Revisiting Online Learning Approach to Inverse Linear Optimization: A Fenchel–Young Loss Perspective and Gap-Dependent Regret Analysis

Shinsaku Sakaue
The University of Tokyo and RIKEN AIP

Han Bao
Kyoto University

Taira Tsuchiya
The University of Tokyo and RIKEN AIP

Abstract

This paper revisits the online learning approach to inverse linear optimization studied by Bärmann et al. (2017), where the goal is to infer an unknown linear objective function of an agent from sequential observations of the agent's input-output pairs. First, we provide a simple understanding of the online learning approach through its connection to online convex optimization of *Fenchel–Young losses*. As a byproduct, we present an offline guarantee on the *suboptimality loss*, which measures how well predicted objective vectors explain the agent's choices, without assuming the optimality of the agent's choices. Second, assuming that there is a gap between optimal and suboptimal objective values in the agent's decision problems, we obtain an upper bound independent of the time horizon T on the sum of suboptimality and *estimate losses*, where the latter measures the quality of solutions recommended by predicted objective vectors. Interestingly, our gap-dependent analysis achieves a faster rate than the standard $O(\sqrt{T})$ regret bound by exploiting structures specific to inverse linear optimization, even though neither the loss functions nor their domains possess desirable properties, such as strong convexity.

1 INTRODUCTION

Linear optimization is arguably the most widely used model of decision-making. Inverse linear optimization is its inverse problem, where the goal is to infer linear objective functions from observed outcomes. Since the early development in geographical science (Tarantola, 1988; Burton and Toint, 1992), inverse linear optimization has been an important subject of study (Ahuja and Orlin, 2001; Heuberger, 2004; Chan et al., 2019) and used in various applications, ranging from route recommendation to healthcare (Chan et al., 2023).

Inverse linear optimization is particularly interesting when forward linear optimization is a decision-making model of a human *agent*.[1] Then, the linear objective function represents the agent's internal preference. If the agent repeatedly takes an action upon facing a set of feasible actions, inverse linear optimization can be seen as online learning of the agent's internal preference from pairs of the feasible sets and the agent's choices. Bärmann et al. (2017) studied this setting and proposed an elegant approach based on online learning, which is the focus of this paper and is described below.

Consider an agent who addresses decision problems of the following linear-optimization form for $t = 1, \dots, T$:

$$\text{maximize } \langle c^*, x \rangle \qquad \text{subject to } x \in X_t, \tag{1}$$

where $c^* \in \mathbb{R}^n$ ($n \in \mathbb{Z}_{>0}$) is the agent's objective vector, and $X_t \subseteq \mathbb{R}^n$ is the set of actions the agent can take in round t, which is not necessarily convex. Let $x_t \in X_t$ be the agent's choice in round t. Informally, we want to learn c^* from the input-output pairs, $\{(X_t, x_t)\}_{t=1}^T$.

To this end, we need learning objectives that quantify how good prediction $\hat{c}$ of c^* is based on the observations, $\{(X_t, x_t)\}_{t=1}^T$. Let $\Theta \subseteq \mathbb{R}^n$ be the space of objective vectors, from which we pick prediction $\hat{c} \in \Theta$. Bärmann et al. (2017) considered the following learning objective:

$$\max\{ \langle \hat{c}, x' \rangle \mid x' \in X_t \} - \langle \hat{c}, x_t \rangle \quad t \in \{1, \dots, T\}. \tag{2}$$

If (non-zero) $\hat{c}$ makes (2) zero for all $t = 1, \dots, T$, then $\hat{c}$ consistently explains why the agent takes $x_t \in X_t$. Since (2) represents the suboptimality of x_t for $\hat{c}$, it is called *suboptimality loss* (Mohajerin Esfahani et al., 2018; Chen and Kılınç-Karzan, 2020; Sun et al., 2023).

Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

[1] An agent is sometimes called an expert, but we here use "agent" to avoid confusion with experts in online learning.

For any $\hat{c} \in \Theta$, letting $\hat{x}_t \in \arg\max_{x' \in X_t} \langle \hat{c}, x' \rangle$, we can write the suboptimality loss (2) as follows:

$$\ell_t^{\mathrm{sub}}(\hat{c}) \coloneqq \langle \hat{c}, \hat{x}_t - x_t \rangle \quad t \in \{1, \ldots, T\}. \tag{3}$$

Relying solely on the suboptimality loss is sometimes insufficient, as the zero suboptimality loss is attained by trivial all-zero prediction $\hat{c} = 0$ (cf. Mishra et al. 2024, Section 6.1).[2] Another natural learning objective free from this issue is the so-called *estimate loss* (Chen and Kılınç-Karzan, 2020; Sun et al., 2023),[3] defined by

$$\ell_t^{\mathrm{est}}(\hat{c}) \coloneqq \langle c^*, x_t - \hat{x}_t \rangle \quad t \in \{1, \ldots, T\}.$$

Ideally, we want to find $\hat{c} \in \Theta$ that attains $\ell_t^{\mathrm{sub}}(\hat{c}) = \ell_t^{\mathrm{est}}(\hat{c}) = 0$ for $t = 1, \ldots, T$. Bärmann et al. (2017) showed that a proxy goal is achievable in an online setting. Specifically, for each round t, given $\{(X_i, x_i)\}_{i=1}^{t-1}$, we can compute prediction $\hat{c}_t \in \Theta$ so that

$$\sum_{t=1}^{T} \underbrace{\left(\ell_t^{\mathrm{sub}}(\hat{c}_t) + \ell_t^{\mathrm{est}}(\hat{c}_t)\right)}_{\text{Total loss}} = \sum_{t=1}^{T} \langle \hat{c}_t - c^*, \hat{x}_t - x_t \rangle \tag{4}$$

grows at the rate of $O(\sqrt{T})$, and hence the average of $\ell_t^{\mathrm{sub}}(\hat{c}_t) + \ell_t^{\mathrm{est}}(\hat{c}_t)$ converges to zero as $T \to \infty$. For convenience, we call the sum of the suboptimality and estimate losses, $\ell_t^{\mathrm{sub}}(\hat{c}_t) + \ell_t^{\mathrm{est}}(\hat{c}_t)$ in (4), the *total loss*. The idea of Bärmann et al. (2017) is based on online learning, which works as follows: for each $t = 1, \ldots, T$, an adversary chooses a convex loss $f_t \colon \Theta \to \mathbb{R}$ and a learner computes $\hat{c}_t \in \Theta$ based on past of observations $\{f_i\}_{i=1}^{t-1}$ so that the *regret* $\sum_{t=1}^{T}(f_t(\hat{c}_t) - f_t(c^*))$ grows only sublinearly in T for any $c^* \in \Theta$. Bärmann et al. (2017) regarded $f_t \colon \hat{c} \mapsto \langle \hat{x}_t - x_t, \hat{c} \rangle$ as a linear loss function by considering $\hat{x}_t$ independent of $\hat{c}$,[4] and used the well-known multiplicative weight update method (MWU); consequently, the $O(\sqrt{T})$ regret bound of MWU translates to the bound on (4).

1.1 Our Contributions

We revisit the online learning approach of Bärmann et al. (2017) and make the following contributions.

Drawing connection to OCO of Fenchel–Young losses. While Bärmann et al. (2017) applied the idea of online learning to the linear loss $f_t \colon \hat{c} \mapsto \langle \hat{x}_t - x_t, \hat{c} \rangle$, we view the problem as online convex optimization (OCO) of *Fenchel–Young losses* (Blondel et al., 2020). Specifically, we first show that the suboptimality loss (3) is a particular type of Fenchel–Young loss, allowing us to take advantage of its useful properties (see Proposition 2.2). Then, we show that the total loss in (4) appears as the *linearized regret* of the Fenchel–Young loss, to which we apply the Follow-The-Regularized-Leader (FTRL) and obtain a bound of $O(\sqrt{T})$ that applies to various situations. We thus offer a simple understanding of the online learning approach to inverse linear optimization from the viewpoint of Fenchel–Young losses. As a byproduct, we obtain an offline guarantee on the suboptimality loss that enjoys broader applicability than a previous result in Bärmann et al. (2020).

Gap-dependent regret analysis. We then discuss going beyond the $O(\sqrt{T})$ bound on the total loss over T rounds (4). We focus on the case where there is a gap between the optimal and suboptimal objective values in the agent's decision problems (1); we parameterize this by $\Delta > 0$. Under this assumption, we obtain a bound of $O(1/\Delta^2)$, which is independent of the time horizon T. Unlike typical gap assumptions in online learning, our assumption is imposed on the agent's decision problems, highlighting that problem structures specific to inverse linear optimization can help accelerate online learning of the agent's objective vector. We discuss situations where the gap assumption is reasonable in Section 5.2.

