Dual adaptive model predictive control ⋆

Tor Aksel N. Heirung [a] B. Erik Ydstie [b] Bjarne Foss [c]

[a] *Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720, United States,* heirung@berkeley.edu.

[b] *Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15123, United States,* ydstie@cmu.edu.

[c] *Department of Engineering Cybernetics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway,* bjarne.foss@ntnu.no.

Abstract

We present an adaptive dual model predictive controller (DMPC) that uses current and future parameter-estimate errors to minimize expected output error by optimally combining probing for uncertainty reduction with control of the nominal model. 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 enables reformulating 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 (QCQP) problem that state-of-the-art algorithms can solve efficiently, providing the exact solution to the probabilistically constrained finite-horizon dual control problem. The adaptive DMPC solves this QCQP at each sampling time on a receding horizon; the adaptation is a result of updating the parameter estimates used by the DMPC to decide the control input. The paper demonstrates the application of DMPC to a single-input single-output (SISO) system with unknown parameters. In the simulation example, the parameter estimates converge quickly and the probing vanishes with increasing accuracy and precision of the estimates, improving the future control performance.

Key words: Dual control; model predictive control; adaptive control; optimal control; stochastic control; probabilistic constraints; parameter estimation; system identification; excitation; active learning.

1 Introduction

This paper addresses the problem of optimal control and learning in the context of stochastic systems and models with stochastic parametric uncertainty and probabilistic constraints. Dual control, as introduced by Feldbaum (1961), is the optimal control under decision-relevant, reducible uncertainty. Dual control problems include the mechanisms for both control and learning in the formulation, and the solution optimally incorporates both aspects in the input to the process.

Using data to progressively reduce uncertainty is often framed as a learning process, in the control community primarily studied in the field of adaptive control. Most adaptive control algorithms are passively adaptive in the sense that learning takes place only as a side effect of the control action. These controllers learn from normal operating data, which can contain very little information. Informally, a control that with nonzero probability affects not only the system state but also the uncertainty (specifically, error covariances or higher-order central moments) has a dual effect on the system; systems in which the control cannot affect this uncertainty are called neutral (see Bar-Shalom and Tse (1974) for a rigorous definition). Note that dual effect and neutrality are properties of the system rather than the control algorithm. For systems in which the control has a dual effect, operating data can be made more informative by actively probing the process (Bar-Shalom, 1981), also known as excitation (Mareels et al., 1987), experimentation (Gevers and Ljung, 1986), exploration (Sutton and Barto, 1998), or active learning (Tse and Bar-Shalom, 1973). An actively adaptive controller is designed to improve the learning by accounting for the dual effect and increasing the amount of information generated. While active learning may fail to improve performance if the level of excitation is insufficient or excessive, the dual control is the optimal control with respect to expected system performance through endogenizing the dual effect in the problem formulation.

Adaptive model predictive control (MPC) has received relatively little attention in the literature (Mayne, 2014). As with most adaptive control approaches, adaptive MPC may suffer from signals that are insufficiently exciting for the controller or model parameters to converge to appropriate values, which may lead to problems such as burst-

⋆ This paper was not presented at any IFAC meeting. Corresponding author: T.A.N. Heirung. Tel. +1 412-805-8359.

ing (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 system by ensuring a certain level of excitation, either constantly or when needed. Shouche et al. (1998), Marafioti et al. (2014), and Larsson et al. (2015) develop algorithms that ensure a prescribed amount of information or excitation be generated, with the potential disadvantage of suboptimal performance resulting from excessive excitation of the process.

Several proposed controllers generate excitation without a specific requirement. Rather, they include a function of information or uncertainty in the MPC cost function and optimize this function together with standard control objectives. Heirung et al. (2015a) propose and compare two such formulations that converge to a standard adaptive certainty-equivalence (CE; Åström and Wittenmark, 1995) MPC formulation as the uncertainty is reduced, and show that the excitation can improve closed-loop performance. Tanaskovic et al. (2014) suggest the addition of an exploring property as a possible extension of their adaptive MPC for finite-impulse-response (FIR) systems. Their approach involves modifying the nominally optimal input sequence by solving a second-stage optimization problem, the objective of which is decreasing the set of possible models at the next time step. Common to these approaches is that the excitation is a consequence of a heuristic modification of the controller, based on the assumption that the resulting excitation will improve overall performance. While this type of algorithm may work well in practice and improve performance over passive-learning approaches (see Heirung et al., 2015a), the excitation is not an implicit consequence of optimizing for performance, which is the case for dual control in the sense of Feldbaum. The algorithm type does, however, illustrate an important distinction: superimposing excitation on a nominally optimal control signal does not generally result in *optimal* performance, and the inputs are consequently not dual.

Feldbaum (1961) identified (stochastic) dynamic programming as an appropriate solution method for dual control problems in his pioneering papers on optimal integration of active learning with multistage decision making under uncertainty. Åström and Helmersson (1986) solved a scalar dual control problem with one unknown parameter, but the "curse of dimensionality" prevents dynamic programming from being a viable solution approach for most dual control problems. This has motivated the use of modern approximate methods (Lee and Lee, 2009; Bayard and Schumitzky, 2010) that directly approximate the dynamic programming equations rather than the problem formulation.

In this article we derive an adaptive dual MPC (DMPC) for systems modeled with orthonormal basis functions with probabilistic 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. This stochastic problem is transformed into an equivalent deterministic form that enables exact evaluation of both the objective function and the probabilistic constraints. The reformulation relies on the future decisions for propagation of the exact conditional distributions over the prediction horizon, which enables determination of the learning outcome of the decision sequence. Consequently, the learning is correctly rewarded in the control algorithm, avoiding heuristic additions to the cost function (as opposed to earlier work by Heirung et al., 2015a, e.g.). We transform the reformulated problem into a quadratically-constrained quadratic-programming (QCQP) problem that can be solved efficiently using state-of-the-art solvers. The proposed DMPC ensures that the system is sufficiently excited for accurate and precise parameter estimation but does not generate a persistently exciting input. Some results in this article are generalizations of ideas by Heirung et al. (2015b), a portion of which are given there without proof. Primarily, this paper considers a more general system type and includes probabilistic output constraints.

The act of exciting, or probing, a system for learning is often seen as conflicting with the control objective (see, e.g., Tse and Bar-Shalom, 1973), and a trade-off between control and probing is frequently discussed (Åström and Kumar, 2014, e.g.). 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.

This article is organized as follows: Section 2 provides the formulation of the stochastic control problem (P) and briefly reviews some necessary statistical background. The main contributions of the paper are in Section 3, where we state and prove the results necessary to reformulate the stochastic optimal-control problem as the equivalent deterministic problem (P′) and subsequently transform this formulation into the QCQP problem (P″). Section 4 contains the dual control algorithm, followed by a simulation example in Section 5. Section 6 concludes the paper with some thoughts for future work.

