Information Aggregation for Constrained Online Control

Tongxin Li
California Institute of Technology
tongxin@caltech.edu

Yue Chen
Tsinghua University
cy11@tsinghua.org.cn

Bo Sun
Hong Kong University of Science and Technology
bsunaa@connect.ust.hk

Adam Wierman
California Institute of Technology
adamw@caltech.edu

Steven Low
California Institute of Technology
slow@caltech.edu

ABSTRACT

We consider a two-controller online control problem where a central controller chooses an action from a feasible set that is determined by time-varying and coupling constraints, which depend on all past actions and states. The central controller's goal is to minimize the cumulative cost; however, the controller has access to neither the feasible set nor the dynamics directly, which are determined by a remote local controller. Instead, the central controller receives only an aggregate summary of the feasibility information from the local controller, which does not know the system costs. We show that it is possible for an online algorithm using feasibility information to nearly match the dynamic regret of an online algorithm using perfect information whenever the feasible sets satisfy a causal invariance criterion and there is a sufficiently large prediction window size. To do so, we use a form of feasibility aggregation based on entropic maximization in combination with a novel online algorithm, named Penalized Predictive Control (PPC).

KEYWORDS

online control; closed-loop control; model predictive control; regret analysis; electric vehicle charging

ACM Reference Format:
Tongxin Li, Yue Chen, Bo Sun, Adam Wierman, and Steven Low. 2021. Information Aggregation for Constrained Online Control. In *Abstract Proceedings of the 2021 ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '21 Abstracts), June 14–18, 2021, Virtual Event, China.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3410220.3461737

1 PROBLEM STATEMENT

We consider a general dynamic model over a discrete time horizon $[T] := \{1, \ldots, T\}$ with *time-varying* and *time-coupling* constraints:

$$x_{t+1} = f_t(x_t, u_t), \quad x_t \in \mathcal{X}_t\left(\mathbf{x}_{<t}, \mathbf{u}_{<t}\right),\ u_t \in \mathcal{U}_t\left(\mathbf{x}_{<t}, \mathbf{u}_{<t}\right), \quad (1)$$

where the deterministic function f_t represents the transition of the state x_t and it satisfies the following assumption:

SIGMETRICS '21 Abstracts, June 14–18, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8072-0/21/06.
https://doi.org/10.1145/3410220.3461737

Assumption 1. *The dynamic $f_t(\cdot,\cdot) : \mathcal{X}_t \times \mathcal{U}_t \to \mathcal{X}_{t+1}$ is a Borel measurable function for $t \in [T]$.*

The dynamical system is governed by a *local controller*, which manages a large fleet of controllable units. The collection of the states of the units is represented by x_t in a state space $\mathsf{X} \subseteq \mathbb{R}^n$. Both the state and action at each time are confined by *safety sets* that maybe time-varying and time-coupling, i.e., $x_t \in \mathcal{X}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t})$ and $u_t \in \mathcal{U}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t})$ for $t \in [T]$. For simplicity, we denote them by $\mathcal{U}_t$ and $\mathcal{X}_t$ in future contexts.

There is a distant *central controller* that communicates with the local controller. The central controller selects an action u_t at each time $t \in [T]$. The actions must be selected from a closed and bounded domain $\mathsf{U} \subseteq \mathbb{R}^m$. The initial point u_0 is assumed to be the origin without loss of generality. The central controller receives time-varying cost functions c_t online from an external environment and each $c_t(\cdot) : \mathsf{U} \to \mathbb{R}_+$ only depends on the action u_t chosen by the central controller. We assume that the local controller does not know the costs and has to choose the action given by the central controller and the central controller cannot access the constraints directly, but some information about the constraints, summarized as a density function $p_t(\cdot) : \mathsf{U} \to [0, 1]$, can be transmitted during the control. The goal of an online control policy in this setting is to make the local and central controllers jointly minimize a cumulative cost $C_T(\mathbf{u}) := \sum_{t=1}^{T} c_t(u_t)$ while satisfying (1). We suppose the online controller has (perfect) predictions of the cost functions and feedback of the current and the next w time slots. Throughout this paper, we make the following regularity and smoothness assumptions on the model.