1.2 Related Work

Inverse optimization. Most studies on inverse optimization are based on the KKT condition (Iyengar and Kang, 2005; Tan et al., 2020; Mishra et al., 2024). The online learning approach of Bärmann et al. (2017) enjoys broader applicability as it does not rely on specific representations of the agent's feasible regions, allowing for even non-connected sets (see Section 2.1). Besbes et al. (2021, 2023) recently obtained an $O(n^4 \ln T)$ regret bound with respect to ℓ_t^{est}, where n is the dimension of the ambient space of objective vectors. More recently, Sakaue et al. (2025) obtained an improved bound of $O(n \ln T)$ on $\ell_t^{\mathrm{est}} + \ell_t^{\mathrm{sub}}$. On the other hand, our bound in Section 5 is independent of T under the gap assumption. Other online-learning-based approaches with different criteria are also widely studied (Dong et al., 2018; Chen and Kılınç-Karzan, 2020; Sun et al., 2023). A minor remark is that a criterion used by Chen and Kılınç-Karzan (2020), called the simple loss, is different from the total loss (4), although they might look similar. Indeed, their simple loss is defined as a linear function (Chen and Kılınç-Karzan, 2020, Lemma 2), while the total loss (4) consists of non-linear terms, such as ℓ_t^{sub}.

[2] Bärmann et al. (2017) avoided this issue by considering Θ that does not contain the all-zero vector—specifically, the probability simplex $\Theta = \{\, c \in \mathbb{R}^n_{\geq 0} \mid \|c\|_1 = 1 \,\}$.

[3] These studies also consider the prediction loss, $\|\hat{c} - c\|^2$, which we cannot use as no direct supervision on $\hat{c}$ is given.

[4] Taking the dependence of $\hat{x}_t$ on $\hat{c}$ into account, f_t coincides with the non-linear suboptimality loss (3). The above linear-loss perspective is explicitly described in Bärmann et al. (2020, Section 3.3), an extended arXiv version of Bärmann et al. 2017 that includes additional results, such as an online-gradient-descent approach and an offline guarantee.

Smart predict, then optimize. *Smart predict, then optimize* (Elmachtoub and Grigas, 2022) is a relevant framework for learning objective functions. Similar settings are studied under the name of *decision-focused learning* (Wilder et al., 2019) and *contextual linear optimization* (Hu et al., 2022). These frameworks are intended for learning models that predict objective functions from contextual information. To this end, they require past objective functions as training data, which is not available in our inverse optimization setting.

Fenchel–Young loss. The Fenchel–Young-loss framework (Blondel et al., 2020) offers a general recipe for designing convex surrogate losses in supervised learning. Its usefulness in a remotely related scenario—learning models that generate objective functions based on contextual information and optimal solutions—was demonstrated by Berthet et al. (2020). Different from their work, we elucidate an unexpected connection between the Fenchel–Young loss and the suboptimality and estimate losses in inverse linear optimization, which is of interest in its own right. Sakaue et al. (2024) used Fenchel–Young losses for online structured prediction, focusing on a different criterion called the *surrogate regret*.

Online convex optimization. Our work is intended to benefit from rich theory of OCO (Orabona, 2023) in inverse linear optimization. In OCO, going beyond the $O(\sqrt{T})$ regret requires additional assumptions, such as the strong convexity and exp-concavity of loss functions (Hazan et al., 2007) and curved decision sets (Huang et al., 2017). Also, gap conditions related to the *learner's* choices sometimes help achieve fast rates in stochastic settings (Lai and Robbins, 1985; Auer et al., 2002). Different from them, our gap condition, detailed in Definition 5.1, is imposed on the *agent's* decision problems (1) and hence specific to inverse optimization.

2 PRELIMINARIES

Notation. Let $\langle \cdot, \cdot \rangle$ be the standard inner product on $\mathbb{R}^n$, or the paring on the primal space $\mathbb{R}^n$ of agent's actions and the dual space $\Theta \subseteq \mathbb{R}^n$ of predictions. Let $\|\cdot\|$ denote a norm on the agent's space $\mathbb{R}^n$ and $\|\cdot\|_\star$ the dual norm on Θ. We sometimes omit the subscript of t when the dependence on t is irrelevant. For example, we use (X, x) to represent a pair of agent's feasible set and output, $\hat{c} \in \Theta$ a prediction, and $\hat{x} \in \arg\max_{x' \in X} \langle \hat{c}, x' \rangle$ an optimal solution in terms of prediction $\hat{c}$.[5] Similarly, let ℓ^{sub} and ℓ^{est} denote the suboptimality and estimate losses, respectively, without specifying the round, t.

[5] As $\hat{x}$ depends on $\hat{c}$, it should be read like $\hat{x}(\hat{c})$. However, we omit the dependence on $\hat{c}$ for brevity. We use the "hat" accent on x to represent the dependence on some prediction $\hat{c}$.

2.1 Problem Setting

As described in Section 1, we aim to learn an unknown objective vector $c^* \in \mathbb{R}^n$ of an agent who addresses problem (1) for $t = 1, \ldots, T$. We assume that every feasible set $X_t \subseteq \mathbb{R}^n$ is a non-empty compact set containing the agent's choice x_t. Until Section 5, $x_t \in X_t$ is not necessarily optimal for c^*, unlike Bärmann et al. (2017, 2020). Also, X_t is not necessarily convex; the only requirement about X_t is that we can compute an optimal solution $x \in \arg\max_{x' \in X_t} \langle c, x' \rangle$ for any $c \in \mathbb{R}^n$. A typical tractable case is when X_t is a polytope specified by linear constraints; then, computing an optimal solution reduces to a linear program. Even if X_t is non-convex (and even non-connected, as in the case of integer linear programs), we can use empirically efficient solvers, such as Gurobi, to find an optimal solution. In Section 5, we will additionally assume that the agent's choices are optimal, i.e., $x_t \in \arg\max_{x' \in X_t} \langle c^*, x' \rangle$, and that the agent's decision problems (1) satisfy a certain gap condition, which we explain later.

We assume that the space of predictions, $\Theta \subseteq \mathbb{R}^n$, is a closed convex set containing the agent's true objective c^*. Additionally, while not strictly necessary, it is common to assume that Θ does not contain the origin; otherwise, trivial all-zero prediction $\hat{c} = 0$ attains $\ell_t^{\mathrm{sub}}(0) = 0$, as discussed in Section 1. For example, Bärmann et al. (2017) studied the case where $\Theta = \{\, c \in \mathbb{R}^n_{\geq 0} \mid \|c\|_1 = 1 \,\}$, i.e., the probability simplex. While our analysis does not use $0 \notin \Theta$, we suppose that this condition holds in Section 4.2, as we focus on the suboptimality loss therein.

Similar to Bärmann et al. (2017), we mainly focus on the online setting, which goes as follows. For each $t = 1, \ldots, T$, we compute $\hat{c}_t \in \Theta$ using past observations $\{(X_i, x_i)\}_{i=1}^{t-1}$. Then, we observe (X_t, x_t) and incur the tth total loss $\ell_t^{\mathrm{sub}}(\hat{c}_t) + \ell_t^{\mathrm{est}}(\hat{c}_t)$. We aim to minimize the total loss over T rounds (4). In Section 4.2, we discuss a stochastic offline setting and give a bound on the suboptimality loss in expectation, which we detail later.

2.2 Fenchel–Young Loss

We briefly describe the basics of the Fenchel–Young loss (Blondel et al., 2020), which is defined as follows.

Definition 2.1. Let $\Omega \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a (possibly non-convex) lower semi-continuous function. For any given $x \in \mathbb{R}^n$ with $\Omega(x) < +\infty$, the Fenchel–Young loss is defined as a function of $c \in \mathbb{R}^n$ as follows:

$$L_\Omega(c; x) = \Omega^*(c) + \Omega(x) - \langle c, x \rangle,$$

where $\Omega^*(c) := \sup_{x \in \mathbb{R}^n} \langle c, x \rangle - \Omega(x)$ is the convex conjugate of Ω.

Many common loss functions in machine learning, such as the squared and logistic losses, can be represented as Fenchel–Young losses by choosing appropriate Ω. We can interpret $L_\Omega(c; x)$ as the discrepancy between primal and dual variables, x and c, measured by the *Fenchel–Young inequality*, $\Omega^*(c) + \Omega(x) \geq \langle c, x\rangle$, a fundamental inequality in convex analysis that readily follows from the definition of Ω^*. In our case, we want to measure the quality of (dual) prediction $\hat{c}$ by using the agent's output x as (primal) feedback. Therefore, the Fenchel–Young loss can be useful to measure the quality of prediction $\hat{c}$ in inverse linear optimization.

We then present some properties of the Fenchel–Young loss that will be useful in the subsequent discussion. We refer the reader to Blondel et al. (2020, Proposition 2) for more information.

Proposition 2.2. *For any $x \in \mathbb{R}^n$, the Fenchel–Young loss $L_\Omega(\cdot; x)$ is non-negative and convex. For any $\hat{c} \in \mathbb{R}^n$, if $\hat{x} \in \arg\max_{x' \in \mathbb{R}^n} \langle \hat{c}, x'\rangle - \Omega(x')$ exists, then the residual vector is a subgradient, i.e., $\hat{x} - x \in \partial L_\Omega(\hat{c}; x)$.*

We emphasize that Proposition 2.2 holds even if Ω is non-convex (and even if non-continuous). These properties are not difficult to confirm: the non-negativity is due to the Fenchel–Young inequality, and the convexity follows from the fact that $\Omega^*(c)$ is the supremum of linear functions $c \mapsto \langle c, x\rangle - \Omega(x)$. The subgradient property follows from Danskin's theorem (Danskin, 1966), which ensures $\hat{x} \in \partial\Omega^*(\hat{c})$.