Notation: $\mathrm{E}[x \mid y]$ denotes the expected value of x, given y. $\Pr[A \mid y]$ is the probability of an event A, given y.

2 Problem formulation and background

This paper considers the output tracking problem for a class of systems of the form

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

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

where $\varphi(t) \in \mathbb{R}^{n_\mathrm{p}}$ is a deterministic regression vector whose elements are functions of past control inputs (deterministic decision variables) u, and $A \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{p}}$ and $B \in \mathbb{R}^{n_\mathrm{p}}$ are known matrices determined by basis functions. The variable $y(t) \in \mathbb{R}^{n_\mathrm{p}}$ is the process output and $v(t) \in \mathbb{R}^{n_p}$ is 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, $\theta = \left[\theta_1, \theta_2, \ldots, \theta_{n_p}\right]^\top$, where $\{\theta_j\}_{j=1}^{n_p}$ are drawn from a multivariate Gaussian distribution at initial 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.

The system (1) is a linear, time-invariant, single-input, single-output (SISO) system, and we assume that the pair (A, B) is controllable and stab; this formulation includes systems modeled by orthogonal basis functions (OBFs). The most well-known member of this model class is the FIR model; other common formulations include the Laguerre (Wahlberg, 1991) and Keutz (Wahlberg, 1994) models; see also Finn et al. (1993) for a combination of the FIR and Laguerre structures. Heuberger et al. (2005) provide a comprehensive overview of modeling and identification with OBFs and demonstrate several of their advantages, including highly accurate models represented by few parameters where high-order FIR models are necessary for acceptable approximations. Linearly parameterized systems of this form continue to find applications related to modeling, identification, and control; one recent example is data-driven and model-based formal verification (Haesaert et al., 2017). This system class is also considered by Sbarbaro et al. (1999), who extend the results of Kulcsár et al. (1996) for FIR systems and derive an unconstrained actively adaptive controller using one-dimensional line search and propagation of future parameter-error covariances. Their approach assumes either that the future controls have one constant value over the control horizon or that they are determined by a certainty-equivalence strategy, which in contrast to the algorithm developed here ignores the learning effect of the future controls.

A standard definition of information (Ljung, 1999) recorded up to and including time t is the set of all past decisions and measurements:

$$\mathcal{Y}(t)=\{u(t), u(t-1),\ldots,u(t_0), y(t), y(t-1),\ldots,y(t_0)\} \quad (2)$$

This definition of information can be expanded to include the initial distribution of θ given by $\hat{\theta}(t_0)$ and $P(t_0)$. The parameter estimate is $\hat{\theta}(t) := \mathrm{E}\left[\theta \,|\, \mathcal{Y}(t)\right] = \left[\hat{\theta}_1(t), \ldots, \hat{\theta}_{n_p}(t)\right]^\top$ and the parameter-estimate-error covariance matrix is $P(t) := \mathrm{E}\left[\tilde{\theta}(t)\tilde{\theta}^\top(t) \,|\, \mathcal{Y}(t)\right]$ with $\tilde{\theta}(t) := \theta - \hat{\theta}(t)$. Let $\{u(k \,|\, t)\}_{k=t}^{t+N-1}$ be a sequence of future control inputs decided at time t ($k \geq t$ unless otherwise noted). The output predictor, which provides the nominal model output, is

$$\begin{aligned} \hat{y}(k+1 \,|\, t) &= \mathrm{E}\left[y(k+1) \;\middle|\; \mathcal{Y}(t)\right], \qquad k \geq t \\ &= \hat{\theta}^\top(t)\varphi(k+1 \,|\, t) \end{aligned} \quad (3)$$

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

$$\varphi(k+1 \,|\, t) = A\varphi(k \,|\, t) + Bu(k \,|\, t), \qquad k \geq t \quad (4)$$

with $\varphi(t \,|\, t) := \varphi(t)$ and $u(t \,|\, t) := u(t)$.

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

$$\mathcal{Y}(k \,|\, t) = \{\underbrace{u(k \,|\, t), u(k-1 \,|\, t), \ldots, u(t+1 \,|\, t), u(t \,|\, t)}_{\text{anticipated information, } k \geq t}, \underbrace{u(t-1), u(t-2), \ldots, u(t_0), y(t), y(t-1), \ldots, y(t_0)}_{\text{past information}}\} \quad (5)$$

Note that the future input sequence $\{u(i \,|\, t)\}_{i=t}^{i=k}$ consists of exogenous decisions and is deterministic at time t, as opposed to uncertain signals like future system outputs, which can be predicted based on $\mathcal{Y}(t)$ (see Equation (3)). Hence, $\mathcal{Y}(k \,|\, t)$ contains no information from the system beyond time t, and the future regressors $\varphi(k \,|\, t)$ in Equation (4) contain deterministic decisions only.

2.1 Optimal control problem

The finite-horizon performance cost considered in this paper is

$$\begin{aligned} J_N(t) = \sum_{k=t}^{t+N-1} \Big\{ &\mathrm{E}\left[(y(k+1) - y^*(k+1 \,|\, t))^2 \;\middle|\; \mathcal{Y}(k \,|\, t)\right] \\ &+ w_2 u^2(k \,|\, t) + w_3(\Delta u(k \,|\, t))^2 \Big\} \end{aligned} \quad (6)$$

where $N \geq 1$ is the length of the prediction horizon, $y^*(k+1 \,|\, t)$ is the output reference sequence at time t, $\Delta u(k \,|\, t) := u(k \,|\, t) - u(k-1 \,|\, t)$ is the control-input rate of change, with $u(t-1 \,|\, t) := u(t-1)$, and $w_2 \geq 0$, $w_3 \geq 0$ are tuning weights. Taking the expectation of the output error with respect to the current information $\mathcal{Y}(t)$, as opposed to the future information $\mathcal{Y}(k \,|\, t)$, results in an output cost that does not reward excitation since future parameter covariances $P(k \,|\, t)$, $k \geq t$, do not appear (see Section 3.1). The current covariance $P(t)$ does appear, and provides a rationale for caution since large uncertainties heavily penalize current and future inputs. In contrast, the objective function we develop in this paper takes into account how future decisions affect uncertainty and as result captures the dual nature of the optimal control. Note that we use only finite N in this paper, and that J_N may be unbounded for an infinite horizon length.

