}_j, \tag{2.21b}$$

$$-y_{ij} \leq \mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} \leq y_{ij}, \tag{2.21c}$$

$$\phi_i \geq \sum_{j\in\mathscr{U}} \left(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj}\right)\cdot\bar{\mathbf{r}}_j, \tag{2.21d}$$

$$\mu_{ij} \geq 0,\ \pi_i \geq 0,\ y_{ij} \geq 0,\ \forall j \in \mathscr{U}. \tag{2.21e}$$

For the inequality (2.17), the maximization problem can be relaxed by a feasible solution

of the above LP problem to yield the following alternative constraints

$$\sum_{j\in\mathscr{U}} \mu_{ij} + \Gamma\pi_i + \phi_i + \sum_k \mathbf{P}_{ik}\eta_k + \sum_k \mathbf{Q}_{ik}\epsilon_k + \sum_{j\in\mathscr{C}} \mathbf{O}_{ij}\bar{\mathbf{r}}_j \le \mathbf{h}_i, \tag{2.22a}$$

$$\mu_{ij} + \pi_i \ge \frac{1}{2}\Big(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} + y_{ij}\Big) - \Big(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj}\Big)\cdot\bar{\mathbf{r}}_j, \tag{2.22b}$$

$$-y_{ij} \le \mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj} \le y_{ij}, \tag{2.22c}$$

$$\phi_i \ge \sum_{j\in\mathscr{U}} \Big(\mathbf{O}_{ij} + \sum_k \mathbf{P}_{ik}M_{kj} + \sum_k \mathbf{Q}_{ik}L_{kj}\Big)\cdot\bar{\mathbf{r}}_j, \tag{2.22d}$$

$$\mu_{ij} \ge 0,\ \pi_i \ge 0,\ y_{ij} \ge 0,\ \forall j \in \mathscr{U}. \tag{2.22e}$$

Furthermore, introducing $\delta_j \in \mathbb{B}$ ($j \in \mathscr{U}$) and $\beta_{ij} \in \mathbb{R}$ ($i \in \mathscr{I}, j \in \mathscr{U}$) and defining $\Gamma := \sum_{j\in\mathscr{U}} \delta_j$, the term $\Gamma\pi_i$ in (2.22a) can be further relaxed via big-M formulation as

$$\Gamma\pi_i = \sum_{j\in\mathscr{U}} \beta_{ij}, \tag{2.23a}$$

$$0 \le \beta_{ij} \le M\delta_j, \tag{2.23b}$$

$$0 \le \pi_i - \beta_{ij} \le M(1-\delta_j), \tag{2.23c}$$

$$\delta_j \in \mathbb{B}, \tag{2.23d}$$

where $M > 0$ is a sufficiently large constant. Finally, constraints (2.7b)–(2.7e) can be replaced by (2.22) and (2.23), which result in the optimization problem (2.9). This completes the proof. □

3

Model Predictive Control for Unlocking Energy Flexibility of Heat Pump and Thermal Energy Storage Systems

Increasing penetration of renewable energy sources (RES) and electrification of energy systems necessitates the engagement of demand-side management (DSM) to help alleviate congestion in the electricity grid. Heat pump and thermal energy storage (HPTES) systems, being energy-efficient solutions, are becoming popular in modern buildings and are promising to contribute to demand-side management (DSM) due to their significant share in household electricity consumption. For typical HPTES systems serving domestic hot tap water demand, this chapter presents a practical energy-flexible model predictive control (MPC) design for exploiting the energy flexibility of the HPTES system. The proposed MPC-based DSM strategy offers an innovative solution for efficient DSM by following a two-step DSM framework. In the first step, flexibility assessment is performed to quantitatively evaluate the flexibility potential of the HPTES system by solving a mixed-integer economic MPC problem. In the second step, flexibility exploitation is achieved through reacting to feasible demand response (DR) requests while respecting system constraints. Both numerical simulations and real-world experiments are performed based on a real HPTES installation to showcase the viability and effectiveness of the proposed design.

This chapter is based on the paper: Tang, W.[1], Li, Y.[1], Walker, S., & Keviczky, T. (2025). Model Predictive Control for Unlocking Energy Flexibility of Heat Pump and Thermal Energy Storage Systems: Experimental Results. arXiv preprint arXiv:2507.17552. ([1]: These authors contributed equally to this work.)

3.1. Introduction

In pursuit of carbon neutrality and energy sustainability, all economies have set to improve the utilization of renewable energy sources (RES), such as solar energy, wind energy, etc. For example, the European Union (EU) has set an ambitious target to increase the share of RES to at least 27% by 2030 [79]. It should be highlighted that the intermittency and volatility of RES pose significant challenges for maintaining a real-time balance between electricity supplement and consumption in power grids. With the aim of reducing congestion, the concept of demand-side management (DSM) has been proposed. Unlike conventional supply-side management, DSM involves strategic adjustments of the energy consumption of end users in response to the need of power grids [55, 80].

Buildings, as major energy consumers, contribute to about 40% of total energy consumption. The heating, ventilation, and air conditioning (HVAC) devices consume more than half of the total household energy consumption. Due to the significant energy consumption of HVAC systems in buildings, there has been growing interest in leveraging their energy flexibility to deliver DSM services. As energy-efficient heating/cooling devices, heat pumps (HPs) have gained increasing popularity. The EU's directive on renewables identified heat pumps as a vital technology for exploiting renewable energy [81]. In recent years, HPs are widely adopted across Europe to provide thermal energy for various building applications, such as domestic hot water usage and floor heating. Currently, 60 million heat pumps are installed in Europe, and the number of HP installations is experiencing a sharp rise. In 2022 alone, there were about 3 million HPs installed, and it is expected that at least 10 million additional heat pumps will be installed by 2027 [82]. A typical application of HP is its integration with thermal energy storage (TES) tanks, referred to as heat pump and thermal energy storage (HPTES) systems. By buffering hot water in water tanks, the integration of TES allows HP to operate flexibly for improving energy efficiency and reducing energy costs without sacrificing the required hot water supply [83]. Given the high power demands created by HPs, it is promising to explore the possibility of exploiting the energy flexibility emanating from HPTES systems to achieve DSM.

In practice, the majority of the HPTES systems operate using rule-based control strategies. For example, HP is turned on when the TES temperature drops below a specified threshold and turned off when it exceeds that threshold. While rule-based approaches are easy to implement, they fail to properly exploit the energy flexibility of the system. It is commonly accepted that advanced control strategies are necessary to harness the energy flexibility of HPTES systems. As an advanced control technique, MPC has become a promising approach in operating HPs and TES systems in building sectors due to its versatility in dealing with system constraints and in incorporating economic factors and predicted system behaviors, see [55, 84–89] and the references therein.

For implementing MPC schemes, prediction models of HP and TES systems are necessary. While there are many existing works developing white-box models of HP and TES systems to give high approximation accuracy, see [90–92] and the references therein, these models are generally too complex to be considered in MPC design. We introduce simplified models for MPC design to balance the approximation accuracy and computational cost. In general, HP models can be simplified as COP functions to bridge the

relationship between the HP's thermal generation and electricity consumption [67, 89, 93, 94]. As for TES systems, they are typically modeled as either a single thermal node, assuming the water within the tank is properly mixed, or as multiple thermal nodes, neglecting the temperature gradient along radial directions [52, 88].

Although the existing literature extensively investigated the viability and effectiveness of MPC in controlling HVAC devices in building sectors, including HP and TES systems, there is still room for improvement, particularly in the context of DSM. For the MPC-based DSM design within the context of TES systems, most of the existing approaches adopt either the price-based or incentive-based programs, in which HVAC devices are incentivized by some price signals to voluntarily adjust their operation patterns in line with power grid requirements [40, 52, 53], relying only on unidirectional communication from the grid operator to the resource manager. However, as noted in [95–97], these approaches, despite being relatively straightforward and easy to implement, often fail to fully exploit the energy flexibility of the system and may not meet the grid's energy reduction goals. This is largely due to insufficient information exchange regarding the system's flexibility potential and expected energy consumption patterns between BMS and grid operators.

The increased penetration of distributed energy sources as well as the electrification of smart building systems enable more advanced DSM strategies to better harness the energy flexibility of energy consumers by establishing privacy-secured information sharing among energy consumers and producers. In order to cater to diverse household devices and different markets and contractual constraints, the European Commission's M/490 mandate has introduced a future-proof energy flexibility interface (EFI) – the S2 standard – for leveraging energy flexibility in the built environment [96, 97]. Within the S2 standard, a bi-directional communication flow between the customer energy manager and the resource manager is developed to facilitate the exchange of information on flexibility capabilities and deployment, which paves the way to a versatile and privacy-secured energy flexibility exploitation architecture.

While there are some recent approaches compatible with the S2 protocol for better exploiting the energy flexibility of HVAC in buildings (e.g., [55, 95, 98]), these results can hardly be generalized to ON-OFF type HPs with nonlinear system dynamics, which is the case investigated in this chapter. Another limitation hindering the practical application of MPC in HPTES systems is the lack of general modeling approaches to develop control-oriented models for typical HPTES systems with balanced model complexity and approximation accuracy. Only specific HPTES configurations with tailored control-oriented models are usually considered in most of the existing literature.

Motivated by the above discussion, this chapter presents a comprehensive approach of MPC design, including system modeling, control algorithm design, numerical simulation and real-world experiments, for HPTES systems to harness the energy flexibility of HPTES systems to mitigate grid congestion issues without degrading system performance. Our key contributions are as follows.

- A practical energy-flexible MPC design is proposed for the HPTES system. Notably, based on classical economic MPC formulations, a two-step DSM scheme is introduced to better harness the energy flexibility of the system. By imposing a set of mixed-integer linear constraints, our proposed approach can quantitatively eval-

uate the energy flexibility of the HPTES system regardless of the explicit system model. The energy flexibility can also be safely exploited via DR request mechanisms.

- Both numerical simulation and real experiments are conducted based on a real-world HPTES installation to validate the viability and effectiveness of the proposed approach. Practical solutions for improving the computational efficiency of MPC are discussed. The results show that adopting the proposed MPC schemes not only leads to reduced energy cost, but also enables a reliable assessment and exploitation of the energy flexibility of the HPTES system while respecting system constraints.

The remaining parts of this chapter are organized as follows. Section 3.2 briefly introduces the control-oriented models of the HPTES system. Section 3.3 discussed the MPC design framework for both the economic operation and DSM of the HPTES system. Numerical and experimental results are detailed in Section 3.4. Finally, Section 3.5 concludes the chapter.

3.2. Control-Oriented Modeling for HPTES System and Problem Formulation

As stated in the introduction, an MPC-based DSM scheme will be designed for the HPTES system, where a prediction model for the HPTES system is required. In this section, a general control-oriented model of HPTES systems will be introduced, accompanied with the system constraints.

3.2.1. Modeling for HPTES System

Based on a real HPTES installation in an office building in the Netherlands, Fig. 3.1 depicts a general configuration of an HPTES system, where red lines denote hot water circulation, and blue lines denote cold water circulation with arrows indicating the direction of water flows. In this system, hot water from the top of Tank 1 is circulated at a mass flow rate $\dot{m}_c$ through connected taps to meet hot water usage demands. The real-time hot water consumption, denoted by $\dot{m}_s$, is assumed to be predictable with sufficient accuracy based on historical data. Whenever hot water is consumed, an equal amount of cold water $\dot{m}_s$ will be simultaneously supplied to the bottom of Tank 2, ensuring that the total water volume within the HPTES system remains constant. If the HP is off, the cold water is directly injected into the bottom of Tank 2. Conversely, when the HP is on, water from the bottom of Tank 2 with a mass flow rate $\dot{m}_p$ will be transmitted to the top of Tank 1 after being heated through the heat exchanger.

Such an HPTES system has two major components: heat pump and thermal energy tanks, which are important in designing an MPC-based DSM scheme. In the following subsections, we will provide general control-oriented models for the HP and the thermal tank, respectively, which will be utilized for our MPC design in Section 3.3.

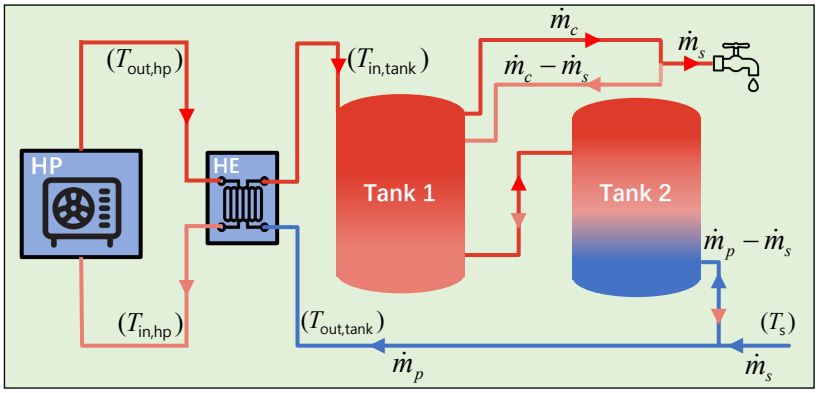


Figure 3.1: Diagram of the considered general HPTES system.

MODELING FOR HEAT PUMP

HPs are responsible for providing thermal energy to the HPTES system by consuming electricity. Accurately modeling the relationship between electricity consumption and thermal energy generation is crucial for designing control strategies to improve the system's energy efficiency and safety. However, a detailed and precise white-box HP model can be complicated and also requires an extensive understanding of the interval working principles of each component of HPs, which demands expert knowledge and will also make the control problem computationally intractable [99]. To balance the system complexity and computational burden, we adopt a grey-box model – coefficient of performance – to build the relationship between the thermal energy output and the electricity consumption of HPs.

The coefficient of performance (COP) is defined as the ratio of heat output to work input [100]. Mathematically, it is expressed as:

$$\text{COP} = \frac{Q_{\text{hp}}}{P_{\text{hp}}}, \tag{3.1}$$

where Q_{hp} denotes the heat output of the HP and P_{hp} its power input. Since the COP generally varies with different operation conditions in a nonlinear relationship, it is approximated with a function $f_{\text{COP}}(\cdot)$ in MPC design. Besides, the HP in our HPTES system is assumed to operate in two modes: on and off, which is common in many practical installations, so that the HP power input is $P_{\text{hp}} = P_r\, u$ with P_r as the rated HP power and $u \in \{0, 1\}$ the on/off HP control signal. Consequently, the thermal energy generated by the HP can be represented as:

$$Q_{\text{hp}} = f_{\text{COP}}(\cdot) P_r\, u, \tag{3.2}$$

where $f_{\text{COP}}(\cdot)$ is an function approximating the COP values under different operation conditions.

The COP is influenced by HP's working conditions, and two main factors are the supply water temperature (condenser outlet) and the ambient heat source temperature [89].

To balance the trade-off between approximation accuracy of the model and the computational complexity of the resulting MPC problem, the following bilinear COP model is utilized

$$f_{\mathrm{COP}}(T_{\mathrm{in,hp}}, T_{\mathrm{amb}}) := a_1 + a_2 T_{\mathrm{in,hp}} + a_3 T_{\mathrm{amb}} + a_4 T_{\mathrm{in,hp}} T_{\mathrm{amb}}, \tag{3.3}$$

where $T_{\mathrm{in,hp}}$ is the inlet water temperature of HP, T_{amb} is the ambient air temperature of HP, and (a_1, a_2, a_3, a_4) are parameters to be identified. Similar COP approximation can also be found in [89, 101].

Compared with the constant COP model, which is widely utilized in existing literature [102], the bilinear model in (3.3) captures the major factors of COP – the heat source (ambient air) and the heat consumer (inlet water to be heated) – that influence the HP performance to provide a more practical and accurate model without demanding high computational burden.

Modeling for Thermal Energy Storage

For the HPTES system, thermal tanks can improve the system's energy efficiency and flexibility by storing hot water. By buffering hot water, the HP can operate more flexibly regardless of the real-time hot water consumption. There are various approaches for modeling the dynamics of thermal storage tanks, with different emphases on model accuracy and simplicity. A three-dimensional model, which is generally based on computational fluid dynamics to capture the detailed internal thermal behaviors, can achieve high approximation accuracy but is computationally expensive [92]. By using a fully mixed tank model or lumped parameter model, the entire water tank is simplified as a single thermal node without considering any internal dynamics [55]. While this mixed tank model is computationally simple, the prediction accuracy is unsatisfactory, especially for large water tanks where internal dynamics cannot be neglected. To balance model fidelity with computational efficiency, one-dimensional models are developed to only consider the vertical internal thermal interactions within the water tank [103, 104]. In our work, a one-dimensional stratified water tank model is adopted to approximate

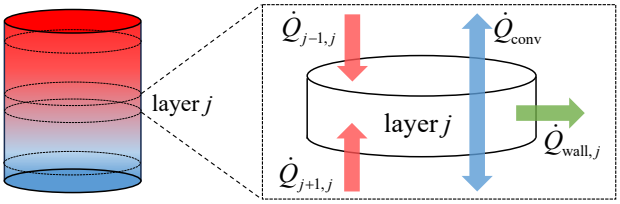


Figure 3.2: Heat flow diagram of a single layer of TES.

the thermal dynamics of water tanks. We divide each water tank into several stratified layers with each layer being modeled as a single thermal node with a uniform temperature. For adjacent water layers, the thermal interactions are considered. The general thermal behavior of each layer is illustrated in Fig. 3.2. For each thermal layer j, its thermal dynamics are driven by three components:

- $\dot{Q}_{j+1,j}$ and $\dot{Q}_{j-1,j}$: thermal conduction between layer j and its adjacent layers $j+1$ and $j-1$.

- $\dot{Q}_{\text{conv}}$: thermal convection due to water flows $\dot{m}_t$ in/out of j-th layer from/to adjacent layers.
- $\dot{Q}_{\text{wall},j}$: thermal loss through the tank wall in layer j.

The thermal dynamics of each thermal node can be expressed by the following equations:

$$m_j c_p \frac{dT_j}{dt} = -\dot{Q}_{\text{wall},j} + \dot{Q}_{j+1,j} + \dot{Q}_{j-1,j} + \dot{Q}_{\text{conv}}, \tag{3.4a}$$

$$\dot{Q}_{\text{conv}} = \begin{cases} \dot{m}_t c_p (T_{j-1} - T_j), & \dot{m}_t \text{ flows from } j-1 \text{ to } j \\ \dot{m}_t c_p (T_{j+1} - T_j), & \dot{m}_t \text{ flows from } j+1 \text{ to } j \end{cases} \tag{3.4b}$$

$$\dot{Q}_{j+1,j} = R_{j,j+1}\left(T_{j+1} - T_j\right), \tag{3.4c}$$

$$\dot{Q}_{j-1,j} = R_{j-1,j}\left(T_{j-1} - T_j\right), \tag{3.4d}$$

$$\dot{Q}_{\text{wall},j} = R_{w,j}\left(T_j - T_{\text{amb}}\right), \tag{3.4e}$$

where m_j is the mass of the j-th node, c_p is the specific heat capacity of water, $\dot{m}_t$ is the water mass flow rate between adjacent tank layers, $R_{j,j+1}$ and $R_{w,j}$ are thermal resistances between the j-th layer and the $(j+1)$-th layer as well as the tank wall, respectively.

Equation (3.4) describes the general thermal behavior of a water layer. Special consideration is needed for layers with hot water extraction, i.e., the top layer of Tank 1, and cold water supply, i.e., the bottom layer of Tank 2. For the top layer of Tank 1, which does not have an upper layer, the thermal conduction term $\dot{Q}_{j-1,j}$ is zero, and the thermal convection from the upper layer can be computed as $\dot{Q}_{\text{conv}} = \dot{m}_t c_p (T_{j-1} - T_j)$ with T_{j-1} as the temperature of the inlet hot water entering Tank 1 from the heat exchanger, and the water mass flow rate $\dot{m}_t$ as the water circulation mass flow rate $\dot{m}_p$. A similar analysis applies to the bottom layer of Tank 2, where no lower layer exists but with a cold water supplement. In addition, for our HPTES system, since the direction of water circulation changes with HP operation status (indicated by the control signal u), the internal mass flow rate of water convection $\dot{m}_t$ is a function of the HP control signal u as well as the mass flow rate of hot water consumption $\dot{m}_s$.

Furthermore, the water pumps driving the water flow of cold water supply $\dot{m}_s$ and water circulation within tanks $\dot{m}_p$ in (3.4) are assumed uncontrollable in our design, which is the case for many practical HPTES installations. It is worth pointing out that the control-oriented model (3.4) can be extended to consider these variables as controllable terms, which can then be optimized in the MPC design of Section 3.3, without changing the model structures. This can be expected to improve system efficiency, but would also increase the computational cost for solving the corresponding MPC problem by inducing extra nonconvex terms. Consequently, this choice should be made based on the hardware availability and the trade-off between expected control performance and computational cost.

Remark 3.1. The HPTES configuration along with the control-oriented model introduced in this section is sufficiently general to cover a wide range of HPTES systems and is flexible enough to be adjusted to cater to other alternative configurations with only minor modifications. For instance, the positions of the hot water extraction and the cold

water supplement, as well as the internal water circulation directions can be flexibly adjusted without breaking the structure of the control-oriented model. Furthermore, this model is computationally tractable to be incorporated into MPC design as a prediction model.

3.2.2. System Constraints

To ensure operational efficiency and safety, the HPTES system has to respect several physical constraints. In this subsection, several typical constraints will be introduced for the HPTES system.

Firstly, the water temperature in the water tanks should be maintained within an admissible range to ensure safety and to provide qualified hot water. The water temperature constraint can be reformulated as the following linear constraint

$$x^l \leq x \leq x^u, \tag{3.5}$$

where x denotes the temperature vector of different water layers and pipelines, x^l and x^u are the corresponding temperature lower bound vector and upper bound vector, respectively. It should be highlighted that, among all water layers within the water tanks, the water temperature at the top layer of Tank 1 is of more importance since it is more related to the water to be supplied to the buildings.

Furthermore, frequent on-off switching of the HP should be prevented as it can lead to excessive wear and tear, reducing the HP's operational lifespan. Consequently, the number of on-off switches within a time period should not exceed a threshold. The constraints for the number of switches can be formulated as the following quadratic constraint

$$\sum_{k=1}^{m} (u_{t-k+1} - u_{t-k})^2 \leq N_{\text{ctrl}}, \tag{3.6}$$

where N_{ctrl} is the maximal number of switches that is allowed during m time steps, and u_{t-k} denotes the control input at the time instant $t-k$ (i.e., k time steps prior to the current time instant t). Since $u_t \in \mathbb{B}$, the left-hand side of (3.6) computes the number of switches that occur during the most recent m time steps. For instance, setting $m = 8$ and $N_{\text{ctrl}} = 1$ implies that, only one switch is allowed within an 8-step time window.

3.3. MPC Design for Economic Operation and Demand-Side Management

In this section, based on the two-step DSM protocol, which is introduced in [95, 98] aligning with the S2 standard introduced by the European Commission for exploiting energy flexibility in the built environment [96], an energy-flexible MPC design framework is proposed for the HPTES system to achieve economical operation and DSM. In the S2 protocal, eight types of distinct flexibility patterns are defined. Our work mainly focuses on "pause a task" for mitigating the grid congestion issue.

Fig. 3.3 provides a comprehensive workflow for our proposed MPC framework to achieve economic operation and DSM with HPTES systems, which consists of two main strategies: 1) economic MPC, and 2) DSM-based MPC. When no DR requests are received to activate DSM, the HPTES system is operated with the economic MPC strategy

for reducing the energy cost. Once the DR requests are detected, the DSM-based MPC strategy will be deployed to provide energy flexibility services. Once the DR request is finished and no further DR requests are in the queue, the control scheme will be switched back to the economic MPC strategy.

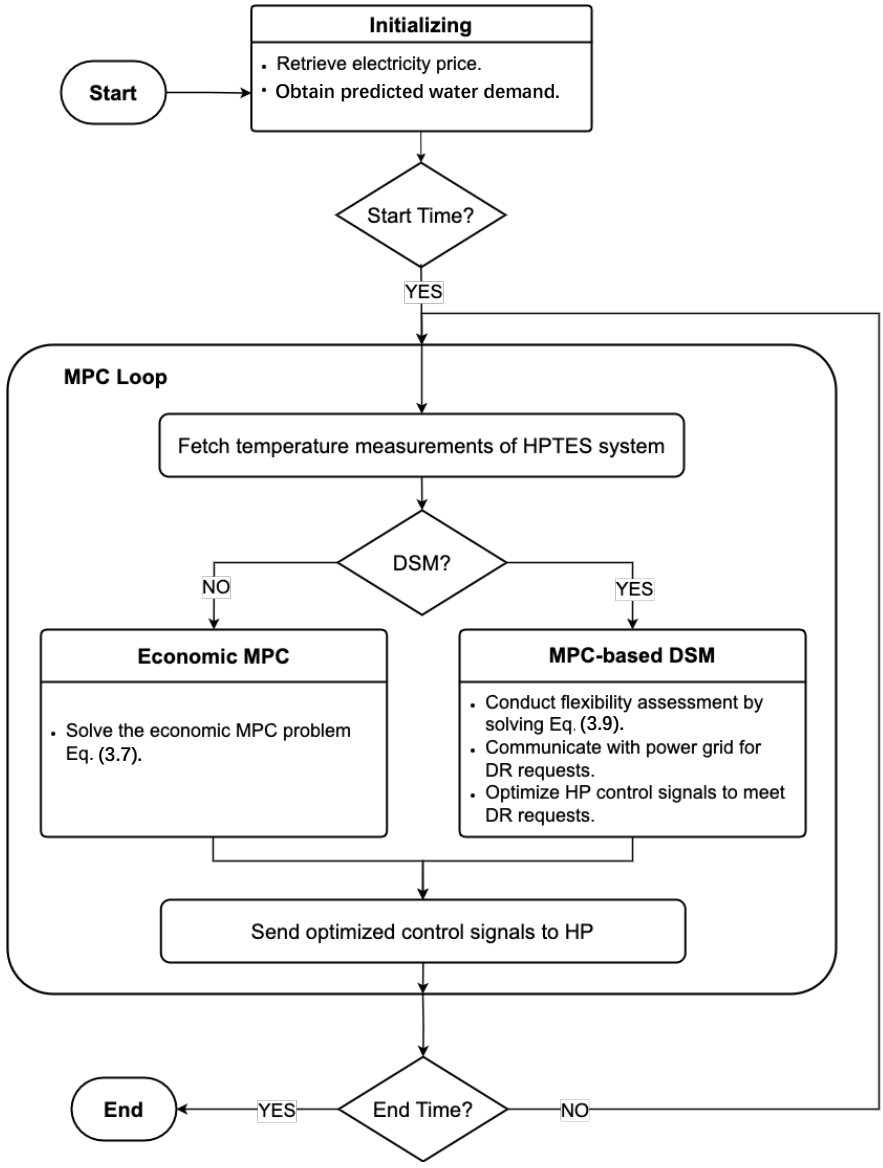


Figure 3.3: Workflow for implementing the proposed MPC schemes.

For the HPTES system, the economic MPC formulation aiming at reducing the en-

ergy cost of the HPTES system is expressed as the following optimization problem

$$\min_{u_t} J_o \tag{3.7a}$$
$$\text{s.t. } x_{t+1} = f(x_t, u_t, d_t), \tag{3.7b}$$
$$(x_t, u_t) \text{ satisfy (3.5) \& (3.6)}, \tag{3.7c}$$
$$\text{for } t = 0, 1, \cdots, N \tag{3.7d}$$

where $J_o := \sum_{t=0}^{N} l(x_t, u_t)$ is the total operational cost function with $l(x_t, u_t)$ as the stage cost and N the length of the prediction horizon, x_t denotes the system states composed of the temperature vector of different water layers and pipelines, $u_t \in \mathbb{B}$ is the HP binary control signal with $u_t = 1$ indicating on status, $f(x_t, u_t, d_t)$ is the control-oriented model introduced in (3.2) and (3.4). In our design settings, the operational cost for the HPTES system mainly comes from electricity consumption. Consequently, the operational cost function is defined as $J_o := \sum_{t=0}^{N} e_t \cdot P_r \cdot u_t$, where e_t is the electricity price signal and P_r is the rated power of HP.

Based on the above economic MPC formulation, in the following, we will show how to further enable energy use flexibility in a DSM scheme. Current mainstream DSM design follows the so-called price-based or incentive-based programs, which are one-step programs and can be inefficient in exploiting the energy flexibility of the system. Instead of following the conventional design framework, our work considers a novel DSM framework based on a bidirectional communication protocol between the building management system (BMS) and the grid operator to achieve better utilization of energy flexibility. Fig. 3.4 shows the diagram of the DSM framework, which consists of two steps. At *Step 1*, called *Flexibility Assessment*, the energy flexibility potential of the HPTES system is quantitatively assessed, and the information of the flexibility potential $\mathscr{F}$ is shared with the power grid. At *Step 2*, called *Flexibility Exploitation*, based on the flexibility potential $\mathscr{F}$, the grid operator will generate a feasible DR request $\mathscr{D}$ to exploit the energy flexibility.

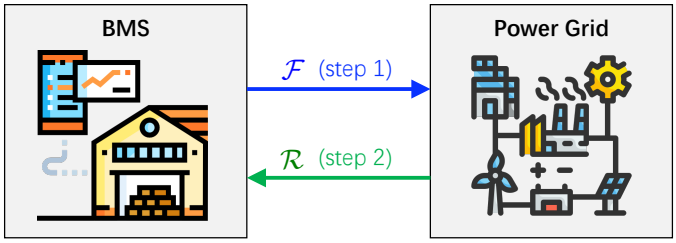


Figure 3.4: Diagram of the two-step DSM scheme.

3.3.1. Flexibility Assessment

For this two-step design framework, the main difficulty lies in the first step: quantitatively characterizing and computing the flexibility potential of the system. Our work mainly focuses on solving grid congestion, which is a challenging issue for the power

grid in the Netherlands. Hence, the DSM design mainly aims at reducing energy consumption during critical hours.

While there is no standard definition of energy flexibility, the flexibility potential of a system can be measured mainly based on three factors that describe the shift in energy use: magnitude, time duration, and cost [105, 106]. In this work, we focus on solving the grid congestion issue by adopting the flexibility pattern "pause a task", i.e., turning off the HP. In our problem setup, the HP can only operate in two modes: on and off, which implies the magnitude of the flexibility reduction at each time instant is constant, the capacity of energy flexibility is thus proportional to the time duration that HP remains off. Consequently, we measure the flexibility capacity of the HPTES system as the time duration during which the HP can remain off without violating system constraints.

Here we denote $\mathcal{T} := \{t_{f_1}, \cdots, t_{f_n}\}$, called flexibility assessment period, as the set of consecutive time indices during which the flexibility of the HPTES system is assessed. Then, given a flexibility period, the time instants within $\mathcal{T}$ can be classified into three categories: the time instant before the flexibility period, at which the corresponding HP control input can be freely adjusted; the time instant within the flexibility period, during which the HP is off; and the time instant after the flexibility period, during which the HP control signal is free to be adjusted. To express the above logic and perform flexibility assessment, two sets of auxiliary decision variables are introduced: $\mathcal{S}_\mathcal{T} := \{s_t, t \in \mathcal{T}\}$ and $\mathcal{Z}_\mathcal{T} := \{z_t, t \in \mathcal{T}\}$. where $s_t \in \mathbb{B}$ and $z_t \in \mathbb{B}$ are binary decision variables. Based on s_t and z_t, the following logic follows

- $s_t = 0$ and $z_t = 0$: the time instant t is before the flexibility period, and $u_t \in \mathbb{B}$.
- $s_t = 1$ and $z_t = 0$: the time instant t is within the flexibility period, and $u_t = 0$.
- $s_t = 0$ and $z_t = 1$: the time instant t is after the flexibility period, and $u_t \in \mathbb{B}$.

The above logic is illustrated by a diagram in Fig. 3.5, and can be expressed as the following mixed-integer linear constraints

$$u_t \le 1 - s_t, \tag{3.8a}$$
$$s_{t+1} \ge s_t - z_{t+1}, \tag{3.8b}$$
$$s_t + z_t \le 1, \tag{3.8c}$$
$$z_{t+1} \ge z_t. \tag{3.8d}$$

Based on the definition of flexibility, the flexibility capacity is proportional to $\sum_{t\in\mathcal{T}} s_t$. Namely, the longer the period that the HP can remain off without violating system constraints, the larger the flexibility capacity. Combining the flexibility assessment with the economic MPC problem (3.7) leads to the following flexibility assessment formulation

$$\min_{u_t, s_k, z_k} \; J_o - J_f \tag{3.9a}$$
$$\text{s.t. } x_{t+1} = f(x_t, u_t, d_t), \tag{3.9b}$$
$$x_t \text{ satisfy (3.5)}, \; u_t \text{ satisfy (3.6)}, \tag{3.9c}$$
$$\text{for } t = 0, 1, 2, \cdots, N-1, \tag{3.9d}$$
$$(u_k, s_k, z_k) \text{ satisfy (3.8)}, \forall k \in \mathcal{T} \tag{3.9e}$$

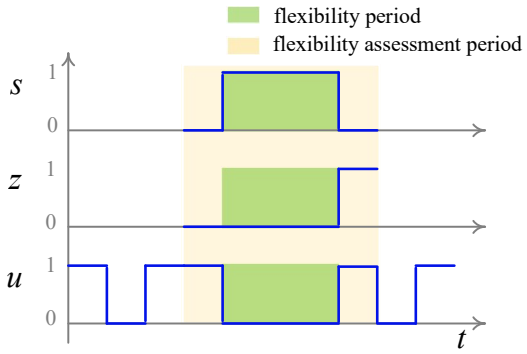


Figure 3.5: Schematic of the logic in (3.8) indicating flexibility periods.

where $J_f := \lambda \sum_{t \in \mathscr{T}} s_t$ is the revenue of exploiting the energy flexibility for DSM services, and $\lambda > 0$ is a weighting parameter. By solving this optimization problem, the optimal flexibility period, which is defined as $\mathscr{T} := \{t | s_t = 1\}$, can be computed. This information will be subsequently used in the *Flexibility Exploitation* stage. The optimization problem (3.9) computes the optimal flexibility commitment by balancing the profits of flexibility provision J_f and the incurred operational cost J_o. Once J_o is omitted, the optimal solution of this optimization problem quantifies the largest flexibility capacity of the system regardless of the incurred increase of operational cost. It should be pointed out that in this work the revenue of flexibility J_f is defined to be proportional to the duration of flexibility, in order to demonstrate the effectiveness of the proposed two-step DSM framework. The objective function in (3.9a) can be modified to consider different revenue formulations depending on different market choices and contractual options, such as maximal demand contract and interruptible load contract.

3.3.2. FLEXIBILITY EXPLOITATION

After finishing the flexibility assessment, the information of the flexibility capacity $\mathscr{T} := \{t | s_t = 1\}$ will be shared with the power grid. Based on $\mathscr{T}$, a feasible DR request $\mathscr{R} = \{t | t \in \mathscr{T}\} \subseteq \mathscr{T}$ will be generated by the grid operator to specify the time instants that the HP needs to be off to provide DSM services. Namely,

$$u_t = 0, \quad \forall t \in \mathscr{R}. \tag{3.10}$$

To achieve this DR request, a classical economic MPC problem with extra constraints (3.10) can be formulated to compute the HP control signals. Since the DR request $\mathscr{D}$ is feasible w.r.t. the flexibility capacity $\mathscr{T}$, there is always an admissible control input sequence, i.e., a solution of the optimization problem (3.9), to meet the DR request without violating system constraints.

Remark 3.2. By solving (3.9), the optimal flexibility period, which is defined as $\mathscr{T} := \{t | s_t = 1\}$, will be computed. If the operational cost J_o is omitted from the objective function, the largest flexibility capacity of the system, i.e., the longest period during which the HP can remain off without violating constraints, is obtained. It should be highlighted

(a) Air-to-water HP.

(b) Two 500L water tanks.

Figure 3.6: The HPTES system used in our numerical and experimental case studies.

that the formulation of flexibility assessment in (3.9) is independent of the system dynamics as well as the economic MPC formulation, which means that the proposed DSM design framework is feasible to any economic MPC design by only considering extra linear constraints (3.8). Hence, this design framework can be easily incorporated into other MPC formulations without inducing high design complexity and computational burden.

3.4. NUMERICAL AND EXPERIMENTAL RESULTS

In this section, the proposed MPC-based DSM strategy will be numerically and experimentally tested. The investigated HPTES system is based on a real installation, as shown in Fig. 3.6, that is designed to provide domestic hot water for fitness rooms and dining rooms in an office building in the Netherlands. This HPTES system consists of two 500-liter water tanks and an air-to-water heat pump. The HP is only operated during working hours from 7:00 to 17:30 on weekdays.

Since the main emphasis of this work is on the development of MPC methodologies, we omit detailed system-modeling results, including the models of HP's COP, water tanks, and hot water prediction. The modeling procedures, results, and parameter values adopted in our simulations and experiments are provided in [107, 108].

The numerical and experimental results are presented in two subsections including the simulation results of the proposed MPC schemes, and real-world experiments of the proposed MPC schemes. All MPC problems for simulations and experiments are modeled using the Python package `Pyomo` [109] and are solved using Gurobi 10.0.1 [74] on a 2.3 GHz 8-Core Intel Core i9 CPU with 32 GB 2667 MHz DDR4 RAM.

3.4.1. SIMULATION AND EXPERIMENT SETTINGS AND MPC CONFIGURATIONS

Fig. 3.7 depicts the diagram for implementing our simulation and experiment of the proposed MPC strategies. At the beginning of the simulation or implementation, electricity price signals and hot water consumption predictions are updated. Then, the MPC solution is computed every 5 minutes. At each sampling instant, the real water tank temperature and water flow rates are used to update the control-oriented model. For each MPC

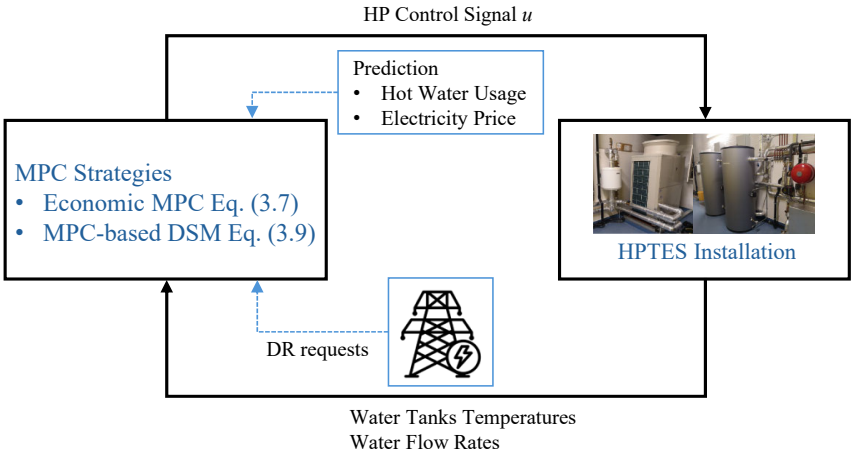


Figure 3.7: Simulation and experiment diagram.

update, if a new DSM request is detected and no existing DR request is being served, the DSM scheme will be activated to meet the requirement of the power grid. Otherwise, the economic MPC scheme will control the HP to optimize system operation.

In our system, the temperature of the hot water supplied to the building is required to be within 55°C - 75°C to mitigate the risk of legionella bacteria and to ensure system safety. Besides, it is preferred to keep the temperature above 60°C for most of the operational time to provide qualified hot water. For the real-time implementation of MPC, it is essential to ensure that the MPC problems in (3.7) and (3.9) are always feasible. Due to the existence of uncertainties, the hard constraints (3.5) and (3.6) might result in infeasible optimization problems. To address this issue, we alternatively consider the following soft constraints

$$55 - \delta_1 \leq T_1 \leq 75 + \delta_1 \tag{3.11a}$$

$$T_1 \geq 60 - \delta_2 \tag{3.11b}$$

where $\delta_1 > 0$ and $\delta_2 > 0$ are slack variables to relax the hard constraints and are penalized in the objective function J_o. The operational cost function J_o then becomes

$$J_o := \sum_{t=1}^{N} e_t P_r u_t + M_1 \delta_1 + M_2 \delta_2 \tag{3.12}$$

where e_t is the electricity price, $M_1 \geq 0$ and $M_2 \geq 0$ are large constants to penalize the slack variables (δ_1, δ_2). The electricity price used in our simulation is depicted in Fig. 3.8, which is a typical electricity price pattern in the Dutch electricity market.

For our proposed MPC strategies, the prediction horizon is selected as 6 hours for the economic MPC scheme (3.7), and 4 hours for flexibility assessment (3.9). This choice is adopted to balance the computational costs and control performance, considering the fact that the HP is operated for about 10 hours per day. In addition, practical DR request

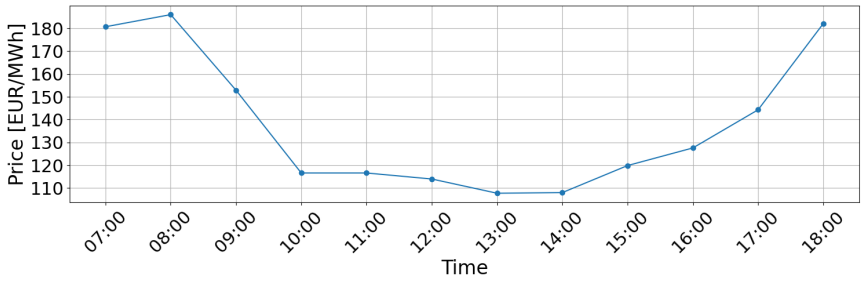


Figure 3.8: Electricity price used in tests.

periods for solving peak hour grid congestion last generally for less than an hour, and the real-time control update period of our MPC schemes is 5 minutes.

One practical challenge in implementing MPC is ensuring that the corresponding optimization problems can be solved efficiently for real-time implementation. Good control performance of MPC generally requires considering a relatively long prediction horizon, which will increase the computational burden. To balance the computational efficiency and control performance, the *move blocking strategy* is utilized in our MPC implementation. Unlike the conventional MPC design, where fixed time steps are applied within the whole prediction horizon, the *moving blocking strategy* involves segmenting the prediction horizon into time windows with different lengths of time steps [110]. For the blocks that are closer to the current time instant, the corresponding time step is smaller. This strategy is based on the premise that the necessity for precise prediction and control diminishes for further predictions in the future since only the first MPC input signal is applied and the control signal at the next time step is recomputed with updated information.

Fig. 3.9 gives a graphical illustration of the standard MPC and *move blocking* MPC. It can be seen that the standard MPC keeps a fixed time step within the prediction horizon. In contrast, *move blocking* MPC has varying-sized time steps. With the same prediction horizon length, *moving blocking* MPC entails fewer decision variables than the standard MPC, and hence can consider a longer prediction horizon with reduced computational burden for solving the corresponding MPC problem. As shown in Fig. 3.9, in our simulations and experiments, the prediction horizon for the economic MPC formulation (3.7) is 6 hours with 3 blocks. The first block has 6 time steps with each time step 20 mins. The second and the third blocks have 4 and 3 steps with corresponding time steps as 30 mins and 40 mins, respectively. As shown in Table 3.1, with the move-blocking strategy, the average computational time for solving the economic MPC problem (3.7) over a simulation day is reduced from 72.96s to 25.77s, highlighting the improved computational efficiency.

3.4.2. SIMULATION RESULTS OF MPC SCHEMES

SIMULATION RESULTS OF MPC FOR ECONOMIC OPERATION

For the proposed MPC strategies, the performance of the economic MPC formulation (3.7) without considering DSM is tested first. To show the efficacy of the proposed MPC

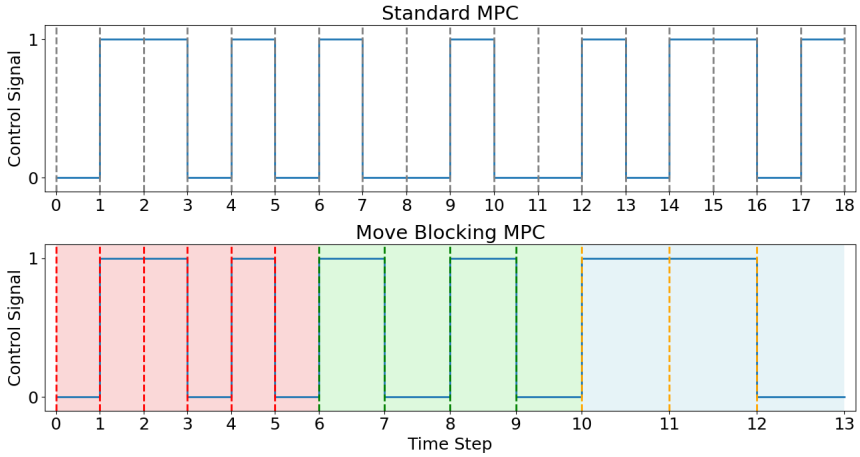


Figure 3.9: Standard MPC and move blocking MPC in a 6-hour prediction horizon.

Table 3.1: Computation time of standard MPC and move blocking MPC.

	Standard MPC	Move Blocking MPC
Minimal computation time (s)	11.74	10.81
Maximal computation time (s)	301.19	56.58
Average computation time (s)	72.96	25.77

approach, the results of the *Rule-Based Approach* that is currently used to operate the real HPTES system are also presented. With the *Rule-Based Approach*, the HP is turned on if the water temperature at the top layer of Tank 1 is lower than 62°C and is turned off until the water temperature at the bottom layer of Tank 2 is higher than 62°C.

Simulation results are shown in Fig. 3.10 and are summarized in Table 3.2. From Fig. 3.10, it is clear that the MPC approach can maintain the supply water temperature within the admissible range [55°C, 75°C]. It can be seen from Table 3.2 that compared with the rule-based approach, the energy cost of MPC is reduced by about 14% and the energy consumption by 13% while also leading to less constraint violations (defined as the violation of the desired lower bound of water temperature 60°C).

Table 3.2: Simulation results of MPC and rule-based schemes for Economic Operation.

	Rule-Based Approach	MPC
Average Water Temperature (°C)	65.53	63.11
Maximal Constraint Violation (°C)	4.43	1.94
Energy Consumption (kWh)	64.17 (100%)	55.83 (87.00%)
Energy Cost (Euro)	8.69 (100%)	7.45 (85.73%)

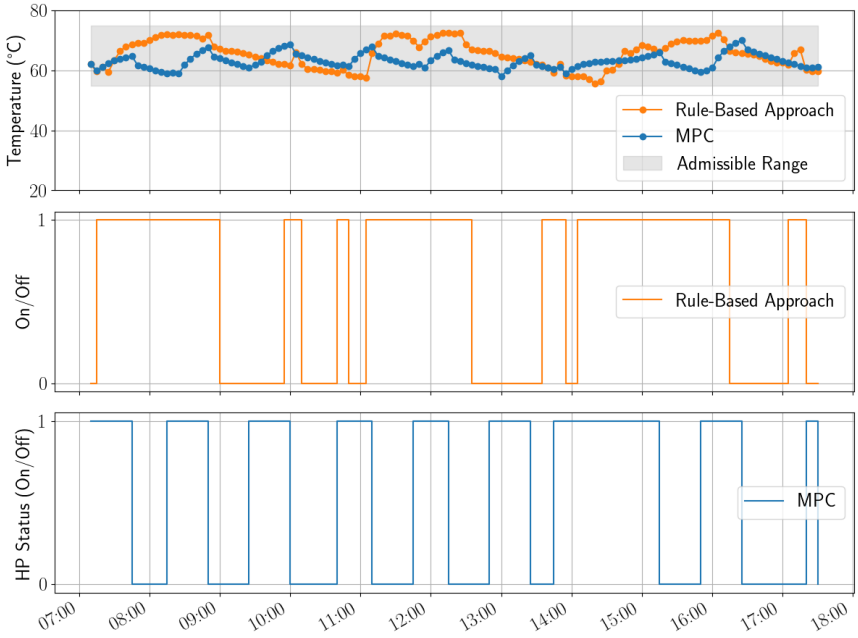


Figure 3.10: Simulation results for Economic Operation: (top) water temperature at the top layer of Tank 1, (middle) HP status with rule-based scheme, (bottom) HP status with MPC.

Table 3.3: Simulation results of MPC and rule-based schemes for DSM.

