

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053

ONLINE DECISION-FOCUSED LEARNING

Anonymous authors

Paper under double-blind review

ABSTRACT

Decision-focused learning (DFL) is an increasingly popular paradigm for training predictive models whose outputs are used in decision-making tasks. Instead of merely optimizing for predictive accuracy, DFL trains models to directly minimize the loss associated with downstream decisions. However, existing studies focus solely on scenarios where a fixed batch of data is available and the objective function does not change over time. We instead investigate DFL in dynamic environments where the objective function and data distribution evolve over time. This setting is challenging for online learning because the objective function has zero or undefined gradients—which prevents the use of standard first-order optimization methods—and is generally non-convex. To address these difficulties, we (i) regularize the objective to make it differentiable and (ii) use perturbation techniques along with a near-optimal oracle to overcome non-convexity. Combining those techniques yields two original online algorithms tailored for DFL, for which we establish respectively static and dynamic regret bounds. These are the first provable guarantees for the online decision-focused problem. Finally, we showcase the effectiveness of our algorithms on a knapsack experiment, where they outperform two standard benchmarks.

1 INTRODUCTION.

Many real-world decision problems involve uncertainty, and a common approach to handling it is through the predict-then-optimize framework (Bertsimas and Kallus, 2020). First, a prediction model is trained on historical data; then, its output is fed into an optimization problem to guide decision-making. This natural strategy has been successfully applied in many operation research (OR) problems, ranging from supply chain management (Acimovic and Graves, 2015; Fisher et al., 2016; Ban and Rudin, 2019; Bertsimas and Kallus, 2020) and revenue management (Farias et al., 2013; Ferreira et al., 2016; Cohen et al., 2017; Chen et al., 2022) to healthcare operation (Bertsimas et al., 2013; Aswani et al., 2019; Gupta et al., 2020; Rath et al., 2017), see Mivsić and Perakis (2020) for an extensive review. It is clear that this approach would yield optimal decisions if the predictions were perfectly accurate. However, in practice, prediction errors are inevitable—and even small inaccuracies can propagate through the optimization process, potentially leading to poor decisions.

To address this limitation, an approach known as *decision-focused learning* (Mandi et al., 2024)—also referred to as *smart predict-then-optimize* (Elmachtoub and Grigas, 2022) or *integrated learning-optimization* (Sadana et al., 2025)—has emerged. Instead of optimizing for prediction accuracy alone, this method trains the predictive model to directly minimize the downstream decision loss. By aligning the learning objective with the decision-making goal, it produces models that are more robust to prediction errors in practical applications. While decision-focused learning yields strong empirical performance (Donti et al., 2017; Verma et al., 2022; 2023; Wang et al., 2023), theoretical development has so far been limited to the batch setting, where models are trained on pre-collected, independently and identically distributed (i.i.d.) data (Wilder et al., 2019; Mandi et al., 2022; Shah et al., 2022; Schutte et al., 2024). This assumption breaks down in many real-world scenarios involving dynamic environments (Cheung et al., 2019; Padakandla et al., 2020) and shifting data distributions (Lu et al., 2018; Quiñonero-Candela et al., 2022).

Online learning (Cesa-Bianchi and Lugosi, 2006; Hazan, 2023) provides a general way to cope with such non-stationarity of data-generating processes. This framework considers a learner who makes decisions sequentially, at each round leveraging data collected from previous rounds to inform its next decision. Crucially, the objective function is allowed to vary over time, either in a stochastic or

adversarial way. This provides a natural framework for the work that we present here, which extends decision-focused learning beyond the i.i.d., batch setting.

Contributions. We develop a theoretical foundation for online decision-focused learning, enabling its application in non-stationary settings. This presents a significant technical challenge, as the inherent difficulties of decision-focused learning—such as the non-differentiability of the objective function or the lack of convexity of the losses due to the bi-level nature of the problem—compound those already present in online learning. Our contributions are as follows:

(i) We formalize the online decision-focused learning problem by assuming that at each round, a decision-maker seeks to solve a linear optimization problem over a polytope but does not have access to the true cost function. Thus the decision-maker has to predict the cost using whatever partial information that it has at hand. The cost function is then revealed, and the decision-maker updates its model in a decision-focused fashion. This results in a bi-level optimization problem, where the inner problem consists in making a decision and the outer problem involves optimizing the resulting decision cost.

(ii) We present two algorithms to tackle this problem, *Decision-Focused Follow-the-Perturbed-Leader* (DF-FTPL) and *Decision-Focused Online Gradient Descent* (DF-OGD). While both rely on regularizing the inner problem to make the resulting decision differentiable, they differ on the way they update the parameters of the prediction model. DF-FTPL uses the FTPL approach (Hutter and Poland, 2005), and DF-OGD leverages a variant of Online Gradient Descent (Zinkevich, 2003). We establish sublinear convergence guarantees for both procedures, in the form of a static regret bound for the former and a dynamic regret bound for the latter. To our knowledge, these are the first provable guarantees for the online decision-focused learning problem.

(iii) Finally, we assess the performance of our algorithms on a knapsack experiment inspired by Mandi et al. (2024). Our simulations demonstrate that our approach outperforms the online version of two popular baselines, namely prediction-focused learning and Smart-Predict-then-Optimize, in both static and dynamic environments.

Additional related work. Several decision-focused approaches have been proposed in the batch setting, among which differentiating the associated KKT conditions Gould et al. (2016); Amos and Kolter (2017); Donti et al. (2017); Wilder et al. (2019); Mandi and Guns (2020), smoothing via random perturbation Berthet et al. (2020), building surrogate losses via duality Elmachtoub and Grigas (2022) or directional gradients Huang and Gupta (2024) and relying on pairwise ranking techniques Mandi et al. (2022), see Mandi et al. (2024); Sadana et al. (2025) for additional references. However, the extension of decision-focused learning to the online setting remains unexplored. While several recent studies address online bi-level optimization Shen et al. (2023); Tarzanagh et al. (2024); Lin et al. (2023), none of them are applicable to decision-focused learning as they rely on restrictive smoothness assumptions, that are incompatible with the structure of decision-focused problems. In particular, the decision-focused objective is typically non-convex and features gradients that are either zero or undefined, owing to its underlying linear structure.

This motivates the use of different online methods, specifically tailored to address these problems. On the one hand, lack of differentiability is usually tackled through zero-th order methods Héliou et al. (2020); Frezat et al. (2023), sub-gradient Duchi et al. (2011), proximal Dixit et al. (2019) or smoothing approaches Abernethy et al. (2014). On the other hand, methods to address non-convexity in online learning often rely on the existence of a near optimal oracle Kalai and Vempala (2005); Agarwal et al. (2019); Suggala and Netrapalli (2020); Xu and Zhang (2024) or additional smoothness conditions Lesage-Landry et al. (2020); Ghai et al. (2022). All those results are derived for *static regret* (Zinkevich, 2003) which compares the learned predictors to the best static strategy and is defined in Section 2. A more challenging criterion, also described in Section 2 is *dynamic regret*, introduced in Zinkevich (2003) and later developed in Hall and Willett (2013); Besbes et al. (2015); Zhao et al.; Zhao and Zhang (2021) among others, which takes into account the evolution of the environment. An alternative approach to obtain guarantees in the online non-convex setting is to weaken static regret to *local regret* Hazan et al. (2017); Aydore et al. (2019); Zhuang et al. (2020); Hallak et al. (2021), which is the sum of the objective gradient norms evaluated in the iterates over time. Minimizing local regret corresponds to encouraging convergence toward stationary points. However as we shall see, this approach is not sensible in our framework, as objective gradients are either zero or undefined.

108 **Outline.** In Section 2, we introduce the online decision-focused problem, our notions of regret
 109 and our assumptions. In Section 3, we present the DF-FTPL and DF-OGD algorithms, before deriving
 110 bounds on the static regret of the former and the dynamic regret of the latter. Finally, we present our
 111 experiment in Section 4.

113 **Notation.** For a differentiable map $\varphi : \mathbb{R}^m \rightarrow \mathcal{Y}$ such that $\mathcal{Y} \subseteq \mathbb{R}$, $\nabla \varphi(x)$ denotes the gradient of
 114 φ at $x \in \mathbb{R}^m$ and $\nabla^2 \varphi(x)$ denotes the Hessian of φ . In the case where $\mathcal{Y} \subseteq \mathbb{R}^d$, $\nabla \varphi(x)$ denotes the
 115 Jacobian of φ . For two vectors $(v, w) \in \mathbb{R}^d \times \mathbb{R}^d$, $v \succ w$ means that $v_i \geq w_i$ for any $i \in [d]$, and
 116 $\langle v, w \rangle = v^\top w$ refers to the standard Euclidian inner product. Also, $\|v\| = \sqrt{\langle v, v \rangle}$ is the standard
 117 euclidian norm. Given a compact convex set $\Theta \subseteq \mathbb{R}^d$, Π_Θ denotes the orthogonal projection onto
 118 Θ . For a matrix $M \in \mathbb{R}^{m \times d}$, $\|M\|_{\text{op}}$ refers to its L^2 -operator norm. In the case $d = m$, $\lambda_{\min}(M)$
 119 refers to its lowest real eigenvalue and $\lambda_{\max}(M)$ its largest real eigenvalue. For $(x, y) \in \mathbb{R}^2$, $x \propto y$
 120 means that there exists $\lambda \in \mathbb{R}^*$ such that $x = \lambda y$.

2 FRAMEWORK

124 **Sequential decision-making.** We consider an online decision-making problem over $T > 0$ peri-
 125 ods, defined for any $t \in [T]$ as

$$\min_{w \in \mathcal{W}} \langle \bar{g}_t(X_t), w \rangle, \quad (1)$$

128 where $\mathcal{W} = \text{Conv}(v_1, \dots, v_K)$ is a bounded convex polytope of \mathbb{R}^d with non-empty interior and
 129 vertices $\{v_i\}_{i=1}^K$. This feasible set appears naturally in many problems such as shortest-path (Gallo
 130 and Pallottino, 1988), portfolio selection (Li and Hoi, 2014) or mixed strategy design in games
 131 (Syrgkanis et al., 2015). In (1), X_t are random covariates and $\bar{g}_t : \mathcal{X} \rightarrow \mathbb{R}^d$, is a deterministic
 132 cost function, which satisfies for instance $\bar{g}_t(X_t) = \mathbb{E}[Z_t | X_t]$ for some hidden state $Z_t \in \mathbb{R}^d$. At
 133 each period t , nature picks both a distribution for X_t and a cost function \bar{g}_t . This corresponds
 134 to the stochastic adversary setting (Rakhlin et al., 2011). Importantly, X_t is revealed at the
 135 beginning of the round, but not $\bar{g}_t(X_t)$ which is only available at the end of the round. While $\bar{g}_t(X_t)$
 136 is unknown to the decision-maker at the decision time, they have access to a family of models
 137 $g : \Theta \times \mathcal{X} \rightarrow \mathbb{R}^d$, parameterized by $\Theta \subset \mathbb{R}^m$, to predict it. Then, the general form of the decision-
 138 making dynamics we consider can be described as follows: at each round $t \in [T]$, for a horizon
 139 $T \in \mathbb{N}$, given the current parameter θ_t ,

1. Nature picks a distribution for $X_t \in \mathcal{X}$ and a cost function $\bar{g}_t : \mathcal{X} \rightarrow \mathcal{Z}$.
2. The decision-maker observes X_t and compute its prediction as $g(\theta_t, X_t)$.
3. Then, they take an action minimizing the resulting predicted cost

$$w_t = w_t^*(\theta_t) \in \operatorname{argmin}_{w \in \mathcal{W}} \langle g(\theta_t, X_t), w \rangle \quad (2)$$

4. Finally, the decision-maker observes $\bar{g}_t(X_t)$ and update θ_{t+1} for the next round.

148 Formally, the decision-maker considers a joint process $(\theta_t, w_t)_{t \in [T]}$, starting from $\theta_1 \in \Theta$ and a his-
 149 tory $\mathcal{H}_0 = \emptyset$, and defined by the following recursion. At round $t \in [T]$, the decision-maker takes the
 150 best action given the collected history \mathcal{H}_{t-1} , the current estimate θ_t and a new feature X_t as in (2).
 151 Then, they observe $\bar{g}_t(X_t)$ and update their history $\mathcal{H}_t = \mathcal{H}_{t-1} \cup \{(X_t, \bar{g}_t(X_t), \theta_t, w_t^*(\theta_t))\}$. Fi-
 152 nally, they update their prediction parameter θ_{t+1} based on \mathcal{H}_t through an algorithm Alg_t . A central
 153 question for the decision-maker is then the choice of algorithms $\{\text{Alg}_t\}_{t \in [T]}$ to fit their regression
 154 model, that is how to pick $\theta_t \in \Theta$ for each $t \in [T]$.

155 **Online decision-focused learning.** From an online perspective, *prediction-focused learning* con-
 156 sists in selecting at each $t \in [T]$ an algorithm Alg_t to estimate $\operatorname{argmin}_{\theta \in \Theta} R_t(\theta)$, where R_t is
 157 a statistical risk based on the historical observations \mathcal{H}_{t-1} , typically chosen as the empirical risk
 158 $R_t(\theta) = \sum_{i=1}^t \ell(\bar{g}_i(X_i), g(\theta, X_i))$, for some loss function ℓ (e.g., cross-entropy or squared error).

160 Alternatively, we consider in this paper, the *decision-focused learning* approach, which selects pa-
 161 rameters $\theta \in \Theta$ by directly minimizing the downstream objective instead of a risk function. Specif-
 162 ically, the decision-maker chooses Alg_t to solve $\operatorname{argmin}_{\theta \in \Theta} \langle \bar{g}_t(X_t), w_t^*(\theta) \rangle$ where w_t^* is defined in

(2). From the previous formulation, we see that the decision-focused learning formulation corresponds to a bilevel optimization problem (Colson et al., 2007; Sinha et al., 2017; Ji et al., 2021).

We emphasize that no stationary assumptions are made on the process $\{X_t, \bar{g}_t(X_t)\}_{t \in [T]}$. **In other words, the adversary may select any distribution for X_t and any function \bar{g}_t (as long as H2 below is satisfied).** To fully accommodate this flexibility, we assess the optimality of $\{\theta_t\}_{t \in [T]}$ with two notions of regret. Let

$$f_t : \theta \in \Theta \mapsto \langle \bar{g}_t(X_t), w_t^*(\theta) \rangle, \quad (3)$$

denotes the loss incurred when taking an action based on the prediction parameter $\theta_t \in \Theta$. Inspired from Zinkevich (2003), we consider the notions of static and dynamic regret to measure the effectiveness of a learning strategy $\{\theta_t\}_{t \in [T]}$, which are respectively:

$$\mathfrak{R}_T^s = \sum_{t \in [T]} f_t(\theta_t) - \inf_{\theta \in \Theta} \sum_{t \in [T]} \mathbb{E}[f_t(\theta)] \quad \text{and} \quad \mathfrak{R}_T^d = \sum_{t \in [T]} F_t(\theta_t) - \sum_{t \in [T]} \inf_{\theta \in \Theta} F_t(\theta), \quad (4)$$

where $F_t : \theta \mapsto \mathbb{E}[f_t(\theta) \mid \mathcal{H}_{t-1}]$. **In our dynamic regret, the sequence of actions is compared against a sequence of oracles, each minimizing the instantaneous loss. Without taking the conditional expectation over \mathcal{H}_{t-1} , each comparator could overfit to the specific realization of X_t resulting in an unrealistically strong and unattainable benchmark. By considering the conditional expectation of the loss, we effectively regularize the dynamic comparators, making them meaningful competitors.** Note that our static regret compare to the best fixed strategy, with respect to the averaged losses. This notion makes sense here as we aim to control those regrets in expectation over the randomness of the process.

We make the following mild regularity assumptions for the rest of the analysis.

H1. (i) $\Theta \subset \mathbb{R}^m$ is a compact and convex set with diameter $D_\Theta < \infty$. (ii) For any $\theta \in \Theta$ and $t \in [T]$, $\nabla_\theta g(\theta, X_t) \in \mathbb{R}^{m \times d}$ is continuously differentiable and $\|\nabla_\theta g(\theta, X_t)\|_{\text{op}} \leq G < \infty$ almost-surely. (iii) For any $t \in [T]$, $\|\bar{g}_t(X_t)\| \leq D_Z < \infty$ almost-surely.

It is common in online learning to assume boundedness of the parameter space, model gradient and prediction space (Boyd, 2004; Bishop and Nasrabadi, 2006). Note that in the well-specified setting, i.e., $\bar{g}_t = g(\bar{\theta}_t, \cdot)$ for some $\bar{\theta}_t$, these two former conditions automatically imply the latter. We emphasize however that we do not assume this realizability condition, in contrast to most of the literature on decision-focused learning (Bennouna et al., 2024).