3 FENCHEL–YOUNG-LOSS PERSPECTIVE

This section describes a Fenchel–Young-loss perspective on inverse linear optimization. Below, let $X \subseteq \mathbb{R}^n$ and $x \in X$ be the agent's feasible set and output, respectively. For convenience, let $L_X(\cdot; x)$ denote the Fenchel–Young loss with $\Omega = I_X$,[6] where $I_X : \mathbb{R}^n \to \{0, +\infty\}$ is the indicator function of X, i.e., $I_X(x') = 0$ if $x' \in X$ and $+\infty$ otherwise. Since X is a non-empty compact set, an optimal solution to $\max_{x' \in X} \langle c, x'\rangle$ exists for any $c \in \mathbb{R}^n$; therefore, $L_X(\cdot; x)$ satisfies Proposition 2.2.

3.1 ℓ^{sub} is a Fenchel–Young Loss

First, we observe that for any agent's choice $x \in X$, $L_X(\cdot; x)$ coincides with the suboptimality loss (3), ℓ^{sub}.

Proposition 3.1. *The suboptimality loss ℓ^{sub} (3) is a Fenchel–Young loss with $\Omega = I_X$, i.e., for any $\hat{c} \in \mathbb{R}^n$,*

$$L_X(\hat{c}; x) = \max_{x' \in X} \langle \hat{c}, x'\rangle - \langle \hat{c}, x\rangle = \langle \hat{c}, \hat{x} - x\rangle = \ell^{\mathrm{sub}}(\hat{c}),$$

where $\hat{x} \in \arg\max_{x' \in X} \langle \hat{c}, x'\rangle$.

Proof. From Definition 2.1 with $\Omega = I_X$ and $\Omega^*(\hat{c}) = \sup_{x' \in \mathbb{R}^n} \langle \hat{c}, x'\rangle - \Omega(x') = \max_{x' \in X} \langle \hat{c}, x'\rangle$, it holds that $L_X(\hat{c}; x) = \Omega^*(\hat{c}) + \Omega(x) - \langle \hat{c}, x\rangle = \max_{x' \in X} \langle \hat{c}, x'\rangle + 0 - \langle \hat{c}, x\rangle = \langle \hat{c}, \hat{x} - x\rangle = \ell^{\mathrm{sub}}(\hat{c})$. □

This connection, although easy to derive, clarifies that the suboptimality loss ℓ^{sub} enjoys the properties of Fenchel–Young losses in Proposition 2.2. Specifically, ℓ^{sub} is non-negative and convex, and the residual $\hat{x} - x$ is a subgradient of $\ell^{\mathrm{sub}}(\hat{c})$.

Remark 3.2. While these properties are also proved in Bärmann et al. (2020, Proposition 3.1) by assuming the agent's optimality $x \in \arg\max_{x' \in X} \langle c^*, x'\rangle$, the above Fenchel–Young-loss perspective offers a comprehensive understanding of the suboptimality loss through convex analysis. It is also worth mentioning that the Fenchel–Young loss, $L_X(\cdot; x)$, can be defined for any $x \in X$, regardless of its optimality for c^*.

3.2 $\ell^{\mathrm{sub}} + \ell^{\mathrm{est}}$ = the Linearized Regret of a Fenchel–Young Loss

In online learning, *linearization* is a common technique for reducing OCO to online linear optimization (OLO) (Orabona, 2023, Section 2.3). For any convex losses $f_t : \Theta \to \mathbb{R}$, predictions $\hat{c}_t \in \Theta$, subgradients $g_t \in \partial f_t(\hat{c}_t)$, and a comparator $c^* \in \Theta$, the *linearized regret* is defined by $\sum_{t=1}^{T} \langle g_t, \hat{c}_t - c^*\rangle$. While this definition depends on the choice of subgradient $g_t \in \partial f_t(\hat{c}_t)$, the convexity of f_t ensures $f_t(\hat{c}_t) - f_t(c^*) \leq \langle g_t, \hat{c}_t - c^*\rangle$ for any $c^* \in \Theta$ and $g_t \in \partial f_t(\hat{c}_t)$. Thus, any upper bound on $\sum_{t=1}^{T} \langle g_t, \hat{c}_t - c^*\rangle$ obtained with an OLO method applies to the original regret, $\sum_{t=1}^{T} (f_t(\hat{c}_t) - f_t(c^*))$.[7]

The following Proposition 3.3 says that if we apply the linearization technique to the regret with respect to the Fenchel–Young loss, $L_X(\hat{c}; x) - L_X(c^*; x)$, the resulting linearized regret equals the total loss, $\ell^{\mathrm{sub}} + \ell^{\mathrm{est}}$, in (4).

Proposition 3.3. *Let $\hat{c} \in \Theta$, $\hat{x} \in \arg\max_{x' \in X} \langle \hat{c}, x'\rangle$, $x \in X$, and $g = \hat{x} - x$. For any $c^* \in \mathbb{R}^n$, $\langle g, \hat{c} - c^*\rangle$ is the linearized regret of $L_X(\hat{c}; x) - L_X(c^*; x)$, which satisfies*

$$L_X(\hat{c}; x) - L_X(c^*; x) \leq \langle g, \hat{c} - c^*\rangle = \ell^{\mathrm{sub}}(\hat{c}) + \ell^{\mathrm{est}}(\hat{c}),$$

[6] When X is a convex set, this Fenchel–Young loss is also called the structured-perceptron loss (Blondel et al., 2020).

[7] One might wonder whether the learner can compete against $g_t \in \partial f_t(\hat{c}_t)$, which is selected in reaction to $\hat{c}_t$. Fortunately, online learning methods are designed to work with such a reactive adversary. This fact is also used in Bärmann et al. (2017) when applying MWU to $f_t : \hat{c} \mapsto \langle \hat{x}_t - x_t, \hat{c}\rangle$, where $\hat{x}_t \in \arg\max_{x' \in X_t} \langle \hat{c}_t, x'\rangle$ depends on $\hat{c}_t$.

where $\ell^{\mathrm{sub}}(\hat{c}) = \langle \hat{c}, \hat{x} - x \rangle$ *and* $\ell^{\mathrm{est}}(\hat{c}) = \langle c^*, x - \hat{x} \rangle$.

Proof. From Proposition 2.2, $L_X(\cdot; x)$ is convex and $g = \hat{x} - x \in \partial L_X(\hat{c}; x)$ holds. Thus, $L_X(\hat{c}; x) - L_X(c^*; x)$ is upper bounded by $\langle g, \hat{c} - c^* \rangle$. Plugging $g = \hat{x} - x$ into this yields the equality. □

We have observed how the suboptimality and total losses relate to the Fenchel–Young loss and its linearized regret, which bridges between inverse linear optimization and OCO. Built on this, Sections 4 and 5 provide theoretical guarantees for inverse linear optimization.

3.3 Discussion on a Difference from the Previous Understanding

We briefly digress to discuss a conceptual difference from Bärmann et al. (2017), who derived the online learning approach to inverse linear optimization by viewing $f_t \colon \hat{c} \mapsto \langle \hat{x}_t - x_t, \hat{c} \rangle$ as a *linear loss*, as mentioned in Section 1. Actually, once we apply the linearization technique to the regret with respect to the suboptimality loss ℓ_t^{sub} as in Proposition 3.3, the resulting linearized regret coincides with the regret with respect to the linear loss, i.e., $\langle g_t, \hat{c}_t - c^* \rangle = \langle \hat{x}_t - x_t, \hat{c}_t - c^* \rangle = f_t(\hat{c}_t) - f_t(c^*)$. Hence, our approach recovers the same result as that of Bärmann et al. (2017), as detailed in Section 4.1. Still, the above Fenchel–Young-loss perspective offers yet another way to understand how the online learning approach works. Below, we discuss a subtle related difference, which suggests that it is more natural to interpret the online learning approach as OCO of the *nonlinear* suboptimality losses, ℓ_t^{sub}.

Let us discuss the same setting as that of Bärmann et al. (2017), where Θ is the probability simplex and each loss f_t is viewed as the linear loss. When considering the regret of predictions $\hat{c}_1, \dots, \hat{c}_T \in \Theta$, it is usual to compare them with an *ex-post* optimal prediction, $c^*_{\min} \in \Theta$, that minimizes the cumulative loss $\sum_{t=1}^T f_t(c)$ in hindsight. Since Θ is the probability simplex and f_t's are linear, $c^*_{\min}$ is a standard basis vector (unless $\sum_{t=1}^T (\hat{x}_t - x_t)$ is normal to a face of Θ). Such $c^*_{\min}$ may not be expressive enough to represent the human agent's preference c^*; this suggests that the linear-loss perspective encounters a dissonance between the ex-post optimal prediction $c^*_{\min}$ and the agent's preference c^*. By contrast, our formulation uses the suboptimality loss ℓ_t^{sub} as the target loss, and reduces OCO of ℓ_t^{sub} to OLO of f_t via linearization. Therein, ℓ_t^{sub} is nonlinear, and the corresponding ex-post optimal prediction $c^*_{\min}$, which minimizes $\sum_{t=1}^T \ell_t^{\mathrm{sub}}(c)$, would be a more plausible representation of the agent's preference c^*. However, we note that this difference does not affect regret bounds discussed below.