This paper considers minimization of $J_N(t)$ in (6) subject to probabilistic output constraints and deterministic bounds on the rate of change and magnitude of the inputs. The predicted outputs and the decision variables are related through the model (3)–(4). The resulting stochastic optimal-control problem is

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

subject to

$$\varphi(k+1 \,|\, t) = A\varphi(k \,|\, t) + Bu(k \,|\, t) \quad \text{(P.b)}$$

$$\hat{y}(k+1 \,|\, t) = \hat{\theta}^\top(t)\varphi(k+1 \,|\, t) \quad \text{(P.c)}$$

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

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

$$u_{\min} \leq u(k \,|\, t) \leq u_{\max} \quad \text{(P.f)}$$

$$\Delta u(k \,|\, t) = u(k \,|\, t) - u(k-1 \,|\, t) \quad \text{(P.g)}$$

$$\Delta u_{\min} \leq \Delta u(k \,|\, t) \leq \Delta u_{\max} \quad \text{(P.h)}$$

$$k \in \{t,\, t+1,\, \ldots,\, t+N-1\} \tag{P.i}$$
$$\varphi(t\,|\,t),\ u(t-1\,|\,t),\ \hat{\theta}(t),\ P(t) \text{ given}$$

where $\Pr[y_{\min} \leq y(k+1)\,|\,\mathcal{Y}(k\,|\,t)]$ and $\Pr[y(k+1) \leq y_{\max}\,|\,\mathcal{Y}(k\,|\,t)]$ are the probabilities, conditioned on $\mathcal{Y}(k\,|\,t)$, that the system outputs stay above $y_{\min}$ and below $y_{\max}$, respectively, and $p_{y,\min}, p_{y,\max} \in (0,1)$ are the minimum probabilities of chance-constraints satisfaction in (P.d)–(P.e).

The stochastic optimal-control problem (P) is solved with initial values $\varphi(t\,|\,t) = \varphi(t), u(t-1), \hat{\theta}(t)$, and $P(t)$; the solution includes an optimal sequence of predicted control inputs $\{u^{\mathrm{o}}(k\,|\,t)\}_{k=t}^{t+N-1}$.

We now describe the evolution and prediction of the parameter-estimate statistics performed at every time t, and then discuss 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 finite N.

2.2 Parameter estimation and statistics

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

$$R(t) = \sum_{k=t_0+1}^{t} r_R^{-1}\lambda^{t-k}\varphi(k)\varphi^{\top}(k) + r_R^{-1}\lambda^{t-t_0}R(t_0) \tag{7}$$

with $t > t_0$, 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 recursively expressed as

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

The conditional mean of θ, given $\mathcal{Y}(t)$, is then

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

The inverse of the information matrix $R(t)$ is the covariance matrix $P(t)$. With the given assumptions, the conditional distribution of θ given $\mathcal{Y}(t)$ is Gaussian with mean $\hat{\theta}(t)$ and covariance $P(t)$; see Theorem 6 in Appendix A, from which the equation set (A.1) with $r = r_R$ can be used to calculate the conditional distribution without inverting $R(t)$ in (8b).

3 Reformulation to a deterministic *QCQP*

The process (1) belongs to a class of systems in which the output is dependent on past inputs but not past outputs. Thus, as noted in Section 2, the predicted future regressors are deterministic since they do not contain future outputs, which are stochastic variables. That is, the future inputs are decision variables, meaning the decision maker is free to decide the future components of $\mathcal{Y}(k\,|\,t)$ for any $k \geq t$. This allows the following theorem (cf. Theorem 6 in Appendix A) for propagation of the future conditional parameter-estimate covariance.

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

$$K(k+1|t) = P(k|t)\varphi(k+1|t)\times \big(r + \varphi^{\top}(k+1|t)P(k|t)\varphi(k+1|t)\big)^{-1} \tag{9a}$$
$$P(k+1|t) = \big(I - K(k+1|t)\varphi^{\top}(k+1|t)\big)P(k|t) \tag{9b}$$

given $\mathcal{Y}(k\,|\,t)$ *and initial covariance* $P(t\,|\,t) := P(t)$.

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 fact that $\mathcal{Y}(k\,|\,t)$ is deterministic and determines $\varphi(k\,|\,t)$, which means that $K(k\,|\,t)$ and $P(k\,|\,t)$ are deterministic for $k \geq t$ given the observed outputs and the past and deterministic future decisions in $\mathcal{Y}(k\,|\,t)$. □

The following theorem, here used to reformulate the optimal-control problem (P), is a consequence of the fact that the deterministic future inputs determine the future covariances $P(k\,|\,t)$.

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

$$\mathrm{E}\big[y^2(k+1)\,|\,\mathcal{Y}(k\,|\,t)\big] = \hat{y}^2(k+1\,|\,t) + \varphi^{\top}(k+1\,|\,t)P(k\,|\,t)\varphi(k+1\,|\,t) + r \tag{10}$$

for all $k \geq t$*.*

PROOF. From the model (1b), the left-hand side of (10) is $\mathrm{E}\big[\big(\theta^{\top}\varphi(k+1) + v(k+1)\big)^2 \;\big|\; \mathcal{Y}(k\,|\,t)\big]$. After expanding the square, using the fact that $\varphi(k+1)$ is deterministic, that $\mathrm{E}\big[v^2(k+1) \;\big|\; \mathcal{Y}(k\,|\,t)\big] = r$, that $v(k+1)$ and θ are uncorrelated, and Theorem 8 in Appendix A, this reduces to

$$\mathrm{E}\big[y^2(k+1) \;\big|\; \mathcal{Y}(k\,|\,t)\big] = \varphi^{\top}(k+1\,|\,t)\big(\hat{\theta}(t)\hat{\theta}^{\top}(t) + P(k\,|\,t)\big)\varphi(k+1\,|\,t) + r \tag{11}$$

Equation (10) follows after expanding the parenthesis and substituting the model (3) for $\hat{\theta}^{\top}(t)\varphi(k+1\,|\,t)$. □

The following corollary extends Theorem 2 to tracking of a time-varying output reference $y^*(k+1\,|\,t)$.

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

$$\begin{aligned}\mathrm{E}[(y(k+1) - y^*(k+1\,|\,t))^2\,|\,\mathcal{Y}(k\,|\,t)] &= (\hat{y}(k+1\,|\,t) - y^*(k+1\,|\,t))^2 \\ &\quad + \varphi^{\top}(k+1\,|\,t)P(k\,|\,t)\varphi(k+1\,|\,t) + r\end{aligned} \tag{12}$$

for all $k \geq t$*.*

PROOF. By expanding the square and using Theorem 2, the left-hand side of (12) becomes $\hat{y}^2(k+1\,|\,t) + \varphi^{\top}(k+1\,|\,t)P(k\,|\,t)\varphi(k+1\,|\,t) + r - 2\hat{y}(k+1\,|\,t)y^*(k+1\,|\,t) + (y^*(k+1\,|\,t))^2$ which is easily rearranged to the right-hand side of (12). □

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

Corollary 4 *For a stochastic process of the form* (1), *the future output variance predicted at time t is*

$$\sigma_y^2(k+1\,|\,t) = \varphi^\top(k+1\,|\,t)P(k\,|\,t)\varphi(k+1\,|\,t) + r \quad (13)$$

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

PROOF. By definition,

$$\begin{aligned}\sigma_y^2(k{+}1\,|\,t) &= \mathrm{E}\big[(y(k{+}1) - \hat{y}(k{+}1\,|\,t))^2 \;\big|\; \mathcal{Y}(k\,|\,t)\big] \\ &= \mathrm{E}\big[y^2(k{+}1) \;\big|\; \mathcal{Y}(k\,|\,t)\big] - \hat{y}^2(k{+}1\,|\,t) \quad (14)\end{aligned}$$

which follows from expanding the square and collecting the terms. The expression for $\sigma_y^2(k{+}1\,|\,t)$ is then obtained directly from Theorem 2. □

3.1 Evaluation of the objective function

Corollary 3 allows the stochastic objective (6) to be reformulated into the equivalent deterministic function in the following theorem.