Without further assumptions on the sequence of costs \bar{g}_t and models g , the problem can still be made arbitrarily hard. Therefore, we make an assumption coming from the classification (Mammen and Tsybakov, 1999; Tsybakov, 2004) and bandit literature (Zeevi and Goldenshluger, 2009; Perchet and Rigollet, 2013) about the *margins* of the cost function. Recall that we denote by $\{v_i\}_{i=1}^K$ the vertices of \mathcal{W} . We define for any $x \in \mathcal{X}$ and $i \in [K]$, $u_i(\theta, x) = \langle g(\theta, x), v_i \rangle$.

H2. There exist $C_0 > 0$ and $\beta \in [0, 1]$, such that almost surely for any $t \in [T]$, $\theta \in \Theta$ and $\varepsilon \in [0, 1]$,

$$\mathbb{P}\left(\inf_{j \neq I_t(\theta)} \{u_j(\theta, X_t) - u_{I_t(\theta)}(\theta, X_t)\} \geq \varepsilon \mid \mathcal{H}_{t-1}\right) \geq 1 - C_0 \varepsilon^\beta,$$

where $I_t(\theta) \in \operatorname{argmin}_{i \in [K]} u_i(\theta, X_t)$.

H2 controls how difficult it is to solve problem (2) since it determines the objective gap between the optimal vertex $v_{I_t(\theta)} \in \mathcal{W}$ and the other vertices. In other words, it quantifies how identifiable is the optimal vertex, $\varepsilon = 0$ meaning it is not distinguishable from the others. If it is satisfied for $\beta \geq 1$, then it is automatically satisfied for $\beta = 1$ thus we do not loose any generality restraining β to $[0, 1]$. This assumption is critical in our analysis, as it allows bounding the expected distance between the actual optimal decision $w_t^*(\theta)$ and the regularized approximation $\tilde{w}_t(\theta)$ introduced in (5) below.

We expect **H2** to hold in a wide variety of classical statistical settings. For example, in Appendix A we show that it is satisfied when $g : (\theta, X) \mapsto X\theta$ where X has i.i.d standard Gaussian columns.

3 ONLINE ALGORITHMS FOR DECISION-FOCUSED LEARNING .

On the need of regularization to get differentiation. From an online learning perspective, a natural approach to minimize the two regrets defined in (4) would be to update θ_{t+1} at each round t

216 using a variant of either the Follow-The-Leader algorithm (FTL, [Kalai and Vempala, 2005](#)) or Online
 217 Gradient Descent (OGD, [Zinkevich, 2003](#)), applied to the objective function f_t defined in (3), thus
 218 specifying the choice of Alg_t . However, these algorithms are designed for single-level optimization
 219 problems and require access to (sub-)gradients of the objective. In our setting, this requirement
 220 is problematic because the function f_t does not admit an informative gradient: by construction,
 221 computing the gradient of f_t involves differentiating the mapping $\theta \mapsto w_t^*(\theta)$, which is generally
 222 not (sub-)differentiable. Indeed, w_t^* minimizes a linear function over the convex polytope \mathcal{W} , thus
 223 $\theta \mapsto w_t^*(\theta)$ ranges in the finite set of vertices $\{v_1, \dots, v_K\}$.

224 To address this issue, following [Wilder et al. \(2019\)](#), we propose to add a regularizer \mathcal{R} to the
 225 objective function in (2). Accordingly, we define for any $\theta \in \Theta$,

$$227 \quad \tilde{w}_t(\theta) \in \operatorname{argmin}_{w \in \mathcal{W}} \{ \langle g(\theta, X_t), w \rangle + \alpha_t \mathcal{R}(w) \}, \quad (5)$$

228 which is a regularized approximation of w_t^* . When this surrogate is continuously differentiable—which holds for our choices of \mathcal{R} in Section 3—we can use $\nabla \tilde{w}_t(\theta)$ in our algorithmic
 229 routine. This approach amounts to minimizing the surrogate
 230

$$232 \quad \tilde{f}_t : \theta \mapsto \langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle, \quad (6)$$

233 instead of f_t , which admits gradients of the form $\nabla \tilde{f}_t(\theta) = \nabla \tilde{w}_t(\theta)^\top \bar{g}_t(X_t)$. Note that there is a
 234 natural trade-off in the choice of the regularization parameter α_t . On the one hand, choosing a large
 235 α_t makes the function $\tilde{w}_t(\theta)$ smoother. On the other hand, the larger α_t is, the more \tilde{w}_t deviates
 236 from the true function w_t^* that we aim to approximate. As we will see later, it is possible to balance
 237 these two extremes by carefully tuning the parameter α_t .
 238

239 **On the choice of regularization on a general polytope.** In what follows, we write \mathcal{W} as the
 240 intersection of n half-spaces, that is $\mathcal{W} = \{w \in \mathbb{R}^d : A^\top w - b \leq 0\}$ where $n > 0$ is the number of
 241 faces, $A \in \mathbb{R}^{d \times n}$ and $b \in \mathbb{R}^n$. We assume that \mathcal{W} is not degenerated, *i.e.* $AA^\top \in \mathbb{R}^{d \times d}$ is full-rank.
 242

243 It remains to determine what regularizer \mathcal{R} to choose in (6). We recall that our aim is to obtain a
 244 $\tilde{w}_t(\theta)$ in (5) which is differentiable for any $\theta \in \Theta$. A possible strategy is to choose \mathcal{R} so \tilde{w}_t remains
 245 in the strict interior of \mathcal{W} . In this case, $\tilde{w}_t(\theta)$ is differentiable in a neighborhood of any $\theta \in \Theta$ and
 246 admits a close-form Jacobian $\nabla \tilde{w}_t(\theta)$ by the implicit function theorem. A natural choice to force
 247 \tilde{w}_t to remain in the interior of \mathcal{W} is the corresponding log-barrier function:

$$248 \quad \mathcal{R} : w \mapsto - \sum_{i=1}^n \ln(b_i - A_i^\top w), \quad (7)$$

251 where $A_i \in \mathbb{R}^d$ is the i -th column of A . With this choice of regularization, we show in Lemma 4
 252 that $\nabla \tilde{w}_t$ has an explicit formulation, allowing its use in practice.
 253

254 **Remark 1.** We remark that in the special case where $\mathcal{W} = \{w \in \mathbb{R}^d : w \succcurlyeq 0, \mathbf{1}^\top w = 1\}$ is the
 255 simplex of \mathbb{R}^d , an alternative choice for \mathcal{R} in (5) is the negative entropy $\mathcal{R}_0 : w \mapsto \sum_{i \in [d]} w_i \ln(w_i)$.
 256 In this case, $\tilde{w}_t(\theta)$ in (5) reduces to the softmax mapping, which is differentiable and admits a close-
 257 form Jacobian. In Appendix C, we theoretically study the performances of our algorithms in this
 258 special case.

259 **Approximate oracles to handle non-convexity.** The bi-level structure of the problem yields
 260 unexpected properties. In particular, even when the regularizer \mathcal{R} is strongly-convex in (26)—
 261 which the case with both log-barriers and negative entropy—and the model g is simple (such as
 262 $g(\theta, X) = X\theta$), $\tilde{f}_t : \theta \mapsto \langle \tilde{w}_t(\theta), \bar{g}_t(X_t) \rangle$ may not be convex, but is lipschitz (see Lemmas 1
 263 and 5). This prevents us from directly using known online convex optimization algorithms ([Hazan, 2023](#)). However, a recent line of research ([Agarwal et al., 2019](#); [Suggala and Netrapalli, 2020](#);
 264 [Xu and Zhang, 2024](#)) has developed online algorithms in the non-convex case, provided the losses
 265 are Lipschitz-continuous—which is the case with log-barriers or negative entropy (see Lemmas 1
 266 and 5). These studies combine near-optimal oracles with perturbation techniques to establish sub-
 267 linear bounds on the expected regret. In line with this literature, we also assume to have access to
 268 an *approximate offline optimization oracle*, a notion that appeared in a slightly modified form in
 269 [Suggala and Netrapalli \(2020\)](#).

270 **Definition 1.** An ξ -approximate offline optimization oracle (or approximate oracle), adapted to a
 271 class \mathcal{C} is a mapping \mathbf{O}_ξ taking a function $f \in \mathcal{C}$ such that $f : \Theta \rightarrow \mathbb{R}$, and outputting $\vartheta = \mathbf{O}_\xi(f) \in$
 272 Θ satisfying:

$$273 \quad 274 \quad f(\vartheta) \leq \inf_{\theta \in \Theta} f(\theta) + \xi. \quad (8)$$

275 The notion of an approximate oracle formalizes that, in non-convex settings, we cannot rely on
 276 subroutines that provably attain a global minimizer of \tilde{f}_t at time t (as offline gradient descent would
 277 in convex problems). Instead, we must settle for local minimizers, whose quality is characterized
 278 by a parameter ξ —which vanishes in favorable loss landscapes. As a concrete example, consider
 279 the stochastic gradient descent (SGD) algorithm as \mathbf{O}_ξ . A large body of work has shown that, even
 280 when Θ is a class of deep neural networks, SGD can converge to local minimizers (Ghadimi and
 281 Lan, 2013; Mertikopoulos et al., 2020; Patel and Zhang, 2021; Cutkosky et al., 2023), justifying that
 282 \mathbf{O}_ξ is indeed an approximate oracle. A more detailed discussion is provided in Appendix B.

283 We conclude this remark by mentioning that it is possible to conduct a non-convex analy-
 284 sis without an approximate oracle by considering the weaker notion of *local regret* defined as
 285 $\sum_{t \in [T]} \|\nabla F_t(\theta_t)\|$; see (Hazan et al., 2017; Aydore et al., 2019; Zhuang et al., 2020; Hallak et al.,
 286 2021). However, such a definition is not meaningful in the DFL framework, where gradients are
 287 zero or undefined due to the structure of the problem.

288 The combined use of approximate oracles and regularization allows us to derive original online
 289 algorithms. In Section 3.1, we focus on a variant of the FTL algorithm, enjoying a static regret
 290 bound, while in Section 3.2 we propose a version of OGD enjoying a dynamic regret bound, both
 291 results holding in the non-convex case.

293 3.1 DECISION-FOCUSED FOLLOW THE PERTURBED LEADER

295 Our first algorithm is displayed in Algorithm 1, where $\text{Exp}(\eta)$ refers to an exponential distribution
 296 with parameter $\eta > 0$. It is inspired from the classic Follow-the-Leader approach (Kalai and Vem-
 297 pala, 2005), which consists in making a decision minimizing the sum of objective functions observed
 298 so far. When losses are non-convex, it is common to inject random noise at each step to regularize
 299 the total cost function, resulting in an approach known as Follow-the-Perturbed-Leader (Hutter and
 300 Poland, 2005; Suggala and Netrapalli, 2020). We employ this strategy as Alg_t to update θ_{t+1} in
 301 Algorithm 1. Note that the oracle in line 6 of Algorithm 1 crucially works with the *regularized*
 302 losses $\tilde{f}_1, \dots, \tilde{f}_t$. This allows to use in practice gradient-based methods to obtain an approximate
 303 minimizer as per our discussion in Section 3, since these surrogate losses are differentiable and
 304 Lipschitz.

306 **Algorithm 1** Decision-Focused Follow The Perturbed Leader DF-FTPL

307 1: **Input:** horizon $T > 0$, initialization $\theta_1 \in \Theta$, ξ -approximate oracle \mathbf{O}_ξ and history $\mathcal{H}_0 = \emptyset$.
 308 2: **for** each $t \in \{1, \dots, T\}$ **do**
 309 3: Observe $X_t \in \mathcal{X}$ and play $w_t^*(\theta_t) = \operatorname{argmin}_{w \in \mathcal{W}} \langle g(\theta_t, X_t), w \rangle$.
 310 4: Observe $\bar{g}_t(X_t) \in \mathcal{Z}$ and update the history \mathcal{H}_t .
 311 5: Draw $\sigma_t \in \mathbb{R}^d$ such that for all $j \in [d]$, the j -th component $\sigma_{j,t} \sim \text{Exp}(\eta)$.
 312 6: Update

$$313 \quad 314 \quad \theta_{t+1} = \mathbf{O}_\xi \left(\sum_{i=1}^t \tilde{f}_i - \langle \sigma_t, \cdot \rangle \right)$$

315 7: **end for**

317 **Note that in line 3, the algorithm takes the best action $w_t^*(\theta_t)$ given the predicted cost $g(\theta_t, X_t)$**
 318 **(the regularized action \tilde{w}_t is only needed to compute \tilde{f}_t).** In most practical settings, w_t^* can be
 319 computed efficiently with standard numerical solvers. However, we show in Appendix F that only
 320 being able to determine $\underline{w}_t(\theta_t) \in \mathcal{W}$ such that $\langle g(\theta_t, X_t), \underline{w}_t(\theta_t) \rangle - \langle g(\theta_t, X_t), w_t^*(\theta_t) \rangle \leq \kappa$ for
 321 some $\kappa > 0$ only shifts the regret bounds of the next section by κ .

323 We provide a theoretical guarantee on the convergence of Algorithm 1 in the form of the following
 324 static regret bound.

324 **Theorem 1.** Assume **H1**, **H2** and having access to an ξ -approximate oracle \mathbf{O}_ξ adapted to
 325 $\{\sum_{i=1}^t \tilde{f}_i - \langle \sigma_t, \cdot \rangle\}_{t \in [T]}$. Let $\{\theta_t\}_{t \in [T]}$ be the output of DF-FTPL (Algorithm 1) instantiated with
 326 learning step $\eta > 0$ and regularization coefficients $\alpha_t = \alpha > 0$ for any t . Then:

$$328 \quad T^{-1} \mathbb{E}[\mathfrak{R}_T^s] = \tilde{\mathcal{O}}\left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi + \alpha n\right),$$

330 where \mathbb{E} denotes the expectation on both data and the intrinsic randomness of DF-FTPL and $\tilde{\mathcal{O}}$
 331 contains polynomial dependency in $\ln(1/\alpha), \ln(\ln(d))$.
 332 Furthermore, taking $\eta \propto m^{1/4} T^{-3/4} n^{-1/2}$ and $\alpha \propto m^{3/4} n^{1/2} T^{-1/4}$ yields:

$$334 \quad T^{-1} \mathbb{E}[\mathfrak{R}_T^s] = \tilde{\mathcal{O}}\left(m^{3/4} \sqrt{n} T^{-1/4} + \xi\right).$$

336 The proof of Theorem 1 can be found in Appendix G.2. In particular, Theorem 1 shows that
 337 DF-FTPL enjoys an average regret bound decaying in $T^{-1/4}$ as long as $\xi = \mathcal{O}(T^{-1/4})$. While
 338 it features a polynomial dependency $m^{3/4}$ on the dimension of Θ , it only depends on the dimension
 339 of the decision space \mathcal{W} through a $\ln \ln(d)$ term. This makes Algorithm 1 a particularly competitive
 340 approach when the decision space is of high dimension.

341 It is informative to compare our guarantee to existing bounds in the literature. [Suggala and Netrapalli \(2020\)](#), who also study non-convex online learning, achieves a rate of $\mathcal{O}(T^{-1/2})$ as long
 342 as their offline oracle satisfies $\xi = \mathcal{O}(T^{-1/2})$, at the cost of a degraded $m^{3/2}$ dependency on the
 343 dimension. Their faster rate in T comes from the fact that they tackle a simpler, single-level problem
 344 as compared to our bi-level, non differentiable setting. Indeed, we need to regularize the inner
 345 problem to overcome non-differentiability as discussed in Section 3. This introduces an additional
 346 trade-off on the regularization strength α on top of the usual trade-off in the learning rate η , which
 347 is reflected in our rate.

350 3.2 DECISION-FOCUSED ONLINE GRADIENT DESCENT

351 While DF-FTPL (Algorithm 1) benefits from a converging static regret bound, the techniques in-
 352 volved in [Suggala and Netrapalli \(2020\)](#) are not enough to reach dynamic regret guarantees, which is
 353 particularly relevant in highly non-stationary environments where the optimal decisions may change
 354 significantly from one round to another. To this end, we go beyond the Follow-the-Leader approach
 355 and propose an original algorithm based on the celebrated Online Gradient Descent ([Zinkevich, 2003](#)). This procedure, which we call DF-OGD, is presented in Algorithm 2.

356 In words, at each time $t \in [T]$, Algorithm 2 first makes a decision $w_t^*(\theta_t) \in \mathcal{W}$ based on the
 357 current parameters $\theta_t \in \Theta$. Then, it observes $\bar{g}_t(X_t) \in \mathcal{Z}$, determines the surrogate objective \tilde{f}_t and
 358 computes a near-minimizer via the offline oracle from Definition 1. Next, to handle non-convexity,
 359 it determines the gradient $\nabla \tilde{w}_t(u_t)$, evaluated in a point u_t drawn uniformly at random in $[\theta_t, \vartheta_t]$.
 360 From there, it updates θ_{t+1} through the classical gradient step scheme of OGD.

361 The main difference with Algorithm 1 is the update of θ_{t+1} through Alg_t . On the one hand, DF-FTPL
 362 invokes \mathbf{O}_ξ to minimize the cumulative loss observed so far and perturbs the entire objective func-
 363 tion. On the other hand, DF-OGD relies only on a near-optimizer of the most recent regularized cost,
 364 $\vartheta_t = \mathbf{O}_\xi(\tilde{f}_t)$, and perturbs the point at which the descent direction (gradient) is evaluated.

