Sub-linear Regret in Adaptive Model Predictive Control

Damianos Tranos [1] **Alexandre Proutiere** [1]

Abstract

We consider the problem of adaptive Model Predictive Control (MPC) for uncertain linear-systems with additive disturbances and with state and input constraints. We present STT-MPC (Self-Tuning Tube-based Model Predictive Control), an online algorithm that combines the certainty-equivalence principle and polytopic tubes. Specifically, at any given step, STT-MPC infers the system dynamics using the Least Squares Estimator (LSE), and applies a controller obtained by solving an MPC problem using these estimates. The use of polytopic tubes is so that, despite the uncertainties, state and input constraints are satisfied, and recursive-feasibility and asymptotic stability hold. In this work, we analyze the regret of the algorithm, when compared to an oracle algorithm initially aware of the system dynamics. We establish that STT-MPC expected regret does not exceed $O(T^{1/2+\epsilon})$, where $\epsilon \in (0,1)$ is a design parameter tuning the persistent excitation component of the algorithm. Our result relies on a recently proposed *exponential decay of sensitivity* property and, to the best of our knowledge, is the first of its kind in this setting. We illustrate the performance of our algorithm using a simple numerical example.

1. Introduction

The problem of optimal decision making for uncertain dynamical systems has been studied in both the control and the learning communities, and is referred to as stochastic optimal control, adaptive control, or reinforcement learning. A fundamental special case of this problem is the online (or adaptive) Linear Quadratic Regulator (LQR) which deals with the optimal control of systems with linear dynamics (Recht, 2019). Early efforts in the control community yielded algorithms with asymptotic performance and stability guarantees (Åström & Wittenmark, 1973; Lai, 1986). Over the last decade, the problem has been revisited in the learning community (see e.g. (Abbasi-Yadkori & Szepesvári, 2011; Mania et al., 2019; Goel & Hassibi, 2022; Jedra & Proutiere, 2022) and references therein), with the aim of devising algorithms with finite-time guarantees on the *regret*, defined as the difference between the cumulative cost of the learning algorithm and that of an oracle which has perfect knowledge of the system.

All aforementioned studies deal with the simple LQR problem and cannot account for constraints on the system state and input. These constraints are important in applications as they capture both the inherent limitations (e.g position, velocity, and actuator limits) as well as limitations imposed due to safety or the cost of operating the system. Unfortunately, it is well known that directly solving a constrained LQR problem is in general intractable because of the infinite time horizon (Scokaert & Rawlings, 1998). A way to circumvent this issue is to use Model Predictive Control (MPC) which solves a finite-time constrained optimization problem in a receding horizon manner.

Early work in the control community led to a better understanding of the connection between MPC and LQR and to algorithms with recursive feasibility and asymptotic stability guarantees (Mayne et al., 2000). In parallel, the tube MPC framework was developed to achieve the robust satisfaction of constraints in the presence of uncertainties (Mayne et al., 2005). Recently, online MPC algorithms, combining tube MPC and adaptive strategies (to cope with the system uncertainty), have been proposed and shown to ensure recursive feasibility and asymptotic stability (Lorenzen et al., 2017; Lu & Cannon, 2019; Lu et al., 2021), and (Tranos et al., 2022). The learning community has also begun to investigate MPC, again with the aim of providing finite-time regret guarantees (Li et al., 2019; Yu et al., 2020; Zhang et al., 2021; Lin et al., 2021; 2022). A key limitation of these analyses is that they are valid only in the absence of constraints which goes against the main motivation of MPC (Bitmead et al., 1990). Furthermore, most of these works assume to have access to (often perfect) system dynamics predictions. And those not making this assumption yield regret bounds having the cost of the oracle as an additive

[1]Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. Correspondence to: Damianos Tranos <tranos@kth.se>.

Workshop on New Frontiers in Learning, Control, and Dynamical Systems at the International Conference on Machine Learning (ICML), Honolulu, Hawaii, USA, 2023. Copyright 2023 by the author(s).

term. In turn, this term may scale linearly with time in the presence of constraints and additive disturbances.

In this work, we present STT-MPC, an adaptive MPC algorithm originally proposed in (Tranos et al., 2022). Inspired by the analysis pipeline of (Lin et al., 2022), we leverage the exponential decay of sensitivity property of the underlying finite-time optimization problem (Shin et al., 2022) to provide upper bounds of the expected regret of the algorithm. Specifically, we show that this regret scales at most as $T^{1/2+\epsilon}$ where $\epsilon \in (0,1)$ is a design parameter tuning the persitent excitation component of the algorithm.

Notation. For a time dependent vector x_t, we denote by $x_{k|t}$ its prediction at time $k+t$ given information at time t. For any two sets $\mathcal{A}$ and $\mathcal{B}$, we define their Minkowski sum as the set $\mathcal{A} \oplus \mathcal{B} := \{a+b : a \in \mathcal{A},\ b \in \mathcal{B}\}$. We also define, for any constant $\lambda \geq 0$, the scaled set $\lambda\mathcal{A} := \{\lambda a,\ a \in \mathcal{A}\}$. For any $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, and $\epsilon > 0$, let $\mathcal{B}(x,\epsilon) = \{y : \|y-x\| \leq \epsilon\}$. Let $\mathcal{B} = \mathcal{B}(0,1)$. For any set $\mathcal{S}$, and any $\varepsilon > 0$, there exists a polytope $\mathcal{P}$ that is an outer approximation of $\mathcal{S}$, i.e., $\mathcal{S} \subset \mathcal{P} \oplus \varepsilon\mathcal{B}$. We refer to this polytope as the outer polyhedral approximation of $\mathcal{S}$. Unless stated otherwise, we use $\|\cdot\|$ to denote the Euclidean norm for vectors and the Frobenius norm for matrices.

