

Robust Dual Control MPC with Guaranteed Constraint Satisfaction

Weiss, A.; Di Cairano, S.

TR2014-110 December 2014

Abstract

We present a robust dual control MPC (RDC-MPC) policy with guaranteed constraint satisfaction for simultaneous closed-loop identification and regulation of state and input-constrained linear systems subject to parametric and additive uncertainty. The uncertain system is modeled as a polytopic Linear Difference Inclusion (pLDI) for which a maximal robust control invariant (RCI) set is calculated. Selecting a control from the associated robust admissible input (RAI) set guarantees constraint satisfaction for all pLDI realizations, and thus guarantees constraint satisfaction during the identification transient when the MPC prediction model is uncertain. The MPC problem is then cast as selecting a control from the RAI set that optimizes the dual objective of identifying the unknown system parameters and regulating the true system, where the tradeoff between the two objectives is adjusted based on the prediction error of the identified system. Numerical examples illustrate the proposed scheme's effectiveness and performance increase, while guaranteeing robust constraint satisfaction.

IEEE Conference on Decision and Control

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Robust Dual Control MPC with Guaranteed Constraint Satisfaction

Avishai Weiss¹ and Stefano Di Cairano²

Abstract—We present a robust *dual control* MPC (RDC-MPC) policy with guaranteed constraint satisfaction for simultaneous closed-loop identification and regulation of state and input-constrained linear systems subject to parametric and additive uncertainty. The uncertain system is modeled as a polytopic Linear Difference Inclusion (pLDI) for which a maximal robust control invariant (RCI) set is calculated. Selecting a control from the associated robust admissible input (RAI) set guarantees constraint satisfaction for all pLDI realizations, and thus guarantees constraint satisfaction during the identification transient when the MPC prediction model is uncertain. The MPC problem is then cast as selecting a control from the RAI set that optimizes the dual objective of identifying the unknown system parameters and regulating the actual system, where the tradeoff between the two objectives is adjusted based on the prediction error of the identified system. Numerical examples illustrate the proposed scheme’s effectiveness and performance increase, while guaranteeing robust constraint satisfaction.

I. INTRODUCTION

This paper develops a *dual control* MPC policy with guaranteed constraint satisfaction for simultaneous identification and control of uncertain systems. While regulation and identification are seemingly conflicting objectives, by identifying (or re-identifying) the system dynamics online, and updating the MPC prediction model, closed-loop performance may be enhanced relative to a standard MPC scheme that uses an inaccurate (or outdated) model. The optimal solution to this trade-off between identification and control is given by a dynamic program [1], which, for many applications, is computationally intractable. As a result, many suboptimal techniques and heuristics have been developed in recent years – often in a *receding horizon* framework such as model predictive control (MPC) [2].

In [3], an approximation of the dynamic program is developed for a linear input-output map with no dynamics. Approaches for more complex systems avoid dynamic programming altogether and, instead, sub-optimally trade between inputs that excite the system and inputs that regulate the state. Excitation signals are often designed to satisfy *persistence of excitation* conditions. For example, a dithering signal may be added on top of the nominal control [4], although difficulties arise in determining the amplitude of the signal, and the dither indiscriminately adds noise to the process. More sophisticated schemes employ *optimal input design*, generally in the frequency domain, where maximizing the Fisher information matrix can be cast as a semidefinite program [5]. However, design in the frequency domain leads to difficulties with constraints that are more naturally addressed

in the time domain, e.g., input magnitude constraints. While the problem formulation in the time domain is highly non-convex, developing such techniques is desirable, and thus the focus of recent work [6]–[10]. In [3] and [11], convex relaxations of the input design problem in the time domain are presented.

Still, most adaptive or learning-based (dual control) MPC strategies that incorporate closed-loop identification have difficulties with state constraints. They either do not include state constraints [6], or they may violate state constraints during the learning transient due to a *certainty equivalence* assumption, that is, using parameter estimates as if they were the true values [7]. Approaches that are able to enforce constraints may be computationally intensive due to solving min-max problems [8], or conservative due to enforcement of open-loop robustness in a receding horizon framework, uncertainty compensation using a fixed feedback law, or only handling additive uncertainty [8]–[10]. Thus, the main goal of this work is to develop a learning MPC policy that robustly guarantees state constraint satisfaction in the presence of both parametric and additive uncertainty.

In our approach, a robust control invariant (RCI) set is calculated for an uncertain linear system modeled by a polytopic Linear Difference Inclusion (pLDI). The set that we compute is the maximal RCI set for the pLDI, that is, the largest set of states for which there exists an input such that, under all possible realizations of the pLDI, state and input constraints can be robustly enforced using *any nonlinear* static feedback law. For any state in the maximal RCI set, the maximal robust admissible input (RAI) set is the set of inputs that can be applied to the system with the guarantee that the resulting state remains in the RCI set.

We formulate a robust dual control MPC (RDC-MPC) policy for simultaneous identification and control that may only cull inputs from the RAI set, thus, guaranteeing that the state will evolve in the RCI set and satisfy constraints at all times. For identification, we excite the system by optimizing for inputs that minimize a functional of the predicted error covariance of the estimated combination of pLDI vertices. We weight the identification functional based on the output prediction error [12]. Thus, when the prediction error is small, the identification functional disappears and the system is regulated to the origin. As in [10], this approach decouples safety and performance, that is, so long as inputs are drawn from the RAI set, we can incorporate any number of closed-loop identification schemes *and* guarantee that state constraints will not be violated.