Assumption 2. *The safety sets $\{\mathcal{U}_t : t \in [T]\}$ and $\{\mathcal{X}_t : t \in [T]\}$ are Borel sets in $\mathbb{R}^m$ and $\mathbb{R}^n$. Furthermore, the safety sets are atoms, i.e., $\mu(\mathcal{X}_t) > 0$ and $\mu(\mathcal{U}_t) > 0$ for all $t \in [T]$ if $\mathcal{X}_t, \mathcal{U}_t \neq \emptyset$.*

Assumption 3. *For each $t \in [T]$, the cost function $c_t(\cdot) : \mathsf{U} \to \mathbb{R}_+$ is Lipschitz continuous. We assume that there exists a Lipschitz constant $L_c > 0$ such that $|c_t(u) - c_t(v)| \leq L_c ||u - v||_2$ for all $u, v \in \mathsf{U}$ and $t \in [T]$.*

Our work is motivated by settings where a local controller governs a large-scale system and a central controller operates remotely. In many situations, full information about the local controllers' dynamics and constraints is not available to the central controller, due to complexity or privacy concerns and the local controller cannot access the system's costs. In such a two-controller system, the central and local controllers each have part of the information needed to control the whole dynamical system online. The task

of designing algorithms is therefore made even more challenging than the single controller case. Note that there are a wide variety of situations that face this challenge, including operator-aggregator coordination in smart grid [2], data center scheduling [3] and fog computing [1].

2 ALGORITHM AND MAIN RESULTS

Our proposed design, termed Penalized Predictive Control (PPC), is a combination of Model Predictive Control (MPC), which is a competitive policy for online optimization with predictions, and the idea of using feasibility information about the safety sets as a penalty term. This design makes a connection between *maximum entropy feedback* (MEF) [2], a special design of p_t satisfying

$$p_t(u|\mathbf{u}_{<t}) = \frac{\mu\left(\mathsf{S}((\mathbf{u}_{<t}, u))\right)}{\mu\left(\mathsf{S}(\mathbf{u}_{<t})\right)}, \quad \forall u \in \mathsf{U} \text{ and } \mathbf{u}_{<t} \in \mathsf{U}^{t-1} \tag{2}$$

where $\mu(\cdot)$ denotes the Lebesgue measure and $\mathsf{S}(\mathbf{u}_{\le t})$ consists of all feasible $T-t$ actions at time $t+1,\dots,T$, given the past actions $\mathbf{u}_{\le t}$, and the well-known MPC scheme. The MEF as a feedback function, only contains feasibility information about the dynamical system in the local controller's side. We present PPC in Algorithm 1, where we use the following notation. Let $t' := \min\{t+w-1, T\}$ and let $\beta > 0$ be a *tuning parameter*. Define a set of time indices $\mathcal{I} := \{t \in [T] : t \equiv 1 \mod w\}$.

Algorithm 1: Closed-loop online control framework

for $t = 1,\dots,T$ **do**
  Central Controller
  if $t \in \mathcal{I}$ **then**
    Generate actions using the PPC:

Penalized Predictive Control

$$\mathbf{u}_{t:t'} = \arg\inf_{\mathbf{u}_{t:t'}} \sum_{\tau=t}^{t'} \left(c_\tau(u_\tau) - \beta \log p_\tau(u_\tau|\mathbf{u}_{<\tau})\right)$$

  end
  Local Controller
    Update system state and compute MEF: $x_{t+1} = f_t(x_t, u_t)$
end

The novel use of MEF as a penalty term in MPC allows PPC to achieve nearly optimal dynamic regret (defined in (3) below), despite having only feasibility information about constraints and dynamics, a setting where no prior algorithms have provable guarantees.

$$\mathsf{Regret}(\mathbf{u}) := \sup_{\mathbf{c}\in\mathsf{C}} \sup_{\mathbf{f}\in\mathsf{F}} \sup_{(\mathsf{U},\mathsf{X})\in\mathcal{I}} C_T(\mathbf{u}) - C_T^* \tag{3}$$

where $C_T^* := \inf_{\mathbf{u}} C_T(\mathbf{u})$ is the offline optimal cost subject to (1) for all $t \in [T]$ and $\mathbf{u}$ is the sequence of actions generated by the online policy π; $\mathbf{f} := (f_1,\dots,f_T)$ denotes a sequence of dynamics chosen from a set of Borel measurable functions F satisfying Assumption 1; $\mathbf{c} := (c_1,\dots,c_T)$ denotes a sequence of cost functions chosen from the set of all Lipschitz continuous functions C; $\mathsf{U} := (\mathcal{U}_1,\dots,\mathcal{U}_T)$ and $\mathsf{X} := (\mathcal{X}_1,\dots,\mathcal{X}_T)$ are the collections of safety constraints. It is important to note that without any restrictions on U and X, $\mathsf{Regret}(\mathbf{u})$ can be no better than $\Omega(T)$ for any deterministic online policy π, even with predictions, as the following theorem shows.