Theorem 5 *For a stochastic process of the form* (1), *the objective function* $J_N(t)$ *in Equation* (6) *can be written*

$$\begin{aligned}J_N(t) = \sum_{k=t}^{t+N-1} \big\{ &(\hat{y}(k+1\,|\,t) - y^*(k+1\,|\,t))^2 \\ &+ \varphi^\top(k+1\,|\,t)P(k\,|\,t)\varphi(k+1\,|\,t) + r \\ &+ w_2 u^2(k\,|\,t) + w_3(\Delta u(k\,|\,t))^2 \big\} \quad (15)\end{aligned}$$

PROOF. This result follows directly from Corollary 3. □

Note that this objective function rewards probing provided $N \geq 2$. $P(t\,|\,t) = P(t)$ is not an optimization variable; it is the current covariance determined prior to evaluating the objective function. On the other hand, $P(t{+}1\,|\,t)$ is a variable, is included in the objective when $N \geq 2$, and represents the covariance one step ahead. The objective function thus rewards uncertainty reduction through the selection of control inputs that reduce future covariances. Furthermore, if the uncertainty represented by $P(k\,|\,t)$, $k \geq t$, goes to zero, the formulation of the output cost $J_N(t)$ in (15) converges to a certainty-equivalence-type output cost (except for the constant term r). This implies that the excitation reward induced by the parameter uncertainty vanishes as the uncertainty is resolved.

3.2 Evaluation of the chance constraints

We now transform the probabilistic constraints (P.d)–(P.e) into deterministic form using Corollary 4. From Åström and Wittenmark (1995) and implied from Corollary 4, the conditional distribution of $y(t+1)$ given $\mathcal{Y}(t)$ is Gaussian with mean $\hat{y}(t+1) = \hat{\theta}^\top(t)\varphi(t+1)$ and variance $\sigma_y^2(t+1) = \varphi^\top(t+1)P(t)\varphi(t+1) + r$. Similarly, the future conditional distribution of $y(k+1)$ given $\mathcal{Y}(k\,|\,t)$ for $k \geq t$ is Gaussian with mean $\hat{y}(k+1\,|\,t)$ and variance $\sigma_y^2(k{+}1\,|\,t)$, given in equations (3) and (13), respectively. Hence, the probabilistic output constraints (P.d)–(P.e) have their respective deterministic equivalents

$$y_{\min} \leq \hat{y}(k+1\,|\,t) - s_{\min}\sigma_y(k+1\,|\,t) \quad (16a)$$

$$\hat{y}(k+1\,|\,t) + s_{\max}\sigma_y(k+1\,|\,t) \leq y_{\max} \quad (16b)$$

where $\Phi(s_{\min}) = p_{y,\min}$ and $\Phi(s_{\max}) = p_{y,\max}$ with Φ the cumulative distribution function (CDF) for the standard normal distribution. The parameters $s_{\min}$ and $s_{\max}$ are determined once and offline. Extending this formulation to time-varying probabilities $p_{y,\min}(k\,|\,t)$ and $p_{y,\max}(k\,|\,t)$ is straight-forward. The only change required is that a larger number of equations $\Phi(s_{\min}(k\,|\,t)) = p_{y,\min}(k\,|\,t)$ and $\Phi(s_{\max}(k\,|\,t)) = p_{y,\max}(k\,|\,t)$ be solved offline.

3.3 The deterministic optimal-control problem

The objective function $J_N(t)$ in Equation (15) can now be minimized by augmenting the constraint set of (P) with (9) from Theorem 1 and (13) from Corollary 4, and replacing the probabilistic constraints (P.d)–(P.e) with their deterministic equivalents (16a)–(16b). The result is a deterministic optimal-control problem that is equivalent to (P):

$$\min J_N(t) \quad (\mathrm{P}'.\mathrm{a})$$

subject to

$$\varphi(k+1\,|\,t) = A\varphi(k\,|\,t) + Bu(k\,|\,t) \quad (\mathrm{P}'.\mathrm{b})$$

$$\hat{y}(k+1\,|\,t) = \hat{\theta}^\top(t)\varphi(k+1\,|\,t) \quad (\mathrm{P}'.\mathrm{c})$$

$$\sigma_y^2(k+1\,|\,t) = \varphi^\top(k{+}1\,|\,t)P(k\,|\,t)\varphi(k{+}1\,|\,t) + r \quad (\mathrm{P}'.\mathrm{d})$$

$$\begin{aligned}K(k+1\,|\,t) &= P(k\,|\,t)\varphi(k+1\,|\,t) \\ &\times \big(r_R + \varphi^\top(k+1\,|\,t)P(k\,|\,t)\varphi(k+1\,|\,t)\big)^{-1} \quad (\mathrm{P}'.\mathrm{e})\end{aligned}$$

$$P(k+1\,|\,t) = \big(I - K(k{+}1\,|\,t)\varphi^\top(k{+}1\,|\,t)\big)P(k\,|\,t) \quad (\mathrm{P}'.\mathrm{f})$$

$$\hat{y}(k+1\,|\,t) \geq y_{\min} + s_{\min}\sigma_y(k+1\,|\,t) \quad (\mathrm{P}'.\mathrm{g})$$

$$\hat{y}(k+1\,|\,t) \leq y_{\max} - s_{\max}\sigma_y(k+1\,|\,t) \quad (\mathrm{P}'.\mathrm{h})$$

$$u_{\min} \leq u(k\,|\,t) \leq u_{\max} \quad (\mathrm{P}'.\mathrm{i})$$

$$\Delta u(k\,|\,t) = u(k\,|\,t) - u(k-1\,|\,t) \quad (\mathrm{P}'.\mathrm{j})$$

$$\Delta u_{\min} \leq \Delta u(k\,|\,t) \leq \Delta u_{\max} \quad (\mathrm{P}'.\mathrm{k})$$

$$k \in \{t,\, t+1,\, \ldots,\, t+N-1\} \quad (\mathrm{P}'.\mathrm{l})$$

$$\varphi(t\,|\,t),\; u(t-1\,|\,t),\; \hat{\theta}(t),\; P(t\,|\,t),\; s_{\min},\; s_{\max} \text{ given}$$

Here $u(t-1\,|\,t) = u(t-1)$ and $P(t\,|\,t) = P(t)$. 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.