The paper is organized as follows. In Section II we introduce the pLDI model and provide a review of RCI sets. In

The authors are with Mitsubishi Electric Research Laboratories, Cambridge, MA, ¹ weiss@merl.com, ² dicairano@ieee.org

Section III we develop the RDC-MPC policy. In Section IV, we introduce a control Lyapunov function (CLF) constraint in the RDC-MPC policy to guarantee stability. Section V highlights the proposed strategy on numerical examples. Finally, we provide concluding remarks and discuss future research directions in Section VI.

II. PRELIMINARIES

In this section we introduce the notation and the main concepts for pLDI systems and RCI sets that are used in the subsequent development of the RDC-MPC policy.

A. Notation

\mathbb{R} , \mathbb{R}_{0+} , \mathbb{R}_+ are the sets of real, nonnegative real, positive real numbers, and \mathbb{Z} , \mathbb{Z}_{0+} , \mathbb{Z}_+ are the sets of integer, non-negative integer, positive integer numbers. Unless otherwise specified, $\|\cdot\|$ indicates either the 1 or ∞ -norm. For a discrete time signal $x \in \mathbb{R}^n$ with sampling period T_s , $x(t)$ is the state at sampling instant t , i.e., at time $T_s t$. The notation x_k denotes the predicted value of x at sample k . By $[x]_i$ we denote the i -th component of x . By I and 0 we denote the identity and the “all-zero” matrices of appropriate size. By $\text{co}\{x_i\}_{i=1}^{n_s}$ we denote the convex hull of $\{x_i\}_{i=1}^{n_s}$. Given $\mathcal{P} \subseteq \mathbb{R}^{n_p}$, $x \in \mathbb{R}^{n_x}$, and $n_p \geq n_x$, the shorthand notation $\text{proj}_x(\mathcal{P})$ denotes the projection of \mathcal{P} onto x . By $\mathcal{P} \oplus \mathcal{Q}$ we denote the Minkowski sum of \mathcal{P} and \mathcal{Q} , and $\mathcal{Q} \equiv \mathcal{P}$ indicates that the sets are equal. A function $\alpha : \mathbb{R} \rightarrow \mathbb{R}_{0+}$ is class \mathcal{K}_∞ if it is monotonically increasing, $\alpha(0) = 0$, and $\lim_{k \rightarrow \infty} \alpha(k) = \infty$.

B. Polytopic Linear Difference Inclusions

Consider ℓ discrete time linear systems

$$x(t+1) = A_i x(t) + B_i u(t), \quad i = 1, \dots, \ell.$$

The polytopic Linear Difference Inclusion (pLDI) based on these (vertex) systems is defined as

$$x(t+1) \in \text{co}\{A_i x(t) + B_i u(t)\}_{i=1}^\ell, \quad (1)$$

and models the family of systems

$$x(t+1) = \sum_{i=1}^\ell [\theta]_i (A_i x(t) + B_i u(t)),$$

parameterized by $\theta \in \mathbb{R}^\ell$, where θ is unknown, $0 \leq [\theta]_i \leq 1$, and $\sum_i [\theta]_i = 1$.

Consider, now, the case where a fixed unknown additive disturbance $w \in \mathcal{W}$ is present. Let \mathcal{W} be a polytope, e.g, $\mathcal{W} = \text{co}\{w_1, \dots, w_p\}$. The disturbed pLDI (dpLDI) is defined as

$$x(t+1) \in \text{co}\{A_i x(t) + B_i u(t)\}_{i=1}^\ell \oplus \text{co}\{B_w w_i\}_{i=1}^\ell, \quad (2)$$

and models the family of systems

$$x(t+1) = \sum_{i=1}^\ell [\theta]_i (A_i x(t) + B_i u(t)) + \sum_{i=1}^p [\eta]_i B_w w_i, \quad (3)$$

parameterized by unknown $\theta \in \mathbb{R}^\ell$, $\eta \in \mathbb{R}^p$, where $0 \leq [\theta]_i \leq 1$, $0 \leq [\eta]_i \leq 1$, and $\sum_i [\theta]_i = \sum_i [\eta]_i = 1$.

Finally, consider an uncertain linear system

$$x(t+1) = A(\xi)x(t) + B(\xi)u(t) + B_w w(t), \quad (4)$$

where $\xi \in \Xi$. Define $\tilde{\xi}_i \in \tilde{\Xi}$, $i = 1, \dots, \ell$, such that

$$[A(\xi) \ B(\xi)] \in \text{co}\{[A(\tilde{\xi}_i) \ B(\tilde{\xi}_i)]\}_{i=1}^\ell, \quad \forall \xi \in \Xi,$$

and let

$$[A_i \ B_i] = [A(\tilde{\xi}_i) \ B(\tilde{\xi}_i)]. \quad (5)$$

Thus, the dpLDI (2) based on the matrices in (5) is said to include the uncertain linear system (4). Note that $\tilde{\Xi}$ may differ from Ξ .

C. Robust Control Invariant Sets

We recall some basic definitions and results on robust control invariant (RCI) sets for constrained systems. Details can be found, e.g., in [13].

