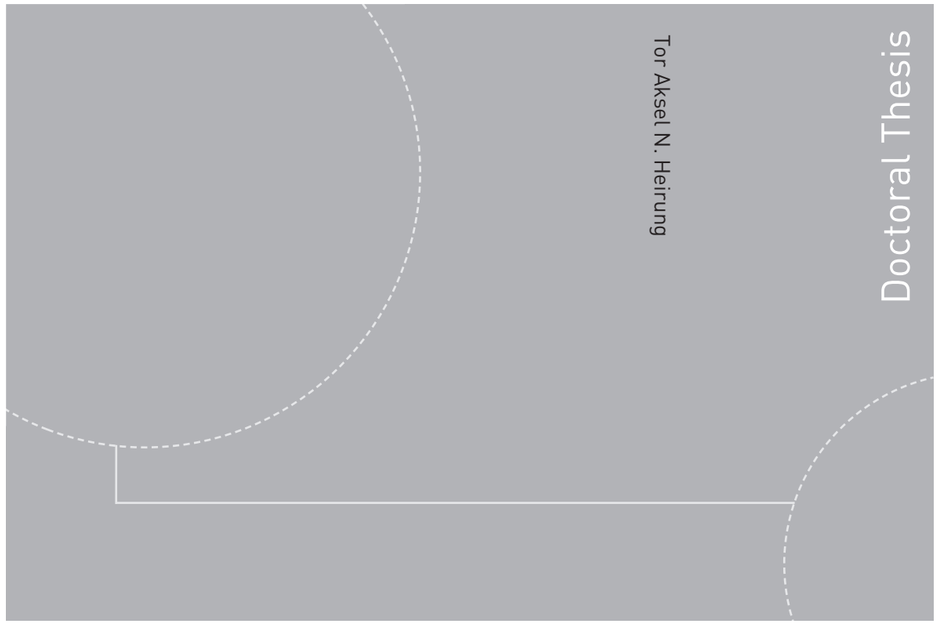
ISBN 978-82-326-1524-7 (printed version)
ISBN 978-82-326-1525-4 (electronic version)
ISSN 1503-8181



Doctoral theses at NTNU, 2016:93

Tor Aksel N. Heirung

Dual control: optimal, adaptive decision-making under uncertainty

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

Doctoral theses at NTNU, 2016:93

NTNU

NTNU
Norwegian University of
Science and Technology

Tor Aksel N. Heirung

Dual control: optimal, adaptive decision-making under uncertainty

Thesis for the degree of Philosophiae Doctor

Trondheim, April 2016

Norwegian University of Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Engineering Cybernetics

ISBN 978-82-326-1524-7 (printed version)
ISBN 978-82-326-1525-4 (electronic version)
ISSN 1503-8181

IKT-report: 2016-8-W
Doctoral theses at NTNU, 2016:93

Printed by Skipnes Kommunikasjon as

DUAL CONTROL: OPTIMAL, ADAPTIVE DECISION-MAKING UNDER UNCERTAINTY

TOR AKSEL N. HEIRUNG

for Lise

SUMMARY

MOST DECISIONS BASED ON predictions from a model have uncertain outcomes. The uncertainty may be exogenous or endogenous to the modeled process, or both, and can greatly affect the degree to which the decision-maker's goals are met. In this thesis I study optimal control problems for systems with endogenous uncertainty that can be reduced by manipulating the system input. When the decision-maker can improve performance through actively reducing uncertainty ("active learning"), there is a dual nature to the optimal sequence of decisions; the decisions or inputs must direct the process toward the desired state and also ensure that information-rich data be generated so that decision-relevant uncertainty is resolved.

The dual-control problem can be defined as that of minimizing the expected output error

$$\mathrm{E}\left[\sum_{k=t}^{t+N-1} (y(k+1) - y^*(k+1 \mid t))^2 \;\middle|\; \mathcal{Y}(t+N-1 \mid t)\right] \qquad (*)$$

where y^* is the output reference, $\mathcal{Y}(t+N-1 \mid t)$ represents future information up to time $t+N-1$ in addition to all past information, and the model for the system output y is not fully known. My coworkers and I propose a novel reformulation technique for a probabilistically-constrained stochastic dual-control problem and show that the optimal strategy for minimizing this cost function involves active exploration of the plant to generate informative data. It is consequently necessary that the model that informs the decision-making include how future data resolve uncertainty. The reformulation permits practical algorithms for true dual control for a class of systems, allows new interpretation of earlier approaches, and may guide approximate dual-control designs where no exact results are obtainable.

In addition to useful algorithms, this thesis contains a number of conceptual insights. In particular, the most recent results provide a foundation from which we argue that the conventional dual-control interpretation involving a trade-off between control and exploration is a false dichotomy. These derivations clearly show, for a specific system class, that control of the nominal model and a specific form of uncertainty reduction are both necessary components of the optimal control,

as opposed to separate entities in which uncertainty reduction can be sacrificed for increased control performance.

I consider the approaches to dual control presented in this thesis as falling into one of the following three categories: minimization of (*i*) a heuristic objective that is different from, yet still reduces, the dual objective ($*$); (*ii*) a systematic approximation of ($*$); and (*iii*) an exact reformulation of ($*$). This main contributions in this thesis are taken from the following three papers:

A Heirung, T.A.N., Foss, B., and Ydstie, B.E. (2015). "MPC-based dual control with online experiment design." *Journal of Process Control* 32, pp. 64–76.

B Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2013b). "An MPC approach to dual control." In: *Dynamics and Control of Process Systems*. Mumbai, India, pp. 69–74.

C Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2015b). "Dual adaptive model-predictive control." *Under revision for resubmission to Automatica*.

The following conference paper presents a possible extension of Paper B to the multivariable case, in addition to experimental results:

D Kumar, K. et al. (2015). "Experimental evaluation of a MIMO adaptive dual MPC." In: *Advanced Control of Chemical Processes*. Whistler, Canada, pp. 546–551.

Each of the three main papers introduce control designs that involve solving a finite-horizon optimal-control problem (O.C.P.) at every sampling instant, with the initial state set to the current state of the plant. Common to the algorithms is their foundation in model-predictive control (M.P.C.). The approaches each involve augmenting the standard O.C.P. in M.P.C. with cost-function terms and constraints, with the result that the controller generates plant inputs with a dual effect. The cost-function terms incentivize plant exploration while the additional constraints model the mechanisms for learning by propagating the uncertainty (or system information) as a function of the signals in the loop. The controller thereby endogenizes the learning process through its capability to predict how the control decisions affect future uncertainty, resulting in semi-closed-loop feedback control. Each of the proposed algorithms are (indirect) adaptive in nature, a consequence of receding-

horizon solution the O.C.P. with updated plant-parameter statistics in the prediction model.

We consider linear single-input single-output (SISO) systems with constant unknown parameters and a Gaussian disturbance. The choice of simple system structures is motivated by our focus on developing a foundation on which to base design choices in approximate dual control for more complex problems. While we have made significant progress toward this goal, the linear, Gaussian, SISO dual-control problem is far from solved, and much remains to be discovered.

In Paper A we introduce two purely heuristic approaches to reducing uncertainty for model improvement while simultaneously minimizing the nominal output error, both developed for autoregressive systems with exogenous input (A.R.X.). We present several candidates for scalar costs of uncertainty and choose two from which we develop practical algorithms for suboptimal dual control. One minimizes a function of the parameter-estimate error-covariance matrix P; the other maximizes a function of the information matrix $R = P^{-1}$ with diminishing marginal return. While the heuristic approaches are suboptimal in nature, they work well and are capable of improving performance compared with a passive-adaptive M.P.C. We apply the second algorithm to the problem of models inadmissible for control and derive a result for guaranteed model identification in minimal time for a two-parameter system. Paper A is an extension of two previous conference papers, and the contribution is primarily of practical interest.

I consider the approach we develop in Paper B an extension of the methods from Paper A and a step toward an exact and more formal approach to dual control. This paper also focuses on A.R.X. systems, but rather than continuing with heuristic objective functions, we develop an exact reformulation for a one-step-ahead cost similar to $(*)$. Applying this reformulation to the entire prediction horizon and adding the necessary constraints for uncertainty propagation results in a semi-exact dual M.P.C.

Paper C contains a more systematic approach to the dual-control problem. We identify and define the subset of the future information that is necessary for exact reformulation of $(*)$ for systems modeled with orthonormal basis functions (O.B.F.s). This subset is the sequence of future decisions (control inputs) in the finite-horizon O.C.P., which we further use for exact reformulation of probabilistic output constraints. We then demonstrate that the deterministic equivalent to the stochastic finite-horizon dual-control problem (with the objective $(*)$ conditional

on future inputs) can be formulated as a quadratically-constrained quadratic program (Q.C.Q.P.), for which there exist efficient global solvers. While this approach does not recover the dynamic-programming solution, Paper C introduces what I argue fully qualifies as (adaptive) dual, model-predictive control (D.M.P.C.). Paper C is an extension of a previous conference paper. We consider the results as having a significant theoretical contribution in addition to their practical importance.

ACKNOWLEDGMENTS

First and foremost I would like to thank Bjarne, my Ph.D. advisor and supervisor. Thank you, Bjarne, for believing in me, for taking me on as your student, for your support and guidance, for all the freedom and trust, and for facilitating so much of my growth both as a researcher and as a person.

A large portion of this work was completed at Carnegie Mellon University in Pittsburgh, Pennsylvania, and I would like to thank my advisor there, Erik Ydstie, without whom none of this would have been possible. Thank you, Erik, for your guidance, advice, patience, dedication, and encouragement, for always cheering me up and for being a great mentor and friend.

I would like to acknowledge financial support from N.T.N.U., the Center for Integrated Operations in the Petroleum Industry at N.T.N.U., and the Peder Sather Center for Advanced Study at the University of California, Berkeley.

I am also very grateful to my Ph.D. thesis committee and opponents: Professor Hans Georg Bock, Dr. Magne Fjeld, and Professor Morten Hovd; it is an honor to be evaluated by this distinguished group of researchers.

I am fortunate to have so many wonderful friends and colleagues at N.T.N.U. Thank you, Brage, for being there through all of it, for ushering me on, for invaluable research discussions, and for everything else. Anne Mai, Anders, Joakim, and Bjarne, thank you for all the great times together, both at and outside of work. Eleni, Mansi, and Serge, thank you for many good discussions about work and everything else. You have all contributed greatly to my time at N.T.N.U.

To all my friends and fellow students at C.M.U., thank you for the welcoming atmosphere and for your part in maintaining an amazing place for work and inspiration. Thank you, Michael, for setting so many things in motion for me; Satya, for interesting discussions about everything; German, for terrific company during late nights at the office; Pablo, for housing and for great times; Krishna, Qi, Francisco, Linlin, Yash, Sree, Juan, Sumit, Edna, Clara and Nick; thank you all for the many good Pittsburgh memories.

I have had the good fortune to collaborate with several researchers from whom I have learned a lot. I would like to thank Sachin Patwardhan, Kunal Kumar, Stefan Streif, Ali Mesbah, Vinay Bavdekar, and Andrew Conn for their time and efforts.

I owe my high-school mathematics teacher Hans Kolsrud a big thank you. Your inspiring command of your subject and approach to teaching were vital to my growing interest in math and science. I frequently send you thankful thoughts.

I am forever grateful to my parents, Lisbeth and Aksel, for their tireless support, unconditional love, inspiring work ethics and resilience, persistent encouragement, and for being exceptional role models.

Emily, thank you for tenaciously challenging my reasoning, inspiring me to clarify my thoughts and articulate them better, improving my English, being a steady voice of reason, helping me sharpen my focus, enduring my various degrees of sanity, the strength you bring out in me, and most of all your love and support.

Finally, thank you, Lise, for being an unfading source of love, inspiration, courage, comfort, and strength; for reminding me I'm loved and helping me persevere; and for always being there for me—in my thoughts and in my memories. Your absence made this more difficult, but I could not have done it without you.

CONTENTS

ACRONYMS AND INITIALISMS

A.D.P.	approximate dynamic programming
A.M.P.C.	adaptive model-predictive control
ARMAX	autoregressive moving average with exogenous input
A.R.X.	auto-regressive with exogenous input
C.E.	certainty equivalence
D.M.P.C.	dual model-predictive control
E.L.S.	extended least squares
F.I.R.	finite impulse response
L.Q.R.	linear-quadratic Gaussian
MIMO	multiple input, multiple output
MISO	multiple input, single output
M.P.C.	model-predictive control
N.L.P.	nonlinear programming
O.B.F.	orthonormal basis function
O.C.P.	optimal-control problem
Q.C.Q.P.	quadratically-constrained quadratic programming
Q.P.	quadratic programming
R.L.S.	recursive least squares
SISO	single input, single output

Part I

INTRODUCTION AND BACKGROUND

This part contains a brief introduction with some intuitive examples for motivation, a chapter on various aspects and manifestations of uncertainty, and a chapter with a fairly general treatment of dual control. Finally, there is a chapter that provides an overview of this thesis, including a list of publications.

INTRODUCTION

1

A MANAGER IN A company has several people working for her, none of which she knows well enough to have a complete picture of their abilities. If the manager only assigns tasks based on her current impression of her staff, she risks losing out by not taking full advantage of their skills. She may learn new things about their capabilities over time, but that is a potentially slow process. She may learn faster, while still getting things done, by challenging the people working for her and assigning work they may or may not be able to complete in a satisfactory manner. This type of experimentation can expedite the learning process and give the manager a much better idea of what the people in her team are capable of, increasing the team's performance and quality of work.

A central bank is faced with the decision of whether to increase, decrease, or leave untouched the national interest rate. Aware of the predictive limitations in their economic models, they know that the outcome of any decision is uncertain. However, a small adjustment for the purpose of learning from the resulting response may provide additional insight into the current workings of the economy and thereby contribute to an improved foundation from which they can base future decisions. Expecting a response that differs from what they predict based on their current knowledge, they make the small adjustment while simultaneously trying to direct the economy in the right direction and minimize any potential downside to their interference. The fact that they anticipate learning from future information and sometimes act accordingly improves their overall performance as governors of the economy.

The citizens of a small town depend on income from industry that pollutes their nearby beautiful lake. Most people in the town want economic growth, but are aware that lakes can exhibit threshold behavior, in which the increased pollution from higher economic activity may suddenly turn a healthy lake unhealthy and murky. Not knowing the details of this threshold response, the inhabitants must weigh economic benefit against an uncertain adverse effect that is difficult quantify. Careful monitoring of the pollution and the relevant health metrics for the lake can provide data from which the citizens can deter-

mine the highest economic activity level the lake can withstand without turning unhealthy. Experimenting with the pollution levels to locate the threshold does carry risk, as returning the lake to a healthy state after crossing the threshold may be expensive. This decision-making problem, analogous to many in climate policy, involves issues of utility maximization, experimenting and learning for uncertainty reduction, modeling disagreement, risk, and possibly irreversibility.

Arctic methane release is an example of a climate phenomenon that exhibits a threshold response.

A driver in a rental car on slippery roads is in a hurry to his destination. The driver is inexperienced and he does not know whether the car has an anti-lock braking system (A.B.S.), how good the brakes are, and how slippery the roads are. A turn is coming up and the driver faces the question of how to best learn about the car and the road conditions without losing too much time, so that he can make the turn fast but responsibly and without excessive caution. Carefully experimenting with small turns and testing the brakes can provoke responses that are valuable to the driver, improving his mental model of the circumstances and in turn his ability to safely and quickly maneuver the vehicle.

These scenarios are simple examples of dual control tasks: problems where the original goal (some form of utilization or exploitation) can best be reached by accounting for the lack of complete knowledge and devising an ideal strategy to reduce the uncertainty (by learning about the system's response to the decision-maker's actions) with the aim of making better decisions from the improved model.

In the driving example, we can assume that there are three alternatives available to the driver: acceleration, gentle braking, and hard braking, each with clear consequences. Accelerating can provide information about how slippery the road is, but not about the type of braking system nor its effectiveness. Gentle braking provides some idea of the responsiveness of the brake pedal but may be insufficient to teach the driver about the road friction or A.B.S., if present. Harder braking that causes the car to lose traction provides a lot of information on the slipperiness of the road and makes it apparent whether the car is equipped with an A.B.S. Failure to learn the characteristics of the car and road can lead the driver to turn over-cautiously and lose valuable time, or not cautiously enough and lead to a dangerous situation.

Neglecting the relationship between the two goals of the optimal decision sequence, (*i*) reaching the destination quickly and safely and (*ii*) learning about the road conditions and car handling, leads to a decision-making approach where the goals can be mistakenly identified as separate and conflicting. Braking for the purpose of learning

means reaching the destination later if one does not account for the future value of the learning outcome (the potential to drive faster after learning the brakes perform well); the best strategy is to drive as fast as achievable if one ignores the limited knowledge about the car and the conditions. A driver will not reach his destination faster by ignoring his lack of familiarity with the car and conditions if the brakes are poor, the tires are worn, and the roads are icy. Hence, it is not meaningful to think of a trade-off between reducing ignorance by testing the brakes and making decisions based on a faulty mental model. The optimal sequence of decisions accounts for the future value of learning, or how current and future data (the decisions and the observed consequences) resolve uncertainty.

In addition to clarifying the dual aspects of an optimal decision sequence, (driving fast and learning about the braking system and road conditions) the simple example above highlights an important difference between two kinds of uncertainty: parametric and structural. The relationship between how hard the driver presses the brake pedal and how quickly the car decelerates under normal circumstances can be described with a parameter, where a large value can indicate significant deceleration from a light press on the pedal while a small value can indicate that pressing the pedal hard is necessary to achieve the same result in braking. The structural uncertainty comes from not knowing whether the car is equipped with an A.B.S. Since an A.B.S. prevents the wheels from locking when the tires lose traction, the braking behavior of an A.B.S.-equipped car is fundamentally different from one without an A.B.S. when the tires lose traction. This difference cannot be captured by different parameter values; rather, two models with different structures are required to capture the two kinds of behavior.

The characteristics of this problem change entirely if the roads are sufficiently dry and free of gravel. The friction between the tires and road can still be determined by braking very hard and there are still things to learn about the effectiveness of the brakes, but most drivers would consider this uncertainty fairly minor and deem experimentation unnecessary; the uncertainty is not relevant for the driving decisions and taking active steps for learning is no longer the best course of action. With sufficiently good road conditions it might be impossible to actively cause a loss of traction that initiates the A.B.S. That is, the presence of an A.B.S. is not only irrelevant for the driving strategy, it is also impossible to determine by experimenting with the car.

A challenge in model development for sequential decision-making and automatic control is the difficulty of obtaining and formulating accurate and precise knowledge about the system. The lack of exact knowledge may manifest itself in uncertain model parameters or in the choice of appropriate model structure. Parameter uncertainty is a common way of describing a lack of "true" parameter values, often in the form of assuming statistical distributions for the parameters. In some cases, the greater modeling challenge is to formulate and represent the lack of knowledge.

The ultimate challenge is to find the sequence of decisions for optimal performance. Finding that sequence involves accounting for uncertainty while anticipating and endogenizing learning. This challenge, known as the dual control problem, is the topic of this thesis.

1.1 THE ORGANIZATION OF THIS THESIS

This chapter presented a handful of intuitive examples that highlight some of the defining characteristics of dual-control problems. In the following two chapters I discuss some fundamental aspects of uncertainty, uncertainty in the context of optimal control and decision-making, the dual control problem in more detail, and some general solution strategies. The fourth chapter contains an outline of the research reported in this thesis, followed by a series of chapters that each contain one research article. I draw some conclusions in the following chapter and share some thoughts on directions for future research.

UNCERTAINTY

2

The topic of this thesis is optimal control and decision-making under uncertainty. Uncertainty is a wide term applied to several distinct phenomena and the following brief discussion, primarily based on an article by Smith and Stern (2011), is intended to review some concepts and different forms of uncertainty and clarify their relevance to engineering contexts.

When discussing uncertainty and estimates, the terms *precise* and *accurate* are sometimes used interchangeably in the literature. This is unfortunate, as the distinction is frequently important. A precise estimate is narrow, but not necessarily true; an accurate statement is true, but not necessarily narrow. As an example, the statement "one liter of water weighs ten to twelve grams" is precise but not accurate. Similarly, "one liter of water weighs between ten grams and ten kilograms" is accurate but not precise. *Imprecise*, a term I define below, is not always an antonym for *precise*.

Smith and Stern (2011) write with great clarity on the nature of uncertainty, its various forms and manifestations, and how these forms affect decision-making differently. Their discussion is primarily focused on the use of models to generate real-world probabilities, but most of the subject matter is more general and widely applicable to other contexts. The following list of terms is adapted from their paper.

- *Imprecision* (Knightian risk, statistical uncertainty): related to values we do not or cannot know precisely, but for which we can make reliable probability statements. This is what is generally referred to as *uncertainty* in the engineering literature.
- *Ambiguity* (Knightian uncertainty): related to values where we cannot make probability statements. Ambiguity is sometimes used to describe uncertainty in a probability statement, for instance uncertainty in the statistical parameters describing a distribution for a model parameter or in which type of distribution is most appropriate. In this case, the term *second-order uncertainty* is frequently used, and sometimes the more ambiguous term *deep uncertainty*.

- *Intractability*: related to relevant computations that we are unable to formulate or perform for reasons such as limitations on the mathematical or computational capacity. Intractability may render ambiguity reduction impossible.
- *Indeterminacy*: related to decision-relevant quantities to which we may assign values, even though no true values exist. This situation can arise with model parameters that have no correspondence to real physical quantities.

The most familiar among these varieties of uncertainty in the engineering literature is *imprecision*. Imprecision is often invoked through the common approach of accounting for parametric uncertainty by assuming distributions for the unknown parameters, with normal and uniform among the most common types. Since the normal distribution is a stable distribution, there exists a large number of results that cannot be obtained for other common distributions. This property, combined with its ability to capture a large range of phenomena acceptably well, makes the distribution an attractive choice in many situations, including cases where its application implies there is a nonzero probability of events we recognize as impossible, for instance a negative mass. Similarly, the uniform distribution is often applied for convenience, despite evidence of a non-uniform probability profile or of there being small, nonzero probabilities of events outside the support.

A deterministic model with wrong parameter values has limited value for predicting the consequences of decisions. Models that account for uncertainty with probability distributions for unknown parameters are generally able to predict a distribution of outcomes, informing the decision-maker that there is a range of possible consequences, some more likely than others. Specifying a distribution for an unknown parameter that appropriately characterizes the uncertainty is typically difficult and is essential to the predictive capabilities of the model. Assuming a distribution that misrepresents the parameter uncertainty is a modeling error that may result in a model that makes less decision-relevant predictions compared with a wrong deterministic model. When we cannot quantify uncertainty using probabilities, or do not know how to formulate the probability statements, we are faced with *ambiguity*. This is a frequent occurrence when using data-driven models, such as the auto-regressive moving-average (ARMA) structure, combined with parameter estimation. These models often contain parameters that have no real-world counterpart (which then involves indeterminacy), and

there are rarely sufficient data available for making informed assumptions on their distributions. Even with a rich data set, it may be prohibitively difficult to characterize the uncertainty in a way that somehow improves the model; in fact, the notion of distributions for parameters in data-driven models is meaningless in many cases.

Ambiguity is also related to the inability to predict the details of an outcome that is virtually inevitable. As an example, a battery will almost certainly break if a high-enough voltage is applied over a sufficient length of time, but predicting the details of the failure may be beyond the capabilities of the best models available. One may be able to establish a probability based on experience and analysis of data, but this probability will itself be uncertain. Similarly, there are no models capable of accurately and precisely predicting the effect charging and usage patterns have on the long-term health of batteries, despite virtual certainty that the capacity will decay with time and use. Optimistic and pessimistic estimates of the future energy capacity may be predicted, but such estimates are themselves uncertain. Models can provide decision-relevant probability distributions for phenomena they simulate realistically and thus reduce ambiguity. In situations where such models run too slow to provide information in a timely manner, when we have no reliable way of solving the equations numerically, or when the relevant equations are not known (as was the case with the orbit of Mercury before Einstein's theory of relativity, noted by Smith and Stern), ambiguity cannot be reduced because of *intractability*.

Many multi-stage decision problems can be formulated compactly with a Bellman equation. However, we can only rarely derive a non-approximate decision-rule from the Bellman equation (see Section 3.3), and numerical solution procedures are usually impractical as they in general suffer from exponential complexity growth. The computations required to answer a question like "given an optimal decision sequence, what is the probability of an event A?" may be intractable, and we might have to resort to approximate solution methods that result in unreliable probability estimates.

An objective in sequential decision processes is maximizing discounted future utility (see Bond and Loomis (2009) for an example I discuss below). Rather than the specific value being unknown, a single correct value of a discount rate does not exist, and the "best value" may vary as a personal opinion. Furthermore, discount rates often have no empirical meaning or counterpart in the real world, and it can be impossible to state anything more definitive than that the discount

rate be strictly between zero and one. The model studied by Bond and Loomis (the source of the lake example in the introduction chapter, also considered by Lempert and Collins (2007)) involves finding the best trade-off between nominal economic development and the resulting environmental damage. Economic damage is then quantified with a cost, the magnitude of which is highly debatable in any real decision-making process of this type. The practice of attaching costs and prices to things that have no inherent monetary value is somewhat controversial; the quantities will differ among experts and non-experts alike; may be a matter of political views, personal preferences, and other subjective factors; and have no objective true value. When an optimal policy is highly sensitive to this type of costs or to the discount rate (Groeneveld, Springborn, and Costello, 2014), the premise of the optimal-control problem should be examined, as well as the extent to which it is ill posed or even meaningful. In such cases, indeterminacy refers not only to the parameter, but to the decision problem in which it is included.

Smith and Stern (2011) discuss the fallacy of confusing model error with imprecision, which is of particular relevance to the topic of this thesis. As an example, they point out how a model based on Newton's laws cannot provide accurate predictions of Mercury's position, and how an ensemble of such models cannot quantify the imprecision of the prediction with any accuracy. This is an illustrative case of how the inadequate nature of the model, not imprecision, leads to inaccurate conclusions.

A related fallacy, also discussed by Smith and Stern, concerns overconfidence in a given model's ability to inform decision-makers, and more specifically the fallacy of mistakenly identifying model-based entities with their real-world counterparts, which they refer to as "Whitehead's fallacy of misplaced correctness." One engineering area where this is of relevance is petroleum-reservoir simulation, where a common prediction approach is to use an ensemble of parameter sets for one model, and then infer probability distributions for various outcomes.

Most control engineers recognize that even high-fidelity models cannot accurately predict system behavior when projecting sufficiently far into the future. The most common sources of uncertainty when modeling for control include imprecision in initial conditions, measurement noise (which may be the source of the imprecision in the initial conditions), exogenous disturbances (process noise), endogenous parametric uncertainty, and endogenous structural uncertainty. These phenomena are all reasons to encourage feedback-based methods over open-loop

control, since feedback to some extent corrects for the difference between model output and system behavior. Hence, control approaches based on open-loop predictions often incorporate feedback through recalculating the predicted behavior at regular intervals, often called open-loop feedback control.

2.1 MODEL SETS AND TRUE MODELS

Gevers (2006), in his overview of the historical developments in system identification, touches upon an important change of focus that occurred around 1976. Up until that point, nearly all activity in the field focused on identifying the *true system* and ensuring convergence to the *true parameters*. However, the attention gradually shifted after the realization that a more intellectually honest framework would have to treat "system identification as an approximation theory," meaning the goal is to find "the best approximation of the true system" within a given model class. This realization is of great importance for a thorough understanding of the dual control problem, which is the problem of optimal adaptive control under reducible, decision-relevant uncertainty. That there are no true parameters and that every model is an approximation are central facts to dual-control theory. Furthermore, realizing that the "true" optimal control is not obtainable without omnipotent system knowledge makes clear that the distance to this true optimal control is determined by the model's veracity and the ability of the algorithm to utilize all the knowledge and uncertainty representation embedded in the model.

Optimal control based on an uncertain model that misrepresents the lack of knowledge, for instance through assuming an inappropriate model structure or set, is arguably not optimal in any meaningful sense of the word, despite qualifying in the traditional mathematical sense. It is arguably more difficult to accurately model uncertainty in a non-deterministic system than it is to comprehensively model a complex deterministic system. Formulating an uncertainty model that is computationally tractable and decision relevant while at the same time provides a realistic representation of the system is even more challenging. This, combined with difficulties in algorithm development, makes optimal, adaptive control under uncertainty a formidable challenge.

2.2 STRUCTURAL UNCERTAINTY

While the control-engineering literature has traditionally focused on modeling uncertainty with unknown parameters and associated distributions, there are types of uncertain phenomena for which this modeling approach is not suitable. An important example is uncertainty regarding the appropriate model structure for some process. This happens, for instance, when one out of several possible chemical reactions may be occurring; the uncertainty here is better captured by a set of model hypotheses, where each hypothesis has its own structure or functional form. This type of uncertainty also appears in Fault Detection (Blanke et al., 2006), where potential faults are modeled with entirely different equations rather than variations in parameters.

2.3 UNCERTAINTY REDUCTION

The standard non-dual approaches to data-based uncertainty reduction in control engineering can be divided into two classes: dedicated experiments and passive learning. Dedicated experiments are generally performed separately from normal operation, and the experiments are often developed according to principles from the field of Optimal Experiment Design (Gevers, Bombois, et al., 2011). A successful experiment generates data that aid the modeling effort and lets the modeler determine, with a high degree of certainty, a model structure with appropriate parameter values. The resulting model is subsequently used for decision-making or automatic control. An alternative to this is often referred to as "passive learning," and can be thought of as learning occurring as a byproduct of normal control operation, in the control community known as (non-active) Adaptive Control (Åström and Wittenmark, 1995).

There is a variety of now standard techniques to determine the parameter values that best fit a given model structure. These techniques have been developed over the last few decades in the System Identification and Parameter Estimation literature and have reached a high degree of maturity; see (Ljung, 1999). The task of determining the most likely among a set of candidate model hypotheses is not widely studied and does not enjoy the same stage of maturity. Common approaches are based on Bayesian statistics, but the control engineering community has not adopted a standard set of techniques for sequential model-

hypothesis testing. The research in this thesis considers parametric uncertainty only; structural uncertainty is beyond its scope.

DUAL CONTROL

3

3.1 DUAL EFFECT

A dual control signal is often considered to be a signal that both excites the system for uncertainty reduction and attempts to achieve desired plant behavior. Desired plant behavior usually involves keeping the states (or outputs) close to a constant or time-varying setpoint, in addition to satisfying the specified constraints. Before discussing the more nuanced aspects of dual control, I find it useful to first investigate the term *dual effect*.

The following definition, after Bar-Shalom and Tse (1974), relates dual effect to the central moments of the unknown quantities in a problem.

DEFINITION 1 (dual effect). *A control with* dual effect *is a signal that with nonzero probability can affect at least one rth-order central moment of a state, with* $r \geq 2$. *Conversely, if the future uncertainty is unaffected by the control with probability one, i.e., there are no central moments of order* $r \geq 2$ *of any state that are affected by the input signal, the control has no dual effect.*

Feldbaum (1961b) used *neutral system* to describe a system with no dual effect. Note that while Bar-Shalom and Tse define dual effect as a property of the control signal, the presence of a dual effect is a system feature, not determined by the decision maker or control algorithm. I use the term *dual effect* in reference to systems in this thesis.

A simple (although somewhat pathological) example of a system with no dual effect is

$$x(t+1) = a(t)x(t) + bu(t) \tag{3.1}$$

where $x(t)$ and $u(t)$ are the state and control at time t, respectively, b is a known constant, and $a(t) \sim \mathcal{U}[a_{\min}, a_{\max}]$ is at every time t drawn from a uniform distribution with bounds at, say, $a_{\min} = 0.2$ and $a_{\max} = 0.8$. Since $a(t)$ takes a new random value at every time t, the control has no way of affecting the uncertainty.

A similar and more common effect is that of a constant, unknown disturbance, also called a bias. Consider the system

$$x(t) = ax(t-1) + bu(t-1) + d \tag{3.2}$$

The conditional distribution of d, given the signals recorded up to and including time t, is Gaussian with mean $\hat{d}(t)$ and variance $P(t)$ that satisfy the equation set

$$\hat{d}(t) = \hat{d}(t-1) + K(t)(x(t) - ax(t-1) - bu(t-1) - \hat{d}(t-1)) \tag{3.3a}$$
$$K(t) = P(t-1)\big(1 + P(t-1)\big)^{-1} \tag{3.3b}$$
$$P(t) = \big(1 - K(t)\big)P(t-1) \tag{3.3c}$$

with initial conditions $\hat{d}(t_0) = \hat{d}_0$ and $P(t_0) = P_0$; see Åström and Wittenmark (1995). The covariance equation can be simplified to

$$P(t) = P(t-1) - P^2(t-1)\big/\big(1 - P(t-1)\big) \tag{3.4}$$

which makes it clear that the second central moment is not affected by the control. That is, there is no dual effect, and the control cannot be used to improve the estimate of d. I find it important to emphasize that the absence of dual effect does not imply that learning is impossible, only that the learning is independently of the control.

It is worthwhile to note that in certain problems one can clearly determine when the input signal does and does not affect the uncertainty. The following example, adapted from Bond and Loomis (2009), helps illustrate this phenomenon. Let $x(t)$ and $u(t)$ be the state and control at time t, respectively, let $a \in [-1, 1]$, b, and x' be known constants, and let $d \sim \mathcal{N}(\mu, \sigma^2)$ be an unknown constant with a normal distribution for which μ and σ are known. The system behaves according to

$$x(t+1) = \begin{cases} ax(t) + bu(t) & \text{if } x(t) < x' \\ ax(t) + bu(t) + d & \text{if } x(t) \geq x' \end{cases} \tag{3.5}$$

This formulation describes a dynamic system where a threshold effect (a disturbance) of unknown magnitude d is activated at a known region in the state space. This magnitude can not be determined as long as the state is below the threshold value x'; in order to reduce uncertainty in the bias magnitude d the control must take the state past the threshold. That is, as long as $x(t) < x'$ and $x(t+1) < x'$ it is clear that the control $u(t)$ can not affect the future uncertainty in d (central moments of order

two or higher). It is equally clear that this uncertainty is reduced whenever the state is above threshold unless the uncertainty is already at its minimum. In general, however, the control *can* affect future uncertainty, but it is up to the decision maker whether to try to bring the system into the region of the state space where this effect occurs. In a more complex system it may be nontrivial to determine the presence of a dual effect, for instance if it is unclear whether the system can be brought into a state where learning occurs.

It is not obvious in the above example model that there is any value in learning the value of d. Consider the case where $x(t_0) < x'$ and the performance measure is $J = \sum_{t=t_0}^{N} \mathrm{E}[(x(t) - x^*)^2]$ with the reference value $x^* < x'$ and N either finite or infinite. In this case it is optimal to stay below the threshold which means the control problem is deterministic and the value of information about d is zero. This leads to the notion of *decision-relevant*, or *control-relevant* uncertainty; uncertainty for which reduction offers no benefit in decision making is not relevant in that context. The decision relevance of uncertainty is primarily a property of the problem, more than of the system or the model: the uncertainty in d becomes relevant for control if the reference value in the above performance measure is $x^* > x'$. The following example contains uncertainty that is irrelevant for the control problem yet fully possible to reduce, and is a better illustration of how not all uncertainty is relevant for control and decision making. We wish to minimize

$$J = \sum_{t=t_0}^{N} \mathrm{E}[x^2(t)] \tag{3.6a}$$

subject to

$$x(t+1) = ax(t) + bu(t) \tag{3.6b}$$
$$x(t_0) = x_0 > 0 \tag{3.6c}$$
$$0 \leq u(t) \leq 1 \tag{3.6d}$$
$$a \sim \mathcal{U}[0.2, 0.8] \tag{3.6e}$$
$$b \sim \mathcal{U}[1, 2] \tag{3.6f}$$

The control task for this stable system ($|a| \leq 1$) is to reduce the state x from $x_0 > 0$ to zero. Since a is positive, the control u is non-negative, and the gain b is positive, the optimal control is a sequence of zeros. While it is trivial to excite the system for the purpose of reducing uncertainty and identifying the values of a and b, this reduction does not aid the decision maker or control algorithm in improving performance for this

particular formulation; there is no decision-relevant uncertainty in this problem.

In certain problems with decision-relevant uncertainty and dual effect it is not optimal, or even acceptable, to actively use the control for uncertainty reduction. Consider a modified version of the ecosystem model (3.5), with d a known constant but unknown, uniformly distributed threshold location $x' \sim \mathcal{U}[x'_{\min}, x'_{\max}]$ with known support $[x'_{\min}, x'_{\max}]$. Assume that the objective is to minimize an expected cost of the form $J = \sum_{t=t_0}^{N} \mathrm{E}[(x(t) - x^*)^2] + ru^2(t)$ where $r > 0$ is a chosen parameter and let $x^* > x'_{\min}$. If $x^* > x'$ and b, d, and r are sufficiently large, the optimal steady-state will be well below x^* since the cost associated with the effort of compensating for the threshold effect is higher than the cost of deviating from x^*. The decision maker is then faced with the question of whether or not to explore the system in order to learn the location of the threshold x'. One strategy may be to investigate if $x^* > x'$ and then make decisions based on the acquired information, realizing that the price of information may be high. However, the problem parameters may have values such that the optimal decision sequence is risk averse, in the sense that the expected benefit of learning the threshold location does not outweigh the expected cost associated with crossing the threshold. In this case, the optimal strategy does not utilize the dual effect. Crossing the threshold may be irreversible if the control is limited in magnitude. This effect further complicates the question of whether it is optimal to reduce the uncertainty.

3.2 THE DUAL CONTROL PROBLEM

Feldbaum (1961a,b,c,d, 1965) introduced and defined dual control (arguably somewhat ambiguously) as the optimal control for systems that exhibit dual effect. That is, the dual control is optimal for problems where the uncertainty can be actively reduced. Feldbaum was also the first to recognize the dual character or purpose of the optimal control for this problem: to direct the system states or outputs to their reference values and generate data that reduces uncertainty. However, this definition does not explicitly account for situations where the uncertainty is not relevant for determining the optimal control. As I demonstrate above, it is not necessarily optimal, or even beneficial, to reduce uncertainty. While any actively reducible uncertainty can be made decision relevant by choice of control objective, the inclusion of decision relevance in the definition better connects the characteristics of the problem

with its solution. For this reason I argue that the following definition is clearer and more useful.

DEFINITION 2 (dual control). *A* dual control *signal is the optimal control for a system with dual effect where the uncertainty is decision relevant for a non-empty subset of the state and parameter space.*

Note that this definition does not imply that the solution to a dual control problem necessarily affect any central moments of order two or higher. As noted above, there are several situations where the optimal course of action is to not use the control for learning, such as when the moments are affected only by the control in a purposefully unexplored region of the state space, or when the uncertainty is so low that the cost of reduction is higher than the benefit of more knowledge. As a consequence, the definition permits solutions where the control reduces uncertainty as a side effect, as opposed to actively exploring for the purpose of affecting the moments. In this case, the control is insensitive to the uncertainty.

While this may appear somewhat paradoxical, it is not practical to define the problem by the features of its solution, and it may be impossible to determine whether learning is beneficial or active uncertainty reduction is optimal prior to solving the dual control problem. Defining a dual control as the solution to a dual control problem then allows dual controls that do not actively excite the process. Accordingly, a control that is designed for improved learning applied to a system with dual effect is not necessarily a dual control, since a deliberately exciting component in no way implies that the control is optimal. As an example, a control signal $u(t)$ generated by superimposing a random signal $w(t)$ on an input generated by a proportional controller,

$$u(t) = K_p e(t) + w(t) \tag{3.7}$$

where $e(t) = y(t) - y^*(t)$ is the output error, is not a dual controller, unless it happens to be the optimal control for a given problem. However, this control signal does actively excite the process and the added term $w(t)$ may increase the information generated in the loop.