Although the solution to (P′) exactly minimizes $J_N(t)$ in (6) over the finite horizon, the feasible area is nonconvex because of the inclusion of the nonlinear equality constraints (P′.d)–(P′.f). This motivates investigation of reformulation approaches that facilitate solving the optimal-control problem. We consider this reformulation a main contribution of the paper.

3.4 Reformulation as a QCQP

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

$$\bar{R}(k+1\,|\,t) := r_R R(k+1\,|\,t) \quad (17)$$

which is recursively expressed as (cf. Equation (8))

$$\bar{R}(k+1\,|\,t) = \bar{R}(k\,|\,t) + \varphi(k+1\,|\,t)\varphi^\top(k+1\,|\,t) \quad (18)$$

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

$$\bar{R}(k\,|\,t)z(k\,|\,t) = \varphi(k+1\,|\,t) \tag{19}$$

or equivalently $r_R z(k\,|\,t) = P(k\,|\,t)\varphi(k+1\,|\,t)$, we write

$$\varphi^\top(k+1\,|\,t)P(k\,|\,t)\varphi(k+1\,|\,t) = r_R \varphi^\top(k+1\,|\,t)z(k\,|\,t) \tag{20}$$

This equation allows simplification of both the objective function (15) from Theorem 5 and the predicted output-variance constraint (P′.d). Furthermore, the nonlinear uncertainty-propagation constraints (P′.e)–(P′.f) can be replaced with the quadratic equations (18) and (19).

The objective function (15) can now be written

$$J_N(t) = \sum_{k=t}^{t+N-1} \big\{ (\hat{y}(k+1\,|\,t) - y^*(k+1\,|\,t))^2 + \sigma_y^2(k+1\,|\,t) + w_2 u^2(k\,|\,t) + w_3(\Delta u(k\,|\,t))^2 \big\} \tag{21}$$

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

$$\min J_N(t) \tag{P″.a}$$

subject to

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

$\varphi(t\,|\,t),\ u(t-1\,|\,t),\ \hat{\theta}(t),\ \bar{R}(t\,|\,t),\ s_{\min},\ s_{\max}$ given

where $\bar{R}(t\,|\,t) = r_R P^{-1}(t)$.

The parameters r_R and r both appear in the objective function (21), 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 reduction of uncertainty, represented by the term $\varphi^\top(k+1\,|\,t)z(k\,|\,t)$, is weighted against reducing the nominal output tracking error.

The formulation (P″) with (21) is nonlinear, but all nonlinearities are quadratic (i.e., either bilinear or square). Specifically, the objective function (21) contains square terms only ($\hat{y}^2(\cdot)$, $\sigma_y^2(\cdot)$, $u^2(\cdot)$, and $(\Delta u(\cdot))^2$) while the constraints (P″.d), (P″.e), and (P″.f) contain both square ($\sigma_y^2(\cdot)$ and the diagonal elements of $\varphi(\cdot)\varphi^\top(\cdot)$) and bilinear ($\varphi^\top(\cdot)z(\cdot)$, the off-diagonal elements of $\varphi(\cdot)\varphi^\top(\cdot)$, and $\bar{R}(\cdot)z(\cdot)$) terms. The optimal-control problem (P″) is therefore a standard QCQP problem (see, e.g., Misener and Floudas, 2013).

Although the quadratic equality constraints render (P″) nonconvex, there are several algorithms that efficiently solve QCQP 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 complexity of the optimal-control problem (P″) increases moderately with the number of unknown model parameters n_p and the length of the prediction horizon N. The output-variance constraint (P″.d) contains N square terms and $n_\mathrm{p}N$ bilinear terms, while there are $n_\mathrm{p}^2 N$ quadratic terms in each of the uncertainty-propagation constraints (P″.e) and (P″.f). The symmetric nature of the quadratic equality constraints can be exploited to reduce the number of added terms, but the quadratic growth cannot be avoided.

4 Dual control algorithm

We now propose a dual control algorithm based on MPC (Mayne et al., 2000). Like in standard MPC, the control input is determined by solving an optimal-control problem at each sampling instant in a receding-horizon fashion. With our DMPC algorithm, the dual control input at time t, $u(t) = u^\mathrm{o}(t\,|\,t)$, is contained in the solution $\{u^\mathrm{o}(k\,|\,t)\}_{k=t}^{t+N-1}$ to the finite-horizon stochastic optimal-control problem (P); the solution is obtained by solving the equivalent QCQP problem (P″). In contrast to standard MPC, where there is feedback from the system state (estimate), our adaptive DMPC depends on feedback from the hyperstate $(\varphi(t),\ u(t-1),\ \hat{\theta}(t),\ P(t))$ (Åström and Wittenmark, 1995). Furthermore, (P″) is not a true open-loop problem since the uncertainty predictions implicitly anticipate a partially closed loop. 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 DMPC a CE controller, since the control action also depends on the current and future estimate covariances, 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 system 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 (A.1)).
(4) Solve (P″) to obtain the solution $\{u^\mathrm{o}(k\,|\,t)\}_{k=t}^{t+N-1}$.
(5) Implement $u(t) = u^\mathrm{o}(t\,|\,t)$.
(6) Set $t \leftarrow t+1$ and go to step 2.

5 Example

We now demonstrate our DMPC algorithm on a small simulation test case that highlights some of its main features. The example system is of FIR type (in the formulation (1), A is a matrix with all elements set to one on the first subdiagonal and all other elements zero, while B is a vector with the first element set to one and the other elements zero) with $n_\mathrm{p} = 4$ parameters. In order

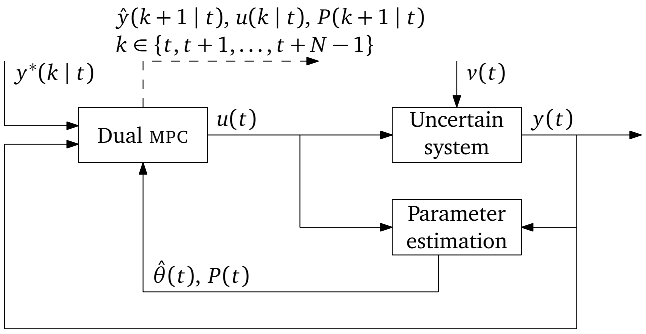

Fig. 1. Block diagram illustrating the adaptive DMPC structure. The dashed line represents variables predicted by the DMPC.