365 Moreover, note that Algorithm 1 is instantiated with a single pair of parameters (α, η) whereas
 366 Algorithm 2 uses on a sequence $(\alpha_t, \eta_t)_{t \in [T]}$. This additional flexibility is crucial for the algorithm
 367 to adapt to the variation of the problem so as to maintain a low dynamic regret.

368 We provide a convergence guarantee for Algorithm 2 with the following dynamic regret bound.

369 **Theorem 2.** Assume **H1**, **H2**, access to a ξ -approximate oracle adapted to $\{\tilde{f}_t\}_{t \in [T]}$. Let $\{\theta_t\}_{t \in [T]}$ be the output of DF-OGD (Algorithm 2) instantiated with the non-increasing sequence $(\eta_t)_{t \in [T]}$ and regularization coefficients $(\alpha_t)_{t \in [T]}$. Then:

$$375 \quad T^{-1} \mathbb{E}[\mathfrak{R}_T^d] = \tilde{\mathcal{O}}\left(\mathbb{E}\left[\frac{1 + P_T}{T \eta_T} + \frac{1}{T} \sum_{t \in [T]} \frac{\eta_t}{\alpha_t^2} + n \alpha_t\right] + \xi\right).$$

Algorithm 2 Decision-Focused Online Gradient Descent (DF-OGD)

```

378 1: Input: horizon  $T > 0$ , initialization  $\theta_1 \in \Theta$  and history  $\mathcal{H}_0 = \emptyset$ .
379 2: for each  $t \in \{1, \dots, T\}$  do
380 3:   Observe  $X_t \in \mathcal{X}$  and play  $w_t^*(\theta_t) = \operatorname{argmin}_{w \in \mathcal{W}} \langle g(\theta_t, X_t), w \rangle$ .
381 4:   Observe  $\bar{g}_t(X_t) \in \mathcal{Z}$  and update the history  $\mathcal{H}_t$ .
382 5:   Get  $\vartheta_t = \mathbf{O}_\xi(f_t)$ .
383 6:   Draw  $\delta_t \sim \text{Unif}([0, 1])$ , compute  $u_t = \vartheta_t + \delta_t(\theta_t - \vartheta_t)$  and  $\tilde{\nabla}_t(u_t) = \nabla \tilde{w}_t(u_t)^\top \bar{g}_t(X_t)$ .
384 7:   Update  $\theta_{t+1} = \Pi_\Theta(\theta_t - \eta_t \tilde{\nabla}_t(u_t))$ .
385 8: end for
386
387
388

```

389 where $P_T = \sum_{t=1}^{T-1} \|\vartheta_{t+1} - \vartheta_t\|$, \mathbb{E} denotes the expectation on both data and the intrinsic randomness of DF-OGD and $\tilde{\mathcal{O}}$ contains polynomial dependency in $\ln(1/\alpha)$, $\ln(\ln(d))$.

390 Furthermore, assume $(t^{-1}(1 + P_t))_{t \geq 1}$ is non-increasing almost surely with $P_t =$
391 $\sum_{s=1}^t \|\vartheta_{s+1} - \vartheta_s\|$. Then, using $\alpha_t \propto n^{-1/2} t^{-1/4} (1 + P_t)^{1/4}$ and $\eta_t \propto n^{-1/2} t^{-3/4} (1 + P_t)^{3/4}$ for
392 any $t \in [T]$ leads to :

$$393 T^{-1} \mathbb{E}[\mathfrak{R}_T^d] = \tilde{\mathcal{O}}\left(\mathbb{E}\left[\sqrt{n}(1 + P_T)^{1/4} T^{-1/4}\right] + \xi\right).$$

394 The complete statement of the main bound, including constants is provided in Equation (48).
395 Our dynamic regret bound naturally depends on P_T , which captures the problem's variability
396 by measuring the cumulative distance between approximate minimizers over time. When $\xi =$
397 $\mathcal{O}((1 + P_T)^{1/4} T^{-1/4})$, the average dynamic regret decreases at the rate $\mathcal{O}((1 + P_T)^{1/4} T^{-1/4})$.
398 Notably, the bound is independent of the dimension of Θ , and depends only mildly on the dimension
399 of \mathcal{W} through a $\ln \ln(d)$ factor. This constitutes a key strength of the optimistic strategy underlying
400 DF-OGD, making it particularly well-suited for high-dimensional settings.

401 It is instructive to compare our guarantee with those established in recent studies on dynamic regret.
402 For instance, [Zhang et al. \(2018\)](#) derive a $\mathcal{O}(T^{-1/2} \sqrt{T(P_T + 1)}$ bound. However, their setting is
403 not directly comparable to ours, as they consider a simpler single-level problem with differentiable
404 and convex objectives. In contrast, our framework involves non-convex, non-differentiable losses
405 due to the bi-level nature of decision-focused learning. More recently, [Huang and Wang \(2025\)](#)
406 obtained a $\mathcal{O}((1 + P_T^\infty)^{1/3} T^{-1/3})$ bound, where $P_T^\infty := \sum_{t=1}^T \|f_{t+1} - f_t\|_\infty$. Yet, this rate is
407 achieved under substantially more favorable conditions: a single-level problem with losses that are
408 strongly convex or Lipschitz, an additional assumption of “quasi-stationary” and a more challenging
409 path involving the full landscapes of the f_t s. By comparison, our losses are neither Lipschitz (indeed,
410 they may even be discontinuous due to the linearity of the lower-level problem) nor required to be
411 stationary over time—the only restriction being [H2](#) to hold. We view the ability of our algorithm
412 to achieve efficient convergence despite these demanding conditions as a core contribution of our
413 work.

4 EXPERIMENTS

414 In this section, we compare the performances of our algorithms DF-FTPL and DF-OGD to two important
415 benchmarks, namely prediction-focused learning and SPO ([Elmachtoub and Grigas, 2022](#)).

416 **Setting.** Our experimental setup is inspired by the knapsack example from [Mandi et al. \(2024\)](#).
417 More precisely, we consider a decision maker who must pick at each $t \in [T]$ an object $v_t \in \mathcal{V}$ among
418 K items denoted $\mathcal{V} = \{1, 2, \dots, K\}$ with respective costs $\bar{g}_t(X) = (\bar{g}_{t,1}(X), \dots, \bar{g}_{t,K}(X)) \in$
419 $[0, 1]^K$ depending on some covariates $X \in \mathcal{X}$. At the beginning of each period, the decision-
420 maker only observe covariates $X_t \in \mathcal{X} \subset \mathbb{R}^p$, and have at their disposal a parametric model $g : \Theta \times \mathcal{X} \rightarrow [0, 1]^K$ to predict \bar{g}_t . Given their current parameter $\theta_t \in \Theta$, they predict item costs
421 as $g(\theta_t, X_t) = (g_1(\theta_t, X_t), \dots, g_K(\theta_t, X_t))$ and pick an item $v_t^*(\theta_t) = \operatorname{argmin}_{i \in [K]} g_i(\theta_t, X_t)$.
422 This setting is depicted in Figure 2 for $K = 2$ in Appendix E. After having made their decision, the
423 decision-maker observes the true item costs $\bar{g}_t(X_t)$, and update θ_{t+1} for the next round based on this
424 feedback. Note that this setting can directly be mapped in the simplex example of Remark 1, since

432 $v_t^*(\theta_t) = \operatorname{argmin}_{w \in \mathcal{W}_0} \langle w, g(\theta_t, X_t) \rangle$, \mathcal{W}_0 being the simplex of \mathbb{R}^K . For this reason, we instantiate
 433 our algorithms DF-OGD and DF-FTPL with the negative entropy regularizer, following Remark 1.
 434

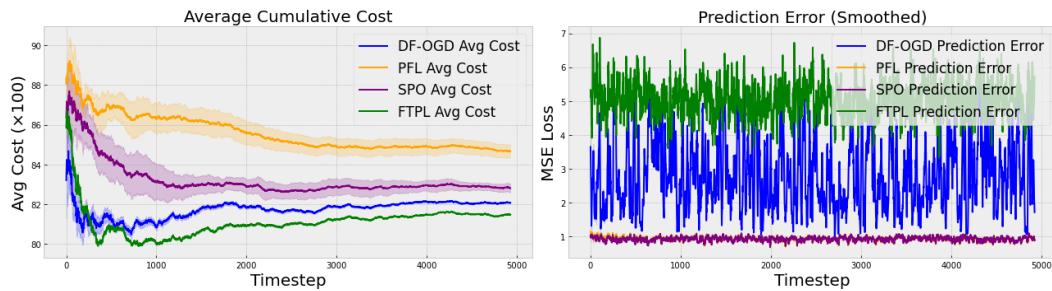
435 **Synthetic data.** We instantiate the previous problem with the following synthetic data. For any
 436 $t \in [T]$, we draw $X_t \in \mathbb{R}^{K \times p}$ with correlated rows, and generate a cost vector $c_t(X_t) \in \mathbb{R}^k$ as :

$$438 \quad c_t(X_t) = A \sin^4((2X_t \theta_t^*)^{-1}) + \varepsilon_t, \quad (9)$$

439 where $A > 0$, $\varepsilon_t \sim N(0, I_K)$ is a Gaussian noise and $\theta_t^* \in \mathbb{R}^p$ is a parameter which satisfies $\theta_t^* =$
 440 $1/2 \theta^* + 1/2 \zeta_t$, where $\zeta_t \sim N(0, I_p)$ for some $\theta^* \in \mathbb{R}^p$. This is a challenging data generating
 441 process, since $\bar{g}_t(X_t) = \sin^4((2X_t \theta_t^*)^{-1})$ is non-stationary and highly non-linear, and features are
 442 correlated. Equation (9) is discussed more in detail in Appendix E. To predict c_t from X_t , we assume
 443 that the decision-maker has access to a class of linear predictors of the form $g : (\theta, X) \mapsto X\theta$.
 444

445 **Benchmarks.** In this setting, we compare the performances of DF-FTPL (Algorithm 1) and
 446 DF-OGD (Algorithm 2) to two benchmarks. First, we implement Prediction-Focused Online Gra-
 447 dient Descent (PF-OGD). This strategy consists in training in an online manner the model g at each
 448 timestep so it minimizes the statistical loss $\ell^{\text{mse}} : (v, X\theta) \mapsto \|v - X\theta\|^2$, irrespective of the down-
 449 stream decision problem. Then, decisions are greedily made based on the predictions of the model.
 450 This approach is formally described in Algorithm 3 in Appendix E. Second, we compare our al-
 451 gorithms to an online version of the Smart Predict-then-Optimize (online SPO, see Algorithm 4)
 452 approach from Elmachtoub and Grigas (2022). This algorithm introduces a differentiable and con-
 453 vex loss which upper-bounds the actual decision-focused loss. Given its effectiveness and versatility,
 454 it is considered as a very important benchmark in the literature.

455 **Results.** In Figure 1, we plot on the left-hand panel the average cumulated cost $t \mapsto$
 456 $t^{-1} \sum_{s=1}^t \bar{g}_{s, v_s}(X_s)$ incurred by DF-OGD, DF-FTPL, PF-OGD and online SPO over 10 runs of
 457 $T = 5000$ timesteps, as well as the associated 95% confidence intervals. On the right-hand panel,
 458 we plot the average Mean Square Error (MSE) resulting from the sequence of prediction parame-
 459 ters $(\theta_t)_{t \in [T]}$. It appears that both DF-FTPL and DF-OGD outperform PF-OGD and online SPO from
 460 a decision point of view, while incurring a higher MSE. This is in line with decision-focused ap-
 461 proach, which cares about decision cost rather than statistical loss. This experiment shows that (i)
 462 DFL outperforms PFL when models are clearly misspecified and (ii) our algorithms also outperforms
 463 the celebrated online SPO. We also mention the presence of additional numerical experiments in



474 Figure 1: Average cumulated cost (left) and prediction errors (right) of DF-OGD, PF-OGD, DF-FTPL
 475 and online SPO.
 476

477 Appendix E, to study how model misspecification affects the relative performances of DFL and
 478 PFL.
 479

480 5 CONCLUSION.

483 Decision-focused learning offers a promising way to integrate prediction into decision making. We
 484 extend its analysis from the batch to the online setting, enabling non-stationary data and varying
 485 objectives. Our algorithms DF-FTPL and DF-OGD comes with provable guarantees, and empirically
 486 outperform prediction-focused learning and online SPO in our experiment.

486 We believe that this work can be extended in several ways. **First, it is plausible that faster rates**
 487 **(e.g., $T^{-1/2}$) could be obtained using more traditional online-learning techniques that bypass**
 488 **the bilevel framework, such as discretizing the parameter space and applying expert aggre-**
 489 **gation over the resulting bins. However, such approaches would likely incur a prohibitive**
 490 **dependence on the parameter dimension m (e.g., exponential in m). Exploring this direc-**
 491 **tion remains however an interesting avenue in future work.** Second, we believe that alterna-
 492 **tive smoothing techniques could be successfully employed, such as Moreau-Yosida transform and**
 493 **proximal operators, to address the more challenging of a general convex set \mathcal{W} . Third, it would be**
 494 **valuable to investigate less adversarial environments, such as those involving i.i.d. data, to explore**
 495 **whether stronger theoretical guarantees could be obtained and whether novel algorithmic designs**
 496 **might emerge. Finally, implementing our method in more ambitious experimental settings would be**
 497 **of great interest from a practical perspective.**

498 REFERENCES

500 J. Abernethy, C. Lee, A. Sinha, and A. Tewari. Online linear optimization via smoothing. In
 501 *Conference on learning theory*, pages 807–823. PMLR, 2014.

503 J. Acimovic and S. C. Graves. Making better fulfillment decisions on the fly in an online retail
 504 environment. *Manufacturing & Service Operations Management*, 17(1):34–51, 2015.

505 N. Agarwal, A. Gonen, and E. Hazan. Learning in non-convex games with an optimization oracle.
 506 In *Conference on Learning Theory*, pages 18–29. PMLR, 2019.

508 B. Amos and J. Z. Kolter. Optnet: Differentiable optimization as a layer in neural networks. In
 509 *International conference on machine learning*, pages 136–145. PMLR, 2017.

510 A. Aswani, Z.-J. M. Shen, and A. Siddiq. Data-driven incentive design in the medicare shared
 511 savings program. *Operations Research*, 67(4):1002–1026, 2019.

513 S. Aydore, T. Zhu, and D. P. Foster. Dynamic local regret for non-convex online forecasting. *Ad-*
 514 *vances in neural information processing systems*, 32, 2019.

515 G.-Y. Ban and C. Rudin. The big data newsvendor: Practical insights from machine learning. *Op-*
 516 *erations Research*, 67(1):90–108, 2019.

518 O. Bennouna, J. Zhang, S. Amin, and A. Ozdaglar. Addressing misspecification in contextual opti-
 519 *mization. arXiv preprint arXiv:2409.10479*, 2024.

520 Q. Berthet, M. Blondel, O. Teboul, M. Cuturi, J.-P. Vert, and F. Bach. Learning with differentiable
 521 perturbed optimizers. *Advances in neural information processing systems*, 33:9508–9519, 2020.

523 D. Bertsimas and N. Kallus. From predictive to prescriptive analytics. *Management Science*, 66(3):
 524 1025–1044, 2020.

525 D. Bertsimas, V. F. Farias, and N. Trichakis. Fairness, efficiency, and flexibility in organ allocation
 526 for kidney transplantation. *Operations Research*, 61(1):73–87, 2013.

528 O. Besbes, Y. Gur, and A. Zeevi. Non-stationary stochastic optimization. *Operations research*, 63
 529 (5):1227–1244, 2015.

531 C. M. Bishop and N. M. Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer,
 2006.

533 S. Boyd. Convex optimization. *Cambridge UP*, 2004.

535 N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

536 X. Chen, Z. Owen, C. Pixton, and D. Simchi-Levi. A statistical learning approach to personalization
 537 in revenue management. *Management Science*, 68(3):1923–1937, 2022.

539 W. C. Cheung, D. Simchi-Levi, and R. Zhu. Learning to optimize under non-stationarity. In *The*
 22nd International Conference on Artificial Intelligence and Statistics, pages 1079–1087. PMLR,
 2019.

540 M. C. Cohen, N.-H. Z. Leung, K. Panchamgam, G. Perakis, and A. Smith. The impact of linear
 541 optimization on promotion planning. *Operations Research*, 65(2):446–468, 2017.
 542

543 B. Colson, P. Marcotte, and G. Savard. An overview of bilevel optimization. *Annals of operations
 544 research*, 153:235–256, 2007.

545 A. Cutkosky, H. Mehta, and F. Orabona. Optimal Stochastic Non-smooth Non-convex Optimization
 546 through Online-to-Non-convex Conversion. In *International Conference on Machine Learning,
 547 ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, 2023.