This argument might be clearer by exploring the alternative. Suppose for the sake of argument that a definition of dual control is any control that actively excites the system for improved adaptation or identification. The question is then what an *optimal* dual controller is. As we have seen, the presence of dual effect does not imply that it is beneficial to actively reduce uncertainty. Accepting this terminology then entails

that an optimal dual controller may not be a dual controller. This leads down a path of confusing nomenclature, and it is clear that a much more informative set of definitions allows dual control to describe a problem of optimal control in the presence of dual effect. This means a dual control problem must be solved with the learning mechanism included in the model, and that the solution may not involve active uncertainty reduction. Hence, a dual control problem must be solved a certain way (accounting for the dual effect), but the solution does not necessarily include active learning (the dual effect is not necessarily exploited).

The following optimal control problem is a fairly general mathematical formulation of the dual control problem in discrete time: find the control sequence $\{u(k)\}_{k=t}^{t+N-1}$ that is the solution to

$$\min_{\{u(k)\}_{k=t}^{t+N-1}} \mathrm{E}\left[\sum_{k=t}^{t+N-1} \ell(y(k+1), u(k)) \;\middle|\; \mathcal{Y}(t+N-1 \mid t)\right] \tag{3.8a}$$

subject to

$$x(k+1) = f(x(k), u(k), \theta(k), v(k)) \tag{3.8b}$$

$$\theta(k+1) = g(\theta(k), e(k)) \tag{3.8c}$$

$$y(k) = h(x(k), w(k)) \tag{3.8d}$$

with given initial conditions and possibly bounds on some or all variables. Here, $\ell(\cdot)$ is the stage cost (a Lagrange term); $\mathrm{E}[\,\cdot \mid \mathcal{Y}(t+N-1 \mid t)]$ denotes expectation conditioned on all past (up to an including time t) and future (from $t+1$ to $t+N-1$) observations or data $\mathcal{Y}(t+N-1 \mid t)$; N is the control horizon, which may be finite or infinite; $x(t)$, $u(t)$, $y(t)$, and $\theta(t)$ are the states, control inputs, outputs, and parameters at time t, respectively; and $v(t)$, $e(t)$, and $w(t)$ are uncorrelated sequences random variables, where the variables in each sequence are independent and with known and identical probability distributions. I discuss the information set $\mathcal{Y}(\cdot)$ in more detail below. A typical example of a stage cost $\ell(\cdot)$ is

$$\ell(y(k+1), u(k)) = y^\top(k+1) Q y(k+1) + u^\top(k+1) R u(k+1) \tag{3.9}$$

where $Q \geq 0$ and $R > 0$ are symmetric matrices that specify performance priorities. Note that the dual control problem (3.8) includes no explicit notion of excitation or active uncertainty reduction, a point I will return to below.

3.3 INFORMATION AND DYNAMIC PROGRAMMING

In his original paper on dual control, Feldbaum recognized dynamic programming (Bellman, 1957) as an appropriate tool for analysis and solution. While dynamic programming can be used to derive explicit solutions for simple control problems like the linear-quadratic regulator (L.Q.R.) (see Kirk (1970) for a very clear derivation), most solution methods based on dynamic programming rely on numerical representations. However, the "curse of dimensionality" — the exponential growth in computational cost — associated with numerical dynamic programming renders the method ill-suited for all but the simplest problems. Despite its shortcomings as a method for obtaining numerical solutions, dual control problems are compactly and elegantly stated through a Bellman equation. Moreover, this formulation of the problem clarifies certain aspects of dual control and helps illustrate some of the more nuanced details.

Analytical evaluation of the conditional expected value of the stage cost $\ell(\cdot)$ is in most cases impossible. The expected stage cost is in general a function of the probability density (or mass) functions, sometimes referred to as *belief states*. Let the hyperstate $\xi(t)$ be a variable that contains all states, inputs, and moments that characterize the system at time t. Note that the hyperstate is not necessarily finite-dimensional.

In order to simplify the following exposition I omit a terminal cost (a Mayer term) from the cost function and assume full noise-free state feedback, so that $y(t) = x(t)$. Let

$$J = \mathrm{E}\left[\sum_{k=t}^{t+N-1} \ell(y(k+1), u(k)) \;\middle|\; \mathcal{Y}(t+N-1 \mid t) \right] \tag{3.10}$$

be the cost function and

$$V(\xi(t'), t'; t) = \min_{\{u(k)\}_{k=t'}^{t+N-1}} \mathrm{E}\left[\sum_{k=t'}^{t+N-1} \ell(y(k+1), u(k)) \;\middle|\; \mathcal{Y}(t+N-1, t' \mid t) \right] \tag{3.11}$$

be the minimal (feasible) expected cost, commonly referred to as the *value function* or the *cost to go* (from time $t' \geq t$ to time $t+N$). Splitting up the sum on the right-hand side gives

$$V(\xi(t'),t';t) = \min_{\{u(k)\}_{k=t'}^{t+N-1}} \mathrm{E}\Bigg[\ell(y(t'+1),u(t')) + \sum_{k=t'+1}^{t+N-1} \ell(y(k+1),u(k)) \;\Bigg|\; \mathcal{Y}(t+N-1,t' \mid t)\Bigg] \quad (3.12)$$

which through the principle of optimality (Bellman, 1957) allows the recursive formulation

$$V(\xi(t'),t';t) = \min_{\{u(k)\}_{k=t'}^{t+N-1}} \mathrm{E}\Big[\ell(y(t'+1),u(t')) + V(\xi(t'+1),t'+1;t) \;\Big|\; \mathcal{Y}(t+N-1,t' \mid t)\Big] \quad (3.13)$$

The recursive equation (3.13) is the *Bellman equation* for the dual control problem. While simple and compact, the recursive relationship is in general very computationally expensive to resolve. The dual control, or the solution to (3.13), takes the form of an optimal hyperstate-feedback control and can be determined through the computed value functions as

$$u^*(\xi(t),t) = \arg\min_{u(t)} \mathrm{E}\big[\ell(y(t+1),u(t)) + V(\xi(t+1),t+1;t) \;\big|\; \mathcal{Y}(t+N-1 \mid t)\big] \quad (3.14)$$

Numeric solution of (3.13) involves quantizing the hyperstate space, i.e., the states, controls, and moments, into a grid and iterating over all values. That is, a continuous space is approximated by a discrete space. Dynamic programming guarantees a globally optimal control *on the discrete approximation of the space*. While this is arguably a disadvantage of the method, the global solution is obtained without any assumptions on the functions (f, g, h, and ℓ) defining the problem.

The recursion (3.13) is started at the final time $t' = t+N-1$, where the cost to go depends on the final stage cost only. A full step of the recursion, calculating the right-hand-side of (3.13) for some stage t', means evaluating the expected value for all combinations of the quantized hyperstate and control values. From the perspective of evaluating the expected value, the actual information observed between time t_0 and

time t is indistinguishable from the hypothetical information assumed for the interval t to t'. That is, the synthetic information appears real for the purpose of finding the expected value of the cost to go from time t' to time N, but the synthetic information is fundamentally different in that it is an approximation that can take only values from the quantized space, as opposed to the continuous space of the actual information. The future information from time t' to time $t' + N$ is incorporated through solving the Bellman equation recursively backwards in time; the future information shares the quantization property since only discrete points in the future hyperstate space are directly considered by the previous recursion steps. At the final stage, the first stage solved by backwards recursion, then involves finding $V(\xi(t+N-1), t+N-1; t)$ for each quantized value of $\xi(t+N-1)$. This operation involves evaluating the expected value of the final stage cost given $\mathcal{Y}(t+N-1, t+N-1 \mid t)$. By the above argument, this is equivalent to the conditional expectation with respect to $\mathcal{Y}(t+N-1)$. Note that $\xi(t+N-1)$ depends on $\mathcal{Y}(t+N-1)$. Hence, the final-stage value function is

$$V(\xi(t+N-1), t+N-1; t) = \min_{u(t+N-1)} \mathrm{E}\Big[\ell(y(t+N), u(t+N-1)) \;\Big|\; \mathcal{Y}(t+N-1)\Big] \quad (3.15)$$

A recursion step is completed when every combination of quantized hyperstate values is explored, and the algorithm can advance to the preceding time stage. Here, that means evaluating the value function for every quantized value of the hyperstate at time $t' = t + N - 2$. The optimal control for each hyperstate depends on actual past information (from prior to time t), synthetic past information (from between time t and time $t' = t + N - 2$), and synthetic future information (from time $t + N - 1$). By the same argument as above, the information relevant for calculating the expected value of current state cost and the cost to go is directly dependent only on the information up to the current stage $t' = t + N - 2$. However, the optimal control is implicitly dependent on the future information used to determine the expected value at the

future stage $t+N-1$, which explains the notation in (3.10) and (3.11). The value function for the penultimate stage is then

$$V(\xi(t+N-2), t+N-2; t) = \min_{u(t+N-2)} \mathrm{E}\Big[\ell(y(t+N-1), u(t+N-2)) + V(\xi(t+N-1), t+N-1; t) \;\Big|\; \mathcal{Y}(t+N-2)\Big] \quad (3.16)$$

where the dependence on the future information $\mathcal{Y}(t+N-1)$ is implicit through $V(\xi(t+N-1), t+N-1; t)$. How the future information enters the recursion can be made clearer by expanding the recursive formulation. For a more compact representation, let $\ell_{t'} := \ell(y(t'+N), u(t'+N-1))$, $\mathcal{Y}_{t'} := \mathcal{Y}(t')$, and $u_{t'} := u(t')$, and write

$$V(\xi(t), t; t) = \min_{u_t} \mathrm{E}\big[\ell_t + \min_{u_{t+1}} \mathrm{E}\big[\ell_{t+1} + \cdots + \min_{u_{t+N-2}} \mathrm{E}\big[\ell_{t+N-2} + \min_{u_{t+N-1}} \mathrm{E}\big[\ell_{t+N-1} \;\big|\; \mathcal{Y}_{t+N-1}\big] \;\big|\; \mathcal{Y}_{t+N-2}\big] \cdots \;\big|\; \mathcal{Y}_{t+1}\big] \;\big|\; \mathcal{Y}_t\big] \quad (3.17)$$

It is clear from this formulation that at time t, $\mathcal{Y}(t)$ and $\mathcal{Y}(t')$ with $t' \geq t+1$ are conceptually different. The actual information $\mathcal{Y}(t)$ represents a single realization or trajectory through hyperstate space; the future information $\mathcal{Y}(t')$, $t' \geq t+1$, represents the set of all feasible candidates for optimal trajectories. The size and characteristics of this set may change, in that the admissible futures in $\mathcal{Y}(t'+k_1)$ at time t' may be different from (that is, a superset of) those admissible futures in $\mathcal{Y}(t'+k_2)$ at time $t'+k_3$, where $k_1 = k_2 + k_3$. Furthermore, any numerical algorithm for dynamic programming has to approximate the future information, often representing the elements with discrete set of values, as opposed the recorded information $\mathcal{Y}(t)$ with elements that typically take any real value (down to machine precision).

This explains the reason for the notation $\mathcal{Y}(t+N-1 \mid t)$ used in the cost function (3.10): the cost function implicitly depends on the future information over the horizon for which it is evaluated. Explicitly noting this dependence is important, and I argue below that the method for representing and approximating future synthetic information is paramount for how close a near-optimal dual control is to the true dual control that is obtainable with dynamic programming and a sufficiently accurate representation of the future.

Solving the Bellman equation involves exploring all paths that are potentially optimal; paths that cannot be optimal are not evaluated. The

cost to go from the next stage to the end of the horizon is a simple table lookup (with interpolation) when the system is deterministic. With a stochastic system where there is noise and unknown parameters, evaluating the expected value of the cost to go is in general a complicated operation since the next state of the process is uncertain; see Åström and Helmersson (1986) for one example, also discussed below.

The following example, taken from Åström and Helmersson (1986) and the first dual control problem solved with numeric dynamic programming reported in the literature, helps clarify some of the discussion above. Åström and Helmersson consider the simplest nontrivial process on a continuous state space: an integrator with unknown gain. Their system is written

$$y(t+1) = y(t) + bu(t) + \sigma e(t+1) \tag{3.18}$$

with $y(t)$ the state, $u(t)$ the control, $e(t)$ a sequence of standard normal variables, σ a known constant, and $b \sim \mathcal{N}(\hat{b}(0), P(0))$ an unknown constant with a known distribution. The authors define information at time t as the sequence of inputs and outputs observed up to and including time t:

$$\mathcal{Y}(t) = \{u(t-1), u(t-2), \ldots, u(0), y(t), y(t-1), \ldots, y(0)\} \tag{3.19}$$

The hyperstate in this problem is the state and the conditional distribution of the unknown parameter b given $\mathcal{Y}(t)$;

$$\xi(t) = \big(y(t),\ \hat{b}(t),\ P(t)\big) \tag{3.20}$$

where the conditional mean is

$$\hat{b}(t) = \mathrm{E}[b \mid \mathcal{Y}(t)] \tag{3.21a}$$

and the conditional covariance is

$$P(t) = \mathrm{E}[(b - \hat{b}(t))^2 \mid \mathcal{Y}(t)] \tag{3.21b}$$

The conditional distribution can be expressed recursively as

$$\hat{b}(t) = \hat{b}(t-1) + K(t)(y(t) - y(t-1) - \hat{b}(t-1)u(t-1)) \tag{3.22a}$$

$$K(t) = P(t-1)u(t-1)(\sigma^2 + P(t-1)u^2(t-1))^{-1} \tag{3.22b}$$

$$P(t) = (1 - K(t)u(t-1))P(t-1) \tag{3.22c}$$

The conditional distribution of the output $y(t)$ given $\mathcal{Y}(t-1)$ is Gaussian with

$$\hat{y}(t) = \hat{b}(t-1)u(t-1) \tag{3.23a}$$

and

$$\sigma_y^2(t) = P(t-1)u^2(t-1) + \sigma^2 \tag{3.23b}$$

the mean and variance, respectively.

Åström and Helmersson state the control problem as minimization of

$$J = \mathrm{E}\left[\sum_{k=t}^{t+N-1} y^2(k+1) \;\middle|\; \mathcal{Y}(t) \right] \tag{3.24}$$

Note that this objective function does not anticipate future information. In fact, the cost (3.24) can be interpreted as the squared *nominal* output error based on the currently available data; see Theorem 6 in Paper C. The value function (conditioned on $\mathcal{Y}(t)$ in Åström and Helmersson's paper) is

$$V(y(t'), \hat{b}(t'), P(t'), t'; t) = \min_{\{u(k)\}_{k=t'}^{t+N-1}} \mathrm{E}\left[\sum_{k=t'}^{t+N-1} y^2(k+1) \;\middle|\; \mathcal{Y}(t+N \mid t) \right] \tag{3.25}$$

with

$$V(y(t'), \hat{b}(t'), P(t'), t'; t) = \min_{\{u(k)\}_{k=t'}^{t+N-1}} \mathrm{E}\left[\sum_{k=t'}^{t+N-1} y^2(k+1) + V(y(t'+1), \hat{b}(t'+1), P(t'+1), t'; t) \;\middle|\; \mathcal{Y}(t+N \mid t) \right] \tag{3.26}$$

the corresponding Bellman equation. Evaluating the expected cost to go amounts to finding the average with respect to the distribution of $y(t+1)$ given $\mathcal{Y}(t+N \mid t)$. The conditional expected value is determined for each point in the gridded hyperstate space, and is therefore equivalent to considering that at any given hyperstate, the information that must have occurred in order for that realization of the hyperstate is real, as opposed to synthetic. The expected value of the cost to go, after Åström and Wittenmark (1995), is

$$\mathrm{E}\big[V(y(t'+1), \hat{b}(t'+1), P(t'+1), t'; t) \;\big|\; \mathcal{Y}(t') \big] = \int_{-\infty}^{\infty} V(\tilde{y}, \hat{b}(t'+1), P(t'+1), t'; t)\, w(\tilde{y}, t'+1)\, \mathrm{d}\tilde{y} \tag{3.27a}$$

where

$$w(\tilde{y}, t'+1) = \frac{1}{\sigma_y(t'+1)\sqrt{2\pi}} \exp \frac{-(\tilde{y} - \hat{y}(t'+1))^2}{2\sigma_y^2(t'+1)} \tag{3.27b}$$

and

$$\hat{b}(t'+1) = \hat{b}(t') + K(t'+1)(\tilde{y} - y(t') - \hat{b}(t')u(t')) \tag{3.27c}$$

$$K(t'+1) = P(t-1)u(t')(\sigma^2 + P(t')/\sigma_y^2(t'+1) \tag{3.27d}$$

$$P(t'+1) = (1 - K(t'+1)u(t'))P(t') \tag{3.27e}$$

This example illustrates the complexity of determining the cost to go, even for a simple dual control problem. It also clarifies how there is no explicit uncertainty reduction; the value of future information and the ability of the control to reduce future uncertainty is implicitly accounted for by modeling the conditional distribution of the states and parameters, the propagation of statistics forward in time made possible through backwards recursion over synthetic future information.

3.4 SOLVING DUAL CONTROL PROBLEMS

From the above discussion it is clear that while the dual control problem is elegantly stated and analyzed with dynamic programming, the framework is impractical for any problems but the very simplest. As noted, the backwards recursion iterates over an approximate characterization of future synthetic information, typically a discrete representation of a continuous space. The computational requirements limit the accuracy of the approximation since the complexity is exponential in the resolution. With infinite memory and an infinitely fast computer the true dual solution could be obtained with dynamic programming on hyperstate-space grid where the quantization distance approaches zero. The absence of such resources raises the question of how best to find a solution that approaches the unobtainable one, and how to design an algorithm that is able to reduce the distance with a reasonable increase in computational expense.

This section contains many terms and categories of controllers. I define the terms informally for a rough idea of how they relate to each other:

- An *actively adaptive* controller takes active steps to improve the adaptation or learning, for instance by adding an excitation signal to the control input.

- An *approximate* dual controller contains a mechanism for active adaptation or uncertainty reduction and computes the control by solving an optimal-control problem that is designed to approximate the dual control problem.
- A *heuristic* dual controller is an approximate dual controller in which the approximation of the dual-control problem is purely heuristic and not necessarily possible to derive from the original problem.
- A *semi-exact* dual controller is an approximate dual controller that computes the control from an optimal-control problem that approximates the dual-control problem through approaches such as reformulations or series expansions.
- An *exact* dual controller makes less significant approximations, which are qualitatively comparable to the quantization in standard numeric dynamic programming.
- An *ideal* dual control is the solution to the non-approximated dual control problem. No general technique exists for obtaining the ideal dual control.

Many early approaches to overcoming the the curse of dimensionality in dynamic programming focused on better quantizing; see, for instance, Larson (1965). A recent example of this type of focus is the algorithm based on adaptive sparse grids for parallel computing developed by Brumm et al. (2015). Approaches such as these are primarily concerned with lowering the cost of increased accuracy in the quantization through more efficient representations of synthetic future information.

Approximate dynamic programming (A.D.P.) is a broad term for a class of algorithms that in general move forward in time, rather than backward, and approximate the value function in the Bellman equation (Bertsekas, 2012). A common technique for approximating the value function is repeated simulation of the hyperstate forward in time, with a wealth of techniques that accelerate convergence. A typical A.D.P. algorithm explores only a subset of the hyperstate space, which in effect is similar to finding the expected cost to go with respect to a subset of the admissible future information. Thus, implicit approximation of future information is a side effect of generating sample paths for approximating the value function. Two notable approximate dual-control algorithms based on A.D.P. are developed by Lee and Lee (2009) and Bayard and Schumitzky (2010).

In his review of dual control, Unbehauen (2000) terms the above solution approaches to dual control problems *implicit*, referring to how

the uncertainty reduction is rewarded implicitly. He contrasts this to *explicit* methods that reward the uncertainty reduction explicitly in the cost function. I discuss some explicit methods below and argue that some of those classified as explicit by Unbehauen are better categorized as, e.g., methods for *actively adaptive* control (Tse, Bar-Shalom, and Meier, 1973).

An alternative to the two approaches above is to directly approximate, or consider only a subset of, the future information $\mathcal{Y}(t+N-1 \mid t)$ in the cost function (3.10). This is the approach my coauthors and I take in Paper C, where we pose the resulting control problem as a nonlinear programming (N.L.P.) problem. For the type of systems we consider there, our proposed subset of future information enables exact propagation of the conditional distribution of the unknown parameters as well as the variance of the output. We are also able to reformulate a probabilistic cost function of the form (3.10) into a deterministic function that includes an explicit uncertainty cost.

The dual control algorithm we develop in Paper B is based on applying the exact reformulation from Paper C to systems where the equivalence no longer holds. Since the optimal-control problem solved by this controller is derived from approximating the dual control problem through applying an exact reformulation beyond its region of validity, it can be classified as a semi-exact dual controller.

Unbehauen (2000) uses the term *explicit dual control* for algorithms that explicitly reward learning or uncertainty reduction. Definition 2 makes no distinction between dual controllers that implicitly and explicitly reward uncertainty reduction. Certain dual-control problems can be reformulated so that an implicit reward becomes explicit, which does not alter the controller in any way. However, heuristically adding a term that measures uncertainty to an existing cost function does not mean the minimizing control is dual. Minimizing the sum of a standard nominal control cost and an uncertainty-dependent function will in most cases not lead to optimal performance.

I argue these approaches are generally better classified as *approximate* dual control, *heuristic suboptimal* dual control, or *actively adaptive* control (Tse, Bar-Shalom, and Meier, 1973). (*Non-heuristic* suboptimal dual control implies an algorithm that finds a control that nearly optimizes performance, possibly with a measure of proximity to the solution.) This class of algorithms does not really approximate a specific aspect of the dual control problem; rather, it solves an alternative problem that exploits the dual effect with the aim of producing a control that is close

to the dual control. Wittenmark (1975) developed an early example of heuristic sub-optimal dual control for A.R.X. systems (autoregressive with exogenous input), or in his own terminology "an active suboptimal dual controller." He suggests adding the one-step-ahead variance of the first (non-delayed) input gain $P_b(t+1) := \mathrm{E}[(b - \hat{b}(t+1))^2 \mid \mathcal{Y}(t)]$, a deterministic function of the current input $u(t)$, to the output cost with the resulting cost function

$$J = \mathrm{E}[(y(t+1) - y^*)^2 \mid \mathcal{Y}(t)] + P_b(t+1) \tag{3.28}$$

where y^* is a reference. Actively reducing the uncertainty in the most important input gain is a well-justified goal, and the excitation generated for this purpose is likely to contribute to uncertainty reduction in other unknown parameters. However, while the controller may work well in a range of cases, there is nothing to suggest it leads to optimal, or even close to optimal, performance. Hence, the controller minimizing (3.28) is not a dual controller. This approach and the algorithms we develop in Paper A fall into the category of heuristic dual control. The controllers in Paper A are based on adding a term to a nominal cost for the purpose of rewarding uncertainty reduction.

A related approach, taken by for instance Rathouský and Havlena (2013), is to optimize for nominal control and excitation separately. Their algorithm first solves a nominal M.P.C. problem and then modifies the nominally optimal input to increase the resulting information content. Another class of controllers that are not dual, but utilize the dual effect, are based on guaranteeing a certain level of excitation. An early algorithm of this type is developed by Genceli and Nikolaou (1996) and guarantees excitation by adding constraints to a standard M.P.C. formulation. A similar approach is taken by Marafioti, Bitmead, and Hovd (2014), where the resulting optimal control problem can be formulated as a quadratic programming problem. These approaches are in this context best described as actively adaptive controllers.

3.5 EXPLORATION VS. EXPLOITATION: A FALSE DICHOTOMY

N-armed bandit problems are a good example of dual-control problems and are frequently, yet misleadingly, used with the catchphrase *exploration versus exploitation*. Åström and Wittenmark (1995) analyze a simple version of the two-armed bandit problem, which is sufficient for the argument I make here. In this problem, a player has N coins and must choose between two slot machines to maximize winnings. One

play on machine 1 or machine 2 gives a unit reward with probabilities p_1 and p_2, respectively. In the simplest instance of the problem, the player knows p_1 but only the distribution of p_2. For the sake of argument, let $p_1 = 0.6$ and let p_2 be uniformly distributed over the interval $[0, 1]$. Playing all coins on machine 1 gives an expected gain of 0.6 per play, while placing all bets on machine 2 gives the expected gain of 0.5 per play. The naïve best strategy, based purely on the prior, is then to spend all coins on machine 1. Since there is a 40 % chance that p_2 is larger than p_1, a smarter player will try to find a better strategy. With a very large number of coins, one could play machine 2 and estimate p_2, eventually obtaining a confident estimate of its value. The choice of which machine to play with the remaining coins is then easy based on the exact estimate of p_1 and the accurate and precise estimate of p_2. With a limited number of coins, the best strategy for maximal exploitation involves playing machine 2 to explore the chances of winning on that machine. The dynamic programming solution shows that the optimal strategy for maximal expected winnings is to play machine 2 until the estimate of p_2 is less than p_1, at which point the player should spend the remaining coins on machine 1. That is, the strategy that results in the best exploitation involves exploration. From this perspective, there is no trade-off or conflict between exploitation and exploration. However, the pervasive expression *exploration versus exploitation* gives the impression that exploration will reduce earnings, or that maximizing earnings excludes exploration. The only trade-off is between exploration and the strategy that ignores the potential for learning. Since the phrasing suggests that exploitation implies naïveté, I find this type of expression counterproductive and a missed opportunity to shed light on the very important realization that learning should not be ignored; sticking to a decision strategy based on the initial belief is rarely the best course of action when the situation permits learning through exploration. A better and more thought-provoking wording is *optimal exploitation may involve exploration*.

An analogous false dichotomy is widespread in adaptive control, from (Feldbaum, 1961b) to the present day (Åström and Kumar, 2014), which is that control and excitation (for learning) are in conflict ("... there is a conflict between the two sides of the controlling process, the investigational and the directional." —Feldbaum (1961b)) or that there is a trade-off between the two ("... the quintessential trade-off implied by the dual roles of control" —Åström and Kumar (2014)). Assume a dual control sequence that gives optimal performance, and that this

sequence contains components whose purpose is excitation, in addition to components that arise from traditional control objectives. The idea that one can reduce the learning component and achieve "better than optimal" performance is not wrong. In Paper C my coauthors and I derive a deterministic function from a dual objective similar to (3.8a) for a specific class of systems. We prove that for these systems,

$$\sum_{k=t}^{t+N-1} \left\{ \mathrm{E}\left[(y(k+1) - y^*(k+1 \mid t))^2 \middle| \mathcal{Y}(k \mid t) \right] \right\} = \sum_{k=t}^{t+N-1} \Big\{ (\hat{y}(k+1 \mid t) - y^*(k+1 \mid t))^2 + \varphi^\top(k+1 \mid t) P(k \mid t) \varphi(k+1 \mid t) + r \Big\} \quad (3.29)$$

where y is the output, P is the parameter-estimate-error covariance, φ is a regression vector of filtered inputs, and r is the constant variance of the Gaussian process noise. The first line in this equation is the expected squared output error conditioned on future information, the second line is the expected square *nominal* output error, and the third line is the output variance as a function of the parameter uncertainty as represented by P. From the premise that the left-hand side is a meaningful control objective, it is clear that introducing weights that allow prioritizing nominal control over uncertainty reduction, or vice versa, cannot improve the overall control performance. Uncertainty reduction is in conflict with neither overall expected control performance nor nominal control performance. This reformulation demonstrates that nominal control and uncertainty reduction are intrinsic parts of the expected performance; it is just as clear that these parts are equally important and cannot be traded off against each other. It is true that a decreased focus on nominal control can improve the learning, but a lower priority on uncertainty reduction does not improve expected performance. Furthermore, while the sum in the first line of Equation (3.29) is similar to the one in the second line, a lower nominal output cost in the second line does not imply improved performance as measured by the left-hand side. Improving precision in the terminology will clarify the sometimes nebulous intricacies and interesting details of the dual-control problem, and will ideally lead to more systematic and rigorous research on dual control.

4 THESIS OVERVIEW

The main contributions in this thesis are the three papers A, B, and C, each included as a separate chapter. Rather than in order of publication, the papers are listed in what I consider the most logical order, which is also the order in which the ideas were developed. Paper A is based on two earlier conference papers (Heirung, Ydstie, and Foss, 2012b, 2013a) and contains two related methods for approximate suboptimal dual control. Paper B is a natural step on the evolution from the heuristic methods in Paper A to the dual control algorithm in Paper C. The semi-exact dual controller in Paper B is a useful approach to dual-like control of A.R.X. systems (autoregressive with exogenous input), but the development reflects what at the time was my incomplete and somewhat immature understanding of the complexity of the problem we consider in the paper, in particular the propagation of uncertainty through a dynamic system and the nature of the associated future information. The dual controller in Paper C represents a big step toward a practical dual control algorithm, and contains several interesting results for orthonormal-basis-function (O.B.F.) models, and is a significant expansion and formalization of the ideas first presented in a conference paper (Heirung, Ydstie, and Foss, 2015c). The results also demonstrate several important nuances of the dual control problem. Paper D expands the ideas from Paper B to a multivariable armax system (autoregressive moving-average with exogenous input) and implements the resulting algorithm on an experimental lab setup.

The chapters that contain the original research are followed by a chapter with conclusions, some comments made in hindsight, and thoughts on future directions and research. A list of literature cited in the text is at the very end.

Since each of the following four chapters are self-contained and contain one article in its original form, there is some overlap and redundant or repeated material. These four chapters begin with a foreword written specifically for this thesis. All referenced material in the research chapters have numbers that are identical to their published (or submitted) counterpart. That is, elements such as sections, equations, figures, and statements are numbered without being prefixed by chapter number.

If referenced from outside of their own chapters, I have included the chapter in the reference; e.g., "Theorem 4 in Chapter C."

4.1 RESEARCH OBJECTIVES AND SCOPE

The objectives of the research included in this thesis are the following:

> *Develop and analyze computationally tractable methods for dual control, including heuristic suboptimal approaches.*

These are fairly wide-ranging objectives, which makes a well-defined scope important. The development of conceptual ideas is more central than comprehensive analysis in the research presented here, and this places natural limits on the types of systems considered. We only study dual control with parametric, as opposed to structural, uncertainty, and do not investigate the effect of measurement noise. Since there is still a lot to discover in dual control of linear systems, none of the presented algorithms are developed for nonlinear systems. With the exception of controller in Paper D, all linear systems have a single input and a single output, constant unknown parameters drawn from known initial Gaussian distributions, and Gaussian white process noise. Papers A and B develop algorithms for A.R.X. systems, while the exact reformulation in Paper C is developed for O.B.F. models.

4.2 CONTRIBUTIONS IN THIS THESIS

The algorithm in Paper C is based on approximating, or rather considering a subset, without quantizing, the elements of the future information set. The resulting problem is solved exactly for O.B.F. models through reformulation techniques. This paper also includes probabilistic output constraints. I consider these results the most important contribution of the thesis.

Applying the reformulation to A.R.X. models, where it is no longer an exact representation of the dual control problem, results in the algorithm we develop in Paper B. As mentioned above, this is a somewhat immature approach, and the development is less rigorous than in Paper C. However, the algorithm is a reasonable approach to semi-exact dual control of A.R.X. systems.

The two control formulations in Paper A are arguably best viewed as approximations or alternatives to a dual controller. That is, the solution utilizes the dual effect in the problem and excites the process for

improved learning, but the heuristic reward for uncertainty reduction means the controllers are not dual since they are not designed from the perspective of optimal performance. Furthermore, the tuning parameters in the objective functions may give a false impression of a trade-off between control and uncertainty reduction.

The semi-exact dual-control algorithm from Paper B is extended to multivariable ARMAX systems and tested in real-time on an experimental setup in Paper D. My involvement in this paper is limited to contributing to extending the algorithm, evaluating the experimental results, and writing parts of the paper. While the experimental data are not conclusive and somewhat difficult to interpret, the results show promise for real-time implementation and performance enhancements for uncertain systems with dual effect. The experiments also demonstrate that this type of near-dual control with data-driven models can work well for fairly complex plants.

4.3 LIST OF PUBLICATIONS

The following three papers form the core of the contributions in this thesis.

A Heirung, T.A.N., Foss, B., and Ydstie, B.E. (2015). "MPC-based dual control with online experiment design." *Journal of Process Control* 32, pp. 64–76.

B Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2013b). "An MPC approach to dual control." In: *Dynamics and Control of Process Systems*. Mumbai, India, pp. 69–74.

C Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2015b). "Dual adaptive model-predictive control." *Under revision for resubmission to Automatica*.

The following paper is the result of a collaboration where I did not do the majority of the work.

D Kumar, K. et al. (2015). "Experimental evaluation of a MIMO adaptive dual MPC." In: *Advanced Control of Chemical Processes*. Whistler, Canada, pp. 546–551.

All of the interesting material from the the first two of the following papers, Papers E and F, is included and expanded in Paper A; the

results from the third of the following three, Paper G, is included and expanded in Paper C.

E Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2012b). "Towards dual MPC." In: *Nonlinear Model Predictive Control*. Noordwijkerhout, the Netherlands, pp. 502–507.
F Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2013a). "An adaptive model predictive dual controller." In: *Adaptation and Learning in Control and Signal Processing*. Caen, France, pp. 62–67.
G Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2015c). "Dual MPC for FIR systems: Information anticipation." In: *Advanced Control of Chemical Processes*. Whistler, Canada, pp. 1034–1039.

4.4 LIST OF PRESENTATIONS

In addition to the conference presentation listed above, I have presented research results from this thesis at the following conferences and workshops.

1 Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2012a). "Optimal input design for parameter identification in dynamic systems using nonlinear programming." In: *Nordic Process Control Workshop*. Kgs Lyngby, Denmark.
2 Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2012c). "Towards model predictive dual control." In: *AIChE Annual Meeting*. Pittsburgh, PA.
3 Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2013c). "MPC-based adaptive dual control." In: *AIChE Annual Meeting*. San Francisco, CA.
4 Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2014). "Optimal control of uncertain systems using dual model predictive control (DMPC)." In: *AIChE Annual Meeting*. Atlanta, GA.
5 Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2015d). "Optimal control of uncertain systems using Dual Model Predictive Control (DMPC)." In: *Nordic Process Control Workshop*. Trondheim, Norway.
6 Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2015a). "A QCQP problem for dual control of FIR systems." In: *The British-French-German Conference on Optimization*. London, UK.
7 Heirung, T.A.N., Morinelly, J.E., et al. (2015). "Dual Model Predictive Control (DMPC): State of the Art." In: *AIChE Annual Meeting*. Salt Lake City, UT.

The invited talks listed below were also based on research contained in this thesis.

1 Heirung, T.A.N. (2014a). "Optimal sequential decision making under uncertainty: anticipatory learning with stochastic dynamic programming." In: *Process system engineering seminar, Carnegie Mellon University*. Pittsburgh, PA.
2 Heirung, T.A.N. (2014b). "Some methods for approximate dual control based on MPC." In: *Seminar at Prof. H.G. Bock's research group at the University of Heidelberg*. Heidelberg, Germany.
3 Heirung, T.A.N. (2015a). "Optimal control of uncertain systems using Dual Model Predictive Control (DMPC)." In: *Seminar at the Ilmenau University of Technology*. Ilmenau, Germany.
4 Heirung, T.A.N. (2015b). "Some approaches to adaptive dual control using MPC." In: *Seminar at Prof. Ali Mesbah's research group at UC Berkeley*. Berkeley, CA.

Part II

RESEARCH ARTICLES

This part contains the main contributions of the thesis, papers A, B, and C, in addition to Paper D. Each paper is included as published (or submitted), with exception of cosmetic improvements, minor changes for better consistency of notation, and corrected typographical errors.

MPC-BASED DUAL CONTROL WITH ONLINE EXPERIMENT DESIGN

Reference information for Paper A:
Heirung, T.A.N., Foss, B., and Ydstie, B.E. (2015). "MPC-based dual control with online experiment design." *Journal of Process Control* 32, pp. 64–76.

A

FOREWORD

This article presents the two heuristic suboptimal control approaches, with objectives that differ from the dual objective. The material here is a formalization and extension of the methods developed in Heirung, Ydstie, and Foss (2012b) and Heirung, Ydstie, and Foss (2013a). The information-maximizing objective first presented in Heirung, Ydstie, and Foss (2013a), here included in Section 4, was applied to optimal experiment design by Janka (2015), who found no advantages over existing methods for the case study considered.

ABSTRACT

We present two dual control approaches to the model maintenance problem based on adaptive model-predictive control (M.P.C.). The controllers employ systematic self excitation and design experiments that are performed under normal operation, resulting in improved control performance with smaller output variance and less control effort. Our control formulations offer a novel approach to the question of how to excite the plant input to generate informative data within the context of M.P.C. and adaptive control. One controller actively tries to reduce the parameter-estimate error covariances; the other controller maximizes the information in the signals for enhanced learning. Our approach differs from existing ones in that we let our controllers converge to standard certainty equivalence (C.E.) M.P.C. when the parameter uncertainty decreases or more information is generated, and as a result we avoid plant excitation when the uncertainty is low or enough informa-

tion has been generated. We demonstrate that the controllers work well with a large number of tuning configurations and also address the issue of models that are not admissible for control design.

1 INTRODUCTION

Model predictive control (M.P.C.) is a widely-adopted control strategy, in large part because it is an inherently multivariable approach and can handle systems that are constrained. Since predictions from a poor model can lead to control action that is far from optimal, high model quality is key to good performance in an M.P.C. loop. The most expensive and time-consuming part of M.P.C. commissioning is frequently cited as modeling, with up to 80 % of the design effort often spent on obtaining a model suitable for M.P.C. (Sun et al., 2013). As the controlled process may change over time, the verisimilitude of the model may decrease and lead to lower performance. There are many possible sources of performance deterioration in M.P.C. loops, including inappropriate setup of constraints, inconsistencies in dynamic optimization in M.P.C. and the higher-level optimization, and poor quality of the input-output and/or disturbance models; among these, the model quality is the most significant for control performance in M.P.C. (Sun et al., 2013).

The active field of controller performance monitoring (C.P.M.) examines questions that are relevant to these issues, among them how to provide information on whether a model is no longer representative of the process behavior (Qin and Yu, 2007; Zagrobelny, Ji, and Rawlings, 2013). In the event of a process model that does not capture the plant dynamics sufficiently, some form of model maintenance must be executed in order to restore the controller's ability to meet its design specifications. Normal operating data recorded in closed loop is rarely informative enough for the purpose of reidentifying the model or for continuous adaptation of the model or controller. However, recently-proposed algorithms are capable of searching through large data sets and locating sufficiently informative segments (Isaksson, 2013) and thus offer a possible alternative to performing dedicated experiments when inadequate models cause poor control performance. Still, model maintenance as a means of improving control performance usually entails conducting a system identification exercise for the entire plant (Sun et al., 2013). Accurate and precise parameter estimates can be obtained by planning the experiment using a model-based experiment design approach, where the current parameters are used to predict the result-

ing information content (Franceschini and Macchietto, 2008). The most common design criteria are based on the parameter-estimate covariance matrix (or its inverse, the information matrix), and include minimizing (maximizing) the determinant, minimizing the largest (maximizing the smallest) eigenvalue, and minimizing (maximizing) the trace. These criteria are commonly referred to as *D*-, *E*-, and *A*-optimal designs, respectively. Franceschini and Macchietto (2008) provide an excellent overview of applications and theoretical developments in model-based experiment design. Their extensive list of applications include many from process systems and closely related fields, including heat and mass transfer, kinetics, fermentation, and biochemical networks. Plant experiments for identification purposes may be expensive for reasons such as the required expertise, their time-consuming nature, and the disruptions to normal operation, all of which lead to a high cost for M.P.C. commissioning or maintenance. Furthermore, the resulting model is not necessarily good enough to increase control performance to a degree that offsets the cost of the experiment.

A common design specification in process control is that the outputs stay close to some optimal operating point while the inputs move as little as possible. When this control goal is met there is a minimum of information being generated in the loop, and standard adaptive control strategies cannot guarantee stability. Hence, in order to ensure informative data, the process must be excited. The conditions for sufficient excitation suggest that the important thing is that *something* be done (excitation), and that it may be less important exactly *how* it is done. There are many simple heuristic approaches to meeting the requirements of persistent excitation; the most recent survey of M.P.C. systems in industry found that most test packages use pseudo-random binary step signals in their identification protocols (Qin and Badgwell, 2003). However, a systematic approach to integrating control design and smart excitation experiments enables the analysis necessary for demonstrating important properties such as convergence and stability.