Algorithm 1 FTRL for inverse linear optimization

1: **for** $t = 1, \dots, T$ **:**
2: Output $\hat{c}_t \in \arg\min_{c \in \Theta} \beta_t \psi(c) + \sum_{i=1}^{t-1} \langle g_i, c \rangle$.
▷ $\hat{c}_1 \in \Theta$ is arbitrary, and $\hat{c}_t = \hat{c}_{t-1}$ if $g_{t-1} = 0$.
3: Observe (X_t, x_t).
4: Compute $\hat{x}_t \in \arg\max_{x' \in X_t} \langle \hat{c}_t, x' \rangle$.
5: Set $g_t = \hat{x}_t - x_t$. ▷ $g_t \in \partial L_{X_t}(\hat{c}_t; x_t)$.

4 THEORETICAL GUARANTEES FOR GENERAL CASES

We discuss the online setting in Section 4.1 and present a bound on the linearized regret, or the total loss over T rounds (4), which we denote by R_T for convenience:

$$R_T \coloneqq \sum_{t=1}^T \langle g_t, \hat{c}_t - c^* \rangle = \sum_{t=1}^T \left(\ell_t^{\mathrm{sub}}(\hat{c}_t) + \ell_t^{\mathrm{est}}(\hat{c}_t) \right). \quad (5)$$

Here, $\hat{c}_t$ is the tth prediction, $\hat{x}_t \in \arg\max_{x' \in X_t} \langle \hat{c}_t, x' \rangle$ is an optimal solution for $\hat{c}_t$, and $g_t = \hat{x}_t - x_t$ is a subgradient of $L_{X_t}(\hat{c}_t; x_t) = \ell_t^{\mathrm{sub}}(\hat{c}_t)$. In Section 4.2, we discuss a stochastic offline setting and present a bound on the suboptimality loss in expectation.

4.1 Online Guarantee on $\ell^{\mathrm{sub}} + \ell^{\mathrm{est}}$

We consider the online setting and give a bound on R_T. We use the well-known FTRL described in Orabona (2023, Section 7), which is shown in Algorithm 1 for our setting for completeness. Let $\lambda > 0$ and $\psi \colon \Theta \to \mathbb{R}$ be a differentiable λ-strongly convex function with respect to the dual norm $\|\cdot\|_\star$. The following bound on R_T is a consequence of the existing analysis of FTRL. For completeness, we provide the proof in Appendix A.

Proposition 4.1 (cf. Orabona 2023, Section 7)**.** *Assume that there exists $B > 0$ such that*

$$\max\left\{ 2^{5/2} \lambda \max_{c, c' \in \Theta} \|c - c'\|_\star^2, \max_{c, c' \in \Theta} \left(\psi(c) - \psi(c') \right) \right\} \le B^2.$$

Let $\hat{c}_1, \dots, \hat{c}_T \in \Theta$ be the outputs of Algorithm 1 with $\beta_t = \frac{2^{1/4}}{B} \sqrt{\frac{\sum_{i=1}^{t-1} \|g_i\|^2}{\lambda}}$. For any $c^* \in \Theta$, the linearized regret (5) (equivalently, the total loss over T rounds (4)) is bounded as follows:

$$R_T = \sum_{t=1}^T \langle g_t, \hat{c}_t - c^* \rangle \le 2^{5/4} B \sqrt{\frac{1}{\lambda} \sum_{t=1}^T \|g_t\|^2}. \quad (6)$$

In particular, if $\|g_t\| = \|\hat{x}_t - x_t\| \le K$ for $t = 1, \dots, T$ for some $K > 0$, we have $R_T \le 2^{5/4} K B \sqrt{T/\lambda}$.

Recovering Bärmann et al. (2017, Theorem 3.3). As a concrete example, we show that Proposition 4.1 implies Bärmann et al. (2017, Theorem 3.3) as a special case. Here, we set $\|\cdot\| = \|\cdot\|_\infty$ and $\|\cdot\|_\star = \|\cdot\|_1$. In their setting, $\Theta = \{ c \in \mathbb{R}^n_{\geq 0} \mid \|c\|_1 = 1 \}$ is the probability simplex, and the ℓ_∞-diameter of the agent's feasible sets, $X_1, \dots, X_T$, is bounded by some $K > 0$; therefore, $\max_{c,c' \in \Theta} \|c - c'\|_1^2 \leq 4$ and $\|g_t\|_\infty = \|\hat{x}_t - x_t\|_\infty \leq K$ $(t = 1, \dots, T)$ hold. Regarding ψ in FTRL, we use the negative Shannon entropy, $\Theta \ni c \mapsto \langle c, \ln c\rangle$, where ln is taken element-wise. From Pinsker's inequality, ψ is 1-strongly convex with respect to $\|\cdot\|_1$, hence $\lambda = 1$. From $\max_{c,c' \in \Theta}(\psi(c) - \psi(c')) \leq \ln n$, it suffices to set $B = 2^{11/4}\sqrt{\ln n}$ (for $n \geq 2$). Thus, Proposition 4.1 implies

$$\sum_{t=1}^{T} \langle \hat{c}_t - c^*, \hat{x}_t - x_t \rangle = R_T \leq 16K\sqrt{T \ln n},$$

achieving a bound of $O(K\sqrt{T \ln n})$ on the total loss over T rounds (4), as with Bärmann et al. (2017, Theorem 3.3).[8] In addition, the online learning methods are essentially identical: Bärmann et al. (2017) used MWU, or the exponentiated gradient method, which is a special case of FTRL with $\psi(c) = \langle c, \ln c\rangle$ (Orabona, 2023, Section 7.5).

Proposition 4.1 also applies to various other settings with different convex sets Θ and convex functions ψ. For example, it is not difficult to recover a similar bound to Bärmann et al. (2020, Theorem 3.11) achieved with the online subgradient descent method (cf. Orabona 2023, Example 7.11). We will also use Proposition 4.1 in the gap-dependent analysis in Section 5.

It is also worth noting that Proposition 4.1 applies to the regret with respect to the suboptimality loss thanks to Propositions 3.1 and 3.3. The following Corollary 4.2, while straightforward, highlights our viewpoint that the online learning approach can be seen as OCO of ℓ^{sub}.

Corollary 4.2. *It holds that*

$$R_T^{\mathrm{sub}} := \sum_{t=1}^{T} \left(\ell_t^{\mathrm{sub}}(\hat{c}_t) - \ell_t^{\mathrm{sub}}(c^*) \right) \leq R_T, \qquad (7)$$

where R_T^{sub} is the regret with respect to the suboptimality loss, and hence Proposition 4.1 also applies to R_T^{sub}.

4.2 Offline Guarantee on ℓ^{sub}

We then show an offline guarantee on the suboptimality loss (3). Having established the connection to the Fenchel–Young loss in Section 3, the following Theorem 4.3 is immediate from the standard *online-to-batch conversion* (Cesa-Bianchi et al., 2004). Still, it is worth mentioning due to its wider applicability than the previous result by Bärmann et al. (2020, Theorem 3.14), as detailed later.

As with Bärmann et al. (2020, Section 3.4), we assume that observations (X, x) follow an unknown distribution $\mathcal{D}$ and that $0 \notin \Theta$ holds, as discussed in Section 2.1. The following Theorem 4.3 states that any bound on the regret with respect to the suboptimality loss, R_T^{sub}, (and hence any bound on R_T) translates into a bound on the expected suboptimality loss as follows.

Theorem 4.3. *Assume that $(X_1, x_1), \dots, (X_T, x_T)$ follow i.i.d. some distribution $\mathcal{D}$. Compute $\hat{c}_1, \dots, \hat{c}_T \in \Theta$ by using any OCO algorithm such that R_T^{sub} in (7) is bounded. Then, for any $c^* \in \Theta$, the average prediction, $\hat{c} = \frac{1}{T}\sum_{t=1}^{T} \hat{c}_t$, satisfies*

$$\mathbb{E}\left[\ell_{X,x}^{\mathrm{sub}}(\hat{c})\right] \leq \mathop{\mathbb{E}}_{(X,x)\sim\mathcal{D}}\left[\ell_{X,x}^{\mathrm{sub}}(c^*)\right] + \mathbb{E}\left[\frac{R_T^{\mathrm{sub}}}{T}\right],$$

where the expectation on the left-hand side is taken over $\{(X_t, x_t)\}_{t=1}^{T} \sim \mathcal{D}^T$ and $(X, x) \sim \mathcal{D}$. Specifically, if we use Algorithm 1 and obtain $R_T = O(KB\sqrt{T/\lambda})$ as in Proposition 4.1, it holds that

$$\mathbb{E}\left[\ell_{X,x}^{\mathrm{sub}}(\hat{c})\right] - \mathop{\mathbb{E}}_{(X,x)\sim\mathcal{D}}\left[\ell_{X,x}^{\mathrm{sub}}(c^*)\right] = O\left(\frac{KB}{\sqrt{T\lambda}}\right).$$

We defer the proof to Appendix C, as it directly follows from the standard online-to-batch conversion applied to the regret with respect to the convex suboptimality loss.

The previous offline result (Bärmann et al., 2020, Theorem 3.14) focuses on the case of $\mathbb{E}_{(X,x)\sim\mathcal{D}}\left[\ell_{X,x}^{\mathrm{sub}}(c^*)\right] = 0$, which requires that every $(X, x) \sim \mathcal{D}$ arising with non-zero probability satisfies $x \in \arg\max_{x' \in X} \langle c^*, x'\rangle$. This condition is sometimes restrictive as agents may not take best actions at all times (e.g., Jagadeesan et al. 2021). By contrast, Theorem 4.3 holds even when $\mathbb{E}_{(X,x)\sim\mathcal{D}}\left[\ell_{X,x}^{\mathrm{sub}}(c^*)\right] > 0$, which quantifies the suboptimality of the agent's choices. This extended applicability stems from the fact that the Fenchel–Young loss, $L_X(\cdot, x)$, can be defined for any $x \in X$, as discussed in Remark 3.2. The fact that the offline guarantee with the broader applicability is easily derived from standard arguments in OCO suggests the merit of viewing inverse linear optimization as OCO of Fenchel–Young losses.