Consider the system

$$x(t+1) = f(x(t), u(t), w(t)), \quad (6)$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $w \in \mathbb{R}^p$ are the state, input, and disturbance vectors, respectively, subject to the constraints

$$x(t) \in \mathcal{X}, \quad u(t) \in \mathcal{U}, \quad \forall t \in \mathbb{Z}_{0+}. \quad (7)$$

An RCI set is a set of states for which there exists a control law such that (6) never violates (7) for any admissible disturbances sequence.

Definition 1 (Robust control invariant set). *A set $\mathcal{C} \subseteq \mathcal{X}$ is said to be a robust control invariant (RCI) set for (6)-(7) if*

$$x(t) \in \mathcal{C} \Rightarrow \exists u(t) \in \mathcal{U} :$$

$$f(x(t), u(t), w(t)) \in \mathcal{C}, \quad \forall w(t) \in \mathcal{W}, \quad \forall t \in \mathbb{Z}_{0+}.$$

The set $\mathcal{C}^\infty \subseteq \mathcal{X}$ is the maximal RCI set if it is an RCI set and contains all the other RCI sets in \mathcal{X} . \square

Definition 2 (Robust admissible input set for $x \in \mathcal{C}$). *Given an RCI set \mathcal{C} for (6)-(7), the robust admissible input (RAI) set for state $x \in \mathcal{C}$ is*

$$\mathcal{C}_u(x) = \{u \in \mathcal{U} : f(x, u, w) \in \mathcal{C}, \quad \forall w \in \mathcal{W}\}.$$

The set $\mathcal{C}_u^\infty(x)$ is the maximal RAI set for $x \in \mathcal{C}^\infty$. \square

RCI and RAI sets for the dpLDI system (2) can be computed, for instance, as in Algorithms 1 and 2 of [14].

III. ROBUST DUAL CONTROL MPC

We now develop the RDC-MPC policy. We first present a robust MPC policy with guaranteed recursive feasibility and constraint satisfaction, and then we modify it for online closed-loop identification.

A. Feasibility and Constraint Satisfaction

Consider the dpLDI (2) subject to the constraints (7), for which the maximal RCI set \mathcal{C}^∞ and the associated maximal RAI set $\mathcal{C}_u^\infty(x)$ have been computed. We formulate an MPC

policy that may only cull inputs from the RAI set, thereby guaranteeing that the state will evolve in the RCI set and satisfy constraints for all time. Moreover, because the state evolves in the RCI set, there always exists an admissible input, and thus the MPC policy is recursively feasible. This result is formalized in Theorem 1.

Theorem 1. Let $[\hat{A} \ \hat{B}] \in \text{co}\{[A_i \ B_i]\}_{i=1}^{\ell}$, $\hat{w} = \sum_{i=1}^p \hat{\eta}_i w_i \in \mathcal{W}$, and B_w be known. Let $x(t)$ be governed by the dpLDI (2) and subject to the constraints (7). Consider the MPC policy that at any $t \in \mathbb{Z}_{0+}$ solves

$$\begin{aligned} \min_U \quad & F_t(x_N) + \sum_{k=1}^N L_t(x_k, u_k), \\ \text{s.t.} \quad & x_{k+1} = \hat{A}x_k + \hat{B}u_k + B_w \hat{w}, \\ & x_0 = x(t), \\ & u_k \in C_u^\infty(x_k), \end{aligned} \quad (8)$$

where $U = [u_0, \dots, u_{N-1}]$, N is the prediction horizon, and F_t, L_t are arbitrary functions of their arguments, and applies $u(t) = u_0^*$, where $[u_0^*, \dots, u_{N-1}^*]$ is the optimal solution of (8). Then, (8) is feasible at time \bar{t} if and only if $x(\bar{t}) \in \mathcal{C}^\infty$. Also, for every $t \geq \bar{t}$, $x(t) \in \mathcal{X}$, $u(t) \in \mathcal{U}$ for any $[A \ B] \in \text{co}\{[A_i \ B_i]\}_{i=1}^{\ell}$ and $w(t) \in \mathcal{W}$.

Proof (sketch): Let (8) be feasible at time t , then, by the definition of the RAI set, $x(t) \in \mathcal{C}^\infty$. Let $x(t) \in \mathcal{C}^\infty$, then by the RAI set properties (8) is feasible. Furthermore, $u(t) = u_0$ for any $u_0 \in C_u^\infty(x_0) \subseteq \mathcal{U}$ ensures that, for any $[A \ B] \in \text{co}\{[A_i \ B_i]\}_{i=1}^{\ell}$ and $w(t) \in \mathcal{W}$, $Ax(t) + Bu(t) + B_w w(t) \in \mathcal{C}^\infty \subseteq \mathcal{X}$. Since $x(t+1) \in \mathcal{C}^\infty$, the procedure can be iterated. Thus, $x(\tau) \in \mathcal{X}$, $u(\tau) \in \mathcal{U}$, for all $\tau \geq t$. \square

B. Online Closed-Loop Identification

We now modify the MPC policy (8) so that the true system is learned, that is, we let the estimated system matrices and estimated disturbance vector be functions of time, $\hat{A} = \hat{A}(t), \hat{B} = \hat{B}(t), \hat{w} = \hat{w}(t)$, and update them so that $[\hat{A}(t) \ \hat{B}(t)] \rightarrow [A \ B]$ and $\hat{w}(t) \rightarrow w$ as $t \rightarrow \infty$. Note that for clarity and to keep the subsequent expressions simple, the additive disturbance vector w is now assumed to be constant.