Larsson, Annergren, et al. (2013) developed an M.P.C. that performs identification experiments while controlling the plant. The experimental aspect of the control signal is a result of adding a constraint on an approximation of the information matrix. By relaxing this constraint they arrive at a convex quadratic programming (Q.P.) problem, which is solved online to obtain the control law. Through an example system, they demonstrated that the input signal excited the system sufficiently and that their approach is significantly better for identifying the plant

than a normal M.P.C. A similar scheme termed Model Predictive Control and Identification (M.P.C.I.) was introduced by Genceli and Nikolaou (1996) and further developed in a number of papers by Nikolaou and coworkers, notably Shouche, Genceli, Vuthandam, et al. (1998) and most recently Shouche, Genceli, and Nikolaou (2002). This approach is based on requiring a persistently exciting input to the process, resulting in a noncovex optimization problem solved to optimality with a branch-and-bound approach developed by the authors.

These algorithms show that M.P.C.-based approaches to integrating control and experiments can be both practically implementable and offer clear advantages over more heuristic means of exciting an unknown process under normal control operation. However, in some sense they require *a priori* that a certain amount of information be generated under closed-loop operation, which may result in unnecessarily invasive experimentation. Furthermore, this type of integration of control and experiment design indirectly results in suboptimal dual controllers.

The dual nature of a signal controlling an unknown plant was first recognized by Feldbaum (1961b). Feldbaum identified the twofold effect of the signal as investigating as well as directing, or as affecting the uncertainty of the unknown parameters as well as the system states. In a series of seminal papers, Feldbaum analyzed the problem and demonstrated that (stochastic) dynamic programming is an appropriate tool for solving dual control problems. Feldbaum's early work on dual control was pioneering in integrating active learning and multistage decision making under uncertainty and inspired it several important directions in stochastic adaptive control.

Despite the conceptually appealing features of dual control, factors such as the limited computational power available in the 1960s and 1970s contributed to a lack of practical algorithms. Åström and Helmersson (1986) were among the first to solve a simple yet nontrivial dual control problem numerically using dynamic programming. The problem they studied had only one unknown parameter but required 180 C.P.U. hours with a time horizon of 30 samples (Åström, 1983). Despite the superior computers available today, the method is ill-suited for solving even moderately large problems because of the curse of dimensionality. The computational complexity and associated difficulties with obtaining dual control laws has led to a limited number of industrial implementations reported in the literature. Allison et al. (1995) reported one of the first applications of suboptimal dual control to a process control problem. Even though several heuristics were used in the control

design, the dual controller improved performance through better identifying gain changes and preventing turn-off. Ismail, Dumont, and Backstrom (2003) successfully implemented a suboptimal dual controller for a paper-coating process. Despite the severe challenges associated with a coupled multivariable process and a gain that drifts over time with frequent sign reversals, their proposed dual controller resulted in a substantial quality improvement over the standard industrial controller.

Lee and Lee (2009) approached the problem of exponential growth in computational requirements using approximate dynamic programming. Their approach uses Monte Carlo simulations with multiple known suboptimal controllers to define a limited region of the hyperstate space and then obtains approximate solutions with dynamic programming within that region. A related method based on sampling and forward dynamic programming with particle filtering was developed by Bayard and Schumitzky (2010).

Only recently has the dual control problem been studied directly within the context of M.P.C. One early M.P.C. for systems with a finite impulse response, similar to those proposed by Shouche, Genceli, and Nikolaou (2002) and Larsson, Annergren, et al. (2013), was developed by Marafioti (2010) and Marafioti, Bitmead, and Hovd (2014). In this approach, the dual character of the control is obtained by guaranteeing the input be persistently exciting. A nonconvex excitation constraint was derived and imposed on the first of the open-loop optimal input variables, resulting in a periodic input signal generated by the controller. Another M.P.C.-based method was developed by Rathouský and Havlena (2013) and later extended by Žáčeková, Prívara, and Pčolka (2013). In the most recent version of the algorithm, the first step is to solve a standard M.P.C. problem with the nominal model; the second step involves finding a perturbation to the first element of the resulting control sequence so that the minimal eigenvalue of the information matrix is maximized without increasing the cost function by more than a prescribed amount. Lucia and Paulen (2014) developed a robust nonlinear M.P.C. capable of actively reducing uncertainty. Deviations from the nominal process output predictions are represented by a scenario tree of possible realizations of uncertainty, and uncertainty is reduced by minimizing an optimal experiment design criterion as part of the algorithm.

We base our approach on certainty-equivalence (C.E.) M.P.C., and expand that structure to allow predictions of how future data will resolve uncertainty. The algorithms we present in the following both reduce pa-

rameter uncertainty and increase the information content of the closed-loop signals. This is accomplished by adding constraints to a standard M.P.C. formulation and augmenting the objective function with a term that explicitly rewards the experimentation. Our formulation does not require that a certain amount of information be generated or that the signals be persistently exciting. The result is controllers that perform experimentation only when there is high uncertainty in the parameters or not enough information for accurate model identification.

This article is organized as follows: Section 2 discusses the problem of uncertainty and learning in the context of predictive and adaptive control and motivates and outlines our proposed design approach; in this section we formulate our problem as well as the M.P.C. framework that we extend, followed by a test problem and a discussion of applying our approach to the admissibility problem. The proposed controllers are presented in Section 3 and Section 4 along with simulation results and sensitivity analyses. In Section 5 we compare and discuss the algorithms. We conclude the article in Section 6 and provide some suggestions for future work. A1 contains a proof of Proposition 1.

2 BACKGROUND AND PROBLEM FORMULATION

2.1 *Uncertainty and active learning in predictive control*

Figure 1 shows the typical adaptive C.E. M.P.C. structure, which consists of two blocks. One block performs parameter estimation while the other updates control inputs using the most recent parameter estimates and an M.P.C. strategy. The parameter estimation block takes the control $u(t)$ and the process output $y(t)$ as inputs and uses these signals to produce an updated estimate of the unknown parameters $\hat{\theta}(t)$ at the discrete time instant t. This enables a form of learning to take place outside of the controller. The parameter estimation procedure typically also generates some measure of confidence in its beliefs about the parameters, such as the parameter-estimate error covariance $P(t)$ in the case of a least-squares algorithm. In a C.E.-type controller, the control input $u(t)$ is calculated using the latest parameter estimates $\hat{\theta}(t)$ as if they were true values; the uncertainty measure (e.g., $P(t)$) does not influence the control signal. This means that the controller treats the estimates as if the uncertainty were zero, and that the control signal does not account for the estimator's beliefs on the precision of the estimates.

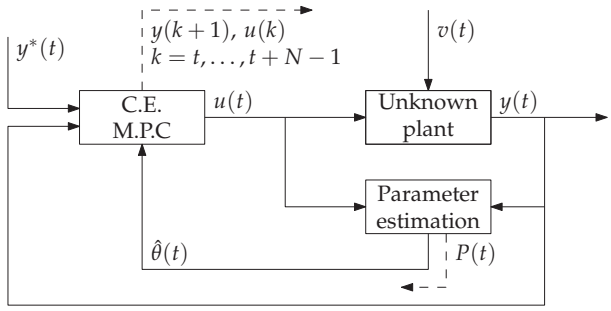


Figure 1: Block diagram illustrating a standard certainty-equivalence M.P.C. structure.

The C.E. M.P.C. uses the current parameters to predict future outputs and control inputs. It does *not* use information about present or future parameter uncertainties. Since the mechanism for learning is exogenous to the controller and in no way included in the prediction model, the controller does not account for the learning that takes place in the loop as a side effect of control. In other words, the controller is unaware of the loop being closed and how future inputs and outputs can generate information since it does not anticipate information acquisition. Therefore, it cannot predict future parameter uncertainty and as a result is unable to take steps for active learning. The prediction is in this case open-loop optimal, meaning the prediction does not account for how future outputs or state trajectories will influence the system through the closed loop. Hence, this particular control structure is often called *open-loop feedback* (feedback is achieved by using the latest output $y(t)$ as initial condition for the prediction calculated at every time t).

It is clear that this strategy can be improved in two ways: first, the control should be robust so that more cautious controls are used if there are large uncertainties; second, it should be explorative to reduce parameter uncertainties. These features are addressed when the controller predicts the future closed-loop response of the system, i.e., not only inputs and outputs, but also how these signals will affect the future information acquisition and hence possible reduction of parameter uncertainty. In the structure outlined in Figure 1, this entails feeding the parameter uncertainty measure $P(t)$ back to the controller, and letting the predictions include the future behavior of this quantity. This design choice leads to a *closed-loop feedback* predictive controller. Now that the controller is aware of the closed loop it anticipates that the

future data can resolve uncertainty and can hence make decisions for active learning; the controller has endogenized the learning. The resulting structure is a model predictive controller with dual features, as illustrated in Figure 2. The input $u(t)$ now both controls and explores through experimentation. The controlling aspect of the signal is motivated by standard control objectives, while the exploring action seeks to excite the plant for improved information acquisition whenever deemed necessary by the controller.

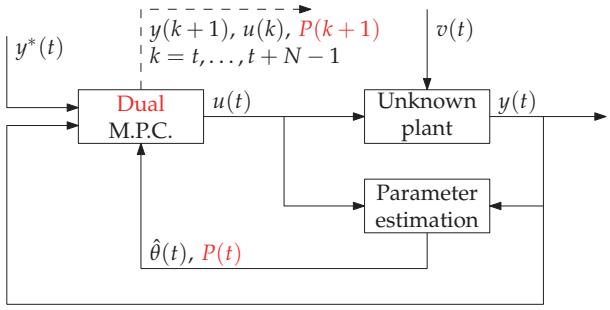


Figure 2: Block diagram illustrating a dual model predictive controller.

2.2 *Problem formulation*

The finite-horizon dual control problem can be formulated as follows: given the recorded process data $\mathcal{Y}(t)$ available at time t, find the sequence of controls $\{u(k)\}_{k=t}^{t+N-1}$ that minimizes the performance measure

This formulation of the dual-control problem is not consistent with that of Chapter 3 or Paper C (see Appendix A2 in Paper C).

$$J_N = \mathrm{E}\left[\sum_{k=t}^{t+N-1} \{w_1 y(k+1)^2 + w_2 u(k)^2\} \;\middle|\; \mathcal{Y}(t)\right] \tag{1}$$

where y is the plant output, $w_1 > 0$ and $w_2 \geq 0$ are weighting parameters, and N is the length of the finite prediction horizon. $\mathrm{E}\{\cdot \mid \mathcal{Y}(t)\}$ is the conditional expectation given the available information. The process model is uncertain and this implies that the causal relationship between the input u and the output y must be inferred based on recorded data. The dynamic properties of the process emerge during exploration, and the resulting information can be expressed in terms of updated model parameter values and corresponding uncertainties. The explicit dual controllers we develop in this article incorporate these features.

The plants we consider in this work are linear and time-invariant with a single input and a single output (SISO), formulated in discrete

time. With these assumptions we formulate our models as autoregressive time-series processes with an exogenous input (A.R.X. processes):

$$y(t) + a_1 y(t-1) + \cdots + a_{n_a} y(t-n_a) = \\ b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) + v(t) \quad (2)$$

Here, t is the discrete time instant; $y(t)$, $u(t)$, and $v(t)$ are the system output, input, and disturbance at time t, respectively; the coefficients $\{a_i\}_{i=1}^{n_a}$ and $\{b_j\}_{j=1}^{n_b}$, $b_1 \neq 0$, are the system parameters, some of which are unknown. The independent and identically distributed Gaussian random variables $v(t)$ have zero mean and variance r.

The A.R.X. model (2) can be written in the more compact form

$$A(q^{-1})y(t) = B(q^{-1})u(t) + v(t) \quad (3)$$

where A and B are polynomials in the backwards shift operator q^{-1}. We assume that the unknown plant is controllable and observable with $A(q^{-1})$ and $B(q^{-1})$ coprime. The unknown model parameters can be determined through analysis of recorded process data using system identification techniques.

We formulate A.R.X. models like (2) in terms of a regression vector $\varphi(t-1)$ and a parameter vector θ;

$$y(t) = \varphi^\top(t-1)\theta + v(t) \quad (4)$$

where

$$\varphi(t-1) = \left[-y(t-1), \ldots, -y(t-n_a), u(t-1), \ldots, u(t-n_b)\right]^\top \quad (5a)$$

contains past inputs and outputs and

$$\theta = \left[a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b}\right]^\top \quad (5b)$$

contains the unknown model parameters. We denote all recorded data available at time t, meaning all past inputs and outputs, by

$$\mathcal{Y}(t) = \left\{u(t), u(t-1), \ldots, y(t), y(t-1), \ldots\right\} \quad (6)$$

The L.Q.G. (linear-quadratic Gaussian) minimizes the performance measure (1) with $w_2 > 0$ when all of the parameters in the model (2) are known. With $w_2 = 0$ and $N = 1$ the optimal performance for a minimum-phase model is obtained by the minimum-variance controller.

In indirect adaptive control we use the past information $\mathcal{Y}(t)$ to estimate the parameters, for instance by solving a least squares problem. The estimates can then be used to determine a control law like the L.Q.G. or the minimum-variance controllers mentioned above knowing that we use the best available estimates given the current information. This is the essence of the certainty-equivalence approach to adaptive control: the estimates are used as if they were the true values (Åström and Wittenmark, 1995). This approach does not consider estimate uncertainty, and it incorporates no mechanism for speeding up convergence and reducing parameter uncertainty by manipulating the control inputs for exploration.

In our proposed control approach we estimate parameters online by minimizing a least-squares criterion recursively. Let $R(t)$ be an information matrix defined as

$$R(t) = \sum_{j=t_0}^{t} r^{-1}\lambda^{t-j}\varphi(j-1)\varphi^\top(j-1) \tag{7}$$

where $\lambda \in (0,1]$ is a forgetting factor. The information matrix $R(t)$ can also be expressed recursively as

$$R(t) = \lambda R(t-1) + r^{-1}\varphi(t-1)\varphi^\top(t-1), \qquad t > t_0 \tag{8a}$$

with $R(t_0)$ given. This matrix then forms the basis for the least-squares estimate of θ at time t, $\hat{\theta}(t)$, with the recursive update equation

$$\hat{\theta}(t) = \hat{\theta}(t-1) + r^{-1}R^{-1}(t)\varphi(t-1)\big(y(t) - \varphi^\top(t-1)\hat{\theta}(t-1)\big) \tag{8b}$$

From the definition of $R(t)$ it is clear that the matrix must be positive semidefinite. When $R(t_0) = 0$, $\operatorname{rank} R(t) \leq t - t_0$ since R undergoes one rank-one update per time step, such that $R(t)$ is invertible at the earliest at time $t_0 + n_a + n_b$.

In order to avoid calculating the inverse of $R(t)$ at every time step, the estimate can be updated through introducing $P(t) = R^{-1}(t)$. The matrix $P(t)$ is positive definite when $R(t)$ is positive definite and is not defined when $R(t)$ is positive semidefinite. The matrix-inversion lemma (see Ljung (1999)) can then be used to derive the recursive algorithm

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\big(y(t) - \varphi^\top(t-1)\hat{\theta}(t-1)\big) \tag{9a}$$

$$K(t) = P(t-1)\varphi(t-1)\big(r\lambda + \varphi^\top(t-1)P(t-1)\varphi(t-1)\big)^{-1} \tag{9b}$$

$$P(t) = \big(I - K(t)\varphi^\top(t-1)\big)P(t-1)(1/\lambda) \tag{9c}$$

where $K(t)$ is commonly referred to as the injection gain. This recursive least-squares (R.L.S.) algorithm can be interpreted as a Kalman filter for estimating the state of a system with the state variable θ constant and output described by Equation (4), provided $\lambda = 1$ (Ljung, 1999). Furthermore, when $\lambda = 1$, the conditional distribution of θ given the information $\mathcal{Y}(t)$ is Gaussian with mean $\hat{\theta}(t)$ and covariance $P(t) = \mathrm{E}[(\theta - \hat{\theta}(t))(\theta - \hat{\theta}(t))^\top]$ as described by the equation set (9) (Åström and Wittenmark, 1995). Note that the value of the estimate $\hat{\theta}(t)$ does not change with r (evident by inserting Equation (7) into Equation (8b)), which makes the inclusion of this variable superfluous when the equation set (9) is used as an estimation algorithm for the parameters θ. However, it is clearly necessary to include r in the equations when we are interested in the variance $P(t)$. We set $r = 1$ in the parameter estimation algorithm when there is no noise present in the system (4).

The parameter estimate $\hat{\theta}(t)$ from the R.L.S. algorithm (9) forms the basis for the one-step-ahead predictor

$$\begin{aligned} \hat{y}(t+1 \mid t) &= \mathrm{E}\big[y(t+1) \;\big|\; \mathcal{Y}(t)\big] \\ &= \varphi^\top(t)\hat{\theta}(t) \end{aligned} \tag{10}$$

which can be used to derive a variety of control laws.

As outlined above, minimizing the dual objective (1) directly can be a formidable computational task for even a moderate number of unknown parameters. Instead of finding approximate solutions to the exact dual problem, we develop control algorithms that find the exact solution to approximations of the dual control problem. These algorithms are based on certainty-equivalence M.P.C., which we formulate next.

2.3 *Certainty-equivalence M.P.C.*

We now introduce the M.P.C. formulation that forms a basis for the controllers we develop in Sections 3 and 4. Consider the quadratic finite-horizon objective function

$$\min_{\{u(t)\}_{k=t}^{t+N-1}} V_{\mathrm{C.E.}}(t) = \sum_{k=t}^{t+N-1} \left\{ w_1 \hat{y}^2(k+1 \mid t) + w_2 u^2(k) \right\} \tag{11}$$

This objective can be modified to include setpoints for both signals, integral action, and control move minimization. The objective function

is minimized at time t and the most recent parameter estimates $\hat{\theta}(t)$ are used to predict the effect of the inputs on the process output using the equation

$$\hat{y}(k+1 \mid t) = \hat{\varphi}^\top(k)\hat{\theta}(t), \qquad k = t, t+1, \ldots, t+N-1 \tag{12}$$

which is a modified version of the predictor (10). Note that the regressor $\hat{\varphi}$ used here contains recorded outputs $y(k)$ for $k \leq t$ and predicted outputs $\hat{y}(k \mid t)$ for $k > t$; similarly, it contains recorded inputs $u(k)$ for $k < t$ and free control variables for $k \geq t$. Most M.P.C. formulations include bounds on inputs and predicted outputs, often called "box constraints." Bounds on the inputs can be formulated as

$$u_{\min} \leq u(k) \leq u_{\max}, \qquad k = t, t+1, \ldots, t+N-1 \tag{13a}$$

where $u_{\min}$ and $u_{\max}$ are the lower and upper bounds on the input, respectively, usually chosen based on physical hardware limitations or design preferences. Since we assume a plant-model mismatch, there is no notion of a nominal guarantee that bounds on the output $y(t)$ will be respected. We therefore constrain the predicted outputs by

$$y_{\min} \leq \hat{y}(k+1 \mid t) \leq y_{\max}, \qquad k = t, t+1, \ldots, t+N-1 \tag{13b}$$

where $y_{\min}$ and $y_{\max}$ are the lower and upper bounds on predicted outputs, respectively. As the model uncertainty decreases and the accuracy of the predictions improves, the constraint (13b) is more likely to ensure that the future plant outputs satisfy

$$y_{\min} \leq y(t) \leq y_{\max} \tag{14}$$

The objective function (11) and the two constraint sets (12) and (13) form a convex Q.P. problem.

At every time instant t, the plant output $y(t)$ and the older signals in the regressor $\varphi(t-1)$ are used in one of the R.L.S. algorithms (8) or (9) to produce an updated parameter estimate $\hat{\theta}(t)$. Subsequently, the open-loop optimization problem formed by (11), (12), and (13) is solved with the output history $\{y(k)\}_{k=t-n_a+1}^{t}$, the input history $\{u(k)\}_{k=t-n_b+1}^{t-1}$, and the current parameter estimate $\hat{\theta}(t)$ given. The solution to the optimal control problem is the open-loop input sequence $\{u^o(k)\}_{k=t}^{t+N-1}$; the first element of this sequence is used as an input to the plant: $u(t) = u^o(t)$. Feedback control is achieved by repeating this process at every sampling time. This control approach is of certainty-equivalence

type, since the control signal $u(t)$ is obtained using the parameter estimate $\hat{\theta}(t)$ as if it were the true parameter set.

Note that the dependence on time t is explicitly noted in the output prediction constraint (12) since the parameter estimate $\hat{\theta}(t)$ appears directly in the equation. Since we write $y(k+1 \mid t)$ in the optimization problem it would be consistent to also write $u(k \mid t)$ and use the notation $u(t) = u^o(t \mid t)$ for the plant input. However, since it is clear from context whether a variable is an optimization variable on the prediction horizon or a physical realization of that variable in the plant, we omit the explicit dependence on t to simplify notation in the following.

2.4 *Extending C.E. M.P.C.: explicit dual predictive control for integrated experiment design*

The following sections present two approximate solutions to the dual control problem. Both approaches result in controllers that excite the process in order to increase the information content of signals and thereby reduce parameter uncertainty, and they both accomplish this through modifying the C.E. M.P.C. framework from Section 2.3 with an augmented objective function and an augmented constraint set. The first algorithm controls the system while explicitly attempting to reduce the parameter estimate error variances; the second controller maximizes the information content of the input and output signals while controlling the system. We refer to the resulting type of controller as explicit dual M.P.C. (D.M.P.C.), since the learning is explicitly rewarded through an added experiment term in the objective function. Early versions of the controllers in Sections 3 and 4 were presented in Heirung, Ydstie, and Foss (2012b) and Heirung, Ydstie, and Foss (2013a), respectively.

The term D.M.P.C. *as used here is not fully consistent with the use in Chapter 3 and Paper C.*

When the experiment term is added to the objective, the parameter uncertainty influences the plant input since part of the control effort is used to reduce predicted uncertainty. Enabling the controller to predict how future data reduces that uncertainty then closes a learning loop in the control structure, effectively enabling endogenous learning in the controller.

2.5 *The admissibility problem*

The admissibility problem (Ydstie, 1997) concerns the fact that certain identified models are not admissible for certainty equivalence control in

the sense that there may be regions in the parameter space that lead to irregular control performance. One example is a pole/zero cancellation in the estimated transfer function, which causes most pole-placement design procedures to fail. An even simpler example is parameters that cause a division by zero in the control law. Similarly, there may be parameters that are non-admissible if the resulting control model is not controllable. See Mareels and Polderman (1996) for a thorough treatment of the admissibility problem, there referred to as the pole/zero cancellation problem.

For our test system (23), a model with $\hat{a}_1 = 0$ and/or $\hat{b}_1 = 0$ is not admissible for control design. $\hat{a}_1 = 0$ hides the fact that the system is dynamic and results in a model with no predictive ability. $\hat{b}_1 = 0$ gives an uncontrollable model and prevents a normal C.E. M.P.C. from producing nonzero controls; other control designs may produce infinitely large control signals. A gain $\hat{b}_1$ very close to zero may also produce unacceptably large inputs.

We address the admissibility problem in Section 4.1, where we provide a proposition stating that our proposed information-maximizing controller will identify a_1 and b_1 in (23) in the worst case of an initial belief state $[R, \hat{\theta}]$ of zero information and all estimates set to zero.

3 MINIMIZING THE VARIANCE OF THE ESTIMATE

As noted above, the conditional distribution of θ given $\mathcal{Y}(t)$ is Gaussian with mean $\hat{\theta}(t)$ and covariance $P(t)$ as described by the equation set (9) when $\lambda = 1$. A natural approach to reducing parameter uncertainty is then to control the system is such a way that the future covariances $P(k+1)$, $k = t, t+1, t+2, \ldots$, are reduced. This entails generating information-rich input and output signals by experimenting with or exciting the plant while minimizing the normal control loss. We approach this by augmenting the control objective with an experiment term, here the function f_P of $P(k+1)$, which we minimize over an experiment horizon N_e, with $N_e \leq N$. The resulting objective function takes the form

$$V_P(t) = w_0 \sum_{k=t}^{t+N_e-1} \delta_e^{k-t} f_P(P(k+1)) + \sum_{k=t}^{t+N-1} \left\{ w_1 \hat{y}^2(k+1\,|\,t) + w_2 u^2(k) \right\} \quad (15)$$

where $w_0 \geq 0$ is a weight parameter and $0 < \delta_e \leq 1$ is a discount factor further discussed below. Note that this objective function reduces to (11) when $w_0 = 0$.

In order to evaluate $P(k+1)$ in the objective function we need to add constraint equations to the optimization problem so that the effect of the input and output on the error covariance is included in the process model used for predictions. This is achieved by adding the two equations

$$K(k+1) = \frac{P(k)\hat{\varphi}(k)}{r + \hat{\varphi}^\top(k)P(k)\hat{\varphi}(k)}, \qquad k = t, t+1, \ldots, t+N_e-1 \quad (16a)$$

$$P(k+1) = (I - K(k+1)\hat{\varphi}^\top(k))P(k), \quad k = t, t+1, \ldots, t+N_e-1 \quad (16b)$$

as constraints to the online optimization problem; $\hat{\varphi}$ is defined as in the predictor (12). The constraint set (16) is highly nonlinear and increases the complexity of the optimization problem in the C.E. M.P.C. given by (11), (12), and (13), turning the Q.P. problem into a nonlinear programming (N.L.P.) problem.

The constraints (16a) and (16b) are identical to (9b) and (9c) from the R.L.S. parameter estimation algorithm, except for the forgetting factor which is set to $\lambda = 1$ in the constraints. Including the forgetting factor would deem a reduction in uncertainty in the near future less valuable than a reduction in the distant future, which is clearly not desirable from a design perspective. A more systematic way of specifying time preference on future uncertainty reduction is to include the discount factor δ_e directly in the objective, providing an intuitive way of formulating that a better parameter estimate at the next time step is more valuable than an improvement further into the future.

There are several possible candidates for the function $f_P(P(k+1))$ in the objective (15), some of which we discuss here. Possible types of matrix functions include various norms, the determinant, and the trace; each of these measure the magnitude of matrices in different ways. Matrix norms involve expensive and nonlinear computations, adding to the complexity already introduced by the constraints responsible for computing $P(k+1)$. The determinant offers an alternative to a matrix norm, and there are various ways of computing the determinant. The simplest expression for the determinant of a matrix is the product of its eigenvalues, but using this formulation in the objective function would necessitate adding nonlinear equality constraints to the N.L.P. problem. One possible approach to supply the eigenvalues is to add a constraint equation that represents the characteristic polynomial of $P(k+1)$. This type of equation is highly nonlinear since the coefficients would be functions of the variable elements of $P(k+1)$, again adding significant

complexity. It is possible to calculate $\det P(k+1)$ without adding equality constraints by instead using the cofactors and minors of $P(k+1)$ to calculate the determinant directly, but this gives a large number of products of the elements of the matrix in the objective function. These large products can be decomposed by adding a bilinear equality constraint for each unique product. Adding a number of these constraints adds substantial complexity, and the number of additional constraints needed becomes intractable with a large number of unknown parameters n_p. One possible way of including a determinant is through the matrix exponential and the relation

$$\det(\exp(P(k+1))) = \exp(\operatorname{tr}(P(k+1))) \tag{17}$$

The right-hand side of this expression is cheap to evaluate since the trace is the sum of the diagonal elements, and the exponential of a scalar is a convex function that is simple to calculate. Since the diagonal elements of $P(k+1)$ are already available, the trace can be evaluated without adding constraints to the optimization problem. While computationally attractive, this function candidate can lead to tuning difficulties as the exponential dominates the other terms in the objective function when uncertainty is high.

We utilize the computational simplicity of the trace by choosing

$$f_P(P(k+1)) = \operatorname{tr} P(k+1) \tag{18}$$

where we retain the advantage of not having to add constraints while avoiding the considerable extra penalty on large variances imposed by the exponential function. The one-stage cost function proposed by Wittenmark (1975) is a special case of the objective function resulting from choosing $\operatorname{tr} P(k+1)$. Since the trace is both the sum of the diagonal elements of the matrix and the sum of the eigenvalues of the matrix, the trace provides the sum of all the parameter estimate error variances and is hence directly related to the original objective of reducing variance. Minimizing the trace of P corresponds to A-optimal design in the field of optimal experiment design (Franceschini and Macchietto, 2008). Note that since $P(k+1)$ is always positive definite, the trace and determinant are similar in that

$$\frac{\partial \det P(k+1)}{\partial \lambda_i} > 0 \qquad \text{and} \qquad \frac{\partial \operatorname{tr} P(k+1)}{\partial \lambda_i} > 0 \tag{19}$$

for all eigenvalues λ_i of $P(k+1)$.

The optimization problem solved at time t can thus be stated as follows:

$$\min_{\{u(k)\}_{k=t}^{t+N-1}} V_P(t) = w_0 \sum_{k=t}^{t+N_e-1} \delta_e^{k-t} \operatorname{tr} P(k+1) + \sum_{k=t}^{t+N-1} \left\{ w_1 \hat{y}^2(k+1 \mid t) + w_2 u^2(k) \right\} \quad (20a)$$

subject to

$$\hat{y}(k+1 \mid t) = \hat{\varphi}^\top(k)\hat{\theta}(t), \qquad k = t, t+1, \ldots, t+N-1 \quad (20b)$$

$$u_{\min} \leq u(k) \leq u_{\max}, \qquad k = t, t+1, \ldots, t+N-1 \quad (20c)$$

$$y_{\min} \leq \hat{y}(k+1 \mid t) \leq y_{\max}, \qquad k = t, t+1, \ldots, t+N-1 \quad (20d)$$

$$K(k+1) = \frac{P(k)\hat{\varphi}(k)}{r + \hat{\varphi}^\top(k)P(k)\hat{\varphi}(k)}, \qquad k = t, t+1, \ldots, t+N_e-1 \quad (20e)$$

$$P(k+1) = (I - K(k+1)\hat{\varphi}^\top(k))P(k), \quad k = t, t+1, \ldots, t+N_e-1 \quad (20f)$$

$$\{y(k)\}_{k=t-n_a+1}^{t}, \{u(k)\}_{k=t-n_b+1}^{t-1}, P(t), \text{ and } \hat{\theta}(t) \text{ given} \quad (20g)$$

The given values in $\varphi(t)$ include the initial value for the output, $y(t)$, along with the n_a most recent outputs y and n_b most recent control inputs u. Hence, the controller takes feedback from the hyperstate defined by $\{y(k)\}_{k=t-n_a+1}^{t}$, $\{u(k)\}_{k=t-n_b+1}^{t-1}$, $P(t)$, and $\hat{\theta}(t)$.

When parameter uncertainty is reduced, $P(t) \to 0$, which means the objective function (20a) converges to the objective function (11) of the C.E. M.P.C. Hence, the degree of experimentation decreases as the parameter uncertainty is reduced.

The objective function (20a) is quadratic and convex, and no computational complexity is introduced by adding the linear experiment term $\operatorname{tr} P(k+1)$ to the objective function in the linear C.E. M.P.C. in Section 2.3. All of the added complexity comes from augmenting the constraint set with the nonlinear equality constraints (20e) and (20f) that represent future parameter uncertainty. The linear C.E. M.P.C. obtains the control input $u(t)$ by solving a Q.P. problem at every time sample; our extension turns the optimization problem into a nonconvex N.L.P. problem.

3.1 *Reformulation for implementation*

The formulation of the optimal control problem (20) can be significantly simplified. We introduce the variables $z(k) \in \mathbb{R}^{n_p \times 1}$ and $\zeta(k) \in \mathbb{R}^1$ defined as

$$z(k) = P(k)\hat{\varphi}(k), \qquad k = t, t+1, \ldots, t+N_e-1 \tag{21a}$$
$$\zeta(k) = \hat{\varphi}^\top(k) z(k), \qquad k = t, t+1, \ldots, t+N_e-1 \tag{21b}$$

and write the optimization problem as

$$\min_{\{u(k)\}_{k=t}^{t+N-1}} V_P(t) = w_0 \sum_{k=t}^{t+N_e-1} \delta_e^{k-t} \operatorname{tr} P(k+1) + \sum_{k=t}^{t+N-1} \left\{ w_1 \hat{y}^2(k+1 \mid t) + w_2 u^2(k) \right\} \tag{22a}$$

subject to

$$\hat{y}(k+1 \mid t) = \hat{\varphi}^\top(k)\hat{\theta}(t), \qquad k = t, t+1, \ldots, t+N-1 \tag{22b}$$
$$u_{\min} \leq u(k) \leq u_{\max}, \qquad k = t, t+1, \ldots, t+N-1 \tag{22c}$$
$$y_{\min} \leq \hat{y}(k+1 \mid t) \leq y_{\max}, \qquad k = t, t+1, \ldots, t+N-1 \tag{22d}$$
$$z(k) = P(k)\hat{\varphi}(k), \qquad k = t, t+1, \ldots, t+N_e-1 \tag{22e}$$
$$\zeta(k) = \hat{\varphi}^\top(k) z(k), \qquad k = t, t+1, \ldots, t+N_e-1 \tag{22f}$$
$$z(k) = rK(k+1) + \zeta(k)K(k+1), \qquad k = t, t+1, \ldots, t+N_e-1 \tag{22g}$$
$$P(k+1) = P(k) - K(k+1)z^\top(k), \qquad k = t, t+1, \ldots, t+N_e-1 \tag{22h}$$
$$\{y(k)\}_{k=t-n_a+1}^{t}, \{u(k)\}_{k=t-n_b+1}^{t-1}, P(t), \text{ and } \hat{\theta}(t) \text{ given} \tag{22i}$$

This reformulated optimization problem has $(n_p+1)N_e$ added equations, and all constraints are now quadratic (or bilinear). With our choice of $f_P(P(k+1)) = \operatorname{tr} P(k+1)$, the problem can be simplified further by only implementing the diagonal elements of the matrix equality constraint (22h). There is then a total of $(3n_p+1)N_e$ constraints added to the standard C.E. M.P.C. formulation for the purpose of achieving the experimentation; the constraints contain a total of $(n_p^2+3n_p)N_e$ bilinear terms. Since the objective function is quadratic, the problem is a quadratically-constrained quadratic programming (Q.C.Q.P.) problem; solvers dedicated to solving this problem class to global optimality have been developed. See Misener and Floudas (2013) for a description of one such algorithm (GLOMIQO) and a review of others.

3.2 *Simulation results*

The proposed controllers are demonstrated on a system of the form

$$y(t) + a_1 y(t-1) = b_1 u(t-1) + v(t) \tag{23}$$

with $a_1 \neq 0$ and $b_1 \neq 0$. This gives the vectors

$$\varphi(t) = \left[-y(t),\ u(t)\right]^\top, \qquad \hat{\theta}(t) = \left[\hat{a}_1(t),\ \hat{b}_1(t)\right]^\top \tag{24}$$

and the predictor

$$\hat{y}(t+1 \mid t) = -\hat{a}_1(t)y(t) + \hat{b}_1(t)u(t) = \varphi^\top(t)\hat{\theta}(t) \tag{25}$$

Unless otherwise noted the simulation examples in the following sections are all performed with $a_1 = -1.1$ (an unstable system), $b_1 = 1$, $y(0) = 1.0$, and noise variance $r = 0.1$. The parameter estimates are initialized with $\hat{a}_1(0) = 0.1$, $\hat{b}_1(0) = -0.1$, and the forgetting factor is $\lambda = 0.99$. In the M.P.C. the objective weights are $w_1 = w_2 = 1$, the prediction horizon is $N = 10$, and the signal bounds are $y_{\min} = -10$, $y_{\max} = 10$, $u_{\min} = -5$, $u_{\max} = 5$.

All results presented below are produced with the two algorithms implemented in MATLAB and GAMS (GAMS Development Corporation, 2015) with IPOPT (Cervantes et al., 2000) as the N.L.P. solver. IPOPT is a state-of-the-art interior-point solver, which exploits the sparse structure of the N.L.P. and is capable of solving very large problems. The system is simulated in MATLAB; GAMS is called at every time step t and returns a (local) solution to the optimal control problem to MATLAB. All solution times are reported in C.P.U. seconds (C.P.U. s) and obtained with a 1.6 GHz Intel Core i5.

Unless otherwise noted, the parameters in this and the following sections are as follows: the covariance matrix is initialized at $P(0) = 10^{-3}I$, the information matrix is initialized at $R(0) = 0$, the experiment weight is $w_0 = 10$, the experiment horizon length is $N_e = 4$, the experiment discount factor is $\delta = 1$, and all simulations are from time $t_0 = 0$ to time $t_f = 10$.

Figure 3 shows a comparison simulation of the variance-minimizing D.M.P.C. defined by (22) with $w_0 = 10$ and the standard C.E. M.P.C. from Section 2.3 (corresponding to $w_0 = 0$). That the parameter estimates are initialized nowhere close to their true values, with wrong signs and at one order of magnitude error, causes the C.E. M.P.C. to excite the system accidentally through the misinformed control decisions.

The D.M.P.C. excites the system because of the experiment term with a weight $w_0 = 10$, and the controller is able to achieve both smaller output variance (1.31, compared with 5.50) and smaller control input energy (3.41, compared with 5.35). As noted above, the objective functions of the two controllers converge as the parameter error variances go to zero, which can be observed here after around $t = 7$, where the plant output and the control input are nearly identical.

The following section provides a better illustration of how the performance of the D.M.P.C. changes as the different tuning parameters are adjusted.

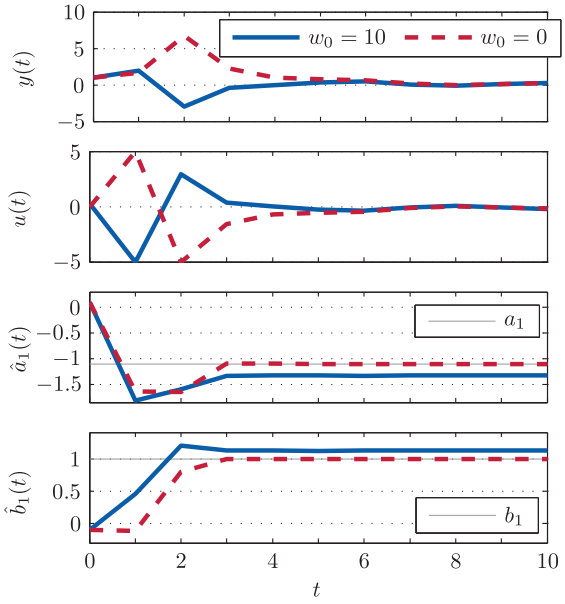


Figure 3: Representative example simulation comparing the variance-minimizing D.M.P.C. defined by (22) with $w_0 = 10$ and the standard C.E. M.P.C. from Section 2.3 (corresponding to $w_0 = 0$).