Also, while Theorem 4.3 pertains to the average prediction $\hat{c}$, we can use *anytime* online-to-batch conversion (Cutkosky, 2019) to ensure the last-iterate convergence. Specifically, by using any OLO algorithm, we can obtain iterates $\hat{c}_t$'s such that subgradients are evaluated at $\hat{c}_t$'s and individual iterate $\hat{c}_t$ enjoys the offline guarantee.

[8] The constant is increased compared to Bärmann et al. (2017, Theorem 3.3) as a trade-off for making (6) depend on $\sum_{t=1}^{T} \|g_t\|^2$, which is crucial in the gap-dependent analysis in Section 5. We can recover exactly the same bound by using FTRL with a different choice of β_t (see Appendix B).

5 GAP-DEPENDENT BOUND ON TOTAL LOSS

We return to the online setting and consider bounding the linearized regret R_T (5), or the total loss over T rounds. In Section 4.1, we have observed that the bound of $O(\sqrt{T})$ is achievable, as is often the case in OCO. In general OCO (or even in OLO), the rate of $O(\sqrt{T})$ is tight (Orabona, 2023, Chapter 5), and improving this requires additional assumptions, such as the strong convexity or exp-concavity of loss functions (Hazan et al., 2007). Unfortunately, the Fenchel–Young loss, $L_X(\cdot; x)$, does not enjoy such properties.[9] Given this, an interesting question is: *can we improve the bound of $O(\sqrt{T})$ by exploiting structures specific to inverse linear optimization?* Below, we show that we can achieve a bound independent of T if the agent's decision problems satisfy a certain gap condition.

First, we formally define our assumption on the gap between optimal and suboptimal objective values.

Definition 5.1 (Δ-gap condition)**.** Let $c^* \in \Theta$ be the agent's objective vector. We say the agent's decision problem (1), which is specified by $c^* \in \Theta$ and $X_t \subseteq \mathbb{R}^n$, satisfies the Δ-*gap condition* for $\Delta > 0$ if, for every $t = 1, \dots, T$, it holds that

$$\langle c^*, x - \hat{x} \rangle \ge \Delta \|x - \hat{x}\| \quad \forall \hat{x} \in X_t$$

for $x = \arg\max_{x' \in X_t} \langle c^*, x' \rangle$.

The above condition implies that x is the unique optimal solution for c^*; otherwise, there exists $\hat{x} \in X_t$ with $\langle c^*, x - \hat{x} \rangle = 0$ and $\|x - \hat{x}\| > 0$, violating the inequality. In Section 5.2, we will discuss situations where the Δ-gap condition is reasonable. Note that the condition is imposed on the agent's decision problems and does not make any explicit assumptions on the loss functions or their domain Θ.

The main result of this section is Theorem 5.2, which offers a bound of $O(1/\Delta^2)$ on the linearized regret (5), or the total loss over T rounds, under the Δ-gap condition.

Theorem 5.2. *Assume the same conditions as Proposition 4.1, i.e., ψ is λ-strongly convex with respect to $\|\cdot\|_\star$, $2^{5/2}\lambda \max_{c,c' \in \Theta} \|c - c'\|_\star^2$ and $\max_{c,c' \in \Theta}(\psi(c) - \psi(c'))$ are bounded by B^2 from above, and the diameter of X_t's with respect to $\|\cdot\|$ is at most K. Additionally, assume the following two conditions: for every $t = 1, \dots, T$, (i) the agent's decision problem (1) satisfies the Δ-gap condition, and (ii) the agent's choice is optimal for c^*, i.e., $x_t \in \arg\max_{x' \in X_t} \langle c^*, x' \rangle$. Then, Algorithm 1 achieves the following bound on the linearized regret R_T (5):*

$$R_T \le \frac{2^{5/4} K B^3}{\lambda^{3/2} \Delta^2}.$$

Notably, the above regret bound is achieved by the same FTRL as that used in Proposition 4.1 without knowing Δ. That is, FTRL automatically adapts to the gap in the agent's decision problems. It should also be noted that, unlike the results presented so far, the agent's outputs x_t's are assumed to be optimal for c^*, which is common in the literature: Besbes et al. (2023) crucially used this condition to derive the $O(\ln T)$ bound on the estimate loss; Bärmann et al. (2017) also assumed the condition to obtain the $O(\sqrt{T})$ bound on the total loss, although it is unnecessary as we have seen in Section 4.

5.1 Proof of Theorem 5.2

Before proving Theorem 5.2, we present a useful inequality derived from the Δ-gap condition.

Lemma 5.3. *Assume that the same conditions as those in Theorem 5.2 hold. For any $t \in \{1, \dots, T\}$, $\hat{c}_t \in \Theta$, and $\hat{x}_t \in \arg\max_{x \in X_t} \langle \hat{c}_t, x \rangle$, it holds that*

$$\|x_t - \hat{x}_t\|^2 \le \frac{KB}{2^{5/4}\sqrt{\lambda}\Delta^2} \langle c^* - \hat{c}_t, x_t - \hat{x}_t \rangle. \tag{8}$$

Proof. Since $\hat{x}_t$ and x_t are optimal for $\hat{c}_t$ and c^*, respectively, we have

$$0 \le \langle \hat{c}_t, \hat{x}_t - x_t \rangle \quad \text{and} \quad \Delta \|x_t - \hat{x}_t\| \le \langle c^*, x_t - \hat{x}_t \rangle,$$

where the latter is due to the Δ-gap condition. Adding these inequalities and squaring both sides yield

$$\Delta^2 \|x_t - \hat{x}_t\|^2 \le \langle c^* - \hat{c}_t, x_t - \hat{x}_t \rangle^2.$$

Since $\langle c^* - \hat{c}_t, x_t - \hat{x}_t \rangle$ is non-negative, the right-hand side is upper bounded by $\frac{KB}{2^{5/4}\sqrt{\lambda}} \langle c^* - \hat{c}_t, x_t - \hat{x}_t \rangle$ due to $\|c^* - \hat{c}_t\|_\star^2 \le \frac{B^2}{2^{5/2}\lambda}$ and $\|x_t - \hat{x}_t\| \le K$, obtaining the desired inequality. ☐

Now we are ready to prove Theorem 5.2.

Proof of Theorem 5.2. We assume that the linearized regret $R_T = \sum_{t=1}^T \langle \hat{x}_t - x_t, \hat{c}_t - c^* \rangle$ is positive; otherwise, it is trivially upper bounded by 0. Recall $g_t = \hat{x}_t - x_t$ for $t = 1, \dots, T$. From (8) in Lemma 5.3, we have

$$\|g_t\|^2 \le \frac{KB}{2^{5/4}\sqrt{\lambda}\Delta^2} \langle g_t, \hat{c}_t - c^* \rangle.$$

Summing over $t = 1, \dots, T$, we obtain

$$\sum_{t=1}^T \|g_t\|^2 \le \frac{KB}{2^{5/4}\sqrt{\lambda}\Delta^2} \sum_{t=1}^T \langle g_t, \hat{c}_t - c^* \rangle = \frac{KB}{2^{5/4}\sqrt{\lambda}\Delta^2} R_T.$$

[9] In general, Fenchel–Young loss L_Ω is strongly convex if Ω is smooth. In our case with $\Omega = I_X$, L_Ω is typically piecewise linear and do not enjoy those properties.

Combining this with the bound (6) on R_T in Proposition 4.1, we obtain

$$R_T \le \frac{2^{5/8}K^{1/2}B^{3/2}}{\lambda^{3/4}\Delta}\sqrt{R_T} \Leftrightarrow R_T \le \frac{2^{5/4}KB^3}{\lambda^{3/2}\Delta^2},$$

thus completing the proof. □

The above proof also utilizes the Fenchel–Young loss perspective at a high level. Observe that the proof is built on two inequalities: the regret bound (6) that depends on $\sum_{t=1}^T \|g_t\|^2$ (or a bound achieved by *adaptive* algorithms (Orabona, 2023, Section 4.2)), namely,

$$\sum_{t=1}^T \left(\ell_t^{\mathrm{sub}}(\hat{c}_t) - \ell_t^{\mathrm{sub}}(c^*)\right) \le \sum_{t=1}^T \langle g_t, \hat{c}_t - c^*\rangle \lesssim \sqrt{\sum_{t=1}^T \|g_t\|^2},$$

and (8) in Lemma 5.3, which, roughly speaking, ensures

$$\|\hat{x} - x_t\|^2 \lesssim \langle \hat{x}_t - x_t, \hat{c}_t - c^*\rangle / \Delta^2.$$

What bridges these two inequalities is the "subgradient as a residual" property of the Fenchel–Young loss in Proposition 2.2, i.e., $g_t = \hat{x}_t - x_t$. With this in mind, we can read (8) as $\|g_t\|^2 \lesssim \langle g_t, \hat{c}_t - c^*\rangle/\Delta^2$ and hence

$$R_T = \sum_{t=1}^T \langle g_t, \hat{c}_t - c^*\rangle \lesssim \sqrt{\sum_{t=1}^T \|g_t\|^2} \lesssim \sqrt{\frac{R_T}{\Delta^2}}.$$

Consequently, R_T is bounded by a term of lower order in R_T itself, obtaining $R_T \lesssim 1/\Delta^2$. This proof strategy, called the *self-bounding technique*, is a powerful tool recently popularized in online learning (Gaillard et al., 2014; Zimmert and Seldin, 2021), and the Fenchel–Young-loss perspective clarifies how it can be adopted.