	Rule-Based Approach	MPC
Average Water Temperature (°C)	65.53	64.85
Maximal Constraint Violation (°C)	4.43	4.12
Energy Consumption (kWh)	64.17 (100%)	62.50 (97.40%)
Energy Cost (Euro)	8.69 (100%)	8.41 (96.78%)

SIMULATION RESULTS OF MPC FOR DSM

Based on the economic MPC formulation, we augment the design to further consider DSM. The data for electricity price and water consumption used are the same as for the economic MPC implementation. For the flexibility assessment problem (3.9), the prediction horizon is selected as 4 hours, and the first 3 hours are defined as the flexibility assessment period. This time interval is chosen to ensure that the flexibility potential of the system can be fully explored while leaving sufficient time to prepare for the next flexibility task. In our simulation and experiments, synthetic DR requests are generated from the power grid every 3 hours so that the flexibility assessment and exploitation tasks are performed (3 tasks in total per day). The objective function for (3.9) is selected as $J_f = \sum_{t\in\mathcal{T}} s_t$ without considering the operational cost J_o, which implies that the largest flexibility potential of the HPTES system, i.e., the longest time period that the HP can be off without violating system constraints, is computed. For flexibility exploitation, virtual

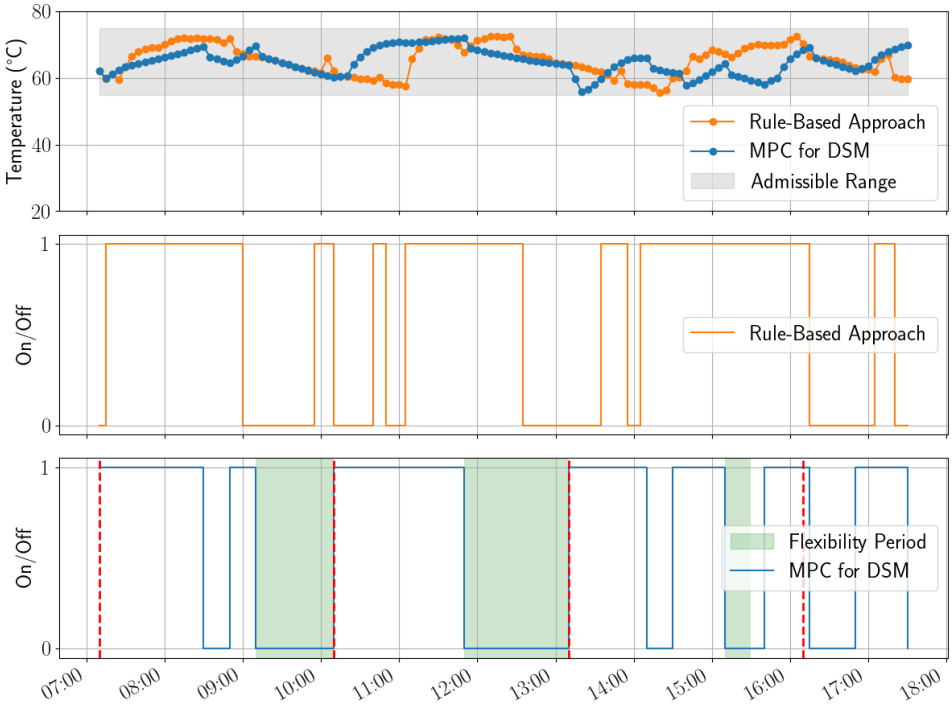


Figure 3.11: Simulation results for DSM: (top) water temperature at the top layer of Tank 1, (middle) HP status with rule-based scheme, (bottom) HP status with MPC.

DR requests are set as $\mathscr{D} = \mathscr{F}$, which request the largest possible energy reduction during the flexibility period. This extreme scenario is selected for testing the robustness of our proposed approach.

Simulation results for DSM are shown in Fig. 3.11 and Table 3.3. In Fig. 3.11, the starting and ending times of flexibility assessment are indicated via dashed red lines, and the flexibility periods as green shaded areas. Within all flexibility periods, the HP is required to remain off to meet the DR requests. It is clear that all DR requests are satisfied while respecting system constraints. Furthermore, as shown in Table 3.3, compared with the rule-based control approach, which is incapable of implementing DSM with safety guarantee, our proposed MPC approach can not only achieve DSM but also lead to less energy consumption. This result highlights the benefits of the proposed MPC-based approach in cost saving, even with the provision of DSM services, which will force the operation of HP to deviate from its economically optimal pattern, and sacrifice economic operational performance.

3.4.3. Experimental Results of MPC Schemes

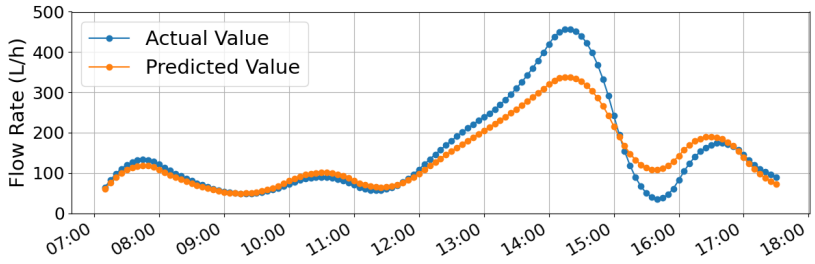


Figure 3.12: Hot water consumption and its prediction in the experiment for economic operation (MAE of prediction: 28.32 L/h).

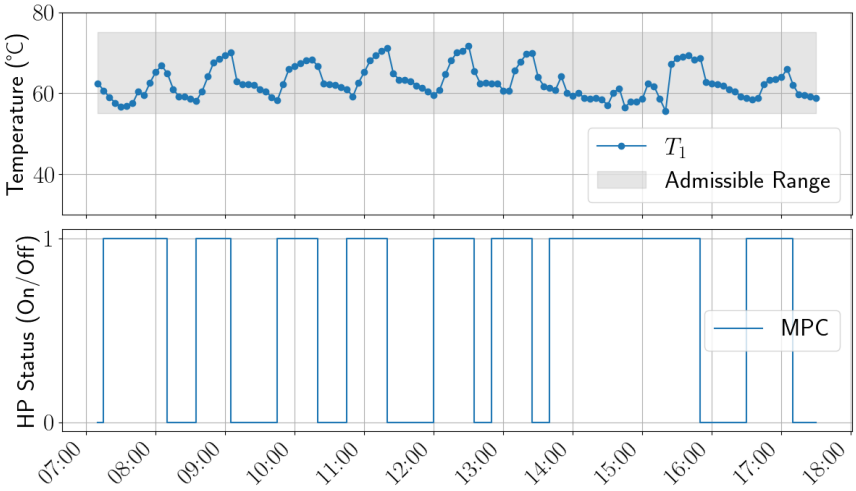


Figure 3.13: Experimental results of MPC for economic operation: (top) water temperature at the top layer of Tank 1; (bottom) HP status.

EXPERIMENTAL RESULTS OF MPC FOR ECONOMIC OPERATION

In order to validate the practical viability of our proposed schemes, experiments are conducted in a real-world HPTES installation as shown in Fig. 3.6. The real hot water consumption and its predictions during the experiment are shown in Fig. 3.12, from which we can conclude that the prediction reflects the real usage reasonably well but with non-negligible prediction error, especially during peak hours. Fig. 3.13 depicts the water temperature of the supplied hot water and the HP control signals. It is clear that the proposed MPC approach successfully maintains the water temperature within the admissible range regardless of the non-negligible prediction error of water usage. In addition, based on our computation, the MPC approach results in an operational cost of 8.64 euros, which is lower than the 9.63 euros achieved by the existing rule-based approach with similar hot water consumption profile.

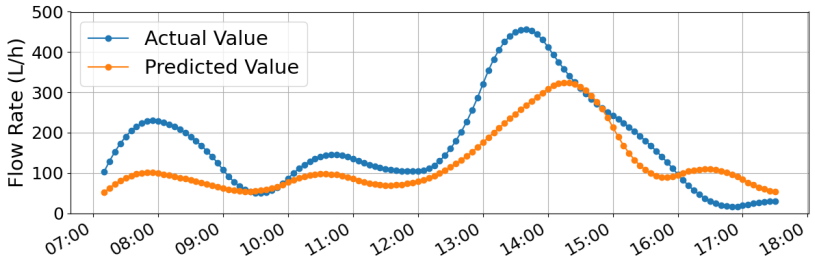


Figure 3.14: Hot water consumption and its prediction in the experiment for DSM (MAE of prediction: 64.93 L/h).

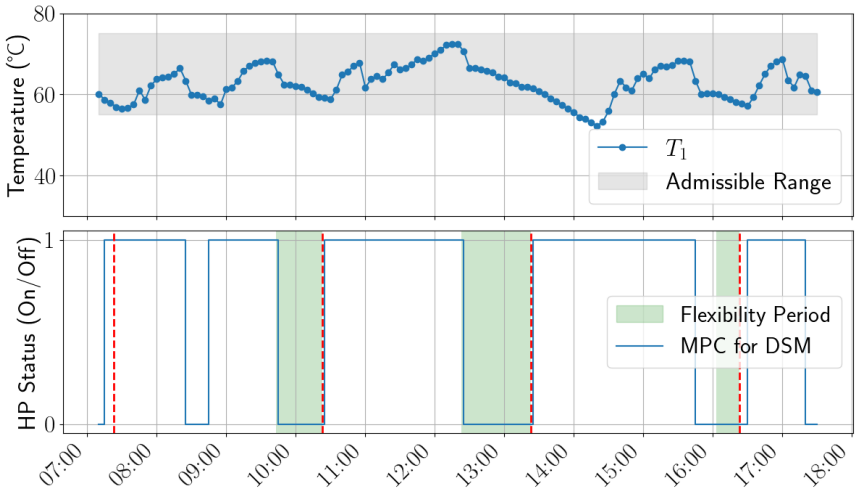


Figure 3.15: Experimental results of MPC for DSM: (top) water temperature at the top layer of Tank 1; (bottom) HP status.

Experimental Results of MPC for DSM

This section presents the experimental results of our proposed approach for DSM in Section 3.3. Fig. 3.14 depicts the real and predicted hot water consumption during the experiment day. For this specific day, an abnormal real hot water consumption happened. Specifically, it can be seen from Fig. 3.14 that the peak hour started earlier than usual so that a relatively larger prediction error occurred compared to our previous case. While the prediction performance of water consumption is unsatisfactory, this scenario provided a good case to test the robustness of our proposed approach. The supplied hot water temperature and the HP control signals are plotted in Fig. 3.14. It can be seen that despite the large water usage prediction error, our proposed approach still performs well. The promised DR services were achieved while keeping the water temperature within the admissible range during almost the entire experiment. The only exception occurred around 14:00, at which the water temperature was slightly below the admissible lower

bound due to the large prediction error of water usage.

3.5. CONCLUSIONS

This chapter proposed an energy-flexible MPC design framework that leverages the energy flexibility of HPTES systems for DSM. Notably, our proposed MPC strategy can quantitatively assess the energy flexibility of HPTES systems by solving a typical economic MPC problem with additional mixed-integer linear constraints, ensuring that any admissible DR requests can be achieved while respecting system constraints. The proposed approach offers a systematic framework for operating HPTES systems using an advanced control strategy to achieve economic operation and DSM. The efficacy of our proposed approach is validated through both numerical and experimental results based on a real-world HPTES installation. The results demonstrate that the proposed MPC schemes are effective in achieving economical operation and energy-flexible DSM. Moreover, they provide a motivating example to confirm the practical viability of the future-proof S2 standard, enabling versatile exploitation of energy flexibility in the built environment.

4

Model Predictive Building Climate Control for Mitigating Heat Pump Noise Pollution

Noise pollution from heat pumps (HPs) has been an emerging concern to their broader adoption, especially in densely populated areas. This chapter explores a model predictive control (MPC) approach for climate control of buildings, aimed at minimizing the perceived noise nuisance from HPs by optimally scheduling their operation to hide HP noise in ambient noise. By exploiting a piecewise linear approximation of HP noise patterns and assuming linear building thermal dynamics, the proposed design can be generalized to handle various HP acoustic patterns with mixed-integer linear programming (MILP). Additionally, two computationally efficient options for defining the noise cost function in the proposed MPC design are discussed. Numerical experiments on a high-fidelity building simulator are performed to demonstrate the viability and effectiveness of the proposed design. Simulation results show that minimizing the excess of HP noise over ambient noise is effective in mitigating the HP noise nuisance. Further, compared with the conventional MPC-based building climate control scheme, the proposed approach can effectively reduce the HP noise pollution with only a minor energy cost increase.

This chapter is based on the paper: Li, Y., Shi, J., Jones, C. N., Yorke-Smith, N., & Keviczky, T. (2025). Model predictive building climate control for mitigating heat pump noise pollution. *European Journal of Control*, 101323.

4.1. Introduction

As an energy-efficient heating/cooling device, heat pumps (HPs) have gained widespread adoption across Europe, driven by the goal of reducing fossil fuel usage and carbon emissions. Currently, approximately 24 million HPs are installed in European buildings, and this number is expected to reach 60 million by 2030 [111]. This growing adoption of HPs to move away from fossil fuels could reduce Europe's gas demand for heating by at least 21 billion cubic meters in 2030, and potentially cut CO2 emissions by 46% [112, 113]. However, despite the benefits of flexible and efficient renewable heating and carbon reduction, a new concern about HPs has been raised: noise.

Noise can induce stress and impact both psychological and physiological well-being. Noise generated by HPs, particularly air source heat pumps (ASHPs) commonly installed in residential areas, has emerged as a primary concern hindering their broader acceptance in these settings [60, 114]. Consequently, HP installation and operation must account for acoustic impacts on the surrounding environment, especially in residential zones where noise levels are subject to legislative noise directives. For example, in the UK, the noise pressure level must be below 42 dB at a distance of one meter from a neighbour's door or window [113, 114]. Similar regulations also apply in other countries as outlined in [115]. Addressing HP noise pollution is thus essential for maintaining acoustic health and fostering the acceptance of HPs, which can further support carbon emissions reduction. Recent initiatives, such as IEA HPT Annex 51 and Annex 63 underscore the growing attention on HP noise concerns [116, 117].

Various solutions for reducing HP noise have been explored, including adding sound-absorbing materials or insulation enclosures, using flexible mountings to dampen vibrations, and implementing active noise cancellation techniques [60, 118, 119]. While these measures can reduce HP noise, they often require intrusive modifications, making it costly or even impractical to retrofit existing HPs for improved acoustic performance.

With advancements in smart metering, computing technology, and building management systems, an alternative approach to reducing HP noise without invasive modifications is to design optimal HP control strategies. The primary noise sources of air source heat pumps are the compressors and, especially, the fans in the outside units of HPs, which significantly contribute to ambient noise [60, 120]. Modern inverter HPs allow for modulation of compressor and fan speeds, enabling noise level adjustments based on thermal output requirements. Thus, it is possible to adjust the HP power inputs to mitigate noise while maintaining a comfortable indoor thermal climate.

Model predictive control (MPC) has shown promise as an advanced control strategy for HP and building climate control, owing to its flexibility in handling system constraints, economic considerations, and predicted weather conditions [26, 121]. The effectiveness of the MPC-based heat pump and building climate control in improving system performance, such as reducing economic cost, increasing indoor comfort, and engaging demand-side management, has been experimentally demonstrated in existing works, see [107, 122] and references therein.

Motivated by the above discussion, this chapter investigates an MPC design to mitigate HP noise within the context of building climate control. The main contributions of this chapter are summarized as:

- The reduction of HP noise pollution is investigated for the first time in optimal building climate control. A general MPC formulation that considers both HP noise pollution reduction and energy cost savings is proposed.
- Leveraging piecewise linear approximation, the proposed design is adaptable to various HP noise patterns through a mixed-integer linear programming (MILP) formulation. Two options for the noise cost function in the MPC design are discussed.
- Numerical experiments using a high-fidelity building simulator are performed to demonstrate the viability and effectiveness of the proposed approach.

The remaining parts of this chapter are organized as follows. Section 4.2 presents the problem setting about building thermal dynamics, HP noise patterns, and our design objective. Section 4.3 delves into the details of the proposed MPC design. The viability and effectiveness of the proposed approach are numerically tested in Section 4.4. Finally, Section 4.5 concludes this chapter.

4.2. PROBLEM SETTING

This work aims to mitigate the noise pollution caused by the operation of HPs while maintaining indoor thermal comfort by redesigning indoor climate control strategies. In the following, we will illustrate our problem setting about indoor thermal dynamics, heat pump noise and ambient noise, as well as the control design objective.

4.2.1. BUILDING THERMAL DYNAMICS

Without loss of generality, the indoor thermal dynamics are assumed to be approximately modeled as the following linear system

$$y_{t+1} = A\mathbf{y}_{t,k_y} + B\mathbf{u}_{t,k_u} + E\mathbf{v}_{t,k_v} \tag{4.1}$$

where $y_{t+1} \in \mathbb{R}^n$ is the predicted indoor temperature vector at time instant $t+1$, $\mathbf{y}_{t,k_y} \in \mathbb{R}^{nk_y}$ is the stacked historical indoor temperature measurements during time period $[t-k_y+1, t]$, which is defined as $\mathbf{y}_{t,k_y} := [y_t^{\mathrm{T}}, y_{t-1}^{\mathrm{T}}, \cdots, y_{t-k_y+1}^{\mathrm{T}}]^{\mathrm{T}}$; similarly, $\mathbf{u}_{t,k_u} \in \mathbb{R}^{mk_u}$ and $\mathbf{v}_{t,k_v} \in \mathbb{R}^{pk_v}$ are the stacked HP power input and ambient climate conditions during time intervals $[t-k_u+1, t]$ and $[t-k_v+1, t]$, respectively; A, B and E are system matrices describing the evolution of indoor temperature at time instant $t+1$ driven by past indoor temperature, HP control input and ambient climate conditions. The above linear model encompasses a wide range of the prediction models for indoor thermal dynamics developed by black-box approaches, e.g., Auto Regressive with eXogeneous inputs (ARX), and gray-box approaches, e.g., RC-network [26, 123, 124].

4.2.2. HEAT PUMP AND AMBIENT NOISES

For HP noise concerns, our primary focus is on the noise generated by the fan in the outside unit. In residential settings, HP fan noise generally ranges from 40-60 decibels, and is typically the main source of HP noise disturbance for nearby residents. As shown in [60, 125], the noise generated by HPs is a nonlinear and nonconvex function of the HP

power, and roughly follows a logarithm-like or sigmoid-like function. However, due to the lack of definitive studies showing that all ASHPs adhere to such a noise pattern, we impose no explicit assumption about the relationship between the HP noise level and power consumption.

In our control scheme design, we only assume the existence of a general noise pattern as defined in (4.2), which can be derived through theoretical analysis or experimental data:

$$L^{\text{hp}} = f(u) \tag{4.2}$$

where L^{hp} is the HP noise level in decibels, u is the HP power input, and $f(\cdot)$ is a function representing the HP noise pattern. The implicit assumption behind the above analysis is that the noise generated by HPs varies with their electrical power consumption, i.e., thermal output, and follows a certain predictable pattern. This assumption requires that the considered HPs should be inverter HPs, whose fan speed can be modulated and power input is adjustable, since the fan speed and power input for ON/OFF controlled HPs are generally fixed to certain nominal values and can not be modulated to adjust the corresponding noise level.

When defining the acoustic nuisance caused by a HP, another factor that should be considered is ambient noise (background noise), which might be caused by traffic noise, alarms, extraneous speech, animal noise, and more. In this work, we impose no specific pattern for ambient noise, assuming only that the predicted ambient noise levels are accessible. This is a practical assumption, and there are many works available that focus on developing ambient noise prediction algorithms, see [126, 127] and references therein. Thus, without loss of generality, in our upcoming MPC design, we assume that the predicted ambient noise level L^{amb} within the MPC prediction horizon is available.

4.2.3. Control Design Objective

The main control objective is to mitigate the acoustic nuisance caused by HP operation in the surrounding environment. It should be noted that reducing the acoustic nuisance of HPs does not equate to minimizing absolute HP noise, which would typically mean shutting down HPs. Instead, it involves reducing the relative impact of HP noise compared to ambient noise levels. Through appropriate HP control, the combined noise from the HP and its surroundings should be dominated by the ambient noise, effectively concealing HP noise within it and thereby mitigating acoustic pollution. In addition to reducing noise pollution, the control objective should also consider indoor comfort and energy costs.

For simplicity in MPC design, our approach does not account for the spectral characteristics of noise signals. Future studies incorporating human factors determining the human-perceived noise nuisance associated with different noise frequencies are warranted.

4.3. Model Predictive Control Design

4.3.1. Model Predictive Control Formulation

In this subsection, a general MPC problem is formulated to adaptively reduce the effect of HP noise on the environment. An MPC problem achieving our design objective can

be formulated as follows

$$\min_{u_t} \underbrace{\sum_{t=0}^{N} l(u_t, y_t)}_{J_o} + \eta \underbrace{\sum_{t=0}^{N} h(L_t^{\text{hp}}, L_t^{\text{amb}})}_{J_n} \tag{4.3a}$$

$$\text{s.t. } y_{t+1} = A\mathbf{y}_{t,k_y} + B\mathbf{u}_{t,k_u} + E\mathbf{v}_{t,k_v}, \tag{4.3b}$$

$$L_t^{\text{hp}} = f(u_t), \tag{4.3c}$$

$$y_t \in \mathscr{Y} \text{ and } u_t \in \mathscr{U}, \ \forall t \in \{0, \cdots, N\}, \tag{4.3d}$$

where N is the length of prediction horizon, J_o is total operational cost with $l(u_t, y_t)$ as the stage cost at sampling instant t, J_n is the total noise cost within the prediction horizon with $h(L_t^{\text{hp}}, L_t^{\text{amb}})$ as the stage noise cost defined based on HP noise L_t^{hp} and ambient noise L_t^{amb}, $\eta \geq 0$ is a user-defined weighting factor for balancing the operational cost and noise nuisance, constraint (4.3b) is the building thermal dynamics defined in (4.1), constraint (4.3c) defines the noise pattern of the HP in (4.2), $\mathscr{Y}$ and $\mathscr{U}$ in (4.3d) are admissible regions of indoor temperature and heat pump power input. This optimization problem defines a general control task to minimize the weighted sum of HP operational cost and HP noise cost while ensuring indoor comfort constraints and HP input constraints. Within the context of the building climate control, the operational cost function J_o is generally defined as the energy cost during the prediction horizon, which is formulated as a linear (convex) function of HP power consumption and will not influence the computational tractability of the proposed scheme, see [121]. Besides, the formulation of hard constraints for indoor comfort requirements (4.3d) can be relaxed as soft constraints by introducing slack variables to ensure the feasibility of the MPC problem in practical implementation.

Remark 4.1. It should be highlighted that the design objective is to mitigate the HP noise nuisance w.r.t. the environment noise. Consequently, the definition of the noise cost function J_n in (4.3a) should reflect the relative noise nuisance of the HP, rather than its absolute value. For example, intuitively, when the ambient environment is noisy, the HP can operate with louder noise, possibly to achieve higher energy efficiency or lower energy bills, without incurring a high acoustic nuisance. Similarly, when the ambient environment is quiet, even a moderate noise level of HP can lead to more nuisance because the HP noise plays a dominating role in the total noise.

4.3.2. PIECEWISE AFFINE APPROXIMATION OF HP NOISE PATTERN

Assuming a general HP noise pattern, this subsection presents a piecewise linear approximation of the HP noise pattern and develops a computationally tractable formulation.

As explained in Section 4.2.2, since the specific noise pattern might vary depending on the individual HP system, we make no explicit assumption about the HP noise pattern, and aim at developing methods that are adaptable to a broader range of HP noise patterns for enhancing the applicability of the proposed approach.

In this work, piecewise affine functions are utilized to approximate the HP noise pattern, which might be nonlinear and nonconvex. Fig. 4.1 shows an example of using three pieces of affine functions to approximate a sigmoid-like noise pattern. To provide a gen-

eral approximation scheme, we assume that affine functions comprising k pieces are used in HP noise approximation. The admissible scope of HP control input is partitioned into k intervals that are defined by $\alpha = [\alpha_0, \cdots, \alpha_k]^{\mathrm{T}}$ with $[\alpha_i, \alpha_{i+1}]$ $(i = 0, \cdots, k-1)$ representing one interval, where the HP noise pattern is approximated via a piecewise affine function. Correspondingly, the vector $\beta := [\beta_0, \cdots, \beta_k]$ is defined with β_i as the HP noise level when its power input $u = \alpha_i$. Then, for any HP power input $u \in [\alpha_i, \alpha_{i+1}]$, there exist real-valued parameters λ_i and λ_{i+1} such that

$$u = \lambda_i \alpha_i + \lambda_{i+1} \alpha_{i+1} \tag{4.4}$$

with $0 \le \lambda_i \le 1$ and $\lambda_i + \lambda_{i+1} = 1$. Correspondingly, the value of the approximated HP noise level $\hat{L}^{\mathrm{hp}}$ is

$$\hat{L}^{\mathrm{hp}} = \lambda_i \beta_i + \lambda_{i+1} \beta_{i+1} \tag{4.5}$$

To denote the approximated noise pattern within the whole admissible input range $[\alpha_0, \alpha_k]$, binary variables $z_i \in \{0, 1\}$ $(i = 1, \cdots, k)$ are introduced with $z_i = 1$ indicating $u \in [\alpha_{i-1}, \alpha_i]$. Finally, the piecewise-affine approximated HP noise pattern can be expressed as the following mixed-integer linear constraints

$$u = \textstyle\sum_{i=0}^{k} \lambda_i \alpha_i, \tag{4.6a}$$
$$\hat{L}^{\mathrm{hp}} = \textstyle\sum_{i=0}^{k} \lambda_i \beta_i, \tag{4.6b}$$
$$\lambda_{i-1} + \lambda_i \le z_i, \ \forall i \in \{1, \cdots, k\} \tag{4.6c}$$
$$\lambda_i \ge 0, \forall i \in \{1, \cdots, k\} \tag{4.6d}$$
$$\textstyle\sum_{i=1}^{k} z_i = 1, \ z_i \in \{0, 1\}. \tag{4.6e}$$

The above mixed-integer linear constraints can replace the HP noise pattern constraint in (4.3c), which might be nonlinear and nonconvex, and enable a universal approximation for various HP noise patterns. It should be noted that approximation accuracy is determined by the number of affine functions utilized in (4.6). Adopting more affine functions leads to a more accurate approximation, but will also introduce more constraints and decision variables in (4.6). Consequently, the number of affine functions k considered in (4.6) should balance the approximation accuracy of HP noise pattern and the corresponding computational overhead.

4.3.3. Noise Cost Function Design

This subsection presents several possible options for defining the noise cost function J_n in (4.3). Recall that our control objective is to mitigate the relative acoustic nuisance in comparison to ambient noise, rather than minimizing the absolute HP noise level, so that the HP noise is hidden in the ambient noise. Accordingly, the design of the noise cost J_n should emphasize relative noise mitigation.

Option 1

The first option of the noise cost is defined as

$$J_n := \textstyle\sum_{t=0}^{N} L_t^{\mathrm{hp}} / L_t^{\mathrm{amb}} \tag{4.7}$$

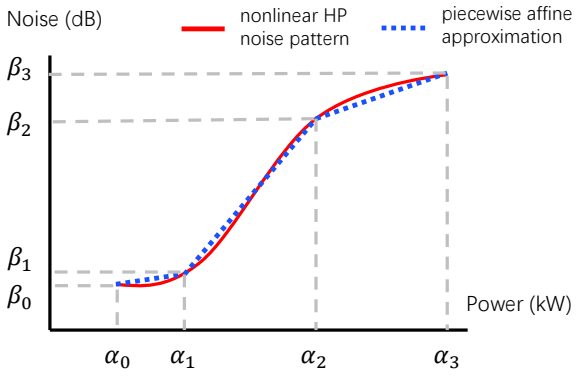


Figure 4.1: Nonlinear heat pump noise pattern and its piecewise affine approximation.

The above cost function penalizes HP noise according to the ambient noise level. Higher ambient noise imposes less penalty on HP noise. Consequently, this cost function incentivizes HP to work at a higher load when the ambient environment is noisy and at a lower load in quieter settings.

OPTION 2

While the cost function defined in (4.7) is straightforward and easy to implement, it fails to impose direct regulation on the mixed noise and may not prevent HP noise from dominating the ambient noise. According to the acoustic properties of combined sounds, the sound level of the mixed noise from HP and ambient sources can be calculated as

$$L^{\text{mix}} = 10 \cdot \log_{10}\left(10^{\frac{L^{\text{amb}}}{10}} + 10^{\frac{L^{\text{hp}}}{10}}\right) \tag{4.8}$$

The above nonlinear function can be used to impose direct constraints on the mixed noise. However, it introduces nonlinear constraints, which might be computationally challenging for certain numerical solvers.

The definition of the mixed noise level in (4.8) suggests that when the ambient noise level exceeds the HP noise, the mixed noise level will be primarily dominated by the ambient noise, due to the power function applied to each noise level. Thus, an alternative approach is to penalize instances when the HP noise exceeds the ambient noise, resulting in the following noise cost function definition

$$J_n := \sum_{t=0}^{N} \delta_t, \quad L_t^{\text{hp}} \leq L_t^{\text{amb}} + \delta_t \tag{4.9}$$

with $\delta_t \geq 0$. The above cost function is equivalent to the nonlinear noise cost function $J_n := \sum_{t=0}^{N} (L_t^{\text{hp}} - L_t^{\text{amb}})^+$, where x^+ is defined as $x^+ = x$ if $x \geq 0$ and otherwise $x^+ = 0$. This cost function incentivizes that HP noise does not exceed ambient noise, allowing the mixed noise to be predominantly influenced by ambient sounds, thereby masking HP noise within the background noise.

Remark 4.2. It is worth noting that the mixed-integer linear formulations in (4.6) and (4.9) are designed to enhance the applicability and computational feasibility of our approaches for a variety of HP noise patterns and for compatibility with most numerical

solvers. However, if the available solvers are capable of handling the specific, possibly nonlinear and nonconvex, HP noise pattern in (4.2) and the mixed noise pressure level definition in (4.8), this nonlinear relationship could be directly incorporated into (4.3), potentially improving control performance.

4.4. Simmulation Results

This section presents numerical simulation results to demonstrate the viability and effectiveness of our proposed design framework. The building model `bestest_hydronic_heat_pump` in the building control test platform `boptest` [128] is utilized as a high-fidelity simulator to test our design. The considered building model is a residential building with a rectangular floor plan 12m×16m, a height of 2.7m, and an air-to-water HP of 15 kW nominal heating capacity for floor heating. See [128] for more details about this building control test platform. The diagram of our simulation is shown in Fig. 4.2. At each sampling instant, the HP control input signal u_t is computed by solving (4.3). With the computed HP control input, the building simulator updates its internal states and returns the updated indoor temperature y_{t+1}. All simulations are performed on an Intel Xeon W-2223 CPU at 3.60GHz with 16G RAM. MPC problems are modeled via the Python package `gurobipy` and solved using `Gurobi 11.0` [74].

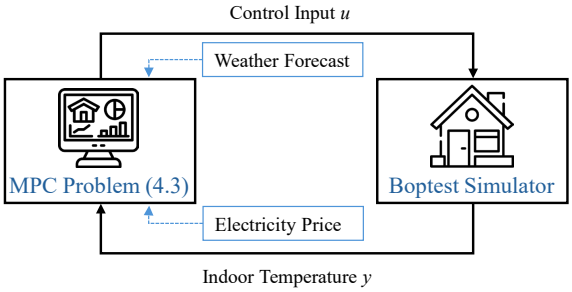


Figure 4.2: Diagram of closed-loop simulation with Boptest simulator.

In our simulation, the sampling period is set as 15-min for both prediction model development and MPC design. An ARX model is identified using randomly generated open-loop control signals. Fig. 4.3 depicts the real indoor temperature profiles and their open-loop predicted values using the ARX model, showing that the ARX model provides satisfactory prediction performance for subsequent MPC design. Due to the limitation of space, the detailed information of the ARX model is provided in the extended version of this paper [129].

4.4.1. Prediction Model of Building Thermal Dynamics

In our simulation, the sampling period is selected as 15 minutes for both prediction model development and MPC design. An ARX model is identified using randomly generated open-loop control signals to approximate the thermal dynamics of the building. Notably, the control signals are forced to be off (on) when the building is overheated (overcooled). The `boptest` simulator utilized in our simulation also involves the inter-

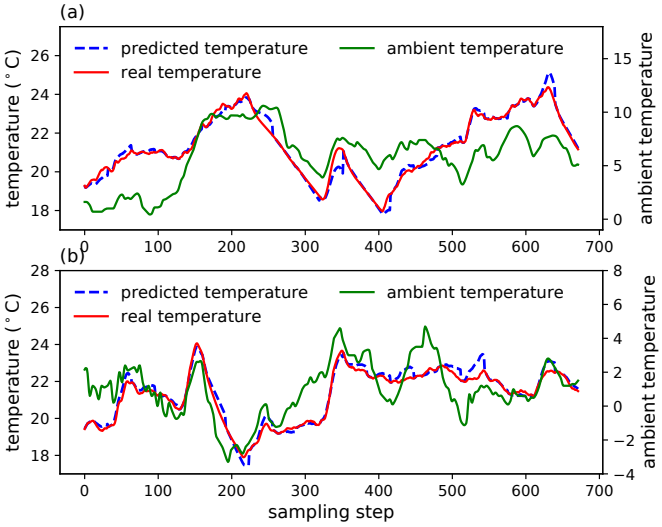


Figure 4.3: Open-loop prediction performance of ARX model: (a) training set (MAE = 0.16°C), (b) test set (MAE = 0.19°C).

nal thermal gains from 5 occupants, but they are not forecast and therefore neglected in our ARX model. The ARX model has the following structure

$$y_t = \sum_{k=1}^{n_a} a_k y_{t-k} + \sum_{k=1}^{n_b} b_k u_{t-k} + \sum_{k=1}^{n_c} c_k T_{t-k} + \sum_{k=1}^{n_d} d_k S_{t-k}$$

where y_t is the indoor temperature at time instant t, T_t denotes the ambient temperature, S_t denotes the solar irradiation. The prediction horizon is set as 8 hours. In our model the parameters (n_a, n_b, n_c, n_d) are set as $(4,1,2,2)$. The parameters (a_k, b_k, c_k, d_k) are identified by minimizing the sum of the squares of the indoor temperature prediction error via `scipy.optimize.least_squares` [130]. For training and testing the ARX model, two independent 7-day of datasets (16th Jan – 22nd Jan, 24th Jan – 30th Jan) are utilized, i.e., 7*24*4 = 672 data points, in both training and test datasets.

Fig. 4.3 depicts the real indoor temperature profiles and their open-loop predicted values using the ARX model for both the training and test datasets. The mean absolute errors (MAEs) for both training and test sets are 0.16 °C and 0.19°C, respectively, which implies that the ARX model provides satisfactory prediction performance for subsequent MPC design.

4.4.2. MODEL PREDICTIVE CONTROL FOR NOISE MITIGATION

For design simplicity, the admissible range of HP power input is scaled to such that $u_t \in [0,1]$. The indoor comfort constraint is set as $19°C \leq y_t \leq 24°C$. For the HP noise pattern, its real value is assumed to be identical with the piece-wise affine approximated value, that are defined with the vectors α and β used in (4.6) as $\alpha = [0, 0.2, 0.7, 1]$ and $\beta = [0, 40, 60, 60]$.

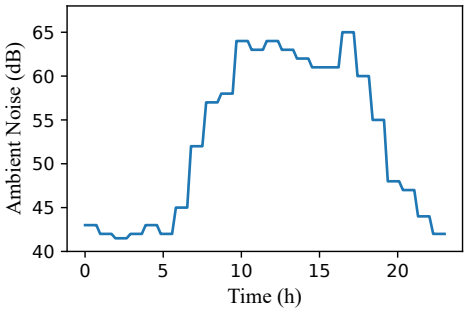


Figure 4.4: Ambient noise profile used in simulation.

The operational cost function, also referred as energy cost, in the MPC design (4.3) is defined as the electricity cost within the prediction horizon, i.e., $J_o := P_{\max} \cdot \sum e_t u_t$, where $P_{\max}$ is the maximal HP power input, e_t is the day-ahead electricity price. The prediction horizon $N = 32$, i.e., 8 hours. The ambient noise pattern used in our simulation is shown in Fig. 4.4, which is generated based on the results in [126, 131]. It can be seen that the environment is quiet during the early morning, evening and night hours, and is noisy in the noon and afternoon, which is consistent with our everyday experience.

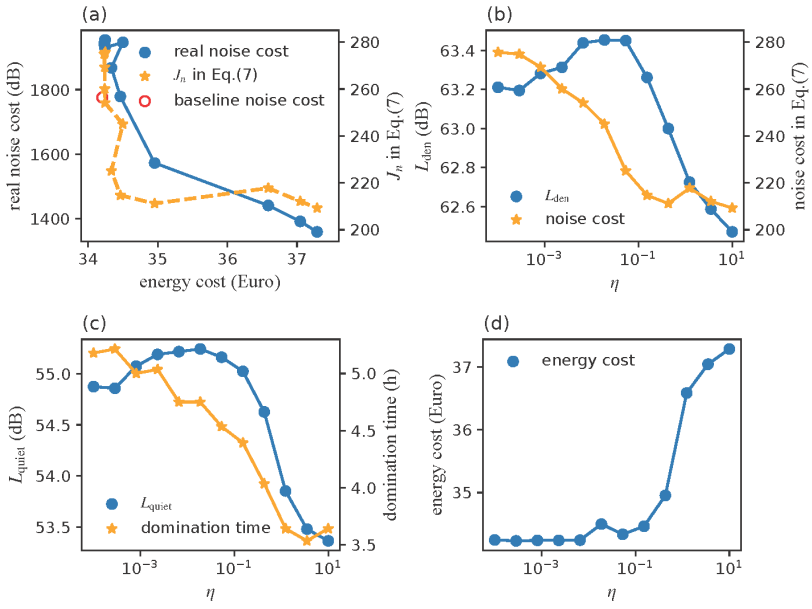


Figure 4.5: Simulation results for noise cost in (4.7): (a) Pareto curves of energy cost and noise cost, (b) L_{den}, (c) L_{quiet} and domination time, (d) energy cost.

In our case studies, the proposed two options of noise cost J_n defined in (4.7) and

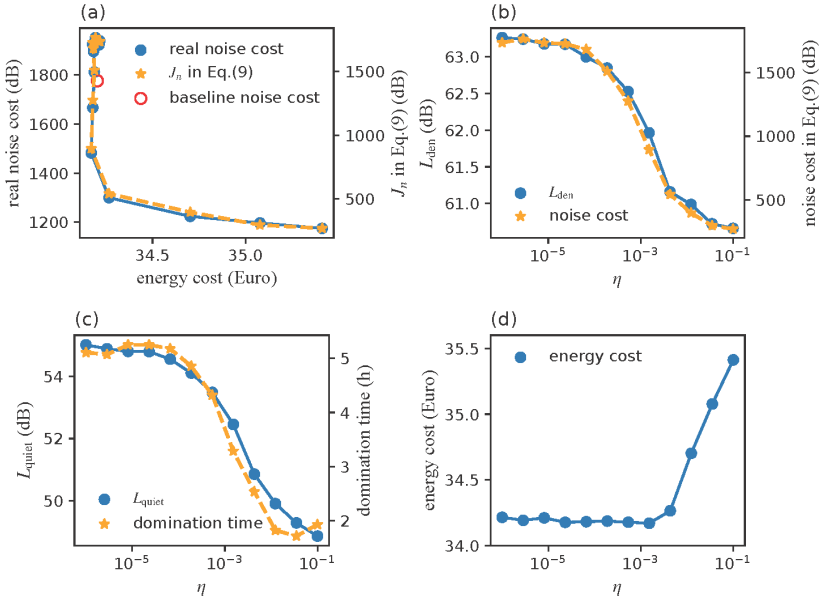


Figure 4.6: Simulation results for noise cost in (4.9): (a) Pareto curves of energy cost and noise cost, (b) L_{den}, (c) L_{quiet} and domination time, (d) energy cost.

(4.9) are tested for the MPC design, respectively.

The explicit MILP formulations for (4.3) with HP noise pattern approximation (4.6), and the noise cost function (4.9) are formulated as follows:

$$\min_{\substack{u_t,\lambda_{i,t},z_{i,t}\\ \delta_t,\hat{L}_t^{hp}}} \quad P_{\max}\cdot\sum_{t=0}^{N} e_t u_t + \eta\cdot\sum_{t=0}^{N}\delta_t \tag{4.11a}$$

$$\text{s.t. } y_{t+1} = A\mathbf{y}_{t,k_y} + B\mathbf{u}_{t,k_u} + E\mathbf{v}_{t,k_v}, \tag{4.11b}$$

$$u_t = \sum_{i=0}^{k}\lambda_{i,t}\alpha_i,\ \hat{L}_t^{\text{hp}} = \sum_{i=0}^{k}\lambda_{i,t}\beta_i, \tag{4.11c}$$

$$\lambda_{i-1,t} + \lambda_{i,t} \le z_{i,t}, \quad \forall i \in \{1,\cdots,k\}, \tag{4.11d}$$

$$\lambda_{i,t} \ge 0, \quad \forall i \in \{0,\cdots,k\}, \tag{4.11e}$$

$$\sum_{i=1}^{k} z_{i,t} = 1, \quad z_{i,t} \in \{0,1\}, \tag{4.11f}$$

$$19 \le y_t \le 24, \quad 0 \le u_t \le 1, \tag{4.11g}$$

$$\hat{L}_t^{\text{hp}} \le L_t^{\text{amb}} + \delta_t, \quad \delta_t \ge 0, \tag{4.11h}$$

$$\forall t \in \{0,\cdots,N\}. \tag{4.11i}$$

where the operation cost J_o is defined as the HP electricity cost $P_{\max}\sum_{t=0}^{N} e_t u_t$. In case of the noise cost function defined in (4.7), the term $\sum_{t=0}^{N}\delta_t$ in the above cost function is replaced by $\sum_{t=0}^{N}\hat{L}_t^{\text{hp}}/L_t^{\text{amb}}$, and constraint (4.11h) is removed.

In the simulation, different values of η in (4.3a) are tested. For each value of η, seven

days of closed-loop simulation are performed. Based on the simulation results, the total noise cost that is defined in (4.7) and (4.9), and energy cost (defined as the electricity cost J_o) during the simulation period are computed. Besides, since the noise costs defined in (4.7) and (4.9) are the convex approximations of the desired noise penalty, which is to ensure that the mixed noise is primarily dominated by ambient noise, we introduce a common metric:

$$\textit{real noise cost} := \sum_t L_t^{\text{mix}} - L_t^{\text{amb}} \tag{4.12}$$

to evaluate different approaches. In addition, the daily-averaged values of the following metrics are also evaluated:

- L_{den} of mixed noise: the day-evening-night noise level. L_{den} is used to measure the overall sound exposure over 24 hours. It is defined as the equivalent sound level with different penalties over different time periods in day, evening and night [132].
- L_{quiet} of mixed noise: L_{quiet} is defined as the equivalent sound level during quiet time (10:00 pm - 7:00 am).
- *domination time*: the total time over 24 hours that the mixed-noise level is dominated by the HP noise.

Furthermore, the *baseline* approach, in which the HP is operated to minimize the energy cost while complying with Switzerland's day-night noise regulation, i.e., 60 dB limit during daytime and 50 dB limit at night [115], is also considered in our case study.

Table 4.1: Performance summary with different noise cost functions.

	noise cost in (4.7)	noise cost in (4.9)
noise cost J_n reduction percentage (%)	24.09	84.48
real noise cost reduction percentage (%)	30.43	39.38
energy cost increase percentage (%)	8.89	3.50
L_{den} reduction (dB)	0.74	2.60
L_{quiet} reduction (dB)	1.51	6.15
domination time reduction (h)	1.54	3.39
average MPC computation time (s)	2.17	0.87

Simulation results are plotted in Figs. 4.5 and 4.6. Table 4.1 summarizes the results in terms of the maximal noise cost reduction percentage and the corresponding *real noise cost* reduction percentage, energy cost increase percentage, L_{den} reduction, L_{quiet} and domination time reduction, and energy cost increase for all considered values of η with both noise cost definitions in (4.7) and (4.9), respectively.

Fig. 4.5 presents the simulation results using the noise cost function J_n defined in (4.7). The Pareto curves of the *real noise cost* in (4.12) and the noise cost J_n in (4.7) w.r.t. energy cost in Fig. 4.5(a), along with Table 4.1, indicates that noise cost J_n can be reduced by 24.09% with an 8.89% increase in energy cost. In the meanwhile, the *real noise cost* is reduced by 30.47%. Fig. 4.5(b) illustrates the variations in L_{den} and noise cost as η in (4.3a) increases. It is observed that while both L_{den} and noise cost J_n generally follow a downward trend, their patterns are not entirely consistent, implying that a reduction in

noise cost does not necessarily correlate with decreased noise nuisance in L_{den}, as also discussed in Section 4.3.3. Similarly, the inconsistency for L_{quiet} and domination time is also visible in Fig. 4.5(c). Fig. 4.5(d) shows that the total energy cost is within 3.

In Fig. 4.6(a), the Pareto curves illustrate the trade-off between noise and energy costs for the noise cost function in (4.9). Together with Table 4.1, it can be observed that the noise cost J_n and the *real noise cost* are reduced by 84.48% and 39.38%, respectively, with only a 3.50% increase in energy cost. Fig. 4.6(b) presents the values of L_{den} and noise cost J_n across various values of η. Similarly, as depicted in Fig. 4.6(c), both L_{quiet} and *domination time* decreases as η increases, achieving a notable 6 dB reduction in L_{quiet} and 3.39 h reduction in *domination time*, which are much larger than the case with J_n defined in (4.7), where a reduction of 1.51 dB in L_{quiet} and 1.54 h in *domination time* are achieved. Notably, Fig. 4.6 shows a much more consistent pattern among the noise cost J_n in (4.9), *real noise cost*, L_{den}, L_{quiet} and *domination time* than in Fig. 4.5 where the noise cost (4.7) is considered. In addition, the energy cost increase shown in Fig. 4.6(d) is also less than the case of Fig. 4.5(d). These results imply that the noise cost in (4.9) is more effective in penalizing the mixed noise level and ensuring the mixed noise is dominated by the ambient noise.

In addition, as shown in Fig. 4.5(a) and Fig. 4.6(a), the *baseline* approach mainly aims at reducing the energy cost and cannot consider the ambient noise profile, which leads to large noise costs. Besides, one might notice that in Fig. 4.5(a) and Fig. 4.6(a) the energy cost does not always increase as the noise cost decreases, which is possibly due to modeling errors causing the HP to deviate from the predicted optimal value for maintaining indoor comfort. This issue could be mitigated by using stochastic or robust optimization-based approaches to enhance the robustness of the MPC solution.

4.5. CONCLUSIONS

This chapter presents the first investigation into HP noise mitigation within the context of building climate control. The proposed approach extends the standard economic MPC design for building climate control to incorporate HP noise reduction. By adopting a piecewise linear approximation, the proposed approach can accommodate diverse HP noise patterns while maintaining computational efficiency by solving MILP problems. The proposed noise cost functions ensure that the HP noise does not dominate the ambient noise, thereby reducing its acoustic impact on surrounding environments. Simulation results using a high-fidelity building simulator show that, with the proposed MPC design, HP noise can be mitigated with only a minor increase in energy costs.

5

MACHINE LEARNING ENABLED UNCERTAINTY SET FOR DATA-DRIVEN ROBUST OPTIMIZATION

The way how the uncertainties are represented by sets plays a vital role in the performance of robust optimization (RO). This chapter presents a novel approach leveraging machine learning (ML) techniques to construct data-driven uncertainty sets from historical uncertainty data for RO problems. The proposed method integrates Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Gaussian Mixture Model (GMM), and Principal Component Analysis (PCA) systematically to eliminate the influence of uncertainty scenarios with low occurrence probability and generate a nonconvex uncertainty set that is a union of multiple basic subsets (box or ellipsoid) without sacrificing its computational tractability. In addition to presenting a comprehensive algorithm for uncertainty set development, this work offers detailed guidelines for parameter tuning and performance analysis. By harnessing the well-established ML packages `scikit-learn`*, a Python-based toolkit for implementing the proposed approach is also provided. Furthermore, a computationally efficient solution for a two-stage linear RO problem with the proposed data-driven uncertainty set is derived, alongside establishing a probabilistic guarantee of constraint satisfaction for out-of-sample uncertainties. Extensive numerical experiments, conducted on both synthetic and real-world datasets as well as an optimization-based control problem, are performed to demonstrate the efficacy of the proposed methodology.*

This chapter is based on the paper: Li, Y., Yorke-Smith, N., & Keviczky, T. (2024). Machine learning enabled uncertainty set for data-driven robust optimization. *Journal of Process Control*, 144, 103339.