2. Problem Setting

We consider the following discrete time, linear, time-invariant system:

$$x_{t+1} = A(\theta^\star)x_t + B(\theta^\star)u_t + w_t, \tag{1}$$

where $x_t, w_t \in \mathbb{R}^{d_x}$ and $u_t \in \mathbb{R}^{d_u}$. The state transition and state-action transition matrices $A(\theta^\star)$ and $B(\theta^\star)$ are initially unknown. The set of possible such matrices is parameterized by $\theta \in \mathbb{R}^{d_\theta}$ (here θ could well parameterize each entry of the matrices, in which case $d_\theta = d_x(d_x+d_u)$). To simplify the notation, for two possible parameters θ_1, θ_2, we define $\|\theta_1 - \theta_2\| := \max(\|A(\theta_1) - A(\theta_2)\|, \|B(\theta_1) - B(\theta_2)\|)$. We make the following assumptions.

Assumption 2.1 (Parameter uncertainty)**.** The decision maker does not know $\theta^\star$, but knows that $\theta^\star \in \Theta_0$ where Θ_0 is a convex polytope. Moreover, there exists $\epsilon_0 > 0$ such that $\mathcal{B}(\theta^*, \epsilon_0) \subset \Theta_0$.

Assumption 2.2 (Additive disturbance)**.** The sequence $(w_t)_{t\geq 0}$ is i.i.d, and for each $t \geq 0$, w_t is zero-mean, isotropic, with support in the ball $\mathcal{B}(0, 3\sigma)$. Hence, w_t is σ^2-sub-gaussian. Further define $\mathcal{W}$, a convex polytope providing a conservative approximation of $\mathcal{B}(0,3\sigma)$, i.e., $\mathcal{B}(0,3\sigma) \subset \mathcal{W}$.

Assumption 2.3 (State and input constraints)**.** The set

$$\mathcal{C} = \{(x,u) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_u} : Fx + Gu \leq \mathbf{1}\}.$$

is compact and contains the origin in its interior. Here, $F \in \mathbb{R}^{d_c \times d_x}$ and $G \in \mathbb{R}^{d_c \times d_u}$ define the state and input constraints respectively. The above inequality holds component-wise, and $\mathbf{1}$ is the vector with all components equal to 1.

Assumption 2.4 (Stabilizing Controller)**.** There exists a known, robustly stabilizing feedback gain K such that $A(\theta) + B(\theta)K$ is stable (i.e., $\rho(A(\theta) + B(\theta)K) < 1$) for all $\theta \in \Theta_0$.

3. Model Predictive Control

We wish to minimize the long-term cost defined as $\limsup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\left[x_t^\top Q x_t + u_t^\top R u_t\right]$, through some positive semi-definite matrices Q, R. To this aim, we use MPC, with a receding horizon N. Specifically, at time t, given the current system state x_t and the past observations used to derive an estimator θ_t of $\theta^\star$, we will identify a control policy $(u_{k|t})_{k=0,\dots,N-1}$ minimizing the cost along a predicted system trajectory $(x_{k|t})_{k=0,\dots,N}$. We use the well-known dual mode prediction paradigm (Kouvaritakis & Cannon, 2016) with the following predicted control sequence,

$$u_{k|t} = \begin{cases} Kx_{k|t} + v_{k|t} & \forall k \in \{0,\dots,N-1\}, \\ Kx_{k|t} & \forall k \geq N, \end{cases}$$

where $\{v_{0|t}, \dots v_{N-1|t}\}$ are the optimization variables to be determined by the MPC. The resulting prediction dynamics will be for $k \in \{0,\dots,N-1\}$,

$$\begin{aligned} x_{0|t} &= x_t, \\ x_{k+1|t} &= \Phi(\theta_t)x_{k|t} + B(\theta_t)v_{k|t}, \end{aligned}$$

where $\Phi(\theta) := A(\theta) + B(\theta)K$ for any $\theta \in \Theta_0$.

3.1. LSE and persistent excitation

Our algorithm starts with an initial parameter $\theta_0 \in \Theta_0$, which is then updated using the LSE. We let $\theta_1 = \theta_0$. For $t \geq 2$, the LSE enjoys the following explicit expression:

$$\hat{\theta}_t = \left(\sum_{k=0}^{t-2} x_{k+1}\begin{bmatrix} x_k \\ u_k \end{bmatrix}^\top\right)\left(\sum_{k=0}^{t-2}\begin{bmatrix} x_k \\ u_k \end{bmatrix}\begin{bmatrix} x_k \\ u_k \end{bmatrix}^\top\right)^\dagger.$$

For $t \geq t^\star(\delta)$ (to be defined later), we align our prediction parameter θ_t to $\hat{\theta}_t$. As shown in (Jedra & Proutiere, 2022), a finite-time analysis of the performance of the LSE is rather intricate but possible even if the feedback controller varies over time. The performance is tightly related to the minimal eigenvalue of the cumulative covariate matrix $\lambda_{\min}(\sum_{s=0}^{t-2} y_s y_s^\top)$ where $y_s = \begin{bmatrix} x_s \\ u_s \end{bmatrix}$. More precisely, for the LSE to lead to a good approximation of $\theta^\star$, we need to

ensure that this eigenvalue grows with time. To this aim, we add an isotropic and bounded noise to the control input. This noise is represented by the random vector ζ_t taken to be the projection of ξ_t on $\mathcal{B}(0, 3\sigma_t)$, where ξ_t is i.i.d. according to a normal distribution, i.e., $\xi_t \sim \mathcal{N}(0, \sigma_t^2 I_{d_u})$. The choice of σ_t directly impacts the performance of the LSE but also the overall performance of the controller (a higher σ_t means higher excitation and hence better LSE, but at the expense of a worse overall controller). Here we set $\sigma_t^2 = \sqrt{d_x}\sigma^2 t^{-\alpha}$ for some $\alpha \in (0, 1)$. This ensures that (i) the LSE θ_t converges to $\theta^\star$ and (ii) the controller converges to that obtained through a classical tube-based MPC framework with known $\theta^\star$. In (Tranos et al., 2022), we have shown that the following *good* event $\mathcal{G}$ holds with probability at least $1 - \delta$:

$$\mathcal{G} = \left(\|\hat{\theta}_t - \theta^\star\| \leq \epsilon_t, \ \forall t \geq t^\star(\delta) \right),$$

where $t^\star(\delta) = c_1 + c_2 \log(1/\delta)$ and $\epsilon_t^2 = c_3 \log(t/\delta)/t^{1-\alpha}$ for some positive constants c_1, c_2, c_3. Next, we define Δ_t as an outer polyhedral approximation of $\mathcal{B}(\theta_t, 2\epsilon_t)$. We further recursively define the *uncertainty sets* as follows: $\Theta_t = \Theta_{t-1} \cap \Delta_t$ for all $t \geq t^\star(\delta)$ and $\Theta_t = \Theta_0$ for $t < t^\star(\delta)$. By construction, the true parameter $\theta^\star$ belongs to the interior of Θ_t with high probability in the following sense:

Lemma 3.1. *Under event $\mathcal{G}$, $\mathcal{B}(\theta^\star, \epsilon_t) \subset \Theta_t$ for all $t \geq 1$.*

Proof. For $t < t^\star(\delta)$ we have $\Theta_t = \Theta_0$ and the result holds by Assumption 2.1. Let $t \geq t^\star(\delta)$. We show that $\mathcal{B}(\theta^*, \epsilon_t) \subset \Delta_t$ and $\mathcal{B}(\theta^*, \epsilon_t) \subset \Theta_{t-1}$. For the first, we have for all $\theta \in \mathcal{B}(\theta^*, \epsilon_t)$, using the triangle inequality: $\|\theta - \theta_t\| \leq \|\theta - \theta^*\| + \|\theta_t - \theta^*\| \leq 2\epsilon_t$, where the second inequality holds under $\mathcal{G}$. Thus $\theta \in \mathcal{B}(\theta_t, 2\epsilon_t)$ and so $\mathcal{B}(\theta^*, \epsilon_t) \subset \mathcal{B}(\theta_t, 2\epsilon_t) \subset \Delta_t$. We prove $\mathcal{B}(\theta^*, \epsilon_t) \subset \Theta_{t-1}$ by induction. Assume that $\mathcal{B}(\theta^*, \epsilon_t) \subset \Theta_{t-1}$. Then we show that $\mathcal{B}(\theta^*, \epsilon_{t+1}) \subset \Theta_t$. Let θ such that $\|\theta - \theta^\star\| \leq \epsilon_{t+1}$. Then $\|\theta - \theta_t\| \leq \|\theta - \theta^*\| + \|\theta_t - \theta^*\| \leq \epsilon_{t+1} + \epsilon_t$. This implies that $\theta \in \Delta_t$. In addition, $\theta \in \Theta_{t-1}$. Indeed, $\|\theta - \theta^\star\| \leq \epsilon_{t+1} \leq \epsilon_t$, and we conclude using the induction assumption. Hence $\theta \in \Theta_t$. □

3.2. Polytopic tubes and associated linear constraints

With the considered control inputs, the system can be rewriten as

$$x_{t+1} = \Phi(\theta^\star)x_t + B(\theta^\star)(v_{0|t} + \zeta_t) + w_t.$$

We apply a tube-based approach, and at time t, we build a polytopic tube based on:

(i) Θ_t, encoding the uncertainty about $\theta^\star$. We denote by m the number of vertices of Θ_t, and the vertices themselves by $\theta_t^{(j)}$, $j = 1, \ldots, m$.

(ii) A polytope $\bar{\mathcal{W}}_t$, handling the uncertainty due to the *noise* $B(\theta^\star)\zeta_t + w_t$, including that due to the persistent excitation. To define $\bar{\mathcal{W}}_t$, let $\bar{B}_t = \max_{\theta \in \Theta_t} \|B(\theta)\|_2$. We define $\mathcal{Z}_t$ as the outer polyhedral approximation to $\mathcal{B}(0, 3\sigma_t)$, and $\bar{\mathcal{W}}_t = \mathcal{W} \oplus \bar{B}_t \mathcal{Z}_t$.

Algorithm 1 STT-MPC

Input: Initial state x_0; confidence δ; estimate θ_0; uncertain parameter set Θ_0
Find stabilizing matrix K for all $\theta \in \Theta_0$
Compute matrices T and H_c
for $t = 1, \ldots, \mathcal{T}$ **do**
  if $t < t^\star(\delta)$ **then**
    Set $\theta_t \leftarrow \theta_{t-1}$ and $\Theta_t \leftarrow \Theta_{t-1}$
  else
    Update $\theta_t \leftarrow \hat{\theta}_t$
    Compute Δ_t and set $\Theta_t \leftarrow \Theta_{t-1} \cap \Delta_t$
  end if
  Compute matrices $H_t^{(j)}$
  Solve $\mathcal{P}_N(x_t, \theta_{\rho(t)})$ and obtain $v_{0|t}^\pi$
  Apply $u_t = Kx_t + v_{0|t}^\pi + \zeta_t$
end for