3.3 *Sensitivity analyses*

Figures 4 and 5 show the sensitivity of the D.M.P.C. control performance with respect to the experiment weight w_0 and the excitation horizon N_e, respectively. Since the graphs illustrating the sensitivity analyses for r and δ provide very little insight beyond the discussion below, they are

not included. All simulations are performed on the example system (23) with parameters as indicated above. The results are based on averaging 100 simulations for each plotted value of the tuning parameters, with a different random seed in each simulation for the noise sequence $v(t)$. In Figures 4 and 5, the average output error E_y, average control effort E_u, and average estimate errors $E_{\hat{a}_1}$ and $E_{\hat{b}_1}$ are defined as follows:

$$E_y := \sum_{t=t_0+1}^{t_f} \frac{y^2(t)}{t_f - t_0} \tag{26a}$$

$$E_u := \sum_{t=t_0}^{t_f-1} \frac{u^2(t)}{t_f - t_0} \tag{26b}$$

$$E_{\hat{a}_1} := \sum_{t=t_0+1}^{t_f} \frac{(a_1 - \hat{a}_1(t))^2}{t_f - t_0} \tag{26c}$$

$$E_{\hat{b}_1} := \sum_{t=t_0+1}^{t_f} \frac{(b_1 - \hat{b}_1(t))^2}{t_f - t_0} \tag{26d}$$

The sensitivity with respect to the experiment weight w_0 is shown in Figure 4. The standard C.E. M.P.C. corresponding to $w_0 = 0$ is included at the left. As the experimentation term is included with a small weight, the most notable change is the improvement in the estimate of the input gain b_1. No notable improvement in output variance or input usage takes place until w_0 is in the range 1–10. In this range there is a significant reduction of output variance, and this is achieved with less use of control input. The simultaneous reduction in these two quantities shows how excitation does not necessarily come at the price of deterioration in output regulation. As the experiment weight is increased toward a value of 1000 the performance suffers, likely because of the large imbalance in the objective weights which can lead to an ill-conditioned problem. The quality of the estimate $\hat{a}_1(t)$ is largely unaffected by the experiment weight, while $\hat{b}_1(t)$ is significantly improved when w_0 goes from 0 to 0.001. That the number of C.P.U. seconds spent solving the optimization problems increases notably for the highest values of w_0 is also likely caused by the poorly scaled objective function.

Consistent with the observation that increasing the experiment weight w_0 up to a moderate value improves performance, increasing the experiment horizon N_e also improves performance up to a point. Figure 5 shows that both output variance and input usage decrease significantly when N_e is increased from 0 (the non-dual C.E. M.P.C.) to 1. Increasing

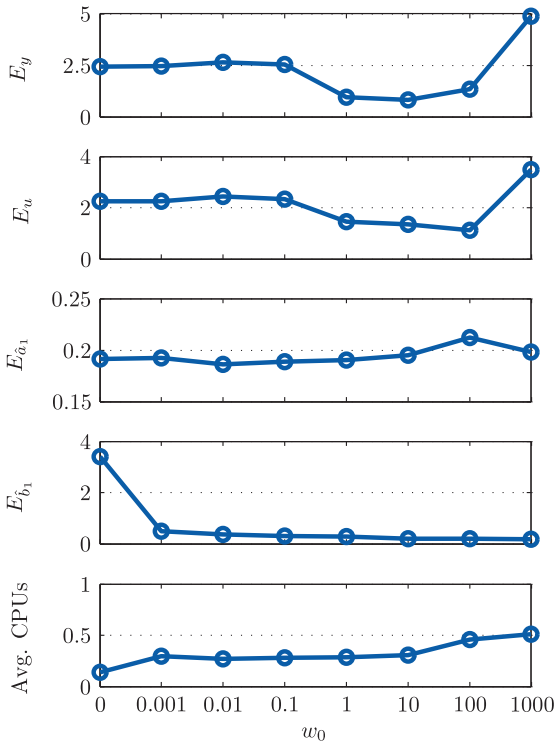


Figure 4: Sensitivity of various performance metrics with respect to the experiment weight w_0 for the variance-minimizing controller defined by (22). Note that the abscissa is logarithmic except for the inclusion of $w_0 = 0$, where the data are obtained using the non-dual C.E. M.P.C. from Section 2.3. The quantities E_y, E_u, $E_{\hat{a}_1}$, and $E_{\hat{b}_1}$ are defined in Equation (26).

N_e to 2 reduces the output variation slightly but increases the control input usage. A further increase in N_e reduces the input usage again with minimal effect on output regulation. Also here the estimate $\hat{b}_1(t)$ benefits from turning on the dual feature by increasing N_e from 0 to 1 while there is very little variation in the quality of $\hat{a}_1(t)$ with different experiment horizon lengths. The reason the time required to solve the optimal control problems increases with N_e is that every increase in N_e adds a number of nonlinear constraints.

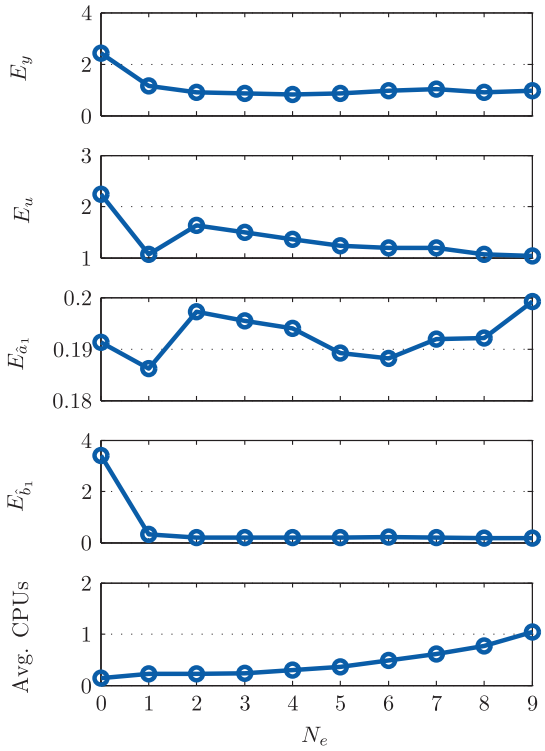


Figure 5: Sensitivity of various performance metrics with respect to the experiment horizon N_e for the variance-minimizing controller defined by (22). The data for $N_e = 0$ are obtained using the non-dual C.E. M.P.C. from Section 2.3. The quantities E_y, E_u, $E_{\hat{a}_1}$, and $E_{\hat{b}_1}$ are defined in Equation (26).

A similar sensitivity analysis with respect to the noise variance r reveals that control performance suffers from increased noise variance. Both output variance and control expenditure increase with the larger

disturbance, and the quality of the parameter estimates suffers similarly. Since the noise variance does not affect the complexity of the optimization problems, the C.P.U. time spent on solving the problems varies little with the noise magnitude.

An investigation of the effect of the experiment discount factor δ reveals that the performance is not sensitive to this parameter. A discount factor in the range 0.95–0.99 improves output regulation slightly compared with larger and smaller values, while the control effort is minimally affected. Heavily discounting future parameter estimate variance reduction is in principle similar to reducing the length of the experiment horizon N_e and decreasing the experimentation weight w_0. It is therefore consistent with the above observations that reducing δ from 0.95 to 0.5 has a negative effect on output regulation and a somewhat smaller negative effect on control input usage. The parameter estimate accuracy does not vary significantly with δ. Since the complexity of the optimization problems is unaffected by the value of the discount factor, the time spent solving the optimization problems is not significantly affected.

4 MAXIMIZING INFORMATION

Motivated by the complexity and the high number of added constraint equations in the optimization problem (22), we propose a different approach that offers computational advantages while solving a similar problem. Instead of using the covariance matrix P to formulate an objective term that rewards exploration and experimentation, we use the information matrix $R = P^{-1}$ in our augmented objective function and constraint set.

We begin by adding an equation to the constraint set of the C.E. M.P.C. Since we aim to express the exploration reward in terms of R, we take Equation (8a) and formulate the equality constraint equation

$$R(k+1) = R(k) + r^{-1}\hat{\varphi}(k)\hat{\varphi}^{\top}(k), \qquad k = t, t+1, \ldots, t+N_e-1 \quad (27)$$

for predicting future information over the experiment horizon N_e. The forgetting factor λ from Equation (8a) is not included here, consistent with the formulation in Section 3. We again augment the C.E. objective

function (11) with an experiment term, now a function f_R of $R(k+1)$, and define the objective V_R as

$$V_R(t) = w_0 \sum_{k=t}^{t+N_e-1} \delta_e^{k-t} f_R(R(k+1)) + \sum_{k=t}^{t+N-1} \left\{ w_1 \hat{y}^2(k+1 \mid t) + w_2 u^2(k) \right\} \quad (28)$$

For expressing the exploration reward function f_R, we start by noting that

$$f_P(P) = \operatorname{tr} P = \sum_{i=1}^{n_p} P_{ii} = \sum_{i=1}^{n_p} \lambda_i^P = \sum_{i=1}^{n_p} \frac{1}{\lambda_i^R} \quad (29)$$

where λ_i^P and λ_i^R are the ith eigenvalues of P and R, respectively, and $n_p = n_a + n_b$ is the number of parameters in the plant model. A possible approach would be to use

$$f_R(R(k+1)) = \sum_{i=1}^{n_p} \frac{1}{\lambda_i^{R(k+1)}} \quad (30)$$

which amounts to substituting this function for $\operatorname{tr} P(k+1)$ in the objective (20a) and replace the constraints (20e) and (20f) with (27), and then add constraints that provide the eigenvalues of $R(k+1)$ for $t \leq k \leq t + N_e - 1$. As mentioned above, the cost of eigenvalue calculations in constraints prevents this from being a computationally viable approach.

Patwardhan and Goapluni (2014) maximize the trace of the information matrix by maximizing a standard M.P.C. objective in their approach to input design for closed loop identification. They demonstrate that maximization of the M.P.C. objective is equivalent to maximizing the trace of the information matrix under mild assumptions, leading to T-optimal (or A-optimal, see Franceschini and Macchietto (2008)) inputs through minimizing a concave objective function.

We can develop a convex experiment function candidate that maximizes the trace of the information matrix by modifying Equation (17) into

$$f_R(R(k+1)) = \det(\exp(-R(k+1))) = \exp(-\operatorname{tr}(R(k+1))) \quad (31)$$

The right-hand side of this equation is cheap to evaluate and there are no additional constraints required beyond Equation (27). An important aspect of maximizing the trace of the information matrix is that the trace can be maximized over the prediction horizon by increasing only one of the diagonal elements of $R(k+1)$, meaning there is no guaranteed incentive to distribute the experimental effort into providing information about all parameters.

Using the function (30) as a basis for a more suitable function candidate, we observe that $\operatorname{tr} P$ is minimized by maximizing all of the eigenvalues of R. This motivates maximizing the trace of R by increasing all of the diagonal elements as opposed to just a few. Furthermore, the function in Equation (30) also captures the diminishing marginal return of increased information through the term $1/\lambda_i^R$, with $\operatorname{tr} P$ approaching its minimum as all eigenvalues of R approach infinity. A computationally simpler function that rewards information and experimentation by evenly increasing the diagonal elements R_{ii} of R is

$$f_R(R(k+1)) = \sum_{i=1}^{n_p} \exp\bigl(-R_{ii}(k+1)\bigr) \tag{32}$$

which also approaches its minimum as the trace of R increases with diminishing return. Hence, this objective too corresponds to an A-optimal experiment design criterion (Franceschini and Macchietto, 2008). This candidate for f_R is convex and does not require additional constraint equations in the optimization problem. Using this function for rewarding process experimentation results in solving the following optimization problem at every time sample t:

$$\begin{aligned}\min_{\{u(k)\}_{k=t}^{t+N-1}} V_R(t) = w_0 \sum_{k=t}^{t+N_e-1} \delta_e^{k-t} \sum_{i=1}^{n_p} \exp\bigl(-R_{ii}(k+1)\bigr) \\ + \sum_{k=t}^{t+N-1} \bigl\{w_1 \hat{y}^2(k+1 \mid t) + w_2 u^2(k)\bigr\}\end{aligned} \tag{33a}$$

subject to

$$\hat{y}(k+1 \mid t) = \hat{\varphi}^\top(k)\hat{\theta}(t), \qquad k = t, t+1, \ldots, t+N-1 \tag{33b}$$

$$u_{\min} \leq u(k) \leq u_{\max}, \qquad k = t, t+1, \ldots, t+N-1 \tag{33c}$$

$$y_{\min} \leq \hat{y}(k+1 \mid t) \leq y_{\max}, \qquad k = t, t+1, \ldots, t+N-1 \tag{33d}$$

$$R(k+1) = R(k) + r^{-1}\hat{\varphi}(k)\hat{\varphi}^\top(k), \quad k = t, t+1, \ldots, t+N_e-1 \tag{33e}$$

$$\{y(k)\}_{k=t-n_a+1}^{t},\ \{u(k)\}_{k=t-n_b+1}^{t-1},\ R(t), \text{ and } \hat{\theta}(t) \text{ given} \tag{33f}$$

Analogous to the objective function (20a), V_R will converge to the C.E. objective (11) as experimentation increases information, as measured by the diagonal elements of R. The convexity of the objective function is conserved as the exponential function is convex. However, the optimization problem is no longer convex because of the introduction of nonlinear equality constraints, and must therefore be solved with

a general N.L.P. solver. Note that only the equations that update the diagonal elements in the constraint (33e) need be implemented. Hence, the experimentation feature is achieved by adding $n_p N_e$ constraints that contain a total of $n_p N_e$ quadratic terms.

4.1 *Application to the admissibility problem*

We now demonstrate some aspects of the information-maximizing controller by investigating its performance in a simple example. We use the system

$$y(t) = -a_1 y(t-1) + b_1 u(t-1) = \varphi^\top(t)\theta(t) \tag{34}$$

where $n_p = 2$ and $v_t \equiv 0$ (so that $r = 1$). The following proposition establishes that the information-maximizing controller identifies the two parameters in the shortest possible time when starting with zero-estimates and zero initial information.

PROPOSITION 1. *Let $N_e = 1$, $\lambda = 1$, $w_0 > w_2$ and $u_{\min} < u_{\max}$. The information-maximizing M.P.C. defined by the optimization problem* (33) *identifies the parameters $a_1 \neq 0$ and $b_1 \neq 0$ in* (34) *in two time steps with zero initial information ($R(t_0)$ a 2×2 zero matrix) and an initial uncontrollable and unobservable process model ($\hat{a}_1(t_0) = 0$, $\hat{b}_1(t_0) = 0$).*

Proof. See Appendix A1. □

The proposition suggests that some care must be taken when choosing the weights w_0 and w_2. This is further explored through sensitivity analyses in Section 4.3.

The simulation example shown in Figure 6 demonstrates Proposition 1. The information-maximizing D.M.P.C. avoids the admissibility problem associated with the estimates initialized at zero and identifies the parameters in two time steps. The C.E. M.P.C. is unable to resolve the situation and the system output hence grows because the system is unstable and the input $u(t)$ remains zero.

4.2 *Simulation results*

The example setup in this section is identical to the one described in Section 3.2. Figure 7 shows a typical simulation example comparing the information-maximizing D.M.P.C. defined by (33) with $w_0 = 10$ and the C.E. M.P.C. from Section 2.3 (corresponding to $w_0 = 0$). The C.E. M.P.C.

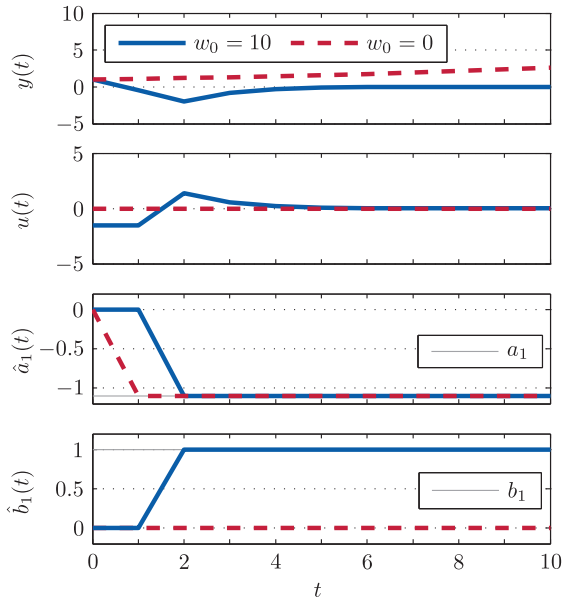


Figure 6: Simulation example demonstrating Proposition 1 with the information-maximizing D.M.P.C. defined by (33) and the C.E. M.P.C. from Section 2.3.

excites the system as a side effect of the poor initial parameter estimates in this case too. However, the information-maximizing D.M.P.C. excites the system in a systematic manner motivated by the experiment term formulated in $R(t)$ and achieves both lower output variance (1.05, compared with 5.50), and smaller input energy (0.30 compared with 5.35). Also here the output and control input trajectories converge around $t = 7$, in this case as the information increases.

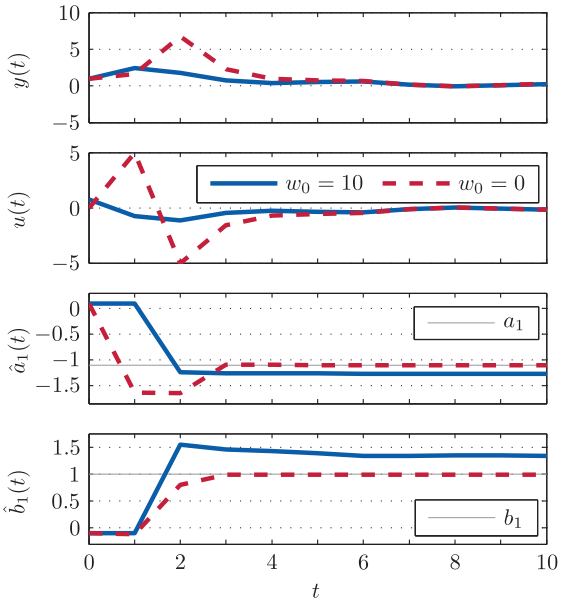


Figure 7: Representative example simulation comparing the information-maximizing D.M.P.C. defined by (33) with $w_0 = 10$ and the standard C.E. M.P.C. from Section 2.3 (corresponding to $w_0 = 0$).

4.3 *Sensitivity analyses*

The sensitivity analyses for the information maximizing controller are based on the same approach as the ones in Section 3.3. Each plotted value of the tuning parameters is based on 100 simulations, each with a different random seed for the noise sequence. The quantities E_y (average output error), E_u (average control effort), and $E_{\hat{a}_1}$ and $E_{\hat{b}_1}$ (average estimate errors) in Figures 8 and 9 are defined in Equation (26).

The sensitivity with respect to the experiment weight w_0 is shown in Figure 8. As suggested by the analysis in Section 4.1, there is little improvement in performance when $w_0 < w_2 = 1$. Additionally, the parameter estimates are very inaccurate when $w_0 < 0.1$. Performance is best when w_0 is in the range 1–10, with a marginal increase in input energy after this point. As long as $w_0 \geq 0.1$ the estimate $\hat{b}_1(t)$ is significantly better than with the non-dual C.E. M.P.C. while $\hat{a}_1(t)$ is worse but acceptable for $w_0 \geq 0.1$. Again, the results show the scope for improving performance by exciting the system appropriately without paying a price in output regulation.

The impact of the experiment horizon length N_e is shown in Figure 9. The biggest improvements come from increasing N_e from 0 (the non-dual C.E. M.P.C.) to 1 and thereby turning on the dual feature. There is very little to gain by increasing N_e beyond 1. When N_e increases from 6 the input energy increases along with the output variance, demonstrating that the controller is not able to "hide" the excitation without increasing the output variance with a suboptimal combination of tuning parameters. Also here the estimate $\hat{b}_1(t)$ benefits greatly from turning on the dual feature by increasing N_e from 0; the estimate $\hat{a}_1(t)$ is somewhat better when $N_e = 0$ and not much affected by an increase in the experiment horizon length beyond $N_e = 1$. Furthermore, increasing N_e adds complexity to the optimization problem (33) by increasing the number of bilinear constraints and exponential terms in the objective, leading to a higher computational cost as shown in the bottom plot in the figure.

Like the variance-minimizing D.M.P.C., the control performances with information-maximizing D.M.P.C. decreases with increasing noise variance r. However, the parameter estimates are not as adversely affected when the information-maximizing D.M.P.C. is used, unless the noise variance increases to a value above 1. The solution times for the optimization problems are not significantly affected by the noise variance.

The information-maximizing D.M.P.C. is almost entirely insensitive to changes in the experiment discount factor δ, with virtually no change in performance in the range $0.9 \leq \delta \leq 1$. When δ is in the range 0.6–0.9 there is a pronounced deterioration of output variance, input cost, and parameter estimate quality. For the lower values 0.5–0.6 the performance metrics are similar to those for $\delta \geq 0.9$. The solution times are largely unaffected by the value of δ.

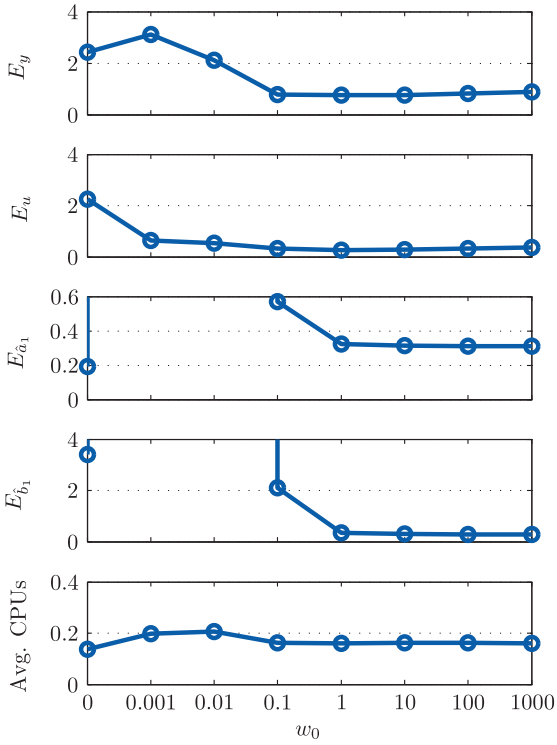


Figure 8: Sensitivity of various performance metrics with respect to the experiment weight w_0 for the information-maximizing controller defined by (33). The average parameter estimate error is not shown for $w_0 = 0.001$ and $w_0 = 0.01$ because the values are disproportionately large. Note that the abscissa is logarithmic except for the inclusion of $w_0 = 0$, where the data are obtained using the non-dual C.E. M.P.C. from Section 2.3. The quantities E_y, E_u, $E_{\hat{a}_1}$, and $E_{\hat{b}_1}$ are defined in Equation (26).

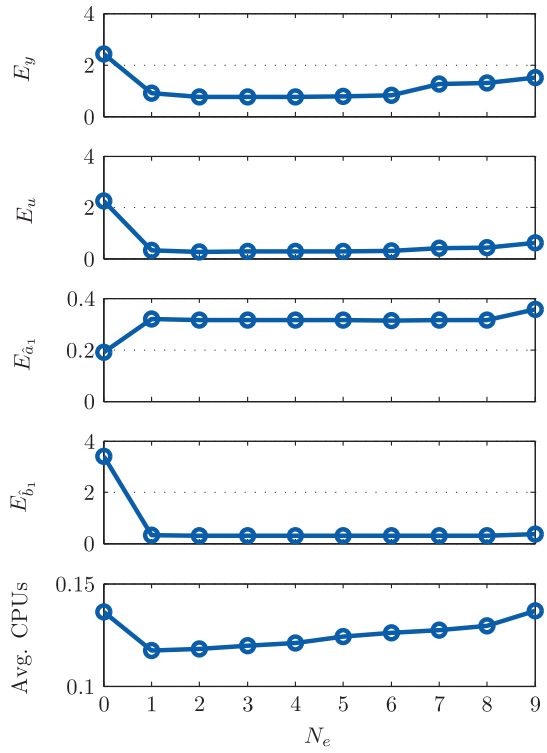


Figure 9: Sensitivity of various performance metrics with respect to the experiment horizon N_e for the information-maximizing controller defined by (33). The data for $N_e = 0$ are obtained using the non-dual C.E. M.P.C. from Section 2.3. The quantities E_y, E_u, $E_{\hat{a}_1}$, and $E_{\hat{b}_1}$ are defined in Equation (26).

The sensitivities with respect to r and δ are not plotted since minimal insight can be gained beyond what is described above.

5 DISCUSSION

Both D.M.P.C. formulations show an improvement in output regulation when N_e is increased from 0 (non-dual M.P.C.) to 1 (D.M.P.C. with myopic experimentation) when applied to our test problem from Section 3.2. Increasing N_e beyond 1 has a positive effect on output regulation with the variance-minimizing D.M.P.C., but that comes at a price of increased control effort. Of the two controllers, the information-maximizing D.M.P.C. benefits the most from increasing the experiment horizon length N_e from 0 to 1. This controller does not benefit much from increasing N_e beyond 1, and the variance-minimizing controller is not able to achieve similar performance by increasing N_e. Since all solutions are found using a local N.L.P. solver, these results may differ if a global solver is used.

The controllers perform similarly with respect to output regulation for experiment weights w_0 in the range 1 to 100, but the information-maximizing controller uses less control effort in this range. The performance of the variance-minimizing D.M.P.C. does not change with w_0 in the range 0 to 0.1; in the same range the information-maximizing controller performs worse than with any $w_0 \geq 1$.

Taken together, these results show that for our test case both proposed D.M.P.C. approaches can achieve the same improvement in output regulation compared with a non-dual C.E. M.P.C. However, the information-maximizing controller improves output regulation with less control effort.

The number of C.P.U. seconds required to solve the optimization problems is significantly different between the two D.M.P.C. algorithms as the experiment horizon N_e increases, which is shown in Figures 5 and 9. Since the number of nonlinear equality constraints in the optimization problems increases with N_e at a higher rate for the variance-minimizing D.M.P.C., the extent to which the variance-minimizing controller is the more complex can be quantified by its larger requirements in computational resources. This difference in required C.P.U. seconds between the two controllers is likely to increase with N_e, suggesting that the information-maximizing controller is a better choice when applied to larger systems if computation time is a possible constraint.

Our experience suggests that the best values of w_0 and N_e largely depend on the plant and that universally good values most likely do not exist. Nevertheless, it seems that choosing a w_0 that is at least as large as w_2 (as suggested by Proposition 1) works well for both controllers in general, and not only for the information-maximizing controller. Our experience furthermore indicates that in most cases w_0 should be at least as large as w_1 and that w_0 should be increased if the noise variance r increases significantly. A good first value for N_e is the number of parameters n_p in the model. However, the value of N_e can be somewhat smaller than n_p if n_p is large, without any noticeable effect on performance.

The approaches presented here do not immediately extend to multivariable systems or real-time control. One possible extension by Kumar et al. (2015) relies on formulating a multiple-input multiple-output linear system as a set of multiple-input single-output ARMAX (Ljung, 1999) models and estimating parameters with an extended least-squares approach. The resulting controller is tested on a benchmark continuously-stirred tank heater system (Thornhill, Patwardhan, and Shah, 2008) with two tanks. The experimental results are promising, but much work remains on evaluating the performance of this approach for real-time control of multivariable systems and there are several theoretical aspects that require further investigation.

6 CONCLUSIONS AND FUTURE WORK

We present two related approaches to dual control with M.P.C. that integrate experiment design and control. The controllers are based on minimization of parameter estimate error variance or maximization of information; different results are achieved although the approaches arise from similar ideas. Both controllers converge to a standard C.E. M.P.C. as uncertainty decreases or there is sufficient information generated, and the experimentation hence decreases with the diminishing need. Sensitivity analyses reveal that the D.M.P.C. algorithms can outperform non-dual C.E. M.P.C. on average, and that the tuning scheme gives intuitive results. The simulations also demonstrate excitation for improved performance without a sacrifice in output regulation and that the information-maximizing approach is the more efficient in terms of reducing output variance with less control effort. The simplicity of the formulation of the information-maximizing D.M.P.C. allows faster control input computation and analysis showing that in certain cases the

admissibility problem in adaptive control can be overcome with simple methods. The more intuitive quantities in the covariance matrix make the variance-minimizing controller a more appealing formulation.

Future work includes research on extensions to multivariable systems and time-varying parameters. As established in the discussion on formulation of the optimization problem, there is scope for global optimization provided the problems are formulated appropriately. We will investigate what can be gained from finding the global optima and to what extent the solution times grow impractical.

A1 PROOF OF PROPOSITION 1

Proof (Proposition 1). Without loss of generality we set $t_0 = 0$, and note that at $t = 0$ the prediction constraint (33b) is

$$\hat{y}(k+1 \mid t) = 0, \quad k = t, \dots, N-1 \tag{35}$$

because $\hat{\theta}(0) = \begin{bmatrix} 0, & 0 \end{bmatrix}^\top$. Furthermore,

$$R(0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad R(1) = \begin{bmatrix} y^2(0) & -y(0)u(0) \\ -y(0)u(0) & u^2(0) \end{bmatrix} \tag{36}$$

The objective function (33a) is in this situation reduced to

$$\begin{aligned} V_R(0) &= w_0 \sum_{k=0}^{0} \delta_e^k \sum_{i=1}^{2} \exp\bigl(-R_{ii}(k+1)\bigr) + \sum_{k=0}^{N-1} \bigl\{ w_1 \cdot 0 + w_2 u^2(k) \bigr\} \\ &= w_0 \exp(-y^2(0)) + w_0 \exp(-u^2(0)) + w_2 \sum_{k=0}^{N-1} u^2(k) \end{aligned} \tag{37}$$

It is then clear that the solution must satisfy $u(k) = 0$ for $k = 1, \dots, N-1$ since the predictor is unable to describe any relationship between the input and the output with the given estimates $\hat{\theta}(0)$. Hence, the only control input that will affect the objective function value is $u(0)$. If $u_{\min}$ and $u_{\max}$ are such that $u_{\min}^2 \leq \ln(w_0/w_2) \leq u_{\max}^2$, a minimizing initial control input must satisfy

$$\frac{\partial V_R(0)}{\partial u(0)} = -2w_0 u(0) \exp\bigl(-u^2(0)\bigr) + 2w_2 u(0) = 0 \tag{38}$$

which is obtained for either

$$u(0) = 0 \qquad \text{or} \qquad u^2(0) = \ln \frac{w_0}{w_2} \tag{39}$$

when $w_0 > w_2$. Since

$$\frac{\partial^2 V_R(0)}{\partial u(0)^2} = 2\big((2u^2(0)-1)w_0 \exp\big(-u^2(0)\big) + w_2\big) \tag{40}$$

is negative for $u(0) = 0$ and positive for $u^2(0) = \ln(w_0/w_2)$, $u(0) = 0$ is a maximizer and $u^2(0) = \ln(w_0/w_2)$ is a minimizer. Furthermore,

$$\begin{aligned} &\frac{\partial V_R(0)}{\partial u(0)} < 0 \ \ \forall \ \ u(0) < -\sqrt{\ln \frac{w_0}{w_2}}, \\ \text{and} \qquad &\frac{\partial V_R(0)}{\partial u(0)} > 0 \ \ \forall \ \ u(0) > \sqrt{\ln \frac{w_0}{w_2}} \end{aligned} \tag{41}$$

so there can be no minimizer at the edge of the feasible region. If $u_{\min}$ and $u_{\max}$ are not such that $u_{\min}^2 \leq \ln(w_0/w_2) \leq u_{\max}^2$, $u(0)$ will take a nonzero value at the bound of the feasible region, and the argument still holds.

The parameter estimates are updated by solving the least squares equation

$$R(1)\hat{\theta}(1) = \varphi(0)y(1) \tag{42}$$

Since rank $R(1) = 1$, this equation has no unique solution, except when $y(0) = 0$, which gives $\hat{b}_1(1) = b_1$. The choice of solution method, like Gaussian elimination or the pseudoinverse, determines the value of $\hat{\theta}(1)$ if one is desired when the information matrix is singular. We here choose to not update the parameter estimate when rank $R < n_p$ and pass $\hat{\theta}(t-1)$ and $R(t-1)$ to the optimization problem instead of $\hat{\theta}(t)$ and $R(t)$. The objective function will now be

$$V_R(1) = w_0 \exp(-y^2(1)) + w_0 \exp(-u^2(1)) + w_2 \sum_{k=1}^{N} u^2(k) \tag{43}$$

Now, $u(k) = 0$ for $k = 2, \ldots, N$ since the parameters were not updated because of insufficient information. Hence, the solution to the control problem is again $u^2(1) = \ln(w_0/w_2)$. The information matrix then becomes

$$R(2) = \begin{bmatrix} y^2(0) + y^2(1) & -y(0)u(0) - y(1)u(1) \\ -y(0)u(0) - y(1)u(1) & u^2(0) + u^2(1) \end{bmatrix} \tag{44}$$

Since $a_1 \neq 0$ and $u^2(1) = \ln(w_0/w_2)$, $y(1) \neq y(0)$ and $R(2)$ has full rank. The least-squares equation

$$R(2)\hat{\theta}(2) = \varphi(0)y(1) + \varphi(1)y(2) \tag{45}$$

then has a unique solution. Writing out (45) to obtain

$$\begin{bmatrix} y^2(0)+y^2(1) & -y(0)u_0-y(1)u_1 \\ -y(0)u_0-y(1)u_1 & u_0^2+u_1^2 \end{bmatrix} \begin{bmatrix} \hat{a}_1(2) \\ \hat{b}_1(2) \end{bmatrix} = \begin{bmatrix} a_1y^2(0)+a_1y^2(1)-b_1y(0)u_0-b_1y(1)u_1 \\ -a_1y(0)u_0-a_1y(1)u_1+b_1u_0^2+b_1u_1^2 \end{bmatrix} \quad (46)$$

reveals that

$$\hat{\theta}(2) = \begin{bmatrix} \hat{a}_1(2) \\ \hat{b}_1(2) \end{bmatrix} = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \quad (47)$$

is the unique solution. Thus, the algorithm has identified the two parameters in two time steps, the shortest possible time. □

As we can see from the proof, $V_R(0)$ and $V_R(1)$ are convex functions with minimizers satisfying $u(0) = 0$ and $u(1) = 0$, respectively, if $w_0 < w_2$. Hence, $w_0 > w_2$ is in this case a condition for a nonzero control signal in the face of a worst-case belief state $[R, \hat{\theta}]$.

AN MPC APPROACH TO DUAL CONTROL

Reference information for Paper B:
Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2013b). "An MPC approach to dual control." In: *Dynamics and Control of Process Systems*. Mumbai, India, pp. 69–74.

B

FOREWORD

This paper is a step from heuristic suboptimal dual control in the direction of an exact deterministic formulation of the dual control problem. We reformulate the first stage of the stochastic objective function and applying the reformulation beyond its temporally valid range. The resulting objective function thus differs from the dual objective, and the controller can be classified as semi-exact. The reformulation we develop Paper C can be applied to the controller we present here. The published version of this article contains some smaller errors that I have been corrected here.

ABSTRACT

We present a model predictive control (M.P.C.) approach to solve the dual adaptive control problem. The cost function minimized by the controller rewards probing the system for information when the parameter estimates are poor. The control algorithm is designed to handle poorly identified models and excites the system so that information can be gathered to achieve the optimal trade-off between process control and identification. This excitation is achieved without requiring the input to be persistently exciting; rather, the probing objective is based on an exact formulation of the expected value of the output error at the first time stage. The resulting expression is also used for the second time stage; this ensures that a proper trade-off between excitation and output regulation is maintained. The algorithm can be viewed as the merging of adaptive control with M.P.C. and its design can easily be implemented with modifications to an existing M.P.C. As an example we consider a

first-order linear process system with two unknown parameters. Our proposed algorithm probes the system even when the output error is small and quickly gathers enough information to correctly identify the unknown plant parameters.

1 INTRODUCTION

Maintaining a good model of a controlled plant is an important challenge in the process industries, model quality being one determining factor for the performance of a model predictive controller (M.P.C.). Performing experiments to generate data suitable for system identification is not always practical due to factors such as time constraints, the expertise needed, and expensive operational disruption. Model parameters are therefore commonly estimated using data collected during normal operation; however, recorded process data may be insufficiently informative for system identification or it can be difficult to locate the informative portions of a large data set.

Feldbaum (1961b) was the first to recognize that an optimal controller for a system with unknown parameters has two conflicting tasks: directing the output toward a reference, and exciting the system for learning purposes so that better control decisions can be made in the future. In its simplest form the dual control problem can be seen as finding the sequence of controls $\{u(t)\colon t = 0, 1, 2, \ldots\}$ that minimizes the control objective

$$J_\infty = \mathrm{E}\left[\sum_{t=0}^{\infty} y^2(t) + \mu u^2(t)\right] \tag{1}$$

given data collected up until $t = 0$, where $\mu \geq 0$ is a weighting parameter. This formulation does not include the notion of a model of the system. Thus we expect that the controller should be capable of exploring the system to find out the causal relationship between input u and output y. The dynamic properties of the system emerge as exploration takes place. This knowledge can be stored, for example in the terms of tables or as parameters in a linear or nonlinear model. We would expect the performance of the control system to improve as more data is gathered. A major contribution of the field of adaptive control was to show that the seemingly impossible problem of even stabilizing an unknown system could be solved for some classes of systems (Mårtensson, 1985). However, claims of optimality were not made. One important property

of a dual controller is that it is optimal in the sense that it finds the best trade-off or balance between control and excitation.

Dual control laws have been computed for very simple systems using dynamic programming (Åström and Helmersson, 1986), but the optimization problem quickly becomes intractable as the number of unknown parameters increases. Lee and Lee (2009) approach the problem of exponential growth in computational requirements using approximate dynamic programming; Bayard and Schumitzky (2010) developed a sampling-based approach to the dual control problem based on forward dynamic programming and particle filtering. Other approximations have been proposed to solve the problem of combined estimation and control. Some of these are motivated by adaptive control and they lead to the idea of generating exiting signals (Radenković and Ydstie, 1995). These approaches led to the idea of iterative adaptive control where instabilities observed in adaptive controllers were controlled by supplying sufficient excitation (Gevers, 2002). Around the same time the problem was approached from the point of view of Model Predictive Control and an approach termed Model Predictive Control and Identification (M.P.C.I.) was introduced by Genceli and Nikolaou (1996) and later extended in several publications by Nikolaou and coworkers, notably Shouche, Genceli, and Nikolaou (2002). The M.P.C.I. approach is based on parameterizing the input as a sum of sinusoids with prescribed frequencies and finding the optimal amplitudes; this leads to an input that is persistently exciting. While persistent excitation guarantees that parameter estimates converge exponentially when a recursive weighted least-squares algorithm is used (Johnstone et al., 1982), the excitation may be excessive. Marafioti, Stoican, et al. (2012) developed a persistently exciting M.P.C. using techniques similar to those used by Genceli and Nikolaou (1996).

The papers above show that the adaptive control and M.P.C. control approaches appear to be converging. However, the previous approaches that have been developed, whether from an M.P.C. or an adaptive control perspective, have not taken advantage of the fact that these perspectives can be blended by mixing M.P.C. and learning directly into the M.P.C. objective as originally proposed by Feldbaum when he developed the idea of dual control. While the original formulation of dual control may be intractable due to strong nonlinearity and the postulation of an infinite horizon, it seems reasonable to believe that good approximations can be obtained by considering finite horizon formulations. This approach has been successfully applied in the areas of

nonlinear M.P.C. and Moving Horizon Estimation in the form of finite horizon approximations to the infinite horizon problem, formulated within the context of dynamic programming. By combining the objectives into one, as suggested by the dual-control formulation, we develop an approach that does not require external excitation to excite the system unnecessarily since excitation is explicit in the sense that excitation is part of the control objective. In this case the objective above is restated so that

$$J_N(t) = \mathrm{E}\left[\sum_{k=0}^{N}\{y^2(t+k)^2 + \mu u^2(t+k) + J(t+N)\} \;\middle|\; \mathcal{Y}(t)\right] \qquad (2)$$

where $\mathrm{E}[\,\cdot \mid \mathcal{Y}(t)]$ denotes the conditional expectation given all data gathered up to time t, $\mathcal{Y}(t)$; $N \geq 1$ is the prediction horizon; and $J(t+N)$ is the cost to go that may not include the exploration component in order to make the problem computationally tractable. This control can then be computed using nonlinear programming and it will be implemented using the idea of receding horizon. Setting $N = 1$ give the classical adaptive control approach since and no exploration is provided. Extending the prediction horizon to 2 and beyond provides control signals that include a trade-off between exploration and control.

The term D.M.P.C. *as used here is not consistent with the use in Chapter 3 and Paper C.*

In this paper we develop a case study to illustrate the application of Dual Model-Predictive Control (D.M.P.C.) and we give particular attention to systems and formulations that cause certainty-equivalence based adaptive controllers to fail. The most common problem is an input gain estimated to be zero; other problems include models of dynamic systems where the dynamic component is missing and pole-zero cancellations in transfer functions.

