Journal of Machine Learning **ISSN**: 2790-2048(e), 2790-203X(p)

A Note on Continuous-Time Online Learning

Lexing Ying [*]

Department of Mathematics, Stanford University, Stanford, CA 94305, USA.

Abstract. In online learning, the data is provided in sequential order, and the goal of the learner is to make online decisions to minimize overall regrets. This note is concerned with continuous-time models and algorithms for several online learning problems: online linear optimization, adversarial bandit, and adversarial linear bandit. For each problem, we extend the discrete-time algorithm to the continuous-time setting and provide a concise proof of the optimal regret bound.

Keywords:
Online learning,
Online optimization,
Adversarial bandits,
Adversarial linear bandits.

Article Info.:
Volume: 4
Number: 1
Pages: 1 - 10
Date: March/2025
doi.org/10.4208/jml.240605

Article History:
Received: 05/06/2024
Accepted: 30/09/2024

Communicated by:
Qianxiao Li

1 Introduction

In online learning, the data is provided in sequential order, and the goal of the learner is to make online decisions to minimize overall regrets. This is particularly relevant for problems with a dynamic aspect. This topic has produced many surprisingly efficient algorithms that are nothing short of magic.

This note is concerned with several important online learning problems:

- online linear optimization,
- adversarial bandit,
- adversarial linear bandit.

For each problem, we define a regret that quantifies how much worse a learning algorithm's performance is compared to the best fixed strategy in hindsight. In most of the existing literature, online learning problems are often placed in the discrete-time setting, and many discrete-time algorithms have been developed to achieve optimal regret bounds. However, there has been relatively little work for online learning in the continuous-time setting. In this note, for each of these problems, we propose a continuous-time model, describe an algorithm motivated by the discrete-time version, and provide a simple proof for the optimal regret bound. The main technical tools are Legendre transform and Ito's lemma.

Several books, reviews, and lecture notes are devoted to online learning [1, 2, 4, 6, 8, 11, 12] in the discrete-time setting. In recent years, there has been a growing interest in

[*] Corresponding author. lexing@stanford.edu

the continuous-time setting [3, 5, 10, 13, 14]. Among them, [3, 13] proposed diffusion approximations for Thompson sampling algorithms for multi-arm bandits; [5, 14] developed continuous models based on Hamilton-Jaboi-Bellman equation for two-armed bandits; and [10] proposed the first continuous prediction models for the experts' advice setting. Our result for the adversarial bandit problem is closely related to the work in [10].

The rest of this note is organized as follows. Section 2 summarizes the main results of the Legendre transform. Section 3 discusses the online linear optimization problem. Section 4 presents the continuous-time model for the adversarial bandit. Section 5 extends the result to the adversarial linear bandit.

2 Legendre transform

Let X be a convex set in $\mathbb{R}^d$ and $F(x)$ be a convex function defined on X. To simplify the discussion, we assume that $F(x)$ is strictly convex.

The Legendre transform [7], denoted by $G(y)$, of $F(x)$, is defined as

$$G(y) \equiv \max_{x \in X} y^\top x - F(x),$$

where the domain Y of the set where $G(y)$ is bounded.

Let $x(y)$ be the point where the maximum is achieved for a given y. Then

$$y = \nabla F\big(x(y)\big).$$

A key result of Legendre transform is that $F(x)$ is also the Legendre transform of $G(y)$

$$F(x) = \max_{y \in Y} x^\top y - G(y)$$

and, similarly for a given x, the maximizer $y(x)$ satisfies

$$x = \nabla G\big(y(x)\big).$$

A trivial but useful inequality is

$$F(x) + G(y) \geq x^\top y.$$

In this note, we are concerned with the following case:

$$X = \Delta^d \equiv \left\{ (x_1, \ldots, x_d) : x_a \geq 0, \ \sum_{i=1}^d x_i = 1 \right\}, \quad Y = \mathbb{R}^d$$

with $F(x)$ and $G(y)$ given by

$$F(x) = \beta^{-1} \sum_{i=1}^d x_i \ln x_i, \quad G(y) = \beta^{-1} \ln \left(\sum_{i=1}^d \exp(\beta y_i) \right) \tag{2.1}$$

with $\beta > 0$.