5.1. INTRODUCTION

Optimization under uncertainties is ubiquitous in real-world engineering problems and has attracted significant research attention. There are two primary approaches – stochastic optimization and robust optimization – to enhance the robustness and reliability of deterministic optimization models amidst uncertainties [6–8, 32]. Within the framework of stochastic optimization, the exact distributional information about uncertainties is deployed, and the expected performance towards the uncertainty distribution is optimized. In reality, however, the distributional information of uncertainties is usually not available, and obtaining this information is also a non-trivial task. One popular solution to solve the implementation difficulty of stochastic optimization is the scenario-based approach [133, 134]. However, to ensure constraint satisfaction with a high confidence level, the scenario-based approach entails a large number of scenario-induced hard constraints, which leads to computational challenges. As an effective alternative, robust optimization (RO) models uncertainties via uncertainty sets without the distributional information of uncertainties and focuses on optimizing the worst-case performance [6–8]. Due to its effectiveness in constraint satisfaction and computational tractability, RO has gained increased popularity.

A crucial component of RO problems is the uncertainty set, which significantly influences both the computational complexity and conservatism of the corresponding RO problems. Common types of uncertain sets include box, ellipsoid, polyhedral and intersections or unions of these basic sets [6, 8, 77, 135, 136]. These conventional methods for constructing uncertainty sets are straightforward to implement and allow for some reduction in the conservatism of the optimal solution through careful adjustment of the set coefficients. However, the selection of these coefficients typically depends on domain-specific knowledge. Moreover, these methods are based on the assumption that each dimension of uncertainties is independently and asymmetrically distributed, which restricts their efficiency in handling correlations among uncertainties and scalability for high-dimensional uncertainties [35].

With the availability of abundant historical uncertainty data and the development of machine learning (ML) techniques, data-driven RO approaches have attracted increasing attention in reducing the conservatism of RO. The main idea of these approaches is to exploit ML techniques, especially unsupervised learning such as support vector clustering (SVC), kernel density estimation (KDE) and principle component decomposition, to extract the latent patterns of uncertainties, which are subsequently used for constructing uncertainty sets. It should be pointed out that many dominant ML techniques do not apply to data-driven RO problems due to the utilization of nonlinear functions, e.g., radial basis function and sigmoid function, which are widely adopted in ML algorithms but might dramatically degrade the computational tractability of the resulting RO problems based on such data-driven uncertainty sets.

Recent literature explores different methods for constructing data-driven uncertainty sets in RO. In [11, 15, 35], a kernel-based support vector clustering (K-SVC) method using a novel piece-wise linear kernel is proposed to develop a non-parametric polyhedral uncertainty set. Subsequently, [137] proposed replacing the piece-wise linear kernel with a deep neural network (DNN) to construct a more compact, though non-convex, uncertainty set. However, the DNN-based sets lead to substantially longer com-

putation times in solving the corresponding RO problems, particularly, even single-stage linear RO problems with NN-based uncertainty sets involve solving mixed-integer quadratic programs iteratively. In [36, 138], the Dirichlet process mixture model is utilized to extract hidden patterns in uncertainty data through a variational inference algorithm, which involves complicated nonconvex optimization that degrades its applicability. In [139], PCA is applied to develop a polyhedral uncertainty set by decomposing the uncertainty into uncorrelated components. However, this method fails to detect low-probability uncertainties, which might result in a conservative uncertainty set. In [140–142], by combining principle components analysis (PCA) with kernel density estimation (KDE), the resultant uncertainty set is able to exclude low-probability uncertainties within the tails of the approximated probability distribution. In [143, 144], PCA is combined with cutting plane methods to reduce the conservatism of the resulting uncertainty set by excluding redundant uncertainty scenarios.

Although the above PCA-based uncertainty sets are computationally efficient due to their polyhedral structure, this simple structure also sacrifices their applicability to complicated and irregular uncertainty distributions. In [62, 145], the PCA-KDE-based approach proposed in [140] is further combined with KMeans clustering to construct uncertainty sets that are applicable to disjunctive uncertainties. However, as will be shown in our simulation results, although this combination offers increased flexibility for handling irregular uncertainty distributions, its performance might be degraded due to the limitation inherent to the KMeans and KDE approaches and lacks robustness against complicated real-world datasets.

In summary, a common limitation of the above data-driven approaches is their inability to effectively balance between the complexity and the conservatism of the uncertainty set, and to adapt to complicated uncertainty distributions. Most of these methods either yield a compact uncertainty set that demands high computational resources for solving the resulting RO problems, or they produce a computationally efficient uncertainty set that, however, lacks adaptivity to complex uncertainty distributions. Besides, while some literature explored the framework of unifying multiple subsets as in this work, their approaches fail to adapt to complicated uncertainty distributions due to the improper selection and integration of ML techniques. In addition, most of the existing works only use simple synthetic datasets for performance evaluation, which might fail to truly reflect the actual practical performance when applied to complex real-world data. Furthermore, detailed guidelines as well as easy-to-use toolkits for the existing approaches are not available.

Motivated by the above discussions, this chapter proposes an ML-enabled data-driven approach for a two-stage adaptive RO problem. The major contributions of this chapter are summarized as follows:

- An ML-based approach is proposed to develop data-driven uncertainty sets by leveraging DBSCAN, GMM and PCA. The resulting uncertainty set is compact regardless of irregular uncertainty distributions and is computationally efficient in solving the resulting RO problems.

- Detailed guidelines for parameter tuning, performance analysis and possible limitations of applying the proposed data-driven uncertainty set are provided to fa-

cilitate its practical usage. A Python-based toolkit to implement the proposed approach is developed [146].

- The conventional affine decision rule is extended to the proposed uncertainty set by exploiting the union of multiple subsets property for a two-stage linear RO problem to give less conservative solution. A probabilistic guarantee of constraint satisfaction for out-of-sample uncertainties is derived.

- The effectiveness of the proposed approach is extensively validated using both synthetic and real-world datasets as well as an optimization-based control problem.

The remainder of this chapter is organized as follows. Section 5.2 introduces the ML techniques used in our work and details the algorithm for developing data-driven uncertainty sets, including guidelines for parameter tuning and performance analysis. Section 5.3 investigates a two-stage linear RO problem employing the proposed uncertainty set. Section 5.4 presents three case studies to demonstrate the efficacy of our proposed approach. Finally, conclusions are drawn in Section 5.5.

Notation: Boldface lowercase letters are used to denote vectors, and boldface uppercase letters denote matrices. Calligraphic uppercase letters denote sets. For a given vector/matrix, $[\cdot]_k$ refers to its k-th element/row. The max/min operators applied to vectorized objective functions imply elementwise maximization/minimization across each function. Equalities/inequalities between two vectors hold elementwise.

5.2. DATA-DRIVEN UNCERTAINTY SET CONSTRUCTION AND ANALYSIS

In this section, we will employ ML techniques to construct data-driven uncertainty sets from historical uncertainty data with the aim of reducing the conservatism of the resulting RO problem while preserving computational efficiency. Specifically, we propose an integrated framework combining Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Gaussian Mixture Models (GMM), and Principal Component Analysis (PCA) to reveal hidden patterns in uncertainty data.

5.2.1. DBSCAN

DBSCAN is a density-based clustering algorithm that groups data into clusters based on high-density areas, which are separated by regions of low density [147]. In contrast to centroid-based or distribution-based clustering methods, such as KMeans, which typically identify spherical or convex clusters, DBSCAN is capable of discovering clusters of arbitrary shapes. In addition, DBSCAN can detect low-probability data samples: data samples residing in low-density areas will not be assigned to any clusters. As a result, in this work, we will adopt DBSCAN to remove low-probability uncertainty samples, thereby reducing the conservatism of our proposed data-driven uncertainty set.

There are two main parameters for DBSCAN: ϵ and $MinPts$. ϵ determines the maximum distance between two samples for one to be considered as a neighbor of the other. Given a data point $\mathbf{p}$, its ϵ-Neighborhood is defined as $N_\epsilon(\mathbf{p}) = \{\mathbf{q} : \text{dist}(\mathbf{p},\mathbf{q})\} \leq \epsilon$. $MinPts$

represents the minimum number of neighbors a sample must have to be classified as a core point. Namely, $|N_\epsilon(\mathbf{p})| \geq MinPts$ if $\mathbf{p}$ is a core point. For a core point $\mathbf{p}$, any data samples that are density-reachable from $\mathbf{p}$ will be grouped in the same cluster. The data samples that do not belong to any clusters will be classified as noise. Generally, increasing ϵ results in fewer samples being classified as outliers, whereas increasing $MinPts$ tends to identify more outliers. For further details about DBSCAN, please see [147, 148].

5.2.2. GAUSSIAN MIXTURE MODEL CLUSTERING

Gaussian mixture model (GMM) clustering, which is a distribution-based clustering approach, assumes all given data samples are generated from a mixture of a finite number of Gaussian distributions. Each distribution in the mixture is characterized by a set of parameters: the mixing probability π_k, mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$ of the k-th model. Given n data samples $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$, these unknown parameters are estimated by maximizing the following log-likelihood function

$$\sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \pi_k \phi(\mathbf{u}_i, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \tag{5.1}$$

where K is the total number of Gaussian distributions, and $\phi(\mathbf{u}_i, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the multivariate Gaussian density function of the k-th Gaussian model. The optimal value of the parameters $(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ can be computed via Expectation Maximization algorithm [149]. Based on the parameters $(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, all data samples can be clustered into K groups, and the data samples in the same group are assigned with an identical label k. The label for a given data sample $\mathbf{u}_i$, denoted as z_i, is $\arg\max_k \{p(z_i = k|\mathbf{u}_i)\}$ where

$$p(z_i = k|\mathbf{u}_i) = \frac{p(z_i = k)p(\mathbf{u}_i|z_i = k)}{p(\mathbf{u}_i)} = \frac{\pi_k \phi(\mathbf{u}_i, \mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k \phi(\mathbf{u}_i, \mu_k, \Sigma_k)}.$$

In our work, the uncertainty samples will be firstly processed by DBSCAN to remove low-possibility scenarios. Then, GMM clustering will be applied to group the remaining uncertainty samples into K clusters for subsequent uncertainty subset construction via PCA.

5.2.3. PRINCIPAL COMPONENT ANALYSIS

In our work, PCA is utilized to construct box-like uncertainty subsets based on each uncertainty cluster generated by GMM. This subsection will briefly introduce the PCA approach and show how PCA is utilized in uncertainty set construction. A detailed tutorial about PCA can be found in [150].

PCA is a popular data analysis technique for dimension reduction and enhancing data interpretability. It achieves this by applying a linear transformation to the original data so that all features of the new data representation are mutually uncorrelated.

Assume that there are K uncertainty clusters $\{\mathbf{U}_1, \cdots, \mathbf{U}_k\}$ after applying DBSCAN and GMM. Then, for each data cluster $\mathbf{U}_k = [\mathbf{u}_1^{\mathrm{T}}, \cdots, \mathbf{u}_{n_k}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbf{R}^{n_k \times m}$, where n_k is the number of uncertainty samples and m is the dimension of the uncertainty, we subtract its mean $\boldsymbol{\beta}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{u}_i$ from each sample in the cluster and obtain the centered training data

set $\bar{\mathbf{U}}_k$. For $\bar{\mathbf{U}}_k$, we have the following approximated covariance matrix

$$\mathbf{C}_k^{\mathbf{U}} = \frac{1}{n_k - 1} \bar{\mathbf{U}}_k^{\mathrm{T}} \bar{\mathbf{U}}_k. \tag{5.2}$$

By performing eigenvalue decomposition, the covariance matrix $\mathbf{C}_k^{\mathbf{U}}$ can be decomposed as

$$\mathbf{C}_k^{\mathbf{U}} = \mathbf{P}_k \mathbf{\Lambda}_k \mathbf{P}_k^{\mathrm{T}} \tag{5.3}$$

where $\mathbf{P}_k = [\mathbf{p}_1, \cdots, \mathbf{p}_m] \in \mathbb{R}^{m \times m}$ is a normalized orthogonal matrix, and $\mathbf{\Lambda}_k = \mathrm{diag}\{\lambda_{1,k}, \cdots, \lambda_{m,k}\}$ is a diagonal matrix. The columns of $\mathbf{P}_k$, denoted as $\mathbf{p}_i$, are eigenvectors of $\mathbf{C}_k^{\mathbf{U}}$, or called principal components. The diagonal entities $\lambda_{i,k}$ of Λ_k are corresponding eigenvalues of $\mathbf{C}_k^{\mathbf{U}}$. Based $\mathbf{P}_k$, we have a new matrix $\mathbf{Y}_k = [\mathbf{y}_1^{\mathrm{T}}, \cdots, \mathbf{y}_{n_k}^{\mathrm{T}}]^{\mathrm{T}} := \bar{\mathbf{U}}_k \mathbf{P}_k$, which is a new representation of the uncertainty samples. $\mathbf{y}_i := \mathbf{P}_k^{\mathrm{T}}(\mathbf{u}_i - \boldsymbol{\beta}_k)$ represents the projection of the centered data sample on the principle components. For $\mathbf{Y}_k$, its covariance matrix is $\mathbf{C}_k^{\mathbf{Y}} := \frac{1}{n_k - 1} \mathbf{Y}_k^{\mathrm{T}} \mathbf{Y}_k = \Lambda_k$, which implies that the components of $\mathbf{y}_i$ are uncorrelated since $\mathbf{\Lambda}_k$ is a diagonal matrix. The property of having uncorrelated features in the transformed data set $\mathbf{Y}_k$ allows for the adoption of basic sets, e.g., box, to construct uncertainty subsets for each cluster $\mathbf{U}_k$ to simplify the complexity of modeling.

5

5.2.4. UNCERTAINTY SETS CONSTRUCTION

In this subsection, we will apply the ML techniques detailed previously to construct a data-driven uncertainty set.

Algorithm 5.1 Data-driven uncertainty sets construction

Input: training data set $\mathbf{U}_{\text{train}}$
Output: uncertainty sets $\mathcal{U}_1, \cdots, \mathcal{U}_K$
Parameters: ϵ, $MinPts$, K

Extreme Uncertainties Removal

1: select values of ϵ and $MinPts$ for DBSCAN
2: apply DBSCAN to $\mathbf{U}_{\text{train}}$ to remove extreme scenarios and generate a cleaned data set $\mathbf{U}_{\text{clean}}$

Uncertainty Samples Clustering

3: select the parameter K
4: apply GMM clustering to $\mathbf{U}_{\text{clean}}$ to generate K clusters $\{\mathbf{U}_1, \cdots, \mathbf{U}_K\}$

Box-like Subsets Construction

5: apply PCA to $\mathbf{U}_k$ to generate $\mathbf{Y}_k$, $\mathbf{P}_k$ and $\mathbf{\Lambda}_k$
6: construct uncertainty sets via (5.4) and (5.5)

Given the training set of uncertainties $\mathbf{U}_{\text{train}}$, implementing DBSCAN and GMM gives K data clusters $\{\mathbf{U}_1, \cdots, \mathbf{U}_K\}$ to be processed by PCA. For each uncertainty cluster $\mathbf{U}_k$, following the PCA process introduced in Section 5.2.3 yields a new representation $\mathbf{Y}_k$ and the mean vector $\boldsymbol{\beta}_k$. Based on the relationship $\mathbf{y}_i = \mathbf{P}_k^{\mathrm{T}}(\mathbf{u}_i - \boldsymbol{\beta}_k)$, for each data cluster

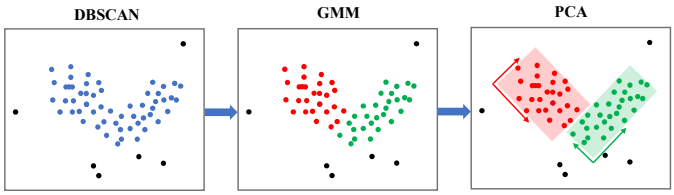


Figure 5.1: Graphical illustration of Algorithm 5.1.

$\mathbf{U}_k$, we construct the following uncertainty set:

$$\mathcal{U}_k = \left\{ \mathbf{u} \middle| \begin{array}{l} \mathbf{u} = \boldsymbol{\beta}_k + \mathbf{P}_k \mathbf{w}, \\ \forall \mathbf{w} : \ \underline{\mathbf{y}}_k \leq \mathbf{w} \leq \bar{\mathbf{y}}_k \end{array} \right\} \tag{5.4}$$

with $\underline{\mathbf{y}}_k := \min\{\mathbf{Y}_k\} \in \mathbb{R}^m$ and $\bar{\mathbf{y}}_k := \max\{\mathbf{Y}_k\} \in \mathbb{R}^m$, where operators min and max are performed columnwise. Since there are K uncertainty clusters, the final uncertainty set $\mathcal{U}$ is the union of K subsets as defined in (5.4). By reformulating the subset in (5.4) as linear constraints, the uncertainty set $\mathcal{U}$ can then be compactly written as

$$\mathcal{U} := \bigcup_{k=1}^{K} \mathcal{U}_k, \quad \mathcal{U}_k := \{\mathbf{u} | \mathbf{D}_k \mathbf{u} \leq \mathbf{d}_k\}, \tag{5.5a}$$

$$\mathbf{D}_k = \begin{bmatrix} \mathbf{P}_k^{\mathrm{T}} \\ -\mathbf{P}_k^{\mathrm{T}} \end{bmatrix}, \quad \mathbf{d}_k = \begin{bmatrix} \bar{\mathbf{y}}_k + \mathbf{P}_k^{\mathrm{T}} \boldsymbol{\beta}_k \\ -\underline{\mathbf{y}}_k - \mathbf{P}_k^{\mathrm{T}} \boldsymbol{\beta}_k \end{bmatrix}. \tag{5.5b}$$

Finally, the construction of the data-driven uncertainty set via DBSCAN, GMM and PCA can be expressed as Algorithm 5.1, and the corresponding graphical illustration of each component of Algorithm 5.1 is depicted in Fig. 5.1.

Remark 5.1. Instead of constructing box uncertainty subset via (5.5), another natural option of uncertainty subset based on Algorithm 5.1 is ellipsoid. By implementing steps 1 - 4 of Algorithm 5.1, the following ellipsoidal subset can be constructed

$$\mathcal{U}_k = \left\{ \mathbf{u} \mid ||\boldsymbol{\Sigma}_k^{-1/2} (\mathbf{u} - \boldsymbol{\mu}_k)||_2 \leq \varrho_k \right\} \tag{5.6a}$$

$$\varrho_k := \max_{\mathbf{u}_i \in \mathbf{U}_k} ||\boldsymbol{\Sigma}_k^{-1/2} (\mathbf{u}_i - \boldsymbol{\mu}_k)||_2. \tag{5.6b}$$

Compared with the box subset, the ellipsoidal subset might be more suitable to the Gaussian distributed uncertainty clusters. However, the robust counterpart (RC) of a linear RO problem with ellipsoidal uncertainty sets is a second-order cone program (SOCP). In contrast, the robust counterpart of a linear RO problem with a box uncertainty set remains a linear program (LP), which is supported by more off-the-shelf solvers and can be solved more efficiently, even for significantly large-scale problems, than SOCP. One advantage of the ellipsoidal uncertainty subset is that the robust counterpart problem does not introduce any extra constraints and decision variables to ensure robust constraint satisfaction. Conversely, for box uncertainty sets, additional constraints and decision variables are introduced in the RC problem, and their numbers are proportional

to the number of constraints defining the uncertainty set and the dimension of the uncertainty, respectively. Therefore, the choice between box and ellipsoidal uncertainty sets should depend on the type of available numerical solvers, the dimension of the uncertainty as well as the size of the RO problem to be solved. In our work, we focus on box uncertainty subsets, but all analysis for box subsets is also applicable to ellipsoidal subsets constructed via (5.6).

Remark 5.2. The enhanced performance of the proposed data-driven uncertainty set stems from the systematic integration of ML techniques such as DBSCAN, GMM and PCA. DBSCAN allows the approach to filter out low-probability uncertainties, independent of their locations. In contrast, many existing methods, such as KDE and K-SVC, primarily exclude extreme uncertainties located at the boundary of the uncertainty cluster. Moreover, the ability of the PCA-based uncertainty subset in (5.5) to reduce conservatism is predicated on the assumption that the uncertainty variables are correlated and Gaussian-distributed. Unlike the centroid-based or density-based clustering approaches used in other existing literature, which fail to ensure this implicit assumption, the GMM clustering approach adopted in Algorithm 5.1 generates Gaussian-distributed data clusters that boost the performance of the subsequent PCA-based subsets.

Remark 5.3. It should be pointed out that the feature of the proposed uncertainty set representation as a union of several basic convex subsets enables a flexible and compact non-convex uncertainty set without sacrificing the computational tractability of the resulting RO problem. On the one hand, the convex subset will guarantee the computational tractability of the resulting RO problem. On the other hand, unifying several subsets increases the flexibility in dealing with irregular uncertainty distributions and reduces the conservatism of the uncertainty set. In comparison, several existing representative data-driven approaches for constructing the uncertainty set, such as PCA-KDE, K-SVC in [140] and [35], try to find a single convex uncertainty set to ensure computational tractability but with sacrificed compactness or a single nonconvex set, such as DNN in [137], to ensure compactness but with high computational demand. In addition, the involved ML techniques – DBSCAN, GMM and PCA – are available in many ML toolboxes, such as `scikit-learn`, and our proposed Algorithm 5.1 can be easily implemented. A Python-based toolkit for implementing our data-driven uncertainty set is developed in [146].

Remark 5.4. Since DBSCAN and GMM are used to generate data clusters for constructing uncertainty subsets, the performance of our proposed uncertainty set is directly influenced by the effectiveness of these clustering methods. This is particularly challenging with high-dimensional uncertainty data, where visual evaluation of clustering quality and the resulting uncertainty set is not feasible. Additionally, high-dimensional data can lead to sparse training data, which can further deteriorate the performance of the ML approaches involved. Moreover, in data-driven RO problems, the absence of labeled data also complicates the validation and testing of the performance of data clustering and the resulting uncertainty set. All these factors make it a complex and nontrivial task to select proper parameters/hyperparameters, such as $(\epsilon, MinPts, K)$, in the context of high-dimensional uncertainty data.

5.2.5. PERFORMANCE ANALYSIS AND UNCERTAINTY SETS CALIBRATION

In this subsection, guidelines for tuning the parameters of Algorithm 5.1 and evaluating the resulting data-driven uncertainty set are provided. In particular, a probabilistic bound of the data coverage for out-of-sample uncertainties is introduced, which can be used to establish a probabilistic guarantee of the out-of-sample performance of the corresponding RO problem.

In Algorithm 5.1, there are three parameters to be selected: ϵ, $MinPts$ and K. A rule of thumb for selecting $MinPts$ is $MinPts \geq m+1$, where m is the dimension of the uncertainty. With a fixed value of $MinPts$, ϵ determines how many uncertainty samples are excluded from the uncertainty set. A smaller ϵ results in fewer low-probability uncertainties included in the uncertainty set. Thus, ϵ can be chosen based on the desired proportion of low-probability uncertainties to be excluded from the uncertainty set. The parameter K is the number of clusters for GMM clustering, which affects both the conservatism and the complexity of the resulting data-driven uncertainty sets. While a larger K tends to yield a more compact uncertainty set, it increases the overall complexity as the total number of constraints defining the uncertainty set is proportional to K, and may bias the uncertainty set towards the training samples.

It is important to note that we assume there are no labeled data to evaluate the clustering performance, unlike typical ML data clustering tasks focusing on high clustering accuracy. Also, accurately clustered uncertainty data do not necessarily result in a favorable uncertainty set for corresponding RO problems. Since the complexity of the uncertainty set significantly influences the computational effectiveness of the corresponding RO problem, the balance between its conservatism and complexity must be considered when selecting the design parameters of Algorithm 5.1.

A suggested sequence of parameter selection for Algorithm 5.1 is as follows. Firstly, begin by setting $MinPts$ based on the rule of thumb, which is at least one more than the dimension of the uncertainty. Next, adjust ϵ to achieve the desired percentage of data coverage for the uncertainty set. Finally, determine an appropriate value of K, balancing the conservatism and complexity of the uncertainty set.

For 2-D or 3-D uncertainty data, we can rely on visual inspection to find out the proper value of design parameters. However, for high-dimensional uncertainty data, visualization of the clustering results and the corresponding uncertainty set is not applicable. As a result, we should rely on some quantitative indicators to support the selection of design parameters. In the ML community, there are several available metrics to evaluate the clustering performance, such as Silhouette Score, Calinski-Harabasz index, and Davies-Bouldin index, which are readily available in well-developed ML packages such as `scikit-learn`. In this work, we choose the Silhouette Score as an example to select the number of uncertainty subsets.

The silhouette Score, whose value belongs to $[-1, 1]$, measures how similar each data is to the cluster it belongs to and how different it is from other remaining clusters. A higher score means dense and well-separated clusters and scores around zero indicate overlapping clusters. As a result, within an acceptable range of K, the value with a larger Silhouette Score is preferable. However, it should be noted that selecting K purely based on such metrics is not wise because the complexity and the conservatism of the uncertainty set need to be balanced.

Recall that the larger the value of K, the more complex the uncertainty set becomes, consequently reducing its conservatism. The level of conservatism of the uncertainty set can be reflected by its data coverage, typically, a less conservative set tends to have a lower coverage. For our proposed Algorithm 5.1, the lower bound of data coverage in the training set is determined by the parameters ϵ and $MinPts$. Therefore, a desirable K should balance a high Silhouette Score, a low training data coverage, and maintain a relatively small value to manage complexity.

Once an uncertainty set is constructed via Algorithm 5.1, another aspect that needs to be evaluated is its consistency in both training and validation/testing sets. Consistency means that the uncertainty sets should achieve consistent data coverage in both training and validation/testing sets so that the uncertainty set is not biased towards the training set. If the data coverage percentage of the training set is much larger than that of the testing set, it implies that the uncertainty sets are biased towards the training data and one might need to reduce the number of data clusters K to decrease the complexity of the uncertainty set.

Furthermore, based on the data coverage in the testing set, the following lemma is applicable to provide a probabilistic bound of data coverage for any I.I.D. uncertainty scenarios. The data coverage, denoted as ρ, is defined as the portion of data that is included in the uncertainty set. $\rho = 1$ means all data are covered by the uncertainty set, and $\rho = 0$ means no data is included in the uncertainty set.

Lemma 5.1. For n I.I.D. samples of testing scenarios $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$, assuming that the data coverage of the uncertainty sets (5.5) for testing data is ρ. Then, for any random sample of uncertainty $\mathbf{u}$, the probability that $\mathbf{u} \in \mathcal{U}$, denoted as $\mathbb{P}_{u\in\mathcal{U}}$, satisfies

$$\mathbb{P}(\mathbb{P}_{\mathbf{u}\in\mathcal{U}} \leq \rho - \tau) \leq \exp(-2n\tau^2) \tag{5.7}$$

with $\tau \geq 0$.

Proof: Define a random variable $\delta(\mathbf{u} \in \mathcal{U})$ as the indicator function of the random event $\mathbf{u} \in \mathcal{U}$. Since the testing scenarios are I.I.D., $\delta(\mathbf{u}_i \in \mathcal{U})$ are n I.I.D. samples of $\delta(\mathbf{u} \in \mathcal{U})$ with $\frac{1}{n}\sum_{i=1}^{n} \delta(\mathbf{u}_i \in \mathcal{U}) = \rho$. Consequently, it follows from Hoeffding's inequality that the inequality (5.7) holds. □

In addition to adopting Lemma 5.1 to construct a probabilistic guarantee of uncertainty coverage for the proposed uncertainty set, the calibration approach using order statistics calculation introduced in [136] is also applicable to our proposed data-driven uncertainty set to develop probabilistic guarantees of uncertainty coverage.

5.3. ROBUST OPTIMIZATION DESIGN AND PERFORMANCE GUARANTEES

In this section, we consider a linear two-stage RO problem with the proposed data-driven uncertainty set. The RO problem investigated in this work has the following struc-

ture:

$$\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \max_{\mathbf{u}\in\mathscr{U}} \min_{\mathbf{z}} \mathbf{b}^{\mathrm{T}}\mathbf{z} \tag{5.8a}$$

$$\text{s.t. } \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{z} + \mathbf{M}\mathbf{u} \le \mathbf{h}, \tag{5.8b}$$

$$\forall \mathbf{u} \in \mathscr{U} \tag{5.8c}$$

where $\mathbf{x} \in \mathbb{R}^s$ is the first-stage decision variables, $\mathbf{z} \in \mathbb{R}^q$ is the second-stage decision variables, or called recourse decision variables, $\mathbf{u} \in \mathbb{R}^m$ is the uncertainty. The uncertainty set $\mathscr{U}$ is assumed to be constructed via our proposed approach, and can be written in the following format

$$\mathscr{U} := \bigcup_{k=1}^{K} \mathscr{U}_k, \quad \mathscr{U}_k := \{\mathbf{u} | \mathbf{D}_k \mathbf{u} \le \mathbf{d}_k\}. \tag{5.9}$$

The row number of $(\mathbf{T}, \mathbf{W}, \mathbf{M})$ is denoted as r, $(\mathbf{c}, \mathbf{b}, \mathbf{h})$ and $(\mathbf{T}, \mathbf{W}, \mathbf{M})$ are parameter vectors and matrices with appropriate dimensions, respectively.

Optimization problem (5.8) is semi-infinite and is computationally intractable since there are infinitely many constraints to be satisfied. To achieve a balance between computational burden and optimality, we adopt the following affine decision rule for the recourse decision variable $\mathbf{z}$:

$$\mathbf{z} = \mathbf{L}_k \mathbf{u} + \mathbf{g}_k, \ \forall \mathbf{u} \in \mathscr{U}_k \tag{5.10}$$

where $\mathbf{L}_k \in \mathbb{R}^{q\times m}$ and $\mathbf{g}_k \in \mathbb{R}^q$ are decision variables to be optimized. Unlike the conventional decision strategy where only a single decision rule is imposed for all possible uncertainties in the uncertainty set, such as the works in [15], the proposed uncertainty set makes it possible to assign different decision policies (5.10) for uncertainties residing in different sets. By exploiting the feature of the proposed uncertainty set for unifying several subsets, it is possible to reduce the conservatism of the optimal RO solution.

Theorem 5.1. For the two-stage robust optimization problem (5.8), assuming that the uncertainty set is defined as (5.9), and the recourse decision variables are determined via (5.10), the optimal solution can be computed by solving (5.11).

$$\min_{\substack{\mathbf{x},\eta,\mathbf{L}_k,\mathbf{g}_k \\ \boldsymbol{\pi}_{0,k},\boldsymbol{\pi}_{i,k}}} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \eta \tag{5.11a}$$

$$\text{s.t. } \mathbf{b}^{\mathrm{T}}\mathbf{g}_k + \boldsymbol{\pi}_{0,k}^{\mathrm{T}}\mathbf{d}_k \le \eta, \tag{5.11b}$$

$$\mathbf{D}_k^{\mathrm{T}}\boldsymbol{\pi}_{0,k} = \mathbf{L}_k^{\mathrm{T}}\mathbf{b}, \tag{5.11c}$$

$$[\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{g}_k - \mathbf{h}]_i + \boldsymbol{\pi}_{i,k}^{T}\mathbf{d}_k \le 0, \tag{5.11d}$$

$$\mathbf{D}_k^{\mathrm{T}}\boldsymbol{\pi}_{i,k} = [\mathbf{W}\mathbf{L}_k + \mathbf{W}]_i, \tag{5.11e}$$

$$\boldsymbol{\pi}_{0,k} \ge 0, \ \boldsymbol{\pi}_{i,k} \ge 0, \tag{5.11f}$$

$$\forall k = 1, \cdots, K, \quad \forall i = 1, \cdots, r. \tag{5.11g}$$

In addition, if the data coverage for n I.I.D. testing uncertainty samples is ρ, the probability that the solution of (5.11) can ensure constraint satisfaction for a random uncertainty $\mathbf{u}$ is greater than $\rho - \tau$ with confidence at least $1 - \exp(-2n\tau^2)$.

Proof: The proof of this theorem is presented in the Appendix 5.A.1.

Remark 5.5. Theorem 5.1 presents a computationally efficient approximation of (5.8) by imposing affine decision policy for the recourse decision variable and solving a linear program. While the uncertainty set in (5.5) is nonconvex, the feature that the uncertainty set is a union of several basic convex subsets ensures the computational efficiency of the corresponding RO problem. In addition, since separate decision rules are applied for each subset, the optimal solution is expected to be less conservative than the typical RO solution with a single uncertainty set and a single decision rule, such as [11, 15, 35, 141]. In addition, beyond the conventional RO solutions, a probabilistic guarantee of constraint satisfaction for out-of-sample uncertainties is provided based on the performance testing of the uncertainty set.

5.4. APPLICATIONS

In this section, the performance of our proposed method is compared with several representative data-driven RO approaches. Specifically, we focus on the following approaches listed in Table 5.1.

Table 5.1: Approaches considered in the case studies.

Approach	Description
Box	box uncertainty set constructed based on the minimal and maximal value of each data dimension.
CH	convex hull of uncertainty data introduced in [139].
K-SVC	kernel-based support vector clustering method proposed in [11, 35].
KPKDE	KMeans-based clustering combined with PCA and KDE approach proposed in [62].
B-DGP	our proposed Algorithm 1 using box subsets in (5.5).
E-DG	one variant of Algorithm 1 where ellipsoidal uncertainty subsets are constructed via (5.6).

The performance of the above approaches is assessed through three case studies. For the first two case studies, the performance of the uncertainty sets is compared in the following aspects: area of uncertainty set, complexity, and computation time. The complexity of polyhedral uncertainty sets in our work refers to the number of linear constraints defining the uncertainty set. This metric, which is generally neglected in the existing literature, is crucial in influencing the computational efficiency of the corresponding RO problem. For a linear RO problem with a polyhedral uncertainty set, the number of decision variables and constraints in its robust counterpart problem is proportional to the number of constraints defining the uncertainty sets. As a result, a high complexity of the uncertainty set leads to increased computational demands for solving the corresponding RO problem. For the last case study, the performance of the proposed approach is assessed via an optimal building climate control problem.

In our upcoming case studies, since K-SVC, KPKDE, B-DGP and E-DG can exclude some uncertainty samples to reduce the conservatism of the corresponding uncertainty sets, the design parameters of these approaches are chosen to exclude about 5% extreme

uncertainty scenarios to reduce the conservatism of the resulting data-driven uncertainty sets. Namely, for the K-SVC approach, the parameter ν is set as 0.05; for the KPKDE approach, the confidence level of each feature is selected as $[(1-0.95^{\frac{1}{m}})/2, (1+0.95^{\frac{1}{m}})/2]$; for the B-DGP and E-DG approaches, the parameter ϵ is adjusted to identify about 5% of the total uncertainty samples as outliers. With the same data coverage, the approach with a smaller set size indicates a less conservative uncertainty set.

All simulations are implemented on an Intel Xeon W-2223 CPU at 3.6GHz with 16G RAM. Optimization problems are modeled using Python package `gurobipy` and solved via `Gurobi 11.0`. The involved ML methods – KDE, DBSCAN, GMM and PCA – are implemented via the Python package `scikit-learn 1.0.2`.

5.4.1. CASE STUDY 1: SYNTHETIC UNCERTAINTY DATA

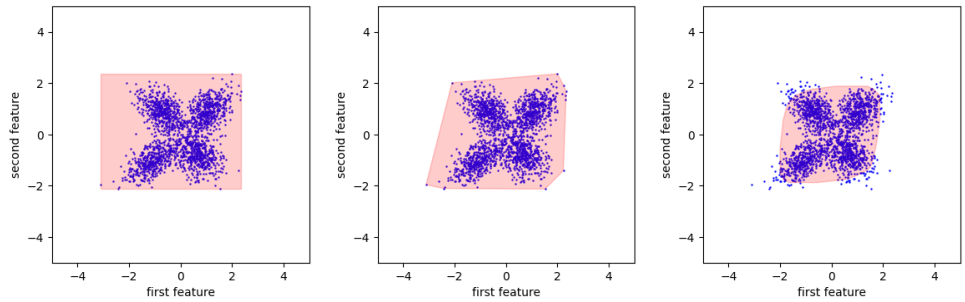
In this case study, we test the performance of the data-driven approaches listed in Table 5.1 for constructing uncertainty sets with synthetic data. A common assumption about uncertainties is that they are Gaussian distributed. In reality, due to different working conditions simultaneously being represented in the dataset, the uncertainties may follow different Gaussian mixture distributions. In order to reflect this, we use synthetic uncertainty data that are generated from 4 different two-dimensional Gaussian distributions with 500 scenarios per distribution.

The uncertainty sets with different data-driven approaches are presented in Fig. 5.2, where the uncertainty sets are the shaded regions. The performance of each data-driven approach is summarized in Table 5.2. It can be observed that among all approaches, the Box approach gives the most conservative uncertainty set since it does not extract and exploit the hidden patterns of the uncertainty data. For the CH approach, it is non-parametric and does not entail any computation to construct the uncertainty set, see [139] for more details, so that its computation time is absent. However, its non-parametric nature also leads to a very high complexity. Similarly, while the K-SVC approach can find a compact uncertainty set by precluding some extreme scenarios, its non-parametric nature also incurs a high complexity. In addition, the convex nature of the K-SVC-induced uncertainty set limits its flexibility in dealing with general nonconvex uncertainty distributions. As indicated in [35], the number of the support vectors, which determines the complexity of the uncertainty set developed by the K-SVC approach, is inversely proportional to the conservatism of the resultant uncertainty set. Hence, with the K-SVC approach, a less conservative uncertainty set implies an uncertainty set with higher complexity. Furthermore, Table 5.2 shows that the K-SVC approach takes much longer computation time than the others since it has to solve a large-scale quadratic programming problem. In the KPKDE approach, KMeans-based clustering fails to produce suitable uncertainty clusters for subsequent PCA and KDE processes, resulting in a relatively conservative uncertainty set. This issue arises because the KMeans algorithm groups data based solely on distances, without considering the data distribution within the same cluster, whereas PCA implicitly assumes a Gaussian data distribution. This discrepancy limits the efficiency of the KPKDE approach. In contrast, our proposed methods B-DGP and E-DG, which combine GMM and PCA, generate more appropriate data clusters and thus yield more compact uncertainty sets. Additionally, the utilization of DBSCAN ensures our proposed uncertainty sets are immune to the influence of unlikely

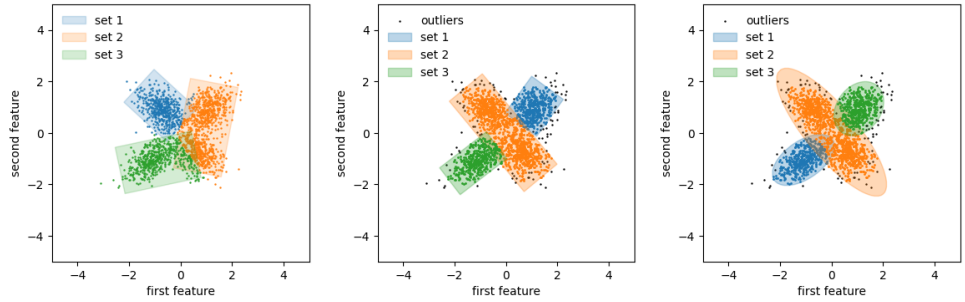
uncertainties. As shown in Table 5.2, the B-DGP approach gives a less conservative uncertainty set compared to the E-DG approach. This is because, in the E-DG approach, while the ellipsoid subset is more suitable for Gaussian distributed data, its symmetric structure and correlated features may also increase its conservatism to some extent.

Table 5.2: Performance summary of different uncertainty sets in Case Study 1.

	existing approaches				our approaches	
	Box	CH	K-SVC	KPKDE	B-DGP	E-DG
uncertainty set size	24.40	20.66	13.10	14.77	12.74	14.26
complexity (# of linear constraints)	4	4001	421	12	12	-
data coverage	1	1	0.95	0.98	0.97	0.98
computation time (s)	< 0.01	-	59.27	0.42	0.62	1.10



(a) Box (b) CH (c) K-SVC



(d) KPKDE (e) B-DGP (f) E-DG

Figure 5.2: Data-driven uncertainty sets using synthetic data in Case Study 1.

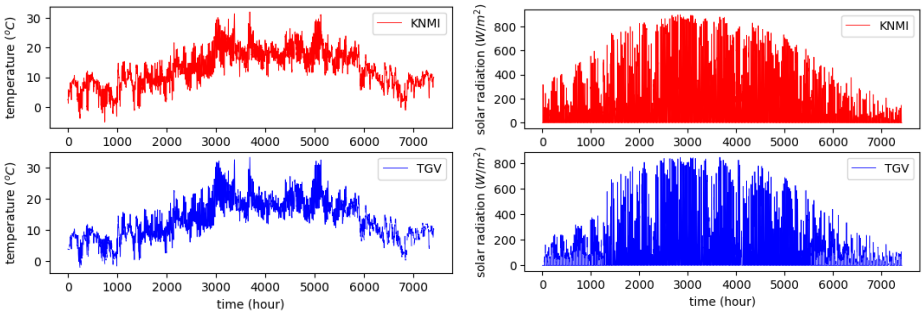
5.4.2. CASE STUDY 2: REAL-WORLD WEATHER DATA

In this case study, the approaches listed in Table 5.1 are evaluated using real-world weather data. The issue of weather uncertainties is common in many applications, such as building climate control, renewable energy management, greenhouse control,

etc. How to properly construct the uncertainty set of weather conditions is important in these application problems.

This case study considers the uncertainties in ambient temperature and solar radiation. The utilized weather data are measured during Jan. 2023 - Dec. 2023 from two weather stations in the Netherlands: 1) the weather station 344 of Koninklijk Nederlands Meteorologisch Instituut (KNMI), a scientific institute of the Dutch government, located in Rotterdam, and 2) the weather station located in The Green Village (TGV) at TU Delft, which is about 7km from the KNMI station. The profiles of corresponding weather data are shown in Fig. 5.3. The weather data from KNMI are utilized as predicted weather conditions, and the data from TGV are used as real local conditions. There are 7416 data points in total. For the sake of visualization, the weather uncertainties are scaled so that the maximum absolute value of each feature is 1, and the scaled data are used for developing uncertainty sets.

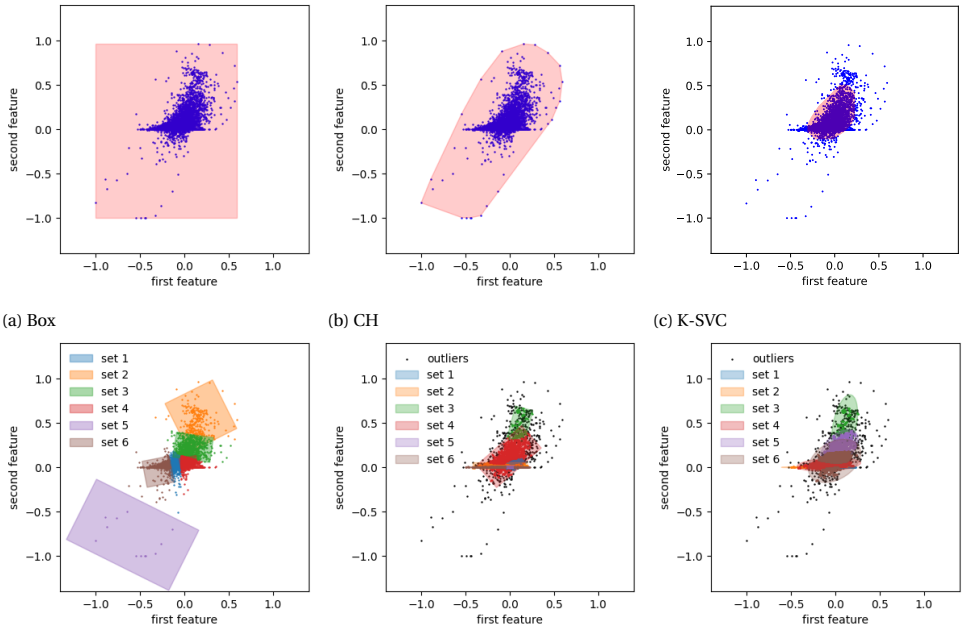
The uncertainty sets with different data-driven approaches are shown in Fig. 5.4 and Table 5.3. From Fig. 5.4, it is clear that the Box approach is the most conservative since no latent feature of the data is utilized. The uncertainty set via the CH approach is also very conservative due to the inclusion of low-probability uncertainty samples. For the K-SVC approach, while the size of the uncertainty set is small, it misses some high-probability scenarios located near the boundary of the uncertainty distribution for the sake of a convex uncertainty set. Also, as shown in Table 5.3, the K-SVC approach takes much longer computation time than the others and incurs a high complexity. While the KPKDE approach can exclude low-probability uncertainties to reduce the conservatism of the uncertainty set, its performance for this irregular uncertainty distribution is far from satisfactory. On the one hand, the KMeans method fails to generate suitable data clusters for subsequent PCA-KDE-based uncertainty set construction. On the other hand, the adoption of KDE fails to remove low-probability uncertainty samples, especially for the uncertainty cluster 5 in Fig. 5.3(d) where uncertainty samples are sparsely distributed. Finally, it can be observed from Fig. 5.4(e) and 5.4(f) as well as Table 5.3 that our proposed B-DGP and E-DG approaches are more versatile than the others in dealing with irregular uncertainty distributions and achieve a notable balance between complexity and conservatism in constructing the uncertainty set.



(a) Ambient temperature profiles.

(b) Solar radiation profiles.

Figure 5.3: Weather data from KNMI and TGV in Case Study 2.



(a) Box (b) CH (c) K-SVC

(d) KPKDE (e) B-DGP (f) E-DG

Figure 5.4: Data-driven uncertainty sets using real weather data in Case Study 2.

Remark 5.6. It should be pointed out that, based on the results of our case studies, the complexity of the uncertainty set constructed via K-SVC can be remarkably influenced by numerical errors. The K-SVC method relies on support vectors to construct the uncertainty set. Based on the value of corresponding Lagrangian multipliers α_i, all uncertainty samples are classified into 3 categories: support vectors with $0 < \alpha_i < \frac{1}{n\nu}$, outliers with $\alpha_i = \frac{1}{n\nu}$ and interior points with $\alpha_i = 0$. Due to the existence of numerical errors, all uncertainty samples may be identified as support vectors, which will dramatically increase the complexity of the resulting RO problem due to the non-parametric property of the uncertainty set. To mitigate the influence of numerical errors, in our case studies, the support vectors are selected based on $0+\varepsilon < \alpha_i < \frac{1}{n\nu} - \varepsilon$, where $\varepsilon > 0$ is a sufficiently small

Table 5.3: Performance summary of different uncertainty sets in Case Study 2.

	existing approaches				our approaches	
	Box	CH	K-SVC	KPKDE	B-DGP	E-DG
uncertainty set size	3.12	1.66	0.25	1.56	0.28	0.34
complexity (# of linear constraints)	4	14833	1493	24	24	-
data coverage	1	1	0.95	0.98	0.98	0.98
computation time (s)	< 0.01	-	1745.85	2.72	2.95	3.44

constant ($\varepsilon = 1 \times 10^{-8}$ in our case studies). The choice of ϵ might strongly influence the number of support vectors and hence the complexity of the uncertainty set. In addition, as shown in [35], the number of the support vectors is proportional to the parameter ν, which implies a less conservative uncertainty set derived via the K-SVC approach has more complexity, and consequently higher computational burden for solving the corresponding RO problem. Furthermore, from Tables 5.2 and 5.3, it can be observed that the computation time of K-SVC is much larger than the others since it entails solving a large-scale QP.