Consider the dpLDI (2) which models the family of systems (3). Our goal is to estimate (or identify) in real time the parameters θ and η , which can be used to reconstruct A, B , and w . To do so, we assume that full state measurements are available and form a linear regression model for the joint parameter vector $\lambda = [\theta^T \ \eta^T]^T$ given by

$$\begin{aligned} y(t+1) &= x(t+1) + \epsilon, \\ &= \sum_{i=1}^{\ell} [\theta]_i (A_i x(t) + B_i u(t)) + \sum_{i=1}^p [\eta]_i B_w w_i + \epsilon, \\ &= \Gamma(x(t), u(t)) \theta + \Upsilon \eta + \epsilon, \\ &= M^T(x(t), u(t)) \lambda + \epsilon, \end{aligned} \quad (9)$$

where ϵ represents measurement noise, $\Gamma(x, u) = [A_1 x + B_1 u, \dots, A_\ell x + B_\ell u]$, $\Upsilon = [B_w w_1, \dots, B_w w_p]$, and

$M^T(x, u) = [\Gamma(x, u) \ \Upsilon]$. We use Recursive Least Squares (RLS) for online parameter estimation, where the parameter estimate is given by

$$\hat{\lambda}(t+1) = \hat{\lambda}(t) + K(t+1) \left(y(t+1) - M^T(x(t), u(t)) \hat{\lambda}(t) \right),$$

where

$$K(t+1) = P(t+1)M(x(t), u(t)), \quad (10)$$

$$P(t+1) = (I - K(t+1)M^T(x(t), u(t)))P(t). \quad (11)$$

Remark 1. If $\dim(\lambda) > \dim(x)$, then λ cannot be uniquely recovered, that is, λ may be shifted by any vector $v \in \ker(M^T)$ and still yield the same y . Thus, $\hat{\lambda}$ may not necessarily converge to λ . Nevertheless, it is still possible for $[\hat{A}(t) \ \hat{B}(t)] \rightarrow [A \ B]$ and $\hat{w}(t) \rightarrow w$ even if $\hat{\lambda} \not\rightarrow \lambda$ since multiple parameter vectors λ may correspond to the same system matrices.

The task now becomes to select proper inputs such that the true system is learned. We modify the optimization problem (8) to minimize a functional of the predicted parameter error covariance in order to enhance persistency of excitation. The optimization problem is thus given by

$$\begin{aligned} \min_U \quad & F_t(x_N) + \sum_{k=1}^N L_t(x_k, u_k, P_k), \\ \text{s.t.} \quad & x_{k+1} = M^T(x_k, u_k) \text{proj}_\lambda(\hat{\lambda}(t)), \\ & x_0 = x(t), \\ & P_0 = P(t), \\ & u_k \in C_u^\infty(x_k), \end{aligned} \quad (12)$$

where the projection ensures that $\hat{\theta}, \hat{\eta}$ are appropriate convex combination vectors, and L_t forms a dual objective, trading regulation and learning. For example,

$$\begin{aligned} L_t(x_k, u_k, P_k) &= x_k^T Q x_k + u_k^T R u_k \\ &+ \gamma \left(y(t) - M^T(x(t-1), u(t-1)) \hat{\lambda}(t-1) \right) \psi(P_k), \end{aligned} \quad (13)$$

where γ is a function of the prediction error, ψ is an information functional of the predicted parameter error covariance P_k , and P_k is predicted over the optimization horizon using (10)-(11). The learning objective is only weighted when the prediction model is poor and, once the error is small, the learning objective vanishes and we recover the original non-learning robust MPC policy (8). Examples of the functional ψ include $\text{Tr}(P_k)$, $\log \det(P_k)$, and the maximum eigenvalue $\lambda_{\max}(P_k)$, where, borrowing the convention from experiment design literature, they correspond to A-optimal, D-optimal, and E-optimal measures, respectively.

Note that (12)-(13) is not a convex problem in the decision variables, i.e. the control inputs. In the simulation section V, we solve the non-convex problem (12)-(13) using Matlab's `fmincon` routine with an interior-point solver. Of course, there is no guarantee of convergence, which is not practical for real world implementation. A more robust algorithm may take the form of a 2-step procedure where, first, the non-

convex closed-loop identification problem

$$\begin{aligned} \min_{U_d} \quad & \sum_{k=1}^N \psi(P_k), \\ \text{s.t.} \quad & x_{k+1} = M^T(x_k, u_{d,k}) \text{proj}_{\lambda}(\hat{\lambda}(t)), \\ & x_0 = x(t), \\ & P_0 = P(t), \\ & u_{d,k} \in \mathcal{C}_u^\infty(x_k), \end{aligned} \quad (14)$$

is solved, generating a desired input sequence $U_d(t)$ for learning the true system. Then, deviations from the desired signal are penalized in (12) by letting L_t take the form

$$\begin{aligned} L_t(x_k, u_k, u_{d,k}) &= x_k^T Q x_k + u_k^T R u_k \\ &+ \gamma \left(y(t) - M^T(x(t-1), u(t-1)) \hat{\lambda}(t-1) \right) \|u_k - u_{d,k}\|_2^2. \end{aligned} \quad (15)$$

The advantage of splitting the algorithm into 2 steps lies in the fact that if (14) fails, (12),(15) still generates an admissible input due to the constraint $u_k \in \mathcal{C}_u^\infty(x_k)$, with $u_{d,k}$ set to $u_{d,k-1}$. The method of autocorrelation described in [15] may also be relied upon as a desired input $U_d(t)$.