5.2 Discussion on the Δ-Gap Condition

We discuss situations where the Δ-gap condition in Definition 5.1 is satisfied. Our Δ-gap condition is inspired by Weed (2018), who analyzed error bounds in linear programs (LPs) with the entropic penalty. In Weed (2018), the feasible region, X_t in our notation, is the set of all vertices of a polytope specified by LP constraints. For such X_t, Weed (2018) assumed that the optimal objective value is better than the others by at least $\Delta' > 0$ (i.e., for any $x' \in X_t$, $\max_{x \in X_t}\langle c^*, x\rangle - \langle c^*, x'\rangle$ is zero or at least Δ'). Our Δ-gap condition is slightly more demanding in that it implicitly requires the uniqueness of the optimal solution. Still, if the assumption of Weed (2018) holds with Δ' and the optimal solution x for c^* is unique, our Δ-gap condition holds with $\Delta = \Delta'/K$, where K is the diameter of X_t with respect to $\|\cdot\|$.

As discussed in Weed (2018), such a gap condition holds in LPs with integral polytopes and integral objectives,

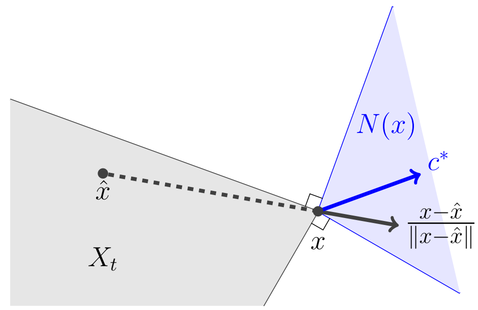

Figure 1: An illustration of the gap condition. The gray area shows polyhedral feasible region X_t. The vertex x is the unique optimal solution for c^* if c^* lies in the interior of the normal cone $N(x)$, shown in blue. The Δ-gap condition requires that the cosine of the angle between c^* and $x - \hat{x}$ is at least $\Delta/\|c^*\|$ for every $\hat{x} \in X_t$; this is true if c^* (the head of the blue arrow) is distant from the boundary of $N(x)$ (the blue lines) by at least Δ.

which often appear in combinatorial optimization, such as the shortest path and bipartite matching problems. Specifically, if $X_t \subseteq \mathbb{Z}^n$ is a set of integral vertices and $c^* \in \mathbb{Z}^n$ holds, there is a gap of $\Delta' \ge 1$ between optimal and suboptimal objective values. Similarly, a gap of $\Delta' \ge 1$ exists in integer LPs with $c^* \in \mathbb{Z}^n$. In such cases, if it additionally holds that the optimal solution x is unique, our Δ-gap condition holds with $\Delta \ge 1/K$.

The Δ-gap condition can be satisfied even if X_t and c^* are not integral. Let $\|\cdot\|$ and $\|\cdot\|_\star$ be the ℓ_2-norm, X_t a polytope, and x a vertex feasible solution, as in Figure 1. Let $N(x) \coloneqq \{\, c \in \mathbb{R}^n \mid \langle c, x - \hat{x}\rangle \ge 0, \forall \hat{x} \in X_t \,\}$ denote the normal cone at x. If c^* lies in the interior of $N(x)$, then x is the unique optimal solution for c^*. The Δ-gap condition, $\langle c^*, \frac{x-\hat{x}}{\|x-\hat{x}\|}\rangle \ge \Delta$ for all $\hat{x} \in X_t$, means that the cosine of the angle between c^* and $x - \hat{x}$ is at least $\Delta/\|c^*\|$. A sufficient condition for this is that the distance between c^* and the closest boundary of $N(x)$ is at least Δ, as in Sakaue et al. (2024, Lemma 18). More precisely, their lemma is concerned with the distance from c^* to the *frontier* defined as follows:

$$F \coloneqq \left\{ c \in \mathbb{R}^n \,:\, \left|\arg\max_{x \in \mathcal{Y}} \langle c, x\rangle\right| \ge 2 \right\},$$

where $\mathcal{Y}$ is the set of vertices of X_t. The is nothing but the boundary of normal cones, $\bigcup_{x \in \mathcal{Y}} \mathrm{bd}(N(x))$, since

$$\begin{aligned} c \in F \iff & \exists x, x' \in \mathcal{Y} \text{ s.t. } x \ne x', \forall \hat{x} \in \mathcal{Y}, \\ & \langle c, x\rangle = \langle c, x'\rangle \ge \langle c, \hat{x}\rangle \\ \iff & \exists x, x' \in \mathcal{Y} \text{ s.t. } x \ne x', \forall \hat{x} \in X_t, \\ & \langle c, x - x'\rangle = 0 \text{ and } \langle c, x\rangle \ge \langle c, \hat{x}\rangle \\ \iff & \exists x \in \mathcal{Y}, c \in \mathrm{bd}(N(x)) \end{aligned}$$

holds, where we used $X_t = \mathrm{conv}(\mathcal{Y})$ and the fact that a (bounded and feasible) LP always has a vertex optimal solution. Therefore, we have $F = \bigcup_{x \in \mathcal{Y}} \mathrm{bd}(N(x))$.

Consequently, Sakaue et al. (2024, Lemma 18) ensures that the Δ-gap condition holds if c^* is distant from boundaries of all $N(x)$ by at least Δ. This consequence is intuitive: c^* close to the boundary of $N(x)$ is ambiguous in that slightly perturbing c^* can lead to a different optimal solution, and the $O(1/\Delta^2)$ regret bound in Theorem 5.2 means that less ambiguous c^* is easier to infer.

6 CONCLUSION

This paper has revisited the online learning approach to inverse linear optimization. We have shown that the suboptimality loss can be seen as a Fenchel–Young loss and that its linearized regret equals the total loss. As a byproduct, we have obtained an offline guarantee on the suboptimality loss without assuming the optimality of the agent's choices. We have also obtained a bound on the cumulative total loss that is independent of T by exploiting the gap condition on the agent's decision problems. We believe that our Fenchel–Young loss perspective, which connects inverse linear optimization and OCO, will be beneficial in synthesizing the research streams and paving the way for further investigations.

There are several directions for future work. For example, analyzing what if we employ other Ω than the indicator function is interesting from the Fenchel–Young loss perspective. Investigating the relation to the *Fitzpatrick loss* (Rakotomandimby et al., 2024), a recent framework for designing tighter surrogate losses than Fenchel–Young losses, will also be interesting. Revealing the lower bound on the cumulative total loss is also an important problem left for future work.

Acknowledgements

The authors are grateful to anonymous reviewers for their constructive feedback. SS is supported by JST ERATO Grant Number JPMJER1903. HB is supported by JST PRESTO Grant Number JPMJPR24K6. TT is supported by JST ACT-X Grant Number JPMJAX210E and JSPS KAKENHI Grant Number JP24K23852.

References

Ahuja, R. K. and Orlin, J. B. (2001). "Inverse optimization". In: *Operations Research* 49.5, pp. 771–783 (cited on page 1).

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). "Finite-time analysis of the multiarmed bandit problem". In: *Machine Learning* 47.2, pp. 235–256 (cited on page 3).

Bärmann, A., Martin, A., Pokutta, S., and Schneider, O. (2020). "An online-learning approach to inverse optimization". In: *arXiv:1810.12997* (cited on pages 2–4, 6).

Bärmann, A., Pokutta, S., and Schneider, O. (2017). "Emulating the expert: Inverse optimization through online learning". In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. PMLR, pp. 400–410 (cited on pages 1–7, 13).

Berthet, Q., Blondel, M., Teboul, O., Cuturi, M., Vert, J.-P., and Bach, F. (2020). "Learning with differentiable pertubed optimizers". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., pp. 9508–9519 (cited on page 3).

Besbes, O., Fonseca, Y., and Lobel, I. (2021). "Online learning from optimal actions". In: *Proceedings of the 34th Conference on Learning Theory*. Vol. 134. PMLR, pp. 586–586 (cited on page 2).

– (2023). "Contextual inverse optimization: Offline and online learning". In: *Operations Research* 0.0, null (cited on pages 2, 7).

Blondel, M., Martins, A. F. T., and Niculae, V. (2020). "Learning with Fenchel–Young losses". In: *Journal of Machine Learning Research* 21.35, pp. 1–69 (cited on pages 2–4).

Burton, D. and Toint, P. L. (1992). "On an instance of the inverse shortest paths problem". In: *Mathematical Programming* 53.1, pp. 45–61 (cited on page 1).

Cesa-Bianchi, N., Conconi, A., and Gentile, C. (2004). "On the generalization ability of on-line learning algorithms". In: *IEEE Transactions on Information Theory* 50.9, pp. 2050–2057 (cited on page 6).