Remark 5.7. With the KPKDE approach, uncertainty samples are first clustered using KMeans. Then, the samples within each cluster are processed using the PCA method. Subsequently, an uncertainty subset is constructed based on the confidence intervals derived from the approximated cumulative density function (CDF) of the PCA-processed data, tailored to a predefined confidence level. As can be seen in Fig. 5.2, the KMeans clustering used in the KPKDE approach can result in unsuitable data clusters for constructing PCA-KDE-based uncertainty sets because PCA implicitly assumes Gaussian distributed data whereas KMeans fails to generate data clusters satisfying this assumption. Furthermore, the efficacy of the KPKDE method is sensitive to the choice of the hyperparameters: the kernel functions and associated bandwidth. Assuming an ideal approximation of the CDF, an uncertainty set formulated with a confidence level γ for each dimension of the PCA-processed data would give a data coverage γ^m, where m represents the uncertainty dimension. Nevertheless, the presence of an approximation error in the CDF, quantified as ϵ, might lead to an actual confidence level $\gamma + \varepsilon$. Consequently, the actual data coverage is $(\gamma+\epsilon)^m$. Given values of $\gamma = 0.98$, $\epsilon = 0.01$ and $m = 10$, the total error of data coverage is 0.087, which means 8.7% of the uncertainty samples are unexpectedly included in the uncertainty set.

Remark 5.8. The importance of using real-world datasets over simple synthetic datasets for performance evaluation lies in two main factors. Firstly, the assumption made when generating synthetic datasets may not accurately represent real-world conditions when applying the approach to practical problems. For instance, a common assumption in building climate control is that the weather uncertainties follow Gaussian distributions. However, as illustrated in Fig. 5.4, our real-world data is clearly non-Gaussian and is more complex to handle. Second, real-world data distributions can be far more intricate and irregular than synthetic ones. Consequently, data-driven approaches that perform well on synthetic datasets may experience significant performance degradation when applied to complex, real-world data. Therefore, conclusions drawn from synthetic datasets may not generalize to actual practical scenarios. Additional supporting materials, demonstrating that the data-driven approaches considered in this chapter are harder to distinguish from each other for simple synthetic datasets, are provided in Appendix 5.A.2 to further validate the aforementioned statement.

5.4.3. CASE STUDY 3: OPTIMIZATION-BASED BUILDING ENERGY CONTROL

In this case study, we will investigate the RO design for the optimization-based building energy control problem considered in [11]. The thermal dynamics of buildings are subject to several weather uncertainties, e.g., ambient temperature uncertainty. Prop-

erly considering these uncertainties in building energy control can improve occupants' comfort and reduce energy usage.

The building thermal dynamics is modeled as the following linear system

$$x_{t+1} = Ax_t + B_u u_t + B_v v_t + B_w w_t, \tag{5.12}$$

where x_t is the state vector consisting of indoor air temperature, wall temperature, roof temperature, and floor temperature; u_t is the heating power injection, v_t is the vector of the predicted value of ambient temperature and underground temperature, w_t is the prediction error of ambient temperature. The system matrices are

$$A = \begin{bmatrix} 0.0167 & 0.0048 & 0.1245 & 0.409 \\ 0.0005 & 0.0002 & 0.0039 & 0.0044 \\ 0.0253 & 0.0073 & 0.3321 & 0.0617 \\ 0.0244 & 0.0070 & 0.0526 & 0.3456 \end{bmatrix},$$

$$B_u = \begin{bmatrix} 0.0986 \\ 0.0029 \\ 0.0288 \\ 0.0275 \end{bmatrix}, B_v = \begin{bmatrix} 0.2536 & 0.4596 \\ 0.0070 & 0.9840 \\ 0.4450 & 0.1287 \\ 0.4477 & 0.1225 \end{bmatrix}, B_w = \begin{bmatrix} 0.2536 \\ 0.0070 \\ 0.4450 \\ 0.4477 \end{bmatrix}.$$

The control objective is to determine the heating power injection u_t over a finite time window to ensure indoor comfort constraints while reducing energy consumption in the presence of weather prediction errors. As a result, the finite-horizon optimal control problem can be formulated as

$$\min_{u_t} \sum_{t=0}^{H} l_t(x_t, u_t) \tag{5.13a}$$

$$\text{s.t. } x_{t+1} = Ax_t + B_u u_t + B_v v_t + B_w w_t, \tag{5.13b}$$

$$x_t \in \mathscr{X},\ u_t \in \mathscr{U}, \tag{5.13c}$$

$$\forall t = 0, 1, \ldots, H-1,\ \forall w_t \in \mathscr{W} \tag{5.13d}$$

where l_t is the stage cost function; $\mathscr{X}$ and $\mathscr{U}$ are the feasible sets of the states and thermal input, respectively; $\mathscr{W}$ is the uncertainty set of the ambient temperature, which are constructed via the data-driven approaches in Table 5.1; H is the length of the prediction horizon.

Similarly to [11], the feasible set of states is defined as keeping the indoor temperature above 21°C during peak occupied hours 7 am - 6 pm, and keeping the indoor temperature above 15°C for off-peak hours to avoid unnecessary energy usage. The length of the prediction horizon is $H = 10$. The admissible control input set is defined as $0 \le u_t \le 150$. The stage cost function is $l_t := u_t$ for minimizing energy consumption. The above finite horizon optimal control problem can be formulated as an RO problem in the format of (5.8). The thermal control input u_t can be regarded as the recourse decision variable, and the prediction errors of ambient temperature are uncertainties. While only one uncertainty factor – ambient temperature – is considered, the dimension of the uncertainty for the resulting RO problem is 10, which is equal to the length of the prediction horizon H, since all uncertainties within the prediction horizon need to be considered.

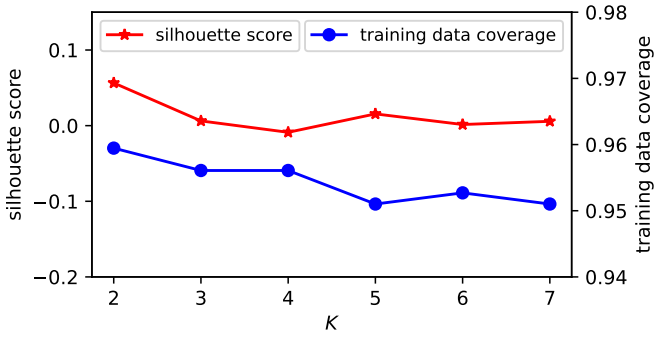


Figure 5.5: Silhouette score and data coverage percentage for different K values (# of clusters) in Case Study 3.

To ensure the causality of input, decision variable $\mathbf{L}_k$ is restricted to be strictly lower triangular. The control inputs within the prediction horizon are applied in an open-loop manner. Namely, the control signals are computed via the affine control policy (5.10) with the parameters $(\mathbf{L}_k, \mathbf{g}_k)$ updated every H time step.

The historical weather data used in this case is from KNMI in *Case Study 2*. 80% of the uncertainty samples are used for developing uncertainty sets, and the remaining 20% uncertainty data are used for implementing simulations of the robust optimization-based control design. The data of ambient temperature in Jan. and Feb., during which heating is needed, are used for simulations. The ground temperature is set as the annual average ambient temperature.

In this case study, we mainly focus on the following data-driven approaches: Box, K-SVC, KPKDE, B-DGP and E-DG. Unlike the previous case studies, this case study considers high-dimensional uncertainties. Hence evaluating the uncertainty sets visually by plotting the uncertainty sets is not suitable anymore. As introduced in Section 5.2.5, we rely on the silhouette score and data coverage measures to find a suitable value of K for Algorithm 5.1, namely the number of box subsets. To ensure the computational efficiency of the resulting RO problem, we restrict $K \in [2,7]$. The silhouette score and the corresponding training data coverage of the B-DGP-based uncertainty set with different K are plotted in Fig. 5.5. It can be observed that $K = 5$ achieves a balance between high silhouette score and low data coverage. Hence, we select $K = 5$. For the KPKDE approach, the number of data clusters is selected as $K = 2$, which is tuned based on the suggestion provided in [145] to yield the highest Calinski-Harabase index. To demonstrate the effectiveness of the proposed decision rule (5.10), the conventional decision rule where only a single decision is applied for all subsets within the proposed uncertainty set is also implemented.

The indoor temperature profiles with different data-driven uncertainty sets and our proposed decision rule are depicted in Fig. 5.6, and the corresponding control performance is summarized in Table 5.4. The conservatism of the uncertainty set and the solution quality of the corresponding RO problem are measured by the value of the average cost. Among all approaches, it can be observed from Table 5.4 that the K-SVC approach and our proposed approaches (B-DGP and E-DG) have relatively low average cost since

Table 5.4: Control performance summary of different data-driven uncertainty sets in Case Study 3.

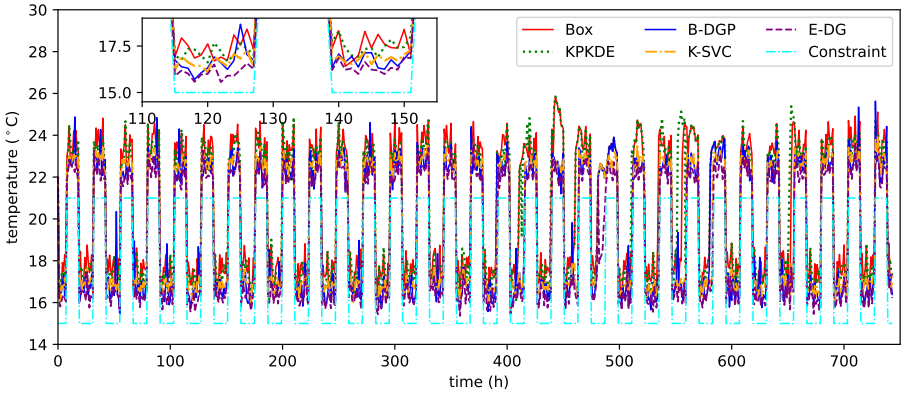
	existing approaches			our proposed approaches			
	Box	K-SVC	KPKDE	B-DGP	E-DG	B-DGP	E-DG
				proposed decision rule		conventional decision rule	
complexity (# of linear constraints)	20	721	40	100	-	100	-
data coverage	1	0.95	0.97	0.95	0.95	0.95	0.95
average cost (W/m^2)	99.62	92.21	98.87	91.29	87.50	86.59	86.30
total constraint violation (°C)	0	0	0	0	0	14.65	6.07
average computation time for solving (5.13) (s)	0.05	0.94	0.10	0.23	0.29	0.05	0.05

they can exclude some low-probability samples. In comparison, the KPKDE results in a much higher average cost (the second highest among all approaches), and consequently a more conservative uncertainty set. While theoretically, the KPKDE approach is also capable of removing low-probability uncertainty samples, its performance is far from satisfactory for this complicated real-world uncertainty distribution. As for the Box approach, it results in the highest average cost since it does not leverage any underlying pattern of the uncertainty data. For the K-SVC approach, while it achieves comparable average cost to our proposed B-DGP approach, its computational time is much longer (4 times greater) than that of the B-DGP approach. This is caused by the high complexity of the uncertainty set due to its non-parametric property. For our proposed B-DGP and E-DG approaches, both have relatively low average cost and short computation time, which means that a favorable balance between the complexity and conservatism of the uncertainty sets is achieved. Notably, the E-DG approach has the lowest average cost compared to all other approaches. It should be pointed out that the E-DG approach requires a SOCP solver, whereas B-DGP only needs an LP solver for solving (5.13).

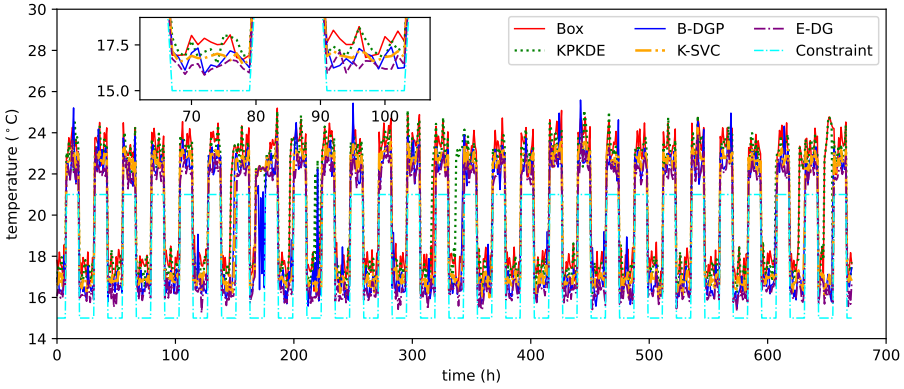
The last two columns of Table 5.4 show the simulation results for our proposed uncertainty sets when a conventional decision rule is applied that implements a single decision rule for all subsets. It can be seen that indoor comfort constraints are violated with the conventional decision rule while our proposed decision rule keeps the indoor temperature within the admissible range during the whole simulation period, which confirms that the proposed decision rule (5.10) is benefical for improving the robustness of the optimal solution.

5.5. CONCLUSIONS

This chapter proposes a novel approach to construct data-driven uncertainty sets, leveraging DBSCAN, GMM, and PCA techniques, for robust decision-making with uncertainties. The proposed approach is flexible in balancing the conservatism and complexity of the uncertainty set while demonstrating resilience to low-probability uncertainty scenarios regardless of the underlying uncertainty distributions. The influence of each design parameter is elucidated, and the performance of the proposed uncertainty set can



(a) Indoor temperature profiles in Jan. 2023 for Case Study 3.



(b) Indoor temperature profiles in Feb. 2023 for Case Study 3.

Figure 5.6: Indoor temperature profiles with different data-driven uncertainty sets in Case Study 3.

be systematically analyzed with the provided guidelines. By adopting well-established ML packages `scikit-learn`, a Python-based toolkit for conveniently and efficiently implementing our proposed data-driven uncertainty set is provided. Furthermore, for a linear two-stage RO problem, a tailored solution with the proposed uncertainty set is derived with a probabilistic guarantee of constraint satisfaction for out-of-sample uncertainties, enhancing the confidence of applicability over conventional RO solutions. Comparative analyses with several existing uncertainty set construction methods highlight the superiority of our methodology in striking a balance between computational efficiency and robustness of hedging against uncertainties.

Future works include designing novel RO formulations to exploit the representation of the proposed data-driven uncertainty set – a union of basic subsets – to further reduce the conservatism typically associated with conventional RO formulations.

5.A. APPENDIX

5.A.1. PROOF OF THEOREM. 5.1

Proof: Substituting the decision policy (5.10) into (5.8) and considering uncertainty sets (5.9) leads to

$$\min_{\mathbf{x},\mathbf{L}_k,\mathbf{g}_k} \left\{ \mathbf{c}^{\mathrm{T}}\mathbf{x} + \max_{1\le k\le K} \max_{\mathbf{u}\in\mathscr{U}_k} \mathbf{b}^{\mathrm{T}}(\mathbf{L}_k\mathbf{u}+\mathbf{g}_k) \right\} \tag{5.14a}$$

$$\text{s.t. } \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{g}_k + (\mathbf{W}\mathbf{L}_k + \mathbf{M})\mathbf{u} \le \mathbf{h}, \tag{5.14b}$$

$$\forall \mathbf{u}\in\mathscr{U}_k,\ \forall k = 1,\cdots,K. \tag{5.14c}$$

The universal constraint satisfaction in (5.14) can be alternatively reformulated as the following worst-case constraint satisfaction

$$\min_{\mathbf{x},\mathbf{L}_k,\mathbf{g}_k} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \eta \tag{5.15a}$$

$$\text{s.t. } \max_{\mathbf{u}\in\mathscr{U}_k} \{\mathbf{b}^{\mathrm{T}}\mathbf{L}_k\mathbf{u}\} + \mathbf{b}^{\mathrm{T}}\mathbf{g}_k \le \eta, \tag{5.15b}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{g}_k + \max_{\mathbf{u}\in\mathscr{U}_k} \{(\mathbf{W}\mathbf{L}_k + \mathbf{M})\mathbf{u}\} \le \mathbf{h}, \tag{5.15c}$$

$$\forall k = 1,\cdots,K. \tag{5.15d}$$

For the worst-case constraint satisfaction in (5.15b), following a standard procedure in RO [8], the maximization problem $\max_{\mathbf{u}\in\mathscr{U}_k}\{\mathbf{b}^{\mathrm{T}}\mathbf{L}_k\mathbf{u}\}$ in (5.4f) can be translated into its dual problem:

$$\min_{\boldsymbol{\pi}_{0,k}} \mathbf{d}_k^{\mathrm{T}}\boldsymbol{\pi}_{0,k} \tag{5.16a}$$

$$\text{s.t. } \mathbf{D}_k^{\mathrm{T}}\boldsymbol{\pi}_{0,k} = \mathbf{L}_k^{\mathrm{T}}\mathbf{b}, \quad \boldsymbol{\pi}_{0,k} \ge 0 \tag{5.16b}$$

where $\boldsymbol{\pi}_{0,k}$ is the Lagrangian multiplier. Consequently, constraint (5.15b) can be relaxed as

$$\mathbf{d}_k^{\mathrm{T}}\boldsymbol{\pi}_{0,k} + \mathbf{b}^{\mathrm{T}}\mathbf{g}_k \le \eta,$$
$$\mathbf{D}_k^{\mathrm{T}}\boldsymbol{\pi}_{0,k} = \mathbf{L}_k^{\mathrm{T}}\mathbf{b}, \quad \boldsymbol{\pi}_{0,k} \ge 0.$$

Similarly, constraint (5.15c) results in constraints (5.11d)-(5.11f).

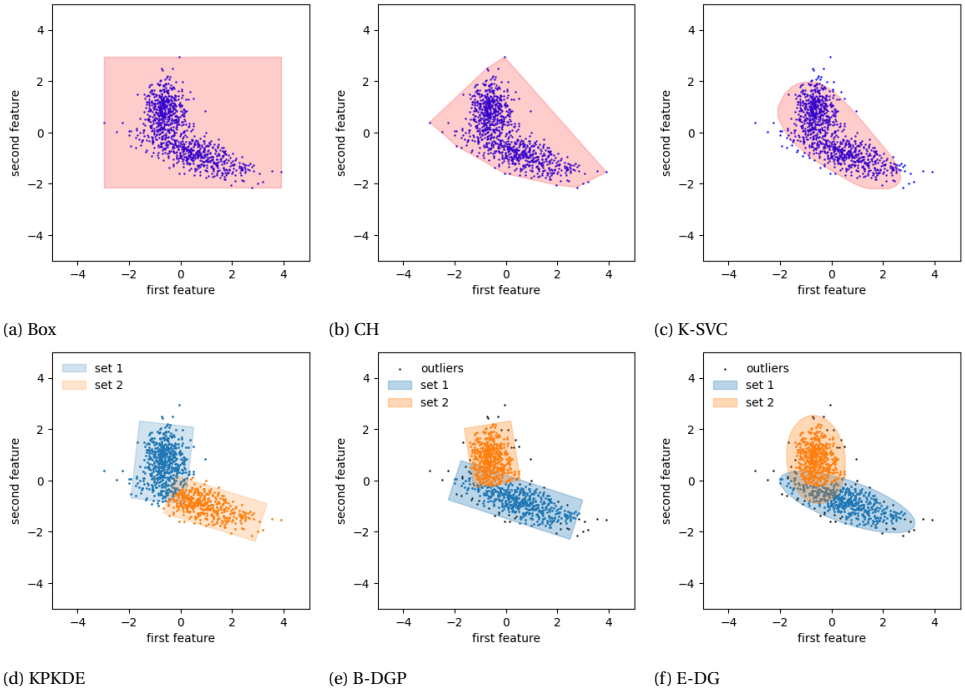
It is clear that any feasible solution of (5.11) can guarantee constraint satisfaction for any uncertainty as long as $\mathbf{u}\in\mathscr{U}$. Based on the I.I.D. assumption of the uncertainty and the data coverage in the testing set, it follows from Lemma 5.1 that $\mathbb{P}(\mathbb{P}_{\mathbf{u}\in\mathscr{U}} \le \rho - \tau) \le \exp(-2n\tau^2)$, which implies that the probability that the solution of (5.11) guarantees constraint satisfaction for any randomly generated uncertainty $\mathbf{u}$ is larger than $\rho-\tau$ with confidence of at least $1-\exp(-2n\tau^2)$. This completes the proof. □

5.A.2. SUPPLEMENTARY SIMULATION RESULTS

Remark 5.8 emphasized the importance of using real-world datasets for performance evaluation instead of simple synthetic datasets. In order to justify this statement and to

further demonstrate our point about real-world data vs synthetic data, a new numerical experiment was performed to test the performance of different data-driven approaches. A synthetic dataset with 400 mixed-Gaussian data samples, which is also considered in the reference [35], is used to test the performance of different data-driven uncertainty sets. Simulation results are shown in Fig. 5.7 and Table 5.5. It can be seen that the performances of different approaches are much more consistent and comparable than those in Case Study 2 of our work, using a real-world dataset. For the simple synthetic dataset shown in Fig. 5.7 and Table 5.5, all considered data-driven approaches work comparably well. However, for the real-world data set (Case Study 2 and 3), the performance of some data-driven approaches degrades a lot.

If only evaluating performance with the results shown in Fig. 5.7 and Table 5.5, one might conclude that these data-driven approaches work comparably well in terms of outliers removal and compactness (which is reflected in the set size and visual inspection). However, the conclusion obtained based on this simple synthetic dataset would be misleading and does not necessarily reflect the true performance when the approaches are applied in practice, where the real-world dataset can be much more irregular and complex. As a result, we want to emphasize that using real-world data for performance evaluation is more meaningful than using simple synthetic datasets to reflect the real situations when the considered approaches are applied in practical scenarios.



(a) Box (b) CH (c) K-SVC

(d) KPKDE (e) B-DGP (f) E-DG

Figure 5.7: Data-driven uncertainty sets using synthetic mixed-Gaussian data samples.

Table 5.5: Performance summary of different uncertainty sets with mixed-Gaussian data.

	existing approaches				our approaches	
	Box	CH	K-SVC	KPKDE	B-DGP	E-DG
uncertainty set size	34.98	17.38	11.07	11.48	11.41	11.42
complexity (# of linear constraints)	4	801	111	8	8	-
data coverage	1	1	0.945	0.965	0.977	0.977
computation time (s)	< 0.01	-	15.53	0.31	1.61	1.83

6

Data-Driven Robust Optimization with Multiple Uncertainty Subsets: Unified Uncertainty Set Representation and Mitigating Conservatism

Constructing uncertainty sets as unions of multiple subsets has emerged as an effective approach for creating compact and flexible uncertainty representations in data-driven robust optimization (RO). This chapter focuses on two separate research questions. The first concerns the computational challenge in applying these uncertainty sets in RO-based predictive control. To address this, a monolithic mixed-integer representation of the uncertainty set is proposed to uniformly describe the union of multiple subsets, enabling the computation of the worst-case uncertainty scenario across all subsets within a single mixed-integer linear programming (MILP) problem. The second research question focuses on mitigating the conservatism of conventional RO formulations by leveraging the structure of the uncertainty set. To achieve this, a novel objective function is proposed to exploit the uncertainty set structure and integrate the existing RO and distributionally robust optimization (DRO) formulations, yielding less conservative solutions than conventional RO formulations, while avoiding the high-dimensional continuous uncertainty distributions and the high computational burden typically associated with existing DRO

This chapter is based on the paper: Li, Y., Yorke-Smith, N., & Keviczky, T. (2026). On data-driven robust optimization with multiple uncertainty subsets: Unified uncertainty set representation and mitigating conservatism. *Journal of Process Control*, 158, 103611.

formulations. Given the proposed formulations, numerically efficient computation methods based on column-and-constraint generation (CCG) are also developed. Extensive simulations across three case studies are performed to demonstrate the effectiveness of the proposed schemes.

6.1. INTRODUCTION

Decision-making under uncertainty has gained increasing attention in both scientific research and engineering applications. One of the solutions for handling uncertainties in decisions is robust optimization (RO). By representing uncertainties via uncertainty sets, RO can ensure constraint satisfaction for all possible uncertainties without incurring an unaffordable computational burden and requiring the complete information of uncertainty distributions [7, 34, 135]. Despite the popularity and effectiveness of RO formulations, it is well-known that the optimal solution of RO problems can be conservative owing to two main factors.

The first factor contributing to the conservatism of the robust optimal solution is how the uncertainty set is represented. In recent years, with the availability of abundant measurement data and machine learning techniques, data-driven approaches have been developed to construct compact uncertainty sets by uncovering and exploiting the latent features of the uncertainty data, see [11, 12, 35, 61, 138, 139] and references therein. Among these data-driven approaches, one innovative methodology is constructing the uncertainty set as a union of multiple subsets, see [61, 138, 151] and references therein. By adopting such a methodology, the data-driven uncertainty set is able to adapt to irregular uncertainty distributions and ensure compactness, which is beneficial for reducing the conservatism of the corresponding robust optimal solutions.

It should be noted that, while the methodology of representing uncertainty sets as unions of multiple subsets has been proven to be effective in several applications, such as power system operation [145] and building climate control [62], this uncertainty set representation together with the existing *column-and-constraint generation* (CCG)-based algorithm could result in computationally demanding optimization problems when applied in robust predictive control settings. This computational tractability in practice is generally neglected in the existing works.

One widely adopted algorithmic solution for solving two-stage RO problems is the so-called CCG algorithm [152, 153]. When applying the existing CCG algorithm to solve the RO problem with multiple uncertainty subsets, each uncertainty subset entails solving one optimization problem to compute the worst-case uncertainty in each iteration of the CCG algorithm [138]. As elaborated in Section 6.3, the total number of uncertainty subsets in RO-based predictive control problems could increase exponentially with the length of the prediction horizon, which leads to computationally demanding or even intractable problems and limits the applicability of the existing approaches.

The second major factor influencing the conservatism of robust optimization solutions lies in how the design objective is formulated. In the classical RO framework, the optimization problem is designed to minimize the cost function while ensuring constraint satisfaction under the worst-case realization of the uncertainty. While this approach provides strong guarantees, it often leads to overly conservative solutions, especially when the worst-case scenario is highly improbable [7, 135, 138, 154]. To alleviate the conservatism of the optimal solution, alternative formulations have been explored. Stochastic optimization (SO) focuses on minimizing the expected performance with respect to a known probability distribution of the uncertainty, thereby potentially achieving better average-case performance [32, 133, 134]. However, SO relies on accurate knowledge of the true distribution of uncertainties, which is rarely available in

practice. Further, it is practically possible that distribution shifts between the empirical distribution and the true distribution might exist to degrade the out-of-sample performance. Distributionally robust optimization (DRO) offers a middle ground by optimizing the worst-case expected performance over an ambiguity set of probability distributions centered around a nominal or empirical distribution [154–159]. This enables DRO to hedge against distributional misspecification while avoiding some of the excessive conservatism inherent in worst-case RO.

Despite the effectiveness of these existing design options, they either require approximating the probability distribution of uncertainties, which could be high-dimensional and is non-trivial to obtain in practice, or entail solving a computationally demanding optimization problem. For example, in the moment-based DRO formulation [155, 158], a semi-definite programming problem needs to be solved at each iteration of the corresponding CCG algorithm; in scenario-based SO [134] and Wasserstein-based DRO formulation [156], the consideration of a large number of uncertainty scenarios leads to large-scale optimization problems and increased computational burden. Moreover, most existing designs of conservatism-reduced objective functions do not exploit the structure of uncertainty sets, especially the promising option of representing the uncertainty set as a union of multiple subsets in data-driven RO settings. In [160, 161], data-driven DRO formulations are proposed with the combination of multiple-source uncertainty modeling, where data-driven ambiguity sets are developed for each uncertainty subset separately. While these approaches exploit the latent property in empirical uncertainty samples and reduce dependence on prior knowledge to construct an ambiguity set, they still inherit the computational limitations in corresponding conventional DRO formulations, as the underlying problem structure remains unchanged.

By considering that the uncertainty set consists of multiple subsets, this chapter investigates two separate research questions: 1) *how to properly represent the uncertainty set to deal with the computational challenge when applying it in RO-based predictive control problems*; and 2) *how to exploit the uncertainty set structure to design new objective function for mitigating the conservatism of the robust optimal solution.* The main contributions of this chapter are summarized as follows:

- A monolithic representation of the uncertainty set consisting of multiple subsets is proposed for RO-based predictive control problems. Given the proposed representation, a computationally efficient CCG algorithm is developed such that only a single MILP problem needs to be solved to compute the worst-case uncertainty scenario across all uncertainty subsets, regardless of the length of the prediction horizon. In comparison, the direct extension of the conventional solutions, e.g., the methods developed in [138, 153], to predictive control problems entails solving a set of optimization problems whose number grows exponentially with the prediction horizon.

- Aiming at reducing the conservatism for optimizing the worst-case performance in conventional RO formulations, a novel objective function exploiting the structure of the uncertainty set, together with a CCG-based computation method, is proposed. By combining conventional RO and DRO formulations, the proposed objective function not only achieves less conservative solutions than conventional

RO formulations but also avoids approximating the continuous and possibly high-dimensional uncertainty distribution and incurring high computational burden as with the existing SO and DRO formulations, especially in the case of fully adaptive recourse decision and large numbers of uncertainty samples.

- Numerical experiments of three case studies are performed to extensively illustrate the issues of the existing approaches with the uncertainty set consisting of multiple subsets and demonstrate the effectiveness of the proposed schemes.

The remainder of this chapter is organized as follows. Section 6.2 briefly describes the preliminaries and research gaps. Section 6.3 clarifies the issues of the existing approach and proposes a new uncertainty set representation with a CCG-based computation method. In Section 6.4, a novel objective function exploiting the structure of the uncertainty set is proposed to mitigate the conservatism of robust optimal solutions. Given the proposed objective function, a CCG-based computation method is also derived. Section 6.5 presents numerical experiment results. Section 6.6 concludes this chapter.

Notation: boldface capital letters denote matrices, and boldface lowercase letters denote vectors. Sets are represented by calligraphic capital letters. $\mathbf{e}$ denotes all-ones vectors with appropriate dimensions. $\otimes$ is the Cartesian product. $\mathbb{E}[\cdot]$ indicates the expectation of a given random variable. The operator $\text{col}(x_1, \cdots, x_n)$ represents the stacked vector $[x_1^{\mathrm{T}}, \cdots, x_n^{\mathrm{T}}]$.

6.2. PRELIMINARIES

This section introduces some preliminaries of the two-stage linear RO problem to be investigated, and briefly discusses the research gaps when considering an uncertainty set consisting of multiple subsets.

The following two-stage linear RO problem is considered in this work

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^{\mathrm{T}}\mathbf{x} + \max_{\mathbf{v}\in\mathcal{V}} \min_{\mathbf{y}\in\mathcal{Y}(\mathbf{x},\mathbf{v})} \mathbf{b}^{\mathrm{T}}\mathbf{y} \right\} \tag{6.1a}$$

$$\text{s.t. } \mathbf{A}\mathbf{x} \le \mathbf{q}, \tag{6.1b}$$

$$\mathcal{Y}(\mathbf{x},\mathbf{v}) := \left\{ \mathbf{y} \mid \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \mathbf{M}\mathbf{v} \le \mathbf{h} \right\}, \tag{6.1c}$$

where $\mathbf{x} \in \mathbb{R}^p$ are the first-stage decision variables (interchangeably called *here-and-now* decision variables), which can contain both continuous and binary elements, $\mathbf{y} \in \mathbb{R}^q$ are the second-stage decision variables (interchangeably called *wait-and-see* decision variables or *recourse* decision variables), which are assumed to be continuous, $\mathbf{v} \in \mathbb{R}^m$ denotes the uncertainties with $\mathcal{V}$ as the corresponding uncertainty set, $\mathcal{Y}(\mathbf{x},\mathbf{v})$ is the admissible set of $\mathbf{y}$, $(\mathbf{c},\mathbf{b},\mathbf{q},\mathbf{h})$ and $(\mathbf{A},\mathbf{T},\mathbf{W},\mathbf{M})$ are parameter vectors and matrices, respectively, with appropriate dimensions. Without loss of generality, in the remaining parts of this chapter, the above two-stage RO problem is assumed to be relatively complete recourse, i.e., the optimization problem is feasible for any possible $\mathbf{x}$ and $\mathbf{v}$ [153]. For problems that do not satisfy this assumption, some other strategic solutions discussed in [152] are also applicable in our proposed design.

To solve the above two-stage RO problem, one commonly adopted approach is the so-called *column-and-constraint generation* (CCG) algorithm. In CCG algorithm, the op-

timal solution of the RO problem is computed by iteratively solving a master problem and several subproblems (the number of the subproblems to be solved is dependent on the number of uncertainty subsets) to update the lower bound and upper bound of the objective function until a predefined optimality gap is achieved [138, 153]. In this chapter, unless mentioned specifically, the existing approach for solving the RO problem (6.1) refers to the CCG algorithm.

For reducing the conservatism of the robust optimal solutions, including but not limited to the two-stage RO problem in (6.1), data-driven approaches exploiting historical uncertainty data and machine learning techniques have been extensively explored for constructing compact uncertainty sets. Among these data-driven approaches, one promising methodology is representing the uncertainty set as a union of multiple subsets, such as polytopes and ellipsoids, which has been shown to be effective in improving the compactness and flexibility of the uncertainty sets in handling irregular and complex uncertainty distributions. An uncertainty set $\mathcal{V}$ represented as a union of multiple subsets can be expressed as

$$\mathcal{V} := \bigcup_{k=1}^{K} \mathcal{V}_k \tag{6.2}$$

where $\mathcal{V}_k$ denotes the k-th uncertainty subset with K as the total number of subsets. The uncertainty subsets can be any basic sets, such as box, ellipsoid, and polytope.

While the existing works have demonstrated the promising effectiveness of formulating an uncertainty set as a union of multiple subsets as in (6.2), there are two research gaps accompanying such uncertainty set representation to be filled.

- *RO-based Predictive Control With Multiple Uncertainty Subsets*: The RO formulation (6.1) has been applied in predictive control problems due to its effectiveness in ensuring robust constraint satisfaction [11]. However, when employing the data-driven uncertainty set representation (6.2), the number of uncertainty subsets could grow exponentially with the prediction horizon. This exponential growth leads to a rapid increase in the number of subproblems solved in each iteration of the existing CCG-based algorithms, making the optimization problem computationally demanding or even intractable. Therefore, developing computationally efficient solutions to address this challenge is crucial for enhancing the practicality of the data-driven RO design with the uncertainty set representation (6.2) in robust predictive control.

- *Conservatism-Reduced Objective Function With Multiple Uncertainty Subsets*: The conventional RO formulation could suffer from conservatism due to the optimization of worst-case performance, as shown in (6.1). Given the existing RO formulations with multiple uncertainty subsets, the structure of the uncertainty set is not fully considered in designing the objective function. Consequently, how to properly exploit the property that the uncertainty set consists of multiple subsets to design novel RO formulations is beneficial to reduce the conservatism of the robust optimal solution.

The above research gaps, together with our proposed solutions, will be further elaborated in detail in the subsequent sections.

6.3. RO-BASED PREDICTIVE CONTROL WITH MULTIPLE UNCERTAINTY SUBSETS

This section details the research gap when applying the uncertainty set structure (6.2) in robust predictive control, and proposes a novel mixed-integer representation to uniformly describe the unified uncertainty set such that the computational efficiency of the CCG algorithm for solving the RO problem is remarkably improved.

6.3.1. PROBLEM FORMULATION

Theoretically, even with the uncertainty set represented as in (6.2), existing data-driven methods for constructing uncertainty sets and the solutions for solving (6.1) can be directly extended to robust predictive control problems. However, practical challenges arise when applying these methods, which have been overlooked in existing works.

For simplicity of illustration, consider the following deterministic formulation of a predictive control problem

$$\min_{\mathbf{u}_t} \sum_{t=1}^{N} l(\mathbf{s}_t, \mathbf{u}_t) \tag{6.3a}$$

$$\text{s.t. } \mathbf{s}_{t+1} = \mathbf{\Phi}\mathbf{s}_t + \mathbf{\Gamma}\mathbf{u}_t + \mathbf{v}_t \tag{6.3b}$$

$$\mathbf{s}_t \in \mathcal{S}, \mathbf{u}_t \in \mathcal{U}, \ \mathbf{v}_t \in \mathcal{V}_t \tag{6.3c}$$

$$t = 1, \cdots, N \tag{6.3d}$$

where $\mathbf{s}_t$ denotes the system states vector, $\mathbf{u}_t$ denotes the control input vector, $\mathbf{v}_t$ the uncertainties, $l(\mathbf{s}_t, \mathbf{u}_t)$ the stage cost function, $\mathcal{S}$ and $\mathcal{U}$ are feasible sets of system states and control inputs, respectively, $\mathcal{V}_t$ is the uncertainty set of $\mathbf{v}_t$, subscript t denotes the t-th time step, and N the length of the prediction horizon. In predictive control settings, it is common that the uncertainties $\mathbf{v}_t$ ($t = 1, \cdots, N$) are assumed to be independent and identically distributed (I.I.D.) [162, 163].

After reformulating the predictive control problem (6.3) into the corresponding RO formulation (6.1), the uncertainty vector $\mathbf{v}$ is the stacked uncertainty sequences $\mathbf{v}_t$ within the prediction horizon, i.e., $\mathbf{v} := \text{col}(\mathbf{v}_1, \cdots, \mathbf{v}_N)$. In the existing data-driven RO framework, the uncertainty set $\mathcal{V}$ is directly constructed for the stacked uncertainty vector $\mathbf{v}$. While this approach is still theoretically feasible, it will incur some issues. On one hand, for a fixed number of uncertainty samples $\mathbf{v}_t$, considering the stacked uncertainty $\mathbf{v}$ reduces the size of the training dataset (shrinking it to $\frac{1}{N}$ of the original dataset for $\mathbf{v}_t$). This reduction may result in insufficient data and degrade the performance of data-driven approaches for constructing uncertainty sets. On the other hand, directly modelling the uncertainty set for the stacked uncertainty $\mathbf{v}$ increases the dimensionality of uncertainty samples, particularly for long prediction horizon N. As shown in existing studies, high-dimensional uncertainties pose challenges in implementing data-driven methods, including difficulties in hyper-parameter tuning and performance validation.

For example, in building climate predictive control, the prediction horizon could span 12 hours with a sampling period of 30 minutes, and a typical uncertainty is the prediction error of ambient temperature. Given one year of historical uncertainty data of $\mathbf{v}_t$, the number of uncertainty samples for $\mathbf{v}_t$ is 365*24*2. However, when consider-

ing the stacked uncertainty samples for $\mathbf{v} = \text{col}(\mathbf{v}_1, \cdots, \mathbf{v}_N)$ with $N = 24$, the training set size reduces to $365 * 2$. Consequently, the dataset for $\mathbf{v}$ may be too small to ensure the effectiveness of data-driven approaches in constructing uncertainty sets.

Under the I.I.D. assumption of $\mathbf{v}_t$, one solution is to construct a data-driven uncertainty set for $\mathbf{v}_t$, denoted as $\mathcal{V}_t$, and then extend the per-step uncertainty set $\mathcal{V}_t$ to construct the uncertainty set of $\mathbf{v}$. Assuming that the uncertainty set for $\mathbf{v}_t$ consists of K subsets and denoting the k-th uncertainty subset as $\mathcal{V}_{t,k}$, the uncertainty set for $\mathbf{v} = \text{col}(\mathbf{v}_1, \cdots, \mathbf{v}_N)$, denoted as $\mathcal{V}$, can be expressed as

$$\mathcal{V} = \underbrace{\mathcal{V}_t \otimes \ldots \otimes \mathcal{V}_t}_{N \text{ times}}, \quad \mathcal{V}_t = \bigcup_{k=1}^{K} \mathcal{V}_{t,k} \tag{6.4}$$

where N is the length of the prediction horizon considered in the predictive control problem (6.3).

Remark 6.1. It can be seen from (6.4) that the number of subsets constituting $\mathcal{V}$ is K^N, which grows exponentially with the prediction horizon N. By adopting the existing CCG algorithm to solve the RO problem (6.1) with the uncertainty set (6.4), each uncertainty subset requires solving a bilinear or mixed-integer linear optimization problem to compute the worst-case uncertainty in every iteration of the algorithm [138]. Consequently, handling K^N subsets translates to solving K^N optimization problems per iteration. In case of large values of K and N, this approach could be computationally demanding and practically inapplicable. Therefore, the computational challenges arising from representing the uncertainty set as a union of multiple subsets in (6.4) highlight the need for an effective formulation of the uncertainty set for $\mathbf{v}$.

6.3.2. UNIFIED UNCERTAINTY SET REPRESENTATION AND COMPUTATIONAL METHOD DESIGN

This section aims to propose a novel representation to uniformly represent the uncertainty set of $\mathbf{v}$ based on the uncertainty subsets of $\mathbf{v}_t$ and derive a computationally efficient solution for solving the corresponding RO problem.

Assumption 6.1. The uncertainty set of $\mathbf{v}_t$, denoted as $\mathcal{V}_t$, consists of K subsets $\mathcal{V}_{t,k}$ with each uncertainty subset $\mathcal{V}_{t,k}$ as a nonempty and bounded polytope defined as $\mathcal{V}_{t,k} := \{\mathbf{v}_t | \mathbf{D}_k \mathbf{v}_t \leq \mathbf{d}_k\}$.

On the basis of Assumption 6.1, the following monolithic mixed-integer formulation of the uncertainty set $\mathcal{V}$ is derived

$$\begin{aligned} \mathcal{V} := \Big\{ \text{col}(\mathbf{v}_1, \cdots, \mathbf{v}_N) \Big| & \sum_{k=1}^{K} \delta_{t,k} \mathbf{D}_k \mathbf{v}_t \leq \sum_{k=1}^{K} \delta_{t,k} \mathbf{d}_k, \\ & \delta_{t,k} \in \mathbb{B}, \ \textstyle\sum_{k=1}^{K} \delta_{t,k} = 1, \ t = 1, \cdots, N \Big\} \end{aligned} \tag{6.5}$$

where $\delta_{t,k} \in \{0,1\}$ are auxiliary variables. For the formulation (6.5), given any subset of $\mathcal{V}$, a set of feasible variables $\{\delta^*_{t,k} \mid t = 1, \cdots, N, k = 1, \cdots, K\}$ can be found to uniquely represent a specific uncertainty subset. Conversely, any feasible $\{\delta^*_{t,k} \mid t = 1, \cdots, N, k = 1, \cdots, K\}$ will also define an admissible uncertainty subset of $\mathcal{V}$. Instead of explicitly representing the uncertainty set $\mathcal{V}$ as a union of multiple subsets as in (6.4), the proposed

mixed-integer representation (6.5) gives a monolithic implicit representation of $\mathcal{V}$ that is feasible for any size of uncertainty subsets K and prediction horizon N.

Theorem 6.1. Considering the uncertainty set (6.5) and decomposing the parameter matrix $\mathbf{M}$ into N column blocks $\{\mathbf{M}_t\}_{t=1}^{N}$ with appropriate dimensions, the RO problem in (6.1) can be reformulated as in (6.6), which can be solved via Algorithm 6.1 within finite iterations.

$$\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \max_{\delta_{t,k},\mathbf{v}_t} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \tag{6.6a}$$

$$\text{s.t. } \mathbf{A}\mathbf{x} \le \mathbf{q}, \tag{6.6b}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \textstyle\sum_{t=1}^{N} \mathbf{M}_t\mathbf{v}_t \le \mathbf{h}, \tag{6.6c}$$

$$\textstyle\sum_{k=1}^{K} \delta_{t,k}\mathbf{D}_k\mathbf{v}_t \le \sum_{k=1}^{K} \delta_{t,k}\mathbf{d}_k, \tag{6.6d}$$

$$\delta_{t,k} \in \mathbb{B},\ k = 1,\cdots,K,\ t = 1,\cdots,N, \tag{6.6e}$$

$$\textstyle\sum_{k=1}^{K} \delta_{t,k} = 1,\ t = 1,\cdots,N. \tag{6.6f}$$

Proof: the proof of this theorem is presented in the Appendix 6.A.1. □

The master problem **MP** and the subproblem **SP**1 entailed in Algorithm 6.1 are defined as:

$$\mathbf{MP}: \quad \min_{\mathbf{x},\eta,\mathbf{y}_i} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \eta \tag{6.7a}$$

$$\text{s.t. } \mathbf{b}^{\mathrm{T}}\mathbf{y}_i \le \eta, \tag{6.7b}$$

$$\mathbf{A}\mathbf{x} \le \mathbf{q}, \tag{6.7c}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y}_i + \textstyle\sum_{t=1}^{N} \mathbf{M}_t\mathbf{v}_{t,i} \le \mathbf{h}, \tag{6.7d}$$

$$i = 1,2,\cdots,r. \tag{6.7e}$$

$$\mathbf{SP}1: \quad \max_{\delta_{t,k},\mathbf{v}_t} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \tag{6.8a}$$

$$\text{s.t. } \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \textstyle\sum_{t=1}^{N} \mathbf{M}_t\mathbf{v}_t \le \mathbf{h}, \tag{6.8b}$$

$$\textstyle\sum_{k=1}^{K} \delta_{t,k}\mathbf{D}_k\mathbf{v}_t \le \sum_{k=1}^{K} \delta_{t,k}\mathbf{d}_k, \tag{6.8c}$$

$$\textstyle\sum_{k=1}^{K} \delta_{t,k} = 1,\ \delta_{t,k} \in \mathbb{B}, \tag{6.8d}$$

$$k = 1,\cdots,K,\ t = 1,\cdots,N. \tag{6.8e}$$

The subproblem (6.8) contains bilinear terms $\delta_{t,k}\mathbf{v}_t$ due to the auxiliary binary decision variable $\delta_{t,k}$, and hence is non-convex and might be intractable to some solvers. These bilinear terms can be further reformulated as mixed-integer linear constraints via

Algorithm 6.1 *column-and-constraint generation* algorithm for solving (6.6).

Input: suboptimality gap ϵ
Output: the optimal decision variable $\mathbf{x}$ and objective function value $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$

1: Set $LB = -\infty$, $UB = \infty$, $r = 0$
2: **while** $|UB - LB| > \epsilon$ **do**
3: Solve **MP** in (6.7) to derive solutions $(\mathbf{x}^*, \eta^*)$ and update $LB = \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$
4: Solve **SP**1 in (6.8) or **SP**2 in (6.9) to derive solutions $\{\mathbf{v}_t^*, \mathbf{y}^*\}$, and update UB as

$$UB = \min\left\{UB, \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \mathbf{b}^{\mathrm{T}}\mathbf{y}^*\right\}$$

5: Create decision variables $\mathbf{y}_r$, set parameters $\mathbf{v}_{t,r} = \mathbf{v}_t^*$ $(t = 1, \cdots, N)$, and add the following constraints to **MP** in (6.7)

$$\begin{cases} \mathbf{b}^{\mathrm{T}}\mathbf{y}_t \le \eta, \\ \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y}_r + \sum_t \mathbf{M}_t \mathbf{v}_{t,r}^* \le \mathbf{h}. \end{cases}$$

6: $r \leftarrow r + 1$
7: **end while**
8: **Return**: $\mathbf{x}^*$ and $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$

6

big-M formulation. The reformulated subproblem **SP**2 is given in (6.9).

$$\mathbf{SP}2:$$

$$\max_{\delta_{t,k}, \mathbf{v}_t, \mathbf{w}_{t,k}} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \tag{6.9a}$$

$$\text{s.t. } \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \sum_{t=1}^{N} \mathbf{M}_t \mathbf{v}_t \le \mathbf{h}, \tag{6.9b}$$

$$\sum_{k=1}^{K} \mathbf{D}_k \mathbf{w}_{t,k} \le \sum_{k=1}^{K} \delta_{t,k} \mathbf{d}_k, \tag{6.9c}$$

$$-\Delta(1 - \delta_{t,k})\mathbf{e} \le \mathbf{v}_t - \mathbf{w}_{t,k} \le \Delta(1 - \delta_{t,k})\mathbf{e}, \tag{6.9d}$$

$$-\Delta\delta_{t,k}\mathbf{e} \le \mathbf{w}_{t,k} \le \Delta\delta_{t,k}\mathbf{e}, \tag{6.9e}$$

$$k = 1, \cdots, K, \ t = 1, 2, \cdots, N \tag{6.9f}$$

where $\Delta > 0$ is a sufficiently large constant, and $\mathbf{e}$ is an all-ones vector with appropriate dimension.

Remark 6.2. It is worth noting that the proposed formulation in (6.6) is equivalent to the conventional RO formulation (6.1) with the uncertainty set (6.4), because the mixed-integer representation of the uncertainty set (6.5) exactly matches (6.4) that unifies K^N subsets. The number of possible parameter combinations $\{\delta_{t,k} \mid t = 1, \ldots, N, \ k = 1, \ldots, K\}$ in (6.5) is equal to the number of uncertainty subsets in (6.4), i.e., K^N. In addition, each feasible parameter combination $\{\delta_{t,k} \mid t = 1, \ldots, N, \ k = 1, \ldots, K\}$ uniquely specifies an uncertainty subset for $\mathbf{v}$ via (6.5).