548

549 O. de Oliveira. The implicit and the inverse function theorems: Easy proofs. *Real Analysis
 550 Exchange*, 39(1):207, 2014. ISSN 0147-1937. doi: 10.14321/realanalexch.39.1.0207. URL
 551 <http://dx.doi.org/10.14321/realanalexch.39.1.0207>.

552 R. Dixit, A. S. Bedi, R. Tripathi, and K. Rajawat. Online learning with inexact proximal online
 553 gradient descent algorithms. *IEEE Transactions on Signal Processing*, 67(5):1338–1352, 2019.

554

555 P. Donti, B. Amos, and J. Z. Kolter. Task-based end-to-end model learning in stochastic optimiza-
 556 tion. *Advances in neural information processing systems*, 30, 2017.

557 J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic
 558 optimization. *Journal of machine learning research*, 12(7), 2011.

559

560 A. N. Elmachtoub and P. Grigas. Smart “predict, then optimize”. *Management Science*, 68(1):9–26,
 561 2022.

562 V. F. Farias, S. Jagabathula, and D. Shah. A nonparametric approach to modeling choice with limited
 563 data. *Management science*, 59(2):305–322, 2013.

564

565 K. J. Ferreira, B. H. A. Lee, and D. Simchi-Levi. Analytics for an online retailer: Demand fore-
 566 casting and price optimization. *Manufacturing & service operations management*, 18(1):69–88,
 567 2016.

568 M. Fisher, C. K. Glaeser, and X. Su. Optimal retail location: Empirical methodology and application
 569 to practice. *Available at SSRN*, 2016.

570

571 H. Frezat, R. Fablet, G. Balarac, and J. L. Sommer. Gradient-free online learning of subgrid-scale
 572 dynamics with neural emulators. *arXiv preprint arXiv:2310.19385*, 2023.

573 G. Gallo and S. Pallottino. Shortest path algorithms. *Annals of operations research*, 13(1):1–79,
 574 1988.

575

576 S. Ghadimi and G. Lan. Stochastic First- and Zeroth-Order Methods for Nonconvex Stochastic
 577 Programming. *SIAM J. Optim.*, 2013.

578

579 U. Ghai, Z. Lu, and E. Hazan. Non-convex online learning via algorithmic equivalence. *Advances
 580 in Neural Information Processing Systems*, 35:22161–22172, 2022.

581

582 S. Gould, B. Fernando, A. Cherian, P. Anderson, R. S. Cruz, and E. Guo. On differentiating param-
 583 eterized argmin and argmax problems with application to bi-level optimization. *arXiv preprint
 584 arXiv:1607.05447*, 2016.

585

586 V. Gupta, B. R. Han, S.-H. Kim, and H. Paek. Maximizing intervention effectiveness. *Management
 587 Science*, 66(12):5576–5598, 2020.

588

589 E. Hall and R. Willett. Dynamical models and tracking regret in online convex programming. In
 590 *International Conference on Machine Learning*. PMLR, 2013.

591

592 N. Hallak, P. Mertikopoulos, and V. Cevher. Regret minimization in stochastic non-convex learn-
 593 ing via a proximal-gradient approach. In *International Conference on Machine Learning*, pages
 4008–4017. PMLR, 2021.

594

595 E. Hazan. Introduction to online convex optimization, 2023. URL <https://arxiv.org/abs/1909.05207>.

594 E. Hazan, K. Singh, and C. Zhang. Efficient regret minimization in non-convex games. In *International Conference on Machine Learning*, pages 1433–1441. PMLR, 2017.

595

596

597 A. Héliou, P. Mertikopoulos, and Z. Zhou. Gradient-free online learning in continuous games with

598 delayed rewards. In *International conference on machine learning*, pages 4172–4181. PMLR,

599 2020.

600 Y. Hsieh, M. R. K. Jaghargh, A. Krause, and P. Mertikopoulos. Riemannian stochastic optimiza-

601 tion methods avoid strict saddle points. In *Advances in Neural Information Processing Systems*

602 *36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New*

603 *Orleans, LA, USA, December 10 - 16, 2023*, 2023.

604 C. Huang and K. Wang. A stability principle for learning under nonstationarity. *Operations Re-*

605 *search*, 2025.

606

607 M. Huang and V. Gupta. Decision-focused learning with directional gradients. *Advances in Neural*

608 *Information Processing Systems*, 37:79194–79220, 2024.

609 M. Hutter and J. Poland. Adaptive online prediction by following the perturbed leader. 2005.

610

611 K. Ji, J. Yang, and Y. Liang. Bilevel optimization: Convergence analysis and enhanced design. In

612 *International conference on machine learning*, pages 4882–4892. PMLR, 2021.

613

614 A. Kalai and S. Vempala. Efficient algorithms for online decision problems. *Journal of Computer*

615 *and System Sciences*, 71(3):291–307, 2005.

616 A. Lesage-Landry, J. A. Taylor, and I. Shames. Second-order online nonconvex optimization. *IEEE*

617 *Transactions on Automatic Control*, 66(10):4866–4872, 2020.

618 B. Li and S. C. Hoi. Online portfolio selection: A survey. *ACM Computing Surveys (CSUR)*, 46(3):

619 1–36, 2014.

620

621 H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein. Visualizing the Loss Landscape of Neural

622 Nets. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural*

623 *Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*,

624 2018.

625 S. Lin, D. Sow, K. Ji, Y. Liang, and N. Shroff. Non-convex bilevel optimization with time-varying

626 objective functions. *Advances in Neural Information Processing Systems*, 36:29692–29717, 2023.

627

628 J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang. Learning under concept drift: A review. *IEEE*

629 *transactions on knowledge and data engineering*, 31(12):2346–2363, 2018.

630

631 E. Mammen and A. B. Tsybakov. Smooth discrimination analysis. *The Annals of Statistics*, 27(6):

632 1808–1829, 1999.

633 J. Mandi and T. Guns. Interior point solving for lp-based prediction+ optimisation. *Advances in*

634 *Neural Information Processing Systems*, 33:7272–7282, 2020.

635

636 J. Mandi, V. Bucarey, M. M. K. Tchomba, and T. Guns. Decision-focused learning: Through the

637 lens of learning to rank. In *International conference on machine learning*, pages 14935–14947.

638 PMLR, 2022.

639 J. Mandi, J. Kotary, S. Berden, M. Mulamba, V. Bucarey, T. Guns, and F. Fioretto. Decision-focused

640 learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial*

641 *Intelligence Research*, 80:1623–1701, 2024.

642

643 P. Mertikopoulos, N. Hallak, A. Kavis, and V. Cevher. On the Almost Sure Convergence of Stochas-

644 tic Gradient Descent in Non-Convex Problems. In *Advances in Neural Information Processing*

645 *Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020,*

646 *December 6-12, 2020, virtual*, 2020.

647 V. V. Mivsić and G. Perakis. Data analytics in operations management: A review. *Manufacturing &*

648 *Service Operations Management*, 22(1):158–169, 2020.

648 S. Padakandla, P. KJ, and S. Bhatnagar. Reinforcement learning algorithm for non-stationary envi-
 649 ronments. *Applied Intelligence*, 50(11):3590–3606, 2020.

650

651 V. Patel and S. Zhang. Stochastic Gradient Descent on Nonconvex Functions with General Noise
 652 Models, 2021.

653 V. Perchet and P. Rigollet. The multi-armed bandit problem with covariates. 2013.

654

655 J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence. *Dataset shift in ma-
 656 chine learning*. Mit Press, 2022.

657 A. Rakhlin, K. Sridharan, and A. Tewari. Online learning: Stochastic, constrained, and smoothed
 658 adversaries. *Advances in neural information processing systems*, 24, 2011.

659

660 S. Rath, K. Rajaram, and A. Mahajan. Integrated anesthesiologist and room scheduling for surgeries:
 661 Methodology and application. *Operations Research*, 65(6):1460–1478, 2017.

662 U. Sadana, A. Chenreddy, E. Delage, A. Forel, E. Frejinger, and T. Vidal. A survey of contextual
 663 optimization methods for decision-making under uncertainty. *European Journal of Operational
 664 Research*, 320(2):271–289, 2025.

665 N. Schutte, K. Postek, and N. Yorke-Smith. Robust losses for decision-focused learning. In *Pro-
 666 ceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 4868–
 667 4875, 2024.

668

669 S. Shah, K. Wang, B. Wilder, A. Perrault, and M. Tambe. Decision-focused learning without
 670 decision-making: Learning locally optimized decision losses. *Advances in Neural Information
 671 Processing Systems*, 35:1320–1332, 2022.

672 L. Shen, N. Ho-Nguyen, and F. Kılınç-Karzan. An online convex optimization-based framework
 673 for convex bilevel optimization. *Mathematical Programming*, 198(2):1519–1582, 2023.

674

675 A. Sinha, P. Malo, and K. Deb. A review on bilevel optimization: From classical to evolutionary
 676 approaches and applications. *IEEE transactions on evolutionary computation*, 22(2):276–295,
 677 2017.

678 A. S. Suggala and P. Netrapalli. Online non-convex learning: Following the perturbed leader is
 679 optimal. In *Algorithmic Learning Theory*, pages 845–861. PMLR, 2020.

680 V. Syrgkanis, A. Agarwal, H. Luo, and R. E. Schapire. Fast convergence of regularized learning in
 681 games. *Advances in Neural Information Processing Systems*, 28, 2015.

682

683 D. A. Tarzanagh, P. Nazari, B. Hou, L. Shen, and L. Balzano. Online bilevel optimization: Re-
 684 gret analysis of online alternating gradient methods. In *International Conference on Artificial
 685 Intelligence and Statistics*, pages 2854–2862. PMLR, 2024.

686

687 A. B. Tsybakov. Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*,
 32(1):135–166, 2004.

688

689 S. Verma, A. Mate, K. Wang, A. Taneja, and M. Tambe. Case study: Applying decision focused
 690 learning in the real world. In *Workshop on Trustworthy and Socially Responsible Machine Learn-
 691 ing, NeurIPS 2022*, 2022.

692 S. Verma, A. Mate, K. Wang, N. Madhiwalla, A. Hegde, A. Taneja, and M. Tambe. Restless multi-
 693 armed bandits for maternal and child health: Results from decision-focused learning. In *AAMAS*,
 694 pages 1312–1320, 2023.

695

696 K. Wang, S. Verma, A. Mate, S. Shah, A. Taneja, N. Madhiwalla, A. Hegde, and M. Tambe. Scal-
 697 able decision-focused learning in restless multi-armed bandits with application to maternal and
 698 child health. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages
 699 12138–12146, 2023.

700 B. Wilder, B. Dilkina, and M. Tambe. Melding the data-decisions pipeline: Decision-focused learn-
 701 ing for combinatorial optimization. In *Proceedings of the AAAI Conference on Artificial Intelli-
 702 gence*, volume 33, pages 1658–1665, 2019.

702 Z. Xu and L. Zhang. Online non-convex learning in dynamic environments. *Advances in Neural*
703 *Information Processing Systems*, 37:51930–51962, 2024.

704

705 A. Zeevi and A. Goldenshluger. Woodrooffe’s one-armed bandit problem revisited. *Annals of Applied*
706 *Probability*, 19(4):1603–1633, 2009.

707

708 L. Zhang, S. Lu, and Z.-H. Zhou. Adaptive online learning in dynamic environments. *Advances in*
709 *neural information processing systems*, 31, 2018.

710

711 P. Zhao and L. Zhang. Improved analysis for dynamic regret of strongly convex and smooth func-
712 tions. In *Learning for Dynamics and Control*, pages 48–59. PMLR, 2021.

713

714 P. Zhao, Y.-J. Zhang, L. Zhang, and Z.-H. Zhou. Dynamic Regret of Convex and Smooth Functions.
715 In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

716

717 Z. Zhuang, Y. Wang, K. Yu, and S. Lu. No-regret non-convex online meta-learning. In *ICASSP 2020-
718 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*,
719 pages 3942–3946. IEEE, 2020.

720

721 M. Zinkevich. Online Convex Programming and Generalized Infinitesimal Gradient Ascent. In *Machine
722 Learning, Proceedings of the Twentieth International Conference (ICML 2003), August
723 21-24, 2003, Washington, DC, USA*, 2003.

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756 A **H2** IS SATISFIED IN THE GAUSSIAN LINEAR CASE.
757758 In this section, we prove that **H2** holds true when \mathcal{W} is the simplex of \mathbb{R}^d , g is a linear model and
759 X has columns distributed according to a Gaussian distribution. In what follows, we denote by
760 $X_j \in \mathbb{R}^d$ the j -th column of X .
761762 **Proposition 1.** *Assume that $g : (\theta, X) \in \Theta \times \mathcal{X} \mapsto X\theta$ where $\Theta = \{\theta \in \mathbb{R}^m : \|\theta\|_2 = 1\}$ and*
763 *$X = (X_1 | \dots | X_m) \in \mathbb{R}^{d \times m}$ where $X_j \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then, **H2** is true for any $\theta \in \Theta$ with*

764
$$765 C_0 = \frac{d(d-1)}{2\sqrt{\pi}} \quad \text{and} \quad \beta = 1. \\ 766$$

767 *Proof.* Let $\theta \in \Theta$. In what follows, we denote $i = I(\theta)$ where $I(\theta)$ is defined as in **H2**. Since \mathcal{W} is
768 the simplex of \mathbb{R}^d , v_j is the j -th vector of the canonical basis of \mathbb{R}^d . For any $\varepsilon > 0$ we have:
769

770
$$\begin{aligned} 771 \mathbb{P}\left(\inf_{j \neq i} |\langle X\theta, v_j \rangle - \langle X\theta, v_i \rangle| \leq \varepsilon\right) &\leq \mathbb{P}\left(\bigcup_{j \neq i} \{|\langle X\theta, v_j \rangle - \langle X\theta, v_i \rangle| \leq \varepsilon\}\right) \\ 772 &\leq \sum_{j \neq i} \mathbb{P}(|\langle \theta, X^\top (v_j - v_i) \rangle| \leq \varepsilon) \\ 773 &= \sum_{j \neq i} \mathbb{P}(|\langle \theta, X_j - X_i \rangle| \leq \varepsilon). \end{aligned} \tag{10}$$

774 Since for any $j \neq i$, $\langle \theta, X_j - X_i \rangle \sim \mathcal{N}(0, 2\|\theta\|_2^2)$ with density f_j , we have:
775

776
$$\mathbb{P}(|\langle \theta, X_j - X_i \rangle| \leq \varepsilon) = \int_{-\varepsilon}^{\varepsilon} f_j(x) dx = (2\sqrt{\pi}\|\theta\|_2)^{-1} \int_{-\varepsilon}^{\varepsilon} \exp\left(-\frac{x^2}{4\|\theta\|_2^2}\right) dx$$

777 and since the integrand reaches its maximum in $x = 0$,
778

779
$$\leq \varepsilon (\sqrt{\pi}\|\theta\|_2)^{-1}. \tag{11}$$

780 Then, plugging (11) in (10) and summing over all pairs (i, j) such that $j \neq i$ yields the desired
781 result. \square 782 Note that the d^2 factor in C_0 comes from the use of an union bound, and could be reduced with a
783 refined analysis.
784785 B EXTENDED DISCUSSION ON THE RELEVANCE OF APPROXIMATE ORACLES.
786787 The notion of an approximate oracle reflects the fact that, in non-convex settings, we cannot rely on
788 subroutines that provably reach a global minimizer of \tilde{f}_t at time t (as offline gradient descent would
789 in convex problems). Instead, we must settle for local minimizers, whose quality is governed by a
790 parameter ξ —which vanishes in favorable loss landscapes.791 As a concrete example, consider **O** as the stochastic gradient descent (SGD) algorithm. Even in the
792 context of deep networks, it is plausible that **O** converges to a local minimizer. Indeed, a large body
793 of work has analyzed SGD in non-convex settings, establishing convergence to stationary points
794 either in expectation (Ghadimi and Lan, 2013) or almost surely (Mertikopoulos et al., 2020; Patel
795 and Zhang, 2021; Cutkosky et al., 2023). Such stationary points may correspond to local/global
796 minima or saddle points.797 Recent studies further show that SGD avoids saddle points. For instance, Mertikopoulos et al.
798 (2020) proved that the trajectories of SGD almost surely avoid all strict saddle manifolds—i.e., sets
799 of critical points x where the Hessian has at least one negative eigenvalue. These manifolds include
800 connected families of non-isolated saddle points, a phenomenon common in the loss landscapes of
801 overparametrized neural networks (Li et al., 2018).

Beyond SGD, similar guarantees extend to more general methods. In particular, the stochastic Riemannian Robbins–Monro method (a broad template encompassing various algorithms) has been shown to converge almost surely to local or global minima (Hsieh et al., 2023).

Finally, we highlight that, in most practical industrial settings, solving the downstream optimization problem is typically far more computationally demanding than updating the prediction model. It is therefore reasonable to assume that the oracle call represents only a small fraction of the overall computational cost.

C SUPPLEMENTARY RESULTS ON THE SIMPLEX.