This article is organized as follows: The control problem is described in Section 2, followed by a discussion of the control algorithm in Section 3. Sections 4 and 5 discuss implementation and provides an example, respectively. We conclude and discuss future work in Section 6.

2 PROBLEM FORMULATION

We consider control of stably invertible, single input, single output, linear, time-invariant systems in discrete time with unknown parameters, disturbed by a sequence of independent zero-mean Gaussian variables

with variance r. Systems of this type can be formulated as autoregressive process with exogenous input (A.R.X.):

$$y(t) + a_1 y(t-1) + \cdots + a_{n_a} y(t-n_a) = \\ b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) + v(t) \quad (3)$$

where t is a discrete time instant (integer), $y(t)$, $u(t)$, and $v(t)$ are the system output, input, and disturbance respectively at time t, and $a_1, \ldots, a_{n_a}$ and $b_1, \ldots, b_{n_b}$ are the unknown system parameters with $b_0 \neq 0$. The system is stably invertible so we can set $\mu = 0$ in the control objective and focus on controllers of the type

$$u(t) = \frac{1}{b_1}(a_1 y(t) + \cdots + a_{n_a} y(t-n_a+1) \\ - b_2 u(t-1) - \cdots - b_{n_b} u(t-n_b+1)) \quad (4)$$

This controller gives the optimal control when the parameters are precisely known. When they are not known we need to consider algorithms that estimate the parameters and implement controllers that in some way trade-off the learning and control.

In adaptive control we use the past information to estimate the parameters, for instance by solving a least squares problem. Subject to the assumptions made above we can guarantee that these estimates are optimal in the sense that they give the Best Linear Unbiased Estimate (BLUE) (Ljung, 1999). We can then use the estimated parameters to calculate the control law in Equation (4) with the knowledge that the estimates we use are the best estimates we can obtain using current information. This is the essence of the certainty-equivalence approach to adaptive control and it has been shown that this in fact can give optimal controls asymptotically. The approach however ignores the possibility of generating exploration signals that can speed up convergence by manipulating control signals to gain better knowledge about the parameters.

The A.R.X. system (3) can be written in the compact form

$$y(t) = \varphi^\top(t-1)\theta + v(t) \quad (5)$$

where

$$\theta = \left[a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b}\right]^\top$$

is a vector containing all system parameters, and

$$\varphi(t-1) = \left[-y(t-1), \ldots, -y(t-n_a), u(t-1), \ldots, u(t-n_b)\right]^\top \quad (6)$$

is a regression vector containing past inputs and outputs. We can also write (3) as

$$A(q^{-1})y(t) = B(q^{-1})u(t) \tag{7}$$

where A and B are polynomials in the backwards shift operator q^{-1}. It is necessary for controllability and observability that $A(q^{-1})$ and $B(q^{-1})$ are coprime.

We let $\mathcal{Y}(t)$ denote all inputs and outputs recorded up until the present time t. That is,

$$\mathcal{Y}(t) = \{u(t), u(t-1), \ldots, y(t), y(t-1), \ldots\} \tag{8}$$

The unknown parameters in θ are estimated with the recursive least-squares (R.L.S.) algorithm (Ljung, 1999)

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\big(y(t) - \varphi^\top(t-1)\hat{\theta}(t)\big) \tag{9a}$$

$$K(t) = P(t-1)\varphi(t-1)\big(\lambda + \varphi^\top(t-1)P(t-1)\varphi(t-1)\big)^{-1} \tag{9b}$$

$$P(t) = \big(I - K(t)\varphi^\top(t-1)\big)P(t-1)(1/\lambda) \tag{9c}$$

where $\hat{\theta}(t)$ is a vector of parameter estimates, $K(t)$ is the injection gain, λ is the forgetting factor, and $P(t)$ is a matrix of parameter estimate error covariances (when $\lambda = 1$, which is used in the remainder of this paper).

Based on the parameter estimate from the R.L.S. algorithm, we introduce the process model or one-step-ahead predictor

$$\begin{aligned}\hat{y}(t+1 \mid t) &= \mathrm{E}\big[y(t+1) \mid \mathcal{Y}(t)\big] \\ &= \varphi^\top(t)\hat{\theta}(t)\end{aligned} \tag{10}$$

2.1 *The dual control objective*

By setting $\mu = 0$ the objective of a dual controller can be be stated as minimizing

$$J_\infty = \sum_{k=0}^{\infty} \mathrm{E}\big[\big(y^*(t+1+k) - y(t+1+k)\big)^2 \mid \mathcal{Y}(t)\big] \tag{11}$$

where $y^*(t+1+k)$ is the output reference. That is, at time t the objective is to minimize future output error $(y^*(t+1+k) - y(t+1+k))^2$, $k = 0, 1, \ldots$, using the information gathered up to time t, $\mathcal{Y}(t)$. This is a fairly standard objective in optimal control; the main challenge comes from

the lack of a model capable of predicting future outputs $y(t+1+k)$, $k = 0, 1, \ldots$. As noted by Feldbaum (1961b), an optimal controller must in this case find the best trade-off between control and excitation.

The objective function (11) can be approximated in a variety of ways. The simplest approximation is to use a one-step-ahead cost function and then replace $y(t+1)$ by the model $\hat{y}(t+1)$; this gives a simple certainty-equivalence controller. Minimization of a one-step-ahead objective without replacing $y(t+1)$ gives a cautious controller (meaning less aggressive control when parameter uncertainty is high). A probing effect appears when the horizon is of length 2 or more, meaning that in addition to moving the output toward y^* the control signal attempts to reduce parameter uncertainty through excitation or probing.

3 CONTROLLER

3.1 *Objective reformulation*

We first rewrite the objective function (11) as

$$J_\infty = \mathrm{E}\big[\big(y^*(t+1) - y(t+1)\big)^2 \mid \mathcal{Y}(t)\big] + \sum_{k=1}^{\infty} \mathrm{E}\big[\big(y^*(t+1+k) - y(t+1+k)\big)^2 \mid \mathcal{Y}(t)\big] \quad (12)$$

The first term is from here on referred to as J_1; that is,

$$J_1 = \mathrm{E}\big[\big(y^*(t+1) - y(t+1)\big)^2 \mid \mathcal{Y}(t)\big] \quad (13)$$

We now subtract and add the predictor (10) and get

$$J_1 = \mathrm{E}\big[(y^*(t+1) - \hat{y}(t+1) + \hat{y}(t+1) - y(t+1))^2 \mid \mathcal{Y}(t)\big] \quad (14)$$

In order to simplify the notation, the condition in the expectation is no longer explicitly stated. The cost (14) can be expanded to

$$J_1 = \mathrm{E}\big[\big(y^*(t+1) - \hat{y}(t+1)\big)^2 - 2\big(y^*(t+1) - \hat{y}(t+1)\big)\big(\varphi^\top(t)\tilde{\theta}(t) + v(t+1)\big) + \big(\varphi^\top(t)\tilde{\theta}(t) - v(t+1)\big)^2 \mid \mathcal{Y}(t)\big] \quad (15)$$

where $\tilde{\theta}(t) = \theta - \hat{\theta}(t)$. The first of the three terms in (15) is deterministic; the second term is zero since both

$$\mathrm{E}\big[\tilde{\theta}(t)\big] = \mathrm{E}\big[\theta - \hat{\theta}(t)\big] = \hat{\theta}(t) - \hat{\theta}(t) = 0 \quad (16)$$

and $\mathrm{E}[v(t+1)] = 0$. Since $\tilde{\theta}(t)$ and $v(t+1)$ are independent,

$$P(t) = \mathrm{E}\big[\tilde{\theta}(t)\tilde{\theta}^\top(t)\big], \qquad \text{and} \qquad r = \mathrm{E}\big[(v(t+1))^2\big] \tag{17}$$

we can write the third term in (15) as

$$\varphi^\top(t)P(t)\varphi(t) + r \tag{18}$$

We then have that (15) can be written

$$J_1 = \big(y^*(t+1) - \varphi^\top(t)\hat{\theta}(t)\big)^2 + \varphi^\top(t)P(t)\varphi(t) + r \tag{19}$$

given $\mathcal{Y}(t)$. The reformulation of the first stage cost highlights the trade-off between output control and *caution*, represented by the first and second terms in (19), respectively. That is, a large $P(t)$ matrix leads to a reward for small signals, meaning that a cautious input reduces the cost. Note that there are no stochastic variables in the one-step-ahead cost (19), which means that a myopic optimal controller can be found by minimizing J_1.

In order to obtain a probing effect, we use an identical cost function for the second time stage. That is, we set the total cost of the first two stages to

$$\begin{aligned} J_2 := \sum_{k=0}^{1} \big\{ &\big(y^*(t+1+k) - \varphi^\top(t+k)\hat{\theta}(t+k)\big)^2 \\ &+ \varphi^\top(t+k)P(t+k)\varphi(t+k) + r\big\} \end{aligned} \tag{20}$$

The probing effect is achieved since a certain choice of input $u(t)$ will reduce the elements of the predicted future parameter estimate error covariance matrix $P(t+1)$ such that the cost is reduced. In other words, an input $u(t)$ that probes or excites the system in a manner that reveals information about the parameters and thereby reduces their estimate uncertainty is rewarded.

The infinite horizon objective function (11) can now be approximated in the following manner:

$$\begin{aligned} J_\infty \approx \sum_{k=0}^{N-1} \big\{ &\big(y^*(t+1+k) - \varphi^\top(t+k)\hat{\theta}(t+k)\big)^2 \\ &+ \varphi^\top(t+k)P(t+k)\varphi(t+k) + r\big\} \\ &+ \sum_{k=N}^{\infty} \mathrm{E}\big[\big(y^*(t+1+k) - y(t+1+k)\big)^2 \,\big|\, \mathcal{Y}(t)\big] \end{aligned} \tag{21}$$

Another simplification of the cost function is obtained by truncating the horizon from infinity to some finite number $N > 1$. This simplified objective function can then be written

$$J_N := \sum_{k=0}^{1} \{ \big(y^*(t+1+k) - \varphi^\top(t+k)\hat{\theta}(t+k)\big)^2 + \varphi^\top(t+k)P(t+k)\varphi(t+k) + r \} + \sum_{k=2}^{N-1} \mathrm{E}\big[\big(y^*(t+1+k) - y(t+1+k)\big)^2 \mid \mathcal{Y}(t)\big] \quad (22)$$

Since we apply a model-predictive-control strategy, the minimizer of the objective function (subject to the specified constraints) will be found at every time instant. The solution to the problem contains an open-loop optimal control sequence; the first input in this sequence is applied to the plant. Hence, it can be argued that the first time stage in the objective is the most important term.

For this reason, we approximate the second sum using the model or predictor (10) instead of the expected value of the output and call the approximated cost V_N. We also add a cost of input usage. This gives

The input cost is missing from both the left-hand side and the first sum on the right-hand side of Equation (23) in the published version of this paper.

$$J_N + \sum_{k=0}^{N-1} w_3 u^2(t+k) \approx V_N := \sum_{k=0}^{1} \{ \big(y^*(t+1+k) - \varphi^\top(t+k)\hat{\theta}(t+k)\big)^2 + \varphi^\top(t+k)P(t+k)\varphi(t+k) + r + w_3 u^2(t+k) \} + \sum_{k=2}^{N-1} \{ w_2 \big(y^*(t+1+k) - \hat{y}(t+1+k)\big)^2 + w_3 u^2(t+k) \} \quad (23)$$

where w_2 and w_3 are cost weights. Note that

$$\hat{y}(t+1+k \mid t) = \varphi^\top(t+k)\hat{\theta}(t) \quad (24)$$

meaning the output is predicted over the horizon using the current parameter estimate.

3.2 *Constraints*

A second contribution in our proposed algorithm is adding an adapted version of the recursive least-squares algorithm (9) as a set of equality

constraints in the optimization problem. We have pursued similar approaches in Heirung, Ydstie, and Foss (2012b) and Heirung, Ydstie, and Foss (2013a), although to different ends. Adding these estimation equations gives the optimization solver a measure of how the input sequence affects parameter estimate uncertainty through predicting future covariances $P(k)$. Before discussing the addition of the modified estimation equations as constraints, we discuss the certainty-equivalence principle used in our controller, as well some standard M.P.C. constraints.

Using the latest parameter estimate $\hat{\theta}(t)$, the output predictor Equation (10) is added as a constraint at all future time instants k in the prediction horizon as

$$\hat{y}(k+1 \mid t) = \hat{\varphi}^\top(k)\hat{\theta}(t), \quad k \in \{t, \dots, t+N-1\} \tag{25}$$

Since the regressor contains predicted future outputs when $k > t$, we use

$$\hat{\varphi}^\top(k) = \left[u(k), \dots, u(k-n_b+1), -\hat{y}(k), \dots, -\hat{y}(k-n_a+1)\right]^\top \tag{26}$$

(cf. (6)) where

$$\hat{y}(i) := y(i) \qquad \text{if} \quad i \leq t,\ i \in \{k-n_a+1, \dots, k\} \tag{27}$$

meaning $\hat{y}(i)$ is a recorded output (as opposed to a prediction) if i corresponds to a current or past time instant.

For $k = t+1$, we set $\hat{\varphi}(k-1) = \varphi(t)$, which contains past inputs and outputs and is passed from the simulator to the optimization problem. Note that in (27), $\hat{\theta}(t)$ is the most recent parameter estimate produced by the estimation algorithm (9) during simulation at time t. The predictor (27) is added as an equality constraint to the optimization problem. Since the output is predicted in this manner our controller becomes a certainty-equivalence type M.P.C. Finally, note that $\hat{\theta}(t)$ is a constant in the optimization problem solved at time t; this means that (27) is a linear equality constraint.

An M.P.C. commonly includes bounds on inputs and outputs as inequality constraints ("box constraints") as part of the online optimization problem. The input bounds are formulated

$$u_{\min} \leq u(k) \leq u_{\max}, \qquad k \in \{t, \dots, t+N-1\} \tag{28}$$

The minimum and maximum values $u_{\min}$ and $u_{\max}$ are usually based on either hardware limitations or similar physical constraints. Since the

model contains unknown parameters, the constraints on the output $y(t)$ cannot be directly included in the optimization problem. Instead, we simply add the bounds the predicted outputs:

$$y_{\min} \leq \hat{y}(k) \leq y_{\max}, \qquad k \in \{t+1, \ldots, t+N\} \tag{29}$$

There is no guarantee that predicted outputs $\hat{y}(k)$ being feasible with respect to the bounds implies that the actual output $y(t)$ would stay within the limits if some open-loop optimal input sequence $u^*(t), \ldots, u^*(t+N-1)$ were implemented on the real process. However, since the parameter estimates improve over time, the box constraint (29) is more likely to ensure that

$$y_{\min} \leq y(t) \leq y_{\max}$$

We now add the necessary constraints for the optimization algorithm to have a way of predicting $P(t+1)$ based on its choice of $u(t)$. Since the future parameter estimates can not be predicted (without future data), we add only equations (9b)–(9c) as equality constraints. We only need P at $t+1$ and hence include the equations in the form

$$K(t+1) = P(t)\hat{\varphi}(t)\big(\lambda + \hat{\varphi}^\top(t)P(t)\hat{\varphi}(t)\big)^{-1} \tag{30a}$$

$$P(t+1) = \big(I - K(t+1)\hat{\varphi}^\top(t)\big)P(t)(1/\lambda) \tag{30b}$$

3.3 *The online optimization problem*

Based on the above discussion, we can now state the full optimization problem solved online at every time stage t:

$$\begin{aligned}\min_{u(k)} V_N = \sum_{k=0}^{1} \Big\{ &\big(y^*(t+1+k) - \varphi^\top(t+k)\hat{\theta}(t+k)\big)^2 \\ &+ \varphi^\top(t+k)P(t+k)\varphi(t+k) + r\Big\} \\ + \sum_{k=2}^{N-1} \Big\{ &w_2\big(y^*(t+1+k) - \hat{y}(t+1+k)\big)^2 \\ &+ w_3 u^2(t+k)\Big\}\end{aligned} \tag{31a}$$

subject to

$$\hat{y}(k \mid t) = \hat{\varphi}^\top(k-1)\hat{\theta}(t), \qquad k = t+1, \ldots, t+N \tag{31b}$$

$$K(t+1) = P(t)\hat{\varphi}(t)\big(\lambda + \hat{\varphi}^\top(t)P(t)\hat{\varphi}(t)\big)^{-1} \tag{31c}$$

$$P(t+1) = \big(I - K(t+1)\hat{\varphi}^\top(t)\big)P(t)(1/\lambda) \tag{31d}$$

$$u_{\min} \leq u(k) \leq u_{\max}, \qquad k = t, \ldots, t+N-1 \tag{31e}$$

$$y_{\min} \leq \hat{y}(k) \leq y_{\max}, \qquad k = t+1, \ldots, t+N \tag{31f}$$

An important consequence of the nonlinear equality constraints (31c)–(31d), which represent the estimation algorithm, is that the optimization problem becomes a nonconvex nonlinear programming (N.L.P.) problem.

The optimization problem has a number of variables fixed; the values of $\varphi(t)$ (except for $u(t)$), $\hat{\theta}(t)$, and $P(t)$ are all passed from the simulation to the optimization algorithm and are fixed parameters in the optimization problem. Note that $\varphi(t+1)$ being fixed implies that $y(t), \ldots, y(t-n_a)$ and $u(t), \ldots, u(t-n_b)$ are all given and fixed in the N.L.P. problem. This means that the controller is neither output nor state feedback; rather, we have feedback from a hyperstate containing the current output $y(t)$, the parameter estimates $\hat{\theta}(t)$, the covariance matrix $P(t)$, as well as a history of inputs and outputs.

4 IMPLEMENTATION

Our implementation of the algorithm in written in MATLAB and GAMS (GAMS Development Corporation, 2012) with IPOPT (Cervantes et al., 2000) as the N.L.P. solver for the online optimization problem. IPOPT is a state-of-the-art open-source interior-point N.L.P. solver, which exploits the sparsity of the N.L.P. and is capable of solving large-scale problems. The system is simulated in MATLAB; GAMS is called at every iteration and returns a (locally) open-loop optimal input sequence $u^*(t), \ldots, u^*(t+N-1)$. The example runs on a standard laptop computer and the optimization problems are solved reasonably fast with solution times ranging from 0.21 to 1.04 C.P.U. seconds. The implementation is not written with a focus on speed of execution.

5 APPLICATION EXAMPLE

As an application example we consider a tank containing a solution of salt dissolved in water. Let $y(t)$ be the amount of salt in the tank, $V(t)$ be the volume of solution in the tank, and $c_{\text{out}}(t) = x(t)/V(t)$ be the concentration in the tank. A solution with constant known concentration c_{in} flows into the tank at a rate $q_{\text{in}}(t) = k_1 u(t)$ where $u(t) \in [0, 1]$ is a valve setting and k_1 is an unknown constant. Solution flows out of the tank at a constant unknown rate q_{out} and with variable concentration $c_{\text{out}}(t)$.

This system can be modeled with the linear first-order differential equation

$$\dot{y} = -\frac{q_{\text{out}}}{V} y(t) + c_{\text{in}} k_1 u(t) \tag{32}$$

A sketch of the system is shown in Figure 1.

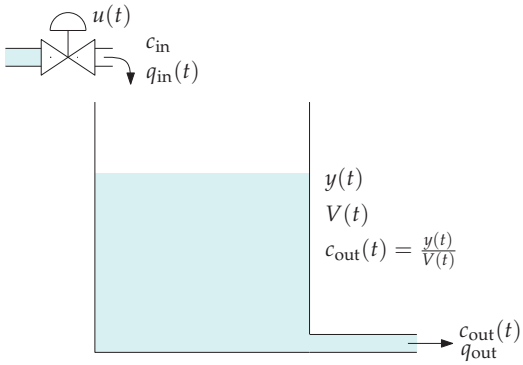


Figure 1: Sketch of the mixture example problem.

We can discretize Equation (32) to obtain a discrete-time equation of the form

$$y(t) = -a_1 y(t-1) + b_1 u(t) \tag{33}$$

We want to control the inlet flow rate to keep the concentration at y^* while identifying the unknown parameters (a_1 and b_1 in Equation (33)).

In the numerical example we use the following parameter values: $y^* = 2.5$, $y(0) = 2.00$, $a_1 = -0.60$, $b_1 = 5.50$, $\hat{a}_1(0) = -0.10$, $\hat{b}_1(0) = 0.00$, $P(0) = 1 \times 10^3 I$, $N = 6$, $w_2 = 1.00 \times 10^{-2}$, $w_3 = 1.00$, $u_{\min} = 0$, $u_{\max} = 1$, $y_{\min} = 0$, $y_{\max} = 10$; the system is simulated for 10 time steps and no noise is used in the simulation.

The results are presented in Figure 2. The most significant result is that the Dual M.P.C. generates a control input that probes or excites the system, even though the output error is very small at $t = 0$. The probing leads to both parameter estimates converging in 2 time steps; the diagonal elements of the covariance matrix decrease accordingly. The output exhibits some oscillation due to the probing, but settles fairly quickly close to the reference of 2.5. The figure also contains results from applying a certainty-equivalence (C.E.) M.P.C. to the same control problem. The initial input gain estimate is zero ($\hat{b}_1(0) = 0$), which causes the C.E. M.P.C. to generate a zero-input since the system appears uncontrollable. This again leads to a complete lack of information about the input gain and the situation is never resolved. This problem is avoided by the Dual M.P.C. since it knows that excitation is necessary for good future performance.

6 CONCLUSIONS AND FUTURE WORK

We present a new model predictive controller exhibiting dual features in that it actively probes the system for information when the parameter estimates are poor. The main contribution is that the cost function is based on an exact reformulation of the output error for the first time step. This resulting expression is also used for the second stage cost and provides the probing reward in the cost function. The objective function is evaluated by the N.L.P. solver with the help of constraints based on a recursive least-squares algorithm for parameter estimation.

Future work includes extending the controller to larger systems with more unknown parameters as well as developing a version that can handle time-varying parameters. The algorithm will also be extended to multivariable systems. Analysis of stability and convergence properties will be investigated at a later point in time. We will also analyze the nonconvexity of the optimization problem and investigate the effect of finding the global solution.

The trade-off between an update constraint for $R(t+1)$ with a more complicated objective function and the update constraint for $P(t+1)$ and a simpler objective function (done here) in the N.L.P. will be the topic of a future paper.

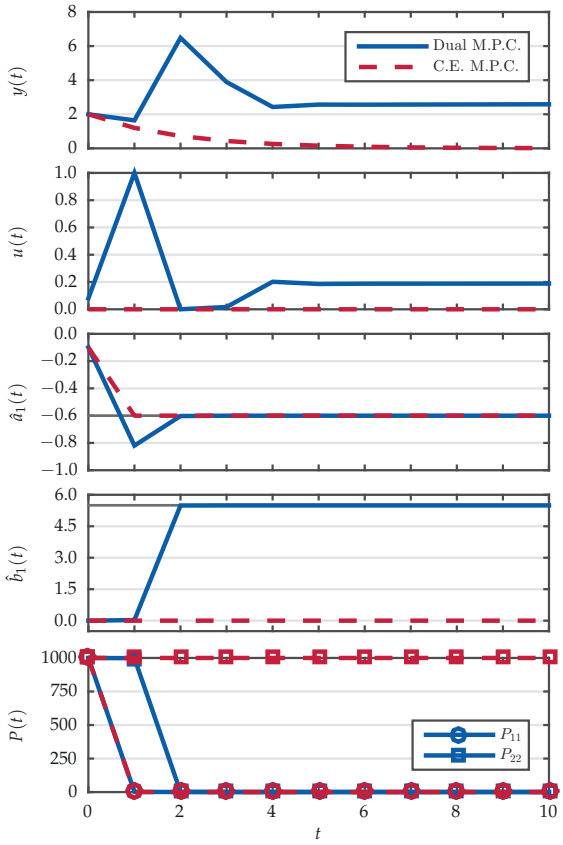


Figure 2: Results from the numerical example. Results obtained with our Dual M.P.C. are in solid blue and results from using a standard certainty-equivalence (C.E.) M.P.C. are in dashed red. The input gain is initially estimated to be zero, causing the C.E. M.P.C. to fail; the Dual M.P.C. excites the system and overcomes the problem.

DUAL ADAPTIVE MODEL-PREDICTIVE CONTROL

Reference information for Paper C:
Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2015b). "Dual adaptive model-predictive control." *Under revision for resubmission to Automatica*.

C

FOREWORD

This paper is based on ideas similar to those used to derive the semi-exact dual controller in Paper B. We here consider orthonormal-basis-function (O.B.F.) models instead of autoregressive processes with exogenous input (A.R.X.), which enables exact reformulation into deterministic form. The concept of future information is also formalized, and we include probabilistic output constraints in the deterministic formulation. The reformulation allows several interesting insights, in particular how nominal control and uncertainty reduction must be combined to obtain the dual control. We further reformulate the problem to a quadratically-constrained quadratic program to facilitate fast solution. Some of the ideas that form the foundation for this paper were presented in simpler form for finite-impulse-response (F.I.R.) systems, without comprehensive proofs, in Heirung, Ydstie, and Foss (2015c).

ABSTRACT

We present an adaptive dual model-predictive controller (D.M.P.C.) that uses current and future parameter-estimate errors to minimize expected output error by optimally combining probing for uncertainty reduction with nominal control. Our novel approach relies on orthonormal-basis-function models to derive expressions for the predicted distributions for the output and unknown parameters, conditional on the future input sequence. Propagating the exact future statistics allows us to reformulate the original stochastic problem into a deterministic equivalent that illustrates the dual nature of the optimal control but is nonlinear and nonconvex. We further reformulate the nonlinear deterministic problem to pose an equivalent quadratically-constrained quadratic-

programming (Q.C.Q.P.) problem that we solve efficiently using state-of-the-art algorithms, providing the exact solution to the probabilistically constrained finite-horizon dual control problem. We implement the adaptive D.M.P.C. by solving the Q.C.Q.P. at each sampling time in a receding-horizon fashion; the adaptation is a result of updating the parameter estimates used by the D.M.P.C. to decide the control input. We demonstrate the application of D.M.P.C. to a single-input single-output (SISO) system with unknown parameters. In the simulation example, the parameter estimates converge quickly and the excitation vanishes with increasing accuracy and precision of the estimates.

1 INTRODUCTION

This paper addresses the problem of optimal control and learning in the context of stochastic systems with parametric uncertainty and probabilistic constraints. Dual control, as introduced by Feldbaum (1961b), is the optimal control under decision-relevant, reducible uncertainty. This control has dual tasks in that it must explore (or excite) the system in order to generate informative data that can reduce the uncertainty while simultaneously directing the outputs toward the reference signal.

Using data to progressively reduce uncertainty is often framed as a learning process. Learning in control has primarily been studied in the field of adaptive control, which in general involves adjusting the controller parameters based on observation of the uncertain system's response to the input (Åström and Wittenmark, 1995). Most adaptive control algorithms are passively adaptive in the sense that learning takes place only as a side effect of the control action and is not actively pursued. This means that the controller learns from normal operating data, which can contain very little information if, for instance, the process is operating at a steady state. The data can be made more informative by actively perturbing the process, also known as excitation (Mareels, Bitmead, et al., 1987), probing (Bar-Shalom, 1981), experimentation (Gevers and Ljung, 1986), exploration (Sutton and Barto, 1998), or active learning (Tse and Bar-Shalom, 1973). Any adaptive controller that is designed to increase the information content in the generated data (or increase the level of excitation), and thus in some sense improve learning, can be classified as an actively adaptive controller. An early example of one such algorithm was developed by Tse and Bar-Shalom (1973). Although the most immediate motivation for active learning is improved performance, this is not necessarily achieved, for instance,

if the design leads to excessive excitation. Active learning can also be motivated by the goal of inferring a more accurate and precise estimate of physical parameters, or more generally, finding model parameters that better fit the data.

Adaptive model-predictive control M.P.C. has received relatively little attention in the literature (Mayne, 2014). As with most types of adaptive control, adaptive M.P.C.s may suffer from signals that are insufficiently exciting for the controller or model parameters to converge, which may lead to problems such as bursting (Anderson, 1985), pole-zero cancellations or inadmissible models (Mareels and Polderman, 1996), and turn-off (Wieslander and Wittenmark, 1971). One way of approaching this issue is to design a controller that actively explores the plant by ensuring a certain level of excitation, either constantly or when needed.

Shouche, Genceli, Vuthandam, et al. (1998) combined M.P.C. with system identification for autoregressive systems with an exogenous input (A.R.X. systems) and added a constraint to the controller that ensures persistent excitation for improved identification. Marafioti, Bitmead, and Hovd (2014) presented a related approach for systems with a finite impulse response (F.I.R.), but accounted for the input history and required only the first predicted output be persistently exciting. Similarly, Larsson, Rojas, et al. (2015) developed an M.P.C. for output-error models that guarantees a specified performance level through the addition of an experiment-design constraint that ensures sufficiently rich signals for model re-identification. Common to these approaches is the requirement that future signals generate a prescribed minimum amount of information or excitation, leading to a constant level of excitation which may be unnecessary in some cases.

Several proposed controllers generate excitation without a specific requirement. Rather, they include a function of information or uncertainty in the M.P.C. cost function and optimize this function together with standard control objectives. Heirung, Foss, and Ydstie (2015) proposed and compared two such formulations, one minimizing a function of parameter-estimate variance and one maximizing a function of generated information, and showed that both controllers converge to a standard adaptive certainty-equivalence M.P.C. as the uncertainty is reduced, but that the excitation can lead to improved performance. Weiss and Di Cairano (2014) developed a robust M.P.C. for polytopic linear difference inclusion models that rewards informative signals through the addition of a cost term that increases with the prediction error, result-

ing in diminishing excitation reward with decreasing prediction errors. Tanaskovic et al. (2014) suggest the addition of an exploring property as a possible extension of their adaptive M.P.C. for finite-impulse-response (F.I.R.) systems. Their approach involves modifying the nominally optimal input sequence by solving a second-stage optimization problem with the objective of decreasing the set of possible models at the next time step. Common to all of these approaches is that the excitation is a consequence of either a heuristic modification or addition to the controller, motivated by the assumption that the resulting excitation will improve overall performance. While this type of algorithm may work very well in practice and improve performance over passive-learning approaches (see Heirung, Foss, and Ydstie (2015)), the excitation does not implicitly arise as a consequence of directly optimizing for performance, which is the case for "ideal" dual control in the sense of Feldbaum (1961b). That is, these approaches are all suboptimal or approximate dual controllers by definition, and they illustrate an important distinction: superimposing excitation on a nominally optimal control signal does not produce the true optimal input, and hence does not give a dual control.

The act of exciting the process is often seen as conflicting with the control objective (see, e.g., Tse and Bar-Shalom (1973)); however, based on the derivations in this article we argue that this is not a correct interpretation and show that excitation is an intrinsic part of the optimal control. That excitation is an inextricable part of the input in dual control means it cannot be derived or rewarded heuristically. Furthermore, the excitation and the nominal output error-minimization are not conflicting goals that can be traded off against each other; rather, they are inseparable components that together constitute the optimal control. Uncertainty reduction cannot be sacrificed for increased control performance.

Feldbaum (1961b) identified (stochastic) dynamic programming as an appropriate solution method for dual control problems in his pioneering papers on integrating active learning with multistage decision making under uncertainty, providing a foundation for much of the later work in stochastic adaptive control. Åström and Helmersson (1986) use solve a dual control problem for a scalar integrator with one unknown parameter, which is among the first reported results obtained with numerical dynamic programming. At the time, they spent 180 C.P.U. hours to obtain the solution using a time horizon of 30 samples (Åström, 1983). Although today's computers are far superior, the "curse of dimension-

ality" prevents dynamic programming from being a viable solution approach for dual control problems. This has motivated investigation of the use of modern approximate methods such as approximate dynamic programming (Lee and Lee, 2009) and forward dynamic programming with particle filtering (Bayard and Schumitzky, 2010). These methods directly approximate the dynamic programming equations rather than coming up with an approximate formulation of the problem.

The optimal strategy for control of systems with resolvable, decision-relevant uncertainty is dual control. This definition, however, is only meaningful for developing algorithms provided the true system can be modeled perfectly. From the mid-1970s, the field of system identification increasingly abandoned the notion of a model set that contains the true system in favor of the more realistic goal of finding the "best approximate model" (Gevers, 2006). A better definition of the dual control, which is the definition we use in this work, is then the optimal control with respect to an a priori chosen model set with resolvable, decision-relevant uncertainty. As with all types of optimal control, the extent to which a dual-control algorithm is able to imitate the optimal strategy is limited by the veracity of the model set, including the assumptions made on parameter uncertainty, if it exists. We thus distinguish between two definitions of dual control: the one that is optimal with respect to the chosen model set, defined above, and the one obtainable with a yet-to-be-invented algorithm combined a with model set that contains the true system and perfectly captures the uncertainty. Although the latter represents the most perfect (and unobtainable) form of dual control, we call the former "ideal" dual control in order to distinguish it from an approximate dual control policy obtained, for instance, through approximation of the Bellman equation. While the importance of this distinction may depend on the problem at hand, it is of particular interest when using dual control with data-driven models but also relevant for high-fidelity models based on first principles. Note that in the simulation examples we present to demonstrate the dual controller we develop, the true system is in fact contained in the model set.

In this article we derive an adaptive dual M.P.C. (D.M.P.C.) for systems modeled with orthonormal basis functions with parametric uncertainty and process noise. We formulate a stochastic optimal-control problem for minimizing expected performance cost, which involves the use of future information to evaluate the conditional expected future tracking error. We transform this stochastic problem into an equivalent deterministic form that enables exact evaluation of both the objective

function and the probabilistic output constraints. This reformulation demonstrates that a particular form of uncertainty reduction is a necessary part of optimizing performance and provides insight into how the uncertainty must be reduced. The reformulation relies on the future decisions to propagate the exact conditional distributions over the prediction horizon, meaning the learning outcome of the input can be exactly evaluated. Consequently, the solution to the problem is an input sequence that provides optimal expected performance by combining excitation with control of the nominal output. We transform this problem into a quadratically-constrained quadratic-programming (Q.C.Q.P.) problem that can be solved efficiently using state-of-the-art solvers. At every sampling time we use the updated conditional distribution for the unknown parameters and solve the optimal-control problem over a finite horizon to obtain the dual control. The controller is therefore of indirect-adaptive receding-horizon type. The proposed D.M.P.C. ensures that the system is sufficiently excited for accurate and precise parameter estimation but does not require a persistently exciting input. The excitation term in the deterministic formulation of the dual objective can be interpreted as a time-varying L-optimal experiment design criterion (see Gevers, Bombois, et al. (2011)).

In addition to providing the foundation for a practical algorithm for dual control, the results clarify aspects of the dual control problem. Most importantly, our reformulated objective function answers the question of how to formulate excitation and output tracking objectives that when minimized together result in optimal expected tracking. The reformulation also demonstrates how the two objectives are not conflicting, and that it is not optimal to eliminate all uncertainty. For the systems we consider here, the dual control is obtained by minimizing a unique combination of two specific functions of nominal output-tracking error and parameter-estimate variance. Minimizing a different combination of two functions of nominal output error and parameter-estimate variance does not give dual control, although the resulting control may achieve almost the same performance in some cases. The reformulated objective can furthermore guide the design of approximate or suboptimal dual controllers for systems where we cannot derive deterministic expressions for the stochastic objective function. Some of the results in this article are generalizations of the work of Heirung, Ydstie, and Foss (2015c), a portion of which are given there without proof. Primarily, we here consider a more general system type and allow probabilistic output constraints.

This article is organized as follows: we formulate the stochastic control problem in Section 2. In Section 3 we review and develop the statistical foundation needed for parameter estimation, for the control problem reformulations, and for propagating the necessary moments of the system output and unknown parameters; we then formulate the stochastic optimal-control problem (P). The main contributions of the paper are in Section 4, where we state and prove a set of theorems and corollaries that are necessary to reformulate the stochastic optimal-control problem as the equivalent deterministic form (P′) and subsequently transform this formulation into the Q.C.Q.P. problem (P″). Section 5 contains the dual control algorithm, followed by a simulation example in Section 6. We provide a brief discussion of the results in Section 7. In Section 8 we conclude the paper and provide some thoughts for future work.

2 FORMULATION OF THE DUAL CONTROL PROBLEM

We consider the output tracking problem for a class of systems that can be written in the form

$$\varphi(t+1) = A\varphi(t) + Bu(t) \tag{1a}$$

$$y(t) = \theta^\top \varphi(t) + v(t) \tag{1b}$$

where $\varphi(t)$ is a deterministic regression vector whose elements are functions of past control inputs (deterministic decision variables) u, and $A \in \mathbb{R}^{n_p \times n_p}$ and $B \in \mathbb{R}^{n_p}$ are known matrices determined by the basis functions. The variable $y(t)$ is the plant output and $v(t)$ an additive, stationary process disturbance assumed to be a sequence of independent and identically distributed Gaussian random variables with zero mean and variance r. The vector $\theta \in \mathbb{R}^{n_p}$ contains the unknown parameters,

$$\theta = \left[\theta_1, \theta_2, \ldots, \theta_{n_p}\right]^\top \tag{2}$$

where $\{\theta_j\}_{j=1}^{n_p}$ are drawn from a multivariate Gaussian distribution at time $t = t_0$ with mean $\hat{\theta}(t_0)$ and covariance $P(t_0)$. The model (1b) is often referred to as a linear regression. Note that if some of the parameters in Equation (1b) are known we can write

$$y(t) = \theta^\top \varphi(t) + v(t) + \mu(t-1) \tag{3}$$

where $\mu(t-1)$ is known at time t.

The system (1) is a linear, time-invariant, single-input, single-output (SISO) system, and the formulation here includes systems modeled by orthogonal basis functions (O.B.F.s). Heuberger, Van den Hof, and Wahlberg (2005) provide a comprehensive overview of models based on O.B.F.s. The most well-known member of this class is the F.I.R. model; other common formulations include the Laguerre (Wahlberg, 1991) and Keutz (Wahlberg, 1994) models; see also Wahlberg and Mäkilä (1996), as well as Finn, Wahlberg, and Ydstie (1993) for a combination of the F.I.R. and Laguerre structures. Appendix A3 contains some examples of the A and B matrices for specific model types. We assume that the pair (A, B) in the system (1) is controllable and stabilizable.

A standard definition of information recorded up to and including time t is the set of all past decisions and measurements:

$$\mathcal{Y}(t) = \{u(t), u(t-1), \ldots, u(t_0), y(t), y(t-1), \ldots, y(t_0)\} \tag{4}$$

We use this to define the parameter estimate

$$\hat{\theta}(t) := \mathrm{E}[\theta \mid \mathcal{Y}(t)] = \left[\hat{\theta}_1(t), \ldots, \hat{\theta}_{n_p}(t)\right]^\top$$

and the parameter-estimate-error covariance matrix $P(t) := \mathrm{E}[\tilde{\theta}(t)\tilde{\theta}^\top(t) \mid \mathcal{Y}(t)]$ with $\tilde{\theta}(t) := \theta - \hat{\theta}(t)$. The definition of information in Equation (4) may also include the initial values for the parameter estimates and the covariances, $\hat{\theta}(t_0)$ and $P(t_0)$. Let $\{u(k \mid t)\}_{k=t}^{t+N-1}$ be a sequence of future control inputs decided at time t; the decision $u(k \mid t)$ may later change, so note that in general, $u(k \mid t') \neq u(k \mid t)$ with $k \geq t' \geq t+1$. Note that we have $k \geq t$ unless otherwise noted. The output predictor is

$$\begin{aligned} \hat{y}(k+1 \mid t) &= \mathrm{E}[y(k+1) \mid \mathcal{Y}(t)], \qquad k \geq t \\ &= \hat{\theta}^\top(t)\varphi(k+1 \mid t) \end{aligned} \tag{5}$$

where $\varphi(k \mid t)$ is the decision regressor defined such that

$$\varphi(k+1 \mid t) = A\varphi(k \mid t) + Bu(k \mid t), \qquad k \geq t \tag{6}$$

where $\varphi(t \mid t) := \varphi(t)$ and $u(t \mid t) := u(t)$.