Remark 6.3. In contrast to the existing solution for solving the RO problem (6.1) with uncertainty set (6.4), in which K^N numbers of subproblems need to be solved in each algorithm iteration, the reformulated uncertainty set (6.5) enables a monolithic description

of the uncertainty set and ensures that only a single subproblem (**SP**1 or **SP**2), whose decision variables and constraints increase linearly with K and N, to be solved in Algorithm 6.1 for computing the worst-case uncertainty. This can remarkably improve the computational efficiency of the proposed approach in the case of large values of K and/or N. Additionally, the proposed approach enables the construction of data-driven uncertainty sets for $\mathbf{v}_t$ instead of the stacked uncertainty $\mathbf{v}$ without incurring a high computational burden for solving the corresponding RO problem (6.1). In comparison with $\mathbf{v}$, $\mathbf{v}_t$ has a lower dimension and more data samples, which will further boost the performance of the data-driven approaches for constructing uncertainty sets. For example, a lower uncertainty dimension makes it easier to tune the parameters/hyper-parameters and evaluate the performance of the data-driven approaches. A larger dataset can also prevent the data-driven approaches from suffering performance degradation, such as overfitting, underfitting, poor model calibration, etc., due to insufficient training data. It is worth highlighting that, while the proposed MILP formulation is NP-hard in theory, it can often be efficiently solved in practice with modern off-the-shelf solvers given the developments in mixed-integer optimization [164].

6.4. MITIGATING CONSERVATISM VIA OBJECTIVE FUNCTION DESIGN

This section proposes an alternative objective function for RO problems by leveraging the structure of uncertainty sets composed of multiple subsets. Additionally, an algorithmic solution is developed to efficiently solve the corresponding RO problem.

6.4.1. PROBLEM FORMULATION

Based on the structure of the uncertainty set – a union of multiple subsets – this section presents a new formulation of the objective function for RO problems to reduce the conservatism of the optimal solution.

The conventional RO formulation in (6.1) could lead to conservative solutions due to the objective of worst-case performance optimization. Many existing works have been done to propose new objective functions to reduce the conservatism of the optimal solution. Two prominent alternative objective functions are given as follows.

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^{\mathrm{T}}\mathbf{x} + Q(\mathbf{v},\mathbf{x}) \right\} \tag{6.10}$$

where two options of $Q(\mathbf{v},\mathbf{x})$ are

$$\text{Option 1: } Q(\mathbf{v},\mathbf{x}) := \mathbb{E}_{f(\mathbf{v})}\left[\min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \right], \tag{6.11a}$$

$$\text{Option 2: } Q(\mathbf{v},\mathbf{x}) := \max_{f(\mathbf{v})\in\mathscr{P}} \mathbb{E}_{f(\mathbf{v})}\left[\min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \right] \tag{6.11b}$$

with $f(\mathbf{v})$ as the probability distribution of the uncertainty $\mathbf{v}$, and $\mathscr{P}$ denotes the ambiguity set defining the admissible set of $f(\mathbf{v})$.

The objective function in (6.11a) seeks to optimize the expected performance to the probability distribution of the uncertainty $\mathbf{v}$, which is the standard setting in stochastic

optimization. This approach can reduce conservatism compared with worst-case formulations, but it relies on explicit and accurate knowledge of the underlying uncertainty distribution. Moreover, closed-form or tractable solutions are generally available only for specific distributional families (e.g., Gaussian) or problem structures (e.g., linear decision rules). When samples of the uncertainty are available, sample-based (randomized) methods are often used to approximate the expectation in (6.11a), resulting in a large-scale deterministic optimization problem whose size grows with the number of samples, which can be computationally demanding. In addition, distribution shifts between the empirical and the true distributions may exist so that the resulting solution of (6.11a) may exhibit poor out-of-sample performance.

To address these issues, distributionally robust optimization (DRO), as in (6.11b), accounts for distributional ambiguity by assuming that the true distribution $f(\mathbf{v})$ is unknown but lies within an ambiguity set $\mathscr{P}$, typically defined around a nominal or empirical distribution. DRO then optimizes the worst-case expected performance over all distributions in $\mathscr{P}$. This approach improves robustness against sampling errors and model misspecification. Nevertheless, achieving both high-probability constraint satisfaction and low conservatism in DRO often requires a sufficiently large and representative set of samples to construct the ambiguity set. Furthermore, as noted in [165], when the decision rule is fully adaptive to the uncertainty, which is the case considered in this work, and the sample size is large, DRO formulations can suffer from the curse of dimensionality, leading to high computational cost.

Motivated by the objective functions in (6.11b) and the property that the uncertainty set consists of multiple subsets, we propose a new objective function

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^{\mathrm{T}}\mathbf{x} + \max_{\mathbf{p}\in\mathscr{P}} \mathbb{E}_{\mathbf{p}} \left[\max_{\mathbf{v}\in\mathcal{V}_k} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \right] \right\} \tag{6.12}$$

where $\mathbf{p} = [\mathbf{p}_1, \mathbf{p}_2 \cdots, \mathbf{p}_K]^{\mathrm{T}}$ is the probability vector with $\mathbf{p}_k := \mathbb{P}(\mathbf{v} \in \mathcal{V}_k)$, $\mathscr{P}$ is the ambiguity set defining the admissible set of $\mathbf{p}$.

Remark 6.4. In contrast to the conventional RO formulation in (6.1), where the worst-case performance among all possible uncertainty scenarios are optimized, the proposed formulation (6.12) optimizes the distributionally robust solution w.r.to the expected worst-case performance $\max_{\mathbf{v}\in\mathcal{V}_k} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y}$ over all uncertainty subsets. In (6.12), the worst-case performance in each uncertainty subset $\mathcal{V}_k$ is weighted based on the probability $\mathbb{P}(\mathbf{v} \in \mathcal{V}_k)$, and hence a less conservative solution is expected with our proposed formulation than the conventional RO formulation.

Remark 6.5. In contrast to the SO and DRO formulations in (6.11), which entail the information of the joint probability distribution $f(\mathbf{v})$ of the continuous uncertainty $\mathbf{v}$, our proposed formulation (6.12) only considers the discrete probability distribution $\mathbb{P}(\mathbf{v} \in \mathcal{V}_k)$. Compared with $f(\mathbf{v})$, $\mathbb{P}(\mathbf{v} \in \mathcal{V}_k)$ is much easier to estimate since $f(\mathbf{v})$ is the continuous probability distribution of random variable $\mathbf{v}$ that might be high dimensional. In contrast, the probability vector $\mathbf{p}$ is the distribution of the discrete random variable $\mathbb{I}(\mathbf{v} \in \mathcal{V}_k)$, where $\mathbb{I}(\cdot)$ is the indicator function. An approximation of $\mathbf{p}$ can be readily computed by counting the frequency of $\mathbf{v} \in \mathcal{V}_k$ for all uncertainty data samples, and is a multinomial distribution when uncertainty subsets $\{\mathcal{V}_k,\ k = 1, \cdots, K\}$ are disjoint. Furthermore, the

uncertainty is still described via an uncertainty set in (6.12), and uncertainty samples are not explicitly considered, which are beneficial for maintaining the computational efficiency as in conventional RO formulation. On the contrary, the resulting optimization problems for the existing SO and DRO formulations in the format of (6.11), e.g., the scenario-based SO formulation [134] and Wasserstein-based DRO formulation [156], need to consider all uncertainty samples and are generally computationally demanding in the presence of large numbers of uncertainty scenarios.

In this work, the ambiguity set $\mathscr{P}$ is defined based on Kullback-Leibler (KL) divergence as

$$\mathscr{P} := \left\{ \mathbf{p} \mid \mathbf{p} \geq 0,\ \mathbf{e}^{\mathrm{T}}\mathbf{p} = 1,\ KL(\bar{\mathbf{p}}, \mathbf{p}) \leq \rho \right\}, \tag{6.13}$$

where $KL(\bar{\mathbf{p}}, \mathbf{p}) := \sum_k \bar{\mathbf{p}}_k \log(\frac{\bar{\mathbf{p}}_k}{\mathbf{p}_k})$ is the KL divergence function measuring the similarity of two probability vectors $\bar{\mathbf{p}}$ and $\mathbf{p}$ with $\bar{\mathbf{p}}$ as the approximated probability vector extracted from historical data, $\rho \geq 0$ is a user-defined parameter determining the size of the ambiguity set.

Considering the objective function (6.12) as well as the ambiguity set definition (6.13) and applying epigraphical reformulation, it gives to the following optimization problem

$$\min_{\mathbf{x},\eta} \ \mathbf{c}^{\mathrm{T}}\mathbf{x} + \eta \tag{6.14a}$$

$$\text{s.t.} \ \max_{\mathbf{p}\in\mathscr{P}} \left\{ \mathbb{E}_{\mathbf{p}} \left[\max_{\mathbf{v}\in\mathscr{V}_k} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \right] \right\} \leq \eta, \tag{6.14b}$$

$$\mathbf{p} \geq 0,\ \mathbf{e}^{\mathrm{T}}\mathbf{p} = 1,\ KL(\bar{\mathbf{p}}, \mathbf{p}) \leq \rho, \tag{6.14c}$$

$$\mathbf{A}\mathbf{x} \leq \mathbf{q}, \tag{6.14d}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \mathbf{M}\mathbf{v} \leq \mathbf{h}. \tag{6.14e}$$

Remark 6.6. For the proposed formulation in (6.14), the conservatism of the optimal solution depends on both the size of the ambiguity set $\mathscr{P}$, determined by the parameter ρ, and the number of uncertainty subsets K. Specifically, conservatism can be tuned in two ways: by adjusting K when constructing the data-driven uncertainty set, and by varying ρ when designing the ambiguity set. In general, a larger ρ yields more conservative solutions, whereas a larger K reduces conservatism. In extreme cases, for a fixed K, a sufficiently large $\mathscr{P}$ leads to the worst-case optimization, as in the RO formulation (6.1), since the ambiguity set then includes the probability vector $\mathbf{p}$ that places unit mass on the subset containing the worst-case sample. Similarly, when the uncertainty set has no subset (i.e., $K = 1$), the formulation reduces to the worst-case optimization regardless of the chosen ρ. In practice, ρ can be tuned through cross-validation or out-of-sample testing, by evaluating different ambiguity set sizes and selecting the one that best balances robustness and performance.

6.4.2. COMPUTATIONAL METHOD DESIGN

The proposed formulation (6.14) contains quadruple-level optimization problems and is computationally intractable to numerical solvers. This section will present a computational method based on the CCG algorithm to efficiently solve the optimization problem (6.14).

Assumption 6.2. For the robust optimization problem (6.14), the uncertainty set $\mathcal{V}$ is a union of K nonempty and bounded subsets $\mathcal{V}_k$ defined as $\mathcal{V}_k := \{\mathbf{v} \mid \mathbf{D}_k\mathbf{v} \le \mathbf{d}_k\}$.

Theorem 6.2. Assuming Assumption 6.2 holds, the optimization problem (6.14) can be solved via Algorithm 6.3 by iteratively solving a master problem $\mathbf{MP}_{\mathrm{DRO}}$ in (6.15) and K subproblems $\mathbf{SP}^k_{\mathrm{DRO}}$ in (6.16) within finite number of iterations.

$\mathbf{MP}_{\mathrm{DRO}}$:

$$\min_{\mathbf{x},\eta,\nu,\mu,\mathbf{y}_{k,i}} \ \mathbf{c}^{\mathrm{T}}\mathbf{x}+\eta \tag{6.15a}$$

$$\text{s.t. } \nu\sum_{k=1}^{K}\bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}_{k,i}-\mu}{\nu}-1\right)+\mu+\rho\nu \le \eta, \tag{6.15b}$$

$$\mathbf{A}\mathbf{x} \le \mathbf{q}, \tag{6.15c}$$

$$\mathbf{T}\mathbf{x}+\mathbf{W}\mathbf{y}_{k,i}+\mathbf{M}\mathbf{v}^*_{k,i} \le \mathbf{h}, \tag{6.15d}$$

$$\nu \ge 0, \tag{6.15e}$$

$$i=1,\cdots,r,\ k=1,\cdots,K. \tag{6.15f}$$

$\mathbf{SP}^k_{\mathrm{DRO}}$:

$$\max_{\mathbf{v}}\min_{\mathbf{y}} \ \mathbf{b}^{\mathrm{T}}\mathbf{y} \tag{6.16a}$$

$$\text{s.t. } \mathbf{T}\mathbf{x}^*+\mathbf{W}\mathbf{y}+\mathbf{M}\mathbf{v} \le \mathbf{h}, \tag{6.16b}$$

$$\mathbf{D}_k\mathbf{v} \le \mathbf{d}_k. \tag{6.16c}$$

Proof. The proof is provided in the Appendix 6.A.2. □

Remark 6.7. It should be noted that the master problem $\mathbf{MP}_{\mathrm{DRO}}$ (6.15) is nonlinear due to the constraint (6.15b) even if the original constraints and objective function in (6.14) are linear. Besides, it can be seen from Algorithm 6.3 that the number of these nonlinear constraints will increase with the iteration of the algorithm, which could degrade the computational efficiency of our proposed scheme. One solution to alleviate this issue is applying the following reformulation for constraint (6.15b):

$$\mu+\rho\nu+\nu\sum_{k}\bar{\mathbf{p}}_k\exp\left(\frac{\phi_k-\mu}{\nu}-1\right) \le \eta, \tag{6.17a}$$

$$\mathbf{b}^{\mathrm{T}}\mathbf{y}_{k,i} \le \phi_k,\ i=1,\cdots,r,\ k=1,\cdots,K, \tag{6.17b}$$

where ϕ_k are auxiliary decision variables. The above reformulation is valid because the left-hand side of (6.15b) is strictly increasing in $\mathbf{b}^{\mathrm{T}}\mathbf{y}_{k,i}$. By replacing (6.15b) via (6.17), only linear constraints are added in each iteration, and the number of nonlinear constraints (6.17a) is fixed regardless of the algorithm iteration, which is beneficial in improving the computational efficiency of Algorithm 6.3.

Remark 6.8. It is worth noting that, while the constraints (6.15b) and (6.17a) are nonlinear, they are convex, and $\mathbf{MP}_{\mathrm{DRO}}$ becomes a convex optimization when the decision variable $\mathbf{x}$ is continuous. The convexity of the constraints (6.15b) and (6.17a) is proved

Algorithm 6.2 *column-and-constraint generation* algorithm for solving (6.14).

Input: suboptimality gap ϵ

Output: optimal decision variable $\mathbf{x}^*$ and objective function value $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$

1: Set $LB = -\infty$, $UB = \infty$, $r = 0$

2: **while** $|UB - LB| > \epsilon$ **do**

3: Solve $\mathbf{MP}_{\mathrm{DRO}}$ (6.15) to obtain solutions $(\mathbf{x}^*, \eta^*, \mu^*, \nu^*)$ and update $LB = \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$

4: Solve K subproblems $\mathbf{SP}^k_{\mathrm{DRO}}$ (6.16) to obtain solutions $\{\mathbf{v}^*_k, \mathbf{y}^*_k\}_{k=1}^K$, and update UB as

$$UB = \min\Big\{UB,\ \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \mu^* + \rho\nu^* + \nu^* \sum_{k=1}^K \tilde{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}^*_k - \mu^*}{\nu^*} - 1\right)\Big\}$$

5: Create decision variables $\{\mathbf{y}_{k,r}\}_{k=1}^K$, set parameters $\mathbf{v}^*_{k,r} = \mathbf{v}^*_k$ $(k = 1, \cdots, K)$, and add the following constraints to $\mathbf{MP}_{\mathrm{DRO}}$ in (6.15)

$$\begin{cases} \mu + \rho\nu + \nu \sum_{k=1}^K \tilde{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}_{k,r} - \mu}{\nu} - 1\right) \le \eta, \\ \mathbf{Tx} + \mathbf{Wy}_{k,r} + \mathbf{Mv}^*_{k,r} \le \mathbf{h},\ k = 1, \cdots, K. \end{cases}$$

6: $r \leftarrow r + 1$

7: **end while**

8: **Return**: $\mathbf{x}^*$ and $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$

in the Appendix 6.A.3. Several off-the-shelf solvers, such as `Gurobi`, `Ipopt` and `MadNLP`, can deal with this type of nonlinearity. The subproblem $\mathbf{SP}^k_{\mathrm{DRO}}$ (6.16) is a bi-level linear optimization problem that is not numerically tractable to solvers. However, by applying strong duality or KKT-based reformulations and big-M approach, the inner-level optimization problem in (6.16) can be eliminated, and this bilevel optimization problem can be reformulated as a mixed-integer linear programming problem, see [152, 153] for more details.

Remark 6.9. Compared to conventional RO formulations, both the proposed formulation (6.14) and existing DRO formulations, e.g., Wasserstein-based DRO formulation [156], result in increased computational burden, though for different reasons. In our proposed formulation, this computational burden arises primarily from the nonlinear constraint (6.15b), particularly when the first-stage decision variables $\mathbf{x}$ include binary components, making the master problem $\mathbf{MP}_{\mathrm{DRO}}$ (6.15) a mixed-integer nonlinear optimization problem. As for the existing DRO formulation, we take the Wasserstein-based DRO [156] as an example. When considering the fully adaptive recourse decision rule and large numbers of uncertainty samples, while the resulting optimization problem still retains the linearity of the original deterministic formulation, the increased computational burden is mainly caused by solving numerous subproblems in every iteration of the CCG algorithm since each uncertainty scenario will incur an optimization problem, see Algorithm 1 in the supplementary material [166]. The curse of dimensionality issue of the Wasserstein-based DRO formulation is also mentioned in [165].

6.5. SIMULATION RESULTS

This section presents three case studies to illustrate the effectiveness of the proposed schemes in this chapter. *Case Study 1* considers robust predictive control of building climate to demonstrate the effectiveness of the proposed approach in Section 6.3. *Case Study 2* and *Case Study 3* showcase the approach designed in Section 6.4 with robust location transportation planning and chemical process network planning problems, respectively.

All simulations are implemented on an Intel Xeon W-2223 CPU at 3.6GHz with 16GB RAM. Optimization problems are modelled via Python package `gurobipy` and solved via `Gurobi 11.0` [74]. The values of the parameters used in our case studies are provided in the Appendix 6.A.4, and the supplementary material [166].

6.5.1. CASE STUDY 1: ROBUST PREDICTIVE CONTROL OF BUILDING CLIMATE

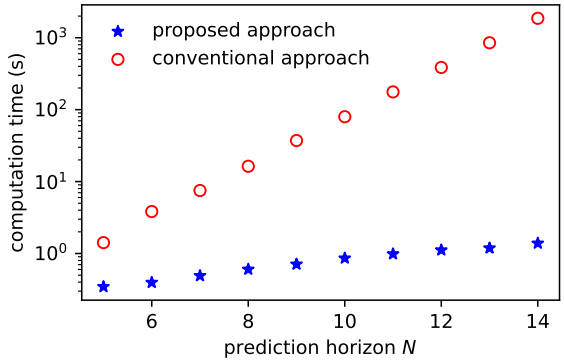
This case study considers robust predictive control of building climate. Building systems suffer from weather uncertainties, such as prediction errors of ambient temperature, solar irradiation, etc. Properly considering these uncertainties can improve indoor climate comfort.

A building climate predictive control problem in its deterministic form can be formulated as [11]

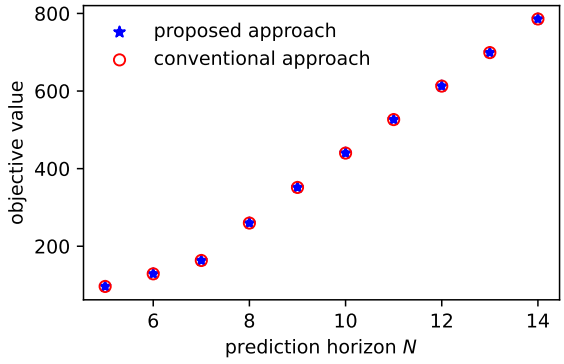
$$\begin{aligned} \min_{\mathbf{u}_t} \quad & \sum_{t=1}^{N} l(\mathbf{s}_t, \mathbf{u}_t) \\ \text{s.t.} \quad & \mathbf{s}_{t+1} = \mathbf{\Phi}\mathbf{s}_t + \mathbf{\Gamma}_u \mathbf{u}_t + \mathbf{\Gamma}_w \mathbf{w}_t + \mathbf{\Gamma}_v \mathbf{v}_t, \\ & \underline{\mathbf{s}}_t \le \mathbf{s}_t \le \bar{\mathbf{s}}_t, \quad \underline{\mathbf{u}}_t \le \mathbf{u}_t \le \bar{\mathbf{u}}_t, \\ & \forall \mathbf{v}_t \in \mathcal{V}_t, \ t = 1, \cdots, N \end{aligned}$$

where $l(\cdot,\cdot)$ is the stage cost function, $\mathbf{s}_t$ is the system states consisting of indoor temperature, roof temperature, wall temperature and floor temperature; $\mathbf{u}_t$ denotes the heating power, $\mathbf{w}_t$ is the predicted ambient conditions, $\mathbf{v}_t$ is the prediction error of ambient temperature, $\mathcal{V}_t$ denotes the uncertainty set, N is the length of prediction horizon, $\underline{\mathbf{s}}_t/\underline{\mathbf{u}}_t$ and $\bar{\mathbf{s}}_t/\bar{\mathbf{u}}_t$ are lower bound and upper bound of system states/control inputs, respectively. The stage cost function is defined as $l(\mathbf{s}_t, \mathbf{u}_t) := \mathbf{u}_t$ to minimize energy usage, system state constraints are defined to keep the indoor temperature above 21°C during 7:00 - 18:00 and above 15°C during the remaining hours. Heating power constraints are $0 \le \mathbf{u}_t \le 150$. The values of system matrices $(\mathbf{\Phi}, \mathbf{\Gamma}_u, \mathbf{\Gamma}_w, \mathbf{\Gamma}_v)$ are adopted from [11].

In the simulation, uncertainties $\mathbf{v}_t$ $(t = 1, \cdots, N)$ are assumed to be I.I.D., and the uncertainty set $\mathcal{V}_t$ for $\mathbf{v}_t$ is a union of 2 subsets. As a result, the total number of subsets for the stacked uncertainty $\mathbf{v} = [\mathbf{v}_1^{\mathrm{T}}, \cdots, \mathbf{v}_N^{\mathrm{T}}]^{\mathrm{T}}$ is 2^N. Two schemes are considered in this case study for solving the building climate control problem. One is the conventional RO formulation with the explicit description of each uncertainty subset (6.4) and is solved via the CCG-based algorithm in [138]. Another one is our proposed formulation (6.6) and is solved via Algorithm 6.1. To fully demonstrate the computational efficiency of the proposed scheme, different values of the prediction horizon N are tested. Simulation results are shown in Fig. 6.1. It can be seen that, with the increase of the prediction horizon N,



(a) computational time.



(b) objective values

Figure 6.1: Simulation results of *Case Study 1* with the proposed approach and the conventional approach.

the computational time of applying the conventional approach increases exponentially and is much larger than that of our proposed formulation since 2^N numbers of subproblems have to be solved in each algorithm iteration. In contrast, the computation time with our proposed formulation (6.6) and Algorithm 6.1 only increases linearly with the prediction horizon N. Besides, Fig. 6.1(b) shows that both approaches give the same objective function, which indicates that the proposed formulation does not sacrifice optimality while remarkably improving the computational efficiency.

6.5.2. CASE STUDY 2: ROBUST LOCATION TRANSPORTATION PLANNING

In this subsection, the effectiveness of the proposed formulation (6.14) and the corresponding Algorithm 6.3 is validated via a robust location transportation planning problem, which is also considered as a benchmark problem in [153]. The deterministic for-

Table 6.1: Computational results of different RO formulations for *Case Study 2*.

	conventional RO formulation (Scheme 1)	existing DRO formulation with Wasserstein metric (Scheme 2)	the proposed RO formulation (6.14)	
			with (6.15b) (Scheme 3)	with (6.17) (Scheme 4)
Optimal Objective	36632	35238	35482	35482
Decision Variable $\mathbf{x}$	[1,0,1,274,0,570]	[1,0,1,324,0,520]	[1,0,1,364,0,480]	[1,0,1,364,0,480]
CPU Time (s)	0.40	425.80	17.52	7.06
Iterations	2	3	2	2

mulation of this problem is

$$\begin{aligned}
\min\ & 400x_1 + 414x_2 + 326x_3 + 18x_4 + 25x_5 + 20x_6 \\
& + 22y_{11} + 33y_{12} + 24y_{13} + 33y_{21} + 23y_{22} + 30y_{23} + 20y_{31} + 25y_{32} + 27y_{33} && \text{(6.18a)} \\
\text{s.t. } & x_{i+3} \le 800x_i, \quad i = 1,2,3, && \text{(6.18b)} \\
& \sum_j y_{ij} \le x_{i+3}, \quad \forall i = 1,2,3, && \text{(6.18c)} \\
& \sum_i y_{ij} \ge d_j + 40 * v_j && \text{(6.18d)} \\
& x_i \in \mathbb{B}, \quad x_{i+3} \le 0, \quad i = 1,2,3, && \text{(6.18e)} \\
& y_{ij} \ge 0, \quad \forall i = 1,2,3, \quad j = 1,2,3 && \text{(6.18f)}
\end{aligned}$$

where binary first-stage decision variables $(x_1, x_2, x_3) \in \mathbb{B}^3$ determine the location of the facilities; continuous first-stage decision variables $(x_4, x_5, x_6) \in \mathbb{R}^3$ denote the facility capacities; recourse continuous decision variables $y_{ij} \in \mathbb{R}$ ($i = 1,2,3,\ j = 1,2,3$) are transportations, $[d_1, d_2, d_3] = [206, 274, 220]$ are basic demands, and $v_j \in \mathbb{R}$ ($j = 1,2,3$) are scaled demand uncertainties.

In our simulation, four schemes are considered:

- Scheme 1: the conventional RO formulation (6.1).
- Scheme 2: the Wasserstein-based DRO formulation [156].
- Scheme 3: the proposed formulation (6.14) and Algorithm 6.3 with the constraints (6.15b).
- Scheme 4: the proposed formulation (6.14) and Algorithm 6.3 with the reformulated constraints (6.17).

The uncertainty set $\mathcal{V}$ for $[v_1, v_2, v_3]^{\mathrm{T}}$ is supposed to have 4 polyhedral subsets. For our proposed schemes (Scheme 3 & 4), the nominal probability distribution for defining the ambiguity set $\mathscr{P}$ is selected as $\bar{\mathbf{p}} = [0.5, 0.1, 0.2, 0.2]$ to represent a practical situation where uncertainty samples are unevenly distributed among different uncertainty subsets. The upper bound of KL divergence for defining the ambiguity set $\mathscr{P}$ in (6.13) is set as $\rho = 0.5$. For Scheme 2, since its implementation entails considering specific uncertainty scenarios, 1000 uncertainty samples in $\mathcal{V}$ are randomly generated, which is a reasonable choice for the uncertainty with dimension 3 [167, 168]. Since we mainly focus on demonstrating the computational efficiency of the DRO formulation, the Wasserstein distance ε, which does not affect its computational efficiency, is set as $\varepsilon = 1$ to give

a comparable performance as with the other schemes. It should be pointed out that the Wasserstein-based DRO formulation (Scheme 2) can be expected to outperform our proposed methods with a well-calibrated radius. Nevertheless, within the Wasserstein-based DRO framework, achieving both low conservatism and strong out-of-sample performance hinges on carefully tuning the Wasserstein radius and considering a large number of representative uncertainty samples to avoid overfitting to the empirical distribution, which will lead to increased computational cost for solving the corresponding optimization problem.

Simulation results are summarized in Table 6.1. It can be seen that these different RO formulations derive distinct first-stage decision variables. Compared with the conventional RO formulation (Scheme 1) for optimizing the worst-case performance, the remaining formulations give less conservative solutions, i.e., smaller optimal objective values, at the price of increased computational burden. As discussed in Remark 8, the increased computational burden for the DRO formulation (Scheme 2) is due to its curse of dimensionality when considering the fully adaptive decision rule and a large number of uncertainty samples [156, 165], while for our proposed schemes (Scheme 3 & 4), the increased computational burden is caused by the nonlinear constraint (6.15b), especially when the first-stage decision variable $\mathbf{x}$ contains integer ingredients. It can be seen that, in comparison with the DRO formulation (Scheme 2), our proposed formulations (Scheme 3 & 4) are much more computationally efficient. Furthermore, the computational time of Scheme 4 is less than that of Scheme 3, which indicates the efficacy of the proposed reformulation (6.17) for improving the computational efficiency of Algorithm 6.3. It is worth noting that, in comparison with Scheme 2, the reduced computation time of our proposed scheme does not simply result from switching the divergence measure. Without altering the problem structure, directly replacing the Wasserstein metric with the KL divergence in Scheme 2 leads to the same issue observed in the Wasserstein-based design. Namely, as shown in Algorithm 1 in the supplementary material, both the number of subproblems and the number of constraints in the master problem are proportional to the number of uncertainty samples.

6.5.3. CASE STUDY 3: CHEMICAL PROCESS NETWORK PLANNING

To further illustrate the viability and effectiveness of the proposed formulation (6.14) and Algorithm 6.3, a chemical process network planning (CPNP) problem is investigated. The CPNP problem is a typical engineering problem that fits into RO settings and has been considered as a benchmark problem in several existing literature, see [29, 35, 61].

A chemical process consists of raw materials, intermediate chemicals, final products, and multiple interconnected processes. The design objective of CPNP is to maximize the net present value (NPV) of the entire network while respecting system constraints for all possible uncertainties. Our case study considers a chemical process network consisting of 8 processes and 7 chemicals. The network sketch is shown in Fig. 6.2, where chemicals are denoted as red circles $(A, B, \cdots, G)$, processes are blue rectangles $(1, 2, \cdots, 8)$, and process flows are indicated as arrows. Among all chemicals, (A, E) are raw materials, and (D, G) are products.

The RO problem for CPNP within our proposed design framework (6.14) can be for-

mulated as

$$\max_{QE_{i,t},Y_{i,t}} \sum_i \sum_t (-\alpha_{i,t} QE_{i,t} - \beta_{i,t} Y_{i,t}) +$$

$$\min_{\mathbf{p}\in\mathscr{P}} \mathbb{E}_{\mathbf{p}} \Big[\min_{\mathbf{v}\in\mathscr{V}_k} \max_{\substack{P_{j,t},QE_{i,t} \\ S_{j,t},W_{i,t}}} \Big(-\sum_i \sum_t \gamma_{i,t} W_{i,t} - \sum_j \sum_t \varphi_{j,t} P_{j,t} + \sum_j \sum_t \tau_{j,t} S_{j,t} \Big) \Big] \tag{6.19a}$$

$$\text{s.t. } qe^L_{i,t} \cdot Y_{i,t} \le QE_{i,t} \le qe^U_{i,t} \cdot Y_{i,t},\ \forall i \in \mathscr{I},\ \forall t \in \mathscr{T} \tag{6.19b}$$

$$Q_{i,t} = Q_{i,t-1} + QE_{i,t},\ \forall i \in \mathscr{I},\ \forall t \in \mathscr{T} \tag{6.19c}$$

$$\sum_t Y_{i,t} \le ce_i,\ \forall i \in \mathscr{I} \tag{6.19d}$$

$$\sum_i \alpha_{i,t} \cdot QE_{i,t} + \beta_{i,t} Y_{i,t} \le cb_t,\ \forall t \in \mathscr{T} \tag{6.19e}$$

$$W_{i,t} \le Q_{i,t},\ \forall i \in \mathscr{I},\ \forall t \in \mathscr{T} \tag{6.19f}$$

$$P_{j,t} - \sum_i \kappa_{i,j} W_{i,t} - S_{j,t} = 0,\ \forall j \in \mathscr{J},\ \forall t \in \mathscr{T} \tag{6.19g}$$

$$P_{j,t} \le su_{j,t},\ S_{j,t} \le du_{j,t}, \forall j \in \mathscr{J},\ \forall t \in \mathscr{T} \tag{6.19h}$$

$$QE_{i,t}, Q_{i,t}, P_{j,t}, W_{i,t}, S_{j,t} \ge 0, Y_{i,t} \in \{0,1\},$$
$$\forall i \in \mathscr{I},\ \forall j \in \mathscr{J},\ \forall t \in \mathscr{T}, \tag{6.19i}$$

$$\mathbf{v} = \{du_{i,t}, su_{i,t}\} \in \mathscr{V},\ \forall i \in \mathscr{I},\ \forall t \in \mathscr{T} \tag{6.19j}$$

where $\mathscr{I} = \{1, \cdots, I\}$, $\mathscr{J} = \{1, \cdots, J\}$, and $\mathscr{T} = \{1, \cdots, T\}$ with I, J and T as the total numbers of processes, chemicals and planning periods, respectively. All notations in the above equations are explained in Table 6.2. The objective function (6.19a) maximizes the NPV consisting of investment cost, operation cost, purchase cost of raw chemicals, and sale profit; constraint (6.19b) specifies the upper and lower bounds of capacity expansion for all processes and time periods; constraint (6.19c) updates the total available capacity of each process; constraint (6.19d) limits the largest process expansion times; constraint (6.19e) ensures the process expansion costs are within available budgets; inequality (6.19f) limits the production level of each process within its total capacity; equality (6.19g) models the mass balance of all chemicals; constraints (6.19h) ensures that the amounts of purchased and sold chemicals are limited by the available market supply and demand, respectively; constraints (6.19i) indicates all non-negative continuous decision variables and binary decision variables. For more detailed explanations of chemical process networks, please refer to [29, 61].

The uncertainty variables considered in our design are the market supply limit and demand limit $(su_{i,t}, du_{i,t})$ for all raw materials and products over the planning period. Expansion decisions $(QE_{i,t}, Y_{i,t})$ are first-stage decision variables, and other remaining variables $(P_{j,t}, S_{j,t}, Q_{i,t}, W_{i,t})$ are recourse decision variables. Our case study considers a 5−year planning period with each planning period as 1 year. The uncertainties $(su_{j,t}, du_{j,t})$ are assumed to be independent and reside in 4 uncertainty sets.

As in *Case Study 2*, we consider the same four schemes. For the DRO formulation in Scheme 2, 1000 uncertainty scenarios are randomly generated, and the Wasserstein distance ε defining the size of the distribution ambiguity set is $\varepsilon = 300$ to give a comparable objective value with the other schemes. It should be noted that the number of uncer-

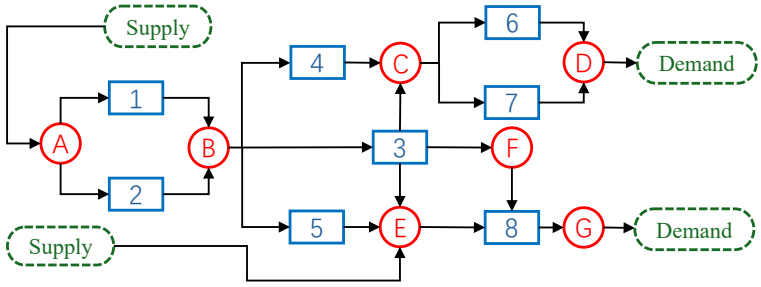


Figure 6.2: The chemical process network for *Case Study 3*.

Table 6.2: Notations in (6.19)

decision variables			
$QE_{i,t}$	amount of capacity expansion	$Y_{i,t}$	binary decision of capacty expansion
$P_{j,t}$	chemical purchase amount	$S_{j,t}$	chemical sale amount
$W_{i,t}$	operating level	$Q_{i,t}$	total process capacity
$su_{i,t}$	market supply limit	$du_{i,t}$	market demand limit
parameters			
$\alpha_{i,t}$	variable investment cost	$\beta_{i,t}$	fixed investment cost
$\gamma_{j,t}$	operating cost	$\varphi_{j,t}$	purchase cost
$\tau_{j,t}$	sale price	$qe^L_{i,t}$	capacity expansion lower bound
$qe^U_{i,t}$	capacity expansion upper bound	ce_i	expansion number limit
cb_t	expansion cost budget	$\kappa_{i,j}$	mass balance coefficient
i	index of the i-th process	j	index of the j-th chemical
t	index of the t-th time period		

tainty samples considered is far smaller than needed for a proper approximation of the uncertainty probability distribution, given the 20-dimensional uncertainty space [167, 168]. However, as shown in our simulation results, even this limited number of samples can be computationally demanding for solving the corresponding DRO problem. For our proposed schemes (Scheme 3 & 4), the parameters defining the ambiguity set $\mathscr{P}$ are $\bar{\mathbf{p}} = [0.5, 0.1, 0.2, 0.2]$ and $\rho = 0.5$.

Simulation results are shown in Table 6.3, from which it can be concluded that consistent results are obtained as in *Case Study 2*. Compared with the conventional RO formulation (Scheme 1), the remaining formulations (Scheme 2, 3 & 4) give less conservative solutions, i.e., a larger value of Max. NPV at the price of increased computational burden (longer CPU time). Compared with the Wasserstein-based DRO formulation (Scheme 2), our proposed formulations (Scheme 3 & 4) are less computationally demanding. In addition, compared with the original constraints (6.15b), the reformulation in (6.17) is effective in improving the computational efficiency of Algorithm 6.3.

Table 6.3: Computational results of different RO formulations for *Case Study 3*.

	conventional RO formulation (Scheme 1)	existing DRO formulation with Wasserstein metric (Scheme 2)	the proposed RO formulation (6.14)	
			with (6.15b) (Scheme 3)	with (6.17) (Scheme 4)
Max. NPV ($MM)	221	284	288	288
CPU Time (s)	17.63	8758.07	195.86	183.17
Iterations	2	5	2	2

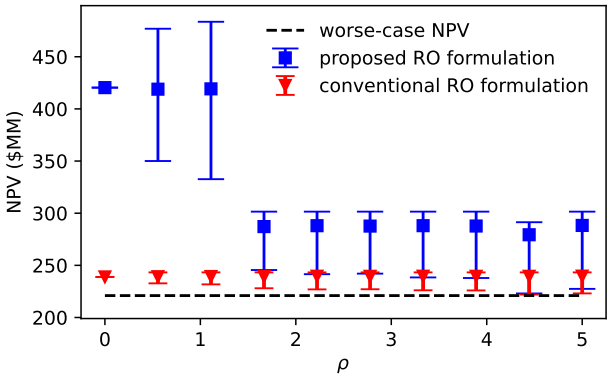


Figure 6.3: Error bar of the expected NPV with different ambiguity set size ρ.

Furthermore, different values of the ambiguity set size ρ are tested to analyze its influence on the conservatism of the optimal solution for our proposed formulation. In the simulation, for each ρ, 5000 random trials are implemented to sample the probability distribution $\mathbf{p}$ that resides in the corresponding ambiguity set. For all admissible samples of $\mathbf{p}$, the corresponding performance metric, which is defined as the expected NPV: $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \sum_k \mathbf{p}_k \cdot \mathbf{b}^{\mathrm{T}}\mathbf{y}_k^*$ with $\mathbf{y}_k^*$ as the optimal recourse variable w.r.t. the worst-case uncertainty in the k-th uncertainty subset, is computed. Fig. 6.3 depicts the error bar (mean value and envelope) of the expected NPV among all feasible samples of $\mathbf{p}$, and the worst-case NPV for all possible uncertainties. It can be seen that, compared with the conventional RO formulation, our proposed approach consistently leads to less conservative solutions for all ρ. Namely, the proposed approach gives a higher mean, minimal, and maximal NPV than the conventional RO formulation. Further, it can be seen from Fig. 6.3 that, with the increase of ρ (the size of the ambiguity set), the minimal NPV of the proposed formulation (6.14) converges to the worse NPV computed via the conventional RO formulation, which is consistent with our statement in Remark 6.

6.6. CONCLUSIONS

This chapter investigates the RO problems where uncertainty sets comprise multiple subsets, focusing on two separate questions: 1) *how to address the computational challenge posed by exponentially increasing uncertainty subsets in RO-based predictive control?* and 2) *how to mitigate the conservatism of the robust optimal solutions by leveraging*

the structure of the uncertainty set?

To address the first question, we propose a monolithic mixed-integer representation of the uncertainty set. Unlike conventional formulations requiring a separate optimization problem for each subset, our method only solves a single mixed-integer optimization problem to compute the worst-case uncertainty scenario over all subsets. This method is particularly advantageous for RO-based predictive control, where the number of uncertainty subsets could increase exponentially with the prediction horizon.

For the second question, we formulate an innovative objective function exploiting the structure of the multi-subset uncertainty set by combining the existing RO and DRO formulations. The proposed formulation achieves less conservative solutions than conventional RO formulations while showing more computational efficiency than conventional DRO formulations. Besides, a CCG-based algorithm is developed to solve the resulting optimization problem efficiently.

Numerical experiments related to the above research questions are conducted to extensively demonstrate the effectiveness of the proposed schemes.

6.A. Appendix

6.A.1. Proof of Theorem 6.1

Proof: The main difference between the proposed Algorithm 6.1 with the conventional CCG algorithm in [138, 153] is the definition of the subproblem (6.8). With the mixed-integer representation of the uncertainty set $\mathcal{V}$, each set of feasible binary variables $\{\delta^*_{t,k} \mid t = 1,\cdots,N, k = 1,\cdots,K\}$ uniquely defines a subset of $\mathcal{V}$, and vice versa. As a result, by solving the subproblem **SP**1 (6.8) or **SP**2 (6.9), the worst-case uncertainty scenario $\mathbf{v}^* = [\mathbf{v}_1^{*\mathrm{T}},\cdots,\mathbf{v}_N^{*\mathrm{T}}]^{\mathrm{T}}$ among all subsets together with the corresponding uncertainty subset, defined by $\{\delta^*_{t,k} \mid t = 1,\cdots,N, k = 1,\cdots,K\}$, that contains the uncertainty scenario $\mathbf{v}^*$ will be computed. Based on Assumption 1, it can be readily concluded that there are finite vertices of the uncertainty set $\mathcal{V}$. Then, it follows a similar proof as shown in [153] that Algorithm 6.1 will solve the RO problem (6.6) within finite iterations. This completes the proof. □

6.A.2. Proof of Theorem 6.2

Proof: For the constraint (6.14b), it can be rewritten as

$$\max_{\mathbf{p}\in\mathscr{P}} \sum_{k=1}^{K} \mathbf{p}_k \cdot \left(\max_{\mathbf{v}\in\mathcal{V}_k} \mathbf{b}^{\mathrm{T}}\mathbf{y}\right) \tag{6.20}$$

Based on Lemma 19.1 in [154], the inequality (6.20) is equivalent to finding μ and $\nu \geq 0$ such that

$$\mu + \rho\nu + \nu \sum_k \bar{\mathbf{p}}_k \exp\Big(\frac{\mathscr{C}(\mathbf{x},\mathcal{V}_k) - \mu}{\nu} - 1\Big) \leq \eta \tag{6.21}$$

where

$$\mathscr{C}(\mathbf{x},\mathcal{V}_k) := \max_{\mathbf{v}\in\mathcal{V}_k} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \tag{6.22a}$$

$$\text{s.t. } \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \mathbf{M}\mathbf{v} \leq \mathbf{h} \tag{6.22b}$$

Assumption 6.2 implies that there are finite vertices of the uncertainty set $\mathcal{V}$. Besides, it can be seen that the left-hand side of (6.21) is a monotonously increasing function w.r.t. $\mathscr{C}(\mathbf{x},\mathcal{V}_k)$, which is LP w.r.t. $\mathbf{v}$. Consequently, it can be concluded that the optimal $\mathbf{v}$ are taken from the vertices of $\mathcal{V}$. By listing all finite vertices of each uncertainty subset $\mathcal{V}_k$, denoted as $\mathscr{S}_k := \{\mathbf{v}_{k,1},\cdots,\mathbf{v}_{k,H_k}\}$ with H_k as the number of all vertices, the RO problem (6.14) can be rewritten as

$$\min_{\substack{\mathbf{x},\mathbf{y}_{k,i_k} \\ \eta,\mu,\nu}} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \eta \tag{6.23a}$$

$$\text{s.t. } \mu + \rho\nu + \nu \sum_k \bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}_{k,i_k} - \mu}{\nu} - 1\right) \leq \eta, \tag{6.23b}$$

$$\mathbf{A}\mathbf{x} \leq \mathbf{q}, \tag{6.23c}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y}_{k,i_k} + \mathbf{M}\mathbf{v}_{k,i_k} \leq \mathbf{h}, \tag{6.23d}$$

$$\forall i_k \in \{1,\cdots,H_k\},\ \forall k \in \{1,\cdots,K\} \tag{6.23e}$$

where $\mathbf{y}_{k,i_k}$ is the optimal recourse decision variable w.r.t. the uncertainty scenario $\mathbf{v}_{k,i_k}$ ($i=1,\cdots,H_k$). Namely,

$$\begin{aligned}\mathbf{y}_{k,i_k} = \arg\min_{\mathbf{y}} & \mathbf{b}^{\mathrm{T}}\mathbf{y} \\ \text{s.t. } & \mathbf{T}\mathbf{x}+\mathbf{W}\mathbf{y}+\mathbf{M}\mathbf{v}_{k,i_k} \le \mathbf{h}.\end{aligned}$$

Since there are finite extreme uncertainty scenarios, the optimization problem (6.23) can be computed by iteratively listing all possible extreme uncertainty scenarios as in the conventional CCG algorithm [152, 153], which leads to Algorithm 6.3.

In the following, we will prove that Algorithm 6.3 can terminate within finite iterations. Namely, Algorithm 6.3 will either find out all extreme uncertainty scenarios or terminate when a repeated uncertainty scenario is observed. Assuming at the r-th iteration of the algorithm, the extreme uncertainty scenarios $\{\mathbf{v}_1^*,\cdots,\mathbf{v}_K^*\}$ computed by solving K subproblems (6.16) are observed in a previous iteration t ($t \le r-1$), then $UB = LB$ and the algorithm terminates.