Alternatively, the non-convex problem (12)-(13) may be relaxed. In [3], the covariance matrix P_k is linearized in U thereby transforming the problem into a convex second order cone program (SOCP). In [11], a convex relaxation of the input design problem is presented that yields an input sequence that is within a tight bound of the global optimum. We leave the convex relaxation of (12), (13) to future work.

Remark 2. If the prediction model in (12) is updated based on $\hat{\lambda}$, but no effort is made to select inputs that increase the information about the system (and thereby decrease the parameter error covariance), e.g., $L_t(x_k, u_k) = x_k^T Q x_k + u_k^T R u_k$, then we deem the policy passive with respect to learning. The passive learning policy can be seen as an example of an indirect adaptive control scheme. Since the control is unlikely to produce persistently exciting signals on its own, such a policy may experience instability in the parameter estimates – otherwise known as the bursting phenomenon [16].

Remark 3. Note that a major feature of the RDC-MPC policy is that inputs are culled from the RAI set $u_k \in \mathcal{C}_u^\infty(x_k)$ in order to guarantee constraint satisfaction. Any online supervised learning scheme that is amenable to selecting inputs in this restricted domain may be substituted in place of RLS and still retain the properties of this approach. Moreover, while not pursued in the present paper, some control problems may not require a dual control approach. An alternative policy might employ 2 separate phases: an identification phase and a control phase, each satisfying $u_k \in \mathcal{C}_u^\infty(x_k)$, where the switch between the phases is based on the prediction error. A switch from the control phase to the identification phase would take place if the prediction error grows beyond a specified threshold.

IV. ROBUST STABILIZATION WITH RDC-MPC

For the uncertain system (2), Theorem 1 guarantees that if at time $\bar{t} \in \mathbb{Z}_{0+}$, $x(\bar{t}) \in \mathcal{C}^\infty$, then, the MPC policy

(12) ensures that $x(t) \in \mathcal{C}^\infty \subseteq \mathcal{X}$ for all $t \geq \bar{t}$. If \mathcal{X} is bounded, the closed-loop system is ultimately bounded. Furthermore, if the cost function in (12) is properly designed (especially the terminal cost, or the horizon length [2]), once the parameter estimates converge to their true values, then the controller will not only guarantee ultimate boundedness, but will also be asymptotically stabilizing. While we leave for future work the stability analysis of such a policy, we now develop a method that guarantees robust asymptotic stability, regardless of the convergence of the estimated system parameters. We use an approach based on control Lyapunov functions (CLFs) [17], where a CLF is introduced in (12) as an additional constraint [18], [19], see also the references therein.

Given a system $x(t+1) = f(x(t), u(t))$, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, if there exists a set $\bar{\mathcal{X}}$, and a function $\mathcal{V} : \mathbb{R}^n \rightarrow \mathbb{R}_{0+}$ such that for $\alpha_1, \alpha_2 \in \mathcal{K}_\infty$, $\rho \in [0, 1)$, and for all $x \in \bar{\mathcal{X}}$,

$$\alpha_1(x) \leq \mathcal{V}(x) \leq \alpha_2(x) \quad (16a)$$

$$\begin{aligned} \exists u \in \mathcal{U} : \quad & f(x, u) \in \bar{\mathcal{X}} \\ & \mathcal{V}(f(x, u)) \leq \rho \mathcal{V}(x), \end{aligned} \quad (16b)$$

the system admits a control law that makes the closed-loop asymptotically stable in $\bar{\mathcal{X}}$ by enforcing (16b).

Consider the pLDI system (1) (i.e., $\mathcal{W} = \{0\}$), and let $\rho \in [0, 1)$ and $\mathcal{V}(x) = \|Px\|_q$, where $q \in \{1, \infty\}$, and $P^{n_p \times n}$ is full column rank, be given. Then, $\mathcal{V}(f(x, u)) \leq \rho \mathcal{V}(x)$ can be rewritten as $\sum_{i=1}^{\ell} [\lambda]_i \|P(A_i x + B_i u)\|_q \leq \rho \sum_{i=1}^{\ell} [\lambda]_i \|Px\|_q$. Thus, for the pLDI (1), if

$$\|P(A_i x + B_i u)\|_q \leq \rho \|Px\|_q, \quad \forall i = 1, \dots, \ell, \quad (17)$$

then $\mathcal{V}(f(x, u)) \leq \rho \mathcal{V}(x)$. Condition (17) can be implemented by linear inequalities via auxiliary variables $\varepsilon_i \in \mathbb{R}_{0+}$, $i = 1, \dots, n_a$, where $n_a = 1$, for $q = \infty$, and $n_a = n_p$, for $q = 1$.