Chan, T. C. Y., Lee, T., and Terekhov, D. (2019). "Inverse optimization: Closed-form solutions, geometry, and goodness of fit". In: *Management Science* 65.3, pp. 1115–1135 (cited on page 1).

Chan, T. C. Y., Mahmood, R., and Zhu, I. Y. (2023). "Inverse optimization: Theory and applications". In: *Operations Research* 0.0, null (cited on page 1).

Chen, V. X. and Kılınç-Karzan, F. (2020). "Online convex optimization perspective for learning from dynamically revealed preferences". In: *arXiv:2008.10460* (cited on pages 1, 2).

Cutkosky, A. (2019). "Anytime online-to-batch, optimism and acceleration". In: *Proceedings of the 36th International Conference on Machine Learning*. Vol. 97. PMLR, pp. 1446–1454 (cited on page 6).

Danskin, J. M. (1966). "The theory of max-min, with applications". In: *SIAM Journal on Applied Mathematics* 14.4, pp. 641–664 (cited on page 4).

Dong, C., Chen, Y., and Zeng, B. (2018). "Generalized inverse optimization through online learning". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., pp. 86–95 (cited on page 2).

Elmachtoub, A. N. and Grigas, P. (2022). "Smart "predict, then optimize"". In: *Management Science* 68.1, pp. 9–26 (cited on page 3).

Gaillard, P., Stoltz, G., and van Erven, T. (2014). "A second-order bound with excess losses". In: *Proceedings of the 27th Conference on Learning Theory*. Vol. 35. PMLR, pp. 176–196 (cited on page 8).

Hazan, E., Agarwal, A., and Kale, S. (2007). "Logarithmic regret algorithms for online convex optimization". In: *Machine Learning* 69.2, pp. 169–192 (cited on pages 3, 7).

Heuberger, C. (2004). "Inverse combinatorial optimization: A survey on problems, methods, and results". In: *Journal of Combinatorial Optimization* 8.3, pp. 329–361 (cited on page 1).

Hu, Y., Kallus, N., and Mao, X. (2022). "Fast rates for contextual linear optimization". In: *Management Science* 68.6, pp. 4236–4245 (cited on page 3).

Huang, R., Lattimore, T., György, A., and Szepesvári, C. (2017). "Following the leader and fast rates in online linear prediction: curved constraint sets and other regularities". In: *Journal of Machine Learning Research* 18.145, pp. 1–31 (cited on page 3).

Iyengar, G. and Kang, W. (2005). "Inverse conic programming with applications". In: *Operations Research Letters* 33.3, pp. 319–330 (cited on page 2).

Jagadeesan, M., Mendler-Dünner, C., and Hardt, M. (2021). "Alternative microfoundations for strategic classification". In: *Proceedings of the 38th International Conference on Machine Learning*. Vol. 139. PMLR, pp. 4687–4697 (cited on page 6).

Lai, T. L. and Robbins, H. (1985). "Asymptotically efficient adaptive allocation rules". In: *Advances in Applied Mathematics* 6.1, pp. 4–22 (cited on page 3).

Mishra, S. K., Raj, A., and Vaswani, S. (2024). "From inverse optimization to feasibility to ERM". In: *Proceedings of the 41st International Conference on Machine Learning*. Vol. 235. PMLR, pp. 35805–35828 (cited on page 2).

Mohajerin Esfahani, P., Shafieezadeh-Abadeh, S., Hanasusanto, G. A., and Kuhn, D. (2018). "Data-driven inverse optimization with imperfect information". In: *Mathematical Programming* 167, pp. 191–234 (cited on page 1).

Orabona, F. (2023). "A modern introduction to online learning". In: *arXiv:1912.13213*. https://arxiv.org/abs/1912.13213v6 (cited on pages 3–8, 12, 13).

Rakotomandimby, S., Chancelier, J.-P., De Lara, M., and Blondel, M. (2024). "Learning with Fitzpatrick losses". In: *Advances in Neural Information Processing Systems*. Vol. 37. Curran Associates, Inc., pp. 79381–79409 (cited on page 9).

Sakaue, S., Bao, H., Tsuchiya, T., and Oki, T. (2024). "Online structured prediction with Fenchel–Young losses and improved surrogate regret for online multiclass classification with logistic loss". In: *Proceedings of the 37th Conference on Learning Theory*. Vol. 247. PMLR, pp. 4458–4486 (cited on pages 3, 8, 9).

Sakaue, S., Tsuchiya, T., Bao, H., and Oki, T. (2025). "Online inverse linear optimization: Improved regret bound, robustness to suboptimality, and toward tight regret analysis". In: *arXiv:2501.14349* (cited on page 2).

Sun, C., Liu, S., and Li, X. (2023). "Maximum optimality margin: A unified approach for contextual linear programming and inverse linear programming". In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. PMLR, pp. 32886–32912 (cited on pages 1, 2).

Tan, Y., Terekhov, D., and Delong, A. (2020). "Learning linear programs from optimal decisions". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., pp. 19738–19749 (cited on page 2).

Tarantola, A. (1988). "Inverse problem theory: Methods for data fitting and model parameter estimation". In: *Geophysical Journal International* 94.1, pp. 167–167 (cited on page 1).

Weed, J. (2018). "An explicit analysis of the entropic penalty in linear programming". In: *Proceedings of the 31st Conference on Learning Theory*. Vol. 75. PMLR, pp. 1841–1855 (cited on page 8).

Wilder, B., Dilkina, B., and Tambe, M. (2019). "Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization". In: *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Vol. 33, pp. 1658–1665 (cited on page 3).

Zimmert, J. and Seldin, Y. (2021). "Tsallis-INF: An optimal algorithm for stochastic and adversarial bandits". In: *Journal of Machine Learning Research* 22.28, pp. 1–49 (cited on page 8).

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] See Algorithm 1 and Section 4.1.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] See Proposition 4.1 for the regret bound. The complexity analysis is omitted since the algorithm is the well-known FTRL.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]
2. For any theoretical claim, check if you include:

(a) Statements of the full set of assumptions of all theoretical results. [Yes] General assumptions are given in Section 2.1. Specific assumptions are stated in each theorem.
(b) Complete proofs of all theoretical results. [Yes] All theoretical results are followed by proofs, where some of them are deferred to the appendix.
(c) Clear explanations of any assumptions. [Yes] General assumptions are explained in Section 2.1, and additional ones needed in gap-dependent analysis are explained in Section 5.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

A Proof of Proposition 4.1

This section presents the proof of Proposition 4.1. The following proof is a slight generalization of the analysis of AdaHedge in Orabona (2023, Section 7.6), which we include here for completeness.

Proof. Based on the notation used in Orabona (2023, Section 7), we define $\psi_t(c) \coloneqq \beta_t(\psi(c) - \min_{c' \in \Theta} \psi(c'))$ and $F_t(c) \coloneqq \psi_t(c) + \sum_{i=1}^{t-1} \langle g_i, c \rangle$ for $t = 1, \dots, T$. Note that replacing the regularizer $\beta_t \psi$ with ψ_t in Algorithm 1 does not affect the choice of $\hat{c}_t$, and hence $\hat{c}_t \in \arg\min_{c \in \Theta} F_t(c)$. Also, $\psi_{t+1}(c) \geq \psi_t(c)$ holds since β_t is non-decreasing.

From Orabona (2023, Lemma 7.1), we have

$$
\begin{aligned}
R_T &= \psi_{T+1}(c^*) - \underbrace{\min_{c \in \Theta} \psi_1(c)}_{=0} + \sum_{t=1}^{T} (F_t(\hat{c}_t) - F_{t+1}(\hat{c}_{t+1}) + \langle g_t, \hat{c}_t \rangle) + \underbrace{F_{T+1}(\hat{c}_{T+1}) - F_{T+1}(c^*)}_{\leq 0} \\
&\leq \beta_{T+1} \left(\psi(c^*) - \min_{c \in \Theta} \psi(c) \right) + \sum_{t=1}^{T} (F_t(\hat{c}_t) - F_{t+1}(\hat{c}_{t+1}) + \langle g_t, \hat{c}_t \rangle) \\
&\leq \beta_{T+1} B^2 + \sum_{t=1}^{T} (F_t(\hat{c}_t) - F_{t+1}(\hat{c}_{t+1}) + \langle g_t, \hat{c}_t \rangle).
\end{aligned} \tag{9}
$$

Below, we derive an upper bound on the second term on the right-hand side.