We define the state tube cross sections as the sets: for $k = 0, \ldots, N$,

$$X_{k|t} = \{x : Tx \leq \alpha_{k|t}\},$$

where the matrix $T \in \mathbb{R}^{d_\alpha \times d_x}$ is chosen such that, for some $\lambda \in [0, 1)$, the set $\{x : Tx \leq \mathbf{1}\}$ is λ-contractive with respect to the system $x_{t+1} = \Phi(\theta)x_t$ for all $\theta \in \Theta_0$. This property is needed to ensure the robust positive invariance of $X_{N|t}$ (see Lemma 5.7 in (Kouvaritakis & Cannon, 2016)). To derive the associated linear constraints, we apply a standard result to ensure inclusion of polyhedra (see Proposition 3.31 in (Blanchini & Miani, 2008)). More precisely, for any $j = 1, \ldots, m$, we have $X_{k|t} \subseteq \{x : \Phi(\theta_t^{(j)})x + B(\theta_t^{(j)})v_{k|t} + w \in X_{k+1|t}\}$ for all $w \in \bar{\mathcal{W}}_t$ if there exists $H_t^{(j)} \geq 0$ such that:

$$H_t^{(j)}T = T\Phi(\theta_t^{(j)}),$$
$$H_t^{(j)}\alpha_{k|t} + TB(\theta_t^{(j)})v_{k|t} + \bar{w}_t \leq \alpha_{k+1|t},$$

where $\bar{w}_t$ is such that $[\bar{w}_t]_i = \max_{w \in \bar{\mathcal{W}}_t} [Tw]_i$ for $i \in \{1, \ldots, d_\alpha\}$. Similarly, we have $X_{k|t} \subseteq \{x : (F + GK)x + G(v_{k|t} + \zeta) \leq \mathbf{1}\}$ for all $\zeta \in \mathcal{Z}_t$, if there exists $H_c \geq 0$, such that:

$$H_c T = F + GK$$
$$H_c \alpha_{k|t} + Gv_{k|t} + \bar{\zeta}_t \leq \mathbf{1},$$

where $\bar{\zeta}_t$ is such that $[\bar{\zeta}_t]_i = \max_{\zeta \in \mathcal{Z}_t} [G\zeta]_i$ for $i \in \{1, \ldots, d_c\}$.

Remark 3.2. For simplicity, we have chosen $\bar{w}_t$ and $\bar{\zeta}_t$ so as to be conservative with respect to the noise we apply. However, notice that we could well pre-sample the noise sequence $(\zeta_t)_{t\geq 0}$ and use it as part of our predictions.

The matrices $H_t^{(j)}$, and H_c are chosen such that for all $i \in \{1, \dots, d_\alpha\}$ and for all $j \in \{1, \dots, m\}$,

$$\left(H_t^{(j)}\right)_i = \arg\min_h \{\mathbf{1}^\top h : h^\top T = (T)_i \Phi(\theta_t^{(j)}),\ h \geq 0\},$$

and for all $i \in \{1, \dots, d_c\}$,

$$\left(H_c\right)_i = \arg\min_h \{\mathbf{1}^\top h : h^\top T = (F + GK)_i,\ h \geq 0\}.$$

Finally, we have the terminal conditions

$$H_t^{(j)} \alpha_N + \bar{w}_t \leq \alpha_N,$$
$$H_c \alpha_N + \bar{\zeta}_t \leq \mathbf{1}.$$

3.3. The tube MPC problem

Let $\boldsymbol{v}_t = \{v_{0|t}, \dots, v_{N-1|t}\}$ and $\boldsymbol{\alpha}_t = \{\alpha_{0|t}, \dots, \alpha_{N|t}\}$. The resulting tube MPC problem, denoted as $\mathcal{P}_N(x_t, \theta)$ for any parameter θ is:

$$\underset{\boldsymbol{v}_t, \boldsymbol{\alpha}_t}{\text{minimize}} \sum_{k=0}^{N-1} \left(x_{k|t}^\top Q x_{k|t} + v_{k|t}^\top R v_{k|t} \right) + x_{N|t}^\top P(\theta) x_{N|t}$$

subject to, for all $j = 1, \dots, m$, and $k = 0, \dots, N-1$:

initial constraints:

$$x_{0|t} = x_t, \tag{6}$$
$$T x_{0|t} \leq \alpha_{0|t}, \tag{7}$$

system constraints,

$$x_{k+1|t} = \Phi(\theta) x_{k|t} + B(\theta) v_{k|t}, \tag{8}$$

tube constraints,

$$H_t^{(j)} \alpha_{k|t} + T B(\theta^{(j)}) v_{k|t} + \bar{w}_t \leq \alpha_{k+1|t}, \tag{9}$$
$$H_c \alpha_{k|t} + G v_{k|t} + \bar{\zeta}_t \leq \mathbf{1}, \tag{10}$$

terminal constraints:

$$H_t^{(j)} \alpha_{N|t} + \bar{w}_t \leq \alpha_{N|t}, \tag{11}$$
$$H_c \alpha_{N|t} + \bar{\zeta}_t \leq \mathbf{1}. \tag{12}$$

where $P(\theta)$ is obtained by solving the Lyapunov equation:

$$P(\theta) - \Phi(\theta)^\top P(\theta) \Phi(\theta) = Q + K^\top R K \tag{13}$$

Remark 3.3. The oracle tube-MPC problem $\mathcal{P}_N(x_t, \theta^\star)$ is similarly defined by setting $\theta = \theta^\star$, $\bar{w}_t = \bar{w}_t^\star$ (with $[\bar{w}_t^\star]_i = \max_{w \in \mathcal{W}} [Tw]_i$, for $i \in \{1, \dots, d_\alpha\}$), and $m = 1$. The later can be achieved by Algorithm 1 by removing redundant constraints at every iteration by solving a linear program or one of the methods outlined in (Paulraj et al., 2010).

The feasibility of the above problem depends on whether $\theta^\star \in \Theta_t$, i.e., on the event $\mathcal{G}$. However, for the sake of the analysis, we would like to ensure that the problem the algorithm solves in each step is always feasible (with probability 1). This is necessary for establishing that $\mathcal{G}$ holds with probability at least $1-\delta$ (Theorem 1 of (Tranos et al., 2022)) as well as for our regret analysis. To ensure that the tube MPC problem is always feasible, i.e., even when $\mathcal{G}$ does not occur, we instead solve $\mathcal{P}_N(x_t, \theta_{\rho(t)})$[1] in Algorithm 1 where,

$$\rho(t) := \max\{\tau \leq t : \mathcal{P}_N(x_t, \theta_\tau) \text{ is feasible}\}.$$

Essentially, in the unlikely event that $\mathcal{P}_N(x_t, \theta_t)$ is not feasible, we instead solve $\mathcal{P}_N(x_t, \theta_\tau)$ using the latest estimates θ_τ and Θ_τ for which the problem is feasible. This modification is of little practical consequence as we can (and typically do) choose δ to be very small.