Combining constraint (17) with the constraint of maintaining the state in the RCI set, we obtain the feasible set

$$\mathcal{F} = \{x \in \mathbb{R}^n : \exists u \in \mathcal{C}_u^\infty(x), \|P(A_i x + B_i u)\|_q \leq \rho \|Px\|_q, \forall i = 1, \dots, \ell\}, \quad (18)$$

where, since $u \in \mathcal{C}_u^\infty(x)$, $\mathcal{F} \subseteq \mathcal{C}^\infty$. In general, \mathcal{F} is not guaranteed to be RCI for (1). Conditions certifying that \mathcal{F} is RCI, and hence that $\mathcal{V}(x) = \|Px\|_q$ is a valid CLF in \mathcal{F} , may be verified. These amount to checking that $\mathcal{F} \equiv \mathcal{C}^\infty$, which, for $q = 1, \infty$, can be done by solving linear problems and convex set operations, and are not discussed here due to space limitations. If \mathcal{F} is verified to be RCI, then the MPC policy (12) with the additional constraint (17) is recursively feasible and asymptotically stabilizing, as formally proved next.

Theorem 2. *Let $P \in \mathbb{R}^{n_p \times n}$, $\rho \in [0, 1)$ be such that (17) is a valid CLF for (1) in $\mathcal{C}^\infty \equiv \mathcal{F}$. Consider the MPC policy that any $t \in \mathbb{Z}_{0+}$ solves*

$$\begin{aligned}
\min_U \quad & F_t(x_N) + \sum_{k=1}^N L_t(x_k, u_k, P_k) \\
\text{s.t.} \quad & x_{k+1} = M^T(x_k, u_k)\hat{\lambda}(t), \\
& x_0 = x(t), \\
& P_0 = P(t), \\
& u_k \in \mathcal{C}_u^\infty(x_k), \\
& \|PM^T(x_0, u_0)e_i\|_\infty \leq \rho\|Px_0\|_\infty, \quad \forall i = 1, \dots, \ell \\
& x_0 = x(t)
\end{aligned} \tag{19}$$

and applies $u(t) = u_0^*$, where e_i is the i^{th} standard basis vector of \mathbb{R}^ℓ (i.e., $[e_i]_i = 1$, $[e_i]_j = 0$, for all $j \neq i$), and possibly $\lambda(t) \neq \hat{\lambda}(t)$. Then, (19) is feasible at time \bar{t} if and only if $x(\bar{t}) \in \mathcal{C}^\infty$. Also, (1) in closed-loop with the MPC policy (19) is such that for every $t \geq \bar{t}$, $x(t) \in \mathcal{X}$, $u(t) \in \mathcal{U}$, and the system is asymptotically stable.

Proof (sketch:) The proof is very similar to that of Theorem 1, and repeated satisfaction of (17) guarantees that the closed-loop system is (robustly) asymptotically stable, as in [18]. \square

Even if $\mathcal{F} \equiv \mathcal{C}^\infty$, constraint (17) generally reduces the degree of freedom for optimizing control and learning performance in (19) relative to (12), because not all the inputs in $\mathcal{C}_u^\infty(x)$ satisfy (17).

If $w \in \mathcal{W}$, $\mathcal{W} \neq \{0\}$, the previous result provides input-

to-state stability of the closed-loop system with respect to $w \in \mathcal{W}$. By computing \mathcal{C}^∞ for the dpLDI system (2), and enforcing (17), where the disturbance is ignored and which is verified to be a CLF for (1), (i.e., for (2) with $\mathcal{W} = \{0\}$), if $x(t) \in \mathcal{C}^\infty$, then $\|Px(t+1)\|_\infty \leq \rho\|Px(t)\|_\infty + \|PB_w w(t)\|_\infty$, and since $u(t) \in \mathcal{C}_u^\infty(x(t))$, then $x(t+1) \in \mathcal{C}_u^\infty(x)$. Also, (17) can be relaxed by adding a constant nonnegative term to the right hand side to achieve ultimate boundedness in the presence of additive disturbances.

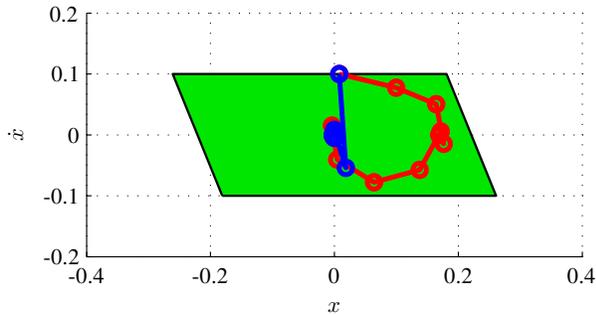
V. NUMERICAL EXAMPLES

We simulate RDC-MPC on a pLDI based on the discretization of the mass-spring-damper system given by

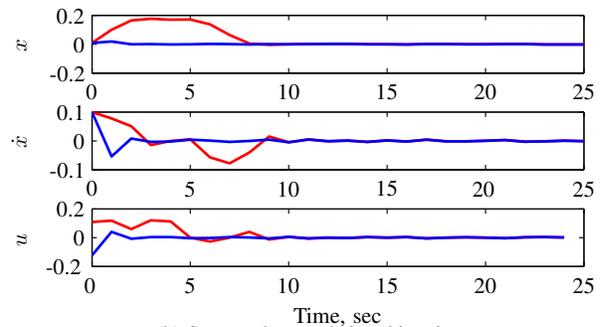
$$m\ddot{x} + c\dot{x} + kx = u, \tag{20}$$

where m is the mass, k is the spring constant, and c is the damping coefficient, and with a sampling period $T_s = 1$. We note that the results developed in this paper only apply to the discretized equations and not the continuous model (20). We consider the case without additive disturbance $\mathcal{W} = \{0\}$, and without enforcing the CLF as in (19) in order to better highlight the learning performance.