For now, let $\beta_t = \frac{1}{\alpha} \sqrt{\sum_{i=1}^{t-1} \|g_i\|^2}$, where $\alpha > 0$ is tuned later, and let $D > 0$ denote the diameter of Θ with respect to $\|\cdot\|_\star$. Since ψ_t is $\lambda\beta_t$-strongly convex, Orabona (2023, Lemma 7.8) implies

$$
F_t(\hat{c}_t) - F_{t+1}(\hat{c}_{t+1}) + \langle g_t, \hat{c}_t \rangle \leq \frac{\|g_t\|^2}{2\lambda\beta_t} + \underbrace{\psi_t(\hat{c}_{t+1}) - \psi_{t+1}(\hat{c}_{t+1})}_{\leq 0} \leq \frac{\alpha \|g_t\|^2}{2\lambda \sqrt{\sum_{i=1}^{t-1} \|g_i\|^2}}, \tag{10}
$$

where the right-hand side is infinite if $\sum_{i=1}^{t-1} \|g_i\|^2 = 0$. Also, as in Orabona (2023, Section 7.6), we have

$$
\begin{aligned}
F_t(\hat{c}_t) - F_{t+1}(\hat{c}_{t+1}) + \langle g_t, \hat{c}_t \rangle &\leq F_t(\hat{c}_{t+1}) - F_{t+1}(\hat{c}_{t+1}) + \langle g_t, \hat{c}_t \rangle \\
&= \psi_t(\hat{c}_{t+1}) + \sum_{i=1}^{t-1} \langle g_i, \hat{c}_{t+1} \rangle - \left(\psi_{t+1}(\hat{c}_{t+1}) + \sum_{i=1}^{t} \langle g_i, \hat{c}_{t+1} \rangle \right) + \langle g_t, \hat{c}_t \rangle \\
&= \underbrace{\psi_t(\hat{c}_{t+1}) - \psi_{t+1}(\hat{c}_{t+1})}_{\leq 0} - \langle g_t, \hat{c}_{t+1} \rangle + \langle g_t, \hat{c}_t \rangle \\
&\leq - \langle g_t, \hat{c}_{t+1} \rangle + \langle g_t, \hat{c}_t \rangle \\
&\leq D \|g_t\|.
\end{aligned}
$$

These two inequalities imply

$$
\begin{aligned}
\sum_{t=1}^{T} (F_t(\hat{c}_t) - F_{t+1}(\hat{c}_{t+1}) + \langle g_t, \hat{c}_t \rangle) &\leq \sum_{t=1}^{T} \min \left\{ \frac{\alpha \|g_t\|^2}{2\lambda \sqrt{\sum_{i=1}^{t-1} \|g_i\|^2}}, D \|g_t\| \right\} \\
&= \sum_{t=1}^{T} \sqrt{ \min \left\{ \frac{\alpha^2 \|g_t\|^4}{4\lambda^2 \sum_{i=1}^{t-1} \|g_i\|^2}, D^2 \|g_t\|^2 \right\} } \\
&\leq \sum_{t=1}^{T} \sqrt{ \frac{2}{\frac{4\lambda^2 \sum_{i=1}^{t-1} \|g_i\|^2}{\alpha^2 \|g_t\|^4} + \frac{1}{D^2 \|g_t\|^2}} } \\
&= \sum_{t=1}^{T} \frac{\sqrt{2} \alpha D \|g_t\|^2}{\sqrt{4\lambda^2 D^2 \sum_{i=1}^{t-1} \|g_i\|^2 + \alpha^2 \|g_t\|^2}},
\end{aligned}
$$

where the second inequality uses the fact that the minimum of positive numbers is at most their harmonic mean. Therefore, if $\alpha \ge 2\lambda D$, which we confirm shortly, we have

$$\sum_{t=1}^{T}(F_t(\hat{c}_t) - F_{t+1}(\hat{c}_{t+1}) + \langle g_t, \hat{c}_t\rangle) \le \frac{\sqrt{2}\alpha}{2\lambda}\sum_{t=1}^{T}\frac{\|g_t\|^2}{\sqrt{\sum_{i=1}^{t}\|g_i\|^2}} \le \frac{\sqrt{2}\alpha}{\lambda}\sqrt{\sum_{t=1}^{T}\|g_t\|^2}, \tag{11}$$

where the last inequality is due to Orabona (2023, Lemma 4.13).

Consequently, from (9) and (11), R_T is bounded as

$$R_T \le \left(\frac{B^2}{\alpha} + \frac{\sqrt{2}\alpha}{\lambda}\right)\sqrt{\sum_{t=1}^{T}\|g_t\|^2}.$$

The right-hand side is minimized when $\alpha = \frac{B\sqrt{\lambda}}{2^{1/4}}$, which satisfies $\alpha \ge 2\lambda D$ due to the assumption of $B^2 \ge 2^{5/2}\lambda D^2$. Therefore, Algorithm 1 with $\beta_t = \frac{2^{1/4}}{B}\sqrt{\frac{\sum_{i=1}^{t-1}\|g_i\|^2}{\lambda}}$ attains $R_T \le 2^{5/4}B\sqrt{\frac{1}{\lambda}\sum_{t=1}^{T}\|g_t\|^2}$. $\square$

B Recovering the Same Bound as Bärmann et al. (2017, Theorem 3.3)

As with Bärmann et al. (2017, Theorem 3.3), we assume that the diameter of X_t's with respect to $\|\cdot\| = \|\cdot\|_\infty$ is at most $K > 0$. We define $H > 0$ as a constant such that $\psi(c^*) - \min_{c\in\Theta}\psi(c) \le H^2$. We consider using FTRL (Algorithm 1) with the following choice of β_t:

$$\beta_t = \frac{1}{H\sqrt{\lambda}}\sqrt{K^2 + \sum_{i=1}^{t-1}\|g_i\|^2},$$

which is also non-decreasing. We let $\beta_{T+1} = \beta_T$, which does not affect the analysis (Orabona, 2023, Remark 7.3).

From (9), (10), and $\beta_{T+1} = \beta_T \le \frac{K}{H}\sqrt{\frac{T}{\lambda}}$, which is due to $\|g_i\| = \|\hat{x}_i - x_i\| \le K$, we have

$$R_T \overset{(9)}{\le} \beta_{T+1}H^2 + \sum_{t=1}^{T}(F_t(\hat{c}_t) - F_{t+1}(\hat{c}_{t+1}) + \langle g_t, \hat{c}_t\rangle) \overset{(10)}{\le} \beta_{T+1}H^2 + \sum_{t=1}^{T}\frac{\|g_t\|^2}{2\lambda\beta_t} \le KH\sqrt{\frac{T}{\lambda}} + \sum_{t=1}^{T}\frac{\|g_t\|^2}{2\lambda\beta_t}.$$

The second term on the right-hand side is bounded as

$$\sum_{t=1}^{T}\frac{\|g_t\|^2}{2\lambda\beta_t} = \sum_{t=1}^{T}\frac{\|g_t\|^2}{\frac{2\sqrt{\lambda}}{H}\sqrt{K^2 + \sum_{i=1}^{t-1}\|g_i\|^2}} \le \frac{H}{2\sqrt{\lambda}}\sum_{t=1}^{T}\frac{\|g_t\|^2}{\sqrt{\sum_{i=1}^{t}\|g_i\|^2}} \le \frac{H}{\sqrt{\lambda}}\sqrt{\sum_{t=1}^{T}\|g_t\|^2} \le KH\sqrt{\frac{T}{\lambda}},$$

where we used $\|g_t\| = \|\hat{x}_t - x_t\| \le K$ in the first and last inequalities, and Orabona (2023, Lemma 4.13) in the second inequality, as in (11). Consequently, we obtain $R_T \le 2KH\sqrt{\frac{T}{\lambda}}$.

As discussed in Section 4.1, if Θ is the probability simplex and $\psi\colon \Theta \to \mathbb{R}$ is the negative Shannon entropy, we have $H = \sqrt{\ln n}$ and $\lambda = 1$, recovering $R_T \le 2K\sqrt{T\ln n}$. This bound is exactly the same as that of Bärmann et al. (2017, Theorem 3.3), including the constant factor of 2.

C Proof of Theorem 4.3

Proof. Since the suboptimality loss ℓ^{sub} is convex due to Propositions 2.2 and 3.1, the claim follows from the standard online-to-batch conversion (e.g., Orabona 2023, Theorem 3.1), which we detail below for completeness. Since $\hat{c}_t$ is independent of $\{(X_i, x_i)\}_{i\ge t}$, the law of total expectation implies

$$\mathbb{E}\left[\ell^{\mathrm{sub}}_{X,x}(\hat{c}_t)\right] = \mathbb{E}\left[\mathop{\mathbb{E}}_{(X_t,x_t)\sim\mathcal{D}}\left[\ell^{\mathrm{sub}}_{(X_t,x_t)}(\hat{c}_t) \;\middle|\; \{(X_i, x_i)\}_{i=1}^{t-1}\right]\right] = \mathbb{E}\left[\ell^{\mathrm{sub}}_t(\hat{c}_t)\right].$$

By using this, together with $\mathbb{E}\big[\ell^{\mathrm{sub}}_{X,x}(\hat{c})\big] \le \mathbb{E}\Big[\frac{1}{T}\sum_{t=1}^{T}\ell^{\mathrm{sub}}_{X,x}(\hat{c}_t)\Big]$, which follows from Jensen's inequality, and $\mathbb{E}_{(X,x)\sim\mathcal{D}}[\ell^{\mathrm{sub}}_{(X,x)}(c^*)] = \mathbb{E}[\ell^{\mathrm{sub}}_t(c^*)]$, we obtain

$$\mathbb{E}\big[\ell^{\mathrm{sub}}_{X,x}(\hat{c})\big] - \mathop{\mathbb{E}}_{(X,x)\sim\mathcal{D}}\big[\ell^{\mathrm{sub}}_{X,x}(c^*)\big] \le \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\big(\ell^{\mathrm{sub}}_t(\hat{c}_t) - \ell^{\mathrm{sub}}_t(c^*)\big)\right] = \mathbb{E}\left[\frac{R^{\mathrm{sub}}_T}{T}\right].$$

The bound with Algorithm 1 follows from $R^{\mathrm{sub}}_T \le R_T$ due to Corollary 4.2. $\square$