With this, we may state the recursive feasibility property of STT-MPC (Theorem 2 in (Tranos et al., 2022)):

Theorem 3.4. *If the optimization problem $\mathcal{P}_N(x_0, \theta_0)$ is feasible for initial state $x_0 \in \mathbb{R}^{d_x}$ and parameter $\theta_0 \in \Theta_0$, then, for all $t > 0$,*
(i) under event $\mathcal{G}$, the problem $\mathcal{P}_N(x_t, \theta_t)$ is feasible;
(ii) the problem $\mathcal{P}_N(x_t, \theta_{\rho(t)})$ is feasible.

4. Regret Analysis

In this section, we present our main result, an upper bound on the expected regret of STT-MPC. To state our main theorem, we introduce the following notation. Let Π be the set of all tube-MPC algorithms (which are parameterized by θ). For an algorithm $\pi \in \Pi$, define $(x_0^\pi, u_0^\pi, \dots, x_t^\pi, u_t^\pi)$ as the sequence of states and control inputs generated under π, with $x_0^\pi = x_0$. We also define by $v_t^\pi(x, \theta) := v_{0|t}^\pi(x, \theta)$, the first element of the solution of the optimization problem $\mathcal{P}_N(x, \theta)$. We denote by $z_t^\pi(x)$ the state of the system (1) at time t under algorithm π given initial state x, and we denote by $y_t^\pi(x)$, the associated output of π. Let $\pi_\star \in \Pi$ be the oracle tube-MPC problem $\mathcal{P}_N(x, \theta^\star)$. We simplify the notation by replacing $\pi_\star$ by $\star$, e.g., $x_t^\star$, $v_t^\star$ and so on. We define the regret $R_T(\pi)$ of an algorithm π as:

$$R_T(\pi) = \sum_{t=0}^{T-1} (\|x_t^\pi\|_Q^2 + \|u_t^\pi\|_R^2) - \sum_{t=0}^{T-1} (\|x_t^\star\|_Q^2 + \|u_t^\star\|_R^2).$$

Theorem 4.1. *The regret of π = STT-MPC with confidence δ and persistent excitation parameter $\alpha := 2\epsilon$, with $\epsilon \in (0, 1)$ satisfies for all $T \geq 1$, with probability at least $1 - \delta$,*

$$R_T(\pi) \leq C \left((\log(T/\delta)^{1/2} + 1) T^{1/2+\epsilon} + \log(1/\delta) + 1 \right),$$

[1] More precisely, the problem is obtained by replacing θ_t by $\theta_{\rho(t)}$, $H_t^{(j)}$ by $H_{\rho(t)}^{(j)}$ for all j in $\mathcal{P}_N(x_t, \theta_t)$.

for some universal constant $C > 0$. Now when the confidence is set to $1/T^2$, we get

$$\mathbb{E}[R_T(\pi) \leq C' \log(T)^{1/2} T^{1/2+\epsilon},$$

where C' is a universal constant.

Proof. We first express the regret as the sum of differences between the states and inputs,

$$\begin{aligned} R_T(\pi) = &\sum_{t=0}^{T-1} \left({x_t^\pi}^\top Q(x_t^\pi - x_t^\star) + (x_t^\pi - x_t^\star)^\top Q x_t^\star \right) \\ &+ \sum_{t=0}^{T-1} \left({u_t^\pi}^\top R(u_t^\pi - u_t^\star) + (u_t^\pi - u_t^\star)^\top R u_t^\star \right) \\ \leq &\sum_{t=0}^{T-1} (\|x_t^\pi\| + \|x_t^\star\|)\|Q\|\|x_t^\pi - x_t^\star\| \\ &+ \sum_{t=0}^{T-1} (\|u_t^\pi\| + \|u_t^\star\|)\|R\|\|u_t^\pi - u_t^\star\| \\ = &\bar{Q} \sum_{t=0}^{T-1} \|x_t^\pi - x_t^\star\| + \bar{R} \sum_{t=0}^{T-1} \|u_t^\pi - u_t^\star\|, \end{aligned}$$

where the first inequality follows from Cauchy-Schwartz inequality, and the second by defining $\bar{Q} := 2\|Q\|\bar{x}$ and $\bar{R} := 2\|R\|\bar{u}$ with $\bar{x} = \max_{(x,u)\in\mathcal{C}}\{\|x\|\}$ and $\bar{u} = \max_{(x,u)\in\mathcal{C}}\{\|u\|\}$. Note that $\bar{Q}$ and $\bar{R}$ are well defined since the constraint set $\mathcal{C}$ is compact by Assumption 2.3. Further note that $(x_t, u_t) \in \mathcal{C}$ is guaranteed by the recursive feasibility of STT-MPC (Theorem 3.4).

We seek to upper bound the quantities $\|x_t^\pi - x_t^\star\|$ and $\|u_t^\pi - u_t^\star\|$ in terms of $\|\theta_t - \theta^\star\|$. To this end, we make use of the following theorem, presented originally in (Shin et al., 2022) (Theorem 4.5) for graph-structured nonlinear optimization problems, and specialized to constrained nonlinear systems by (Lin et al., 2022) (Theorem H.1.).

Theorem 4.2. *For any tube-MPC algorithm $\pi \in \Pi$, there exists $R > 0$ such that, for $x_i, x_i' \in B(x_i^\star, R)$, where $i \in \{0, \dots, t\}$, the following perturbation bounds hold:*

$$\|z_t^\pi(x_i, \theta) - z_t^\pi(x_i', \theta)\| \leq q_1(t-i)\|x_i - x_i'\|, \tag{14}$$

and

$$\|v_t^\pi(x_i, \theta) - v_t^\pi(x_i, \theta')\| \leq q_2(t-i)\|\theta - \theta'\|, \tag{15}$$

where, for $j \in \{1, 2\}$, $q_j(t)$ is a function such that $\lim_{t\to\infty} \lim q_j(t) = 0$ and $\sum_{t=0}^{\infty} q_j(t) \leq C_j$, for some constant $C_j \geq 1$.