to isolate the capabilities of the DMPC algorithm and better demonstrate its features, no disturbance is included in the simulations, meaning $r = 0$ and $r_R = 1$. The unknown system parameters are first $\theta = \theta^1 = [2.5,\ 1.8,\ 1.6,\ 0.9]^\top$ for $t \leq -3$ and then change so that $\theta = \theta^2 = [4.2,\ 2.2,\ -0.2,\ 0.6]^\top$ for $t \geq -2$. Given perfect information on θ, the optimal steady-state input for the example system is $u_{\text{ss}}(t;\theta) = y^*(t) \;/\; \left(\sum_{j=1}^{n_\text{p}} \theta_j\right)$, with $\sum_{j=1}^{n_\text{p}} \theta_j \neq 0$. The initial hyperstate at time $t = t_\text{i} = -5$ is given by $u(t) = u_{\text{ss}}(t;\theta^1)$ for all $t \leq t_\text{i}$ (meaning all elements in the initial regressor are $u_{\text{ss}}(t;\theta^1)$), $\hat{\theta}(t_\text{i}) = \theta^1$, and $P(t_\text{i}) = 0_{n_\text{p}\times n_\text{p}}$, with $y(t_\text{i}) = 5.0$. The system constraints are specified as $y_{\min} - 3.0$, $y_{\max} = 6.0$, $p_{y,\min} = p_{y,\max} = 0.5$ (which corresponds to $s_{\min} = s_{\max} = 0$), $-u_{\min} = u_{\max} = 1.0$, and $-\Delta u_{\min} = \Delta u_{\max} = \infty$. The DMPC and estimation parameters are $N = 8$, $w_2 = 0$, $w_3 = 10^{-3}$, and $\lambda = 1$.

We consider a scenario with the system having been in steady state for some time to best demonstrate the qualitative behavior of the DMPC and to show the effect of reinitializing the controller with different sets of values in the covariance matrix $P(t_0)$. The simulations start at $t = t_\text{i}$ and end at $t = t_\text{f} = 20$; they are are identical up to time $t = t_0 = 0$, with the history from $t = t_\text{i}$ shown for clearer illustration of the decisions made by the DMPCs. The initial parameter estimates are exact ($\hat{\theta}(t_\text{i}) = \theta^1$) with a point distribution ($P(t) = 0$ for $t < t_0$). This means the DMPC reduces to a CE MPC, which gives optimal control performance in this situation (correct parameter estimates with no uncertainty). A shift in the system parameters from θ^1 to θ^2 occurs at $t = -2$, after which the DMPC and the CE MPC are both able to keep the output at the reference without changing the input. Consequently, the parameter estimates do not change. The DMPC is reinitialized at $t = t_0$ by setting $P(t_0)$ to a positive-definite matrix. At time $t = 12$ the output reference setpoint changes from the initial value $y^*(t) = 5.0$ to $y^*(t) = -2$.

The simulations compare DMPC to adaptive CE MPC and also show how DMPC performance is affected by different initial error-covariance magnitudes $P(t_0)$. For illustration purposes we show one case with very large variances in a DMPC at time t_0 to trigger a strong probing action that results from uncertainty (Figure 2). This behavior is contrasted with that resulting from more moderate variance values of $P(t_0)$ (Figure 3).

Figures 2 and 3 together show the consequence of reinitializing the DMPC at time t_0 by changing P from a zero-matrix to a positive-definite matrix, and how the resulting probing action reduces parameter-estimate error enough for the controller to perform well at a subsequent change in output reference.

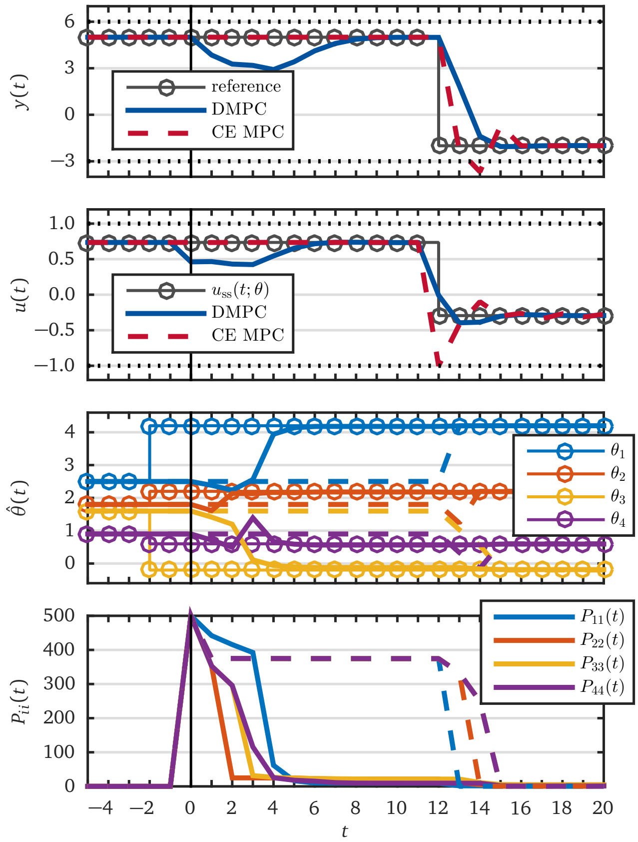

Fig. 2. A comparison of DMPC (solid, heavy lines for all variables) and CE MPC (dashed, heavy lines for all variables). From top to bottom, the four plots show the outputs, including the reference $y^*(t)$ and the bounds $y_{\min}$ and $y_{\max}$ (dotted lines); the inputs, including the optimal steady-state input $u_{\text{ss}}(t;\theta)$ and the bounds $u_{\min}$ and $u_{\max}$ (dotted lines); the true parameters θ (solid lines with circles) and both sets of estimates $\hat{\theta}(t)$; and the diagonal elements of the covariance matrix $P(t)$, which are the variance terms.

Figure 2 shows how a DMPC with $P(t_0) = 500I$ diverges from an identically-tuned CE MPC, provided the same operational history at $t = t_0$. The CE MPC (dashed red line) continues applying a control identical to the optimal steady-state input $u_{\text{ss}}(t;\theta)$ after the change in parameter values, keeping the output at the reference without changing the control signal. 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 system. Hence, the parameter estimates (dashed lines) 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

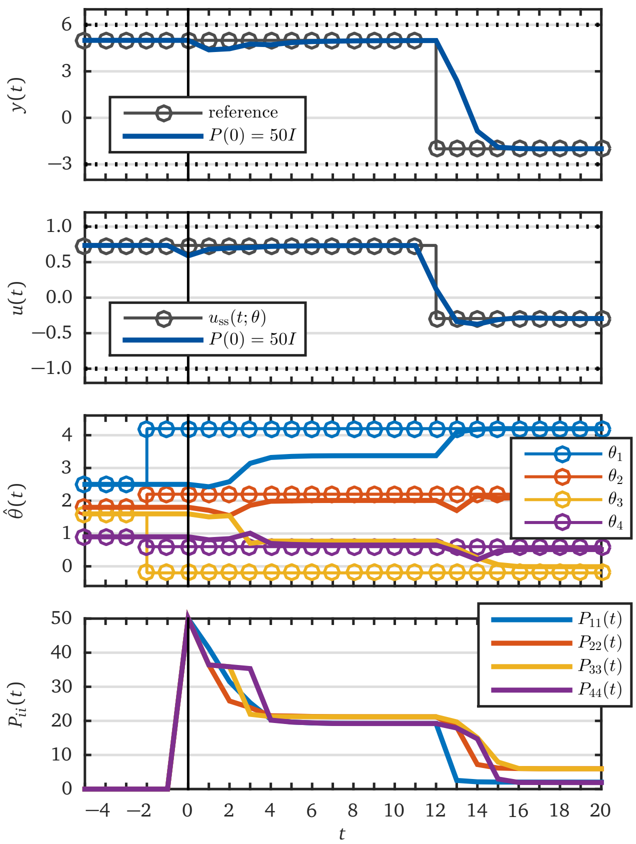

Fig. 3. A DMPC setup reinitialized with a smaller covariance at $t = t_0$: $P(t_0) = 50I$ (solid, heavy lines) as opposed to $500I$ (Figure 2). See the caption of Figure 2 for further description of the plots.

bound and significant oscillations before the output settles at the new reference value. Note that none of the controllers have prior knowledge of the reference change.