The results in this work depend on extending the definition of $\mathcal{Y}(t)$ to include future decisions (first introduced by Heirung, Ydstie, and Foss (2015c)). We define

$$\begin{aligned} \mathcal{Y}(k \mid t) = \{ &\underbrace{u(k \mid t), u(k-1 \mid t), \ldots, u(t+1 \mid t), u(t \mid t)}_{\text{anticipated information, } k \geq t}, \\ &\underbrace{u(t-1), u(t-2), \ldots, u(t_0), y(t), y(t-1), \ldots, y(t_0)}_{\text{past information}}\} \end{aligned} \tag{7}$$

Note that the future input sequence $\{u(i \mid t)\}_{i=t}^{i=k}$ consists of exogenous decisions and is deterministic, in contrast to other predicted variables that are subject to uncertainty, such as the predicted outputs $\hat{y}(k \mid t)$ that are based on $\mathcal{Y}(t)$ (see Equation (5)). Hence, $\mathcal{Y}(k \mid t)$ contains no information from the plant beyond time t. Similarly, the future regressors $\varphi(k \mid t)$ in Equation (6) contain deterministic decisions only, as opposed to uncertain signals like future plant outputs or noise terms.

We define the finite-horizon performance cost

$$J_N(t) = \sum_{k=t}^{t+N-1} \big\{ \mathrm{E}\big[(y(k+1) - y^*(k+1 \mid t))^2 \mid \mathcal{Y}(k \mid t)\big] + w_2 u^2(k \mid t) + w_3 (\Delta u(k \mid t))^2 \big\} \quad (8)$$

where $1 \leq N \leq \infty$ is the length of the prediction horizon, $y^*(k+1 \mid t)$ is the output reference sequence at time t, $\Delta u(k \mid t) := u(k \mid t) - u(k-1 \mid t)$ is the control input change, with $u(t-1 \mid t) := u(t-1)$, and $w_2 \geq 0$, $w_3 \geq 0$ are tuning weights. Since J_N may be unbounded for an infinite horizon length, we use only finite N in this paper. Note that we take the conditional expectation of the squared output-tracking error with respect to the future decision sequence. If we use only the current information $\mathcal{Y}(t)$ we obtain the output cost

$$\sum_{k=t}^{t+N-1} \big\{ (\hat{y}(k+1|t) - y^*(k+1|t))^2 + \varphi^\top(k+1|t) P(t) \varphi(k+1|t) + r \big\} \quad (9)$$

as a special case of (8). This cost function does not reward excitation since the future parameter covariances are not included; see Theorem 6 and the discussion in Appendix A2. The inclusion of $P(t)$ does provide a rationale for caution, or a form of robustness with respect to parameter-estimate errors, since large uncertainties heavily penalize current and future inputs $u(k \mid t)$, $k \in \{t, t+1, \ldots, t+N-1\}$. However, this formulation does not anticipate information and as a result the controller does not excite the system to reduce future uncertainty. Hence, there is no dual effect. In contrast, the objective function we develop in this paper takes future decisions into account and as result captures the dual nature of the optimal control.

We want to minimize $J_N(t)$ in (8) subject to a set of constraints. The outputs and the decision variables are given by the model (5)–(6). We further specify probabilistic output constraints and deterministic bounds on the magnitude and rate of change of the inputs. The resulting

moving-horizon stochastic optimal-control problem we solve at time t is

$$\min J_N(t) \tag{P.a}$$

subject to

$$\varphi(k+1 \mid t) = A\varphi(k \mid t) + Bu(k \mid t) \tag{P.b}$$
$$\hat{y}(k+1 \mid t) = \hat{\theta}^\top(t)\varphi(k+1 \mid t) \tag{P.c}$$
$$\Pr[y_{\min} \leq y(k+1) \mid \mathcal{Y}(k \mid t)] \geq p_{y,\min} \tag{P.d}$$
$$\Pr[y(k+1) \leq y_{\max} \mid \mathcal{Y}(k \mid t)] \geq p_{y,\max} \tag{P.e}$$
$$u_{\min} \leq u(k \mid t) \leq u_{\max} \tag{P.f}$$
$$\Delta u(k \mid t) = u(k \mid t) - u(k-1 \mid t) \tag{P.g}$$
$$\Delta u_{\min} \leq \Delta u(k \mid t) \leq \Delta u_{\max} \tag{P.h}$$
$$k \in \{t, t+1, \ldots, t+N-1\} \tag{P.i}$$
$$\varphi(t \mid t),\ u(t-1 \mid t),\ \hat{\theta}(t),\ P(t) \text{ given} \tag{P.j}$$

where $\varphi(t \mid t) = \varphi(t)$, $u(t-1 \mid t) = u(t-1)$. Furthermore, $\Pr[y_{\min} \leq y(k+1) \mid \mathcal{Y}(k \mid t)]$ and $\Pr[y(k+1) \leq y_{\max} \mid \mathcal{Y}(k \mid t)]$ are the probabilities, conditioned on $\mathcal{Y}(k \mid t)$, that the predicted outputs stay above $y_{\min}$ and below $y_{\max}$, respectively, and $p_{y,\min}, p_{y,\max} \in (0,1)$ are the specified constraint satisfaction probabilities. Hence, the constraints (P.d)–(P.e) are probabilistic (chance) constraints on the output. The stochastic programming problem is solved with $\hat{\theta}(t)$, $\varphi(t)$, $u(t-1)$, and $P(t)$ as parameters; the solution includes an optimal sequence of predicted control inputs $\{u^o(k \mid t)\}_{k=t}^{t+N-1}$.

We now describe the evolution and prediction of the parameter-estimate statistics at each time t and then discuss a how we evaluate the objective function $J_N(t)$ and the chance constraints (P.d)–(P.e) to transform (P) into a tractable, deterministic problem for $N < \infty$.

3 UNCERTAINTY PROPAGATION, PARAMETER STATISTICS, AND ESTIMATION

With the class of systems we consider here, in which the regressor $\varphi(t)$ is deterministic, we can formulate the following theorem for propagation of the parameter statistics.

THEOREM 1. *For a system of the form* (1)*, the conditional covariance of* θ*,* $P(k \mid t)$*, can be propagated forward in time with* $k \geq t$ *through the recursive relations*

$$K(k+1|t) = P(k|t)\varphi(k+1|t)(r+\varphi^\top(k+1|t)P(k|t)\varphi(k+1|t))^{-1} \quad (10a)$$

$$P(k+1|t) = \big(I - K(k+1|t)\varphi^\top(k+1|t)\big)P(k|t) \quad (10b)$$

given $\mathcal{Y}(k \mid t)$*.*

Proof. The proof of Theorem 1 is identical to a textbook proof of the Kalman Theorem (see Åström and Wittenmark (1995)), except we here state the result for future time. The equivalence follows from the deterministic nature of $\mathcal{Y}(k \mid t)$ and the fact that $\varphi(k \mid t)$ is a function of $\mathcal{Y}(k \mid t)$, which means that $K(k \mid t)$ and $P(k \mid t)$ are deterministic for $k \geq t$ given the recorded outputs and the past and deterministic future decisions in $\mathcal{Y}(k \mid t)$. □

For the systems we consider here, $P(t)$, the covariance of θ at time t given $\mathcal{Y}(t)$, becomes as a special case of $P(k \mid t)$ in Equation (10b). With $\mathcal{Y}(t)$ available we can also determine the conditional mean $\hat{\theta}(t)$. These two quantities fully describe the temporal evolution of the conditional distribution of θ at time t, as described in the following theorem.

THEOREM 2. *For a system of the form* (1)*, the conditional distribution of* θ *given* $\mathcal{Y}(t)$ *is Gaussian with mean* $\hat{\theta}(t)$ *and covariance* $P(t)$ *satisfying the recursive equation set*

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\big(y(t) - \hat{\theta}^\top(t-1)\varphi(t)\big) \quad (11a)$$

$$K(t) = P(t-1)\varphi(t)\big(r\lambda + \varphi^\top(t)P(t-1)\varphi(t)\big)^{-1} \quad (11b)$$

$$P(t) = \big(I - K(t)\varphi^\top(t)\big)P(t-1)(1/\lambda) \quad (11c)$$

with $\lambda = 1$ *and the initial conditions* $\hat{\theta}(t_0)$ *and* $P(t_0)$*.*

Proof. The proof is found many standard texts on stochastic and adaptive control; see, e.g., Åström and Wittenmark (1995). □

Note that with $\lambda = 1$ the equation set (11) can be interpreted as a Kalman filter for estimating the state of a system with the constant state variable θ (no dynamics and no process noise) and output equation (1b).

We now derive (after Ljung (1999)) a standard recursive least-squares algorithm for estimating θ using past data. Let $R(t)$ be the information matrix

$$R(t) = \sum_{k=t_0+1}^{t} r_R^{-1}\lambda^{t-k}\varphi(k)\varphi^\top(k) + r_R^{-1}\lambda^{t-t_0}R(t_0), \quad t > t_0 \quad (12)$$

with the forgetting factor $\lambda \in (0, 1]$ and $R(t_0) = P^{-1}(t_0)$ given; $r_R = r$ when $r \neq 0$ and $r_R = 1$ when $r = 0$. $R(t)$ is better expressed recursively, as

$$R(t) = \lambda R(t-1) + r_R^{-1}\varphi(t)\varphi^\top(t), \qquad t > t_0 \tag{13a}$$

We can now update the conditional mean (the parameter estimate) $\hat{\theta}(t)$ according to

$$\hat{\theta}(t) = \hat{\theta}(t-1) + R^{-1}(t)\varphi(t)\big(y(t) - \hat{\theta}^\top(t-1)\varphi(t)\big) \tag{13b}$$

The inverse of the information matrix $R(t)$ is the covariance matrix $P(t)$. Instead of calculating the inverse $P(t) = R^{-1}(t)$ at every time t we update the least squares parameter estimate recursively using (11) as an estimation procedure where we set $r = r_R$. The equation set (11) can be derived directly from (13) using the matrix-inversion lemma (see Ljung (1999)).

Note the possibility of bias in the parameter estimate $\hat{\theta}(t)$ when using the algorithm for identification under closed-loop control. The issue of possible bias is beyond the scope of this paper, although sufficiently complex controllers should in general allow experiments that generate enough information to eliminate bias (Ljung, 1999). Van den Hof, Heuberger, and Bokor (1995) analyze bias and variance errors in system identification with generalized orthonormal basis functions.

4 REFORMULATION TO A DETERMINISTIC Q.C.Q.P.

The process (1) belongs to a class of systems in which the output is dependent on past inputs but not past outputs. This means the future regressors are deterministic since they do not contain future outputs, which are stochastic variables. We have the freedom to choose the future decisions, meaning we effectively decide $\mathcal{Y}(k \mid t)$ for any $k \geq t$. As demonstrated above, the future covariances $P(k \mid t)$ are deterministic functions of these deterministic future inputs. The following theorem, which we use to reformulate the optimal-control problem (P), is a consequence of this feature.

THEOREM 3. *For a stochastic process of the form* (1),

$$\begin{aligned}\mathrm{E}\big[y^2(k+1) \mid \mathcal{Y}(k \mid t)\big] = \hat{y}^2(k+1 \mid t) \\ + \varphi^\top(k+1 \mid t)P(k \mid t)\varphi(k+1 \mid t) + r\end{aligned} \tag{14}$$

for all $k \geq t$.

Proof. From the definition of variance,

$$\mathrm{E}\big[y^2(k+1) \mid \mathcal{Y}(k \mid t)\big] = \mathrm{E}\big[\big(\theta^\top \varphi(k+1) + v(k+1)\big)^2 \mid \mathcal{Y}(k \mid t)\big] \quad (15)$$

Expanding the square gives

$$\begin{aligned}\mathrm{E}\big[y^2(k+1) \mid \mathcal{Y}(k \mid t)\big] = \mathrm{E}\big[&\varphi^\top(k+1)\theta\theta^\top\varphi(k+1) \\ &+ v^2(k+1) + 2v(k+1)\theta^\top\varphi(k+1) \mid \mathcal{Y}(k \mid t)\big]\end{aligned} \quad (16)$$

Since $\varphi(k+1)$ is deterministic, this reduces to

$$\begin{aligned}&\mathrm{E}\big[y^2(k+1) \mid \mathcal{Y}(k \mid t)\big] \\ &\quad = \varphi^\top(k+1 \mid t)\,\mathrm{E}\big[\theta\theta^\top \mid \mathcal{Y}(k \mid t)\big]\varphi(k+1 \mid t) \\ &\qquad + \mathrm{E}\big[v^2(k+1) \mid \mathcal{Y}(k \mid t)\big] \\ &\qquad + 2\,\mathrm{E}\big[v(k+1)\theta^\top \mid \mathcal{Y}(k \mid t)\big]\varphi(k+1 \mid t)\end{aligned} \quad (17)$$

where the second term is the variance of the noise, r, and last term is zero since $v(k+1)$ and θ are uncorrelated. Using Theorem 5, we then have

$$\begin{aligned}\mathrm{E}\big[y^2(k+1) \mid \mathcal{Y}(k \mid t)\big] = \varphi^\top(k+1 \mid t)\big(&\hat{\theta}(t)\hat{\theta}^\top(t) \\ &+ P(k \mid t)\big)\varphi(k+1 \mid t) + r\end{aligned} \quad (18)$$

which after expanding the parenthesis can be written

$$\begin{aligned}\mathrm{E}\big[y^2(k+1) \mid \mathcal{Y}(k \mid t)\big] = &\big(\hat{\theta}^\top(t)\varphi(k+1 \mid t)\big)^2 \\ &+ \varphi^\top(k+1 \mid t)P(k \mid t)\varphi(k+1 \mid t) + r\end{aligned} \quad (19)$$

Inserting the model (5) for $\hat{\theta}^\top(t)\varphi(k+1 \mid t)$ completes the proof. □

The following corollary extends Theorem 3 to tracking of a time-varying output reference $y^*(t)$.

COROLLARY 1. *For a stochastic process of the form* (1)*,*

$$\begin{aligned}&\mathrm{E}[(y(k+1) - y^*(k+1 \mid t))^2 \mid \mathcal{Y}(k \mid t)] \\ &\qquad = (\hat{y}(k+1 \mid t) - y^*(k+1 \mid t))^2 \\ &\qquad\quad + \varphi^\top(k+1 \mid t)P(t)\varphi(k+1 \mid t) + r\end{aligned} \quad (20)$$

for all $k \geq t$.

Proof. We have that

$$\begin{aligned}
&\mathrm{E}\big[(y(k+1) - y^*(k+1 \mid t))^2 \mid \mathcal{Y}(k \mid t)\big] \\
&\quad = \mathrm{E}\big[y^2(k+1) - 2y(k+1)y^*(k+1 \mid t) \\
&\qquad\qquad + (y^*(k+1 \mid t))^2 \mid \mathcal{Y}(k \mid t)\big] \\
&\quad = \mathrm{E}\big[y^2(k+1) \mid \mathcal{Y}(k \mid t)\big] \\
&\qquad - 2\,\mathrm{E}\big[y(k+1) \mid \mathcal{Y}(k \mid t)\big] y^*(k+1 \mid t) \\
&\qquad\qquad + (y^*(k+1 \mid t))^2 \qquad (21)
\end{aligned}$$

Using Theorem 3, we can write

$$\begin{aligned}
&\mathrm{E}\big[(y(k+1) - y^*(k+1 \mid t))^2 \mid \mathcal{Y}(k \mid t)\big] \\
&\quad = \hat{y}^2(k+1 \mid t) + \varphi^\top(k+1 \mid t) P(k \mid t) \varphi(k+1 \mid t) + r \\
&\qquad - 2\hat{y}(k+1 \mid t) y^*(k+1 \mid t) + (y^*(k+1 \mid t))^2 \\
&\quad = (\hat{y}(k+1 \mid t) - y^*(k+1 \mid t))^2 \\
&\qquad\qquad + \varphi^\top(k+1 \mid t) P(k \mid t) \varphi(k+1 \mid t) + r \qquad (22)
\end{aligned}$$

which is the desired result. □

We now define the future conditional output variance $\sigma_y^2(k+1 \mid t) := \mathrm{E}\big[(y(k+1) - \hat{y}(k+1 \mid t))^2 \mid \mathcal{Y}(k \mid t)\big]$; the following corollary follows from Theorem 3 and states the explicit expression for the output variance at time k predicted at time t ($k \geq t$).

COROLLARY 2. *For a stochastic process of the form* (1)*, the future output variance* $\sigma_y^2(k+1 \mid t)$ *predicted at time* t *is*

$$\sigma_y^2(k+1 \mid t) = \varphi^\top(k+1 \mid t) P(k \mid t) \varphi(k+1 \mid t) + r \qquad (23)$$

for all $k \geq t$ *given* $\mathcal{Y}(k \mid t)$*.*

Proof. We let $\sigma_y^2(k+1 \mid t)$ denote the anticipated conditional output variance and by definition have that

$$\begin{aligned}
\sigma_y^2(k+1 \mid t) &= \mathrm{E}\big[(y(k+1) - \hat{y}(k+1 \mid t))^2 \mid \mathcal{Y}(k \mid t)\big] \\
&= \mathrm{E}\big[y^2(k+1) \mid \mathcal{Y}(k \mid t)\big] - \hat{y}^2(k+1 \mid t) \qquad (24)
\end{aligned}$$

which follows from expanding the square and collecting the terms in the same manner as in the proof of Lemma 1. The expression for $\sigma_y^2(k+1 \mid t)$ is then obtained directly from Theorem 3. □

4.1 *Evaluating the objective function*

Corollary 1 allows the stochastic objective (8) to be reformulated into the equivalent deterministic function in the following theorem. We consider this one of the main contributions of the paper.

THEOREM 4. *For a stochastic process of the form* (1)*, the objective function* $J_N(t)$ *in Equation* (8) *can be written*

$$J_N(t) = \sum_{k=t}^{t+N-1} \{ (\hat{y}(k+1 \mid t) - y^*(k+1 \mid t))^2 + \varphi^\top(k+1 \mid t) P(k \mid t) \varphi(k+1 \mid t) + r + w_2 u^2(k \mid t) + w_3 (\Delta u(k \mid t))^2 \} \quad (25)$$

Proof. This reformulation follows directly from Corollary 1. □

Note that this objective function rewards probing (excitation) provided $N \geq 2$, since $P(t \mid t) = P(t)$ is not a variable, but rather a parameter or initial condition. On the other hand, $P(t+1 \mid t)$ is a variable that represents the covariance one step ahead, which has to be included if a reduction of uncertainty is to be rewarded.

4.2 *Evaluation of the chance constraints*

We now transform the probabilistic constraints (P.d)–(P.e) into deterministic form using Corollary 2. From Åström and Wittenmark (1995) and implied from Theorem 2 we know that the conditional distribution of $y(t+1)$ given $\mathcal{Y}(t)$ is Gaussian with mean and variance

$$\hat{y}(t+1) = \hat{\theta}^\top(t) \varphi(t+1) \quad (26a)$$

$$\sigma_y^2(t+1) = \varphi^\top(t+1) P(t) \varphi(t+1) + r \quad (26b)$$

Similarly, the future conditional distribution of $y(k+1)$ given $\mathcal{Y}(k \mid t)$, $k \geq t$, is Gaussian with mean

$$\hat{y}(k+1 \mid t) = \hat{\theta}^\top(t) \varphi(k+1) \quad (27a)$$

and covariance

$$\sigma_y^2(k+1 \mid t) = \varphi^\top(k+1 \mid t) P(k \mid t) \varphi(k+1 \mid t) + r \quad (27b)$$

which follows from Corollary 2.

Hence, the constraint

$$\Pr[y_{\min} \leq y(k+1) \mid \mathcal{Y}(k \mid t)] \geq p_{y,\min} \tag{P.d}$$

is equivalent to the deterministic constraint

$$y_{\min} \leq \hat{y}(k+1 \mid t) - s_{\min}\sigma_y(k+1 \mid t) \tag{28}$$

where $\Phi(s_{\min}) = p_{y,\min}$ and Φ is the cumulative distribution function (C.D.F.) for the standard normal distribution. Similarly,

$$\Pr[y(k+1) \leq y_{\max} \mid \mathcal{Y}(k \mid t)] \geq p_{y,\max} \tag{P.e}$$

corresponds to

$$\hat{y}(k+1 \mid t) + s_{\max}\sigma_y(k+1 \mid t) \leq y_{\max} \tag{29}$$

where $\Phi(s_{\max}) = p_{y,\max}$. The parameters $s_{\min}$ and $s_{\max}$ are determined once and offline.

Taken together, the probabilistic constraints (P.d)–(P.e) can be written in deterministic form as

$$y_{\min} + s_{\min}\sigma_y(k+1 \mid t) \leq \hat{y}(k+1 \mid t) \leq y_{\max} - s_{\max}\sigma_y(k+1 \mid t) \tag{30}$$

Note that specifying $p_{y,\min} = p_{y,\max} = 0.5$ simplifies the control problem since the chance constraints then are linear. This is because $\Phi(s = 0) = 0.5$, which renders the quadratic variance expression (27b) superfluous since the constraints in this case correspond to requiring that the output means $\hat{y}(k+1 \mid t)$ stay within the bounds.

Extending this formulation to time-varying probabilities $p_{y,\min}(k \mid t)$ and $p_{y,\max}(k \mid t)$ is trivial. The only change required is that a larger number of equations $\Phi(s_{\min}(k \mid t)) = p_{y,\min}(k \mid t)$ and $\Phi(s_{\max}(k \mid t)) = p_{y,\max}(k \mid t)$ must be solved once and offline.

4.3 *The deterministic optimal-control problem*

The objective function $J_N(t)$ in Equation (25) can now be minimized by augmenting the constraint set of (P) with (10) from Corollary 1 and (23) from Theorem 2, and replacing the probabilistic constraints (P.d)–(P.e) with the deterministic equivalents (28)–(29). The result is a deterministic optimal-control problem that is equivalent to (P):

$$\min J_N(t) \tag{P′.a}$$

subject to

$$\varphi(k+1 \mid t) = A\varphi(k \mid t) + Bu(k \mid t) \tag{P′.b}$$
$$\hat{y}(k+1 \mid t) = \hat{\theta}^\top(t)\varphi(k+1 \mid t) \tag{P′.c}$$
$$\sigma_y^2(k+1 \mid t) = \varphi^\top(k+1 \mid t)P(k \mid t)\varphi(k+1 \mid t) + r \tag{P′.d}$$
$$K(k+1 \mid t) = P(k \mid t)\varphi(k+1 \mid t) \times \big(r_R + \varphi^\top(k+1 \mid t)P(k \mid t)\varphi(k+1 \mid t)\big)^{-1} \tag{P′.e}$$
$$P(k+1 \mid t) = \big(I - K(k+1 \mid t)\varphi^\top(k+1 \mid t)\big)P(k \mid t) \tag{P′.f}$$
$$\hat{y}(k+1 \mid t) \geq y_{\min} + s_{\min}\sigma_y(k+1 \mid t) \tag{P′.g}$$
$$\hat{y}(k+1 \mid t) \leq y_{\max} - s_{\max}\sigma_y(k+1 \mid t) \tag{P′.h}$$
$$u_{\min} \leq u(k \mid t) \leq u_{\max} \tag{P′.i}$$
$$\Delta u(k \mid t) = u(k \mid t) - u(k-1 \mid t) \tag{P′.j}$$
$$\Delta u_{\min} \leq \Delta u(k \mid t) \leq \Delta u_{\max} \tag{P′.k}$$
$$k \in \{t, t+1, \ldots, t+N-1\} \tag{P′.l}$$
$$\varphi(t \mid t),\ u(t-1 \mid t),\ \hat{\theta}(t),\ P(t \mid t),\ s_{\min},\ s_{\max} \text{ given} \tag{P′.m}$$

where $\varphi(t \mid t) = \varphi(t)$, $u(t-1 \mid t) = u(t-1)$, and $P(t \mid t) = P(t)$. The constraints (P′.c)–(P′.d) and (P′.e)–(P′.f) deterministically propagate the complete statistics (the two first moments) of the system output and the variance (the second moment) of the parameters, respectively.

Note that if the uncertainty represented by $P(t)$ goes to zero, the formulation of the output cost in the objective $J_N(t)$ in (25) converges to the certainty-equivalence output cost (48) (except for the constant term r). This implies that the excitation reward induced by the parameter uncertainty vanishes as the uncertainty is resolved. This property highlights the main idea of adaptive certainty-equivalence control, which is the assumption that the parameter estimates give the correct representation of the plant dynamics; see Åström and Wittenmark (1995).

Although the solution to (P′) exactly minimizes

$$J_N(t) = \sum_{k=t}^{t+N-1} \big\{ \mathrm{E}\left[(y(k+1) - y^*(k+1 \mid t))^2 \mid \mathcal{Y}(k \mid t) \right] + w_2 u^2(k \mid t) + w_3 (\Delta u(k \mid t))^2 \big\} \tag{8}$$

over the finite horizon $k \in \{t, t+1, \ldots, t+N-1\}$, the feasible area is nonconvex because of the inclusion of the nonlinear equality constraints (P′.e)–(P′.f) and (P′.d) for propagating the parameter variance and output variance, respectively. This motivates investigation of refor-

mulation approaches that facilitate solving the optimal-control problem. We consider this reformulation the other main contribution of the paper.

4.4 *Reformulation as a Q.C.Q.P.*

In order to reduce the complexity of the deterministic optimal-control problem (P′) we introduce a set of new variables. First, we define the scaled, noise-invariant, predicted information matrix $\bar{R}(k+1 \mid t)$,

$$\bar{R}(k+1 \mid t) := r_R R(k+1 \mid t) \tag{31}$$

which is then recursively expressed as (cf. (13))

$$\bar{R}(k+1 \mid t) = \bar{R}(k \mid t) + \varphi(k+1 \mid t)\varphi^\top(k+1 \mid t) \tag{32}$$

Accordingly, the covariance matrix $P(k+1 \mid t)$ (which is positive definite) can be expressed in terms of $\bar{R}(k+1 \mid t)$ as $P(k+1 \mid t) = r_R \bar{R}^{-1}(k+1 \mid t)$. By introducing the variable $z(k \mid t)$ defined through

$$\bar{R}(k \mid t) z(k \mid t) = \varphi(k+1 \mid t) \tag{33}$$

or equivalently $z(k \mid t) := r_R^{-1} P(k \mid t)\varphi(k+1 \mid t)$, we write

$$\begin{aligned} &\varphi^\top(k+1 \mid t) P(k \mid t) \varphi(k+1 \mid t) \\ &\quad = r_R \varphi^\top(k+1 \mid t) \bar{R}^{-1}(k \mid t) \varphi(k+1 \mid t) \\ &\quad = r_R \varphi^\top(k+1 \mid t) z(k \mid t) \end{aligned} \tag{34}$$

Using this equation we can simplify both the objective function (25) from Theorem 4 and the predicted output variance constraint (P′.d). Furthermore, the highly nonlinear uncertainty-propagation constraints (P′.e)–(P′.f) can be replaced with the quadratic (bilinear) equations (32) and (33).

The objective function (25) is then equivalent to

$$\begin{aligned} J_N(t) = \sum_{k=t}^{t+N-1} \big\{ &(\hat{y}(k+1 \mid t) - y^*(k+1 \mid t))^2 \\ &+ r_R \varphi^\top(k+1 \mid t) z(k \mid t) + r \\ &+ w_2 u^2(k \mid t) + w_3 (\Delta u(k \mid t))^2 \big\} \end{aligned} \tag{35}$$

Accordingly, the optimal-control problem (P′) is equivalent to

$$\min J_N(t) \tag{P″.a}$$

subject to

$$\varphi(k+1 \mid t) = A\varphi(k \mid t) + Bu(k \mid t) \tag{P″.b}$$
$$\hat{y}(k+1 \mid t) = \hat{\theta}^\top(t)\varphi(k+1 \mid t) \tag{P″.c}$$
$$\sigma_y^2(k+1 \mid t) = r_R\varphi^\top(k+1 \mid t)z(k \mid t) + r \tag{P″.d}$$
$$\bar{R}(k+1 \mid t) = \bar{R}(k \mid t) + \varphi(k+1 \mid t)\varphi^\top(k+1 \mid t) \tag{P″.e}$$
$$\bar{R}(k \mid t)z(k \mid t) = \varphi(k+1 \mid t) \tag{P″.f}$$
$$\hat{y}(k+1 \mid t) \geq y_{\min} + s_{\min}\sigma_y(k+1 \mid t) \tag{P″.g}$$
$$\hat{y}(k+1 \mid t) \leq y_{\max} - s_{\max}\sigma_y(k+1 \mid t) \tag{P″.h}$$
$$u_{\min} \leq u(k \mid t) \leq u_{\max} \tag{P″.i}$$
$$\Delta u(k \mid t) = u(k \mid t) - u(k-1 \mid t) \tag{P″.j}$$
$$\Delta u_{\min} \leq \Delta u(k \mid t) \leq \Delta u_{\max} \tag{P″.k}$$
$$k \in \{t, t+1, \ldots, t+N-1\} \tag{P″.l}$$
$$\varphi(t \mid t),\ u(k-1 \mid t),\ \hat{\theta}(t),\ \bar{R}(t \mid t),\ s_{\min},\ s_{\max} \text{ given} \tag{P″.m}$$

where $\varphi(t \mid t) = \varphi(t)$, $u(k-1 \mid t) = u(t-1)$, and $\bar{R}(t \mid t) = r_R P^{-1}(t)$.

The parameters r_R and r both appear in the objective function in this formulation, and can there be interpreted as parameters determining the performance cost incurred by the noise sequence. In addition, r_R can be interpreted as the optimal choice for how uncertainty reduction, represented by the term $\varphi^\top(k+1 \mid t)z(k \mid t)$, should be weighted against reducing the nominal output tracking error.

The formulation (P″) with (35) is nonlinear, but all nonlinearities are now quadratic; that is, either bilinear (e.g., $\varphi^\top(\cdot)z(\cdot)$) or square (e.g., $\sigma_y^2(\cdot)$). If we collect all the variables in (P″) in a single vector x (the variable $R(t+N \mid t)$ is not needed to obtain the solution and can be omitted), the objective function can be written in the form

$$J_N(t) = x^\top Q_0(t)x + \alpha_0^\top(t)x \tag{36}$$

where $x^\top Q_0(t)x$ contains all square ($\hat{y}^2(\cdot)$, $w_2u^2(\cdot)$, and $w_3(\Delta u(\cdot))^2$) and bilinear ($r_R\varphi^\top(\cdot)z(\cdot)$) terms, and $\alpha_0^\top(t)x$ contains all linear terms ($-2y^*(\cdot)\hat{y}(\cdot)$). The square and bilinear terms are represented in $Q_0(t)$ by diagonal and off-diagonal entries, respectively. The constant terms ($\hat{y}^2(\cdot)$ and r) do not affect the solution and need not be included in the implementation.

The constraints (P″.d) and (P″.e) contain linear and quadratic terms (square as well as bilinear), while the quadratic terms in (P″.f) are all bilinear; hence, these constraints can all be written in the form

$$\beta_{\min,i} \leq x^\top Q_i x + \alpha_i^\top x \leq \beta_{\max,i}, \qquad i = 1, \ldots, n_q \tag{37}$$

where n_q is the number of quadratic constraints. Since there are both square and bilinear terms in (P″.d) and (P″.e), the corresponding Q_i matrices have nonzero elements both on and off the diagonals. All quadratic terms in (P″.f) are bilinear (no square terms), so the corresponding Q_i matrices have all zeros on their diagonals. We set $\beta_{\min,i} = \beta_{\max,i} = r$ for the constraints that correspond to (P″.d); similarly, we set $\beta_{\min,i} = \beta_{\max,i} = 0$ for the constraints that correspond to (P″.e) and (P″.f) to obtain equality constraints with a zero right-hand side. The remaining constraints are all linear and can be written in the form

$$\beta_{\min,i} \leq \alpha_i^\top(t)x \leq \beta_{\max,i}, \qquad i = n_q + 1, \ldots, n_q + n_\ell \tag{38}$$

where n_ℓ is the number of linear constraints.

The optimal-control problem (P″) can then be written as a standard quadratically-constrained quadratic programming (Q.C.Q.P.) problem (cf. Misener and Floudas (2013)):

$$\min_x \; x^\top Q_0(t)x + \alpha_0^\top(t)x \tag{39a}$$

subject to

$$\beta_{\min,i} \leq x^\top Q_i x + \alpha_i^\top x \leq \beta_{\max,i}, \qquad i = 1, \ldots, n_q \tag{39b}$$

$$\beta_{\min,i} \leq \alpha_i^\top(t)x \leq \beta_{\max,i}, \qquad i = n_q{+}1, \ldots, n_q{+}n_\ell \tag{39c}$$

$$\text{some elements of } x \text{ given} \tag{39d}$$

The structures of the matrices Q_i ($i = 0, \ldots, n_q$) and the vectors α_i ($i = 0, \ldots, n_q + n_\ell$), $\beta_{\min,i}$, and $\beta_{\max,i}$ ($i = 0, \ldots, n_q + n_\ell$), depend on the organization of variables in the vector x.

The optimal-control problem (P″) increases moderately in complexity with the number of unknown model parameters n_p and the length of the prediction horizon N. There are $n_p N$ bilinear terms of the form $\varphi_j(\cdot)z_j(\cdot)$ (j denotes vector element) in the objective function (35), while there are $n_p^2 N$ quadratic terms in each of the uncertainty-propagation constraints (P″.e) and (P″.f). The output-variance constraint (P″.d) contains N square terms and $n_p N$ bilinear terms. The symmetric nature of the quadratic equality constraints can be exploited in implementation to reduce this number, but the quadratic growth cannot be avoided.

Although the quadratic equality constraints imply that (P″) is a nonconvex Q.C.Q.P. problem, there are several algorithms that efficiently solve Q.C.Q.P. problems to ϵ-global optimality (Tawarmalani and Sahinidis, 2002). Two such algorithms are BARON (Tawarmalani and Sahinidis, 2005) and GLOMIQO (Misener and Floudas, 2013); the latter reference also provides a good overview of the Q.C.Q.P. problem class.

5 DUAL CONTROL ALGORITHM

We now propose a dual control algorithm based on Model Predictive Control (M.P.C., see Mayne et al. (2000)). Just as in standard M.P.C., we find the control by solving an optimal-control problem at each sampling instant in a receding-horizon fashion. With our D.M.P.C. algorithm, the dual control input at time t, $u(t)$, is contained in the solution to the finite-horizon stochastic optimal-control problem (P), which we obtain by solving the equivalent Q.C.Q.P. problem (P″). The solution is the control sequence $\{u^o(k \mid t)\}_{k=t}^{t+N-1}$, the first element of which is implemented as the control input: $u(t) = u^o(t \mid t)$. In contrast to standard M.P.C., where there is feedback from the system state or state estimate, our adaptive D.M.P.C. depends on feedback from the hyperstate defined as $(\varphi(t), u(t-1), \hat{\theta}(t), P(t))$ (Åström and Wittenmark, 1995). Furthermore, (P″) is not a true open-loop problem since the uncertainty predictions implicitly anticipate a closed loop, which we discuss in Section 7. The (indirect) adaptive feature of the algorithm is a consequence of solving the optimal-control problem using the latest parameter estimate $\hat{\theta}(t)$. Note that this does not make the D.M.P.C. a certainty-equivalence controller, since the control action also depends on the estimate covariance, meaning the estimates are not used as if they were the true values. The algorithm is illustrated in Figure 1 and can be summarized as follows:

1 Initialize at time $t = t_0$: specify the hyperstate $(\varphi(t_0), u(t_0-1), \hat{\theta}(t_0), P(t_0))$.
2 At time t, collect plant data: measure $y(t)$ and $u(t-1)$.
3 Update the hyperstate $\big(\varphi(t), u(t-1), \hat{\theta}(t), P(t)\big)$ (the conditional distribution of θ is updated using (11)).
4 Solve (P″) to obtain the solution $\{u^o(k \mid t)\}_{k=t}^{t+N-1}$.
5 Implement $u(t) = u^o(t \mid t)$.
6 Set $t \leftarrow t+1$ and go to step 2.

Our algorithm differs from standard M.P.C. in that the D.M.P.C. not only needs feedback from the system state, but from the hyperstate $\big(\varphi(t), u(t-1), \hat{\theta}(t), P(t)\big)$. While the approach is similar to an adaptive certainty-equivalence controller in that it uses parameter estimates (expected values) in the prediction model, an important difference is that the D.M.P.C. does not assume that the expected values equal the true parameter values.

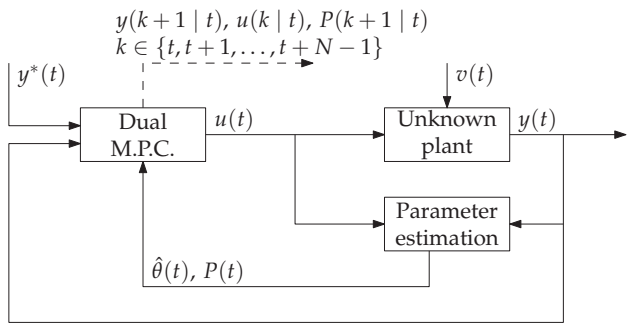


Figure 1: Block diagram illustrating the adaptive D.M.P.C. structure. The dashed line illustrates variables predicted by the D.M.P.C.

6 EXAMPLE

We now demonstrate our D.M.P.C. algorithm on a small simulation test case that highlights some of its main features. The simulations compare adaptive certainty-equivalence M.P.C. with D.M.P.C. (Figure 2) and also demonstrate how D.M.P.C. performance is affected by different initial error-covariance magnitudes (Figure 3). To best demonstrate the qualitative behavior of the D.M.P.C. we consider a scenario where the system has been in steady state for some time and show the effect of reinitializing the controller with different sets of values in the covariance matrix $P(t_0)$. We use one case with very large variances at time t_0 to trigger a strong dual action that results from uncertainty. This behavior is contrasted with an adaptive certainty-equivalence M.P.C. as well as more moderate variance values at t_0.

All four simulations are identical up to time $t_0 = 0$, with the history from $t = -5$ included for clearer illustration of the decisions made by the D.M.P.C.s. The parameter estimates are perfect and there is no uncertainty (P is a matrix of zeros) prior to time $t = -3$. This means the D.M.P.C. reduces to a certainty-equivalence M.P.C., which gives optimal control performance in this situation (correct parameter estimates with no uncertainty). A shift in the model parameters occurs at $t = -2$, but the D.M.P.C. and the certainty-equivalence M.P.C. are both able to keep the ouput at the reference without changing the input. Figures 2 and 3 together show the consequence of reinitializing the D.M.P.C. at time $t_0 = 0$ by setting $P(t_0)$ to a nonzero, positive-semi-definite matrix. From top to bottom, the four plots show the outputs, including the reference $y^*(t)$; the inputs, including the optimal steady-state input

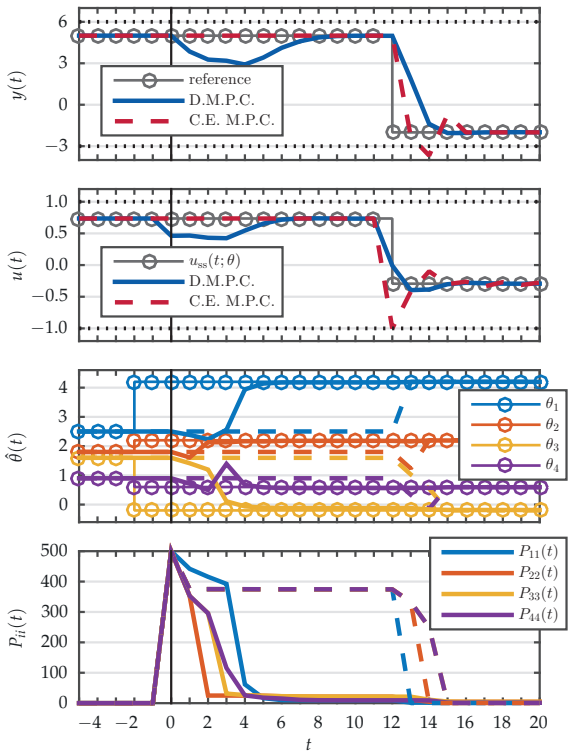


Figure 2: A comparison of certainly-equivalence (C.E.) M.P.C. (dashed, heavy lines) and D.M.P.C. (solid, heavy lines).