The bound (15) can be applied directly to the input error (along the trajectory induced by π) $e_t := \|v_t^\pi - v_t^\star(x_t^\pi)\|$, provided that $x_t^\pi \in \mathcal{B}(x_t^\star, R)$. Therefore, we show inductively that if $e_t \leq R/(\|B(\theta^\star)\|C_1^2)$ then $x_t^\pi \in \mathcal{B}(x_t^\star, R/C_1)$. First, observe that

$$\begin{aligned} \|x_t^\pi - z_t^\star(x_{t-1}^\pi)\| &= \|B(\theta^\star)(v_{t-1}^\pi - v_{t-1}^\star(x_{t-1}^\pi))\| \\ &\leq \|B(\theta^\star)\| e_{t-1}, \end{aligned} \tag{16}$$

and so the condition holds for the base case of $t = 0$, since $x_0^\pi = x_0^\star$. Let our induction hypothesis be that it holds for $0, \dots, t-1$. We will have

$$\begin{aligned} &\|x_t^\pi - x_t^\star\| \\ &= \|x_t^\pi - z_t^\star(x_0)\| \\ &\leq \|x_t^\pi - z_t^\star(x_{t-1})\| + \sum_{i=1}^{t-1} \|z_t^\star(x_{t-i}) - z_t^\star(x_{t-i-1})\| \\ &= \|x_t^\pi - z_t^\star(x_{t-1})\| + \sum_{i=1}^{t-1} \|z_t^\star(x_{t-i}) - z_t^\star(z_{t-1}^\star(x_{t-i-1}))\| \\ &\leq \|x_t^\pi - z_t^\star(x_{t-1})\| + \sum_{i=1}^{t-1} q(i)\|x_{t-i} - z_{t-i}^\star(x_{t-i-1})\| \\ &\leq \sum_{i=0}^{t-1} q_1(i)\|x_{t-i} - z_{t-i}^\star(x_{t-i-1})\| \\ &\leq \|B(\theta^\star)\| \sum_{i=0}^{t-1} q_1(t) e_{t-i-1}, \end{aligned}$$

where the first inequality is a straightforward application of the triangle inequality, the second equality follows by the definition of $z_t^\star$, and the third inequality is a direct application of (14) (noting that $x_{t-i}^\pi \in \mathcal{B}(x_{t-i}^\star, R/C_1)$ by the induction hypothesis). The last inequality follows from (16).

Now if we substitute $e_{t-i} \leq R/(\|B(\theta^\star)\|C_1^2)$, we obtain

$$\|x_t^\pi - x_t^\star\| \leq \|B(\theta^\star)\| \frac{R}{\|B(\theta^\star)\|C_1^2} \sum_{i=0}^{t-1} q_1(t) \leq \frac{R}{C_1},$$

which proves the induction step.

Proceeding similarly, we obtain the following bound,

$$\begin{aligned} &\|v_t^\pi - v_t^\star\| \\ &= \|v_t^\pi - y_t^\star(x_0)\| \\ &\leq \|v_t^\pi - y_t^\star(x_{t-1})\| + \sum_{i=0}^{t-1} \|v_t^\star(x_{t-i}) - y_t^\star(x_{t-i-1})\| \\ &\leq \|v_t^\pi - v_t^\star(x_{t-1})\| + \sum_{i=0}^{t-1} q_1(i)\|x_{t-i} - z_{t-i}^\star(x_{t-i-1})\| \\ &\leq e_t + \|B(\theta^\star)\| \sum_{i=0}^{t-1} q_1(t) e_{t-i-1}, \end{aligned}$$

where again, we have used the triangle inequality for the first inequality, and the bound (14) for the second. Finally, we have

$$\begin{aligned}\|u_t^\pi - u_t^\star\| &= \|v_t^\pi + \zeta_t - v_t^\star\| \\ &\leq \|v_t^\pi - v_t^\star\| + \|\zeta_t\|.\end{aligned}$$

With these bounds in hand, we first account for the regret up to time $t^\star(\delta)$ while the event $\mathcal{G}$ holds. Note that a crude bound is sufficient to show that it is finite:

$$\begin{aligned}R_{t^\star(\delta)}(\pi)_\mathcal{G} &\leq \bar{Q}\sum_{t=0}^{t^\star(\delta)}\|x_t^\pi - x_t^\star\| + \bar{R}\sum_{t=0}^{t^\star(\delta)}\|u_t^\pi - u_t^\star\| \\ &\leq (2\bar{Q}\bar{x} + 2\bar{R}\bar{u})t^\star(\delta) \\ &\leq \bar{C}_2 + \bar{C}_3\log(1/\delta).\end{aligned}$$

Where $\bar{C}_2 = (2\bar{Q}\bar{x} + 2\bar{R}\bar{u})c_2$ and $\bar{C}_3 = (2\bar{Q}\bar{x} + 2\bar{R}\bar{u})c_3$.