Suppose that the optimal decision variables by solving $\mathbf{MP}_{\mathrm{DRO}}$ (6.15) at the r-th iteration are $(\mathbf{x}^*,\mu^*,\eta^*,\nu^*)$, which further lead to the optimal decision variables of K subproblems $\mathbf{SP}_{\mathrm{DRO}}^k$ $\{\mathbf{v}_1^*,\cdots,\mathbf{v}_K^*,\mathbf{y}_1^*,\cdots,\mathbf{y}_K^*\}$. It readily gives that

$$LB \le UB \le \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \mu^* + \rho\nu^* + \nu^*\sum_k \bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}_k^* - \mu^*}{\nu^*} - 1\right)$$

Since the worst-case uncertainty scenarios $\{\mathbf{v}_1^*,\cdots,\mathbf{v}_K^*\}$ was observed at a previous iteration, the optimal solution of the master problem $\mathbf{MP}_{\mathrm{DRO}}$ at $(r+1)$-th iteration will be the same as with the r-th iteration and satisfy

$$LB \ge \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \mu^* + \rho\nu^* + \nu^*\sum_k \bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}_k^* - \mu^*}{\nu^*} - 1\right)$$

As a result, it can be concluded that $LB = UB$ at the $(r+1)$-th iteration, and the algorithm terminates. This completes the proof. □

6.A.3. PROOF OF THE CONVEXITY OF CONSTRAINT (6.15b)

Proof: For brevity of analysis, it is sufficient to prove that the nonlinear function $f(x,y) := x\exp(\frac{y}{x}-1)$ with $x \ge 0$ is convex since the nonlinear constraint (6.15b) can be reformulated as

$$\mu + \rho\nu + \sum_k \bar{\mathbf{p}}_k f(\nu, g_{k,i}) \le \eta, \tag{6.24a}$$

$$g_{k,i} = \mathbf{b}^{\mathrm{T}}\mathbf{y}_{k,i} - \mu. \tag{6.24b}$$

For the nonlinear function $f(x,y) := x\exp(\frac{y}{x}-1)$, its Hessian matrix is

$$\nabla^2 f = \begin{bmatrix} \frac{y^2}{x^3}\exp(\frac{y}{x}-1) & -\frac{y}{x^2}\exp(\frac{y}{x}-1) \\ -\frac{y}{x^2}\exp(\frac{y}{x}-1) & \frac{1}{x}\exp(\frac{y}{x}-1) \end{bmatrix}. \tag{6.25}$$

It can be easily verified that $\nabla^2 f$ is positive semidefinite when $x \ge 0$. As a result, it follows from the *second-order condition* of convexity [78] that the nonlinear function $f(x,y)$ is convex, and hence constraint (6.15b) is convex. This completes the proof. □

6

6.A.4. Supplementary Materials for Simulation

Parameter Values Used in *Case Study 1*

The basic uncertainty set for the uncertainty $\mathbf{v}_t$ in each time period t consists two subsets $\mathcal{V}_{t,1}$ and $\mathcal{V}_{t,2}$ that are defined as

$$\mathcal{V}_{t,1} := \{\mathbf{v}_t | \mathbf{D}_1 \mathbf{v}_t \leq \mathbf{d}_1\}, \quad \mathcal{V}_{t,2} := \{\mathbf{v}_t | \mathbf{D}_2 \mathbf{v}_t \leq \mathbf{d}_2\} \tag{6.26}$$

where

$$\mathbf{D}_1 = \mathbf{D}_2 = [1,-1]^{\mathbf{T}}, \quad \mathbf{d}_1 = [2,0]^{\mathrm{T}}, \quad \mathbf{d}_2 = [0,2]^{\mathrm{T}}. \tag{6.27}$$

The system matrices defining the building thermal dynamics are

$$\mathbf{\Phi} = \begin{bmatrix} 0.0167 & 0.0048 & 0.1245 & 0.409 \\ 0.0005 & 0.0002 & 0.0039 & 0.0044 \\ 0.0253 & 0.0073 & 0.3321 & 0.0617 \\ 0.0244 & 0.0070 & 0.0526 & 0.3456 \end{bmatrix}, \mathbf{\Gamma}_u = \begin{bmatrix} 0.0986 \\ 0.0029 \\ 0.0288 \\ 0.0275 \end{bmatrix}, \mathbf{\Gamma}_w = \begin{bmatrix} 0.2536 & 0.4596 \\ 0.0070 & 0.9840 \\ 0.4450 & 0.1287 \\ 0.4477 & 0.1225 \end{bmatrix}, \mathbf{\Gamma}_v = \begin{bmatrix} 0.2536 \\ 0.0070 \\ 0.4450 \\ 0.4477 \end{bmatrix}.$$

Parameter Values Used in *Case Study 2*

The uncertainty set for the demand uncertainties consists of four uncertainty sets $\mathcal{V}_k := \{\mathbf{v} \mid \mathbf{D}_k \mathbf{v} \leq \mathbf{d}_k\}$ with parameters $(\mathbf{D}_k, \mathbf{d}_k)$ set as

$$\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D}_3 = \mathbf{D}_4 = [\mathbf{I}_3, -\mathbf{I}_3]^{\mathrm{T}} \tag{6.28}$$

$$\mathbf{d}_1 = [0.3, 0.3, 0.3, 0, 0, 0]^{\mathrm{T}}, \quad \mathbf{d}_2 = [1.2, 1.2, 1.2, -1, -1, -1]^{\mathrm{T}}, \tag{6.29}$$

$$\mathbf{d}_3 = [1, 0.3, 0.3, -0.8, 0, 0], \quad \mathbf{d}_4 = [0.3, 1, 0.3, 0, -0.7, 0]^{\mathrm{T}}. \tag{6.30}$$

Parameter Values and Algorithms Used in *Case Study 3*

The mass balance coefficient matrix is

$$\boldsymbol{\kappa} = \begin{bmatrix} 0.63 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0.58 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.25 & -4 & 0 & -1 & -1.5 & 0 \\ 0 & 0.1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0.24 & -1 & 0 & 0 & 0 \\ 0 & 0 & 2.3 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.93 & 0.77 & -1 \end{bmatrix} \tag{6.31}$$

The uncertainty set consists of four subsets that are defined as

$$\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D}_3 = \mathbf{D}_4 = \mathbf{I}_5 \otimes \begin{bmatrix} \mathbf{I}_{\text{in}} & \mathbf{0} \\ -\mathbf{I}_{\text{in}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\text{out}} \\ \mathbf{0} & -\mathbf{I}_{\text{out}} \end{bmatrix} \tag{6.32}$$

where $\mathbf{I}_{\text{in}}$ and $\mathbf{I}_{\text{out}}$ are identity matrices with the number of columns as the number of chemical process inputs and outputs, respectively.

$$\mathbf{d}_1 = \mathbf{e}_5 \otimes [130, 112, -110, -88, 35, 49, -25, -41]^{\mathrm{T}}, \tag{6.33}$$

$$\mathbf{d}_2 = \mathbf{e}_5 \otimes [76, 68, -64, -56, 40, 37, -30 - 27]^{\mathrm{T}}, \tag{6.34}$$

$$\mathbf{d}_3 = \mathbf{e}_5 \otimes [65, 35, -55, -25, 23, 14, -17, -6]^{\mathrm{T}}, \tag{6.35}$$

$$\mathbf{d}_4 = \mathbf{e}_5 \otimes [52, 52, -48, -48, 27, 27, -23, -23]^{\mathrm{T}}. \tag{6.36}$$

CCG ALGORITHM FOR SOLVING WASSERSTEIN METRIC-BASED DRO PROBLEM

For the DRO problem

$$\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \max_{f(\mathbf{v})\in\mathscr{P}} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \tag{6.37a}$$

$$\text{s.t. } \mathbf{A}\mathbf{x} \le \mathbf{q}, \tag{6.37b}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \mathbf{M}\mathbf{v} \le \mathbf{h} \tag{6.37c}$$

where the ambiguity set is defined as

$$\mathscr{P} = \{\mathbb{Q} \in M(\mathscr{V}) : d_W(\mathbb{Q}, \mathbb{Q}_0) \le \varepsilon\}$$

$$d_W(\mathbb{Q}, \mathbb{Q}_0) = \inf_{\Pi}\left(\int_{\mathscr{V}^2} ||\mathbf{v} - \mathbf{v}^0|| \, \Pi(\mathrm{d}\mathbf{v}, \mathrm{d}\mathbf{v}^0) \,\middle|\, \begin{array}{l} \Pi \text{ is a joint distribution of} \\ \mathbf{v} \text{ and } \mathbf{v}^0 \text{ with marginals } \mathbb{Q} \text{ and } \mathbb{Q}_0 \end{array} \right)$$

with ε as the Wasserstein distance. According to the results in [156], the above optimization can be reformulated as

$$\min_{\mathbf{x},\lambda,\mathbf{z}_i,s_i} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i \tag{6.38a}$$

$$\text{s.t. } \sup_{\mathbf{v}\in\mathscr{V}} \left[\min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} - \mathbf{z}_i^{\mathrm{T}}\mathbf{v} \right] + \mathbf{z}_i^{\mathrm{T}}\hat{\mathbf{v}}_i \le s_i, \ i = 1, 2, \cdots, N, \tag{6.38b}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \mathbf{M}\mathbf{v} \le \mathbf{h}, \tag{6.38c}$$

$$\mathbf{A}\mathbf{x} \le \mathbf{q}, \tag{6.38d}$$

$$||\mathbf{z}_i||_* \le \lambda, \ i = 1, 2, \cdots, N, \tag{6.38e}$$

where N is the number of uncertainty samples $\hat{\mathbf{v}}_i$. Then, based on the above formulation and applying a *column-and-constraint generation* approach, the DRO problem can be solved via Algorithm 6.3.

MP:

$$\min_{\mathbf{x},\lambda,s_i,\mathbf{z}_{t,i},\mathbf{y}_{t,i}} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i \tag{6.39a}$$

$$\text{s.t. } \mathbf{b}^{\mathrm{T}}\mathbf{y}_{t,i} - \mathbf{z}_i^{\mathrm{T}}\mathbf{v}_{t,i}^* + \mathbf{z}_i^{\mathrm{T}}\hat{\mathbf{v}}_i \le s_i, \tag{6.39b}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y}_{t,i} + \mathbf{M}\mathbf{v}_{t,i} \le \mathbf{h}, \tag{6.39c}$$

$$\mathbf{A}\mathbf{x} \le \mathbf{q}, \tag{6.39d}$$

$$i = 1, \cdots, N, \quad t = 1, \cdots, r. \tag{6.39e}$$

SP:

$$\mathscr{C}(\mathbf{x}^*, \mathbf{z}_i^*) := \max_{\mathbf{v}_i} \left[\min_{\mathbf{y}_i} \mathbf{b}^{\mathrm{T}}\mathbf{y}_i - \mathbf{v}_i^{\mathrm{T}}\mathbf{z}_i^* \right] \tag{6.40a}$$

$$\text{s.t. } \mathbf{T}\mathbf{x}^* + \mathbf{W}\mathbf{y}_i + \mathbf{M}\mathbf{v}_i \le \mathbf{h}. \tag{6.40b}$$

Algorithm 6.3 *column-and-constraint generation* algorithm for DRO problem (6.38).

Input: suboptimality gap ϵ

Output: the optimal decision variable $\mathbf{x}^*$ and objective function value $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \lambda^*\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i^*$

Set $LB = -\infty$, $UB = \infty$, $r = 0$

while $|UB - LB| > \epsilon$ **do**

  Solve **MP** in (6.39) to derive solutions $\{\mathbf{x}^*, \lambda^*, s_i^*, \mathbf{z}_i^*\}$ and update $LB = \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \lambda^*\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i^*$

  for $i = 1, \cdots, N$ **do**

    Solve **SP** $\mathscr{C}(\mathbf{x}^*, \mathbf{z}_i^*)$ in (6.40) to derive solutions $\{\mathbf{v}_i^*, \mathbf{y}_i^*\}$

  end for

  Update UB as

$$UB = \min\left\{UB, \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \lambda^*\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{b}^{\mathrm{T}}\mathbf{y}_i^* - \mathbf{z}_i^{*\mathrm{T}}\mathbf{v}_i^* + \mathbf{z}_i^{*\mathrm{T}}\hat{\mathbf{v}}_i\right)\right\}$$

  Create decision variables $\{\mathbf{y}_{r,i} \mid i = 1, \cdots, N\}$, set parameters $\mathbf{v}_{r,i} = \mathbf{v}_i^*$ $(i = 1, \cdots, N)$, and add the following constraints to **MP** in (6.39)

$$\begin{cases} \mathbf{b}^{\mathrm{T}}\mathbf{y}_{r,i} - \mathbf{z}_i^{\mathrm{T}}\mathbf{v}_{r,i} + \mathbf{z}_i^{\mathrm{T}}\hat{\mathbf{v}}_i \le s_i, \\ \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y}_{r,i} + \mathbf{M}\mathbf{v}_{r,i}^* \le \mathbf{h}. \end{cases}$$

  $r \leftarrow r + 1$

end while

Return: $\mathbf{x}^*$ and $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \lambda^*\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i^*$

7

DISTRIBUTIONALLY ROBUST SYSTEM LEVEL SYNTHESIS WITH OUTPUT FEEDBACK AFFINE CONTROL POLICY

This chapter studies the finite-horizon robust optimal control of constrained linear systems subject to model mismatch and additive stochastic disturbances. Utilizing the system level synthesis (SLS) parameterization, we propose a novel SLS design using an output-feedback affine control policy and extend it to a distributionally robust setting to improve system resilience by minimizing the cost function while ensuring constraint satisfaction against the worst-case uncertainty distribution. The scopes of model mismatch and stochastic disturbances are quantified using the 1-norm and a Wasserstein metric-based ambiguity set, respectively. For the closed-loop dynamics, we analyze the distributional shift between the predicted output-input response – computed using nominal parameters and empirical disturbance samples – and the actual closed-loop distribution, highlighting its dependence on model mismatch and SLS parameterization. Assuming convex and Lipschitz continuous cost functions and constraints, we derive a tractable reformulation of the distributionally robust SLS (DR-SLS) problem by leveraging tools from robust control and distributionally robust optimization (DRO). Numerical experiments validate the performance and robustness of the proposed approach.

This chapter is based on the paper: Li, Y., Shi, J., Jones, C. N., Yorke-Smith, N., & Keviczky, T. (2025). Distributionally Robust System Level Synthesis With Output Feedback Affine Control Policy. arXiv preprint arXiv:2508.05466.

7.1. Introduction

The growing complexity of modern control systems has increasingly challenged the applicability and effectiveness of traditional control strategies. In contrast, the advancement in onboard computational capabilities and numerical solvers has opened up new opportunities for real-time optimization-based control approaches. Among them, Model Predictive Control (MPC) has emerged as a powerful and versatile framework, thanks to its ability to explicitly handle multi-objective performance criteria while systematically accounting for system constraints. MPC's flexibility in incorporating prediction models, actuator limitations, and safety requirements makes it particularly well-suited for complex and safety-critical applications such as autonomous systems, robotics, and energy management [37, 39, 169]. These advantages have driven a growing interest in both theoretical developments and practical implementations of MPC across various domains.

One of the fundamental challenges in control design, including MPC, is the effective handling of uncertainties. In practice, systems are inevitably subject to various sources of uncertainty, such as model mismatch and external disturbances. The performance and reliability of a control scheme critically depend on how well these uncertainties are addressed.

Broadly speaking, methodologies for dealing with uncertainty can be categorized into two main approaches. The robust approach treats uncertainties as belonging to predefined deterministic sets and seeks to optimize control performance under the worst-case scenario [8]. While this method is computationally efficient in general and does not require explicit knowledge of the uncertainty distributions, it often leads to overly conservative solutions. Alternatively, the stochastic approach models uncertainties probabilistically and aims to optimize the expected performance [32]. This can reduce conservatism but typically relies on strong assumptions about the underlying uncertainty distributions—closed-form or tractable solutions are generally available only in specific cases. To address this limitation, the randomized (sample-based) approach has gained traction when a large number of uncertainty samples are available. It approximates the stochastic optimization problem by solving a large-scale optimization problem constructed from sampled scenarios [33, 134].

To leverage the complementary strengths of both robust and stochastic control methods, the distributionally robust (DR) framework has been proposed. In this setting, the uncertainty is assumed to follow an unknown probability distribution that lies within a specified ambiguity set, which captures plausible distributions based on prior knowledge or data. The control objective is then to optimize the expected performance under the worst-case distribution within this ambiguity set [43, 156, 170]. Recent studies, e.g., [45–50], have explored the integration of DR methods into predictive control schemes. These works primarily focus on mitigating the impact of additive disturbances by assuming perfect knowledge of the system dynamics and optimizing open-loop control actions accordingly.

To enhance robustness and reduce conservatism in optimal control design, closed-loop feedback policies are generally preferred over open-loop control actions in both robust optimal decision-making and Model Predictive Control (MPC) frameworks [8, 171]. However, incorporating feedback policies directly into online MPC optimization

introduces significant computational challenges, as it results in nonconvex optimization problems – even when the system dynamics and control laws are linear. For linear systems, the system level synthesis (SLS) framework provides a parameterization of the closed-loop system. Other than the conventional MPC design focusing on optimizing controllers, SLS-based design alternatively focuses on directly optimizing the whole closed-loop performance by solving convex optimization problems and reconstructing control policies from the optimized closed-loop response via linear transformations [172]. This shift in perspective has facilitated progress in distributed and robust predictive control design [172–174].

As noted in [175], a limitation of existing SLS methods is their reliance on purely linear control policies, primarily due to their tractability. However, linear policies lack the expressiveness to approximate piecewise-affine solutions, which frequently arise in Model Predictive Control (MPC) problems [171]. Moreover, current linear-policy-based SLS frameworks cannot be directly extended to the affine case without incurring computational issues, see [172, 176–178]. Notable exceptions are the works in [175, 179]. In [175], the authors developed an SLS framework using state feedback control law under the assumption of full state measurability and perfect knowledge of system models. In [179], a linear-quadratic SLS approach with linear output feedback was proposed, with a discussion on its extension to affine policies. However, this method is restricted only to LQR problems and to systems with additive input disturbances.

Motivated by the above limitations, this chapter investigates SLS design with output feedback affine control policies for uncertain linear systems subject to model mismatch and additive disturbances. This setting presents several major challenges: 1) the use of affine output feedback policies complicates the construction of an SLS parameterization while maintaining computational tractability; 2) model mismatch and stochastic disturbances lead to inaccurate predictions of system dynamics, increasing the difficulty of designing effective feedback policies; 3) model mismatch induces distributional shifts between the true closed-loop behavior and the predicted closed-loop response, further complicating the solution of the DR control formulation.

In order to address the above challenges, we propose a novel DR-SLS method with output feedback affine policies. The contributions of this chapter are summarized as follows:

- A novel SLS parameterization for a class of linear systems using output feedback affine control policy is proposed. Compared to existing SLS methods, the proposed scheme introduces fewer constraints and decision variables and accommodates a broader class of systems.

- The impact of model mismatch on the proposed SLS parameterization is analyzed. Besides, a new DR-SLS formulation is proposed to account for the influence of model mismatch and stochastic disturbances.

- The distribution shift between the prediction and true closed-loop responses is analyzed, highlighting its dependence on model uncertainties and the chosen SLS parameterization. An upper bound on this shift is derived in the sense of the Wasserstein metric.

- A computationally tractable relaxation of the proposed DR-SLS problem is developed by incorporating tools from robust SLS and DRO, resulting in a convex reformulation. Simulation results are provided to validate the effectiveness of the proposed approach.

The remainder of this chapter is organized as follows. Section 7.2 introduces the control problem under consideration. Section 7.3 proposes an SLS design with output feedback affine policies and extends it to the case of inexact models. In Section 7.4, a distributionally robust SLS (DR-SLS) formulation is introduced, along with a computationally tractable solution. Section 7.5 provides simulation results that validate the effectiveness of the proposed approach. Finally, Section 7.6 concludes this chapter.

Notation: Lowercase letters denote vectors or scalars; uppercase letters denote matrices. Boldface lowercase letters represent the stacked version of the corresponding vectors. Norm $||\cdot||$ refers to the 1-norm for vectors and the induced 1-norm for matrices. The notation A^{-1} denotes the inverse of matrix A. The operator $\text{col}(x_1, x_2, \cdots, x_n)$ represents the stacked vector $[x_1^{\mathrm{T}}, \cdots, x_n^{\mathrm{T}}]^{\mathrm{T}}$. Calligraphic capital letters represent sets or collections. $\text{blkdiag}(M_1, \cdots, M_n)$ denotes the block-diagonal matrix with $(M_1, \cdots, M_n)$ on its diagonal block entries. I_n represents a $n \times n$ identity matrix, and the subscript n will be omitted when clear from the context.

7.2. PROBLEM FORMULATION

Consider the following uncertain linear system in the so-called innovation form

$$x_{t+1} = Ax_t + Bu_t + Le_t, \tag{7.1a}$$
$$y_t = Cx_t + Du_t + e_t, \tag{7.1b}$$

where $x_t \in \mathbb{R}^n$ is the system state, $u_t \in \mathbb{R}^m$ is the control input, $e_t \in \mathbb{R}^q$ is the innovation process, and subscript t is time index. The matrices A, B, L, C and D are the respective system, input, Kalman, output and feedforward matrices, respectively. More details about the innovation form of linear systems can be found in [180–182].

For this linear system, the following finite-horizon optimal control problem will be investigated:

$$\min_{\pi_t} \sum_{t=0}^{T} h_t(y_t, u_t) \tag{7.2a}$$
$$\text{s.t. system dynamics (7.1)}, \tag{7.2b}$$
$$g_t(y_t, u_t) \le 0, \ t = 0, \cdots, T, \tag{7.2c}$$
$$u_t = \pi_t(y_0, \cdots, y_{t-1}), \tag{7.2d}$$

where T is the length of prediction horizon, $h_t(\cdot)$ and $g_t(\cdot)$ are stage cost and constraint functions, respectively, $\pi(\cdot)$ is the control policy, which is restricted to be strictly causal. In our design, h_t and g_t are assumed to be convex and Lipschitz continuous functions.

In this work, only the output signal y_t is assumed to be measurable for control design. To balance computational tractability and optimality, the following strictly causal output

feedback affine control policy is considered

$$\pi_t(y_0, \cdots, y_{t-1}) = \sum_{k=0}^{t-1} K_{t,k} y_k + p_t, \tag{7.3}$$

where $(K_{t,k}, p_t)$ are control policy parameters to be optimized. To ensure the causality of the control policy, the control input u_t is only parameterized via the output signals y_k $(k < t)$ prior to the time instant t.

It should be mentioned that directly solving the optimization problem (7.2) and considering the control policy (7.3) is computationally challenging since it leads to a nonconvex optimization problem w.r.t. the control policy parameters $K_{t,k}$ and p_t. Instead of directly optimizing the control parameters, the SLS parameterization provides an alternative solution by optimizing the closed-loop responses via solving a convex optimization problem. Then, the control parameters are computed based on the linear relationship between closed-loop responses and control policy parameters.

In the upcoming sections, the SLS design for system (7.1) under control law (7.3) will be investigated. It is important to note that incorporating the affine control policy introduces computational challenges for existing SLS methods. Specifically, the presence of the bias term in the affine policy renders a direct extension of the existing SLS methods developed for linear control policies nonconvex, see [176, 177], and thus computationally demanding. In addition to developing an SLS design for system (7.1) under the assumption of perfect knowledge of system models, we further investigate the impact of model uncertainties for the proposed SLS design. To address the influence of model uncertainties, we propose a novel DR-SLS framework to improve the robustness of the optimal control solution against model mismatch and stochastic disturbances.

7.3. SYSTEM LEVEL SYNTHESIS WITH OUTPUT FEEDBACK AFFINE CONTROL POLICY

In this section, the SLS-based method for solving (7.2) will be presented. Firstly, a new SLS parameterization for the linear system with output feedback affine control policy is derived. Then, the effects of model mismatch on the proposed SLS formulation are analyzed.

7.3.1. SYSTEM LEVEL SYNTHESIS WITH EXACT MODEL

System (7.1) can be rewritten as the following predictor form

$$x_{t+1} = (A - LC)x_t + (B - LD)u_t + Ly_t, \tag{7.4a}$$
$$e_t = y_t - Cx_t, \tag{7.4b}$$

and for any $\tau \in \mathbb{Z}_+$ (also called past horizon), we see that

$$x_t = (A - LC)^\tau x_{t-\tau} + \sum_{k=1}^{\tau} \left[\Psi_k^u u(t-k) + \Psi_k^y y(t-k)\right], \tag{7.5}$$

where $\Psi_k^u = (A - LC)^{k-1}(B - LD)$ and $\Psi_k^y = (A - LC)^{k-1}L$.

Supposing that $(A-LC)$ is Schur stable and denoting $\lambda_{\max}$ ($|\lambda_{\max}|<1$) as the eigenvalue of $(A-LC)$ with the largest absolute value, it was shown in [180] that

$$x_t = \Psi \begin{bmatrix} \mathbf{u}_t^- \\ \mathbf{y}_t^- \end{bmatrix} + \underbrace{O(|\lambda_{\max}|^\tau)}_{\to 0 \text{ as } \tau \to +\infty} \tag{7.6}$$

where $\Psi := [\Psi_\tau^u, \cdots, \Psi_1^u, \Psi_\tau^y, \cdots, \Psi_1^y]$, $\mathbf{u}_t^- = \mathrm{col}(u_{t-\tau}, \cdots, u_{t-1})$, and $\mathbf{y}_t^- = \mathrm{col}(y_{t-\tau}, \cdots, y_{t-1})$.

By selecting τ sufficiently large such that $O(|\lambda_{\max}|^\tau)$ can be neglected, and denoting the current time instant as $t=0$, we have

$$\mathbf{x} = \mathbf{T}_u \mathbf{u} + \mathbf{T}_e \mathbf{e} + \Gamma \Psi \begin{bmatrix} \mathbf{u}_0^- \\ \mathbf{y}_0^- \end{bmatrix}, \tag{7.7a}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} + \mathbf{e}, \tag{7.7b}$$

where $\mathbf{x} = \mathrm{col}(x_0, x_1, \cdots, x_T)$ is the stacked state vector over the prediction horizon, $\mathbf{y} := \mathrm{col}(y_0, y_1, \cdots, y_T)$ the stacked output vector, $\mathbf{u} = \mathrm{col}(u_0, u_1, \cdots, u_T)$ the stacked control input, $\mathbf{e} := \mathrm{col}(e_0, \cdots, e_T)$ the stacked innovation process, the matrices $\mathbf{C}$, Γ and the Toeplitz matrices $\mathbf{T}_u$ and $\mathbf{T}_e$ are defined as

$$\mathbf{C} = \mathrm{blkdiag}(\underbrace{C, \cdots, C}_{T+1 \text{ times}}),\ \mathbf{D} = \mathrm{blkdiag}(\underbrace{D, \cdots, D}_{T+1 \text{ times}}), \tag{7.8a}$$

$$\Gamma = \begin{bmatrix} I \\ A \\ \vdots \\ A^T \end{bmatrix},\ \mathbf{T}_u = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ B & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ A^{T-1}B & A^{T-2}B & \cdots & B & 0 \end{bmatrix}, \tag{7.8b}$$

$$\mathbf{T}_e = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ L & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ A^{T-1}L & A^{T-2}L & \cdots & L & 0 \end{bmatrix}. \tag{7.8c}$$

Combining (7.7a) and (7.7b) gives

$$\mathbf{y} = \mathbf{G}\mathbf{u} + \mathbf{y}_0 + \Theta \mathbf{e}, \tag{7.9}$$

where

$$\mathbf{G} = \mathbf{C}\mathbf{T}_u + \mathbf{D}, \tag{7.10a}$$

$$\mathbf{y}_0 = \mathbf{C}\Gamma\Psi \begin{bmatrix} \mathbf{u}_0^- \\ \mathbf{y}_0^- \end{bmatrix}, \tag{7.10b}$$

$$\Theta = \mathbf{C}\mathbf{T}_e + I. \tag{7.10c}$$

It is worth noting that since $\mathbf{T}_u$ is block strictly-lower-triangular, and $\mathbf{C}$ and $\mathbf{D}$ are block diagonal matrices, the matrix $\mathbf{G}$ is a block lower-triangular matrix.

Remark 7.1. By considering the system dynamics in the innovation form, our proposed scheme applies to a broader class of systems compared to the approach in [176, 179], which is limited to the case of dynamic disturbances modeled as input disturbances. Additionally, the use of (7.5) eliminates the need of knowledge of the initial state when predicting system outputs over the prediction horizon. Instead, only past input-output data over a finite horizon τ are required. A common heuristic for selecting τ is to choose it large enough such that the term $(A-LC)^{\tau}x_{t-\tau}$ becomes negligible, which is standard in many subspace algorithms [180, 183].

According to the control policy in (7.3), the control inputs over the prediction horizon satisfy

$$\mathbf{u} = \mathbf{K}\mathbf{y} + \mathbf{p}, \tag{7.11}$$

where $\mathbf{K}$ is defined as

$$\mathbf{K} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ K_{1,0} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ K_{T,0} & \cdots & K_{T,T-1} & 0 \end{bmatrix}, \tag{7.12}$$

which is block strictly-lower-triangular, and $\mathbf{p} := \mathrm{col}(p_0, \cdots, p_T)$ is the bias term vector.

Combing (7.9) and (7.11) gives the following closed-loop representation

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} (I-\mathbf{G}\mathbf{K})^{-1} & (I-\mathbf{G}\mathbf{K})^{-1}\mathbf{G}\mathbf{p} \\ \mathbf{K}(I-\mathbf{G}\mathbf{K})^{-1} & (I-\mathbf{K}\mathbf{G})^{-1}\mathbf{p} \end{bmatrix} \begin{bmatrix} \mathbf{y}_0 + \Theta\mathbf{e} \\ 1 \end{bmatrix}. \tag{7.13}$$

Since $\mathbf{G}$ is block lower-triangular and $\mathbf{K}$ is block strictly-lower-triangular, it can be guaranteed that $(I-\mathbf{G}\mathbf{K})$ and $(I-\mathbf{K}\mathbf{G})$ are always invertible.

Based on the above analysis, we have the following theorem.

Theorem 7.1. Consider the linear system (7.1) and the affine control policy (7.3), the following statements hold

1. the affine subspace defined by

$$\begin{bmatrix} I & -\mathbf{G} \end{bmatrix} \begin{bmatrix} \Phi_y & \phi_y \\ \Phi_u & \phi_u \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix} \tag{7.14}$$

with Φ_y and Φ_u satisfying the following block-lower-triangular and block-strictly-lower-triangular structures, respectively,

$$\Phi_y = \begin{bmatrix} \Phi_y^{0,0} & 0 & \cdots & 0 \\ \Phi_y^{1,0} & \Phi_y^{1,1} & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \Phi_y^{T,0} & \Phi_y^{T,1} & \cdots & \Phi_y^{T,T} \end{bmatrix} \tag{7.15a}$$

$$\Phi_u = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \Phi_u^{1,0} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \Phi_u^{T,0} & \cdots & \Phi_u^{T,T-1} & 0 \end{bmatrix} \tag{7.15b}$$

parameterizes all possible closed-loop output-input responses as

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\end{bmatrix} = \begin{bmatrix}\Phi_y & \phi_y\\ \Phi_u & \phi_u\end{bmatrix}\begin{bmatrix}\mathbf{y}_0+\Theta\mathbf{e}\\ 1\end{bmatrix}. \tag{7.16}$$

Namely, all closed-loop response parameterizations in the format of (7.16) with Φ_y and Φ_u having structures shown in (7.15) satisfy (7.14).

2. for any matrices $\{\Phi_y, \Phi_u\}$ and vectors $\{\phi_y, \phi_u\}$ satisfying (7.14) and (7.15), the corresponding control law $\mathbf{u} = \mathbf{K}\mathbf{y} + \mathbf{p}$ with

$$\mathbf{K} = \Phi_u \Phi_y^{-1}, \tag{7.17a}$$

$$\mathbf{p} = \phi_u - \mathbf{K}\phi_y \tag{7.17b}$$

achieves the desired closed-loop output-input responses in (7.16).

Proof: The proof is provided in the Appendix 7.A.1.

Remark 7.2. Theorem 7.1 leads to a concise SLS design with output feedback affine control policy. In the existing literature, almost all SLS designs only consider linear feedback policies, based either on state or output signals, since linear policies are less theoretically challenging to deal with in the SLS framework. Compared with linear policies, affine policies show increased expressiveness and can lead to improved control performance. The exception considering the SLS design with output feedback affine policy is the work proposed in [179]. It is worth noting that, in comparison with [179], the SLS design proposed in Theorem 1 provides a more concise design with fewer constraints and decision variables, and avoids computing an extra matrix inverse.

Remark 7.3. The existing SLS design approach with linear policies, e.g., [179], could be extended to affine policies by augmenting the output signal and feedback matrix as $\bar{y}_t := [y_t^{\mathrm{T}}, 1]^{\mathrm{T}}$ and $\bar{K}_{t,k} = [K_{t,k}, p_k^{\mathrm{T}}]$, respectively. However, this extension introduces potential issues. Specifically, considering the augmented output signal and feedback gains causes the matrix $\mathbf{K}$ in (7.12) to be block-lower-triangular rather than strictly block-lower-triangular. Consequently, the invertibility of $(I - \mathbf{G}\mathbf{K})$, which is required in (7.34), is no longer guaranteed. However, if the system (7.1) is strictly proper, i.e., $D = 0$, the matrix $\mathbf{G}$ defined in (7.10) becomes strictly-block-lower-triangular, and the existing approach with linear control policy is applicable to the extended system.

For notational brevity, we denote the closed-loop output-input vector as $\boldsymbol{\eta} := \mathrm{col}(\mathbf{y}, \mathbf{u})$ in the remaining parts of this work. Based on the SLS formulation in Theorem 7.1, the stochastic version of the finite-horizon predictive control problem (7.2) can be formulated as

$$\min_{\Phi_y, \Phi_u, \phi_y, \phi_u} \ \mathbb{E}^{\boldsymbol{\eta}\sim\mathbb{P}^{\mathscr{M}}}\left[h(\boldsymbol{\eta})\right] \tag{7.18a}$$

$$\text{s.t. constraints (7.14) and (7.15)}, \tag{7.18b}$$

$$\mathbb{E}^{\boldsymbol{\eta}\sim\mathbb{P}^{\mathscr{M}}}\left[g(\boldsymbol{\eta})\right] \le 0, \tag{7.18c}$$

$$\boldsymbol{\eta} = \begin{bmatrix}\Phi_y & \phi_y\\ \Phi_u & \phi_u\end{bmatrix}\begin{bmatrix}\mathbf{y}_0+\Theta\mathbf{e}\\ 1\end{bmatrix}, \tag{7.18d}$$

where $\mathbb{P}^{\cdot\mathscr{M}}$ denotes the probability distribution of the closed-loop output-input response $\boldsymbol{\eta}$ with the system model $\mathscr{M}$, which influences the parameters $(\mathbf{G}, \Theta, \mathbf{y}_0, \mathbf{e})$, the functions $h(\cdot)$ and $g(\cdot)$ are constructed based on $h_t(\cdot)$ and $g_t(\cdot)$ defined in (7.2), respectively. The stochastic uncertainty considered in the above formulation is the innovation process vector $\mathbf{e}$.

Suppose that the system model $\mathscr{M}$ is perfectly available and the stochastic uncertainty $\mathbf{e}$ can be replaced by its nominal prediction. By exploiting the SLS-based formulation proposed in Theorem 7.1, the deterministic version of the above optimization problem leads to a convex optimization problem, and the corresponding control policy can be computed via (7.17).

7.3.2. SLS PARAMETERIZATION WITH MODEL MISMATCH

The SLS formulation proposed in Section 7.3.1 is based on the assumption that the system model, i.e., the set of parameters $(\mathbf{G}, \Theta, \mathbf{y}_0, \mathbf{e})$, is perfectly known. However, obtaining the exact model information is nontrivial and challenging. It is more practical to consider that only an approximate model is available for SLS design.

It is worth noting that among the above-listed parameters, $\mathbf{G}$ plays the most important role in influencing the SLS design since it will directly influence the SLS design via (7.14). In this section, the SLS formulation proposed in Theorem 7.1 of Section 7.3.1 is extended to the case where only the approximated parameter $\hat{\mathbf{G}}$ is available. For ease of notation, the approximation error of $\mathbf{G}$ is defined as $\Delta := \mathbf{G} - \hat{\mathbf{G}}$.

Theorem 7.2. Assuming a nominal approximation parameter $\hat{\mathbf{G}}$ is adopted for SLS-based design in (7.14) and (7.17), the true closed-loop output-input response $\boldsymbol{\eta} := \mathrm{col}(\boldsymbol{y}, \boldsymbol{u})$ with the true parameters $(\mathbf{G}, \mathbf{y}_0, \Theta, \mathbf{e})$ is

$$\boldsymbol{\eta} = \Phi(\mathbf{y}_0 + \Theta\mathbf{e}) + \phi \tag{7.19}$$

with

$$\Phi := \begin{bmatrix} \Phi_y \\ \Phi_u \end{bmatrix} = \begin{bmatrix} \hat{\Phi}_y \\ \hat{\Phi}_u \end{bmatrix} \underbrace{(I - \Delta\hat{\Phi}_u)^{-1}}_{R_\Phi}, \tag{7.20a}$$

$$\phi := \begin{bmatrix} \phi_y \\ \phi_u \end{bmatrix} = \underbrace{\begin{bmatrix} I & \hat{\Phi}_y(I - \Delta\hat{\Phi}_u)^{-1}\Delta \\ 0 & I + \hat{\Phi}_u(I - \Delta\hat{\Phi}_u)^{-1}\Delta \end{bmatrix}}_{R_\phi} \begin{bmatrix} \hat{\phi}_y \\ \hat{\phi}_u \end{bmatrix}, \tag{7.20b}$$

where $\{\Phi_y, \Phi_u, \phi_y, \phi_u\}$ is the true closed-loop response parametrization, $\{\hat{\Phi}_y, \hat{\Phi}_u, \hat{\phi}_y, \hat{\phi}_u\}$ is the predictive closed-loop response parameterization derived via (7.14) by replacing $\mathbf{G}$ with $\hat{\mathbf{G}}$, and (R_Φ, R_ϕ) are the uncertainties of closed-loop parameterization introduced by the model mismatch Δ.

Proof: The proof is presented in the Appendix 7.A.2

7.4. DISTRIBUTIONALLY ROBUST SLS DESIGN AND TRACTABLE REFORMULATION

Based on the results shown in Section 7.3, this section proposes a novel DR-SLS framework to systematically enhance the robustness of the optimal control solution against

model mismatch and additive stochastic disturbances.

7.4.1. DISTRIBUTIONALLY ROBUST SLS FORMULATION

Solving the SLS-based problem (7.18) is generally intractable when assuming the existence of model mismatch for parameters $(\mathbf{G}, \Theta, \mathbf{y}_0)$, and uncertain probability distribution $\mathbb{P}_{\mathbf{e}}$. One possible solution for solving (7.18) is adopting the sample average approximation (SAA) and certainty equivalence (CE) principles, in which the unknown system parameters $(\mathbf{G}, \Theta, \mathbf{y}_0)$ are replaced by their nominal values $(\hat{\mathbf{G}}, \hat{\Theta}, \hat{\mathbf{y}}_0)$ and the true innovation process distribution $\mathbb{P}_{\mathbf{e}}$ is approximated by its empirical counterpart $\overline{\mathbb{P}}_{\mathbf{e}}$. In this way, the true closed-loop response distribution $\mathbb{P}^{\mathscr{M}}$ is replaced by its empirical approximated distribution $\overline{\mathbb{P}}^{\hat{\mathscr{M}}}$.

Although this solution is straightforward and easy to implement, its efficacy is built upon the expectation that the true closed-loop distribution $\mathbb{P}^{\mathscr{M}}$ can be properly approximated by $\overline{\mathbb{P}}^{\hat{\mathscr{M}}}$. However, the complex dependency of the closed-loop system behaviour on the true model $\mathscr{M}$, as opposed to the nominal predictive model $\hat{\mathscr{M}}$, undermines the justification for applying CE and SAA. Moreover, when only a limited number of uncertainty samples are available, the SAA approach may lead to a poor out-of-sample performance.

In the following, a distributionally robust SLS (DR-SLS) design will be presented, considering the uncertainties introduced by the model mismatch between $(\mathbf{G}, \Theta, \mathbf{y}_0)$ and $(\hat{\mathbf{G}}, \hat{\Theta}, \hat{\mathbf{y}}_0)$ as well as the estimation error of the empirical innovation process distribution $\overline{\mathbb{P}}_{\mathbf{e}}$. The approximation errors of $(\mathbf{G}, \Theta, \mathbf{y}_0)$ are denoted as $\Delta := \mathbf{G} - \hat{\mathbf{G}}$, $\tilde{\Theta} := \Theta - \hat{\Theta}$, and $\tilde{\mathbf{y}}_0 := \mathbf{y}_0 - \hat{\mathbf{y}}_0$, respectively. Before presenting the main results of this section, the following definitions are introduced.

Definition 7.1. (Wasserstein Distance [156]) The Wasserstein distance $d_W : \mathscr{H}(\Xi) \times \mathscr{H}(\Xi) \to \mathbb{R}$ is defined as

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \inf_{\Pi} \left\{ \int_{\Xi^2} ||\xi_1 - \xi_2|| \Pi(d\xi_1, d\xi_2) \right\},$$

where Π is a joint distribution of the random variables ξ_1 and ξ_2 with marginal distributions $\mathbb{Q}_1$ and $\mathbb{Q}_2$, respectively, and $\mathscr{H}(\Xi)$ is the space of all distributions $\mathbb{Q}$ supported on Ξ such that $\int_{\Xi} ||\xi|| \mathbb{Q}(d\xi) < +\infty$.

Definition 7.2. (Wasserstein Ambiguity Set [156]) The Wasserstein ambiguity set $\mathscr{B}_\varepsilon(\mathbb{Q}_0)$ is defined as the ball of radius ε centered at the empirical distribution $\mathbb{Q}_0$

$$\mathscr{B}_\varepsilon(\mathbb{Q}_0) := \left\{ \mathbb{Q} \in \mathscr{H}(\Xi) \mid d_W(\mathbb{Q}, \mathbb{Q}_0) \leq \varepsilon \right\}.$$

For notational brevity, the true system model and closed-loop parameterizations are compactly represented as $\mathscr{M} := \{\mathbf{G}, \Theta, \Phi_y, \Phi_u, \phi_y, \phi_u, \mathbf{y}_0\}$. Similarly, the nominal predictive closed-loop model and the corresponding parameterizations are denoted as $\hat{\mathscr{M}} := \{\hat{\mathbf{G}}, \hat{\Theta}, \hat{\Phi}_y, \hat{\Phi}_u, \hat{\phi}_y, \hat{\phi}_u, \hat{\mathbf{y}}_0\}$.

Assumption 7.1. For the innovation process $\mathbf{e}$, N samples of historical realizations $\{\mathbf{e}_i\}_{i=1}^N$ are available.

The DR-SLS problem is formulated as

$$\min_{\substack{\hat{\Phi}_y,\hat{\Phi}_u\\ \hat{\phi}_y,\hat{\phi}_u}} \sup_{\mathbb{Q}\in\mathscr{B}_\varepsilon(\overline{\mathbb{P}}^{\hat{\mathscr{M}}})} \mathbb{E}^{\boldsymbol{\eta}\sim\mathbb{Q}}\left[h(\boldsymbol{\eta})\right] \tag{7.21a}$$

$$\text{s.t.} \quad \sup_{\mathbb{Q}\in\mathscr{B}_\varepsilon(\overline{\mathbb{P}}^{\hat{\mathscr{M}}})} \mathbb{E}^{\boldsymbol{\eta}\sim\mathbb{Q}}[g(\boldsymbol{\eta})] \le 0, \tag{7.21b}$$

$$\begin{bmatrix} I & -\hat{\mathbf{G}} \end{bmatrix} \begin{bmatrix} \hat{\Phi}_y & \hat{\phi}_y \\ \hat{\Phi}_u & \hat{\phi}_u \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix}, \tag{7.21c}$$

$$\text{constraints (7.19) and (7.20)}, \tag{7.21d}$$

where $\mathbb{Q}$ is the probability distribution of the true closed-loop output-input responses $\boldsymbol{\eta}$, $\overline{\mathbb{P}}^{\hat{\mathscr{M}}}$ is the empirical predictive probability distribution of closed-loop dynamics based on $(\hat{\mathbf{G}},\hat{\Theta},\hat{\mathbf{y}}_0,\overline{\mathbb{P}}_{\mathbf{e}})$, $\mathscr{B}_\varepsilon(\overline{\mathbb{P}}^{\hat{\mathscr{M}}})$ is the ambiguity set of $\mathbb{Q}$.

Given N samples of the innovation process $\{\mathbf{e}_i\}_{i=1}^N$, the empirical distribution $\overline{\mathbb{P}}_{\mathbf{e}}$ can be constructed as $\overline{\mathbb{P}}_{\mathbf{e}} := \frac{1}{N}\sum_{i=1}^N \delta(\mathbf{e}_i)$, where $\delta(\cdot)$ is the Dirac distribution. Accordingly, the empirical predictive distribution of the output-input response $\boldsymbol{\eta}$ with the nominal model $\hat{\mathscr{M}}$ is

$$\overline{\mathbb{P}}^{\hat{\mathscr{M}}} = \frac{1}{N}\sum_{i=1}^N \delta(\hat{\boldsymbol{\eta}}_i), \tag{7.22a}$$

$$\hat{\boldsymbol{\eta}}_i = \hat{\Phi}(\hat{\mathbf{y}}_0 + \hat{\Theta}\mathbf{e}_i) + \hat{\phi}, \tag{7.22b}$$

where $(\hat{\Phi},\hat{\phi})$ are the predictive closed-loop response parameterization derived with $\hat{\mathbf{G}}$ via (7.21c), and $\hat{\boldsymbol{\eta}}_i$ is the empirical prediction of output-input response using $(\hat{\Phi},\hat{\phi},\hat{\Theta},\hat{\mathbf{y}}_0,\mathbf{e}_i)$.

7.4.2. CHARACTERIZATION OF DISTRIBUTION SHIFT

In order to properly formulate and solve the DR-SLS problem in (7.21), the distribution shift between $\mathbb{P}^{\mathscr{M}}$ and $\overline{\mathbb{P}}^{\hat{\mathscr{M}}}$, which is measured by the Wasserstein distance $d_W(\overline{\mathbb{P}}^{\hat{\mathscr{M}}},\mathbb{P}^{\mathscr{M}})$ and the associated ambiguity set $\mathscr{B}_\varepsilon(\overline{\mathbb{P}}^{\hat{\mathscr{M}}})$, should be properly considered. An excessively large radius ε of the ambiguity set will lead to a conservative solution, and an excessively small one would result in an aggressive and fragile solution. Importantly, the Wasserstein distance $d_W(\overline{\mathbb{P}}^{\hat{\mathscr{M}}},\mathbb{P}^{\mathscr{M}})$ is determined by the closed-loop model mismatch between $\mathscr{M}$ and $\hat{\mathscr{M}}$, and is therefore a function of the predictive closed-loop parameterization $\{\hat{\Phi}_y,\hat{\Phi}_u,\hat{\phi}_y,\hat{\phi}_u\}$. Consequently, simply setting the radius of the ambiguity set with a prescribed constant does not reflect its intrinsic dependence on the SLS parameters and can lead to unsatisfactory performance.

Based on the results in Theorem 7.2, the true closed-loop output-input responses $\boldsymbol{\eta}\sim\mathbb{P}^{\mathscr{M}}$ are

$$\boldsymbol{\eta} = \hat{\Phi}R_\Phi(\mathbf{y}_0 + \Theta\mathbf{e}) + R_\phi\hat{\phi}. \tag{7.23}$$

Compared with the empirical prediction in (7.22), there are two factors contributing to the distribution shift: 1) the uncertainties caused by the model mismatch between

$(\mathbf{G}, \Theta, \mathbf{y}_0)$ and $(\hat{\mathbf{G}}, \hat{\Theta}, \hat{\mathbf{y}}_0)$, and 2) the approximation error between the true distribution $\mathbb{P}_{\mathbf{e}}$ and its empirical approximation $\overline{\mathbb{P}}_{\mathbf{e}}$.

In the following, we analyze the Wasserstein metric $d_W(\overline{\mathbb{P}}_{\pi}^{\hat{\mathcal{M}}}, \mathbb{P}_{\pi}^{\mathcal{M}})$, and provide an upper bound for the distribution shift. The triangle inequality of Wasserstein distance gives

$$d_W(\overline{\mathbb{P}}^{\hat{\mathcal{M}}}, \mathbb{P}^{\mathcal{M}}) \leq d_W(\overline{\mathbb{P}}^{\hat{\mathcal{M}}}, \overline{\mathbb{P}}^{\mathcal{M}}) + d_W(\overline{\mathbb{P}}^{\mathcal{M}}, \mathbb{P}^{\mathcal{M}}), \tag{7.24}$$

where $\overline{\mathbb{P}}^{\mathcal{M}}$ is the empirical version of the true output-input distribution $\mathbb{P}^{\mathcal{M}}$. Namely,

$$\overline{\mathbb{P}}^{\mathcal{M}} = \frac{1}{N}\sum_{i=1}^{N} \delta(\boldsymbol{\eta}_i), \tag{7.25a}$$

$$\boldsymbol{\eta}_i = \hat{\Phi} R_{\Phi}(\mathbf{y}_0 + \Theta \mathbf{e}_i) + R_{\phi}\hat{\phi}. \tag{7.25b}$$

Applying (7.24) decomposes the distribution shift into two parts: a) the shift caused by model mismatch $d_W(\overline{\mathbb{P}}^{\hat{\mathcal{M}}}, \overline{\mathbb{P}}^{\mathcal{M}})$, and b) the shift caused by the emprical approximation error $d_W(\overline{\mathbb{P}}^{\mathcal{M}}, \mathbb{P}^{\mathcal{M}})$. The following lemmas provide upper bounds for the above distribution shifts.