We now provide convergence guarantees for Algorithms 1 and 2 when \mathcal{W} is the simplex. To do so we choose another regularizer than done in the main document.

C.1 ADDITIONAL FRAMEWORK

We focus on the case where \mathcal{W} is the simplex of \mathbb{R}^d , that is

$$\mathcal{W}_0 = \{w \in \mathbb{R}^d : w \succcurlyeq 0, \mathbf{1}^\top w = 1\}.$$

This setting encompasses various decision-making scenarios such as portfolio selection Li and Hoi (2014) or mixed strategy design in multiplayer games Syrgkanis et al. (2015). In this case, we choose the negative entropy

$$\mathcal{R}_0 : w \mapsto \sum_{i \in [d]} w_i \ln(w_i),$$

as the regularizer in (5). The minimizer $\tilde{w}_t(\theta) \in \mathcal{W}_0$ in (5) with $\mathcal{R} = \mathcal{R}_0$ can easily be shown to be the softmax mapping, that is it satisfies for any $i \in [d]$:

$$\tilde{w}_{t,i}(\theta) = \frac{\exp(-\alpha^{-1}g_i(\theta, X_t))}{\sum_{k \in [d]} \exp(-\alpha^{-1}g_k(\theta, X_t))}.$$

It is clear from this expression that \tilde{w}_t is differentiable, and that for any $\theta \in \Theta$:

$$\nabla \tilde{w}_t(\theta) = -\frac{1}{\alpha} [\text{diag}[\tilde{w}_t(\theta)] - \tilde{w}_t(\theta) \tilde{w}_t(\theta)^\top] \nabla_\theta g(\theta, X_t). \quad (12)$$

C.2 RESULTS

DF-FTPL

Proposition 2. *Assume **H1**, **H2** and having access to an ξ -approximate optimization oracle \mathbf{O}_ξ adapted to $\left\{ \sum_{i=1}^t \tilde{f}_i - \langle \sigma_t, \cdot \rangle \right\}_{t \in [T]}$. Fix $\mathcal{W} = \mathcal{W}_0$, $\mathcal{R} = \mathcal{R}_0$. Let $\{\theta_t\}_{t \in [T]}$ be the output of DF-FTPL (Algorithm 1) instantiated with learning step $\eta > 0$ and regularization coefficients $\alpha_t = \alpha > 0$ for any t . Then:*

$$T^{-1} \mathbb{E}[\mathfrak{R}_T^s] \leq \tilde{\mathcal{O}} \left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi + \alpha \ln(d) \right),$$

where \mathbb{E} denotes the expectation on both data and the intrinsic randomness of DF-FTPL and $\tilde{\mathcal{O}}$ contains polynomial dependency in $\ln(1/\alpha), \ln(\ln(d))$.

The proof of this result is postponed to Appendix G.4.

DF-OGD

Proposition 3. *Assume **H1**, **H2**, and having access to an ξ -approximate optimization oracle \mathbf{O}_ξ adapted to $\left\{ \tilde{f}_i \right\}_{t \in [T]}$. Fix $\mathcal{W} = \mathcal{W}_0$, $\mathcal{R} = \mathcal{R}_0$. Let $\{\theta_t\}_{t \in [T]}$ be the output of DF-OGD (Algorithm 2)*

864 instantiated with the non-increasing sequence $(\eta_t)_{t \in [T]}$ and regularization coefficients $(\alpha_t)_{t \in [T]}$.
 865 Then:

$$867 T^{-1} \mathbb{E}[\mathfrak{R}_T^d] \leq \frac{D_\Theta^2 + D_\Theta \mathbb{E}[P_T]}{2T\eta_T} + \sum_{t \in [T]} \frac{25D_Z G\eta_t}{32T\alpha_t^2} \\ 868 + 2D_Z \sum_{t \in [T]} \alpha_t [1 + 2\ln(d)C_0 \{\ln(2\alpha_t^{-1}) + (1 - \beta)\ln^2(\alpha_t \ln d)\}] + \xi,$$

872 where $P_T = \sum_{t=1}^{T-1} \|\vartheta_{t+1} - \vartheta_t\|$.

874 The proof of this result is postponed to Appendix G.5.

876 Note that the principal gain, compared to the general polytope case, is to avoid the dependency in n
 877 the number of faces and enjoys a nice $\ln(d)$ dependency in the dimension of the decision space.

879 D BACKGROUND ON THE FTPL ALGORITHM

881 **Approximate Optimization Oracle.** The recent work of [Suggala and Netrapalli \(2020\)](#) proposed
 882 an online algorithm for nonconvex losses $(\ell_t)_{t \in [T]}$ with static regret guarantees. They rely on an
 883 *approximate optimization oracle* \mathcal{O} which takes as input any function ℓ , a d -dimensional vector
 884 σ and returns an approximate minimizer of $\ell - \langle \sigma, \cdot \rangle$. An optimization oracle is called " (ξ, χ) -
 885 approximate optimization oracle" if it returns $\mu^* \in \mathcal{K}$ such that

$$886 \ell(\mu^*) - \langle \sigma, \mu^* \rangle \leq \inf_{\mu \in \mathcal{K}} [\ell(\mu) - \langle \sigma, \mu \rangle] + (\xi + \chi \|\sigma\|_1). \quad (13)$$

888 We denote the output of such an optimization oracle by $\mu^* = \mathcal{O}_{\xi, \chi}(\ell - \langle \sigma, \cdot \rangle)$. Note the notion
 889 of oracle described in (13) is very close from the Definition 1 we made in Section 3 on the expert
 890 sequence $(\vartheta_t)_{t \in [T]}$ (in this case $\chi = 0$ as no σ is involved).

891 **Remark 2.** Note that, in our work, we made the choice to fix $\chi = 0$. This stronger assumption is due
 892 to the will of having a unifying framework encompassing both the setup for DF-FTPL and DF-OGD.

894 **Follow The Perturbed Leader.** Given access to an (ξ, χ) -approximate optimization oracle, [Sug-
 895 gala and Netrapalli \(2020\)](#) study the FTPL algorithm which is described by the following recursion.
 896 Starting from $\hat{\mu}_1$, at each time steps t , $\hat{\mu}_t \in \mathcal{K}$ is updated as follows:

$$898 \hat{\mu}_{t+1} = \mathcal{O}_{\xi, \chi} \left(\sum_{i=1}^{t-1} \ell_i - \langle \sigma_t, \cdot \rangle \right) \quad (14)$$

901 where $\sigma_t \in \mathbb{R}^d$ is a random perturbation such that $\sigma_{t,j}$, the j -th coordinate of σ_t , is sampled from
 902 $\text{Exp}(\eta)$, the exponential distribution with parameter η .

904 Following this update route, the addition of an exponential noise allowing to handle the non-
 905 convexity of the losses, they reach the following static regret bound.

906 **Proposition 4** (Theorem 1 of [Suggala and Netrapalli \(2020\)](#)). *Let D be the diameter of \mathcal{K} . Suppose
 907 that ℓ_t is L -Lipschitz w.r.t L1 norm, for all $t \in [T]$. For any fixed η , FTPL (Equation (14)) with
 908 access to a (ξ, χ) -approximate" optimization oracle satisfies the following static regret bound:*

$$909 \frac{1}{T} \sum_{t=1}^T \ell_t(\hat{\mu}_t) - \inf_{\mu} \sum_{t=1}^T \ell_t(\mu) \leq \mathcal{O} \left(\eta m^2 D L^2 + \frac{m(\chi T + D)}{\eta T} + \xi + \chi m L \right),$$

912 where \mathbb{E} denotes the expectation over $\sigma_1 \cdots \sigma_T$.

914 E ADDITIONAL EXPERIMENTAL MATERIAL

916 **The motivating example from [Mandi et al. \(2024\)](#).** In this paragraph, we recall the example that
 917 motivates our experiment. [Mandi et al. \(2024\)](#) illustrates the interest of decision-focused learning

918 with a simple problem where a decision maker seeks to pick, between two objects, the one with
 919 the lowest cost. Before picking an object, they do not know costs but have at their disposal a
 920 model to predict it. This situation is depicted in Figure 2. For instance, if the true cost $\bar{g}_t =$
 921 $(\bar{g}_{t,1}, \bar{g}_{t,K})$ is $(*)$, any prediction $\hat{g}_t = (\hat{g}_{t,1}, \hat{g}_{t,K})$ lying in the blue-shaded area, such as $(+)$, induces
 922 the optimal decision $v_t = 1$. As observed in Mandi et al. (2024), a prediction-focused approach
 923 may underperform compared to a decision-focused one in this setting. This is because generating
 924 a prediction \hat{v}_t that closely approximates the true value v_t —in the sense of minimizing statistical
 925 loss—does not necessarily ensure that \hat{v}_t falls on the same side of the 45° line as v_t . For example,
 926 the prediction (\times) is satisfactory from a predictive point of view, but induces a sub-optimal action.
 927 On the contrary, one would expect the decision-focused approach to produce predictions that lie
 928 further away from $(*)$ since it does not minimize prediction error, but on the right side of the 45°
 929 line; see for instance (\blacktriangle) or (\blacktriangledown) on Figure 2.
 930

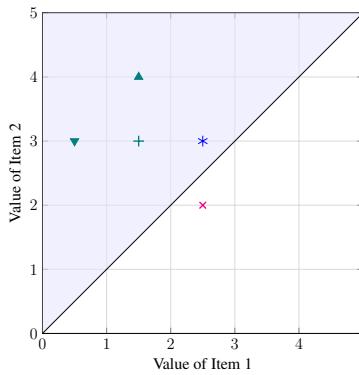


Figure 2: Figure 2 from Mandi et al. (2024).

946 **Details on the experimental setup.** In this paragraph, we provide more detail about the experimental
 947 setup in Section 4 and perform additional numerical simulations. We start by explaining
 948 more precisely how the data used in our experiment are drawn. First, for any $t \in [T]$, $X_t \in \mathbb{R}^{K \times p}$
 949 is constructed as follows: (i) we generate a Toeplitz covariance matrix $\Sigma = (\rho^{|i-j|})_{(i,j) \in [K]^2}$ for
 950 some $\rho \in (0, 1)$, (ii) apply a Cholesky decomposition to it: $\Sigma = LL^\top$, and (iii) generate a matrix
 951 \bar{X}_t with standard Gaussian entries. Then, X_t is defined as $X_t = L\bar{X}_t$. This introduces correlation
 952 between features, which makes convergence harder for an online linear model. Second, we generate
 953 $v_t \in \mathbb{R}^K$ as follows:

$$954 \quad c_t = \min(\max(\tilde{c}_t, 0), 1) \quad \text{with} \quad \tilde{c}_t = A \sin^4((2X_t \theta_t^*)^{-1}) + \varepsilon. \\ 955$$

956 In the above equation, $A > 0$ is a constant, θ_t^* satisfies:

$$957 \quad \theta_t^* = \frac{1}{2}\theta^* + \frac{1}{2}\zeta_t, \quad \text{where} \quad \zeta_t \sim N(0, I_K) \quad \text{and} \quad \theta^* \in \mathbb{R}^K, \\ 958$$

959 and $\varepsilon_t \sim N(0, I_K)$ is Gaussian noise. The fact that θ_t^* varies throughout time and that the relationship
 960 between v_t and X_t is highly linear makes learning hard for a linear model.

961 We now present the values used for the different parameters in our experiment. We consider $K = 5$
 962 items and $p = 10$ features. The horizon is set to $T = 5000$. For each plot, we run $N = 10$ times
 963 DF-OGD and PF-OGD. The reported error bars are 95% Gaussian confidence intervals (from the CLT).
 964 In (9), $A = 45$ and the correlation coefficient for X_t is $\rho = 0.8$. Our algorithms are instantiated
 965 with the following parameters. First, DF-OGD runs with (α_t) and (η_t) as suggested by our theoretical
 966 analysis. The oracle from H2 is obtained through a SGD algorithm performing $t_\vartheta = 10$ steps at
 967 each iteration with a learning rate set to $\eta_\vartheta = 0.01$, while being initialized at ϑ_{t-1} . On the other
 968 hand, PF-OGD runs with a learning rate $\eta = 10$ determined by grid search. Our code is implemented
 969 with pytorch.
 970

Benchmark algorithms.

Algorithm 3 Prediction-Focused Online Gradient Descent (PF-OGD).

```

972
973 1: Input: horizon  $T > 0$ , initialization  $\theta_1 \in \Theta$ .
974 2: for each  $t \in \{1, \dots, T\}$  do
975 3:   Observe  $X_t$ , predict  $c_t = g(\theta_t, X_t)$ 
976 4:   Play  $v_t^*(\theta_t) = \operatorname{argmin}_{i \in [K]} c_{t,i}$ , observe  $\bar{g}_t(X_t)$ .
977 5:   Update  $\theta_{t+1} = \Pi_\Theta(\theta_t - \eta \nabla \ell^{\text{mse}}(\bar{g}_t(X_t), X_t \theta_t))$ .
978 6: end for
979
980
981

```

Deviation from realizability. As argued in Section 1, prediction-focused learning would be optimal if the prediction model made no mistake, since in this case problem (2) would perfectly align with (1). As a consequence, one would expect prediction-focused learning to perform very well in the realizable setting, that is when for any $t \in [T]$, there exists $\bar{\theta}_t \in \Theta$ such that $\bar{g}_t(X_t) = g(\bar{\theta}_t, X_t)$. On the flip side, it is likely to struggle in a misspecified setting. This is in contrast with DF-OGD, whose theoretical performance analysis does not highlight any peculiar dependence on realizability. We therefore hypothesize that PFL should outperform DFL in the realizable case, and DFL should take the upper hand when the prediction model is highly misspecified. To test this hypothesis, we run an alternative experiment where we interpolate between the well-specified, linear case, and ill-specified, non-linear one. More precisely, for $\gamma \in [0, 1]$, we generate $(X_t, v_t^{(\gamma)})_{t \in [T]}$ as follows:

$$v_t^{(\gamma)} = (1 - \gamma)X_t \theta_t^* + \gamma \sin^4((2X_t \theta_t^*)^{-1}) + \varepsilon_t.$$

In Figure 3, we make γ vary from 0 to 1. At each value, we plot the average cumulated cost gap between DF-OGD and PF-OGD, averaged over 50 runs with horizon $T = 1000$. This experiment provides empirical support to the previous conjecture, namely, DFL becomes a competitive option when the prediction model is ill-specified.

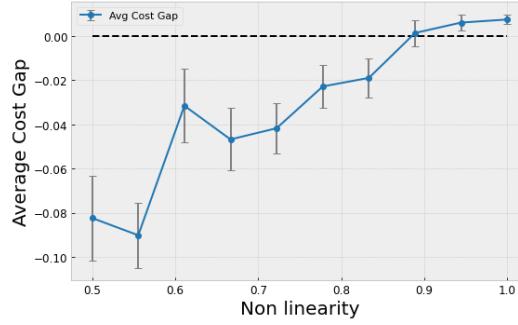


Figure 3: Average cumulated cost gap PFL - DFL as a function of γ .

Algorithm 4 Online Smart Predict-then-Optimize Online Gradient Descent (Online SPO).

```

1009 1: Input: horizon  $T > 0$ , initialization  $\theta_1 \in \Theta$ .
1010 2: for each  $t \in \{1, \dots, T\}$  do
1011 3:   Observe  $X_t$ , predict  $\hat{c}_t = g(\theta_t, X_t)$ .
1012 4:   Play  $v_t^*(\theta_t) = \operatorname{argmin}_{v \in \mathcal{V}} \hat{c}_t^\top v$ , observe  $c_t$ .
1013 5:   Compute  $v^*(\hat{c}_t) = \operatorname{argmin}_{v \in \mathcal{V}} \hat{c}_t^\top v$  and  $v^*(c_t) = \operatorname{argmin}_{v \in \mathcal{V}} c_t^\top v$ .
1014 6:   Define SPO+ surrogate:
1015   
$$\ell^{\text{SPO}+}(\hat{c}_t, c_t) = 2c_t^\top v^*(\hat{c}_t) - c_t^\top v^*(c_t) - \hat{c}_t^\top v^*(\hat{c}_t).$$

1016 7:   Update  $\theta_{t+1} = \Pi_\Theta(\theta_t - \eta \nabla_\theta \ell^{\text{SPO}+}(\hat{c}_t, c_t))$ , where the gradient is computed via chain rule
1017   through  $\hat{c}_t = g(\theta_t, X_t)$ .
1018 8: end for
1019
1020
1021
1022
1023
1024
1025

```

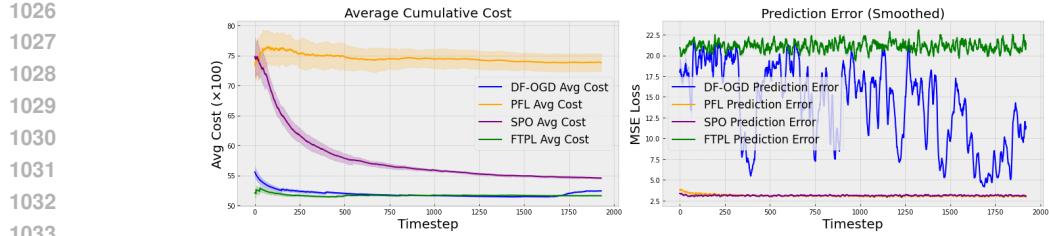


Figure 4: Experiment results with 80 items.