We then bound the regret from $t^\star(\delta)$ onward under the event $\mathcal{G}$:

$$\begin{aligned}&(R_T(\pi) - R_{t^\star(\delta)}(\pi))_\mathcal{G} \\ &\leq \bar{Q}\sum_{t=t^\star(\delta)}^{T-1}\|x_t^\pi - x_t^\star\| + \bar{R}\sum_{t=t^\star(\delta)}^{T-1}\|u_t^\pi - u_t^\star\| \\ &\leq \bar{Q}\sum_{t=0}^{T-1}\|x_t^\pi - x_t^\star\| + \bar{R}\sum_{t=0}^{T-1}\|v_t^\pi - v_t^\star\| + \|\zeta_t\| \\ &\leq \bar{Q}\sum_{t=0}^{T-1}\|B(\theta^\star)\|e_{t-1} \\ &+ \bar{R}\sum_{t=0}^{T-1}(e_t + \|B(\theta^\star)\|\sum_{i=0}^{t-1}q_1(t)e_{t-i-1}) + 3\sigma_t \\ &\leq (\bar{Q} + \bar{R}C_1)\|B(\theta^\star)\|\sum_{t=0}^{T-1}e_t + 3\sigma_t \\ &\leq (\bar{Q} + \bar{R}C_1)\|B(\theta^\star)\|\sum_{t=0}^{T-1}q_2(0)\|\theta_t - \theta^\star\| + 3\sigma_t \\ &\leq \bar{C}\sum_{t=0}^{T-1}\epsilon_t + 3\sigma_t \\ &\leq \bar{C}\sqrt{c_3}\sum_{t=0}^{T-1}\log(t/\delta)^{1/2}t^{-(1-\alpha)/2} + 3d_x^{1/4}\sigma\sum_{t=0}^{T-1}t^{-\alpha/2} \\ &\leq \frac{2\bar{C}\sqrt{c_3}}{\alpha+1}\log(T/\delta)^{1/2}T^{(\alpha+1)/2} + \frac{6d_x^{1/4}}{\alpha+1}\sigma T^{(\alpha+1)/2},\end{aligned}$$

with $\bar{C} = (\bar{Q} + \bar{R}C_1)\|B(\theta^\star)\|C_1$.

Putting it together, we have under the event $\mathcal{G}$:

$$\begin{aligned}R_T(\pi)_\mathcal{G} =&(R_T(\pi) - R_{t^\star(\delta)}(\pi))_\mathcal{G} + R_{t^\star(\delta)}(\pi)_\mathcal{G} \\ \leq&\frac{2\bar{C}\sqrt{c_3}}{\alpha+1}\log(T/\delta)^{1/2}T^{(\alpha+1)/2} \\ &+ \frac{2d_x^{1/4}}{\alpha+1}\sigma T^{(\alpha+1)/2} + \bar{C}_2\log(1/\delta) + \bar{C}_1.\end{aligned}$$

To obtain a bound on the expected regret, first note that

$$\mathbb{E}[R_T(\pi)] = (1-\delta)R_T(\pi)_\mathcal{G} + \delta R_T(\pi)_{\neg\mathcal{G}}$$

where $R_T(\pi)_{\neg\mathcal{G}}$ is the regret when the event $\mathcal{G}$ does not hold and can be taken to be linear in time with some constant c. This follows again from the recursive feasibility of STT-MPC (Theorem 3.4), which ensures that x_t^π and u_t^π are bounded in the set $\mathcal{C}$. Letting $\delta = 1/T^2$ leads to

$$\begin{aligned}\mathbb{E}[R_T(\pi)] &= (1-\delta)R_T(\pi)_\mathcal{G} + \delta R_T(\pi)_{\neg\mathcal{G}} \\ &= R_T(\pi)_\mathcal{G} + \frac{1}{T^2}R_T(\pi)_\mathcal{G} + \frac{c}{T} \\ &\lesssim \frac{2\sqrt{3}\bar{C}\sqrt{c_3}}{\alpha+1}\log(T)^{1/2}T^{(\alpha+1)/2}.\end{aligned}$$

Here, the notation $\lesssim$ hides the universal constant such that the inequality holds. □

5. Numerical Example

We illustrate the performance of STT-MPC using the following second-order linear system:

$$A = \begin{bmatrix}0.6 & 0.2\\ -0.1 & 0.4\end{bmatrix}, \quad B = \begin{bmatrix}1\\ 0.6\end{bmatrix}, \quad \sigma = 0.01, \tag{17}$$

The initial state is $x_0 = (6, 3)$, and we consider Θ_0 to be a 6-dimensional hypercube centered on $\theta_0 = \begin{bmatrix}0.57 & 0.17 & -0.12 & 0.42 & 0.95 & 0.65\end{bmatrix}$ with side length 0.14. Consequently, the resulting stabilizing feedback gain is $K = \begin{bmatrix}-0.426 & -0.290\end{bmatrix}$. We consider the following state and input constraints: $[x_t]_1 \geq -0.15, [x_t]_2 \geq -1.1$ and $u_t \leq 0.5$. We inject a persistent excitation signal with a standard deviation of $\sigma_t = \sqrt{2}\sigma(t+1)^{-\alpha}$.

The matrix T is computed according to the relation (5.98) in (Kouvaritakis & Cannon, 2016), with $\lambda = 0.999$. We consider $Q = I_{d_x}$, $R = I_{d_u}$, and, for simplicity, we let the worst case noise realization be $\bar{w}_t = \bar{w}_0$, for all $t \geq 1$. All the simulations were performed in Python 3.9, using the CVXPY library (Diamond & Boyd, 2016) and the MOSEK solver.

In Figure 1, we present the regret of STT-MPC versus the oracle tube-MPC algorithm the cases where α is $0.01, 0.5$, and 0.99. For all three cases, the scaling is logarithmic and is thus over-estimated by the bound of Theorem 4.1. Note

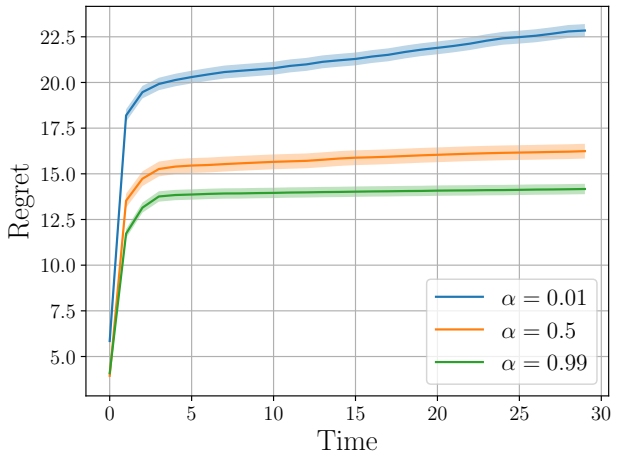

Figure 1. Regret vs. time of STT-MPC averaged over 100 runs (the shaded area corresponds to the standard error of the mean). Each curve corresponds to a different decay rate of the persistent excitation signal ζ_t.

the lower the value of α, the slower the decay rate of the excitation, which improves the transient estimation accuracy but negatively impacts the performance of the controller. The result suggests that the LSE converges very quickly to the true parameter (already before $t = 5$) even for high values of α, and so the performance of the controller (rather, the decay rate of the excitation signal) dominates the regret.