Lemma 7.1. (distribution shift by model mismatch) The distribution shift caused by the model mismatch $d_W(\overline{\mathbb{P}}^{\hat{\mathcal{M}}}, \overline{\mathbb{P}}^{\mathcal{M}})$ can be upper-bounded as

$$d_W(\overline{\mathbb{P}}^{\hat{\mathcal{M}}}, \overline{\mathbb{P}}^{\mathcal{M}}) \leq \frac{1}{N}\sum_{i=1}^{N}\Big(||\hat{\Phi}(R_{\Phi} - I)(\hat{\mathbf{y}}_0 + \hat{\Theta}\mathbf{e}_i)|| + ||\hat{\Phi} R_{\Phi}(\tilde{\mathbf{y}}_0 + \tilde{\Theta}\mathbf{e}_i)||\Big) + ||(R_{\phi} - I)\hat{\phi}||. \tag{7.26}$$

Proof: The Wassertein distance $d_W(\overline{\mathbb{P}}^{\hat{\mathcal{M}}}, \overline{\mathbb{P}}^{\mathcal{M}})$ is defined as

$$d_W(\overline{\mathbb{P}}^{\hat{\mathcal{M}}}, \overline{\mathbb{P}}^{\mathcal{M}}) := \inf_{\Pi} \int_{\Xi^2} ||\hat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i|| \Pi(d\hat{\boldsymbol{\eta}}_i, d\boldsymbol{\eta}_i), \tag{7.27}$$

where Π is the joint probability distribution of $(\hat{\boldsymbol{\eta}}_i, \boldsymbol{\eta}_i)$. By selecting the suboptimal adjoint probability distribution $\Pi^*(\hat{\boldsymbol{\eta}}_i, \boldsymbol{\eta}_i) = \frac{1}{N}\delta(\hat{\boldsymbol{\eta}}_i, \boldsymbol{\eta}_i)$, and considering the expression of $\hat{\boldsymbol{\eta}}_i$ and $\boldsymbol{\eta}_i$ in (7.22) and (7.25) together with (7.20), and applying the norm triangular inequality completes the proof. □

Lemma 7.2. (distribution shift by empirical approximation error of $\mathbb{P}_{\mathbf{e}}$) The Wasserstein distance between the empirical closed-loop output-input distribution $\overline{\mathbb{P}}^{\mathcal{M}}$ and the real closed-loop output-input distribution $\mathbb{P}^{\mathcal{M}}$ is upper-bounded as

$$d_W(\overline{\mathbb{P}}^{\mathcal{M}}, \mathbb{P}^{\mathcal{M}}) \leq ||\hat{\Phi} R_{\Phi}(\hat{\Theta} + \tilde{\Theta})|| d_W(\overline{\mathbb{P}}_{\mathbf{e}}, \mathbb{P}_{\mathbf{e}}). \tag{7.28}$$

Proof: Based on the definition of Wasserstein distance, we have

$$\begin{aligned} d_W(\overline{\mathbb{P}}^{\mathcal{M}}, \mathbb{P}^{\mathcal{M}}) &:= \inf_{\Pi} \int_{\Xi^2} ||\boldsymbol{\eta}' - \boldsymbol{\eta}|| \Pi(d\boldsymbol{\eta}', d\boldsymbol{\eta}) \\ &= \inf_{\Pi} \int_{\mathcal{V}^2} ||\hat{\Phi} R_{\Phi}(\hat{\Theta} + \tilde{\Theta})(\mathbf{e}' - \mathbf{e})|| \Pi(d\mathbf{e}', d\mathbf{e}) \\ &\leq ||\hat{\Phi} R_{\Phi}(\hat{\Theta} + \tilde{\Theta})|| \underbrace{\inf_{\Pi} \int_{\mathcal{V}^2} ||\mathbf{e}' - \mathbf{e}|| \Pi(d\mathbf{e}', d\mathbf{e})}_{d_W(\overline{\mathbb{P}}_{\mathbf{e}}, \mathbb{P}_{\mathbf{e}})} \end{aligned}$$

where the above relationships are based on the fact that the innovation process $\mathbf{e}$ is coupled with $\boldsymbol{\eta}$ only through Φ in (7.19) and (7.20), and the sub-multiplicativity property of 1-norm. This completes the proof. □

Proposition 7.1. The Wasserstein distance between the empirical predictive output-input distribution $\overline{\mathbb{P}}^{\hat{\mathcal{M}}}$ and the true output-input distribution $\mathbb{P}^{\mathcal{M}}$ is upper-bounded by:

$$d_W(\overline{\mathbb{P}}^{\hat{\mathcal{M}}}, \mathbb{P}^{\mathcal{M}}) \leq \frac{1}{N}\sum_{i=1}^{N}\left(||\hat{\Phi}(R_\Phi - I)(\hat{\mathbf{y}}_0 + \hat{\Theta}\mathbf{e}_i)|| + ||\hat{\Phi}R_\Phi(\tilde{\mathbf{y}}_0 + \tilde{\Theta}\mathbf{e}_i)||\right) + ||(R_\phi - I)\hat{\phi}|| \\ + ||\hat{\Phi}R_\Phi(\hat{\Theta} + \tilde{\Theta})||d_W(\overline{\mathbb{P}}_{\mathbf{e}}, \mathbb{P}_{\mathbf{e}}). \tag{7.29}$$

Proof: Applying the triangle inequality (7.24) and the results in Lemmas 7.1 and 7.2 completes the proof. □

7.4.3. TRACTABLE REFORMULATION

While the DR-SLS formulation in (7.21) could enhance the robustness of the optimal solution of the SLS design, it is challenging to solve the corresponding optimization problem when considering the Wasserstein distance (7.29). In this section, a tractable reformulation is proposed. To this end, the following assumptions are made.

Assumption 7.2. There exist constants $\alpha > 1$ and $\beta > 0$ such that

$$\mathbb{E}^{\mathbf{e}\sim\mathbb{P}_{\mathbf{e}}}\left[e^{\beta||\mathbf{e}||^\alpha}\right] < +\infty \tag{7.30}$$

Assumption 7.3. The Wasserstein distance between the empirical distribution $\overline{\mathbb{P}}_{\mathbf{e}}$ and the true distribution $\mathbb{P}_{\mathbf{e}}$ is bounded as $d_W(\overline{\mathbb{P}}_{\mathbf{e}}, \mathbb{P}_{\mathbf{e}}) \leq \kappa$.

Assumption 7.4. The model mismatch is bounded as $||\Delta|| \leq \gamma_1$, $||\tilde{\Theta}|| \leq \gamma_2$, and $||\tilde{\mathbf{y}}_0|| \leq \gamma_3$.

Assumption 7.5. The cost function $h(\boldsymbol{\eta})$ and the system constraint $g(\boldsymbol{\eta})$ in (7.21) are convex and Lipschitz continuous, with Lipschitz constants l_h and l_g, respectively.

Remark 7.4. Assumption 7.2 ensures that the distribution $\mathbb{P}_{\mathbf{e}}$ has light tails, which is standard in DRO problems to guarantee well-posedness. Assumption 7.3 imposes a bound on the Wasserstein distance. The radius κ can be selected using, for example, cross-validation. Assumption 7.4 provides upper bounds on model mismatch, which is standard in robust control design (e.g., [176, 179, 184]). If uncertainty bounds are originally given for the system matrices (A, B, C, D), the corresponding bounds for $(\Delta, \tilde{\Theta}, \tilde{\mathbf{y}}_0)$ can be obtained via either numerical simulation or theoretical analysis as shown in [185]. Assumption 7.5 ensures that the cost and constraint functions are tractable by assuming convexity and Lipschitz continuity. This covers a broad class of functions, including linear functions, pointwise maximum of affine functions, and norms such as the l_1-norm and l_2-norm.

Theorem 7.3. Supposing Assumptions 7.1-7.5 hold, and letting $\gamma_1||\hat{\Phi}_u|| \leq \rho$ with $\rho \in [0, 1)$ and $||\hat{\Phi}_y|| \leq \sigma$ with $\sigma > 0$, a convex relaxation of the DR-SLS problem in (7.21) can be

formulated as

$$\min_{\bar{\varepsilon},\hat{\Phi},\hat{\phi},s_i,q_i} \quad l_h\bar{\varepsilon}+\frac{1}{N}\sum_{i=1}^{N}s_i \tag{7.31a}$$

$$\text{s.t.}\begin{bmatrix} I & -\hat{\mathbf{G}}\end{bmatrix}\begin{bmatrix}\hat{\Phi}_y & \hat{\phi}_y\\ \hat{\Phi}_u & \hat{\phi}_u\end{bmatrix}=\begin{bmatrix} I & 0\end{bmatrix}, \tag{7.31b}$$

$$\hat{\boldsymbol{\eta}}_i=\begin{bmatrix}\hat{\Phi}_y & \hat{\phi}_y\\ \hat{\Phi}_u & \hat{\phi}_u\end{bmatrix}\begin{bmatrix}\hat{\mathbf{y}}_0+\hat{\Theta}\mathbf{e}_i\\ 1\end{bmatrix}, \tag{7.31c}$$

$$\hat{\Phi}_y \text{ and } \hat{\Phi}_u \text{ satisfy the structure in (7.15)}, \tag{7.31d}$$

$$h(\hat{\boldsymbol{\eta}}_i)\le s_i,\quad \forall i\in\{0,\cdots,N\}, \tag{7.31e}$$

$$l_g\bar{\varepsilon}+\frac{1}{N}\sum_{i=1}^{N}q_i\le 0, \tag{7.31f}$$

$$g(\hat{\boldsymbol{\eta}}_i)\le q_i,\quad \forall i\in\{0,\cdots,N\}, \tag{7.31g}$$

$$\begin{aligned}\bar{\varepsilon}\ge\frac{||\hat{\Phi}||}{N(1-\rho)}\Big[\sum_{i=1}^{N}\big(\rho||\hat{\mathbf{y}}_0+\hat{\Theta}\mathbf{e}_i||+\gamma_2||\mathbf{e}_i||\big)+N\gamma_3\Big]\\ +\left(\frac{\rho}{\gamma_1}+\sigma\right)\cdot\frac{\gamma_1}{1-\rho}||\hat{\phi}_u||+\frac{\kappa}{1-\rho}||\hat{\Phi}||\cdot(||\hat{\Theta}||+\gamma_2),\end{aligned} \tag{7.31h}$$

$$||\hat{\Phi}_y||\le\sigma,\ \gamma_1||\hat{\Phi}_u||\le\rho,\ \rho\in[0,1), \tag{7.31i}$$

where constraints (7.31b)-(7.31d) are the SLS-based parameterization of closed-loop output-input response, constraint (7.31e) is for reformulating the objective function in (7.21a), and constraints (7.31f) and (7.31g) are for reformulating the constraint in (7.21b), constraint (7.31h) is for upper-bounding the Wasserstein distance $d_W(\overline{\mathbb{P}}^{\hat{\mathcal{M}}},\mathbb{P}^{\mathcal{M}})$.

Proof: The proof is presented in the Appendix 7.A.3

Remark 7.5. By fixing the parameters ρ and σ, which influence the bounds of $||\hat{\Phi}_u||$ and $||\hat{\Phi}_y||$, respectively, the optimization problem (7.31) becomes convex and can be efficiently solved. The optimal values of ρ and σ can be determined via a grid search by iteratively solving a sequence of convex optimization problems. Alternatively, if ρ and σ are treated as decision variables rather than fixed parameters, the problem (7.31) becomes nonconvex due to the presence of bilinear terms. Such nonconvexities can still be dealt with using off-the-shelf solvers, such as `Gurobi` and `Ipopt`. Furthermore, the bound derived in (7.31h) possesses an asymptotic consistency property. In the absence of uncertainties, i.e., when $\gamma_1=\gamma_2=\gamma_3=\kappa=0$, the last two terms in (7.31h) vanish. Moreover, the condition $\gamma_1=0$ implies that ρ can be chosen arbitrarily small, thereby allowing the Wasserstein distance bound $\bar{\varepsilon}$ to be made arbitrarily close to zero.

Remark 7.6. In (7.21), the system constraint is formulated to guarantee that its expected value is within a prescribed bound. To enhance constraint satisfaction and provide probabilistic guarantees, our approach can be readily extended–following the method in [184]–to incorporate a conditional value at risk (CVaR) formulation. This extension preserves the structure of the proposed scheme and does not increase computational complexity. Natually, using a CVaR-based formulation to achieve higher probability guarantees results in a more conservative solution.

Remark 7.7. It is worth highlighting that while the parameters $(\mathbf{G}, \Theta, \mathbf{y}_0)$ considered in Theorem 7.1 are related to system matrices (A, B, C, D), the proposed design does not rely on the explicit knowledge of these system matrices. Instead, $(\mathbf{G}, \Theta, \mathbf{y}_0)$ can be possibly estimated from system input-output data. As in many existing subspace algorithms (e.g., [182]), the input-output relationships in (7.7) can be leveraged to identify approximate values of $(\mathbf{G}, \Theta, \mathbf{y}_0)$. Unlike the case in traditional subspace-based system identification methods, our approach avoids the explicit reconstruction of the system matrices (A, B, C, D), a task that is typically complex and challenging.

7.5. SIMULATION RESULTS

In this section, numerical simulation results are presented to demonstrate the effectiveness of the proposed approach. The numerical study is based on the linear system considered in [180], with a more challenging modification by setting $B_2 = 0$ such that small perturbations of B_2 will change the sign of the control input, resulting in different behaviours. The true system dynamics used in simulation are

$$x_{t+1} = \begin{bmatrix} 0.7326 & -0.0861 \\ 0.1722 & 0.9909 \end{bmatrix} x_t + \begin{bmatrix} 0.0609 \\ 0 \end{bmatrix} u_t + w_t, \tag{7.32a}$$

$$y_t = \begin{bmatrix} 0 & 1.4142 \end{bmatrix} x_t + v_t, \tag{7.32b}$$

where $w_t \in \mathbb{R}^2$ and $v_t \in \mathbb{R}$ are assumed to follow uniform distributions with each element belonging to $\mathscr{U}(-0.01, 0.01)$. For constructing the innovation process, the feedback matrix L in (7.1) is set as $[0.1, 0.1]^{\mathrm{T}}$, and $A - LC$ is Schur stable. For the SLS design, its prediction horizon is set as $T = 15$, and $\tau = 25$. The cost function is set as

$$h(\boldsymbol{\eta}) := \left\| \begin{bmatrix} \mathbf{Q} & \\ & \mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \end{bmatrix} \right\|, \tag{7.33}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are diagonal matrices with diagonal elements as 1 and 0.1, respectively. System constraints at each time step $k = 0, \cdots, T$ are defined as

$$g_k(\boldsymbol{\eta}) := \max\{-y_k - 0.01, u_k - 1, -u_k - 1\} <= 0.$$

This constraint limits the output and input signals satisfying $y_t \geq -0.01$ and $-1 \leq u_t \leq 1$, respectively.

The upper bounds of uncertainties are $\gamma_1 = \gamma_2 = \gamma_3 = 0.01$. In the simulation, 50 samples of system matrices $(\hat{A}, \hat{B}, \hat{C})$ are randomly generated such that the corresponding parameters $(\hat{\mathbf{G}}, \hat{\Theta}, \hat{\mathbf{y}}_0)$ reside in the prescribed scope. Random input signals within $[-1, 1]$ are used to generate 100 samples of the innovation process signal $\mathbf{e}$. The upper bound of the Wasserstein distance $d_W(\overline{\mathbb{P}}_{\mathbf{e}}, \mathbb{P}_{\mathbf{e}})$ is set as $\kappa = 0.005$.

In our simulation, we mainly investigate two SLS approaches:

- N-SLS: nominal system level synthesis design adopting the CE and SAA solutions, where the nominal parameters $(\hat{\mathbf{G}}, \hat{\Theta}, \hat{\mathbf{y}}_0)$ and the averaged historical samples of $\mathbf{e}$ are used in Theorem 7.1 of Section 7.3.
- DR-SLS: distributionally robust system level synthesis design proposed in this work by solving (7.31).

For each sample of the parameters $(\hat{\mathbf{G}}, \hat{\Theta}, \hat{\mathbf{y}}_0)$, after solving the SLS design problems with the above two approaches, the control policy parameters $(\mathbf{K}, \mathbf{p})$ are used for performing the closed-loop simulation with the true system dynamics in (7.32).

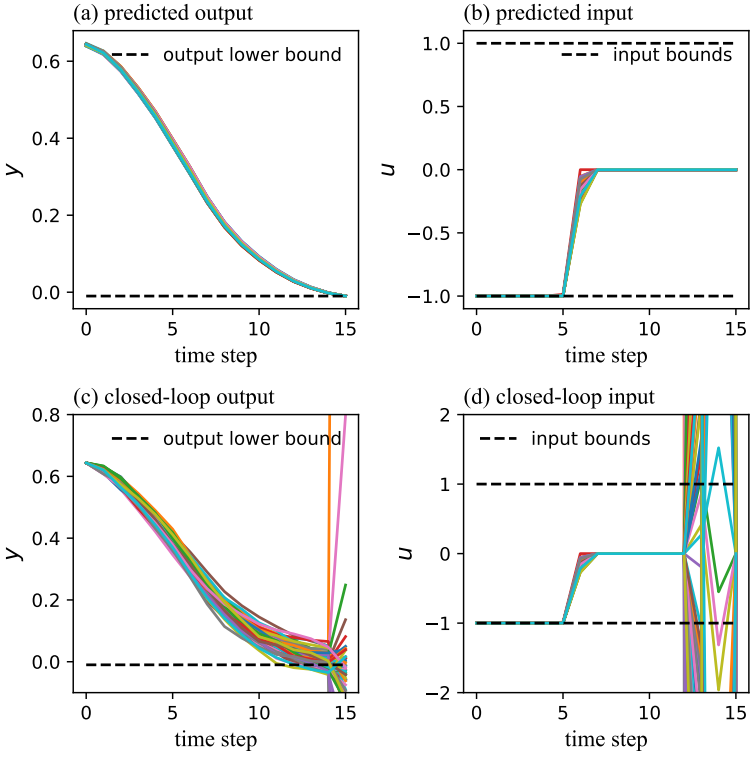


Figure 7.1: Output and input trajectories with N-SLS approach: (a) open-loop output prediction, (b) open-loop input prediction, (c) closed-loop output, (d) closed-loop input.

Simulation results are shown in Figs. 7.1–7.3. Fig. 7.1 depicts the output and input trajectories adopting the N-SLS approach for both open-loop prediction and closed-loop simulation. It can be observed that the N-SLS approach is aggressive in optimizing the cost function such that the output trajectory is close to the predefined lower bound. However, while the predicted output and input profiles with the approximated system parameters satisfy all system constraints and converge to the expected setpoints, shown in Fig. 7.1(a-b), the closed-loop response with the true model becomes unstable and the output-input trajectories diverge and violate system constraints, see Fig. 7.1(c-d).

Fig. 7.2 presents the simulation results adopting the proposed DR-SLS approach. It can be observed from Fig. 7.2(a-b) that the output and input predictions all satisfy system constraints and leave margins to the respective permissible boundaries. In comparison with the case of the N-SLS approach shown in Fig. 7.1, the input-output predictions do not exactly converge to the origin, and small offsets are introduced to enhance robustness against possible model uncertainties. Fig. 7.2(c-d) show the true closed-loop

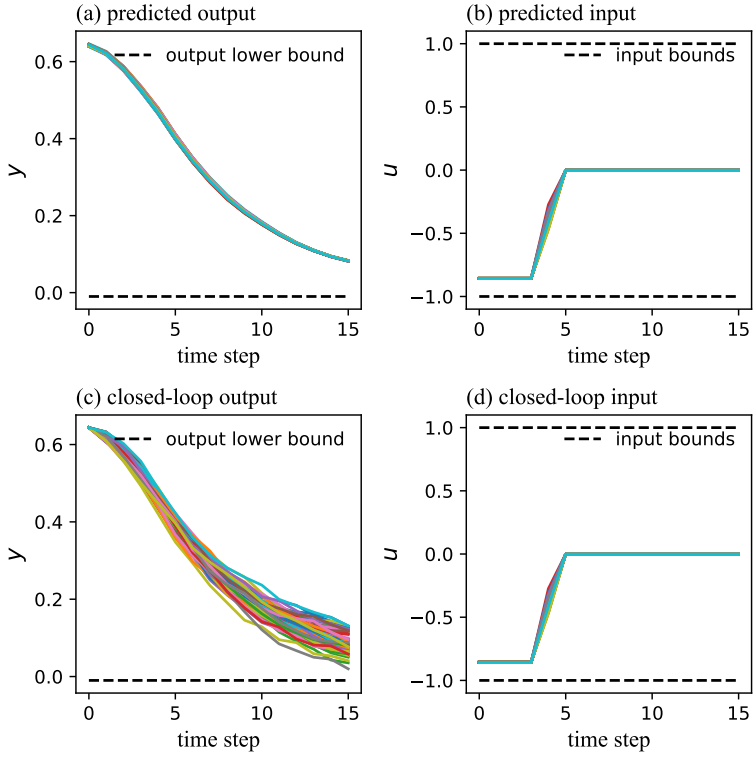


Figure 7.2: Output and input trajectories with DR-SLS approach: (a) open-loop output prediction, (b) open-loop input prediction, (c) closed-loop output, (d) closed-loop input.

trajectories. In contrast to the unstable behavior observed with the N-SLS approach, the DR-SLS approach ensures closed-loop stability while satisfying all system constraints, demonstrating its robustness and effectiveness.

Fig. 7.3 compares the cost values and constraint violation ratios of the two approaches for all samples of model parameters. It is clear from Fig. 7.3(a) that the predictive cost values of the N-SLS approach are lower than those of the DR-SLS approach. However, as shown in Fig. 7.3(b), when implementing the corresponding control policies in closed-loop simulation, the closed-loop cost value of the N-SLS approach is tremendously larger than the DR-SLS approach since the closed-loop responses with the control policy computed via the N-SLS approach become divergent. In addition, Fig. 7.3(c) shows that the N-SLS approach results in constraint violations in all closed-loop simulations, whereas the proposed DR-SLS approach robustly enforces constraint satisfaction in the presence of model mismatch and additive disturbances.

In summary, compared to the nominal SLS approach that adopts SAA and CE solutions to handle model uncertainties, the proposed DR-SLS formulation provides a systematic and principled method for enhancing robustness. Naturally, this improved robustness comes at the cost of increased conservatism in the resulting optimal solution.

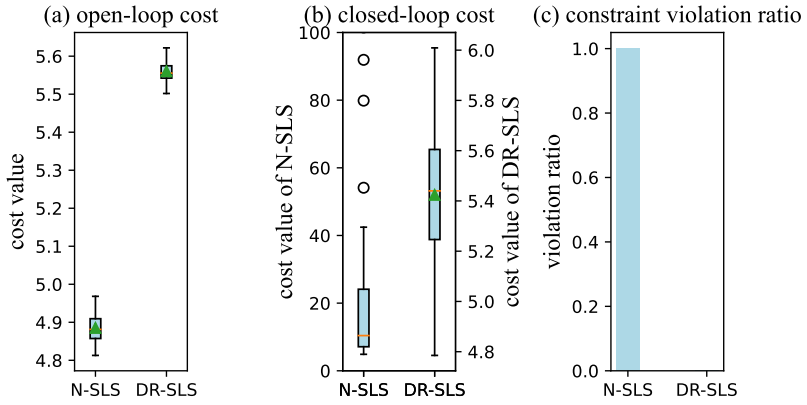


Figure 7.3: Cost values and ratios of constraint violation of N-SLS and DR-SLS: (a) open-loop cost values, (b) closed-loop cost values, (c) ratios of constraint violation.

7.6. CONCLUSION

In this chapter, we presented a distributionally robust system level synthesis design with output feedback affine control policy for uncertain linear systems subject to additive model mismatch and stochastic disturbances. By designing the SLS parameterization for the linear system in its innovation form, the proposed SLS scheme, compared to existing methods, reduces the number of decision variables and constraints and accommodates a broader class of system structures. The DR-SLS problem is cast as a robust optimization task that optimizes the expectation of the cost function while ensuring constraint satisfaction under the worst-case uncertainty distribution. We analyzed the impact of model mismatch and the empirical estimation error of stochastic uncertainties on the distribution shift between predictive closed-loop response and real closed-loop response, and derived an upper bound in the sense of the Wasserstein metric. Leveraging tools from robust SLS and DRO, we developed a tractable convex reformulation for the proposed DR-SLS problem. Numerical results demonstrate that the proposed scheme ensures safe and effective control despite the model mismatch and uncertainties.

7.A. APPENDIX

7.A.1. PROOF OF THEOREM 7.1

Proof: Proof of 1): It can be readily verified that the closed-loop responses shown in (7.13) with any strictly-lower-block-triangular matrix $\mathbf{K}$ satisfy (7.14) and (7.15).

Proof of 2): Firstly, notice that the linear constraint in (7.14) imposes that all diagonal blocks of Φ_y are identity matrices, which implies that Φ_y is invertible. Then, to complete the proof, it is sufficient to show that with the control law (7.17), the following equalities hold

$$(I-\mathbf{GK})^{-1}=\Phi_y,\ \mathbf{K}(I-\mathbf{GK})^{-1}=\Phi_u, \tag{7.34a}$$

$$(I-\mathbf{GK})^{-1}\mathbf{Gp}=\phi_y,\ (I-\mathbf{KG})^{-1}\mathbf{p}=\phi_u. \tag{7.34b}$$

The proof of the two equalities in (7.34a) follows a standard SLS design with linear policies, as shown in [176, 179]. By substituting $\mathbf{K}=\Phi_u\Phi_y^{-1}$, we obtain

$$(I-\mathbf{GK})^{-1}=(I-\mathbf{G}\Phi_u\Phi_y^{-1})^{-1} \tag{7.35a}$$

$$=(I-(\Phi_y-I)\Phi_y^{-1})^{-1}=\Phi_y, \tag{7.35b}$$

where the second equality exploits the property $\mathbf{G}\Phi_u=\Phi_y-I$ in (7.14). Similarly, one has

$$\mathbf{K}(I-\mathbf{GK})^{-1}=\Phi_u\Phi_y^{-1}\Phi_y=\Phi_u.$$

Unlike in the existing SLS design with linear policies, the parameters ϕ_y and ϕ_u in (7.34b) quantify the contribution of the bias term of the affine policy to the closed-loop output and input responses, respectively. Substituting $\mathbf{p}=\phi_u-\mathbf{K}\phi_y$ into the first equality in (7.34b) gives

$$(I-\mathbf{GK})^{-1}\mathbf{Gp}=(I-\mathbf{GK})^{-1}\mathbf{G}(\phi_u-\mathbf{K}\phi_y) \tag{7.37a}$$

$$=(I-\mathbf{GK})^{-1}(\phi_y-\mathbf{GK}\phi_y)=\phi_y, \tag{7.37b}$$

where the second equality builds upon the property $\mathbf{G}\phi_u=\phi_y$ in (7.14).

Similarly, it can be derived that

$$\begin{aligned}(I-\mathbf{KG})^{-1}\mathbf{p}&=(I+\mathbf{K}(I-\mathbf{GK})^{-1}\mathbf{G})\mathbf{p}\\&=\mathbf{p}+\mathbf{K}(I-\mathbf{GK})^{-1}\mathbf{Gp}\\&=\mathbf{p}+\mathbf{K}\phi_y=\phi_u,\end{aligned}$$

where the first equality exploits the matrix inverse lemma (Sherman-Morrison-Woodbury formula) [186], the second equality adopts the results derived in (7.37), and the last equality utilizes the relationship in (7.17). This completes the proof. □

7.A.2. PROOF OF THEOREM 7.2

Proof: Based on Theorem 7.1, we have

$$\begin{bmatrix}I & -\hat{\mathbf{G}}\end{bmatrix}\begin{bmatrix}\hat{\Phi}_y\\ \hat{\Phi}_u\end{bmatrix}=I,$$

which further leads to

$$\begin{bmatrix} I & -\mathbf{G} \end{bmatrix} \begin{bmatrix} \hat{\Phi}_y \\ \hat{\Phi}_u \end{bmatrix} = I - \Delta\hat{\Phi}_u.$$

Consequently, we arrive at

$$\begin{bmatrix} I & -\mathbf{G} \end{bmatrix} \begin{bmatrix} \hat{\Phi}_y \\ \hat{\Phi}_u \end{bmatrix} \left(I - \Delta\hat{\Phi}_u\right)^{-1} = I, \tag{7.39}$$

where the invertibility of $(I - \Delta\hat{\Phi}_u)$ is guaranteed since Δ is block-lower-triangular and $\hat{\Phi}_u$ is strictly-block-lower-triangular. The above equality provides an option for selecting $\Phi_y = \hat{\Phi}_y(I - \Delta\hat{\Phi}_u)^{-1}$ and $\Phi_u = \hat{\Phi}_u(I - \Delta\hat{\Phi}_u)^{-1}$ that ensures $\mathbf{K} = \Phi_u\Phi_y^{-1} = \hat{\Phi}_u\hat{\Phi}_y^{-1}$.

Since $\hat{\phi}_y$ and $\hat{\phi}_u$ satisfy

$$\begin{bmatrix} I & -\hat{\mathbf{G}} \end{bmatrix} \begin{bmatrix} \hat{\phi}_y \\ \hat{\phi}_u \end{bmatrix} = 0, \tag{7.40}$$

it follows that

$$\begin{bmatrix} I & -\mathbf{G} \end{bmatrix} \begin{bmatrix} \hat{\phi}_y \\ \hat{\phi}_u \end{bmatrix} = -\Delta\hat{\phi}_u. \tag{7.41}$$

Since $(\Phi_y, \Phi_u, \phi_y, \phi_u)$ are defined as true closed-loop parameterizations under the control policy parameters $(\mathbf{K}, \mathbf{p})$ derived via (7.17), Theorem 7.1 implies the following equalities

$$\begin{bmatrix} I & -\mathbf{G} \end{bmatrix} \begin{bmatrix} \phi_y \\ \phi_u \end{bmatrix} = 0, \tag{7.42a}$$

$$\mathbf{p} = \phi_u - \mathbf{K}\phi_y = \hat{\phi}_u - \mathbf{K}\hat{\phi}_y, \tag{7.42b}$$

where the relationship in (7.42b) is built upon the fact that the control policy parameter $\mathbf{p}$ is identical in both cases of the true closed-loop parameterization and nominal closed-loop parameterization.

Denoting $r_y := \phi_y - \hat{\phi}_y$ and $r_u := \phi_u - \hat{\phi}_u$, and combining (7.41) and (7.42) results in

$$\begin{cases} r_u - \mathbf{K}r_y = 0 \\ r_y - \mathbf{G}r_u = \Delta\hat{\phi}_u \end{cases} \tag{7.43}$$

Solving (7.43) gives

$$r_u = \mathbf{K}(I - \mathbf{G}\mathbf{K})^{-1}\Delta\hat{\phi}_u = \Phi_u\Delta\hat{\phi}_u, \tag{7.44a}$$

$$r_y = (I - \mathbf{G}\mathbf{K})^{-1}\Delta\hat{\phi}_u = \Phi_y\Delta\hat{\phi}_u, \tag{7.44b}$$

where the relationships in (7.34) are exploited.

Combining the results in (7.39) and (7.44) and considering the definition of (r_y, r_u) leads to the closed-loop parameterizations in (7.19) and (7.20). For these parameters, it can be verified that the condition (7.14) in Theorem 7.1 is satisfied, which justifies that the parameters in (7.20) are true closed-loop parameterizations.

Furthermore, based on the analysis shown in (7.39)-(7.44), it can be verified that the parameters $\{\Phi_y, \Phi_u, \phi_y, \phi_u\}$ together with the true model $\mathbf{G}$ satisfy (7.14), and

$$\mathbf{K} = \Phi_u \Phi_y^{-1} = \hat{\Phi}_u \hat{\Phi}_y^{-1}, \tag{7.45a}$$

$$\mathbf{p} = \phi_u - \mathbf{K}\phi_y = \hat{\phi}_u - \mathbf{K}\hat{\phi}_y. \tag{7.45b}$$

Given the results in Theorem 7.1, we can conclude that $\{\Phi_y, \Phi_u, \phi_y, \phi_u\}$ defined in (7.20) are the true closed-loop response parameterizations under the control policy $\mathbf{u} = \mathbf{K}\mathbf{y} + \mathbf{p}$, where the control policy parameters $(\mathbf{K}, \mathbf{p})$ are derived based the SLS design with the nominal model $\hat{\mathbf{G}}$ via (7.14) and (7.17). This completes the proof. □

7.A.3. PROOF OF THEOREM 7.3

Proof: The key step in proving Theorem 7.3 is to derive the computationally tractable relaxation in (7.31h) for the upper bound of the Wasserstein distance $d_W(\overline{\mathbb{P}}^{\hat{\mathscr{M}}}, \mathbb{P}^{\mathscr{M}})$ proposed in Proposition 7.1. It can be observed that the right-hand side of (7.29) consists of four terms. In the following, computationally tractable relaxations for these terms are derived.

Term 1: $||\hat{\Phi}(R_\Phi - I)(\hat{\mathbf{y}}_0 + \hat{\Theta}\mathbf{e}_i)||$. The norm sub-multiplicativity can be exploited to show that

$$||\hat{\Phi}(R_\Phi - I)(\hat{\mathbf{y}}_0 + \hat{\Theta}\mathbf{e}_i)|| \leq ||\hat{\Phi}|| \cdot ||(R_\Phi - I)|| \cdot ||\hat{\mathbf{y}}_0 + \hat{\Theta}\mathbf{e}_i||. \tag{7.46}$$

The constraint (7.31i) together with Assumption 7.4 implies that $||\Delta\hat{\Phi}_u|| \leq \rho < 1$. Then, based on Neumann series formulation, norm triangle inequality, and norm sub-multiplicativity, we can derive that

$$||R_\Phi - I|| = ||\sum_{k=1}^{\infty} (\Delta\hat{\Phi}_u)^k|| \tag{7.47a}$$

$$\leq \sum_{k=1}^{\infty} ||(\Delta\hat{\Phi}_u)^k|| \leq \sum_{k=1}^{\infty} ||\Delta\hat{\Phi}_u||^k \leq \frac{\rho}{1-\rho}. \tag{7.47b}$$

Substituting (7.47) into (7.46) gives

$$||\hat{\Phi}(R_\Phi - I)(\hat{\mathbf{y}}_0 + \hat{\Theta}\mathbf{e}_i)|| \leq \frac{\rho}{1-\rho} ||\hat{\Phi}|| \cdot ||\hat{\mathbf{y}}_0 + \hat{\Theta}\mathbf{e}_i||. \tag{7.48}$$

Term 2: $||\hat{\Phi}R_\Phi(\tilde{\mathbf{y}}_0 + \tilde{\Theta}\mathbf{e}_i)||$. Following a similar analysis as in Term 1, and exploiting Assumption 7.4, it can be derived that

$$||\hat{\Phi}R_\Phi(\tilde{\mathbf{y}}_0 + \tilde{\Theta}\mathbf{e}_i)|| \leq \frac{1}{1-\rho} ||\hat{\Phi}|| \big(\gamma_2 ||\mathbf{e}_i|| + \gamma_3\big). \tag{7.49}$$

Term 3: $||(R_\phi - I)\hat{\phi}||$. Given the definition of R_ϕ in (7.20b) and Assumption 7.4, we can show that

$$||(R_\phi - I)\hat{\phi}|| = ||[\hat{\Phi}_y^{\mathrm{T}}, \hat{\Phi}_u^{\mathrm{T}}]^{\mathrm{T}} (I - \Delta\hat{\Phi}_u)^{-1} \Delta\hat{\phi}_u|| \tag{7.50a}$$

$$\leq (||\hat{\Phi}_y|| + ||\hat{\Phi}_u||) \cdot ||(I - \Delta\hat{\Phi}_u)^{-1}|| \cdot ||\Delta|| \cdot ||\hat{\phi}_u|| \tag{7.50b}$$

$$\leq \left(\frac{\rho}{\gamma_1} + \sigma\right) \cdot \frac{\gamma_1}{1-\rho} ||\hat{\phi}_u||. \tag{7.50c}$$

Term 4: $||\hat{\Phi} R_\Phi(\hat{\Theta}+\tilde{\Theta})||d_W(\overline{\mathbb{P}}_{\mathbf{e}}, \mathbb{P}_{\mathbf{e}})$. Considering the definition of R_Φ in (7.20a) and applying similar analysis together with Assumption 7.3 leads to

$$||\hat{\Phi} R_\Phi(\hat{\Theta}+\tilde{\Theta})||d_W(\overline{\mathbb{P}}_{\mathbf{e}}, \mathbb{P}_{\mathbf{e}}) \leq \frac{\kappa}{1-\rho}||\hat{\Phi}|| \cdot \big(||\hat{\Theta}|| + \gamma_2\big). \tag{7.51}$$

Combining the results in (7.48)-(7.51) leads to $d_W(\overline{\mathbb{P}}^{\hat{\mathscr{M}}}, \mathbb{P}^{\mathscr{M}}) \leq \bar{\varepsilon}$ and constraints (7.31h). Finally, for the DRO problem in (7.21) with $\mathbb{Q} \in \mathscr{B}_{\bar{\varepsilon}}(\overline{\mathbb{P}}^{\hat{\mathscr{M}}})$, applying the results of Lemma A.2 in [51] by choosing $b = 1$, then exploiting Theorem 4.2 in [156] and considering the SLS parameterization in (7.14) and (7.15) with $\mathbf{G}$ replaced by $\hat{\mathbf{G}}$ lead to the optimization problem (7.31). This completes the proof. □

8

Conclusions and Recommendations

This chapter summarizes the main contents of this thesis, emphasizes the associated contributions, and recommends some possible future research directions.

8.1. CONCLUSIONS

This thesis investigates three topics: 1) building climate control and demand-side management (DSM), 2) data-driven robust optimization (RO), and 3) distributionally robust system-level synthesis (DR-SLS). The objective is to mitigate decision-making fragility and conservatism in optimization and predictive control of linear systems, and to enhance energy flexibility and environmental sustainability in smart building energy management. The main topics covered in this thesis are summarized below:

- A robust optimal control (ROC) framework has been developed to model the "reserve and provision" problem with binary HVAC control inputs. A new metric was introduced to quantify the flexibility scope, defined as the number of flexible switches relative to the nominal input without violating system constraints. To balance optimality and computational efficiency, affine control policies were adopted. By exploiting strong duality and big-M reformulation, the resulting ROC problem was transformed into a mixed-integer linear program, effectively addressing the computational challenges associated with adjustable binary uncertainties.
- By introducing a two-step DSM framework, we have proposed an energy-flexible MPC design to achieve DSM of HPTES systems. The proposed scheme enables assessing the energy flexibility of the HPTES system via a mixed-integer economic MPC formulation, and a realizable and efficient flexibility exploitation via DR requests. Both simulation and experimental results confirm the practical viability and effectiveness of the proposed design, as a motivating example for the future-proof EU S2 standard.
- We have developed an MPC design to provide a control-oriented solution to mitigate the acoustic nuisance of HPs. By applying piece-wise linear approximation to provide a universal representation of HP noise patterns, the proposed design is adaptable to various HP noise patterns while maintaining computational efficiency. Numerical studies show that the HP noise can be mitigated via the proposed design without a notable increase in energy cost.
- A ML-enabled approach has been proposed to construct data-driven uncertainty sets for robust optimization by integrating DBSCAN, GMM, and PCA. By representing the uncertainty set as a union of multiple subsets, the proposed approach enables an improved balance between compactness and computational tractability, and ensures adaptability to irregular data distributions. The influence of design parameters was analyzed, and detailed calibration guidelines were provided. Extensive case studies validated the effectiveness of the proposed method and highlighted the limitations and neglected issues of existing approaches.
- Based on the multi-subset uncertainty set representation, two separate schemes have been proposed:

    1. A monolithic mixed-integer representation of uncertainty set, together with a corresponding CCG-based algorithm, to address the computational challenge of exponentially increasing subsets in RO-based predictive control.

With the proposed approach, only a single mixed-integer linear optimization problem needs to be solved to find the worst-case uncertainty across all uncertainty subsets, while the conventional approach requires solving a separate optimization problem for each subset.

2. A new objective function, together with a corresponding CCG-based algorithm, to exploit the structure of the uncertainty set for reduced conservatism. By integrating the existing RO and DRO formulations, the proposed formulation mitigates the conservatism of conventional RO models while avoiding the computational complexity of existing DRO formulations.

- We have proposed a DR-SLS design for constrained uncertain linear systems to improve control robustness against both additive model mismatch and disturbances. The proposed SLS parameterization provides a concise design for MPC with affine output feedback policies. Compared with the existing approaches, the proposed design accommodates a wider class of systems and decision structures. Besides, by leveraging techniques in robust SLS and DRO, the distributional shift between the real closed-loop response and predicted closed-loop response was analyzed, and a convex relaxation of the DR-SLS formulation was derived to ensure computational tractability.

8.2. CONTRIBUTIONS OF THIS THESIS

The main contributions of the approaches and results developed in this thesis and their impact in a broader context can be summarized as follows:

- **New ROC Paradigm for "Reserve and Provision" Problems**: Chapter 2 introduces a novel ROC framework that establishes a new paradigm for modeling "reserve and provision" decisions in modern energy-system operations. Unlike prior ROC–AUS studies that primarily handle continuous uncertainties, this work extends the framework to incorporate on–off HVAC devices, offering a meaningful step toward realistic DSM applications. The proposed formulations broaden the applicability of ROC to hybrid and mixed-integer decision environments and open pathways for possible future research on robustness analysis in power systems with switchable transmission assets, discrete actuators, or other hybrid control architectures.

- **Practical MPC Solution for Unlocking the Flexibility of HPTES Systems**: Chapter 3 develops a practical and scalable MPC-based control scheme capable of exploiting the latent energy flexibility of HPTES systems. The proposed design and experimental results showcase the practical viability and effectiveness of using MPC for reduced energy cost and improved DSM participation, providing a motivating example for the EU S2 household flexibility interface. The real-world experiments validate the robustness and deployability of the approach and have already motivated our industrial partner to explore integrating and modularizing MPC tools within their commercial BMS APIs – showcasing direct technology-transfer potential.

- **A Control-Oriented Perspective on Mitigating Heat Pump Noise**: Chapter 4 provides one of the first control-oriented approaches to address noise nuisance in

heat pumps. Existing noise-mitigation solutions typically rely on intrusive retrofits or costly acoustic enclosures; in contrast, our MPC strategy regulates acoustic emissions through intelligent control of system operation. This work demonstrates that noise-aware control is both feasible and impactful, representing a promising complementary pathway to traditional engineering methods. It establishes a foundation for future development of control algorithms that simultaneously improve acoustic performance, user comfort, and environmental sustainability of heat pump technologies.

- **New Formulations and Algorithmic Contributions to Data-Driven RO**: The data-driven uncertainty sets and robust optimization formulations in Chapters 5 and 6 provide a systematic methodology for reducing conservatism and computational burden in RO models. These contributions enable scalable and realistic robust decision-making across a wide range of engineering applications, including transportation planning, power system operation, chemical processes, etc. In addition, this thesis highlights several previously overlooked issues of the existing approaches – numerical sensitivity, high computational cost, and performance degradation in complex datasets – that may hinder the practicality of existing methods. Addressing these challenges offers valuable guidance for future RO research. The accompanying Python toolkits, benchmark datasets, and algorithm-tuning guidelines serve as reusable resources for subsequent studies and application implementations.

- **New Formulations for Robust Model Predictive Control**: Chapter 7 proposes a DR-SLS framework that advances the formulation and solution of robust MPC under model mismatch and disturbance uncertainty. By jointly handling model uncertainties, exogenous disturbances, and closed-loop output-feedback policies, this work expands the theoretical frontier of robust MPC for uncertain linear systems. The provided analysis offers a foundation for developing computationally efficient and scalable robust predictive controllers. Beyond the specific application studied, the proposed formulations are extensible to a broad class of engineering systems requiring performance guarantees under uncertainty, and the proposed analysis lays the corresponding theoretical guarantees.

8.3. RECOMMENDATIONS FOR FUTURE RESEARCH

Based on the results of this thesis, we have the following recommendations for future research.

8.3.1. RECOMMENDATIONS IN TERMS OF APPLICATIONS

- **Energy Flexibility Exploitation under the Vehicle-to-Home Framework**

  The rapid adoption of electric vehicles (EVs) in residential settings offers new opportunities for integrated energy management. The vehicle-to-home (V2H) concept enables bidirectional energy exchange between EVs and household systems, allowing vehicles to act as mobile energy storage units. Leveraging the electricity

buffering capacity of EVs can significantly enhance the energy flexibility of buildings, particularly in managing peak loads and renewable generation variability. Future research should therefore develop coordinated control and optimization strategies that effectively harness the flexibility potential under the V2H framework, considering factors such as battery degradation, user driving schedules, and electricity pricing mechanisms.

- **Comprehensive Control Design for Diverse Flexibility Patterns**

  According to the S2 standard, eight distinct flexibility patterns have been defined to characterize household demand-side flexibility. The studies presented in Chapters 2 and 3 primarily addressed two of these patterns – "pause a task" and "alternative power profiles". However, to fully exploit the potential of household energy flexibility, future work should aim to design comprehensive and unified control frameworks capable of accommodating all defined patterns. This would involve developing adaptive and modular control strategies that can dynamically coordinate multiple flexibility types while ensuring user comfort, device interoperability, and compliance with grid requirements. Such work would also support the large-scale adoption of the emerging EU household flexibility interface.

- **Experimental Study of MPC for HP Noise Mitigation**

  In Chapter 4, we examined the feasibility of applying an MPC-based design to mitigate HP noise. Although numerical simulations demonstrated the effectiveness of the proposed scheme, experimental validation is required to further substantiate its practicality and robustness in real-world conditions. In addition, human factors influencing the perception of acoustic discomfort should also be further studied in the MPC design. For example, even under the same sound pressure level, humans exhibit varying sensitivities to different sound frequencies.

- **Extending Data-Driven Approaches to Other Engineering Domains**

  The data-driven uncertainty modeling and RO methods proposed in Chapters 5 & 6 offer a systematic framework for constructing compact and computationally efficient uncertainty sets, and developing computationally efficient solutions for solving RO problems. These techniques effectively balance robustness and performance, mitigating conservatism while maintaining tractability. Beyond the context of building energy management, the proposed approaches hold strong potential for broader applications in various engineering domains, such as power system operation and transportation network planning. Future research could explore these extensions to validate the generality, scalability, and practical impact of the developed data-driven RO schemes.

8.3.2. RECOMMENDATIONS IN TERMS OF THEORY

- **General Models and Additional Constraints on Uncertainties**

  In Chapter 2, the ROC-AUS problem is formulated under the assumption of linear system dynamics with affine control policies. While this assumption ensures

tractability, it inevitably limits the generality of the results. Future theoretical work could extend this framework to accommodate more general model structures—such as convex or nonlinear dynamics—and more flexible decision policies. Such extensions would enhance the applicability of the ROC framework to a broader range of systems and potentially yield less conservative control strategies.

Moreover, the current formulation simplifies the treatment of binary uncertainties by assuming that they are independent and unconstrained. A promising theoretical direction would be to incorporate coupling or structural constraints among uncertain variables, reflecting realistic conditions such as correlated demand response (DR) requests, limited resource availability, or operational interdependencies. This would bridge the gap between theoretical formulations and practical DR coordination scenarios.

- **Data-Driven Design of DR-SLS**

  In Chapter 7, the proposed DR-SLS framework relies on the availability of a known or nominal state-space model for controller synthesis. However, in practical applications, model identification could be a nontrivial and time-consuming task. A valuable theoretical extension would be to explore a data-driven variant of the DR-SLS design, where controller synthesis is based directly on measured input–output data rather than explicit model identification. Recent advances in behaviour system theory and data-enabled predictive control could provide a promising solution for such approaches. Investigating how these data-driven paradigms can be integrated within the distributionally robust predictive control framework would not only simplify implementation but also enhance the adaptivity and robustness of MPC designs in data-rich environments.

ACKNOWLEDGEMENTS

It is time to draw my doctoral journey to a close. As I begin writing these acknowledgements, countless memories come to mind. The past four years in Delft have been rewarding and transformative. Completing this journey marks one of the most significant milestones in my life. Throughout these years, I have been fortunate to receive immense help, support and encouragement from many people. I would like to take this opportunity to express my sincere gratitude to all of you.

First and foremost, I would like to express my heartfelt gratitude to my supervisors, Tamas and Neil. Thank you for giving me the opportunity, four years ago, to begin this journey under your guidance. I consider myself incredibly fortunate to have had you as my advisors. Your encouragement, insight feedback, patient guidance, and unwavering trust have made my PhD experience both smooth and deeply enjoyable. You consistently offered support whenever I encountered difficulties, while also granting me to the freedom exploring my research interests independently. Tamas, thank you for your thoughtful suggestions on my research, as well as your sincere advice regarding my career development. Neil, I am equally grateful for your warm invitation to share meals with your lovely family and the STAR group. Over the past four years, I have often shared with my friends how fortunate I am to have such exceptional and supportive supervisors. Although my doctoral studies are now coming to an end, your mentorship will continue to influence the way I approach research and teaching in the years ahead.