Higher dimension problem. Finally, the following figure shows the performance of our algorithms in a higher dimension problem, with $K = 80$ items. We see that our algorithms still significantly beat the baselines, suggesting that they remain effective in moderately high dimension problems.

F REGRET BOUND WITH AN APPROXIMATE OPTIMAL SOLUTION.

In this section, we show that when we cannot access $w_t^*(\theta_t) \in \mathcal{W}$ but only $\underline{w}_t(\theta_t) \in \mathcal{W}$ such that $\langle g(\theta, X_t), \underline{w}_t(\theta_t) - w_t^*(\theta_t) \rangle \leq \kappa$ for $\kappa > 0$, our regret bounds are only shifted by κ . The following proof is for the general convex polytope case. The simplex case is established in the exact same way.

Proposition 5. *Assume the assumptions from Theorem 2 and that at each step, $w_t^*(\theta_t) \in \mathcal{W}$ is replaced by $\underline{w}_t(\theta_t) \in \mathcal{W}$ which satisfies*

$$\langle g(\theta, X_t), \underline{w}_t(\theta_t) - w_t^*(\theta_t) \rangle \leq \kappa ,$$

where $\kappa > 0$. Then,

$$\begin{aligned} \mathbb{E}[\mathfrak{R}_T^d] &\leq \frac{D_\Theta^2 - D_\Theta \mathbb{E}[P_T]}{2\eta_T} + \sum_{t \in [T]} \frac{\eta_t}{2} \left(\frac{GD_Z \sup_{w \in \mathcal{W}_1} \|b - Aw\|_\infty}{\alpha_t \lambda_{\min}(AA^\top)} \right)^2 \\ &\quad + 2D_Z \sum_{t \in [T]} \alpha_t [1 + 2n \max_{k \in [K]} \|v_k\|_1 C_0 (\ln(1\alpha_t^{-1}) + (1 - \beta)^2 \ln^2(\alpha_t \ln d))] + \xi + \kappa . \end{aligned}$$

The proof Proposition 5 is deferred to Appendix G.6.

G PROOFS

G.1 PROOFS OF PRELIMINARY LEMMAS.

G.1.1 SIMPLEX

Lemma 1. *Assume H1, $\mathcal{W} = \mathcal{W}_0$ and $\mathcal{R} = \mathcal{R}_0$. Then for any $t \in [T]$, \tilde{f}_t is K_t -Lipchitz almost surely, with $K_t = 5D_Z G(4\alpha_t)^{-1}$.*

Proof. Let $(\theta, \theta') \in \Theta^2$. We have:

$$\begin{aligned} |\tilde{f}_t(\theta) - \tilde{f}_t(\theta')| &= \langle \bar{g}_t(X_t), \tilde{w}_t(\theta) - \tilde{w}_t(\theta') \rangle \\ &\leq \|\bar{g}_t(X_t)\| \|\tilde{w}_t(\theta) - \tilde{w}_t(\theta')\| \\ &= \|\bar{g}_t(X_t)\| \int_{t=0}^1 \|\nabla \tilde{w}_t(\theta + t(\theta' - \theta))(\theta - \theta')\| dt \\ &\leq D_Z \int_{t=0}^1 \|\nabla \tilde{w}_t(\theta + t(\theta' - \theta))\|_{\text{op}} \|\theta - \theta'\| dt \quad (\text{by H1-(ii)}) \end{aligned} \tag{15}$$

1080 Now, denoting $\zeta = \theta + t(\theta' - \theta)$, we have:

$$\begin{aligned}
 \|\nabla \tilde{w}_t(\zeta)\|_{\text{op}} &= \alpha_t^{-1} \|(\text{diag}[\tilde{w}_t(\zeta)] - \tilde{w}_t(\zeta)\tilde{w}_t(\zeta)^\top) \nabla_\theta g(\zeta, X_t)\|_{\text{op}} \\
 &\leq \alpha_t^{-1} \|\text{diag}[\tilde{w}_t(\zeta)] - \tilde{w}_t(\zeta)\tilde{w}_t(\zeta)^\top\|_{\text{op}} \|\nabla_\theta g(\zeta, X_t)\|_{\text{op}} \\
 &\leq G\alpha_t^{-1} \|\text{diag}[\tilde{w}_t(\zeta)] - \tilde{w}_t(\zeta)\tilde{w}_t(\zeta)^\top\|_{\text{op}} \quad (\text{by H1-(iii)}) \tag{16}
 \end{aligned}$$

1086 Since $\text{diag}[\tilde{w}_t(\zeta)] - \tilde{w}_t(\zeta)\tilde{w}_t(\zeta)^\top$ is symmetric, its operator norm equals its largest eigenvalue,
1087 denoted λ_{\max} . It then follows that:

$$\begin{aligned}
 \|\text{diag}[\tilde{w}_t(\zeta)] - \tilde{w}_t(\zeta)\tilde{w}_t(\zeta)^\top\|_{\text{op}} &= \lambda_{\max}(\text{diag}[\tilde{w}_t(\zeta)] - \tilde{w}_t(\zeta)\tilde{w}_t(\zeta)^\top) \\
 &\leq \max_{i \in [d]} \sum_{j \in [d]} |\text{diag}[\tilde{w}_t(\zeta)]_{i,j} - \tilde{w}_t(\zeta)_{i,j}\tilde{w}_t(\zeta)_{i,j}^\top| \\
 &\leq \max_{i \in [d]} \{ \tilde{w}_t(\zeta)_{i,i}(1 - \tilde{w}_t(\zeta)_{i,i}) + \tilde{w}_t(\theta)_{i,i} \sum_{j \neq i} \tilde{w}_t(\zeta)_{i,j} \} \\
 &\leq \frac{1}{4} + 1 = \frac{5}{4},
 \end{aligned}$$

1088 where the two last inequalities hold because $\tilde{w}_t \in [0, 1]^d$. Therefore, we obtain from (16) that
1089 $\|\nabla \tilde{w}_t(\zeta)\|_{\text{op}} \leq 5G(4\alpha_t)^{-1}$, and plugging this in (15) yields:

$$\left| \tilde{f}_t(\theta) - \tilde{f}_t(\theta') \right| \leq \frac{5GD_{\mathcal{Z}}}{4\alpha_t} \int_0^1 \|\theta - \theta'\| \, dt = \frac{5GD_{\mathcal{Z}}}{4\alpha_t} \|\theta - \theta'\|. \tag*{\square}$$

1103 **Lemma 2.** Assume **H2**, $\mathcal{W} = \mathcal{W}_0$ and $\mathcal{R} = \mathcal{R}_0$. Then for any $\theta \in \Theta$ and $t \in [T]$, for any $\theta \in \Theta$

$$\mathbb{E}[\|\tilde{w}_t(\theta) - w_t^*(\theta)\|_1 \mid \mathcal{H}_{t-1}] \leq \alpha_t \left(1 + 2 \ln(d) C_0 \left(\ln\left(\frac{2}{\alpha_t}\right) + (1 - \beta) \ln^2(\alpha_t \ln(d)) \right) \right).$$

1107 *Proof.* Let $\theta \in \Theta$ and $t \in [T]$. Since $\mathcal{W} = \mathcal{W}_0$, denoting for any $j \in [d]$ e_j the j -th element of
1108 the canonical basis, we have $u_j(\theta, X_t) = \langle g(\theta, X_t), e_j \rangle$. We recall that we denote $u_{I_t(\theta)}(\theta, X_t) =$
1109 $\min_{j \in [d]} u_j(\theta, X_t)$. We first prove the following lemma.

1111 **Lemma 3.** Assume that there exists $\varepsilon > 0$ such that $u_{I_t(\theta)}(\theta, X_t) + \varepsilon \leq u_j(\theta, X_t)$ for any $j \in$
1112 $\{1, \dots, K\} \setminus \{I_t(\theta)\}$. Then,

$$\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 \leq \frac{2\alpha_t \ln(d)}{\varepsilon}.$$

1116 *Proof.* In this proof, $I_t(\theta)$ from **H2** is denoted I_t to lighten notation. We have by assumption:

$$\begin{aligned}
 \langle g(\theta, X_t), \tilde{w}_t(\theta) \rangle &= \left\langle g(\theta, X_t), \sum_{j=1}^d \tilde{w}_{t,j}(\theta) e_j \right\rangle \\
 &= u_{I_t}(\theta, X_t) \tilde{w}_{t,I_t}(\theta) + \sum_{j \neq I_t} u_j(\theta, X_t) \tilde{w}_{t,j}(\theta) \\
 &\geq u_{I_t}(\theta, X_t) \tilde{w}_{t,I_t}(\theta) + (u_{I_t}(\theta, X_t) + \varepsilon) \sum_{j \neq I_t} \tilde{w}_{t,j}(\theta) \\
 &= u_{I_t}(\theta, X_t) \tilde{w}_{t,I_t}(\theta) + (u_{I_t}(\theta, X_t) + \varepsilon)(1 - \tilde{w}_{t,I_t}(\theta)). \tag{17}
 \end{aligned}$$

1126 The last equality holds because $\|\tilde{w}_t(\theta)\|_1 = 1$. On the other hand, using the facts that $\mathcal{R}_0 \geq 0$,
1127 $\tilde{w}_t(\theta) = \text{argmin} \langle g(\theta, X_t), w \rangle + \alpha_t \mathcal{R}_0(w)$ and $u_{I_t}(\theta, X_t) = \langle g(\theta, X_t), w_t^*(\theta) \rangle$ yields:

$$\begin{aligned}
 \langle g(\theta, X_t), \tilde{w}_t(\theta) \rangle &\leq \langle g(\theta, X_t), \tilde{w}_t(\theta) \rangle + \alpha_t \mathcal{R}_0(\tilde{w}_t(\theta)) \\
 &\leq \langle g(\theta, X_t), w_t^*(\theta) \rangle + \alpha_t \mathcal{R}_0(w_t^*(\theta)) \\
 &= u_{I_t}(\theta, X_t) + \alpha_t \mathcal{R}_0(w_t^*(\theta)) \\
 &\leq u_{I_t}(\theta, X_t) + \alpha_t \ln(d). \tag{18}
 \end{aligned}$$

1134 The last line holds because $\mathcal{R}_0(w) \leq \ln(d)$ for any $w \in \mathcal{W}_0$. Combining (17) and (18) gives:
 1135

$$1136 \quad u_{I_t}(\theta, X_t) \tilde{w}_{t, I_t}(\theta) + (u_{I_t} + \varepsilon)(1 - \tilde{w}_{t, I_t}(\theta)) \leq u_{I_t}(\theta, X_t) + \alpha_t \ln(d).$$

1138 Re-organizing the terms then yields $\tilde{w}_{t, I_t}(\theta) \geq 1 - \frac{\alpha_t \ln(d)}{\varepsilon}$ and again, as $\|\tilde{w}_t(\theta)\|_1 = 1$, we have:
 1139 $\sum_{j \neq I_t} \tilde{w}_{t, j}(\theta) \leq \frac{\alpha_t \ln(d)}{\varepsilon}$.
 1140

1141 Combining those inequalities yields:

$$1142 \quad \|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 = \|e_{I_t} - \tilde{w}_t(\theta)\|_1 = 1 - \tilde{w}_{t, I_t}(\theta) + \sum_{j \neq I_t} \tilde{w}_{t, j}(\theta) \leq \frac{2\alpha_t \ln(d)}{\varepsilon}.$$

□

1144

1145

To conclude the proof, we first write the expected distance as follows:

$$1146 \quad \mathbb{E}[\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 \mid \mathcal{H}_{t-1}] = \int_0^{+\infty} \mathbb{P}(\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 > y \mid \mathcal{H}_{t-1}) dy,$$

1147 Since both $w_t^*(\theta)$ and $\tilde{w}_t(\theta)$ belong to the simplex, we can restrict the integral to:

$$1148 \quad = \int_0^2 \mathbb{P}(\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 > y \mid \mathcal{H}_{t-1}) dy \quad (19)$$

$$1149 \quad = \int_0^{\alpha_t} \mathbb{P}(\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 > y \mid \mathcal{H}_{t-1}) dy \quad (20)$$

$$1150 \quad + \int_{\alpha_t}^2 \mathbb{P}(\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 > y \mid \mathcal{H}_{t-1}) dy \quad (21)$$

1151 We now apply the change of variable $y = 2\alpha_t \ln(d)\varepsilon^{-1}$ to obtain:

$$1152 \quad \mathbb{E}[\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 \mid \mathcal{H}_{t-1}] \\ 1153 \quad = \alpha_t + 2\alpha_t \ln(d) \int_{(\alpha_t \ln(d))}^{(2 \ln(d))} \mathbb{P}\left(\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 > 2 \frac{\alpha_t \ln(d)}{\varepsilon} \mid \mathcal{H}_{t-1}\right) \frac{d\varepsilon}{\varepsilon^2}. \quad (22)$$

1154 Moreover, for any $\varepsilon > 0$ we have by H2 and Lemma 3 that:

$$1155 \quad 1 - C_0 \varepsilon^\beta \leq \mathbb{P}\left(\inf_{j \neq I_t} \{u_j(\theta, X_t) - u_{I_t}(\theta, X_t)\} \geq \varepsilon \mid \mathcal{H}_{t-1}\right) \\ 1156 \quad \leq \mathbb{P}(\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 \leq 2\alpha_t \ln(d)\varepsilon^{-1} \mid \mathcal{H}_{t-1}),$$

1157

1158

so it follows that:

$$1159 \quad \mathbb{P}_t\left(\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 > 2 \frac{\alpha_t \ln(d)}{\varepsilon} \mid \mathcal{H}_{t-1}\right) = 1 - \mathbb{P}_t\left(\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 \leq 2 \frac{\alpha_t \ln(d)}{\varepsilon} \mid \mathcal{H}_{t-1}\right) \\ 1160 \quad \leq C_0 \varepsilon^\beta, \quad (23)$$

1161 Hence, plugging (23) in (22) gives:

$$1162 \quad \mathbb{E}[\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1 \mid \mathcal{H}_{t-1}] \leq \alpha_t + 2\alpha_t \ln(d) C_0 \int_{\alpha_t \ln(d)}^{2 \ln(d)} \varepsilon^{\beta-2} d\varepsilon.$$

1163 We now proceed to bound $\varphi(\beta) = \int_{\alpha_t \ln(d)}^{2 \ln(d)} \varepsilon^{\beta-2} d\varepsilon$ for $\beta \in (0, 1)$. With $a = \alpha_t \ln(d)$ and $b = 2 \ln(d)$, we have:

$$1164 \quad \varphi(\beta) = (1 - \beta)^{-1} (a^{\beta-1} - b^{\beta-1}) = (1 - \beta)^{-1} [e^{-(1-\beta) \ln(a)} - e^{-(1-\beta) \ln(b)}],$$

1165 and since for any $x > 0$, $1 - x \leq e^{-x} \leq 1 - x + x^2$:

$$1166 \quad \varphi(\beta) \leq (1 - \beta)^{-1} [1 - (1 - \beta) \ln(a) + (1 - \beta)^2 \ln^2(a) - 1 + (1 - \beta) \ln(b)] \\ 1167 \quad = \ln\left(\frac{b}{a}\right) + (1 - \beta) \ln^2(a).$$

1168 Substituting in the values of a and b yields the desired result. □

1188 G.1.2 GENERAL POLYTOPE
11891190 In this section, we denote by \mathcal{W}_1 the general polytope described in ?? and \mathcal{R}_1 the log-barrier regu-
1191 larization described in Equation (7).1192 **Lemma 4.** *Assume **H1**, $\mathcal{W} = \mathcal{W}_1$ and $\mathcal{R} = \mathcal{R}_1$. Then for any $t \in [T]$, \tilde{f}_t is differentiable and for
1193 any $\theta \in \Theta$, $\nabla \tilde{f}_t(\theta) = \nabla \tilde{w}_t(\theta) \bar{g}_t(X_t)$ where:*

1194
1195
$$\nabla \tilde{w}_t(\theta) = -\alpha_t^{-1} \left(\sum_{i=1}^n (b_i - A_i^\top \tilde{w}_t(\theta))^{-2} A_i A_i^\top \right) \nabla_\theta g(\theta, X_t) .$$

1196
1197

1198 *Proof.* Let $\theta \in \mathbb{R}$ and $t \in [T]$. With

1199
$$h_t : (w, \theta) \mapsto \langle g(\theta, X_t), w \rangle + \alpha_t \mathcal{R}_1(w) ,$$