$u_{ss}(t;\theta)$ (defined below for the example system); the true parameters θ and both sets of estimates $\hat{\theta}(t)$; and the diagonal elements of the covariance matrix $P(t)$, which are the variance terms.

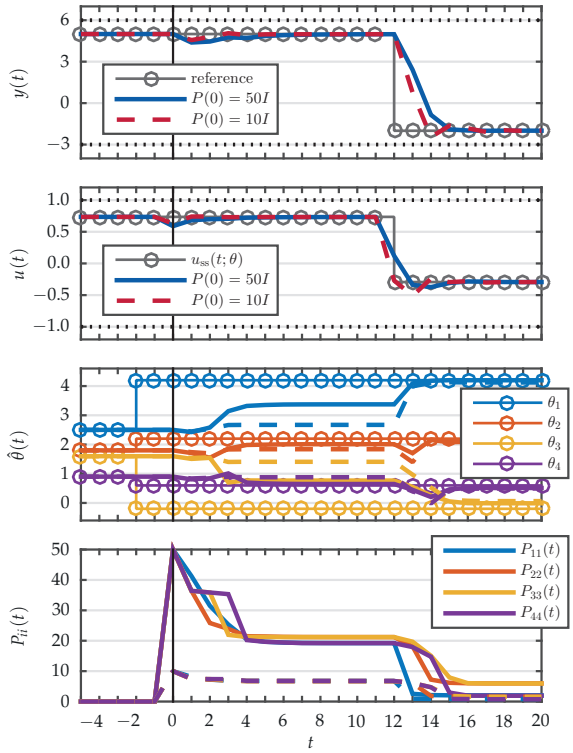


Figure 3: A comparison of two D.M.P.C. setups that differ only in the choice of covariance at $t = t_0$: $P(t_0) = 10I$ (dashed, heavy lines) and $P(t_0) = 50I$ (solid, heavy lines).

Figure 2 shows how a D.M.P.C. with $P(t_0) = 500I$ diverges from an identically-tuned certainty-equivalence M.P.C., provided the same operational history. The certainty-equivalence M.P.C. (dashed red line) continues applying a control identical to the optimal steady-state input $u_{ss}(t;\theta)$ after the change in parameter values, keeping the output at the reference without changing the control signal. Note the dependence on the true model parameters θ in the optimal steady-state input. A consequence of the status-quo input-output situation is that no new data is generated with the result that there is no indication that a change occurred in the plant. Hence, the parameter estimates do not change and

are wrong when the subsequent change in output reference occurs at $t = 12$, resulting in a violation of the lower output bound and significant oscillations before the output settles at the new reference value. Note that none of the controllers anticipate the setpoint change.

The consequence of reinitializing the D.M.P.C. (solid blue line) with $P(t_0) = 500I$ is clear when contrasted with the certainty-equivalence M.P.C. The output from the D.M.P.C. is reduced in magnitude at $t = t_0 = 0$ in order to reduce the covariance matrix, in turn reducing the cost incurred from the term $\varphi^\top(k \mid t)P(k \mid t)\varphi(k \mid t)$ in the objective (25). This probing action causes a reduction in output magnitude, and subsequently a change in the parameter estimates (solid lines) and a reduction of the variance (the diagonal elements of $P(t)$ are shown). The excitation, or probing action, continues until about $t = 5$, which can be understood as a consequence of $P(5)$ being sufficiently small (the variances are two orders of magnitude smaller than their initial values) so that further uncertainty reduction is not worth the price paid in control deterioration in the nominal sense. At this time, the parameter estimates are very close to the true values and the input from the D.M.P.C. in converging toward the optimal steady-state value. Consequently, the output is converging back to the setpoint. When the subsequent change of setpoint occurs, the D.M.P.C. has learned enough about the new plant parameters to successfully move the output toward the new reference, without the constraint violation and oscillations in the output seen with the certainty-equivalence M.P.C.

Figure 3 shows how the D.M.P.C. performs when reinitialized with smaller diagonal values in $P(t_0)$ ($10I$ and $50I$) and all other factors kept identical. We observe the same qualitative behavior with these smaller covariance matrices that represent less system uncertainty, but the excitation vanishes sooner and decreases in magnitude with the smaller variance quantities. Less excitation means less informative data, which leads to parameter estimates that are further from the true values. Despite the very moderate excitation, both of these D.M.P.C. setups are able to direct the output to the new reference without constraint violations and significant oscillations.

The example system is of F.I.R. type with $n_p = 4$ parameters, and the matrices A and B in the model formulation (1) are given in Appendix A3. In order to isolate the capabilities of the D.M.P.C. algorithm we are not including noise in the simulations, meaning $r = 0$ and $r_R = 1$. All simulations start at $t_i = -5$ and end at $t_f = 20$ with $t_0 = 0$ (the time at which we set P to a nonzero matrix in the D.M.P.C.);

the output setpoints are $y^*(t) = 5.0$ for $t \leq 11.0$ and $y^*(t) = -2$ for $t \geq 12$; the unknown model parameters are $\theta = \theta_1 = [2.5, 1.8, 1.6, 0.9]^\top$ for $t \leq -3$ and $\theta = \theta_2 = [4.2, 2.2, -0.2, 0.6]^\top$ for $t \geq -2$; the initial data are $u(t) = u_{\mathrm{ss}}(t;\theta_1)$ for all $t \leq t_i$, $\hat{\theta}(t_i) = \theta_1$, $P(t_i) = 0_{n_p \times n_p}$, and $y(t_i) = 5.0$; the algorithm parameters are $N = 8$, $w_2 = 0$, $w_3 = 10^{-3}$, $y_{\min} - 3.0$, $y_{\max} = 6.0$, $s_{\min} = s_{\max} = 0$ (which corresponds to $p_{y,\min} = p_{y,\max} = 0.5$), $-u_{\min} = u_{\max} = 1.0$, $-\Delta u_{\min} = \Delta u_{\max} = \infty$, and $\lambda = 1$. The optimal steady-state input for the example system is $u_{\mathrm{ss}}(t;\theta) = y^*(t) \,/\, \left(\sum_{j=1}^{n_p} \theta_j\right)$, which depends on the true parameter values θ.

The example simulations are implemented in MATLAB and the Q.C.Q.P. problems are solved using the local N.L.P. solver SNOPT 7.2 (Gill, Murray, and Saunders, 2005) under GAMS (GAMS Development Corporation), which uses automatic differentiation to provide gradients to the solver. We run the simulations on a standard laptop computer and the Q.C.Q.P. problems all take between 0.05 s and 5.02 s to solve, with a mean of 0.47 s and more than 90 % of the solutions obtained in less than 0.70 s.

7 DISCUSSION

The dual-control algorithm we develop here is based on exact transformation of a stochastic optimal-control problem, formulated for orthonormal-basis-function systems, into deterministic form. Starting from the stochastic dual output cost in (8), which is the square of the future tracking error, we are able to arrive at two distinct, deterministic components: 1) future nominal tracking error, and 2) future output covariance. Minimizing the sum of these quantities results in the minimal expected tracking error, both conditioned on the future information. That is, these two quantities cannot be traded off against each other; less excitation does not improve the expected tracking error, and neither does prioritizing the nominal tracking error over excitation. Rather, the specific reward functions for excitation and nominal output error are the unique expressions that at the minimum of their sum result in the dual control, which is the control that minimizes the expected output error. That is, excitation and nominal output tracking are not tasks in conflict: they are inseparable means to an end and must be rewarded together in specific forms when the goal is to determine the optimal control. One insight from this line of argument is that any objective function with its minimum different from that of (8) does not result in the dual control. At best, this type of design leads to an approximate dual control. Neverthe-

less, such strategies may give improved performance over non-exciting control algorithms. For more general system formulations it may be impossible to obtain a closed-form deterministic function that can be systematically minimized in order to obtain the dual control. However, this type of insight may aid the control designer to better approximate the unknown deterministic dual objective.

For the systems we consider here, we can formulate the equivalent deterministic dual-control problem as a Q.C.Q.P. problem of modest complexity. Solving this problem in a receding-horizon manner results in a dual feedback controller, or a D.M.P.C. algorithm.

We demonstrate the main features of D.M.P.C. through simulation examples. When there is uncertainty in the parameter estimates, as represented by a nonzero covariance matrix, the D.M.P.C. takes active steps to reduce uncertainty and improve the parameter estimates. The controller accomplishes this through systematic self-excitation, the magnitude and length of which increases with the estimate covariance. The uncertainty is not reduced to zero and the parameter estimates do not converge completely to the true values; rather, the excitation is such that further uncertainty reduction decreases the expected performance. That is, the cost of the additional excitation required for zero uncertainty and perfect estimates is greater than the expected marginal return of the improved model. The control objective of the D.M.P.C. converges to that of the certainty-equivalence M.P.C. as the uncertainty goes to zero, resulting in identical input decisions when the D.M.P.C. is not actively perturbing the system for learning purposes. The simulations show how active learning, or excitation for uncertainty reduction, directly enables the D.M.P.C. to later move the output to a different setpoint without encountering problems such as constraint violations and significant oscillations, both if which are experienced with the passive-learning approach in the certainty-equivalence M.P.C.

The D.M.P.C. is an indirect adaptive controller since we update the plant-parameter statistics in the prediction model used by the algorithm. This is not a design choice; it is a natural consequence of updating the statistics and making the control decisions based on current and future information, resulting in a closed loop. Furthermore, the D.M.P.C. accounts for the effect of the control actions on learning since the mechanisms for learning in the closed loop are part of the prediction model. This form of endogenized learning is a type of feedback control with a partially closed prediction loop model.

The results presented here rely on the assumption that both the parameters θ and the process noise $v(t)$ are Gaussian; the reformulations are not valid if these assumptions are not met. However, we can trivially extend the framework we develop here to time-varying parameters modeled as the Gauss-Markov process

$$\theta(t+1) = \Phi\theta(t) + w(t) \tag{40}$$

where Φ is a known, constant matrix and $w(t)$ is a sequence of independent and identically distributed Gaussian random vectors with zero mean and variance r_w. Extending the approach to multivariable control is possible by using ideas similar to those presented by Kumar et al. (2015), where an approximate D.M.P.C. for scalar systems is extended to the multivariable case.

8 CONCLUSIONS AND FUTURE WORK

Our reformulation (P′) of the probabilistically-constrained stochastic optimal-control problem (P) provides clear insight into how specific functions of excitation and nominal control together result in dual control when their sum is minimal, as well as a foundation for practical control-algorithm design. Through careful choices in the control-problem reformulation we arrive at a Q.C.Q.P. problem (P″) that can be solved efficiently. Solving this Q.C.Q.P. on a receding horizon using the future and current information results in the D.M.P.C. algorithm. Global solutions to the optimal-control problem can be obtained efficiently, which is the topic of a future paper. The reformulation allows for easy incorporation of exact probabilistic constraints with only a small increase in problem complexity.

Future work includes stability analysis, derivation of tight variable bounds to help facilitate global solutions, and investigation of when dual control is and is not beneficial.

A1 OTHER RESULTS

The following lemma is a standard result, but we list it here as it is omitted from several basic statistics texts.

LEMMA 1. *For a stochastic variable vector $X \in \mathbb{R}^n$, the variance of X is*

$$\mathrm{Var}(X) = \mathrm{E}\left[XX^\top\right] - \hat{X}\hat{X}^\top \tag{41}$$

where $\hat{X} = \mathrm{E}[X]$.

Proof. Since $X\,\mathrm{E}[X]^\top$ is symmetric, we have that

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\big[(X - \mathrm{E}[X])(X - \mathrm{E}[X])^\top\big] \\
&= \mathrm{E}\big[XX^\top - X\,\mathrm{E}[X]^\top - \mathrm{E}[X]X^\top + \mathrm{E}[X]\,\mathrm{E}[X]^\top\big] \\
&= \mathrm{E}\big[XX^\top - 2X\,\mathrm{E}[X]^\top + \mathrm{E}[X]\,\mathrm{E}[X]^\top\big] \\
&= \mathrm{E}\big[XX^\top\big] - 2\,\mathrm{E}[X]\,\mathrm{E}[X]^\top + \mathrm{E}[X]\,\mathrm{E}[X]^\top \\
&= \mathrm{E}\big[XX^\top\big] - \mathrm{E}[X]\,\mathrm{E}[X]^\top \\
&= \mathrm{E}\big[XX^\top\big] - \hat{X}\hat{X}^\top
\end{aligned} \tag{42}$$

□

For the systems we consider in this paper, the expected value of the unknown parameter vector θ given current information and future decisions, $\mathcal{Y}(k \mid t)$, is the same as given only current information. We state this formally in the following lemma.

LEMMA 2. *For a stochastic process of the form* (1)*,*

$$\mathrm{E}\big[\theta \;\big|\; \mathcal{Y}(k \mid t)\big] = \mathrm{E}\big[\theta \;\big|\; \mathcal{Y}(t)\big] = \hat{\theta}(t), \qquad k \geq t \tag{43}$$

Proof. From (11a) it is apparent that the conditional mean of θ at time t depends on $y(t)$. Thus, $\mathcal{Y}(k \mid t)$ does not contain any information relevant for $\hat{\theta}(k)$, $k \geq t$, beyond $\mathcal{Y}(t)$, so the conditional mean of θ given $\mathcal{Y}(k \mid t)$ is simply $\mathrm{E}\big[\theta \;\big|\; \mathcal{Y}(t)\big] = \hat{\theta}(t)$. □

A consequence of this is that we can write

$$\hat{\theta}(k \mid t) := \mathrm{E}\big[\theta \;\big|\; \mathcal{Y}(k \mid t)\big] = \hat{\theta}(t), \qquad k \geq t \tag{44}$$

The following theorem states an expression for the expected value of the matrix $\theta\theta^\top$ given the anticipated information $\mathcal{Y}(k \mid t)$ in terms of deterministic quantities.

THEOREM 5. *For a stochastic process of the form* (1)*,*

$$\mathrm{E}\left[\theta\theta^\top \;\middle|\; \mathcal{Y}(k \mid t)\right] = \hat{\theta}(t)\hat{\theta}^\top(t) + P(k \mid t), \quad k \geq t \tag{45}$$

Proof. Using Lemma 2 and the fact that $\hat{\theta}(t)\theta^\top$ is symmetric, we start from the definition of the anticipated covariance matrix $P(k \mid t)$ and get

$$\begin{aligned} P(k \mid t) &:= \mathrm{E}\big[\big(\theta - \mathrm{E}\big[\theta \mid \mathcal{Y}(k \mid t)\big]\big)\big(\theta - \mathrm{E}\big[\theta \mid \mathcal{Y}(k \mid t)\big]\big)^\top \mid \mathcal{Y}(k \mid t)\big] \\ &= \mathrm{E}\big[\big(\theta - \hat{\theta}(t)\big)\big(\theta - \hat{\theta}(t)\big)^\top \mid \mathcal{Y}(k \mid t)\big] \\ &= \mathrm{E}\big[\theta\theta^\top - \theta\hat{\theta}^\top(t) - \hat{\theta}(t)\theta^\top + \hat{\theta}(t)\hat{\theta}^\top(t) \mid \mathcal{Y}(k \mid t)\big] \\ &= \mathrm{E}\big[\theta\theta^\top - 2\theta\hat{\theta}^\top(t) + \hat{\theta}(t)\hat{\theta}^\top(t) \mid \mathcal{Y}(k \mid t)\big] \\ &= \mathrm{E}\big[\theta\theta^\top \mid \mathcal{Y}(k \mid t)\big] - 2\,\mathrm{E}\big[\theta \mid \mathcal{Y}(k \mid t)\big]\hat{\theta}^\top(t) + \hat{\theta}(t)\hat{\theta}^\top(t) \\ &= \mathrm{E}\big[\theta\theta^\top \mid \mathcal{Y}(k \mid t)\big] - 2\hat{\theta}(t)\hat{\theta}^\top(t) + \hat{\theta}(t)\hat{\theta}^\top(t) \\ &= \mathrm{E}\big[\theta\theta^\top \mid \mathcal{Y}(k \mid t)\big] - \hat{\theta}(t)\hat{\theta}^\top(t) \end{aligned} \tag{46}$$

Rearranging the equation completes the proof. □

An immediate consequence is the expression obtained when the expectation is taken with respect to the current information $\mathcal{Y}(t)$:

COROLLARY 3. *For a stochastic process of the form* (1),

$$\mathrm{E}\big[\theta\theta^\top \mid \mathcal{Y}(t)\big] = \hat{\theta}(t)\hat{\theta}^\top(t) + P(t) \tag{47}$$

Proof. The proof follows trivially from the proof of Theorem 5. □

A2 ALTERNATIVE OBJECTIVE FUNCTIONS

The simplest output cost for an uncertain system of the form (1) is obtained by minimizing the squared difference between the predicted model output and the output reference. This amounts to taking the expected value of the output with respect to current information, resulting in the output cost

$$\begin{aligned} &\sum_{k=t}^{t+N-1} \big(\mathrm{E}[y(k+1) \mid \mathcal{Y}(t)] - y^*(k+1 \mid t)\big)^2 \\ &\qquad = \sum_{k=t}^{t+N-1} (\hat{y}(k+1 \mid t) - y^*(k+1 \mid t))^2 \end{aligned} \tag{48}$$

This cost penalizes the output error as if the most recent parameter estimate $\hat{\theta}(t)$, as defined by (5), were exact. That means this output cost forms the basis for a certainty-equivalence-type M.P.C. Combining

this idea with system identification then gives the indirect adaptive predictive controller. Neither the current nor future uncertainty affect the output cost, meaning this type of controller has no caution or probing features (see Bar-Shalom (1981)). Hence, the controller is not risk averse in the face of large uncertainty, and takes no active steps to explore the plant.

If we take the conditional expectation of the squared output-tracking error with respect to the current information $\mathcal{Y}(t)$, as opposed to also including the future decision sequence through $\mathcal{Y}(k \mid t)$, we can derive an output cost that rewards cautious controls, as stated in the following theorem.

THEOREM 6. *For a stochastic process of the form* (1),

$$\begin{aligned}\sum_{k=t}^{t+N-1} \mathrm{E}\left[(y(k+1) - y^*(k+1 \mid t))^2 \mid \mathcal{Y}(t)\right] \\ = \sum_{k=t}^{t+N-1} \big\{ (\hat{y}(k+1 \mid t) - y^*(k+1 \mid t))^2 \\ + \varphi^\top(k \mid t) P(t) \varphi(k \mid t) + r \big\} \end{aligned} \tag{49}$$

Proof. This proof is near identical to that of Theorem 3. We have that

$$\begin{aligned}&\mathrm{E}[(y(k+1) - y^*(k+1 \mid t))^2 \mid \mathcal{Y}(t)] \\ &\quad = \mathrm{E}[y^2(k+1) - 2y(k+1)y^*(k+1 \mid t) \\ &\qquad\qquad + (y^*(k+1 \mid t))^2 \mid \mathcal{Y}(t)] \\ &\quad = \varphi^\top(k+1 \mid t)\, \mathrm{E}\left[\theta\theta^\top \mid \mathcal{Y}(t)\right] \varphi(k+1 \mid t) \\ &\qquad\qquad + \mathrm{E}\left[v^2(k+1) \mid \mathcal{Y}(t)\right] \\ &\qquad + 2\,\mathrm{E}\left[v(k+1)\theta^\top \mid \mathcal{Y}(t)\right] \varphi(k+1 \mid t) \\ &\qquad - 2\,\mathrm{E}\left[y(k+1) \mid \mathcal{Y}(t)\right] y^*(k+1 \mid t) \\ &\qquad\qquad + (y^*(k+1 \mid t))^2 \end{aligned} \tag{50}$$

where the second term r and the third term is zero since $v(k+1)$ and θ are uncorrelated. We here use Corollary 3 for the first term to arrive at the desired result:

$$\begin{aligned}
&\mathrm{E}[(y(k+1) - y^*(k+1 \mid t))^2 \mid \mathcal{Y}(t)] \\
&\quad = \left(\hat{\theta}(t)\varphi^\top(k \mid t)\right)^2 + \varphi^\top(k \mid t)P(t)\varphi(k \mid t) + r \\
&\qquad\qquad - 2\hat{y}(k+1 \mid t)y^*(k+1 \mid t) + (y^*(k+1 \mid t))^2 \\
&\quad = \left(\hat{\theta}^\top(t)\varphi(k \mid t)\right)^2 + \varphi^\top(k \mid t)P(t)\varphi(k \mid t) + r \\
&\qquad\qquad - 2\hat{\theta}^\top(t)\varphi(k \mid t)y^*(k+1 \mid t) + (y^*(k+1 \mid t))^2 \\
&\quad = (\hat{y}(k+1 \mid t) - y^*(k+1 \mid t))^2 + \varphi^\top(k \mid t)P(t)\varphi(k \mid t) + r \qquad (51)
\end{aligned}$$

□

The output cost (49) (which is identical to (9)) rewards caution because the current covariance $P(t)$ is included in the quadratic form $\varphi^\top(t)P(t)\varphi(k)$. The current covariance can be interpreted as a weight for a 2-norm of the regressors: $\|\varphi^\top(k \mid t)\|^2_{P(t)}$. The same covariance matrix $P(t)$ is thus penalizing all future regressors $\varphi^\top(k \mid t)$ on the prediction horizon, $k \geq t$. Large covariances then lead to smaller inputs, or caution, while there is no incentive to explore the plant since the effect of future decisions on the future uncertainty is not modeled.

Note that neither of the two objectives above requires the additional constraints for uncertainty propagation. (The chance constraints do of course require the uncertainty propagation constraints.)

A3 STATE-SPACE FORMULATIONS FOR MODELS WITH ORTHONORMAL BASIS FUNCTIONS

The simplest and most common model of the form (1) is the finite-impulse-response (F.I.R.) model. $A \in \mathbb{R}^{n_p \times n_p}$ and $B \in \mathbb{R}^{n_p \times 1}$ in (1a) are in this case

$$A = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ 1 & \ddots & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \qquad (52)$$

Note that all the eigenvalues of A are at the origin in the F.I.R. case. For a Laguerre model, we can write $\bar{E}\varphi(t+1) = \bar{A}\varphi(t) + \bar{B}u(t)$ (see, e.g, Wahlberg and Lindskog (1990)) with

$$\bar{E} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ a & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a & 1 \end{bmatrix}, \qquad \bar{A} = \begin{bmatrix} a & 0 & \cdots & \cdots & 0 \\ 1 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & a \end{bmatrix} \tag{53a}$$

$$\bar{B}^\top = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \end{bmatrix}^\top \sqrt{1-a^2} \tag{53b}$$

We can also write the Laguerre model in the form (1) with $A = \bar{E}^{-1}\bar{A} \in \mathbb{R}^{n_p \times n_p}$ and $B = \bar{E}^{-1}\bar{B} \in \mathbb{R}^{n_p \times 1}$. A and B are then

$$A = \begin{bmatrix} a & 0 & 0 & \cdots & 0 \\ 1-a^2 & a & 0 & \cdots & 0 \\ (-a)(1-a^2) & 1-a^2 & a & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ (-a)^{n_p-2}(1-a^2) & (-a)^{n_p-3}(1-a^2) & \cdots & 1-a^2 & a \end{bmatrix} \tag{54a}$$

$$B = \begin{bmatrix} 1 & (-a) & (-a)^2 & \cdots & (-a)^{n_p-1} \end{bmatrix}^\top \sqrt{1-a^2} \tag{54b}$$

Note that an F.I.R. model is a Laguerre model with the pole $a = 0$. A state-space realization in terms of $\bar{E}$, $\bar{A}$, and $\bar{B}$ for Kautz models can be found in the literature, for instance in Wahlberg and Lindskog (1990), which also provides illustrations of network structures for different models based on orthonormal basis functions. The state-space representation for Kautz models is more involved than for the F.I.R. and Laguerre models, and is not repeated here.

EXPERIMENTAL EVALUATION OF A MIMO ADAPTIVE DUAL MPC

Reference information for Paper D:
Kumar, K. et al. (2015). "Experimental evaluation of a MIMO adaptive dual MPC." In: *Advanced Control of Chemical Processes*. Whistler, Canada, pp. 546–551.

D

FOREWORD

This paper is a result of a collaborative project with researchers at I.I.T. Bombay in India. The project was led by Prof. S.C. Patwardhan at I.I.T. and the implementation and experiment work was primarily done by K. Kumar for his master's thesis research. We extend the algorithm developed in Paper B to multivariable ARMAX systems and test the controller on an experimental setup. My involvement in the project includes contributions to extending the algorithm, evaluation of the experimental results, and writing parts of the paper.

I have made some minor edits to this paper for inclusion in the thesis. These edits include changing some notation for better consistency with the rest of the material, as well smaller changes to the language and typography.

ABSTRACT

Maintaining uniformly satisfactory control performance of an M.P.C. scheme in the face of changing operating conditions is a difficult task. An adaptive M.P.C. scheme that directs the output towards a reference and simultaneously injects a probing signal to get more information about the system for better model identification appears to be ideally suited for achieving this goal. In this work, taking motivation from the dual control problem originally developed by Feldbaum (1961b), a MIMO adaptive dual M.P.C. (adaptive D.M.P.C.) formulation has been proposed, which does not require external probing signals to improve the model parameter estimates. The objective function of the M.P.C. is

The term D.M.P.C. *as used here is not consistent with the use in Chapter 3 and Paper C.*

modified to include terms that ensure that sufficient excitation is injected into the system while performing the control tasks. The efficacy of the proposed adaptive D.M.P.C. formulation is evaluated by conducting experimental studies on the benchmark heater mixer system. The experimental results demonstrate that the proposed formulation is able to inject probing inputs of small magnitude while meeting the desired servo and regulatory control objectives.

1 INTRODUCTION

Over the last three decades, model predictive control (M.P.C.) has emerged as the most effective multivariable control scheme and has been used to control a wide variety of processes (Qin and Badgwell, 2003). The quality of the model has a large effect on the closed loop performance of a linear M.P.C. scheme. Maintaining a high quality model so as to achieve good control performance in the face of changing operating conditions poses a difficult challenge in the process industry. This problem has been dealt with mainly by (a) incorporating robustness in the controller design (b) employing multiple model-based controller designs and (c) updating the parameters of the linear prediction model either intermittently or on-line (Morari and Lee, 1999). While (c) appears to be an attractive option, due to various constraints, such as the time required for model identification and the cost associated with the model identification exercise, the model updates are carried out infrequently. If it is desired to update model parameters online, then a variety of recursive least square algorithms are available in the system identification literature (Åström and Wittenmark, 1995; Söderström and Stoica, 1989). These approaches, though extensively studied in the system identification and adaptive control literature, have not received the attention they deserve in the industrial applications of M.P.C. Qin and Badgwell (2003), in their review of industrial M.P.C., noticed that only two adaptive M.P.C. algorithms had reached the marketplace by 2003 despite strong market incentive for self-tuning M.P.C.

The observed lack of interest in employing these adaptive M.P.C. (A.M.P.C.) formulations on industrial systems can be attributed to the reliability of on-line parameter estimation schemes, which are at the heart of any A.M.P.C. strategy. Ydstie (1997) has indicated that *instability of the parameter estimator* or the *parameter drift* is an important issue that needs to be addressed while developing an adaptive control scheme.

This drift can be avoided by adding deliberate perturbations to the manipulated inputs.

In the present work we focus on maintaining high quality parameter estimates by deliberately letting manipulated inputs perturb the plant. The persistent excitation guarantees the convergence of the parameter estimates. However, in practice, the perturbation signal are often chosen through some heuristic means and this can lead to excessive excitation. Ideally, an optimal controller must direct the output towards a reference and simultaneously inject a probing signal to get more information about the system for improved model identification, so that better control can be achieved in the future. This type of controller is referred to as a dual controller. In the dual control formulation, external persistent excitation is not required because the controller itself optimally excites the process when needed. The concept of dual control was first introduced by Feldbaum (1961b) as a result of an attempt to formulate optimal control problems which would give an adaptive control law. This dual character of the control law refers to the two tasks of directing the output towards specified values and investigating the plant for learning. That is, the controller finds a balance between control and excitation. A dual controller optimally probes the system when the model is poor, which generates sufficient excitation to improve the model and, in turn, the closed loop performance.

Larsson, Annergren, et al. (2013) developed an M.P.C. that experiments with the plant for identification processes while simultaneously controlling the plant. The excitation is introduced through a constraint on the predicted information matrix. A similar M.P.C. with dual features was developed by Marafioti, Bitmead, and Hovd (2014). The excitation is here guaranteed by requiring that the first element of the open-loop optimal input sequence be persistently exciting. Žáčeková, Prívara, and Pčolka (2013) suggested an approach where a standard M.P.C. problem is solved first, followed by a procedure for finding an optimal perturbation to the nominally optimal input so that the resulting control increases the minimal eigenvalue of the information matrix.

Adaptive M.P.C. with dual control features has emerged as an attractive approach to the problem of control loop performance degradation due to model plant mismatch. Recently, Heirung, Ydstie, and Foss (2012b, 2013b) developed a dual control formulation based on certainty-equivalence adaptive M.P.C. In their most recent approach the first stage cost is reformulated from a stochastic expression into a deterministic one. The same expression is used for the next stage

cost in order to ensure that excitation is rewarded by the controller. The result is a level of excitation that excites the process enough to improve the quality of the parameter estimates and thereby improves closed loop performance. However, they considered only single-input single-output (SISO) systems subjected to zero-mean white disturbances; an extension to the more general MIMO case is not obvious. Since the ability to handle multivariable systems are among the main advantages of M.P.C. in industrial applications, we here extend the formulation proposed by Heirung, Ydstie, and Foss (2013b) to deal with multiple input multiple output (MIMO) systems subjected to colored unmeasured disturbances. Our proposed extension is based on multiple MISO ARMAX models, which are better suited for colored disturbances. The efficacy of the proposed adaptive dual M.P.C. is evaluated through experimental studies on a heater-mixer system (Thornhill, Patwardhan, and Shah, 2008).

This paper is organized into four sections. In the next section, development of the proposed A.M.P.C. formulation motivated by the dual control approach is presented. The analysis of the experimental results is presented in Section 3. The major conclusions reached from the analysis are summarized in Section 4.

2 A DUAL-CONTROL APPROACH TO ADAPTIVE M.P.C.

In this section we extend the M.P.C.-based approach to dual control proposed by Heirung, Ydstie, and Foss (2013b) to handle MIMO systems subjected to unmeasured stochastic disturbances. Heirung, Ydstie, and Foss (2013b) used an A.R.X. model to formulate an algorithm for adaptive dual M.P.C. for SISO systems. The proposed extension can in principle be carried out using MISO or MIMO versions of the A.R.X. model, but the conventional A.R.X. models have certain limitations when the system under consideration is subjected to colored unmeasured disturbances (Söderström and Stoica, 1989). For example, consider a system governed by a SISO ARMAX model of the form

$$y(k) = \frac{B(q^{-1})}{A(q^{-1})}u(k) + \frac{C(q^{-1})}{A(q^{-1})}e(k) \tag{1}$$

where $e(k)$ is a zero mean white noise sequence and $C(q^{-1})$ has all roots inside the unit circle. An equivalent A.R.X. model for this system can be obtained by rearranging the ARMAX model as

$$y(k) = \frac{B(q^{-1})/C(q^{-1})}{A(q^{-1})/C(q^{-1})}u(k) + \frac{1}{A(q^{-1})/C(q^{-1})}e(k) \tag{2}$$

and by truncating $\tilde{A}(q^{-1})=A(q^{-1})/C(q^{-1})$ and $\tilde{B}(q^{-1})=B(q^{-1})/C(q^{-1})$ after the coefficients of $\tilde{A}(q^{-1})$ and $\tilde{B}(q^{-1})$ become insignificant. The truncation order depends on the locations of the roots of $C(q^{-1})$. If $C(q^{-1})$ has root(s) close to the unit circle, then an A.R.X. model of high order is needed to adequately capture the noise dynamics. Thus, even for a SISO system, it may be necessary to estimate a relatively large number of model parameters to adequately capture the system dynamics when an A.R.X. structure is used to model a system subjected to colored noise. This difficulty is further compounded for a MIMO system. A model with a large number of parameters can lead to difficulties with on-line parameter estimation as a large data set is needed to keep the variance errors small (Madakyaru, Narang, and Patwardhan, 2009). In other words, the plant needs to be perturbed for a longer time to estimate the parameters accurately. From the viewpoint of parsimony of model parameters, a better option is to employ MISO models with either ARMAX or Box-Jenkins (B.J.) structure (Ljung, 1999). In this work we propose to capture the system dynamics using an ARMAX model structure to keep the development simple.

2.1 ARMAX *models and online parameter estimation*

Consider a MIMO system with r outputs and m manipulated inputs. The system under consideration is assumed to have a local linear approximation in the neighborhood of a desired operating point such that the approximate linear model is stably invertible. We propose to model the system as r MISO ARMAX models of the form

$$A_i(q^{-1})y_i(k) = \sum_{j=1}^{m} B_{ij}(q^{-1})u_j(k) + C_i(q^{-1})e_i(k) \tag{3}$$

where $i = 1, \ldots, r$. Here, $A_i(q^{-1})$, $B_{ij}(q^{-1})$, and $C_i(q^{-1})$ are polynomials in the backward shift operator q^{-1} and $\{e_i(k)\}$ represents a zero mean

white noise sequence. To simplify the notation, the index i is dropped in this subsection and the ith MISO model is represented as

$$A(q^{-1})y(k) = \sum_{j=1}^{m} B_j(q^{-1})u_j(k) + C(q^{-1})e(k) \tag{4}$$

For the purpose of online parameter estimation, this model can be expressed as

$$y(k) = \phi^{\top}(k-1)\theta(k-1) + e(k) \tag{5}$$

where

$$\theta = [a_1, \ldots, a_n, b_{11}, \ldots, b_{1n}, \ldots, b_{m1}, \ldots, b_{mn}, c_1, \ldots, c_n]^{\top} \tag{6}$$

is a vector containing the MISO model parameters and

$$\begin{aligned}\phi(k-1) = [&-y(k-1), \ldots, -y(k-n), u_1(k-1), \\ &\ldots, u_m(k-n), e(k-1), \ldots, e(k-n)]^{\top}\end{aligned} \tag{7}$$

represents the regressor vector, which consists of inputs, outputs, and noise inputs from the past.

REMARK. *For the sake of simplifying the notation we assume that the orders of $A(q^{-1})$, $B_j(q^{-1})$, and $C(q^{-1})$ are equal (cf. equation (6)). However, the orders of the $B_j(q^{-1})$ and $C(q^{-1})$ polynomials may in general differ, and they can be of any (positive) order less than or equal to the order of $A(q^{-1})$.*

A difficulty in using (7) for recursive parameter estimation is that the noise sequence $\{e(k)\}$ is unknown. This difficulty can be alleviated if we employ the extended least square (E.L.S.) approach (also known as pseudo-linear regression or approximate M.L. method) for the model parameter estimation (Åström and Wittenmark, 1995; Söderström and Stoica, 1989). In this approach, $e(k)$ is replaced by the estimated prediction error. Thus, the regressor vector is in the E.L.S. approach modified to

$$\begin{aligned}\varphi(k-1) = [&-y(k-1), \ldots, -y(k-n), u_1(k-1), \\ &\ldots, u_m(k-n), \varepsilon(k-1), \ldots, \varepsilon(k-n)]^{\top}\end{aligned} \tag{8}$$

where $e(k-i)$ has been replaced by $\varepsilon(k-i)$, which is the past innovations. Here, the innovation $\varepsilon(k)$ at instant k is

$$\varepsilon(k) = y(k) - \varphi^{T}(k-1)\hat{\theta}(k-1) \tag{9}$$

where $\hat{\theta}(k-1)$ represents the parameter estimate obtained at instant $(k-1)$ using the E.L.S. method. The E.L.S. method can be summarized as

$$\hat{\theta}(k) = \hat{\theta}(k-1) + L(k)\varepsilon(k) \tag{10a}$$

$$L(k) = P(k-1)\varphi(k-1)(\lambda + \varphi^T(k-1)P(k-1)\varphi(k-1))^{-1} \tag{10b}$$

$$P(k) = (I - L(k)\varphi^T(k-1))P(k-1)/\lambda \tag{10c}$$

To carry out model identification, r MISO estimators are used in parallel. Thus, we obtain the estimates $\hat{\theta}^{(i)}(k)$ and the corresponding covariance matrices $P^{(i)}(k)$ for $i = 1, 2, \ldots, r$, which are then used in the proposed adaptive dual M.P.C. (D.M.P.C.) formulation. Let $\mathcal{Y}(k)$ denote the set of inputs and outputs recorded up to time instant k; i.e.,

$$\mathcal{Y}(k) := \{u(k), u(k-1), \ldots, y(k), y(k-1), \ldots\} \tag{11}$$

Based on the E.L.S. parameter estimates, let the one step ahead prediction for the model be

$$\hat{y}_i(k+1) = \mathrm{E}\big[y_i(k+1) \mid \mathcal{Y}(k)\big] = \big[\varphi^{(i)}(k)\big]^\top \hat{\theta}^{(i)}(k) \tag{12}$$

Here, $\mathrm{E}(\cdot)$ represents the expectation operator. To facilitate the development of the dual controller we further assume that the E.L.S. algorithm asymptotically generates unbiased (or consistent) estimates of the model parameters. A sufficient condition for the convergence of the pseudo-linear regression type methods for an ARMAX model (under the ideal conditions) can be found in Söderström and Stoica (1989). The convergence of the E.L.S. estimates to the true parameters implies that the innovation sequence $\{\varepsilon(k)\}$ asymptotically converges to $\{e(k)\}$. In practice, however, the model error may approach a very small value if the model order is chosen appropriately.