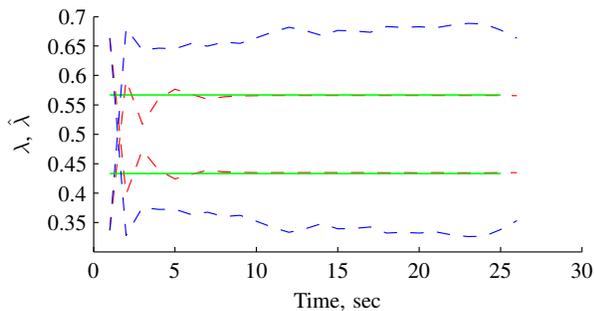
We first show that the RDC-MPC policy (12)-(13) effectively learns the system, whereas the non-learning robust policy (8) does not. Let $k = 1$, $c = 1$, and $m \in [1, 5]$ represent the uncertainty in the model, and let the state and input constraints be given by $|x| \leq 1$, $|\dot{x}| \leq 0.1$, and $|u| \leq 1$.



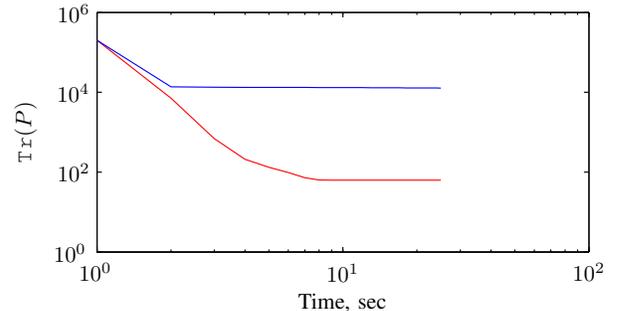
(a) Phase plane trajectories. The allowable states $x \in \mathcal{C}^\infty$ are shaded green.



(b) State and control time histories.



(c) Estimated parameter time histories. The green lines are the true values of λ .



(d) Trace of the predicted parameter error covariance matrix.

Fig. 1: Mass-Spring-Damper system with uncertain mass. In blue: trajectories and time histories under closed-loop control using the robust policy without learning (8). In red: trajectories and time histories under closed-loop control using RDC-MPC (12)-(13).

We form the pLDI (1) based on the vertices of m and compute the corresponding RCI and RAI sets. We let $N = 5$, $\psi(P_k) = \text{Tr}(P_k)$, $Q = I$, $R = 0.1I$, $P(0) = 10^5 I$, and $x(0) = [0.008 \ 0.1]^T$. We apply both the RDC-MPC policy (12)-(13) and the robust policy (8), where the corresponding optimization problems are solved using Matlab's `fmincon` routine with an interior-point solver. The results are presented in Figure 1. In Figure 1a we see the phase plane trajectories for both policies and notice that, as a result of the learning objective (13), RDC-MPC takes a more circuitous path to the origin. Neither policy violates the RCI set constraints shown by the shaded region in green. In Figure 1b we see the state and control time histories. The control input u remains well within the input constraints. Figure 1c shows that RDC-MPC has learned the model (i.e., $\hat{\lambda} \rightarrow \lambda$), whereas by applying RLS to the signals obtained under the non-learning robust policy, the parameters are not properly estimated due to lack of excitation. Finally, in Figure 1d we see that RDC-MPC only minimizes $\text{Tr}(P)$ when the prediction error is large, smoothly changing to the regulation objective once the model is learned.

We now show that RDC-MPC can provide enhanced performance relative to a non-learning policy. Consider the same mass-spring-damper system (20) as before, but with the additional spring constant uncertainty $k \in [0.1, 3]$. We let $c = 0.01$ and $R = 0.001I$. The rest of the parameters are as given previously. The initial condition $x(0)$ is set to a random combination of the vertices of the maximal RCI set \mathcal{C}^∞ . Every 10 seconds the state is reset to another random combination (during which the learning objective is temporarily frozen so as to not introduce an artificial disturbance in the prediction error). We apply both the RDC-MPC policy (12)-(13) and the robust policy (8). In Figure 2 we see

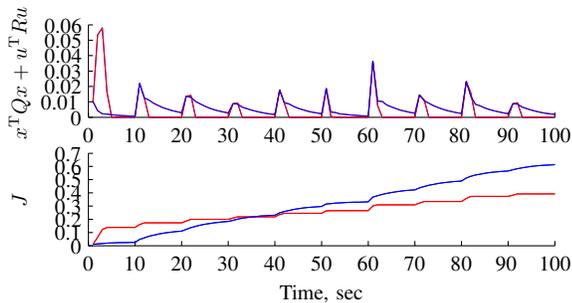


Fig. 2: Mass-Spring-Damper system with uncertain mass and spring constant. Top: Instantaneous stage cost. Bottom: Cumulative cost, $J = \sum x^T Q x + u^T R u$. In blue: time history under closed-loop control using the robust policy without learning (8). In red: time history under closed-loop control using RDC-MPC (12)-(13).