1200

1201 we have $h_t(w) \rightarrow +\infty$ as $w \rightarrow \text{bdry}(\mathcal{W}_1)$. Since $h_t(\tilde{w}_t(\theta), \theta) = \min_{w \in \mathcal{W}_1} h_t(w, \theta)$, we deduce
1202 that $\tilde{w}_t(\theta) \in \text{int}(\mathcal{W}_1)$. It follows from the first order condition and the implicit function theorem
1203 (de Oliveira, 2014, Theorem 2) that $\tilde{w}_t : \theta \mapsto \tilde{w}_t(\theta)$ is differentiable, and

1204
$$\begin{aligned} \nabla \tilde{w}_t(\theta) &= -(\nabla_{ww}^2 h_t(\tilde{w}_t(\theta), \theta))^{-1} \nabla_{\theta w}^2 h_t(\tilde{w}_t(\theta), \theta) \\ &= -\frac{1}{\alpha_t} [\nabla^2 \mathcal{R}_1(\tilde{w}_t(\theta))]^{-1} \nabla_\theta g(\theta, X_t) . \end{aligned} \quad (24)$$

1205
1206

1207 Since $\mathcal{R}_1(w) = -\sum_{i=1}^n \ln(b_i - A_i w)$, simple computations give:

1208
1209
$$\nabla^2 \mathcal{R}_1(\tilde{w}_t(\theta)) = \sum_{i=1}^n (b_i - A_i^\top \tilde{w}_t(\theta))^{-2} A_i A_i^\top , \quad (25)$$

1210
1211

1212 Moreover, since $\text{rank}(AA^\top) = d$ by assumption, $\nabla^2 \mathcal{R}_1(\tilde{w}_t(\theta))$ is invertible. Plugging (25) in (24)
1213 gives the result. \square 1214 **Lemma 5.** *Assume **H1** and $\mathcal{R} = \mathcal{R}_1$. For any $t \in [T]$, \tilde{f}_t is K_t -Lipschitz almost-surely, with*

1215
1216
$$K_t = \frac{GD_{\mathcal{Z}} \sup_{w \in \mathcal{W}_1} \|b - Aw\|_\infty}{\alpha_t \lambda_{\min}(AA^\top)} .$$

1217
1218

1219 *Proof.* Let $(\theta, \theta') \in \Theta^2$. We proved in (15) that:

1220
1221
$$\left| \tilde{f}_t(\theta) - \tilde{f}_t(\theta') \right| \leq D_{\mathcal{Z}} \int_{t=0}^1 \|\nabla \tilde{w}_t(\theta + t(\theta' - \theta))\|_{\text{op}} \|\theta - \theta'\| dt .$$

1222

1223 Denoting $\zeta = \theta + t(\theta' - \theta)$, by (24), since $\|\cdot\|_{\text{op}}$ is sub-multiplicative:

1224
1225
$$\begin{aligned} \|\nabla \tilde{w}_t(\zeta)\|_{\text{op}} &= -\alpha_t^{-1} \|[\nabla^2 \mathcal{R}_1(\tilde{w}_t(\zeta))]^{-1} \nabla_\theta g(\theta, X_t)\|_{\text{op}} \\ &\leq -\alpha_t^{-1} \|[\nabla^2 \mathcal{R}_1(\tilde{w}_t(\zeta))]^{-1}\|_{\text{op}} G \quad (\text{by H1-(iii)}) . \end{aligned}$$

1226
1227

1228 Moreover, we have:

1229
1230
$$\begin{aligned} \|[\nabla^2 \mathcal{R}_1(\zeta)]^{-1}\|_{\text{op}} &= \left\| \left(\sum_{i=1}^n (b_i - A_i^\top \tilde{w}_t(\zeta))^{-2} A_i A_i^\top \right)^{-1} \right\|_{\text{op}} \\ &\leq \sup_{w \in \mathcal{W}_1} \|b - Aw\|_\infty^2 \|(AA^\top)^{-1}\|_{\text{op}} , \end{aligned}$$

1231
1232
1233

1234 Since AA^\top is symmetric, $\|(AA^\top)^{-1}\|_{\text{op}} = \lambda_{\min}(AA^\top)^{-1}$, where $\lambda_{\min}(M)$ is the lowest eigenvalue
1235 of M . We therefore obtain by (15) and the previous inequalities that almost-surely:

1236
1237
$$\begin{aligned} \left| \tilde{f}_t(\theta) - \tilde{f}_t(\theta') \right| &\leq \frac{GD_{\mathcal{Z}} \sup_{w \in \mathcal{W}_1} \|b - Aw\|_\infty^2}{\alpha_t \lambda_{\min}(AA^\top)} \int_{t=0}^1 \|\theta - \theta'\| dt \\ &= \frac{GD_{\mathcal{Z}} \sup_{w \in \mathcal{W}_1} \|b - Aw\|_\infty^2}{\alpha_t \lambda_{\min}(AA^\top)} \|\theta - \theta'\| . \end{aligned}$$

1238
1239
1240
1241

 \square

1242 **Lemma 6.** Assume **H2**, $\mathcal{W} = \mathcal{W}_1$ and $\mathcal{R} = \mathcal{R}_1$. Then for any $t \in [T]$ and $\theta \in \Theta$,

$$1244 \mathbb{E}[\|\tilde{w}_t(\theta) - w_t^*(\theta)\|_1 \mid \mathcal{H}_{t-1}] \leq \alpha_t \left(1 + 2n \max_{k \in [K]} \|v_k\|_1 C_0 \ln \left(\frac{2}{\alpha_t} \right) + (1 - \beta) \ln^2(\alpha_t \ln(d)) \right).$$

1246

1247 *Proof.* Let $\theta \in \Theta$ and $t \in [T]$. Since $\mathcal{W}_1 = \text{Conv}(v_1, \dots, v_K)$, there exists
1248 $(\lambda_{t,I_t}(\theta), \dots, \lambda_{t,I_t}(\theta)) \in [0, 1]^K$ such that

$$1249 \tilde{w}_t(\theta) = \sum_{i=1}^K \lambda_{t,i}(\theta) v_i \quad \text{and} \quad \sum_{i=1}^K \lambda_{t,i}(\theta) = 1.$$

1252 We recall that $u_j(\theta, X_t) = \langle g(\theta, X_t), v_j \rangle$ denotes the value of the objective function on the vertex
1253 $v_j \in \mathcal{W}$, and $u_{I_t}(\theta, X_t) = \min_{j \in [K]} u_j(\theta, X_t)$ so $v_{I_t} = w_t^*(\theta)$. We start by proving the following
1254 lemma:

1255 **Lemma 7.** If there exists $\varepsilon > 0$ such that $u_{I_t} + \varepsilon \leq u_j$ for any $j \in \{1, \dots, K\} \setminus \{I_t\}$, then,

$$1257 \|\tilde{w}_t(\theta) - w_t^*(\theta)\|_1 \leq \frac{2n\alpha_t}{\varepsilon} \max_{k \in [K]} \|v_k\|_1.$$

1259

1260 *Proof.* On the one hand, $\tilde{w}_t(\theta)$ is by definition the solution to the problem:

$$1262 \tilde{w}_t(\theta) = \underset{w \in \mathcal{W}}{\text{argmin}} \langle g(\theta, X_t), w \rangle - \alpha_t \sum_{i=1}^d \ln(b_i - A_i w), \quad (26)$$

1264 which is uniquely determined by strong convexity of the objective. We know by (Boyd, 2004, page
1265 566) that:

$$1266 \langle g(\theta, X_t), \tilde{w}_t(\theta) \rangle - \inf_{w \in \mathcal{W}_1} \langle g(\theta, X_t), w \rangle \leq n\alpha_t,$$

1268 that is

$$1269 \langle g(\theta, X_t), \tilde{w}_t(\theta) \rangle \leq u_{I_t} + n\alpha_t. \quad (27)$$

1270 On the other hand, by assumption we know that for any $j \neq I_t$, $u_j \geq u_{I_t} + \varepsilon$ so:

$$1272 \begin{aligned} \langle g(\theta, X_t), \tilde{w}_t(\theta) \rangle &= \left\langle g(\theta, X_t), \sum_{j=1}^K \lambda_{t,j}(\theta) v_j \right\rangle \\ 1273 &= \lambda_{t,I_t}(\theta) u_{I_t} + \sum_{j \neq I_t} \lambda_{t,j}(\theta) u_j \\ 1274 &\geq \lambda_{t,I_t}(\theta) u_{I_t} + \sum_{j \neq I_t} \lambda_{t,j}(\theta) (u_{I_t} + \varepsilon) \\ 1275 &= \lambda_{t,I_t}(\theta) u_{I_t} + (u_{I_t} + \varepsilon)(1 - \lambda_{t,I_t}(\theta)). \end{aligned} \quad (28)$$

1276 Combining (27) and (28) yields:

$$1277 \lambda_{t,I_t}(\theta) \geq 1 - \frac{n\alpha_t}{\varepsilon}, \quad (30)$$

1278 and it follows from (30) that:

$$1279 \begin{aligned} \|\tilde{w}_t(\theta) - w_t^*(\theta)\|_1 &= \|v_{I_t} - \tilde{w}_t(\theta)\|_1 = \left\| \sum_{j=1}^K (\mathbb{1}_{\{j=I_t\}} - \lambda_{t,j}(\theta)) v_j \right\|_1 \\ 1280 &\leq [(1 - \lambda_{t,I_t}(\theta)) + \sum_{j \neq I_t} \lambda_{t,j}(\theta)] \max_{k \in [K]} \|v_k\|_1 \\ 1281 &\leq \frac{2n\alpha_t}{\varepsilon} \max_{k \in [K]} \|v_k\|_1. \end{aligned}$$

1282

1283 The result follows from combining **H2** and Lemma 7 according to the same lines of computation as
1284 in the proof of Lemma 2. \square

1296 G.1.3 OTHER LEMMAS
1297

1298 **Lemma 8.** Assume that for any $t \in [T]$, \tilde{f}_t is K_t -Lipschitz almost-surely and that the sequence of
1299 steps $(\eta_t)_{t \geq 1}$ is non-increasing. Denote $\mathfrak{R}_T^d = \sum_{t \in [T]} \tilde{f}_t(\theta_t) - \tilde{f}_t(\vartheta_t)$. Then it holds almost surely
1300 that:

$$1301 \mathbb{E}[\mathfrak{R}_T^d] \leq \frac{1}{2\eta_T} \left(D_\Theta^2 + 2D_\Theta \sum_{t=1}^T \|\vartheta_{t+1} - \vartheta_t\| \right) + \sum_{t=1}^T \frac{\eta_t}{2} K_t^2,$$

1304 where the expectation is taken over the sequence (u_1, \dots, u_T) in Algorithm 2.

1305 *Proof.* In what follows, we write \mathbb{E}_{u_t} the expectation under the distribution $\text{Unif}([0, 1])$ of u_t , and
1306 $\mathbb{E}_{u_{1:T}}$ the expectation under the joint distribution $\text{Unif}([0, 1])^{\otimes T}$ of (u_1, \dots, u_T) . For any $t \in [T]$,
1307 we have:

$$1309 \begin{aligned} 1310 \sum_{t=1}^T \tilde{f}_t(\theta_t) - \tilde{f}_t(\vartheta_t) &= \sum_{t=1}^T \int_{u=0}^1 \left\langle \nabla \tilde{f}_t(\vartheta_t + u(\theta_t - \vartheta_t)), \theta_t - \vartheta_t \right\rangle du \\ 1311 &= \sum_{t=1}^T \left\langle \mathbb{E}_{u_t} [\nabla \tilde{f}_t(\vartheta_t + u_t(\theta_t - \vartheta_t))], \theta_t - \vartheta_t \right\rangle \\ 1312 &= \sum_{t=1}^T \left\langle \mathbb{E}_{u_t} [\tilde{\nabla}_t(u_t)], \theta_t - \vartheta_t \right\rangle, \end{aligned} \quad (31)$$

1318 where $\tilde{\nabla}(u_t) = \nabla \tilde{f}_t(\vartheta_t + u_t(\theta_t - \vartheta_t))$ as in Algorithm 2. To control (31) we note that by definition
1319 of θ_{t+1} in Algorithm 2,

$$1321 \begin{aligned} 1322 \mathbb{E}_{u_t} [\|\theta_{t+1} - \vartheta_t\|^2] &= \mathbb{E}_{u_t} \left[\left\| \Pi_\Theta[\theta_t - \eta_t \tilde{\nabla}_t(u_t)] - \vartheta_t \right\|^2 \right] \\ 1323 &\leq \mathbb{E}_{u_t} \left[\left\| \theta_t - \vartheta_t - \eta_t \tilde{\nabla}_t(u_t) \right\|^2 \right] \\ 1324 &\leq \|\theta_t - \vartheta_t\|^2 - 2\eta_t \left\langle \mathbb{E}_{u_t} [\tilde{\nabla}_t(u_t)], \theta_t - \vartheta_t \right\rangle + \eta_t^2 K_t^2 \end{aligned} \quad (32)$$

1327 where we have used that $\|\tilde{\nabla}_t(u_t)\|^2 \leq K_t^2$ because \tilde{f}_t is K_t -Lipschitz. Re-arranging (32) and
1328 taking the expectation over (u_1, \dots, u_T) yields:

$$1331 \mathbb{E}_{u_{1:T}} \left[\left\langle \tilde{\nabla}_t(u_t), \theta_t - \vartheta_t \right\rangle \right] \leq \mathbb{E}_{u_{1:T}} \left[\frac{1}{2\eta_t} \left(\|\theta_t - \vartheta_t\|^2 - \|\theta_{t+1} - \vartheta_t\|^2 \right) + \frac{\eta_t K_t^2}{2} \right], \quad (33)$$

1334 Now notice that, for any $t \in \{1, \dots, T-1\}$,

$$1336 \begin{aligned} 1337 \|\theta_{t+1} - \vartheta_{t+1}\|^2 &= \|\theta_{t+1} - \vartheta_t\|^2 + \|\theta_{t+1} - \vartheta_{t+1}\|^2 - \|\theta_{t+1} - \vartheta_t\|^2 \\ 1338 &= \|\theta_{t+1} - \vartheta_t\|^2 \\ 1339 &\quad + (\|\theta_{t+1} - \vartheta_{t+1}\| + \|\theta_{t+1} - \vartheta_t\|)(\|\theta_{t+1} - \vartheta_{t+1}\| - \|\theta_{t+1} - \vartheta_t\|) \\ 1340 &\leq \|\theta_{t+1} - \vartheta_t\|^2 + 2D_\Theta \|\vartheta_{t+1} - \vartheta_t\|, \end{aligned} \quad (34)$$

1341 where we have used **H1**-(i) and the reversed triangular inequality in the last line. It follows that:

$$1343 -\|\theta_{t+1} - \vartheta_t\|^2 \leq -\|\theta_{t+1} - \vartheta_{t+1}\|^2 + 2D_\Theta \|\vartheta_{t+1} - \vartheta_t\| \quad (35)$$

1344 Plugging (35) into (33) then gives:

$$1346 \begin{aligned} 1347 \mathbb{E}_{u_{1:T}} \left[\left\langle \tilde{\nabla}_t(u_t), \theta_t - \vartheta_t \right\rangle \right] &\leq \mathbb{E}_{u_{1:T}} \left[\frac{1}{2\eta_t} \left(\|\theta_t - \vartheta_t\|^2 - \|\theta_{t+1} - \vartheta_{t+1}\|^2 \right) \right. \\ 1348 &\quad \left. + \frac{1}{\eta_t} D_\Theta \|\vartheta_{t+1} - \vartheta_t\| + \frac{\eta_t K_t^2}{2} \right]. \end{aligned}$$

1350 Thus, summing over T and plugging the resulting sum in (31) gives:
 1351

$$\begin{aligned}
 1352 \mathbb{E}_{u_{1:T}} \left[\sum_{t=1}^T \tilde{f}_t(\theta_t) - \tilde{f}_t(\vartheta_t) \right] &\leq \mathbb{E}_{u_{1:T}} \left[\sum_{t=1}^T \frac{\|\theta_t - \vartheta_t\|^2}{2} \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) \right. \\
 1353 &\quad \left. + \sum_{t=1}^T \frac{1}{\eta_t} D_\Theta \|\vartheta_{t+1} - \vartheta_t\| + \sum_{t=1}^T \frac{\eta_t K_t^2}{2} \right] . \\
 1354 &\leq \frac{D_\Theta^2}{2} \sum_{t=1}^T \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) + \frac{D_\Theta}{\eta_T} \sum_{t=1}^T \|\vartheta_{t+1} - \vartheta_t\| + \sum_{t=1}^T \frac{\eta_t K_t^2}{2} .
 \end{aligned}$$

1361 where we used that for all t , $\|\theta_t - \vartheta_t\| \leq D_\Theta$ by **H1**-(i) and that for all $t \in [T]$, $\eta_{t-1} \geq \eta_t \geq \eta_T$.
 1362 Then, we have by telescoping the first sum:

$$\mathbb{E}_{u_{1:T}} \left[\sum_{t=1}^T \tilde{f}_t(\theta_t) - \tilde{f}_t(\vartheta_t) \right] \leq \frac{D_\Theta^2 + 2D_\Theta}{2\eta_T} \sum_{t=1}^T \|\vartheta_{t+1} - \vartheta_t\| + \sum_{t=1}^T \frac{\eta_t K_t^2}{2} .$$

□

1363 G.2 PROOF OF THEOREM 1.

1370 **Theorem 1.** *Assume **H1**, **H2** and having access to an ξ -approximate oracle \mathbf{O}_ξ adapted to
 1371 $\{\sum_{i=1}^t \tilde{f}_i - \langle \sigma_t, \cdot \rangle\}_{t \in [T]}$. Let $\{\theta_t\}_{t \in [T]}$ be the output of DF-FTPL (Algorithm 1) instantiated with
 1372 learning step $\eta > 0$ and regularization coefficients $\alpha_t = \alpha > 0$ for any t . Then:*

$$1374 T^{-1} \mathbb{E}[\mathfrak{R}_T^s] = \tilde{\mathcal{O}} \left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi + \alpha n \right) ,$$

1377 where \mathbb{E} denotes the expectation on both data and the intrinsic randomness of DF-FTPL and $\tilde{\mathcal{O}}$
 1378 contains polynomial dependency in $\ln(1/\alpha), \ln(\ln(d))$.