The consequence of reinitializing the DMPC (solid blue line) with $P(t_0) = 500I$ is clear when contrasted with the CE MPC. The output from the DMPC is reduced in magnitude at $t = t_0$ in order to reduce the future covariances, in turn reducing the cost incurred from the term $\varphi^\top(k \mid t)P(k \mid t)\varphi(k \mid t)$ over the prediction horizon in the objective (15). This probing action causes a reduction in output magnitude and thus generates data, which leads to a change in the parameter estimates (solid lines) and a reduction of the variance (the diagonal elements of $P(t)$ are shown in bottom plot). The 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 at $t = t_0$) so that the expected reward from further uncertainty reduction is not higher than the cost of increased reference-tracking error for the nominal output. At this time, the parameter estimates are very close to their true values and the input from the DMPC is converging back toward the optimal steady-state value. Consequently, the output is converging back to the reference setpoint. When the subsequent change in output reference occurs, the DMPC has learned enough about the new system-parameter values to successfully move the output toward the new reference, without the constraint violation and oscillations in the output seen with the CE MPC.

Figure 3 shows how the DMPC performs when reinitialized with smaller variances ($P(t_0) = 50I$) and all other factors kept identical. Compared with the above case of $P(t_0) = 500I$, the DMPC with less model uncertainty, represented by smaller variances, exhibits the same qualitative behavior, but the probing is lower in magnitude and vanishes sooner. As a consequence there is less information in the generated data, which leads to parameter estimates that are further from their true values than those obtained when reinitializing with a larger $P(t_0)$. Despite the moderate excitation, the DMPC is able to direct the output to the new reference with no constraint violations or oscillations.

The example simulations are implemented in MATLAB and the QCQP problems are solved using the local NLP solver SNOPT 7.2 (Gill et al., 2005) under GAMS (GAMS Development Corporation), which uses automatic differentiation to provide gradients to the solver. A standard laptop computer runs the simulations and the QCQP 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. Using BARON for verification, all local solutions found by SNOPT are within a 1 % relative global-optimality gap.

6 Conclusions and future work

Our reformulation (P′) of the probabilistically-constrained stochastic optimal-control problem (P) provides insight into how specific functions of excitation and nominal output error together result in dual control when their sum is minimal, as well as a foundation for practical control-algorithm design. The reformulated objective can furthermore guide the design of approximate or suboptimal dual controllers for systems where deterministic expressions for the stochastic objective function cannot be derived. Further reformulation of the control problem leads to the QCQP problem (P″) that can be solved efficiently. Solving this QCQP on a receding horizon using the future and current information results in the DMPC algorithm. The reformulation allows for easy incorporation of exact probabilistic constraints with a moderate increase in problem complexity. Note that unlike deterministic MPC on an infinite horizon for systems with no uncertainty, the proposed DMPC does not recover the dynamic programming solution since the effect of the future outputs on the future parameter estimates are not predicted by the controller.

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 developed here to time-varying parameters modeled as the Gauss-Markov process $\theta(t + 1) = \Theta\theta(t) + w(t)$ where Θ is a known, constant matrix and $w(t)$ is a sequence of independent and identically distributed Gaussian random vectors with zero mean and known variance.

While the number of quadratic terms in (P″) grows quadratically in the number of model parameters n_p, one of the advantages of using OBF models other than FIR is that the required number of parameters is typically much lower without sacrificing prediction accuracy (Heuberger et al., 2005). One possible approach for extending the control algorithm to multivariable systems is to use ideas similar to those presented by Kumar et al. (2015), where an approximate dual MPC for scalar systems is extended to the multivariable case.

While the expected value of the system output is bounded when the control inputs are bounded, stability analysis of the algorithm is complicated by the infinite support of the stochastic disturbance and the boundedness of the controls, and is the subject of a future paper. Another topic of future work is derivation of tight variable bounds and careful exploitation of problem structure to efficiently solve (P″) to global optimality.

A Parameter statistics and least-squares estimation

For the systems we consider here the covariance of θ at time t given $\mathcal{Y}(t)$, $P(t)$, becomes as a special case of $P(k\,|\,t)$ in Equation (9b). 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 6 *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*

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

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

$$P(t) = \big(I - K(t)\varphi^\top(t)\big)P(t-1)(1/\lambda) \quad \text{(A.1c)}$$

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 (A.1) 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).

The expected value of the unknown parameter vector θ given current information and future decisions, $\mathcal{Y}(k\,|\,t)$, is the same as given only current information. The following lemma states this formally.

Lemma 7 *For a stochastic process of the form* (1)*,*

$$\mathrm{E}\big[\theta \;\big|\; \mathcal{Y}(k\,|\,t)\big] = \mathrm{E}\big[\theta \;\big|\; \mathcal{Y}(t)\big] = \hat{\theta}(t), \qquad k \geq t \quad \text{(A.2)}$$

PROOF. From (A.1a) it is apparent that the conditional mean of θ at time t depends on $y(t)$. Thus, $\mathcal{Y}(k\,|\,t)$ does not contain any information relevant for $\hat{\theta}(k)$, $k \geq t$, beyond $\mathcal{Y}(t)$, so the conditional mean of θ given $\mathcal{Y}(k\,|\,t)$ is simply $\mathrm{E}\big[\theta \;\big|\; \mathcal{Y}(t)\big] = \hat{\theta}(t)$. □

A consequence of Lemma 7 is that

$$\hat{\theta}(k\,|\,t) := \mathrm{E}\big[\theta \;\big|\; \mathcal{Y}(k\,|\,t)\big] = \hat{\theta}(t), \qquad k \geq t \quad \text{(A.3)}$$

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

Theorem 8 *For a stochastic process of the form* (1)*,*

$$\mathrm{E}\big[\theta\theta^\top \;\big|\; \mathcal{Y}(k\,|\,t)\big] = \hat{\theta}(t)\hat{\theta}^\top(t) + P(k\,|\,t), \quad k \geq t \quad \text{(A.4)}$$