I would also like to express my sincere appreciation to Prof. Colin Jones for providing me the invaluable opportunity to conduct a research visit at EPFL. My time in Lausanne was both intellectually stimulating and personally enriching. I truly cherish the experience of learning from you and collaborating with your group. Thank you for your guidance and support during my stay, and also for kindly serving as a member of my doctoral committee.

I also want to extend my sincere thanks to my friends and (former) colleagues in DCSC: Changrui, Dingshan, Diyou, Jingwei, Jianfeng, Kanghui, Luyao, Meichen, Ruiyuan, Shijie, Shengling, Tian, Xiaoyu, Ximan, Yang, Yixuan, Ying, Yichen, and Zhixin. Thank you for building such a warm Chinese community within DCSC. In particular, I would like to thank Xiaoyu, Dingshan, Kanghui, Changrui, Yixuan, and Diyou. The many meals, BBQs, karaoke nights, board games, trips, and long conversations we shared together have become some of my most treasured memories. The common experiences and mutual support we have shared over the years are far too meaningful to be captured in just a few lines. My appreciation also goes to Max, Amin, Tolga, Lotfi, Reza, Filippo, Ioannis, Maarten, Mohammad, and Giray, in DCSC, for the enjoyable moments of conference travels, gym sessions, coffee breaks, dinners and snacks we experienced together. I would also like to express my gratitude to Weihong. Thank you for choosing me as your Master's thesis co-supervisor. It has been a pleasure and a privilege to work with you, and I truly enjoyed our collaboration. I also grateful to the members of STAR

lab in EWI – Yanyan, Haohui, Xueying, Leila, Mahsa, Carlos, Ngan, Noah – for making the group meetings engaging and enjoyable. Furthermore, I would also like to extend my gratitude to the secretaries of DCSC for their kind assistance and administrative support throughout my PhD journey.

My sincere thanks also go to some friends I met in Delft: Xinxin (PME), Yifei (AE), Jiao (INSA Lyon), Kai (PME), Jingyi (AE), Dinghao (Leiden University), Zhenzhen (AS), Xinhan (EWI), Nan (BK), Qiaorui (BK) and Ziao (BK). Although we met by chance through mutual friends, your companionship and the many moments we have shared – over meals, board games, gatherings, and conversations – have greatly enriched my life in Delft. Xinxin, thank you for hosting lovely dinner parties at your home and for always being willing to serve as my model while I practiced portrait photography. The conference trips we shared – such as CCTA, Benelux Meeting, and ECC – with Xinxin and Yifei are also among my most memorable experiences. Jiao, I am grateful to you and Xiaoyu for hosting so many warm and unforgettable parties. Nan, thank you for inviting me to several hometown gatherings, through which I had the opportunity in Delft to meet many outstanding people from our hometown.

I would like to thank Xue and Lingyu, my "happy brothers" in the "big white building". I still vividly remember our first trip in Paris to celebrate the New Year together. The dinners, gatherings, drinks, and long conversations and travels we shared among the past four years, have been a constant source of joy for my life. My gratitude also extends to Chen and Shuoyan. Thank you for your kindness and thoughtfulness. Our grocery shopping trips, get-togethers, and long conversations brought warmth to my final PhD stage. I truly wish we had met earlier. Shuoyan, I am grateful for your attentiveness to my inner thoughts and emotions and for always being willing to listen.

I would also like to thank my "badminton buddies" and "Avalon buddies": Yaqi, Mingkai, Sifeng, Linqiu, Na, Rui, Ruijian, Jinzhang, Yang and Yixuan. It is a pity that we met only in the final year of my PhD. Nevertheless, the joyful moments we shared – playing badminton and board games, enjoying hotpot and delicious Chinese dishes, and celebrating festivals together – greatly brightened my days in the dull winter. I wish you all ever-improving badminton skills.

To my "travel buddies" – Qi, Liangyu, and Yun (all from WUR) – what started as an unexpected encounter during our travels eventually turned into a meaningful and lasting friendship. Your care and companionship made every journey we shared truly memorable. Our footprints across Luxembourg, Austria, Germany, Sicily, and the Netherlands mark the moments of unforgettable happiness in my life.

I am also grateful to my friends in Lausanne: Bobo, Dingding, Fenglong, Jicheng, Jiawei, Shaohui, Wenbin, Wenjie, Yiwen, Yi, Yuning, and Zhaoming. Thank you for the lunches and dinners, pingpang games, karaoke, hikes, and conference trips we experienced together. Because of you, I never felt lonely during my time in Lausanne.

Finally, my deepest gratitude goes to my parents and my younger brother. Without your unconditional love and unwavering support, I could not have achieved what I have today. Wherever life may lead me, I know I will always have a home with you.

Yun Li
Delft, February 2026

BIBLIOGRAPHY

[1] Abdul Afram and Farrokh Janabi-Sharifi. “Theory and applications of HVAC control systems–A review of model predictive control (MPC)”. In: *Building and Environment* 72 (2014), pp. 343–355.

[2] Michaela Killian and Martin Kozek. “Ten questions concerning model predictive control for energy efficient buildings”. In: *Building and Environment* 105 (2016), pp. 403–412.

[3] Ye Yao and Divyanshu Kumar Shekhar. “State of the art review on model predictive control (MPC) in Heating Ventilation and Air-conditioning (HVAC) field”. In: *Building and Environment* 200 (2021), p. 107952.

[4] Alberto Bemporad and Manfred Morari. “Robust model predictive control: A survey”. In: *Robustness in identification and control*. Springer, 2007, pp. 207–226.

[5] Virginie Gabrel, Cécile Murat, and Aurélie Thiele. “Recent advances in robust optimization: An overview”. In: *European Journal of Operational Research* 235.3 (2014), pp. 471–483.

[6] Dimitris Bertsimas, David B Brown, and Constantine Caramanis. “Theory and applications of robust optimization”. In: *SIAM Review* 53.3 (2011), pp. 464–501.

[7] Aharon Ben-Tal, Arkadi Nemirovski, and Laurent El Ghaoui. *Robust optimization*. Princeton University Press, 2009.

[8] Dimitris Bertsimas and Dick Den Hertog. *Robust and adaptive optimization*. Dynamic Ideas, 2022.

[9] Paul J Goulart, Eric C Kerrigan, and Jan M Maciejowski. “Optimization over state feedback policies for robust control with constraints”. In: *Automatica* 42.4 (2006), pp. 523–533.

[10] Paul J Goulart, Eric C Kerrigan, and Daniel Ralph. “Efficient robust optimization for robust control with constraints”. In: *Mathematical Programming* 114.1 (2008), pp. 115–147.

[11] Chao Shang and Fengqi You. “A data-driven robust optimization approach to scenario-based stochastic model predictive control”. In: *Journal of Process Control* 75 (2019), pp. 24–39.

[12] Dimitris Bertsimas, Vishal Gupta, and Nathan Kallus. “Data-driven robust optimization”. In: *Mathematical Programming* 167 (2018), pp. 235–292.

[13] Ugo Rosolia, Xiaojing Zhang, and Francesco Borrelli. “Data-driven predictive control for autonomous systems”. In: *Annual Review of Control, Robotics, and Autonomous Systems* 1.1 (2018), pp. 259–286.

[14] Georgios Darivianakis, Angelos Georghiou, Roy S Smith, and John Lygeros. “The power of diversity: Data-driven robust predictive control for energy-efficient buildings and districts”. In: *IEEE Transactions on Control Systems Technology* 27.1 (2017), pp. 132–145.

[15] Chao Shang, Wei-Han Chen, Abraham Duncan Stroock, and Fengqi You. “Robust model predictive control of irrigation systems with active uncertainty learning and data analytics”. In: *IEEE Transactions on Control Systems Technology* 28.4 (2019), pp. 1493–1504.

[16] Chao Ning and Fengqi You. “Data-driven adaptive robust unit commitment under wind power uncertainty: A Bayesian nonparametric approach”. In: *IEEE Transactions on Power Systems* 34.3 (2019), pp. 2409–2418.

[17] European Commission. *Renewable energy targets*. 2023. URL: https://energy.ec.europa.eu/topics/renewable-energy/renewable-energy-directive-targets-and-rules/renewable-energy-targets_en.

[18] Xiaodong Liang. “Emerging power quality challenges due to integration of renewable energy sources”. In: *IEEE Transactions on Industry Applications* 53.2 (2016), pp. 855–866.

[19] Simon R Sinsel, Rhea L Riemke, and Volker H Hoffmann. “Challenges and solution technologies for the integration of variable renewable energy sources—a review”. In: *Renewable Energy* 145 (2020), pp. 2271–2285.

[20] Peter Palensky and Dietmar Dietrich. “Demand side management: Demand response, intelligent energy systems, and smart loads”. In: *IEEE Transactions on Industrial Informatics* 7.3 (2011), pp. 381–388.

[21] Linas Gelazanskas and Kelum AA Gamage. “Demand side management in smart grid: A review and proposals for future direction”. In: *Sustainable Cities and Society* 11 (2014), pp. 22–30.

[22] Anjukan Kathirgamanathan, Mattia De Rosa, Eleni Mangina, and Donal P Finn. “Data-driven predictive control for unlocking building energy flexibility: A review”. In: *Renewable and Sustainable Energy Reviews* 135 (2021), p. 110120.

[23] Deyslen Mariano-Hernández, Luis Hernández-Callejo, A Zorita-Lamadrid, O Duque-Pérez, and F Santos García. “A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis”. In: *Journal of Building Engineering* 33 (2021), p. 101692.

[24] Waqar A Qureshi, Nirmal-Kumar C Nair, and Mohammad M Farid. “Impact of energy storage in buildings on electricity demand side management”. In: *Energy Conversion and Management* 52.5 (2011), pp. 2110–2120.

[25] Mohammed Arif, Martha Katafygiotou, Ahmed Mazroei, Amit Kaushik, Esam Elsarrag, et al. “Impact of indoor environmental quality on occupant well-being and comfort: A review of the literature”. In: *International Journal of Sustainable Built Environment* 5.1 (2016), pp. 1–11.

[26] Ján Drgoňa, Javier Arroyo, Iago Cupeiro Figueroa, David Blum, Krzysztof Arendt, Donghun Kim, Enric Perarnau Ollé, Juraj Oravec, Michael Wetter, Draguna L Vrabie, et al. "All you need to know about model predictive control for buildings". In: *Annual Reviews in Control* 50 (2020), pp. 190–232.

[27] Zhihong Yuan, Bingzhen Chen, Gürkan Sin, and Rafiqul Gani. "State-of-the-art and progress in the optimization-based simultaneous design and control for chemical processes". In: *AIChE Journal* 58.6 (2012), pp. 1640–1659.

[28] Claudia Archetti, Lorenzo Peirano, and M Grazia Speranza. "Optimization in multimodal freight transportation problems: A Survey". In: *European Journal of Operational Research* 299.1 (2022), pp. 1–20.

[29] Fengqi You and Ignacio E Grossmann. "Stochastic inventory management for tactical process planning under uncertainties: MINLP models and algorithms". In: *AIChE Journal* 57.5 (2011), pp. 1250–1277.

[30] Raul Banos, Francisco Manzano-Agugliaro, Francisco G Montoya, Consolacion Gil, Alfredo Alcayde, and Julio Gómez. "Optimization methods applied to renewable and sustainable energy: A review". In: *Renewable and sustainable energy reviews* 15.4 (2011), pp. 1753–1766.

[31] Panos Xidonas, Ralph Steuer, and Christis Hassapis. "Robust portfolio optimization: a categorized bibliographic review". In: *Annals of Operations Research* 292.1 (2020), pp. 533–552.

[32] John R Birge and Francois Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, 2011.

[33] Marco C Campi and Simone Garatti. *Introduction to the scenario approach*. SIAM, 2018.

[34] Laurent El Ghaoui, Francois Oustry, and Hervé Lebret. "Robust solutions to uncertain semidefinite programs". In: *SIAM Journal on Optimization* 9.1 (1998), pp. 33–52.

[35] Chao Shang, Xiaolin Huang, and Fengqi You. "Data-driven robust optimization based on kernel learning". In: *Computers & Chemical Engineering* 106 (2017), pp. 464–479.

[36] Chao Ning and Fengqi You. "Data-driven stochastic robust optimization: General computational framework and algorithm leveraging machine learning for optimization under uncertainty in the big data era". In: *Computers & Chemical Engineering* 111 (2018), pp. 115–133.

[37] Max Schwenzer, Muzaffer Ay, Thomas Bergs, and Dirk Abel. "Review on model predictive control: An engineering perspective". In: *The International Journal of Advanced Manufacturing Technology* 117.5 (2021), pp. 1327–1349.

[38] Jiefeng Hu, Yinghao Shan, Josep M Guerrero, Adrian Ioinovici, Ka Wing Chan, and Jose Rodriguez. "Model predictive control of microgrids–An overview". In: *Renewable and Sustainable Energy Reviews* 136 (2021), p. 110422.

[39] Sergio Vazquez, Jose I Leon, Leopoldo G Franquelo, Jose Rodriguez, Hector A Young, Abraham Marquez, and Pericle Zanchetta. “Model predictive control: A review of its applications in power electronics”. In: *IEEE Industrial Electronics Magazine* 8.1 (2014), pp. 16–31.

[40] Meisam Farrokhifar, Hamidreza Bahmani, Behdad Faridpak, Amin Safari, David Pozo, and Marco Aiello. “Model predictive control for demand side management in buildings: A survey”. In: *Sustainable Cities and Society* 75 (2021), p. 103381.

[41] David Mayne. “Robust and stochastic model predictive control: Are we going in the right direction?” In: *Annual Reviews in Control* 41 (2016), pp. 184–192.

[42] James Blake Rawlings, David Q Mayne, and Moritz Diehl. *Model predictive control: theory, computation, and design*. 2017.

[43] Hamed Rahimian and Sanjay Mehrotra. “Distributionally robust optimization: A review”. In: *arXiv preprint arXiv:1908.05659* (2019).

[44] Fengming Lin, Xiaolei Fang, and Zheming Gao. “Distributionally robust optimization: A review on theory and applications”. In: *Numerical Algebra, Control and Optimization* 12.1 (2022), pp. 159–212.

[45] Bart PG Van Parys, Daniel Kuhn, Paul J Goulart, and Manfred Morari. “Distributionally robust control of constrained stochastic systems”. In: *IEEE Transactions on Automatic Control* 61.2 (2015), pp. 430–442.

[46] Bin Li, Yuan Tan, Ai-Guo Wu, and Guang-Ren Duan. “A distributionally robust optimization based method for stochastic model predictive control”. In: *IEEE Transactions on Automatic Control* 67.11 (2021), pp. 5762–5776.

[47] Bahar Taskesen, Dan Iancu, Çağıl Koçyiğit, and Daniel Kuhn. “Distributionally robust linear quadratic control”. In: *Advances in Neural Information Processing Systems* 36 (2023), pp. 18613–18632.

[48] Liviu Aolaritei, Marta Fochesato, John Lygeros, and Florian Dörfler. “Wasserstein tube MPC with exact uncertainty propagation”. In: *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE. 2023, pp. 2036–2041.

[49] Robert D McAllister and Peyman Mohajerin Esfahani. “Distributionally robust model predictive control: Closed-loop guarantees and scalable algorithms”. In: *IEEE Transactions on Automatic Control* 70.5 (2024), pp. 2963–2978.

[50] Astghik Hakobyan and Insoon Yang. “Wasserstein distributionally robust control of partially observable linear stochastic systems”. In: *IEEE Transactions on Automatic Control* 69.9 (2024), pp. 6121–6136.

[51] Jeremy Coulson, John Lygeros, and Florian Dörfler. “Distributionally robust chance constrained data-enabled predictive control”. In: *IEEE Transactions on Automatic Control* 67.7 (2021), pp. 3289–3304.

[52] Francesco D’Ettorre, Mattia De Rosa, Paolo Conti, Daniele Testi, and Donal Finn. “Mapping the energy flexibility potential of single buildings equipped with optimally-controlled heat pump, gas boilers and thermal storage”. In: *Sustainable Cities and Society* 50 (2019), p. 101689.

[53] Hessam Golmohamadi, Kim Guldstrand Larsen, Peter Gjøl Jensen, and Imran Riaz Hasrat. "Integration of flexibility potentials of district heating systems into electricity markets: A review". In: *Renewable and Sustainable Energy Reviews* 159 (2022), p. 112200.

[54] Mente Konsman and Ewoud Werkman. *S2 standard*. 2023. URL: https://s2standard.org/.

[55] Felix Bünning, Joseph Warrington, Philipp Heer, Roy S Smith, and John Lygeros. "Robust MPC with data-driven demand forecasting for frequency regulation with heat pumps". In: *Control Engineering Practice* 122 (2022), p. 105101.

[56] Mohsen Kalantar-Neyestanaki, Fabrizio Sossan, Mokhtar Bozorg, and Rachid Cherkaoui. "Characterizing the reserve provision capability area of active distribution networks: A linear robust optimization method". In: *IEEE Transactions on Smart Grid* 11.3 (2019), pp. 2464–2475.

[57] Omid Nohadani and Kartikey Sharma. "Optimization under decision-dependent uncertainty". In: *SIAM Journal on Optimization* 28.2 (2018), pp. 1773–1795.

[58] Xiaojing Zhang, Maryam Kamgarpour, Angelos Georghiou, Paul Goulart, and John Lygeros. "Robust optimal control with adjustable uncertainty sets". In: *Automatica* 75 (2017), pp. 249–259.

[59] Yeojun Kim, Xiaojing Zhang, Jacopo Guanetti, and Francesco Borrelli. "Robust model predictive control with adjustable uncertainty sets". In: *2018 IEEE Conference on Decision and Control (CDC)*. IEEE. 2018, pp. 5176–5181.

[60] E Langerova, J Kralicek, and M Kucera. "Air-to-water heat pump noise in residential settings: A comprehensive review". In: *Renewable and Sustainable Energy Reviews* 207 (2025), p. 114968.

[61] Chao Ning and Fengqi You. "Data-driven stochastic robust optimization: General computational framework and algorithm leveraging machine learning for optimization under uncertainty in the big data era". In: *Computers & Chemical Engineering* 111 (2018), pp. 115–133.

[62] Guoqing Hu and Fengqi You. "Multi-zone building control with thermal comfort constraints under disjunctive uncertainty using data-driven robust model predictive control". In: *Advances in Applied Energy* 9 (2023), p. 100124.

[63] Bin Li, Tao Guan, Li Dai, and Guang-Ren Duan. "Distributionally robust model predictive control with output feedback". In: *IEEE Transactions on Automatic Control* 69.5 (2023), pp. 3270–3277.

[64] Marta Fochesato and John Lygeros. "Data-driven distributionally robust bounds for stochastic model predictive control". In: *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE. 2022, pp. 3611–3616.

[65] Anushri Dixit, Mohamadreza Ahmadi, and Joel W Burdick. "Distributionally robust model predictive control with total variation distance". In: *IEEE Control Systems Letters* 6 (2022), pp. 3325–3330.

[66] Altuğ Bitlislioğlu, Tomasz T Gorecki, and Colin N Jones. "Robust tracking commitment". In: *IEEE Transactions on Automatic Control* 62.9 (2017), pp. 4451–4466.

[67] Evangelos Vrettos, Frauke Oldewurtel, and Göran Andersson. “Robust energy-constrained frequency reserves from aggregations of commercial buildings”. In: *IEEE Transactions on Power Systems* 31.6 (2016), pp. 4272–4285.

[68] Nikolaos H Lappas and Chrysanthos E Gounaris. “Robust optimization for decision-making under endogenous uncertainty”. In: *Computers & Chemical Engineering* 111 (2018), pp. 252–266.

[69] Yun Li, Neil Yorke-Smith, and Tamas Keviczky. “Robust Optimal Control With Inexact State Measurements and Adjustable Uncertainty Sets”. In: *IFAC-PapersOnLine* (2023).

[70] Long Zhao and Bo Zeng. “Vulnerability analysis of power grids with line switching”. In: *IEEE Transactions on Power Systems* 28.3 (2013), pp. 2727–2736.

[71] Aharon Ben-Tal, Alexander Goryashko, Elana Guslitzer, and Arkadi Nemirovski. “Adjustable robust solutions of uncertain linear programs”. In: *Mathematical Programming* 99.2 (2004), pp. 351–376.

[72] Dimitris Bertsimas and Angelos Georghiou. “Binary decision rules for multistage adaptive mixed-integer optimization”. In: *Mathematical Programming* 167 (2018), pp. 395–433.

[73] Jean Jacod and Philip Protter. *Probability essentials*. Springer Science & Business Media, 2004.

[74] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. 2022. URL: https://www.gurobi.com.

[75] Michael L. Bynum, Gabriel A. Hackebeil, William E. Hart, Carl D. Laird, Bethany L. Nicholson, John D. Siirola, Jean-Paul Watson, and David L. Woodruff. *Pyomo–optimization modeling in python*. Third. Vol. 67. Springer Science & Business Media, 2021.

[76] Jose Ramirez-Calderon, V Jorge Leon, and Barry Lawrence. “Robust binary linear programming under implementation uncertainty”. In: *Engineering Optimization* (2022), pp. 1–21.

[77] Dimitris Bertsimas and Melvyn Sim. “The price of robustness”. In: *Operations Research* 52.1 (2004), pp. 35–53.

[78] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[79] Dutch Government. *Klimaatnota 2022*. 2022. URL: https://www.rijksoverheid.nl/documenten/publicaties/2022/11/01/klimaatnota-2022.

[80] European Commission, Directorate General for Energy. *Clean energy for all Europeans package*. 2019. URL: https://energy.ec.europa.eu/topics/energy-strategy/clean-energy-all-europeans-package_en.

[81] European Parliament and Council of the European Union. *Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC*. 2009. URL: https://energy.ec.europa.eu/topics/renewable-energy/renewable-energy-directive-targets-and-rules/renewable-energy-directive_en.

[82] European Commission. *Heat pumps are key to enabling the clean energy transition and achieving the EU's carbon neutrality goal by 2050*. 2022. URL: https://energy.ec.europa.eu/topics/energy-efficiency/heat-pumps_en.

[83] Conrado Ermel, Marcus VA Bianchi, Ana Paula Cardoso, and Paulo S Schneider. "Thermal storage integrated into air-source heat pumps to leverage building electrification: A systematic literature review". In: *Applied Thermal Engineering* (2022), p. 118975.

[84] Sebastian Kuboth, Florian Heberle, Theresa Weith, Matthias Welzl, Andreas König-Haagen, and Dieter Brüggemann. "Experimental short-term investigation of model predictive heat pump control in residential buildings". In: *Energy and Buildings* 204 (2019), p. 109444.

[85] Francesco D'Ettorre, Paolo Conti, Eva Schito, and Daniele Testi. "Model predictive control of a hybrid heat pump system and impact of the prediction horizon on cost-saving potential and optimal storage capacity". In: *Applied Thermal Engineering* 148 (2019), pp. 524–535.

[86] Mikkel Urban Kajgaard, Jesper Mogensen, Anders Wittendorff, Attila Todor Veress, and Benjamin Biegel. "Model predictive control of domestic heat pump". In: *2013 American Control Conference*. IEEE. 2013.

[87] Alice Mugnini, Maarten Evens, and Alessia Arteconi. "Model predictive controls for residential buildings with heat pumps: Experimentally validated archetypes to simplify the large-scale application". In: *Energy and Buildings* 320 (2024), p. 114632.

[88] Cara R Touretzky and Michael Baldea. "Integrating scheduling and control for economic MPC of buildings with energy storage". In: *Journal of Process Control* 24.8 (2014), pp. 1292–1300.

[89] Joan Tarragona, Anna Laura Pisello, Cèsar Fernández, Luisa F Cabeza, Jorge Payá, Javier Marchante-Avellaneda, and Alvaro de Gracia. "Analysis of thermal energy storage tanks and PV panels combinations in different buildings controlled through model predictive control". In: *Energy* 239 (2022), p. 122201.

[90] Elias Kinab, Dominique Marchio, Philippe Rivière, and Assaad Zoughaib. "Reversible heat pump model for seasonal performance optimization". In: *Energy and Buildings* 42.12 (2010), pp. 2269–2280.

[91] Mi-Soo Shin, Hey-Suk Kim, Dong-Soon Jang, Sang-Nam Lee, Young-Soo Lee, and Hyung-Gi Yoon. "Numerical and experimental study on the design of a stratified thermal storage system". In: *Applied Thermal Engineering* 24.1 (2004), pp. 17–27.

[92] Brecht Baeten, Thomas Confrey, Sébastien Pecceu, Frederik Rogiers, and Lieve Helsen. "A validated model for mixing and buoyancy in stratified hot water storage tanks for use in building energy simulations". In: *Applied Energy* 172 (2016), pp. 217–229. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2016.03.118.

[93] Hessam Golmohamadi, Kim Guldstrand Larsen, Peter Gjøl Jensen, and Imran Riaz Hasrat. "Optimization of power-to-heat flexibility for residential buildings in response to day-ahead electricity price". In: *Energy and Buildings* 232 (2021), p. 110665.

[94] Roel De Coninck and Lieve Helsen. "Quantification of flexibility in buildings by cost curves–Methodology and application". In: *Applied Energy* 162 (2016), pp. 653–665.

[95] Yun Li, Neil Yorke-Smith, and Tamas Keviczky. "Unlocking Energy Flexibility From Thermal Inertia of Buildings: A Robust Optimization Approach". In: *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE. 2023, pp. 2555–2562.

[96] Mente Konsman, Ewoud Werkman, and in collaboration with TC 205 WG 18 members. *S2 White Paper*. 2023. URL: https://s2standard.org/.

[97] Mente J Konsman, Wilco E Wijbrandi, and George B Huitema. "Unlocking residential Energy Flexibility on a large scale through a newly standardized interface". In: *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE. 2020, pp. 1–5.

[98] Yun Li, Neil Yorke-Smith, and Tamas Keviczky. "Robust Optimal Control With Binary Adjustable Uncertainties". In: *2024 European Control Conference (ECC)*. IEEE. 2024, pp. 3721–3727.

[99] Fabián Correa and Cristian Cuevas. "Air-water heat pump modelling for residential heating and domestic hot water in Chile". In: *Applied Thermal Engineering* 143 (2018), pp. 594–606.

[100] Na Zhu, Pingfang Hu, Wei Wang, Jianming Yu, and Fei Lei. "Performance analysis of ground water-source heat pump system with improved control strategies for building retrofit". In: *Renewable energy* 80 (2015), pp. 324–330.

[101] Björn Felten and Christoph Weber. "The value(s) of flexible heat pumps–Assessment of technical and economic conditions". In: *Applied Energy* 228 (2018), pp. 1292–1319.

[102] Patrick J Luickx, Lieve M Helsen, and William D D'haeseleer. "Influence of massive heat-pump introduction on the electricity-generation mix and the GHG effect: Comparison between Belgium, France, Germany and The Netherlands". In: *Renewable and Sustainable Energy Reviews* 12.8 (2008), pp. 2140–2158.

[103] Iván De la Cruz-Loredo, Daniel Zinsmeister, Thomas Lickleder, Carlos E. Ugalde-Loo, Daniel A. Morales, Héctor Bastida, Vedran S. Perić, and Arslan Saleem. "Experimental validation of a hybrid 1-D multi-node model of a hot water thermal energy storage tank". In: *Applied Energy* 332 (2023), p. 120556. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2022.120556.

[104] Soroush Rastegarpour, Mahshid Ghaemi, and Luca Ferrarini. “A predictive control strategy for energy management in buildings with radiant floors and thermal storage”. In: *2018 SICE International Symposium on Control Systems (SICE ISCS)*. IEEE. 2018, pp. 67–73.

[105] Rune Grønborg Junker, Armin Ghasem Azar, Rui Amaral Lopes, Karen Byskov Lindberg, Glenn Reynders, Rishi Relan, and Henrik Madsen. “Characterizing the energy flexibility of buildings and districts”. In: *Applied Energy* 225 (2018), pp. 175–182.

[106] Søren Østergaard Jensen, Anna Marszal-Pomianowska, Roberto Lollini, Wilmer Pasut, Armin Knotzer, Peter Engelmann, Anne Stafford, and Glenn Reynders. “IEA EBC annex 67 energy flexible buildings”. In: *Energy and Buildings* 155 (2017), pp. 25–34.

[107] Weihong Tang, Yun Li, Shalika Walker, and Tamas Keviczky. “Model Predictive Control Design for Unlocking the Energy Flexibility of Heat Pump and Thermal Energy Storage Systems”. In: *2024 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE. 2024, pp. 433–439.

[108] Weihong Tang, Yun Li, Shalika Walker, and Tamas Keviczky. “Model Predictive Control for Unlocking Energy Flexibility of Heat Pump and Thermal Energy Storage Systems: Experimental Results”. In: *arXiv preprint arXiv:2507.17552* (2025).

[109] William E Hart, Jean-Paul Watson, and David L Woodruff. “Pyomo: modeling and solving mathematical programs in Python”. In: *Mathematical Programming Computation* 3.3 (2011), pp. 219–260.

[110] Tim Schwickart, Holger Voos, Mohamed Darouach, and Souad Bezzaoucha. “A flexible move blocking strategy to speed up model-predictive control while retaining a high tracking performance”. In: *2016 European Control Conference (ECC)*. IEEE. 2016, pp. 764–769.

[111] European Heat Pump Association (EHPA). *European Heat Pump Market and Statistics Report 2024*. 2024. URL: https://www.ehpa.org/market-data/.

[112] European Commission. *Heat Pumps*. 2023. URL: https://energy.ec.europa.eu/topics/energy-efficiency/heat-pumps_en.

[113] Jillian Ambrose. *Will my heat pump be a noise nuisance to my neighbours?* 2024. URL: https://www.theguardian.com/business/article/2024/jul/09/will-my-heat-pump-be-a-noise-nuisance-to-my-neighbours.

[114] M Torjussen. *Sound from domestic air source heat pumps: a case study*. 2020. URL: https://www.ioa.org.uk/system/files/proceedings/5._torjussen._m.pdf.

[115] European Heat Pump Association (EHPA). *Heat pumps and sound*. 2020. URL: https://www.ehpa.org/wp-content/uploads/2023/07/HEAT_PUMPS_AND_SOUND_-_WHITE_PAPER-compressed.pdf.

[116] Christoph Reichl. *Annex 51: Acoustic signatures of heat pumps*. URL: https://heatpumpingtechnologies.org/annex51/.

[117] Christoph Reichl. *Annex 63: Placement impact on heat pump acoustics*. URL: https://heatpumpingtechnologies.org/annex63/.

[118] Sebastian Wagner, Xavier Carniel, Jens Rohlfing, Karlheinz Bay, and Henrik Hellgren. *Overview on Heat Pump Component Noise and Noise Control Techniques*. 2020. URL: https://heatpumpingtechnologies.org/annex51/wp-content/uploads/sites/59/2021/10/iea-hpt-annex-51-d3.pdf.

[119] Anja Thielecke, Santiago Peréz Brovia, Johannes Hübel, David Gänzle, Jürgen Herbst, and Stefan Becker. "Active electric machine control to reduce compressor noise in heat pumps". In: *Applied Acoustics* 213 (2023), p. 109601.

[120] M Torjussen, J Harvie-Clark, A Lamacraft, and P Rogers. *Noise from ASHPS - what we know?* 2019. URL: https://www.apexacoustics.co.uk/wp-content/uploads/2023/10/NOISE-FROM-ASHPS-WHAT-WE-KNOW.pdf.

[121] Frauke Oldewurtel, Alessandra Parisio, Colin N Jones, Dimitrios Gyalistras, Markus Gwerder, Vanessa Stauch, Beat Lehmann, and Manfred Morari. "Use of model predictive control and weather forecasts for energy efficient building climate control". In: *Energy and Buildings* 45 (2012), pp. 15–27.

[122] Jicheng Shi, Yingzhao Lian, Christophe Salzmann, and Colin N Jones. "Adaptive data-driven prediction in a building control hierarchy: A case study of demand response in Switzerland". In: *Energy and Buildings* 333 (2025), p. 115498.

[123] Zequn Wang and Yuxiang Chen. "Data-driven modeling of building thermal dynamics: Methodology and state of the art". In: *Energy and Buildings* 203 (2019), p. 109405.

[124] Peder Bacher and Henrik Madsen. "Identifying suitable models for the heat dynamics of buildings". In: *Energy and Buildings* 43.7 (2011), pp. 1511–1522.

[125] Haglund Caroline Stignor, Ola Gustafsson, and Hendrik Hellgren. *Heat pump noise - operational dependence and seasonal averaging*. 2024. URL: https://heatpumpingtechnologies.org/annex51/wp-content/uploads/sites/59/2019/09/heat-pump-noise-operation-dependence-and-seasonal-averaging.pdf.

[126] Xueqi Zhang, Meng Zhao, and Rencai Dong. "Time-series prediction of environmental noise for urban IoT based on long short-term memory recurrent neural network". In: *Applied Sciences* 10.3 (2020), p. 1144.

[127] Jérémy Renaud, Ralph Karam, Michel Salomon, and Raphaël Couturier. "Deep learning and gradient boosting for urban environmental noise monitoring in smart cities". In: *Expert Systems with Applications* 218 (2023), p. 119568.

[128] David Blum, Javier Arroyo, Sen Huang, Ján Drgoňa, Filip Jorissen, Harald Taxt Walnum, Yan Chen, Kyle Benne, Draguna Vrabie, Michael Wetter, et al. "Building optimization testing framework (BOPTEST) for simulation-based benchmarking of control strategies in buildings". In: *Journal of Building Performance Simulation* 14.5 (2021), pp. 586–610.

[129] Yun Li, Jicheng Shi, Colin N Jones, Neil Yorke-Smith, and Tamas Keviczky. *Model Predictive Building Climate Control for Mitigating Heat Pump Noise Pollution (Extended Version)*. 2025. URL: https://arxiv.org/abs/2504.04182.

[130] Eric Jones, Travis Oliphant, Pearu Peterson, et al. *SciPy: Open source scientific tools for Python*. 2001. URL: http://www.scipy.org/.

[131] Pieter Thomas, Francesco Aletta, Tara Vander Mynsbrugge, Karlo Filipan, Arne Dijckmans, Lieven De Geetere, Dick Botteldooren, Mirko Petrovic, Patricia De Vriendt, Dominique Van De Velde, et al. "Evaluation and improvement of the acoustic comfort in nursing homes: a case study in Flanders, Belgium". In: *11th European Congress and Exposition on Noise Control Engineering (Euronoise 2018)*. 2018, pp. 405–412.

[132] European Environment Agency. *Lden*. 2001. URL: https://www.eea.europa.eu/help/glossary/eea-glossary/lden.

[133] Marco C Campi and Simone Garatti. "The exact feasibility of randomized solutions of uncertain convex programs". In: *SIAM Journal on Optimization* 19.3 (2008), pp. 1211–1230.

[134] Giuseppe Carlo Calafiore and Marco C Campi. "The scenario approach to robust control design". In: *IEEE Transactions on Automatic Control* 51.5 (2006), pp. 742–753.

[135] Aharon Ben-Tal and Arkadi Nemirovski. "Robust convex optimization". In: *Mathematics of Operations Research* 23.4 (1998), pp. 769–805.

[136] L Jeff Hong, Zhiyuan Huang, and Henry Lam. "Learning-based robust optimization: Procedures and statistical guarantees". In: *Management Science* 67.6 (2021), pp. 3447–3467.

[137] Marc Goerigk and Jannis Kurtz. "Data-driven robust optimization using deep neural networks". In: *Computers & Operations Research* 151 (2023), p. 106087.

[138] Chao Ning and Fengqi You. "Data-driven adaptive nested robust optimization: general modeling framework and efficient computational algorithm for decision making under uncertainty". In: *AIChE Journal* 63.9 (2017), pp. 3790–3817.

[139] Meysam Cheramin, Richard Li-Yang Chen, Jianqiang Cheng, and Ali Pinar. "Data-driven robust optimization using scenario-induced uncertainty sets". In: *arXiv preprint arXiv:2107.04977* (2021).

[140] Chao Ning and Fengqi You. "Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods". In: *Computers & Chemical Engineering* 112 (2018), pp. 190–210.

[141] Wei-Han Chen and Fengqi You. "Smart greenhouse control under harsh climate conditions based on data-driven robust model predictive control with principal component analysis and kernel density estimation". In: *Journal of Process Control* 107 (2021), pp. 103–113.

[142] Chenhan Zhang, Zhenlei Wang, and Xin Wang. “Machine learning-based data-driven robust optimization approach under uncertainty”. In: *Journal of Process Control* 115 (2022), pp. 1–11.

[143] Yi Zhang, Xuanzhi Jin, Yiping Feng, and Gang Rong. “Data-driven robust optimization under correlated uncertainty: a case study of production scheduling in ethylene plant”. In: *Computers & Chemical Engineering* 109 (2018), pp. 48–67.

[144] Shulei Zhang, Runda Jia, Dakuo He, and Fei Chu. “Data-driven robust optimization based on principle component analysis and cutting plane methods”. In: *Industrial & Engineering Chemistry Research* 61.5 (2022), pp. 2167–2182.

[145] Ning Zhao and Fengqi You. “Sustainable power systems operations under renewable energy induced disjunctive uncertainties via machine learning-based robust optimization”. In: *Renewable and Sustainable Energy Reviews* 161 (2022), p. 112428.

[146] Yun Li. *DGP_Set*. 2024. URL: https://github.com/li-yun/DGP_Set.

[147] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. “A density-based algorithm for discovering clusters in large spatial databases with noise”. In: *kdd*. Vol. 96. 34. 1996, pp. 226–231.

[148] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[149] Todd K Moon. “The expectation-maximization algorithm”. In: *IEEE Signal Processing Magazine* 13.6 (1996), pp. 47–60.

[150] Jonathon Shlens. “A tutorial on principal component analysis”. In: *arXiv preprint arXiv:1404.1100* (2014).

[151] Yun Li, Neil Yorke-Smith, and Tamas Keviczky. “Machine learning enabled uncertainty set for data-driven robust optimization”. In: *Journal of Process Control* 144 (2024), p. 103339.

[152] Long Zhao and Bo Zeng. *An exact algorithm for two-stage robust optimization with mixed integer recourse problems*. 2012. URL: https://optimization-online.org/wp-content/uploads/2012/01/3310.pdf.

[153] Bo Zeng and Long Zhao. “Solving two-stage robust optimization problems using a column-and-constraint generation method”. In: *Operations Research Letters* 41.5 (2013), pp. 457–461.

[154] Dimitris Bertsimas and Dick den Hertog. *Robust and Adaptive Optimization*. Belmont, MA: Dynamic Ideas LLC, 2022.

[155] Grani A Hanasusanto, Vladimir Roitch, Daniel Kuhn, and Wolfram Wiesemann. “A distributionally robust perspective on uncertainty quantification and chance constrained programming”. In: *Mathematical Programming* 151 (2015), pp. 35–62.

[156] Peyman Mohajerin Esfahani and Daniel Kuhn. “Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations”. In: *Mathematical Programming* 171.1 (2018), pp. 115–166.

[157] Dimitris Bertsimas, Xuan Vinh Doan, Karthik Natarajan, and Chung-Piaw Teo. “Models for minimax stochastic linear optimization problems with risk aversion”. In: *Mathematics of Operations Research* 35.3 (2010), pp. 580–602.

[158] Wei Wei. “Tutorials on advanced optimization methods”. In: *arXiv preprint arXiv:2007.13545* (2020). URL: https://arxiv.org/abs/2007.13545.

[159] Daniel Kuhn, Soroosh Shafiee, and Wolfram Wiesemann. “Distributionally robust optimization”. In: *Acta Numerica* 34 (2025), pp. 579–804.

[160] Longyan Li, Chao Ning, Haifeng Qiu, Wenli Du, and Zhaoyang Dong. “Online data-stream-driven distributionally robust optimal energy management for hydrogen-based multimicrogrids”. In: *IEEE Transactions on Industrial Informatics* 20.3 (2023), pp. 4370–4384.

[161] Xutao Ma, Chao Ning, Longyan Li, Haifeng Qiu, Wei Gu, and Zhaoyang Dong. “Bayesian nonparametric two-stage distributionally robust unit commitment optimization: From global multimodality to local trimming-wasserstein ambiguity”. In: *IEEE Transactions on Power Systems* 39.5 (2024), pp. 6702–6715.

[162] Giuseppe C Calafiore and Lorenzo Fagiano. “Robust model predictive control via scenario optimization”. In: *IEEE Transactions on Automatic Control* 58.1 (2012), pp. 219–224.

[163] Ali Mesbah. “Stochastic model predictive control: An overview and perspectives for future research”. In: *IEEE Control Systems Magazine* 36.6 (2016), pp. 30–44.

[164] Dimitris Bertsimas and Jack Dunn. *Machine learning under a modern optimization lens*. Dynamic Ideas LLC Waltham, 2019.

[165] Yuwei Wang, Yuanjuan Yang, Liu Tang, Wei Sun, and Bingkang Li. “A Wasserstein based two-stage distributionally robust optimization model for optimal operation of CCHP micro-grid under uncertainties”. In: *International Journal of Electrical Power & Energy Systems* 119 (2020), p. 105941.

[166] Yun Li, Neil Yorke-Smith, and Tamas Keviczky. *Supplementary material*. 2024. URL: https://github.com/li-yun/data_driven_ro_supplementary.

[167] Harald Johan van Heerde. “Models for sales promotion effects based on store-level scanner data”. In: (1999).

[168] David W Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

[169] James B Rawlings. “Tutorial overview of model predictive control”. In: *IEEE Control Systems Magazine* 20.3 (2002), pp. 38–52.

[170] Daniel Kuhn, Soroosh Shafiee, and Wolfram Wiesemann. “Distributionally robust optimization”. In: *Acta Numerica* 34 (2025), pp. 579–804.

[171] Francesco Borrelli. *Constrained optimal control of linear and hybrid systems*. Springer, 2003.

[172] James Anderson, John C Doyle, Steven H Low, and Nikolai Matni. "System level synthesis". In: *Annual Reviews in Control* 47 (2019), pp. 364–393.

[173] Jerome Sieber, Samir Bennani, and Melanie N Zeilinger. "A system level approach to tube-based model predictive control". In: *IEEE Control Systems Letters* 6 (2021), pp. 776–781.

[174] Shaoru Chen, Han Wang, Manfred Morari, Victor M Preciado, and Nikolai Matni. "Robust closed-loop model predictive control via system level synthesis". In: *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE. 2020, pp. 2152–2159.

[175] Lukas Schüepp, Giulia De Pasquale, Florian Dörfler, and Carmen Amo Alonso. "System Level Synthesis for Affine Control Policies: Model Based and Data-Driven Settings". In: *arXiv preprint arXiv:2504.01677* (2025).

[176] Luca Furieri, Andrea Martin, Baiwei Guo, and Giancarlo Ferrari-Trecate. *Learning Robustly Safe Output Feedback Controllers from Noisy Data with Performance Guarantees*. URL: https://infoscience.epfl.ch/server/api/core/bitstreams/e67849e9-5653-400e-a5f7-4a1a8b4dabbf/content.

[177] Luca Furieri, Yang Zheng, Antonis Papachristodoulou, and Maryam Kamgarpour. "An input–output parametrization of stabilizing controllers: Amidst Youla and system level synthesis". In: *IEEE Control Systems Letters* 3.4 (2019), pp. 1014–1019.

[178] Riccardo Cescon, Andrea Martin, and Giancarlo Ferrari-Trecate. "Data-driven Distributionally Robust Control Based on Sinkhorn Ambiguity Sets". In: *arXiv preprint arXiv:2503.20703* (2025).

[179] Luca Furieri, Baiwei Guo, Andrea Martin, and Giancarlo Ferrari-Trecate. "Near-optimal design of safe output-feedback controllers from noisy data". In: *IEEE Transactions on Automatic Control* 68.5 (2022), pp. 2699–2714.

[180] Valentina Breschi, Alessandro Chiuso, and Simone Formentin. "Data-driven predictive control in a stochastic setting: a unified framework". In: *Automatica* 152 (2023), p. 110961.

[181] Yibo Wang, Keyou You, Dexian Huang, and Chao Shang. "Data-driven output prediction and control of stochastic systems: An innovation-based approach". In: *Automatica* 171 (2025), p. 111897.

[182] Gijs Van der Veen, Jan-Willem van Wingerden, Marco Bergamasco, Marco Lovera, and Michel Verhaegen. "Closed-loop subspace identification methods: an overview". In: *IET Control Theory & Applications* 7.10 (2013), pp. 1339–1358.

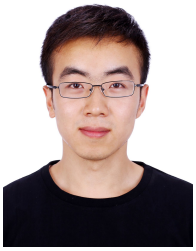
[183] Alessandro Chiuso. "The role of vector autoregressive modeling in predictor-based subspace identification". In: *Automatica* 43.6 (2007), pp. 1034–1048.

[184] Francesco Micheli, Anastasios Tsiamis, and John Lygeros. "Data-Driven Distributionally Robust System Level Synthesis". In: *arXiv preprint arXiv:2405.18142* (2024).

[185] Changrui Liu, Shengling Shi, and Bart De Schutter. “Stability and Performance Analysis of Model Predictive Control of Uncertain Linear Systems”. In: *arXiv preprint arXiv:2405.15552* (2024).

[186] William W Hager. “Updating the inverse of a matrix”. In: *SIAM review* 31.2 (1989), pp. 221–239.

CURRICULUM VITÆ



Yun Li was born on April 15, 1993, in Qingyang, Gansu, China. He obtained the B.Sc. degree from Beijing University of Chemical Technology, Beijing, China, in 2015, and the M.Sc. degree from Beihang University, Beijing, China, in 2018. During 2019–2021, he studied at New York University and University of British Columbia as a doctoral researcher.

Since Jan. 2022, he has been pursuing his doctoral degree in Delft Center for Systems and Control at Delft University of Technology, delft, the Netherlands, under the supervision of Prof. Tamas Keviczky and Prof. Neil Yorke-Smith. From September to November 2024, he conducted a research visit to the Automatic Control Laboratory, EPFL, Lausanne, Switzerland, hosted by Prof. Colin Jones.

His Ph.D. research focuses on developing robust optimization and predictive control algorithms and their applications in the energy management of smart buildings. His research interests include model predictive control, robust optimization, data-driven control, learning-based control, and applications to energy systems.

LIST OF PUBLICATIONS

9. Li, Y., Yorke-Smith, N., & Keviczky, T. (2026). On data-driven robust optimization with multiple uncertainty subsets: Unified uncertainty set representation and mitigating conservatism. *Journal of Process Control*, 158, 103611.

8. Li, Y., Shi, J., Jones, C. N., Yorke-Smith, N., & Keviczky, T. (2025). Distributionally Robust System Level Synthesis With Output Feedback Affine Control Policy. *Submitted to journal*, under review.

7. Li, Y., Shi, J., Jones, C. N., Yorke-Smith, N., & Keviczky, T. (2025). Model predictive building climate control for mitigating heat pump noise pollution. *European Journal of Control*, 101323.

6. Tang[1], W., Li[1], Y., Walker, S., & Keviczky, T. (2025). Model Predictive Control for Unlocking Energy Flexibility of Heat Pump and Thermal Energy Storage Systems: Experimental Results. *Submitted to journal*, under review. ([1]: These authors contributed equally to this work.)

5. Li, Y., Yorke-Smith, N., & Keviczky, T. (2024). Machine learning enabled uncertainty set for data-driven robust optimization. *Journal of Process Control*, 144, 103339.

4. Tang, W., Li, Y., Walker, S., & Keviczky, T. (2024, August). Model Predictive Control Design for Unlocking the Energy Flexibility of Heat Pump and Thermal Energy Storage Systems. In 2024 IEEE Conference on Control Technology and Applications (CCTA) (pp. 433-439). IEEE.

3. Li, Y., Yorke-Smith, N., & Keviczky, T. (2024, June). Robust Optimal Control With Binary Adjustable Uncertainties. In 2024 European Control Conference (ECC) (pp. 3721-3727). IEEE.

2. Li, Y., Yorke-Smith, N., & Keviczky, T. (2023). Robust optimal control with inexact state measurements and adjustable uncertainty sets. IFAC-PapersOnLine, 56(2), 562-569.

1. Li, Y., Yorke-Smith, N., & Keviczky, T. (2023, December). Unlocking energy flexibility from thermal inertia of buildings: A robust optimization approach. In 2023 62nd IEEE Conference on Decision and Control (CDC) (pp. 2555-2562). IEEE.

: included in this thesis.

Delft University of Technology

ISBN 978-94-6518-276-6