2.2 *An objective function for dual M.P.C.*

The objective for M.P.C.-based dual control can be stated as finding the control sequence $\{u(k), u(k+1), \ldots\}$ that minimizes

$$J_\infty = \mathrm{E}\Bigg[\sum_{j=k+1}^{\infty}\Bigg\{\sum_{i=1}^{r} w_i E_i(j)^2 + \sum_{i=1}^{r}\mu_i u_i(j-1)^2\Bigg\} \;\Bigg|\; \mathcal{Y}(k)\Bigg] \tag{13}$$

given data obtained up to time k, where

$$E_i(j) = r_i(j) - y_i(j) \tag{14}$$

Here, $w_i > 0$ and $\mu_i \geq 0$ are weighting parameters and $r(j)$ represents the output reference or the setpoint vector at the future time instant j. Since we assume that the system is stably invertible we can set $\mu_i = 0$ for all i and work with the control objective

$$J_\infty = \mathrm{E}\Big[\sum_{j=k+1}^{\infty} \sum_{i=1}^{r} w_i E_i(j)^2 \;\Big|\; \mathcal{Y}(k) \Big] \tag{15}$$

The main difficulty in using this objective function is the lack of a model that can accurately predict future outputs $y(k+j)$. Thus, the objective function needs to be reformulated for simultaneous probing and control. To achieve this we first rewrite the objective function as

$$J_\infty = \mathrm{E}\Big[\sum_{j=k+1}^{k+2} \sum_{i=1}^{r} w_i E_i(j)^2 \;\Big|\; \mathcal{Y}(k) \Big] + \mathrm{E}\Big[\sum_{j=k+3}^{\infty} \sum_{i=1}^{r} w_i E_i(j)^2 \;\Big|\; \mathcal{Y}(k) \Big] \tag{16}$$

Now, consider the first term, $J_1 = \sum_{i=1}^{r} w_i J_{1i}$ where

$$J_{1i} = \mathrm{E}\big[E_i(k+1)^2 \mid \mathcal{Y}(k) \big] \tag{17}$$

By adding and subtracting the model $\hat{y}_i(k+1)$, we can rewrite equation (17) as

$$J_{1i} = \mathrm{E}\big[(r_i(k+1) - \hat{y}_i(k+1) + \delta y_i(k+1))^2 \mid \mathcal{Y}(k) \big] \tag{18}$$

where

$$\delta y_i(k+1) = \hat{y}_i(k+1) - y_i(k+1) \tag{19}$$

Dropping the conditional expectation notation for the sake of simplicity and using the assumption that $\varepsilon_i(k) \to e_i(k)$ asymptotically, the term J_{1i} can be expressed as

$$\begin{aligned} J_{1i} = \mathrm{E}\big[&(r_i(k+1) - \hat{y}_i(k+1))^2 + (\delta y_i(k+1))^2 \\ &- 2(r_i(k+1) - \hat{y}_i(k+1))(\delta y_i(k+1)) \big] \end{aligned} \tag{20}$$

where

$$\delta y_i(k+1) = \big[\varphi^{(i)}(k) \big]^\top \delta\theta^{(i)}(k) - \varepsilon_i(k+1) \tag{21}$$

and $\delta\theta^{(i)}(k) = \hat{\theta}^{(i)}(k) - \theta^{(i)}(k)$. The first term is deterministic and the third term is zero since the E.L.S. algorithm is assumed to be an asymptotically unbiased estimator; i.e.,

$$\mathrm{E}\big[\delta\theta^{(i)}(k) \big] = 0 \quad \text{and} \quad \mathrm{E}\big[\varepsilon_i(k+1) \big] = 0 \tag{22}$$

Since $\delta\theta^{(i)}(k)$ and $\varepsilon_i(k+1)$ are independent, using

$$\mathrm{Cov}\big[\delta\theta^{(i)}(k)\big] =: P^{(i)}(k) \qquad \text{and} \qquad \mathrm{Cov}\big[\varepsilon_i(k)\big] =: \sigma_i^2 \tag{23}$$

we can write

$$J_{1i} = \big(r_i(k+1) - \big[\varphi^{(i)}(k)\big]^\top \hat{\theta}^{(i)}(k)\big)^2 + \big[\varphi^{(i)}(k)\big]^\top P^{(i)}(k)\varphi^{(i)}(k) + \sigma_i^2 \tag{24}$$

given $\mathcal{Y}(k)$. Note that σ_i is not a known quantity, but since it appears as a constant term in the objective function its value does not matter and can be treated as zero. To achieve the probing effect, the same reformulation that is used to approximate the cost function is used in the second stage; i.e.,

$$J_{2i} \approx \big(r_i(k+2) - \big[\varphi^{(i)}(k+1)\big]^\top \theta^{(i)}(k+1)\big)^2 + \big[\varphi^{(i)}(k+1)\big]^\top P^{(i)}(k+1)\varphi^{(i)}(k+1) + \sigma_i^2$$

We can further simplify J_∞ by truncating the infinite horizon to some finite number N. The modified approximate cost function can be expressed as

$$J_N \approx \sum_{j=k}^{k+1}\sum_{i=1}^{r}\Big\{ w_i\big(r_i(j+1) - \big[\varphi^{(i)}(j)\big]^\top \theta^{(i)}(j)\big)^2 + w_i\big(\big[\varphi^{(i)}(j)\big]^\top P^{(i)}(j)\varphi^{(i)}(j) + \sigma_i^2\big)\Big\} + \mathrm{E}\Big[\sum_{j=k+2}^{k+N}\sum_{i=1}^{r} w_i\,(E_i(j+1))^2 \;\Big|\; \mathcal{Y}(k)\Big] \tag{25}$$

Since we intend to use M.P.C., we further approximate the last term in J_N using the model predictions instead of expected values of the outputs, which yields a cost function

$$V_N = \sum_{j=k}^{k+1}\sum_{i=1}^{r}\Big\{ w_i\big(r_i(j+1) - \big[\varphi^{(i)}(j)\big]^\top \theta^{(i)}(j)\big)^2 + w_i\big(\big[\varphi^{(i)}(j)\big]^T P^{(i)}(j)\varphi^{(i)}(j) + \sigma_i^2\big)\Big\} + \sum_{j=k+2}^{k+N}\sum_{i=1}^{r} w_i\,(r_i(j+1) - \hat{y}_i(j+1 \mid k))^2 \tag{26}$$

Note that the covariance matrices $\{P^{(i)}(k+1) : i = 1, \ldots, r\}$ and future regressor vectors $\{\varphi^{(i)}(k+1) : i = 1, \ldots, r\}$ are functions of $u(k)$. As a consequence, the modified optimization objective rewards inputs that reduce the future covariance $P^{(i)}(k+1)$. In other words, the controller injects inputs that improve the quality of the parameter estimates and thereby reduce the parameter uncertainty.

2.3 *Output prediction*

In the proposed adaptive M.P.C. formulation, the identified models are used for predicting future outputs. Consider a scenario at the kth sampling instant, when given the N future inputs

$$\mathcal{U}_k := \{u(k \mid k), u(k+1 \mid k), \ldots, u(k+N-1 \mid k)\} \tag{27}$$

we want to predict outputs over time window $[k+1, k+N]$. Since the future parameter vectors and future innovations are unavailable at time k we have to make further simplifying assumptions to carry out predictions using the proposed model.

- Given the information at time k, the expected value of the unknown parameters in model i is $\hat{\theta}^{(i)}(k)$. Hence, the model outputs are predicted with
$$\hat{\theta}^{(i)}(k+j \mid k) = \hat{\theta}^{(i)}(k) \text{ for } j > 0 \text{ and for all } i \tag{28}$$
- Consistent with conventional M.P.C. formulations, we assume the following for the future innovations for output prediction:
$$\varepsilon_i(k+j+1) = \varepsilon_i(k+j) \text{ for } i = 1, 2, \ldots, r \tag{29}$$
where $j = 0, 1, \ldots, N-1$. However, a difficulty with this approach is that the sequence $\{\varepsilon_i(k)\}$ contains high frequency noise, which can lead to noisy predictions. Thus, to eliminate the effect of the high frequency noise on the predictions and limit the frequency range of the model plant mismatch, we use a unity gain innovation filter for each innovation sequence (Madakyaru, Narang, and Patwardhan, 2009):
$$\varepsilon_{f,i}(k) = \alpha_i \varepsilon_{f,i}(k-1) + (1-\alpha_i)\varepsilon_i(k) \tag{30}$$
for $i = 1, \ldots, r$ and with $0 < \alpha_i < 1$ being tuning parameters. Thus, the future innovation terms in $\varphi^{(i)}(k+j)$ are estimated as
$$\varepsilon_i(k+j \mid k) = \varepsilon_{f,i}(k) \text{ for } j > 0 \tag{31}$$

With the above simplifying assumptions, the predicted output for the ith MISO ARMAX model at time $k+j$ can be expressed

$$\hat{y}_i(k+j+1 \mid k) = \left[\varphi^{(i)}(k+j \mid k)\right]^\top \hat{\theta}^{(i)}(k) \tag{32}$$

where the predicted regressor vector is

$$\begin{aligned}\hat{\varphi}^{(i)}(k+j \mid k) = \big[&-\hat{y}_i(k+j \mid k), \ldots, -\hat{y}_i(k+1 \mid k),\\ &-y_i(k), \ldots, -y_i(k+j-n+1),\\ &u_1(k+j \mid k), \ldots, u_m(k+j \mid k), \ldots, u_m(k+j-n),\\ &\varepsilon_i(k+j \mid k), \ldots, \varepsilon_i(k+j-n)\big]^\top\end{aligned} \tag{33}$$

for $j = 0, 1, \ldots, N-1$ and $i = 1, 2, \ldots, r$. Note that $\varepsilon_i(k+j)$ for $j \leq 0$ are available at instant k and are directly used in formulating $\varphi^{(i)}(k+j \mid k)$.

2.4 *Adaptive D.M.P.C. formulation*

Based on the modified cost function (26) and the proposed prediction model (32), an adaptive M.P.C. scheme is proposed as follows

$$\begin{aligned}\min_{\mathcal{U}_k} V_N(k) = &\sum_{j=k+1}^{k+N} E(j)^\top W_E E(j)\\ &+ \sum_{j=k}^{k+1} \sum_{i=1}^{r} w_i \left[\varphi^{(i)}(j \mid k)\right]^\top P^{(i)}(j) \varphi^{(i)}(j \mid k)\\ &+ \sum_{j=k+3}^{k+N} \Delta u^\top(j) W_{\Delta u} \Delta u(j)\end{aligned} \tag{34}$$

where $\Delta u(j) = u(j) - u(j-1)$, $E(j) = r(j) - \hat{y}(j \mid k)$, and

$$\hat{y}_i(j \mid k) = \left[\varphi^{(i)}(j-1 \mid k)\right]^\top \hat{\theta}^{(i)}(k) \tag{35}$$

for $i = 1, \ldots, r$, subject to the following constraints

$$P^{(i)}(k+1) = \left[I - L_i(k+1)\varphi^{(i)}(k))\right] P^{(i)}(k) \tag{36a}$$

$$L_i(k+1) = P^{(i)}(k)\varphi^{(i)}(k) \times \left[1 + \varphi^{(i)}(k)^\top P^{(i)}(k) \varphi^{(i)}(k)\right]^{-1} \tag{36b}$$

$$\text{for } i = 1, 2, \ldots, r$$

$$\Delta u_{\min} \leq \Delta u(j) \leq \Delta u_{\max} \tag{36c}$$

$$\Delta u(j) = 0, \text{ for } j = k + N_c, \ldots, k + N - 1 \tag{36d}$$

Here, N_c is the control horizon and $W_E = \text{diag}[w_1, \ldots, w_r]$ is the tracking error weighting matrix. Note that we introduced an input-move suppression term with a corresponding tuning matrix $W_{\Delta u} \geq 0$, and that this can be used to adjust the intensity of the probing effect. That is, the weighting matrix $W_{\Delta u}$ can be used to counteract excessively large input changes that can otherwise occur when $P^{(i)}(k)$ is large. Also note that the proposed adaptive D.M.P.C. formulation results in a constrained non-convex optimization problem that has to be solved with a nonlinear programming (N.L.P.) solver.

3 EXPERIMENTAL EVALUATION

We now demonstrate an experimental evaluation of the proposed adaptive D.M.P.C. algorithm carried out using the benchmark Continuous Stirred Tank Heater (C.S.T.H.) system (Thornhill, Patwardhan, and Shah, 2008) at the Automation Lab in the Chemical Engineering Department at I.I.T. Bombay.

3.1 *Plant description*

The C.S.T.H. setup consists of two tanks in series as shown in Figure 1. The cold water flow (F_1) from the reservoir is heated using a 4 kW h heating coil in Tank 1. The water level in Tank 1 remains constant and the hot water overflows to Tank 2 where it is mixed with cold water flow F_2. The water in Tank 2 can be heated using another 3.5 kW h heating coil. To make the system more complex and interactive, a recycle flow (F_R) is set up from the bottom of Tank 2 to Tank 1 using a metering pump. Cold water inflows to both the tanks can be manipulated using pneumatic control valves CV-1 and CV-2. Also, the heat input to both heaters can be manipulated using two thyristor power controller (T.P.C.) systems, which are driven by 4 mA to 20 mA current inputs. From a control viewpoint the C.S.T.H. is a MIMO system with three manipulated inputs (4 mA to 20 mA current inputs to TPC 1 (u_4), TPC 2 (u_5), and to CV-2 (u_2)), and three controlled outputs (temperature in Tank 1 (T_1), temperature in Tank 2 (T_2) and water level in Tank 2 (h_2)). The current input to CV-1, which can be used to manipulate the cold water flow to Tank 1, and the temperature of the cold water inflows both act as unmeasured disturbances. This setup is controlled with a P.C. (with an Intel Core i5 processor and 8 GiB RAM) using a combination of LabView

version 2012 and MATLAB. A sampling interval of 5 seconds is used in this work for carrying out identification and control studies.

In the experimental study, the level in the second tank (h_2) is maintained at 50 % (i.e., 20 cm) using a P.I. controller ($k_c = 1.723$ and $\tau_I = 2$ min), which manipulates current input to control valve (CV-2). The inputs to CV-1 and the recycle flow metering pump are kept constant at 50 % levels. Thus, for the evaluation of the adaptive D.M.P.C., the system is reduced to a 2×2 configuration with u_4 and u_5 as manipulated inputs and the tank temperatures (T_1 and T_2) as the controlled outputs.

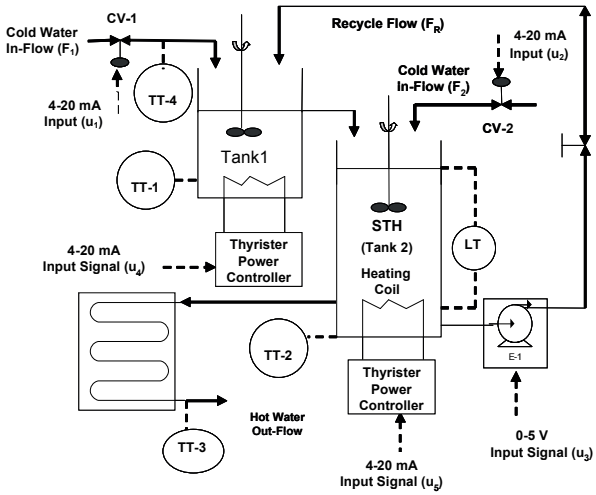


Figure 1: Schematic diagram of the C.S.T.H.

3.2 *Closed-loop studies*

Before implementing the proposed adaptive D.M.P.C. it is necessary to decide a suitable ARMAX model structure. We perturbed the C.S.T.H. system in open loop by simultaneously introducing low frequency pseudo-random binary sequences (P.B.R.S.) in the heating inputs to the tanks. We used the resulting data to identify an ARMAX model using the System Identification Toolbox in MATLAB. Using first order MISO ARMAX models were sufficient for ensuring that the innovation sequences $\{e_i(k)\}$ are white noise for each output. However, a minimum of 9th and 12th order MISO A.R.X. models were needed to obtain white noise innovation sequences for T_1 and T_2, respectively. This may be attributed to fact that

the C polynomials in the identified MISO ARMAX models have a pole close to 0.88.

We developed and implemented the adaptive D.M.P.C. on the C.S.T.H. system using two MISO second order ARMAX models. Each of these ARMAX models is of the form

$$\left(1+\sum_{i=1}^{2} a_i q^{-i}\right) y(k) = \sum_{j=1}^{2}\sum_{i=1}^{2} b_{ji} q^{-i} u_j(k) + \left(1+\sum_{i=1}^{2} c_i q^{-i}\right) e(k) \quad (37)$$

The tuning parameters used for the adaptive D.M.P.C. formulation are set to $N = 60$, $N_c = 6$, $\alpha_i = 0.9$ for all i, $W_E = I$, and $W_{\Delta U} = \mathrm{diag}\big[2,\ 1\big]$,

$$u_{\min} = \big[4,\ 4\big]^\top, \quad \text{and} \quad u_{\max} = \big[20,\ 20\big]^\top \quad (38)$$

The initial model identified from the open loop data was used to initialize the parameter estimators and the initial covariance matrices were selected as $P^{(1)}(0) = P^{(2)}(0) = 10^4 I$. We deliberately set the initial covariances to high numbers and the adaptive D.M.P.C. was started when the parameter estimates stabilized and the covariances reduced significantly. The system was controlled using the conventional (non-adaptive) M.P.C. that the initial model employed for predictions prior to starting the adaptive D.M.P.C. The adaptive D.M.P.C. was implemented using the constrained N.L.P. solver FMINCON from the MATLAB Optimization Toolbox. Average computation time for the adaptive D.M.P.C. computations at each sampling instant was found to be 0.5482 seconds.

The closed loop experiments consist of (a) a sequence of positive and negative setpoint changes in both tank temperatures (a servo problem) and (b) a large magnitude step change in the cold water inflow to Tank 1 (a regulatory problem). Performance of the adaptive D.M.P.C. for the servo problem is presented in Figure 2 and the corresponding profiles of the manipulated inputs are presented in Figure 3. As shown in Figure 2, the controller is able to achieve quick transitions to the desired setpoint and settle the reference temperatures without any offset. The probing effect of the proposed adaptive D.M.P.C. formulation is visible in Figure 3, where time-varying low-amplitude perturbations are introduced after switching to adaptive D.M.P.C. from conventional M.P.C. These perturbations of varying intensity are continuously introduced throughout the experiment with adaptive D.M.P.C. Since the high-frequency excitation may increase actuator wear, an operator may consider turning off the dual feature if it is deemed unnecessary based on some performance criterion. Since the high-frequency excitation may increase

actuator wear, an operator may consider turning off the dual feature if it is deemed unnecessary based on some performance criterion. Note that the manipulated input profiles generated by the conventional M.P.C. are smoother and without any such excitation.

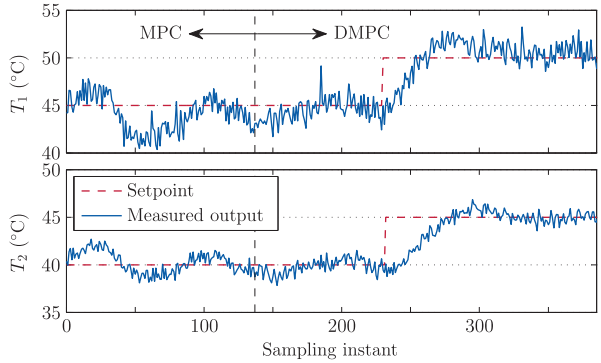


Figure 2: C.S.T.H. experiment — controller performance: setpoint tracking.

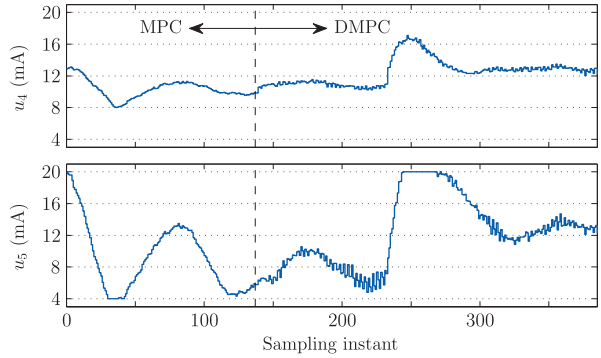


Figure 3: C.S.T.H. experiment — controller performance: manipulated inputs.

4 CONCLUSION

In this work, we develop a MIMO adaptive D.M.P.C. using ARMAX models. The efficacy of the proposed control scheme is evaluated by conducting experimental studies on the benchmark heater-mixer setup. We show that despite the complexity of the algorithm, we are able to implement the controller for real-time control with a fairly standard implementation and achieve fast control input computation. Analysis

of the experimental results reveals that if the tuning parameters are selected carefully, the proposed adaptive D.M.P.C. is able to inject input perturbations that are sufficient for maintaining the health of the on-line parameter estimators. When the system is operating at a fixed setpoint, these fluctuations are found to be of variable and low amplitudes, thereby introducing minimal disturbance in the plant operation. Though initial experimental studies have shown promising results, a number of issues remain to be resolved. The ARMAX structure leads to nonlinear parameter models and the E.L.S. algorithm is a nonlinear estimator. Thus, alternate model structures that are parsimonious in parameters are currently being examined for cases in which the system is subjected to correlated unmeasured disturbances.

Part III

AFTERWORD

CONCLUSIONS, HINDSIGHT, AND THE FUTURE

CONCLUDING REMARKS AND HINDSIGHT

In this thesis I present three similar, yet fundamentally different, approaches to (heuristic) dual control, each of which is based on the receding-horizon principle. The approach from Paper A solves a problem that is different from, while inspired by, the dual; it improves performance through utilizing the dual effect and actively exciting the process. We develop a semi-exact dual controller in Paper B through transforming the stage cost and applying the result beyond its temporally valid region. In Paper B we suggest an extension of this algorithm to MIMO ARMAX models and apply the result to an experimental setup, showing that there is potential for using multivariable M.P.C. with dual effect on moderately challenging real-time problems. The dual controller my coworkers and I develop in Paper C includes probabilistic chance constraints and is based on an exact reformulation of the dual objective with respect to a specific subset of the future information. I consider this last contribution the most significant and see great potential for future research that may result in useful extensions and insights.

The terminology used in the dual control literature does, arguably, suffer from a lack of consensus and standardization. I have attempted to clarify some of the terms and associated nuances with Definitions 1 and 2 in Chapter 3. While one may argue that there are other definitions and different terminology that serve the field better, I maintain that greater consensus and clarity would be of significant help in both conducting and communicating research. While the discussion in Chapter 3 on implicit dependence on future information in the Bellman equation is brief and somewhat superficial, I consider this a fair attempt at clarifying certain aspects of an important issue worthy of more attention.

In hindsight, I should have directed more of my research focus at an earlier point in time toward the concept of future information and its role in the dual control problem. In my opinion, this type of approach shows great promise, but much remains to be learned about its extendibility as well as its stability and convergence properties. The most valuable results from the research presented in this thesis is in

my judgment the introduction of the subset of future information, the reformulation in Theorem 4 and the formulation of the Q.C.Q.P. (P″), all in Paper C. The resulting insight and the fact that the cost functions for nominal control and uncertainty reduction so clearly appear is remarkable, makes a strong statement on the lack of conflicting goals, and can serve as inspiration for heuristic methods for systems where no exact results exist. It is also remarkable that the optimal-control problem can be stated as a Q.C.Q.P. with the introduction of a small number of variables.

THOUGHTS AND RECOMMENDATIONS FOR FUTURE WORK

I see the most potential in continued research on the dual controller from Paper C. The formulation is at a degree of maturity where it makes sense to investigate convergence and stability; Martingale convergence theory (Williams, 1991) is to me the obvious place to start. The performance of the algorithm for larger systems and with other types of orthonormal basis functions should be investigated. There is a strong possibility that the importance of global solutions to the Q.C.Q.P. problem increases with system size, and there is great potential for improving the performance of global optimization software for this problem, both in terms of refining the problem formulation and tailoring the optimization algorithm. The possibility of better representing the future through a larger subset of future information should be investigated; the algorithm does not account for future system outputs in its current form. I would like to see a closer link established between this algorithm and the Bellman equation, possibly providing insights that guide the development of future improvements.

The case for dual control would be easier to make with successful applications to demonstrate its effectiveness. There are many systems with reducible decision-relevant uncertainty that are unlikely to benefit from dual control and more effort should be put into identifying when dual control comes with the largest potential for improving performance.

For optimal control of linear systems with a quadratic cost, the dynamic-programming solution can be obtained with a receding-horizon solution and quadratic programming (M.P.C.). The numeric dynamic-programming solution to dual-control problems is the closest we can get to the ideal dual solution. An important question that warrants thorough investigation is whether this solution can be obtained with a receding-horizon approach. My sense is that it is not possible to obtain

the dynamic-programming solution using nonlinear programming on a receding horizon, which does not mean the receding-horizon solution cannot get very close.

BIBLIOGRAPHY

Allison, B.J., Ciarniello, J.E., Tessier, P.J.C., and Dumont, G.A. (1995). "Dual adaptive control of chip refiner motor load." *Automatica* 31(8), pp. 1169–1184.

Anderson, B.D.O. (1985). "Adaptive systems, lack of persistency of excitation and bursting phenomena." *Automatica* 21(3), pp. 247–258.

Åström, K.J. (1983). "Theory and applications of adaptive control—a survey." *Automatica* 19(5), pp. 471–486.

Åström, K.J. and Helmersson, A. (1986). "Dual control of an integrator with unknown gain." *Computers and Mathematics with Applications* 12(6), pp. 653–662.

Åström, K.J. and Kumar, P.R. (2014). "Control: A perspective." *Automatica* 50(1), pp. 3–43.

Åström, K.J. and Wittenmark, B. (1995). *Adaptive Control*. 2nd ed. Dover.

Bar-Shalom, Y. (1981). "Stochastic dynamic programming: Caution and probing." *IEEE Transactions on Automatic Control* 26(5), pp. 1184–1195.

Bar-Shalom, Y. and Tse, E. (1974). "Dual effect, certainty equivalence, and separation in stochastic control." *IEEE Transactions on Automatic Control* 19(5), pp. 494–500.

Bayard, D.S. and Schumitzky, A. (2010). "Implicit dual control based on particle filtering and forward dynamic programming." *International Journal of Adaptive Control and Signal Processing* 24(3), pp. 155–177.

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

Bertsekas, D.P. (2012). *Dynamic Programming and Optimal Control: Approximate Dynamic Programming*. 4th ed. Vol. 2. Athena Scientific.

Blanke, M., Kinnaert, M., Lunze, J., and Staroswiecki, M. (2006). *Diagnosis and Fault-Tolerant Control*. 2nd ed. Springer.

Bond, C.A. and Loomis, J.B. (2009). "Using numerical dynamic programming to compare passive and active learning in the adaptive management of nutrients in the adaptive management of nutrients in shallow lakes." *Canadian Journal of Agricultural Economics* 57(4), pp. 555–573.

Brumm, J., Mikushin, D., Scheidegger, S., and Schenk, O. (2015). "Scalable high-dimensional dynamic stochastic economic modeling." *Journal of Computational Science* 11, pp. 12–25.

Cervantes, A.M., Wächter, A., Tütüncü, R.H., and Biegler, L.T. (2000). "A reduced space interior point strategy for optimization of differential algebraic systems." *Computers and Chemical Engineering* 24(1), pp. 39–51.

Feldbaum, A.A. (1961a). "Dual control theory. II." *Automation and Remote Control* 21(11), pp. 1033–1039.

——— (1961b). "Dual-control theory. I." *Automation and Remote Control* 21(9), pp. 874–880.

——— (1961c). "The theory of dual control. III." *Automation and Remote Control* 22(1), pp. 1–12.

——— (1961d). "The theory of dual control. IV." *Automation and Remote Control* 22(2), pp. 109–121.

——— (1965). *Optimal Control Systems*. Elsevier.

Finn, C.K., Wahlberg, B., and Ydstie, B.E. (1993). "Constrained predictive control using orthogonal expansions." *AIChE Journal* 39(11), pp. 1810–1826.

Franceschini, G. and Macchietto, S. (2008). "Model-based design of experiments for parameter precision: State of the art." *Chemical Engineering Science* 63(19), pp. 4846–4872.

GAMS Development Corporation (2012). *The General Algebraic Modeling System*.

——— (2015). *The General Algebraic Modeling System*.

Genceli, H. and Nikolaou, M. (1996). "New approach to constrained predictive control with simultaneous model identification." *AIChE Journal* 42(10), pp. 2857–2868.

Gevers, M. (2002). "A decade of progress in iterative process control design: from theory to practice." *Journal of Process Control* 12(4), pp. 519–531.

——— (2006). "A personal view of the development of system identification." *IEEE Control Systems Magazine* 26(6), pp. 93–105.

Gevers, M., Bombois, X., Hildebrand, R., and Solari, G. (2011). "Optimal experiment design for open and closed-loop system identification." *Communications in Information and Systems* 11(3), pp. 197–224.

Gevers, M. and Ljung, L. (1986). "Optimal experiment designs with respect to the intended model application." *Automatica* 22(5), pp. 543–554.

Gill, P.E., Murray, W., and Saunders, M.A. (2005). "SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization." *SIAM Review* 47(1), pp. 99–131.

Groeneveld, R.A., Springborn, M., and Costello, C. (2014). "Repeated experimentation to learn about a flow-pollutant threshold." *Environmental and Resource Economics* 58(4), pp. 627–647.

Heirung, T.A.N. (2014a). "Optimal sequential decision making under uncertainty: anticipatory learning with stochastic dynamic programming." In: *Process system engineering seminar, Carnegie Mellon University*. Pittsburgh, PA.

——— (2014b). "Some methods for approximate dual control based on MPC." In: *Seminar at Prof. H.G. Bock's research group at the University of Heidelberg*. Heidelberg, Germany.

——— (2015a). "Optimal control of uncertain systems using Dual Model Predictive Control (DMPC)." In: *Seminar at the Ilmenau University of Technology*. Ilmenau, Germany.

——— (2015b). "Some approaches to adaptive dual control using MPC." In: *Seminar at Prof. Ali Mesbah's research group at UC Berkeley*. Berkeley, CA.

Heirung, T.A.N., Foss, B., and Ydstie, B.E. (2015). "MPC-based dual control with online experiment design." *Journal of Process Control* 32, pp. 64–76.

Heirung, T.A.N., Morinelly, J.E., Foss, B., and Ydstie, B.E. (2015). "Dual Model Predictive Control (DMPC): State of the Art." In: *AIChE Annual Meeting*. Salt Lake City, UT.

Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2012a). "Optimal input design for parameter identification in dynamic systems using nonlinear programming." In: *Nordic Process Control Workshop*. Kgs Lyngby, Denmark.

——— (2012b). "Towards dual MPC." In: *Nonlinear Model Predictive Control*. Noordwijkerhout, the Netherlands, pp. 502–507.

——— (2012c). "Towards model predictive dual control." In: *AIChE Annual Meeting*. Pittsburgh, PA.

——— (2013a). "An adaptive model predictive dual controller." In: *Adaptation and Learning in Control and Signal Processing*. Caen, France, pp. 62–67.

——— (2013b). "An MPC approach to dual control." In: *Dynamics and Control of Process Systems*. Mumbai, India, pp. 69–74.

——— (2013c). "MPC-based adaptive dual control." In: *AIChE Annual Meeting*. San Francisco, CA.

——— (2014). "Optimal control of uncertain systems using dual model predictive control (DMPC)." In: *AIChE Annual Meeting*. Atlanta, GA.

Heirung, T.A.N., Ydstie, B.E., and Foss, B. (2015a). "A QCQP problem for dual control of FIR systems." In: *The British-French-German Conference on Optimization*. London, UK.

——— (2015b). "Dual adaptive model-predictive control." *Under revision for resubmission to Automatica*.

——— (2015c). "Dual MPC for FIR systems: Information anticipation." In: *Advanced Control of Chemical Processes*. Whistler, Canada, pp. 1034–1039.

——— (2015d). "Optimal control of uncertain systems using Dual Model Predictive Control (DMPC)." In: *Nordic Process Control Workshop*. Trondheim, Norway.

Heuberger, P.S.C., Van den Hof, P.M.J., and Wahlberg, B. (2005). *Modelling and Identification with Rational Orthogonal Basis Functions*. Springer.

Isaksson, A.J. (2013). "Some aspects of industrial system identification." In: *Dynamics and Control of Process Systems*. Mumbai, India, pp. 153–159.

Ismail, A., Dumont, G.A., and Backstrom, J. (2003). "Dual adaptive control of paper coating." *IEEE Transactions on Control Systems Technology* 11(3), pp. 289–309.

Janka, D. (2015). "Sequential quadratic programming with indefinite Hessian approximations for nonlinear optimum experimental design for parameter estimation in differential-algebraic equations." PhD thesis. Heidelberg University.

Johnstone, R.M., Johnson Jr., C.R., Bitmead, R.R., and Anderson, B.D.O. (1982). "Exponential convergence of recursive least squares with exponential forgetting factor." *Systems & Control Letters* 2(2), pp. 77–82.

Kirk, D.E. (1970). *Optimal Control Theory: An Introduction*. Dover.

Kumar, K., Heirung, T.A.N., Patwardhan, S.C., and Foss, B. (2015). "Experimental evaluation of a MIMO adaptive dual MPC." In: *Advanced Control of Chemical Processes*. Whistler, Canada, pp. 546–551.

Larson, R.E. (1965). "Dynamic programming with reduced computational requirements." *IEEE Transactions on Automatic Control* 10(2), pp. 135–143.

Larsson, C.A., Annergren, M., Hjalmarsson, H., Rojas, C.R., Bombois, X., Mesbah, A., and Modén, P.E. (2013). "Model predictive control with integrated experiment design for output error systems." In: *European Control Conference*. Zürich, Switzerland, pp. 3790–3795.

Larsson, C.A., Rojas, C.R., Bombois, X., and Hjalmarsson, H. (2015). "Experimental evaluation of model predictive control with excitation

(MPC-X) on an industrial depropanizer." *Journal of Process Control* 31, pp. 1–16.

Lee, J.M. and Lee, J.H. (2009). "An approximate dynamic programming based approach to dual adaptive control." *Journal of Process Control* 19(5), pp. 859–864.

Lempert, R.J. and Collins, M.T. (2007). "Managing the risk of uncertain threshold responses: comparison of robust, optimum, and precautionary approaches." *Risk analysis* 27(4), pp. 1009–1026.

Ljung, L. (1999). *System Identification. Theory for the User*. 2nd ed. Prentice-Hall.

Lucia, S. and Paulen, R. (2014). "Robust nonlinear model predictive control with reduction of uncertainty via robust optimal experiment design." In: *IFAC World Congress*. Cape Town, South Africa, pp. 1904–1909.

Madakyaru, M., Narang, A., and Patwardhan, S.C. (2009). "Development of ARX models for predictive control using fractional order and orthonormal basis filter parametrization." *Industrial & Engineering Chemistry Research* 48(19), pp. 8966–8979.

Marafioti, G. (2010). "Enhanced Model Predictive Control: Dual Control Approach and State Estimation Issues." PhD thesis. Norwegian University of Science and Technology.

Marafioti, G., Bitmead, R.R., and Hovd, M. (2014). "Persistently exciting model predictive control." *International Journal of Adaptive Control and Signal Processing* 28(6), pp. 536–552.

Marafioti, G., Stoican, F., Bitmead, R.R., and Hovd, M. (2012). "Persistently exciting model predictive control for SISO systems." In: *IFAC Nonlinear Model Predictive Control Conference*. Noordwijkerhout, the Netherlands, pp. 448–453.

Mareels, I., Bitmead, R.R., Gevers, M., Johnson, C.R., Kosut, R.L., and Poubelle, M.A. (1987). "How exciting can a signal really be?" *Systems & Control Letters* 8(3), pp. 197–204.

Mareels, I. and Polderman, J.W. (1996). *Adaptive Systems: An Introduction*. Birkhäuser.

Mårtensson, B. (1985). "The order of any stabilizing regulator is sufficient a priori information for adaptive stabilization." *Systems & Control Letters* 6(2), pp. 87–91.

Mayne, D.Q. (2014). "Model predictive control: Recent developments and future promise." *Automatica* 50(12), pp. 2967–2986.

Mayne, D.Q., Rawlings, J.B., Rao, C.V., and Scokaert, P.O.M. (2000). "Constrained model predictive control: Stability and optimality." *Automatica* 36(6), pp. 789–814.

Misener, R. and Floudas, C.A. (2013). "GloMIQO: Global mixed-integer quadratic optimizer." *Journal of Global Optimization* 57(1), pp. 3–50.

Morari, M. and Lee, J.H. (1999). "Model predictive control: Past, present and future." *Computers and Chemical Engineering* 23, pp. 667–682.

Patwardhan, R.S. and Goapluni, R.B. (2014). "A moving horizon approach to input design for closed loop identification." *Journal of Process Control* 24(3), pp. 188–202.

Qin, S.J. and Badgwell, T.A. (2003). "A survey of industrial model predictive control technology." *Control Engineering Practice* 11(7), pp. 733–764.

Qin, S.J. and Yu, J. (2007). "Recent developments in multivariable controller performance monitoring." *Journal of Process Control* 17(3), pp. 221–227.

Radenković, M.S. and Ydstie, B.E. (1995). "Using persistent excitation with fixed energy to stabilize adaptive controllers and obtain hard bounds for the parameter estimation error." *SIAM Journal on Control and Optimization* 33(4), p. 1224.

Rathouský, J. and Havlena, V. (2013). "MPC-based approximate dual controller by information matrix maximization." *International Journal of Adaptive Control and Signal Processing* 27(11), pp. 974–999.

Shouche, M.S., Genceli, H., and Nikolaou, M. (2002). "Effect of on-line optimization techniques on model predictive control and identification (MPCI)." *Computers & Chemical Engineering* 26(9), pp. 1241–1252.

Shouche, M.S., Genceli, H., Vuthandam, P., and Nikolaou, M. (1998). "Simultaneous constrained model predictive control and identification of DARX processes." *Automatica* 34(12), pp. 1521–1530.

Smith, L.A. and Stern, N. (2011). "Uncertainty in science and its role in climate policy." *Philosophical Transactions of The Royal Society, Series A* 369(1956), pp. 1–24.

Söderström, T. and Stoica, P. (1989). *System Identificatin*. Prentice Hall.

Sun, Z., Qin, S.J., Singhal, A., and Megan, L. (2013). "Performance monitoring of model-predictive controllers via model residual assessment." *Journal of Process Control* 23(4), pp. 473–482.

Sutton, R.S. and Barto, A.G. (1998). *Reinforcement Learning. An Introduction*. The MIT Press.

Tanaskovic, M., Fagiano, L., Smith, R., and Morari, M. (2014). "Adaptive receding horizon control for constrained MIMO systems." *Automatica* 50(12), pp. 3019–3029.

Tawarmalani, M. and Sahinidis, N.V. (2002). *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming. Theory, Algorithms, Software, and Applications*. Kluwer.

——— (2005). "A polyhedral branch-and-cut approach to global optimization." *Mathematical Programming* 103(2), pp. 225–249.

Thornhill, N.F., Patwardhan, S.C., and Shah, S.L. (2008). "A continuous stirred tank heater simulation model with applications." *Journal of Process Control* 18(3-4), pp. 347–360.

Tse, E. and Bar-Shalom, Y. (1973). "An actively adaptive control for linear systems with random parameters via the dual control approach." *IEEE Transactions on Automatic Control* 18(2), pp. 109–117.

Tse, E., Bar-Shalom, Y., and Meier, L. (1973). "Wide-sense adaptive dual control for nonlinear stochastic systems." *IEEE Transactions on Automatic Control* 18(2), pp. 98–108.

Unbehauen, H. (2000). "Adaptive dual control systems: a survey." In: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pp. 171–180.

Van den Hof, P.M.J., Heuberger, P.S.C., and Bokor, J. (1995). "System identification with generalized orthonormal basis functions." *Automatica* 31(12), pp. 1821–1834.

Wahlberg, B. (1991). "System Identification Using Laguerre Models." *IEEE Transactions on Automatic Control* 36(5), pp. 551–562.

——— (1994). "System identification using Kautz models." *IEEE Transactions on Automatic Control* 39(6), pp. 1276–1282.

Wahlberg, B. and Lindskog, P. (1990). "Approximate modeling by means of orthonormal functions." In: *Modeling, Estimation and Control of Systems with Uncertainty*. Sopron, Hungary, pp. 449–467.

Wahlberg, B. and Mäkilä, P.M. (1996). "On approximation of stable linear dynamical systems using Laguerre and Kautz functions." *Automatica* 32(5), pp. 693–708.

Weiss, A. and Di Cairano, S. (2014). "Robust dual control MPC with guaranteed constraint satisfaction." In: *IEEE Conference on Decision and Control*. Los Angeles, CA.

Wieslander, J. and Wittenmark, B. (1971). "An approach to adaptive control using real time identification." *Automatica* 7(2), pp. 211–217.

Williams, D. (1991). *Probability with Martingales*. Cambridge University Press.

Wittenmark, B. (1975). "An active suboptimal dual controller for systems with stochastic parameters." *Automatic Control Thoery & Applications* 3(1), pp. 13–19.

Ydstie, B.E. (1997). "Certainty equivalence adaptive control: what's new in the gap." In: *Chemical Process Control, CPC-V*. Ed. by J.C. Kantor, C.E. Garcia, and B. Carnaham. Tahoe City, CA: AIChE Symposium Series, pp. 9–23.

Žáčeková, E., Prívara, S., and Pčolka, M. (2013). "Persistent excitation condition within the dual control framework." *Journal of Process Control* 23(9), pp. 1270–1280.

Zagrobelny, M., Ji, L., and Rawlings, J.B. (2013). "Quis custodiet ipsos custodes?" *Annual Reviews in Control* 37(2), pp. 260–270.