PROOF. Using Lemma 7, Equation (A.3), and the fact that $\hat{\theta}(t)\theta^\top$ is symmetric, we start from the definition of the anticipated covariance matrix $P(k\,|\,t)$ and get

$$\begin{aligned} P(k\,|\,t) :=& \,\mathrm{E}\big[\big(\theta - \mathrm{E}[\theta\,|\,\mathcal{Y}(k\,|\,t)]\big)\times \\ &\qquad \big(\theta - \mathrm{E}[\theta\,|\,\mathcal{Y}(k\,|\,t)]\big)^\top \;\big|\; \mathcal{Y}(k\,|\,t)\big] \\ =& \,\mathrm{E}\big[\theta\theta^\top\,|\,\mathcal{Y}(k\,|\,t)\big] - \hat{\theta}(t)\hat{\theta}^\top(t) \end{aligned} \quad \text{(A.5)}$$

through simple algebra. Rearranging the equation completes the proof. □

Acknowledgements

We gratefully acknowledge the funding provided by the Center of Integrated Operations in the Petroleum Industry at NTNU, Trondheim, Norway (www.iocenter.no). We are also grateful to the five referees whose insightful comments and suggestions improved the paper.

References

Anderson, B.D.O. Adaptive systems, lack of persistency of excitation and bursting phenomena. *Automatica*, 21(3):247–258, 1985.

Åström, K.J. and Helmersson, A. Dual control of an integrator with unknown gain. *Computers and Mathematics with Applications*, 12(6):653–662, 1986.

Åström, K.J. and Kumar, P.R. Control: A perspective. *Automatica*, 50(1):3–43, 2014.

Åström, K.J. and Wittenmark, B. *Adaptive Control*. Dover, second edition, 1995.

Bar-Shalom, Y. Stochastic dynamic programming: Caution and probing. *IEEE Transactions on Automatic Control*, 26(5):1184–1195, 1981.

Bar-Shalom, Y. and Tse, E. Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Transactions on Automatic Control*, 19(5):494–500, 1974.

Bayard, D.S. and Schumitzky, A. Implicit dual control based on particle filtering and forward dynamic programming. *International Journal of Adaptive Control and Signal Processing*, 24(3):155–177, 2010.

Feldbaum, A.A. Dual-control theory. I. *Automation and Remote Control*, 21(9):874–880, 1961.

Finn, C.K., Wahlberg, B., and Ydstie, B.E. Constrained predictive control using orthogonal expansions. *AIChE Journal*, 39(11):1810–1826, 1993.

GAMS Development Corporation. The General Algebraic Modeling System. 2016. `https://www.gams.com/`.

Gevers, M. and Ljung, L. Optimal experiment designs with respect to the intended model application. *Automatica*, 22(5):543–554, 1986.

Gill, P.E., Murray, W., and Saunders, M.A. SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization. *SIAM Review*, 47(1):99–131, 2005.

Haesaert, S., Van den Hof, P.M.J., and Abate, A. Data-driven and model-based verification via bayesian identification and reachability analysis. *Automatica*, 2017. https://www.cs.ox.ac.uk/people/alessandro.abate/publications/HVA17.pdf.

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

Heirung, T.A.N., Ydstie, B.E., and Foss, B. Dual MPC for FIR systems: Information anticipation. In *Advanced Control of Chemical Processes*, pages 1034–1039. Whistler, Canada, 2015b.

Heuberger, P.S.C., Van den Hof, P.M.J., and Wahlberg, B. *Modelling and Identification with Rational Orthogonal Basis Functions*. Springer, 2005.

Kulcsár, C., Pronzato, L., and Walter, E. Dual control of linearly parameterised models via prediction of posterior densities. *European Journal of Control*, 2(2):135–143, 1996.

Kumar, K., Heirung, T.A.N., Patwardhan, S.C., and Foss, B. Experimental evaluation of a MIMO adaptive dual MPC. In *Advanced Control of Chemical Processes*, pages 546–551. Whistler, Canada, 2015.

Larsson, C.A., Rojas, C.R., Bombois, X., and Hjalmarsson, H. Experimental evaluation of model predictive control with excitation (MPC-X) on an industrial depropanizer. *Journal of Process Control*, 31:1–16, 2015.

Lee, J.M. and Lee, J.H. An approximate dynamic programming based approach to dual adaptive control. *Journal of Process Control*, 19(5):859–864, 2009.

Ljung, L. *System Identification. Theory for the User*. Prentice-Hall, second edition, 1999.

Marafioti, G., Bitmead, R.R., and Hovd, M. Persistently exciting model predictive control. *International Journal of Adaptive Control and Signal Processing*, 28(6):536–552, 2014.

Mareels, I., Bitmead, R.R., Gevers, M., Johnson, C.R., Kosut, R.L., and Poubelle, M.A. How exciting can a signal really be? *Systems & Control Letters*, 8(3):197–204, 1987.

Mareels, I. and Polderman, J.W. *Adaptive Systems: An Introduction*. Birkhäuser, 1996.

Mayne, D.Q. Model predictive control: Recent developments and future promise. *Automatica*, 50(12):2967–2986, 2014.

Mayne, D.Q., Rawlings, J.B., Rao, C.V., and Scokaert, P.O.M. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, 2000.

Misener, R. and Floudas, C.A. GLOMIQO: Global mixed-integer quadratic optimizer. *Journal of Global Optimization*, 57(1):3–50, 2013.

Sbarbaro, D., Filatov, N.M., and Unbehauen, H. Adaptive predictive controllers based on orthonormal series representation. *International Journal of Adaptive Control and Signal Processing*, 13(7):621–631, 1999.

Shouche, M.S., Genceli, H., Vuthandam, P., and Nikolaou, M. Simultaneous constrained model predictive control and identification of DARX processes. *Automatica*, 34(12):1521–1530, 1998.

Sutton, R.S. and Barto, A.G. *Reinforcement Learning. An Introduction*. The MIT Press, 1998.

Tanaskovic, M., Fagiano, L., Smith, R.S., and Morari, M. Adaptive receding horizon control for constrained MIMO systems. *Automatica*, 50(12):3019–3029, 2014.

Tawarmalani, M. and Sahinidis, N.V. *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming. Theory, Algorithms, Software, and Applications*. Kluwer, 2002.

Tawarmalani, M. and Sahinidis, N.V. A polyhedral branch-and-cut approach to global optimization. *Mathematical Programming*, 103(2):225–249, 2005.

Tse, E. and Bar-Shalom, Y. An actively adaptive control for linear systems with random parameters via the dual control approach. *IEEE Transactions on Automatic Control*, 18(2):109–117, 1973.

Wahlberg, B. System identification using Laguerre models. *IEEE Transactions on Automatic Control*, 36(5):551–562, 1991.

Wahlberg, B. System identification using Kautz models. *IEEE Transactions on Automatic Control*, 39(6):1276–1282, 1994.

Wieslander, J. and Wittenmark, B. An approach to adaptive control using real time identification. *Automatica*, 7(2):211–217, 1971.