both the instantaneous performance $x^T Q x + u^T R u$ (top) and cumulative performance (bottom). The non-learning policy outperforms RDC-MPC initially, due to the initial learning transient, but over time, as RDC-MPC learns the true system, it recovers and outperforms the non-learning policy. Note that because $\dim(\lambda) > \dim(x)$, $\hat{\lambda}$ does not converge to λ . Nevertheless, $[\hat{A}(t) \ \hat{B}(t)] \rightarrow [A \ B]$ after 3 seconds.

VI. CONCLUSION

We have presented a robust dual control MPC policy for simultaneous identification and control of uncertain linear systems. Input and state constraint satisfaction is guaranteed even when the MPC prediction model is inaccurate. Asymptotic stability can be enforced by a control Lyapunov function. Future work will investigate less conservative stabilization techniques such as proving that the MPC value function is a Lyapunov function, and the exploitation of other supervised learning strategies, as well as convex relaxations of the input design problem. Finally, future work will address online adaptation of the RCI set, where the acquired knowledge of the system parameters may be exploited to enlarge the RAI set.

REFERENCES

- [1] A. Feldbaum, "Dual control theory," *Automation and Remote Control*, vol. 21, no. 9, pp. 874–1039, 1960.
- [2] J. Maciejowski, *Predictive control with constraints*. Englewood Cliffs, NJ: Prentice Hall., 2002.
- [3] M. S. Lobo and S. Boyd, "Policies for simultaneous estimation and optimization," in *Proc. American Contr. Conf.*, 1999, pp. 958–964.
- [4] O. A. Sotomayor, D. Odloak, and L. F. Moro, "Closed-loop model re-identification of processes under MPC with zone control," *Control Engineering Practice*, vol. 17, no. 5, pp. 551 – 563, 2009.
- [5] H. Jansson and H. Hjalmarsson, "Input design via LMIs admitting frequency-wise model specifications in confidence regions," *IEEE Trans. Automatic Control*, vol. 50, no. 10, pp. 1534–1549, 2005.
- [6] J. Rathoušký and V. Havlena, "MPC-based approximate dual controller by information matrix maximization," *International Journal of Adaptive Control and Signal Processing*, vol. 27, no. 11, pp. 974–999, 2013. [Online]. Available: <http://dx.doi.org/10.1002/acs.2370>
- [7] T. A. N. Heirung, B. E. Ydstie, and B. Foss, "An adaptive model predictive dual controller," in *Adaptation and Learning in Control and Signal Processing*, vol. 11, no. 1, 2013, pp. 62–67.
- [8] V. Adetola, D. DeHaan, and M. Guay, "Adaptive model predictive control for constrained nonlinear systems," *Systems & Control Letters*, vol. 58, no. 5, pp. 320–326, 2009.
- [9] A. González, A. Ferramosca, G. Bustos, J. Marchetti, and D. Odloak, "Model predictive control suitable for closed-loop re-identification," in *Proc. American Contr. Conf.* IEEE, 2013, pp. 1709–1714.
- [10] A. Aswani, H. Gonzalez, S. S. Sastry, and C. Tomlin, "Provably safe and robust learning-based model predictive control," *Automatica*, vol. 49, no. 5, pp. 1216–1226, 2013.
- [11] I. R. Manchester, "Input design for system identification via convex relaxation," in *Proc. 49th IEEE Conf. on Decision and Control*, 2010, pp. 2041–2046.
- [12] L. Ljung, "Prediction error estimation methods," *Circuits, Systems and Signal Processing*, vol. 21, no. 1, pp. 11–21, 2002.
- [13] E. C. Kerrigan, "Robust constraints satisfaction: Invariant sets and predictive control," Ph.D. dissertation, Dep. of Engineering, University of Cambridge, Cambridge, UK, 2000.
- [14] S. Di Cairano, A. Ulusoy, and S. Haghghat, "Soft-landing control by control invariance and receding horizon control," in *Proc. American Contr. Conf.* IEEE, 2014, pp. 784–789.
- [15] P. Hägg, C. Larsson, and H. Hjalmarsson, "Robust and adaptive excitation signal generation for input and output constrained systems," in *Proc. European Contr. Conf.*, 2013.
- [16] B. Anderson, "Adaptive systems, lack of persistency of excitation and bursting phenomena," *Automatica*, vol. 21, no. 3, pp. 247–258, 1985.
- [17] E. D. Sontag, "A Lyapunov-like characterization of asymptotic controllability," *SIAM Journal of Control and Optimization*, vol. 21, pp. 462–471, 1983.
- [18] P. O. M. Scokaert, D. Q. Mayne, and J. B. Rawlings, "Suboptimal model predictive control (feasibility implies stability)," *IEEE Trans. Automatic Control*, vol. 44, no. 3, pp. 648–654, 1999.
- [19] M. Lazar, "Model predictive control of hybrid systems: Stability and robustness," Ph.D. dissertation, Eindhoven University of Technology, The Netherlands, 2006.