Overall the logarithmic scaling is surprising as it is better than the $\sqrt{T}$ bound shown both theoretically and experimentally by (Jedra & Proutiere, 2022) in the case of the LQR. We conjecture that the presence of constraints has a benign property on the regret and that it should be possible to tighten our upper bound to also scale as $\log(T)$.

6. Conclusions

We proposed STT-MPC which combines least-squares estimaton with a polytopic tube-based MPC method to ensure robust constraint satisfaction while learning the system dynamics. Persistent excitation is ensured by injecting a truncated noise signal which decays at a rate $t^{-\alpha}$, with $\alpha > 0$ controlling the trade-off between (transient) estimation accuracy and controller performance. Importantly, we asymptotically recover the performance of the oracle tube-based MPC which has full knowledge of the dynamics.

We provided guarantees on the expected regret of our proposed algorithm by leveraging performance bounds on the LSE, the exponential decay of sensitivity property of the optimization problem, and the recursive feasibility of our algorithm. We show theoretically that the expected regret of STT-MPC scales at a rate of $T^{1/2+\epsilon}$ with $\epsilon \in (0, 1)$ and also demonstrated its performance via a numerical example.

The logarithmic rate demonstrated in simulation suggests that our bound can be further tightened, and we will investigate this further in future work.

References

Abbasi-Yadkori, Y. and Szepesvári, C. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pp. 1–26. JMLR Workshop and Conference Proceedings, 2011.

Åström, K. J. and Wittenmark, B. On self tuning regulators. *Automatica*, 9(2):185–199, 1973.

Bitmead, R. R., Gevers, M., and Wertz, V. *Adaptive optimal control the thinking man's GPC*. Prentice Hall, 1990.

Blanchini, F. and Miani, S. *Set-theoretic methods in control*, volume 78. Springer, 2008.

Diamond, S. and Boyd, S. Cvxpy: A python-embedded modeling language for convex optimization. *The Journal of Machine Learning Research*, 17(1):2909–2913, 2016.

Goel, G. and Hassibi, B. The power of linear controllers in lqr control. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 6652–6657. IEEE, 2022.

Jedra, Y. and Proutiere, A. Minimal expected regret in linear quadratic control. In *International Conference on Artificial Intelligence and Statistics*, pp. 10234–10321. PMLR, 2022.

Kouvaritakis, B. and Cannon, M. Model predictive control. *Switzerland: Springer International Publishing*, 38, 2016.

Lai, T. L. Asymptotically efficient adaptive control in stochastic regression models. *Advances in Applied Mathematics*, 7(1):23–45, 1986. ISSN 0196-8858.

Li, Y., Chen, X., and Li, N. Online optimal control with linear dynamics and predictions: Algorithms and regret analysis. *Advances in Neural Information Processing Systems*, 32, 2019.

Lin, Y., Hu, Y., Shi, G., Sun, H., Qu, G., and Wierman, A. Perturbation-based regret analysis of predictive control in linear time varying systems. *Advances in Neural Information Processing Systems*, 34:5174–5185, 2021.

Lin, Y., Hu, Y., Qu, G., Li, T., and Wierman, A. Bounded-regret mpc via perturbation analysis: Prediction error, constraints, and nonlinearity. *arXiv preprint arXiv:2210.12312*, 2022.

Lorenzen, M., Allgöwer, F., and Cannon, M. Adaptive model predictive control with robust constraint satisfaction. *IFAC-PapersOnLine*, 50(1):3313–3318, 2017.

Lu, X. and Cannon, M. Robust adaptive tube model predictive control. In *2019 American Control Conference (ACC)*, pp. 3695–3701. IEEE, 2019.

Lu, X., Cannon, M., and Koksal-Rivet, D. Robust adaptive model predictive control: Performance and parameter estimation. *International Journal of Robust and Nonlinear Control*, 31(18):8703–8724, 2021.

Mania, H., Tu, S., and Recht, B. Certainty equivalence is efficient for linear quadratic control. *Advances in Neural Information Processing Systems*, 32, 2019.

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

Mayne, D. Q., Seron, M. M., and Raković, S. Robust model predictive control of constrained linear systems with bounded disturbances. *Automatica*, 41(2):219–224, 2005.

Paulraj, S., Sumathi, P., et al. A comparative study of redundant constraints identification methods in linear programming problems. *Mathematical Problems in Engineering*, 2010, 2010.

Recht, B. A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:253–279, 2019.

Scokaert, P. O. and Rawlings, J. B. Constrained linear quadratic regulation. *IEEE Transactions on automatic control*, 43(8):1163–1169, 1998.

Shin, S., Anitescu, M., and Zavala, V. M. Exponential decay of sensitivity in graph-structured nonlinear programs. *SIAM Journal on Optimization*, 32(2):1156–1183, 2022.

Tranos, D., Russo, A., and Proutiere, A. Self-tuning tube-based model predictive control. *arXiv preprint arXiv:2210.00502*, 2022.

Yu, C., Shi, G., Chung, S.-J., Yue, Y., and Wierman, A. The power of predictions in online control. *Advances in Neural Information Processing Systems*, 33:1994–2004, 2020.

Zhang, R., Li, Y., and Li, N. On the regret analysis of online lqr control with predictions. In *2021 American Control Conference (ACC)*, pp. 697–703. IEEE, 2021.