Theorem 2.1 (Fundamental limit). *Suppose $\mathcal{I}$ is a collection of safety sets satisfying Assumption 2. For any sequence of actions $\mathbf{u} \in \mathsf{S}$ generated by a deterministic online policy that has full information about the safety sets, $\mathsf{Regret}(\mathbf{u}) = \Omega\left(d(T-w)\right)$ for any $w \ge 1$, where $d := \mathsf{diam}(\mathsf{U}) := \sup\{||u-v||_2 : u, v \in \mathsf{U}\}$ is the diameter of the action space U, w is the prediction window size and T is the total number of time slots.*

Therefore, the focus of this paper is to find conditions on (U,X) so that given enough predictions, the regret can be bounded by a sub-linear (in T) function. This motivates the following *causal invariance criterion*. Let $\overline{\mathbf{u}}_{\le t} = (\overline{u}_1,\dots,\overline{u}_t)$ be a subsequence of optimal actions that maximizes the volume of the set of feasible actions, defined as $\overline{\mathbf{u}}_{\le t} := \arg\sup_{\mathbf{u}\in\mathsf{U}^t} \mu\left(\mathsf{S}(\mathbf{u})\right)$. Define the maximizing length-k subsequence of actions as $\overline{\mathbf{u}}_{t+1:t+k} := \arg\sup_{\mathbf{u}\in\mathsf{U}^k} \mu\left(\mathsf{S}_k\left(\mathbf{u}_{\le t}, \mathbf{u}\right)\right)$.

***Definition* 2.1** ((k,δ,λ)-causal invariance)**.** *The safety sets are (k,δ,λ)-causally invariant if there exist constants $\delta, \lambda > 0$ such that the following holds:*

(1) *For all $t \in [T]$ and sequences of actions $\mathbf{u}_{\le t}$ and $\mathbf{v}_{\le t}$,*

$$d_{\mathsf{H}}\left(\mathsf{S}_k(\mathbf{u}_{\le t}), \mathsf{S}_k(\mathbf{v}_{\le t})\right) \le \delta \left(\frac{|\mu(\mathsf{S}(\mathbf{u}_{\le t})) - \mu(\mathsf{S}(\mathbf{v}_{\le t}))|}{\mu\left(\mathcal{B}\right)}\right)^{1/((T-t)m)}$$

where $\mathcal{B}$ denotes the unit ball in $\mathbb{R}^{m\times(T-t)}$.

(2) *For all $t \in [T]$ and sequences of actions $\mathbf{u}_{\le t}$,*

$$\frac{\mu\left(\mathsf{S}\left(\mathbf{u}_{\le t}\right)\right)}{\mu\left(\mathsf{S}\left(\overline{\mathbf{u}}_{\le t}\right)\right)} \le \lambda \left(\frac{\mu\left(\mathsf{S}\left((\mathbf{u}_{\le t}, \overline{\mathbf{u}}_{t+1:t+k})\right)\right)}{\mu\left(\mathsf{S}\left(\overline{\mathbf{u}}_{\le t+k}\right)\right)}\right)^{\frac{T-t}{T-t-k}}.$$

We are now ready to present our main result, which bounds the dynamic regret by a decreasing function of the prediction window size under the assumption that the safety sets are causally invariant.

Theorem 2.2 (Dynamic Regret Bound). *Suppose the safety sets are (w,δ,λ)-causally invariant and suppose p_t satisfies (2) for all $t \in [T]$. The sequence of actions $\mathbf{u} = (u_1,\dots,u_T)$ generated by the PPC always satisfies $\mathbf{u} \in \mathsf{S}$. The dynamic regret for the sequence of actions $\mathbf{u}$ given by PPC is bounded from above by*

$$\mathsf{Regret}(\mathbf{u}) = O\left(dT\left(\frac{\delta\log\lambda}{\sqrt{w}} + \frac{\sqrt{\delta}}{w^{1/4}}\right)\right)$$

where d denotes the diameter of the action space U, w is the prediction window size and T is the total number of time slots.

This theorem implies that, with additional assumptions on the safety constraints, a *sub-linear* (in T) dynamic regret is achievable, given a sufficiently large prediction window size $w = \omega(1)$ (in T). The effectiveness of our online algorithm for closed-loop coordination between central and local controllers is validated via an electric vehicle charging application in power systems.

REFERENCES

[1] Tianyi Chen and Georgios B Giannakis. Bandit convex optimization for scalable and dynamic iot management. *IEEE Internet of Things Journal*, 6(1):1276–1286, 2018.

[2] Tongxin Li, Steven H Low, and Adam Wierman. Real-time flexibility feedback for closed-loop aggregator and system operator coordination. In *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, pages 279–292, 2020.

[3] Hao Yu, Michael Neely, and Xiaohan Wei. Online convex optimization with stochastic constraints. In *Advances in Neural Information Processing Systems*, pages 1428–1438, 2017.