1379 Furthermore, taking $\eta \propto m^{1/4} T^{-3/4} n^{-1/2}$ and $\alpha \propto m^{3/4} n^{1/2} T^{-1/4}$ yields:

$$1380 T^{-1} \mathbb{E}[\mathfrak{R}_T^s] = \tilde{\mathcal{O}} \left(m^{3/4} \sqrt{n} T^{-1/4} + \xi \right) .$$

1383 *Proof.* For the sake of conciseness we use the notation \mathbb{E}_t to denote $\mathbb{E}[\cdot \mid \mathcal{H}_{t-1}]$ In what follows, we
 1384 consider, for any θ , the intermediary regret: $R_T^s(\theta) := \sum_{t=1}^T F_t(\theta_t) - F_t(\theta)$, where $F_t : \theta \mapsto$
 1385 $\mathbb{E}[f_t(\theta) \mid \mathcal{H}_{t-1}]$. We have, for any θ , the following decomposition of the static regret
 1386

$$1387 R_T^s(\theta) = \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta_t) - \tilde{w}_t(\theta_t) \rangle] \tag{36}$$

$$1388 + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) \rangle] - \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] \tag{37}$$

$$1389 + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] - \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) \rangle] \tag{38}$$

1394 Then, taking the sup over Θ in Equations (36) and (38), and defining

$$1397 \text{Reg}_T^S := \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) \rangle] - \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] ,$$

1399 in Equation (37) yields for any θ :

$$1402 R_T^s(\theta) \leq \text{Reg}_T^S + 2 \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) - \tilde{w}_t(\theta) \rangle]$$

1404 Using the fact that $\|\cdot\|_\infty$ is dual to $\|\cdot\|_1$:

$$1406 \quad R_T^s(\theta) \leq \text{Reg}_T^S + 2 \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t [\|\bar{g}_t(X_t)\|_\infty \|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1]$$

1408 Since for any $t \in [T]$, $\|\bar{g}_t(X_t)\|_\infty \leq \|\bar{g}_t(X_t)\|_2 \leq D_Z$ by **H1**-(ii),

$$1409 \quad R_T^s(\theta) \leq \text{Reg}_T^S + 2D_Z \sum_{t \in [T]} \mathbb{E}_t \sup_{\theta \in \Theta} \|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1$$

1412 First, notice that, by Lemma 6, and because for any t , $\alpha_t = \alpha$ we have for all t :

$$1414 \quad \mathbb{E}[\|\tilde{w}_t(\theta) - w_t^*(\theta)\|_1 \mid \mathcal{H}_{t-1}] \leq \alpha \left(1 + 2n \max_{k \in [K]} \|v_k\|_1 C_0 \ln \left(\frac{2}{\alpha} \right) + (1 - \beta) \ln^2(\alpha_t \ln(d)) \right). \quad (39)$$

1417 Second, taking the expectation (denoted by \mathbb{E}) over $(\sigma_t)_{t \in [T]}$ and X_1, \dots, X_T on both sides gives:

$$1419 \quad \mathbb{E}[R_T^s(\theta)] \leq \mathbb{E}[\text{Reg}_T^S] + 2D_Z \alpha T \left(1 + 2n \max_{k \in [K]} \|v_k\|_1 C_0 \ln \left(\frac{2}{\alpha} \right) + (1 - \beta) \ln^2(\alpha_t \ln(d)) \right). \quad (40)$$

1422 To bound the first term in the right-hand-side of Equation (40). by the definition of conditional expectation, we have:

$$1425 \quad \mathbb{E}[\text{Reg}_T^S] = \mathbb{E}_{X_1, \dots, X_T} \mathbb{E}_{\sigma_1, \dots, \sigma_T} \left[\sum_{t \in [T]} \langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) \rangle - \langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle \right]$$

1428 Then,

$$1430 \quad \mathbb{E}[\text{Reg}_T^S] \leq \mathbb{E}_{X_1, \dots, X_T} \mathbb{E}_{\sigma_1, \dots, \sigma_T} \left[\sum_{t \in [T]} \tilde{f}_t(\theta_t) - \inf_{\theta \in \Theta} \sum_{t \in [T]} \tilde{f}_t(\theta) \right].$$

1433 One recognises (up to a factor T) the left-hand side of Proposition 4 on the loss sequence $(\tilde{f}_t)_{t \in [T]}$.

1434 Furthermore, we can use this proposition as, given our choice of \mathcal{R} , for any $t \in [T]$, \tilde{f}_t is
1435 L -Lipschitz with $L = 5D_Z G(4\alpha)^{-1}$ almost surely by Lemma 1 (with $\chi = 0$). We then have:

$$1437 \quad \mathbb{E}_{\sigma_1, \dots, \sigma_T} \left[\frac{1}{T} \sum_{t \in [T]} \tilde{f}_t(\theta_t) - \inf_{\theta \in \Theta} \frac{1}{T} \sum_{t \in [T]} \tilde{f}_t(\theta) \right] \leq \mathcal{O} \left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi \right). \quad (41)$$

1440 Dividing Equation (40) by T , and plugging Equation (41) gives: for all $\theta \in \Theta$:

$$1442 \quad T^{-1} \mathbb{E}[R_T^s(\theta)] \leq \mathcal{O} \left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi \right) \\ 1443 \quad + 2D_Z \alpha \left(1 + 2 \ln(d) C_0 \left(\ln \left(\frac{2}{\alpha} \right) + (1 - \beta) \ln^2(\alpha \ln(d)) \right) \right), \\ 1444 \quad = \tilde{\mathcal{O}} \left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi + \alpha n \right).$$

1449 Finally remark that, by the definition of the conditional expectation (thus of F_t), and because θ_t is
1450 \mathcal{F}_{t-1} -measurable, we have for any θ :

$$1453 \quad \mathbb{E}[R_T^s(\theta)] = \mathbb{E} \left[\sum_{t=1}^T f_t(\theta_t) - f_t(\theta) \right] \\ 1454 \quad = \mathbb{E} \left[\sum_{t=1}^T f_t(\theta_t) \right] - \mathbb{E}[f_t(\theta)]$$

1458 Then taking the infimum over θ yields:
 1459

$$1460 \quad T^{-1}\mathbb{E}[\mathfrak{R}_T^s] \leq \tilde{\mathcal{O}}\left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi + \alpha n\right).$$

1462 This concludes the proof. The second equation consists in simply plugging the proposed value of
 1463 η, α in this bound. \square
 1464

1465 **G.3 PROOF OF THEOREM 2**
 1466

1467 **Theorem 2.** *Assume **H1**, **H2**, access to a ξ -approximate oracle adapted to $\{\tilde{f}_t\}_{t \in [T]}$. Let $\{\theta_t\}_{t \in [T]}$ be the output of DF-OGD (Algorithm 2) instantiated with the non-increasing sequence $(\eta_t)_{t \in [T]}$ and regularization coefficients $(\alpha_t)_{t \in [T]}$. Then:*

$$1470 \quad 1471 \quad T^{-1}\mathbb{E}[\mathfrak{R}_T^d] = \tilde{\mathcal{O}}\left(\mathbb{E}\left[\frac{1+P_T}{T\eta_T} + \frac{1}{T} \sum_{t \in [T]} \frac{\eta_t}{\alpha_t^2} + n\alpha_t\right] + \xi\right).$$

1474 where $P_T = \sum_{t=1}^{T-1} \|\vartheta_{t+1} - \vartheta_t\|$, \mathbb{E} denotes the expectation on both data and the intrinsic randomness of DF-OGD and $\tilde{\mathcal{O}}$ contains polynomial dependency in $\ln(1/\alpha), \ln(\ln(d))$.
 1475

1476 Furthermore, assume $(t^{-1}(1 + P_t))_{t \geq 1}$ is non-increasing almost surely with $P_t = \sum_{s=1}^t \|\vartheta_{s+1} - \vartheta_s\|$. Then, using $\alpha_t \propto n^{-1/2}t^{-1/4}(1+P_t)^{1/4}$ and $\eta_t \propto n^{-1/2}t^{-3/4}(1+P_t)^{3/4}$ for any $t \in [T]$ leads to :

$$1480 \quad T^{-1}\mathbb{E}[\mathfrak{R}_T^d] = \tilde{\mathcal{O}}\left(\mathbb{E}\left[\sqrt{n}(1+P_T)^{1/4}T^{-1/4}\right] + \xi\right).$$

1482 *Proof.* In this proof, for the sake of conciseness, we use the notation \mathbb{E}_t to denote $\mathbb{E}[\cdot | \mathcal{H}_{t-1}]$.
 1483 Observe that the dynamic regret can be decomposed as follows:
 1484

$$1485 \quad \mathfrak{R}_T^d = \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta_t) - \tilde{w}_t(\theta_t) \rangle] + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) - \tilde{w}_t(\vartheta_t) \rangle] \quad (42)$$

$$1486 \quad + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\vartheta_t) \rangle] - \inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] \quad (43)$$

$$1487 \quad + \sum_{t \in [T]} \left[\inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] - \inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) \rangle] \right] \quad (44)$$

1493 First, we remark that for any t , given $\vartheta_t = \mathbf{O}(\tilde{f}_t)$, we control (43) by Jensen:
 1494

$$1495 \quad \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\vartheta_t) \rangle] - \inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] \leq \mathbb{E}_t\left[\tilde{f}_t(\vartheta_t) - \inf_{\theta \in \Theta} \tilde{f}_t(\theta)\right] \\ 1496 \quad \leq \xi.$$

1498 Then, taking the sup over Θ for each summand of the first sum of (42) (which is valid as θ_t is
 1499 \mathcal{F}_{t-1} -measurable), defining by $\text{Reg}_T := \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) - \tilde{w}_t(\vartheta_t) \rangle]$, and noticing for
 1500 (44) that $\inf_{\theta \in \Theta} \mathbb{E}_t[f_t(\theta)] - \inf_{\theta \in \Theta} \mathbb{E}_t[\tilde{f}_t(\theta)] \leq \sup_{\theta \in \Theta} \mathbb{E}_t[f_t(\theta) - \tilde{f}_t(\theta)]$ leads to:
 1501

$$1502 \quad \mathfrak{R}_T^d \leq \text{Reg}_T + 2 \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) - \tilde{w}_t(\theta) \rangle] + \xi T$$

1504 Using the fact that $\|\cdot\|_\infty$ is dual to $\|\cdot\|_1$:

$$1506 \quad \mathfrak{R}_T^d \leq \text{Reg}_T + 2 \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t[\|\bar{g}_t(X_t)\|_\infty \|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1] + \xi T$$

1509 Since for any $t \in [T]$, $\|\bar{g}_t(X_t)\|_\infty \leq \|\bar{g}_t(X_t)\|_2 \leq D_Z$ almost surely by **H1**-(ii),
 1510

$$1511 \quad \mathfrak{R}_T^d \leq \text{Reg}_T + 2D_Z \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t[\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1] + \xi T \quad (45)$$

1512 First, by Lemma 6, we know have for any t :
 1513

$$1515 \sup_{\theta \in \Theta} \mathbb{E}_t[\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1] \leq \alpha_t \left(1 + 2n \max_{k \in [K]} \|v_k\|_1 C_0 \ln\left(\frac{2}{\alpha_t}\right) + (1 - \beta) \ln^2(\alpha_t \ln(d)) \right).$$

1518 Thus combining the two last equations and taking the expectation (denoted by \mathbb{E}) over $(u_t)_{t \in [T]}$ and
 1519 X_1, \dots, X_T on both sides gives:
 1520

$$1522 \mathbb{E}[\mathfrak{R}_T^d] \leq \mathbb{E}[\text{Reg}_T] \\ 1523 + 2D_{\mathcal{Z}} \sum_{t \in [T]} \alpha_t \left(1 + 2n \max_{k \in [K]} \|v_k\|_1 C_0 \ln\left(\frac{2}{\alpha_t}\right) + (1 - \beta) \ln^2(\alpha_t \ln(d)) \right) + \xi T. \quad (46)$$

1527 Now, to bound the first term in Equation (46) note that the definition of conditional expectation
 1528 implies:
 1529

$$1530 \mathbb{E}[\text{Reg}_T] = \mathbb{E}_{X_1 \dots X_T} \mathbb{E}_{u_1, \dots, u_T} \left[\sum_{t \in [T]} \tilde{f}_t(\theta_t) - \tilde{f}_t(\vartheta_t) \right].$$

1534 One recognizes the definition of $\tilde{\mathfrak{R}}_T^d$ of Lemma 8. Since \tilde{f}_t is K_t -Lipschitz with $K_t =$
 1535 $\frac{GD_{\mathcal{Z}} \sup_{w \in \mathcal{W}_1} \|b - Aw\|_{\infty}}{\alpha_t \lambda_{\min}(AA^{\top})}$ almost surely for any t by Lemma 5, we deduce from Lemma 8 that:
 1537

$$1538 \mathbb{E}_{u_1, \dots, u_T} \left[\sum_{t \in [T]} \tilde{f}_t(\theta_t) - \tilde{f}_t(\vartheta_t) \right] \leq \frac{D_{\Theta}^2 - D_{\Theta} P_T}{2\eta_T} + \sum_{t \in [T]} \frac{\eta_t}{2} \left(\frac{GD_{\mathcal{Z}} \sup_{w \in \mathcal{W}_1} \|b - Aw\|_{\infty}^2}{\alpha_t \lambda_{\min}(AA^{\top})} \right)^2. \quad (47)$$

1541 Plugging Equation (47) into Equation (46) and then dividing by $T > 0$ on both sides gives:
 1542

$$1544 T^{-1} \mathbb{E}[\mathfrak{R}_T^d] \leq \frac{D_{\Theta}^2 + D_{\Theta} \mathbb{E}[P_T]}{2T\eta_T} + \sum_{t \in [T]} \frac{\eta_t}{2} \left(\frac{GD_{\mathcal{Z}} \sup_{w \in \mathcal{W}_1} \|b - Aw\|_{\infty}^2}{\alpha_t \lambda_{\min}(AA^{\top})} \right)^2 \\ 1545 + 2D_{\mathcal{Z}} \sum_{t \in [T]} \alpha_t [1 + 2n \max_{k \in [K]} \|v_k\|_1 C_0 (\ln(2\alpha_t^{-1}) + (1 - \beta) \ln^2(\alpha_t \ln d))] + \xi. \quad (48)$$

1551 Rewriting Equation (48) using $\tilde{\mathcal{O}}$ concludes the proof.
 1552

1553 Concerning the second bound, We know by Equation (48) that the regret of Algorithm 2 satisfies the
 1554 following bound:
 1555

$$1556 \mathbb{E}[\mathfrak{R}_T^d] = \tilde{\mathcal{O}} \left(\mathbb{E} \left[\frac{1 + P_T}{\eta_T} + \sum_{t \in [T]} \frac{\eta_t}{\alpha_t^2} + n \sum_{t \in [T]} \alpha_t + \xi T \right] \right).$$

1559 With $\alpha_t \propto n^{-1/2} t^{-1/4} (1 + P_t)^{1/4}$ and $\eta_t \propto \alpha_t t^{-1/2} (1 + P_t)^{1/2} \propto n^{-1/2} t^{-3/4} (1 + P_t)^{3/4}$, the
 1560 sequence $(\eta_t)_{t \in [T]}$ is non-increasing because so is $(t^{-1} (1 + P_t))_{t \in [T]}$ by assumption. Moreover, the
 1561 first term satisfies:
 1562

$$1564 \frac{1 + P_T}{\eta_T} = \mathcal{O} \left(n^{1/2} \frac{T^{3/4}}{(1 + P_T)^{3/4}} (1 + P_T) \right) \text{ so } \frac{1 + P_T}{T\eta_T} = \mathcal{O} \left(n^{1/2} \left(\frac{1 + P_T}{T} \right)^{1/4} \right).$$

1566 For the second term,

$$1568 \sum_{t=1}^T \frac{\eta_t}{\alpha_t^2} = \sum_{t=1}^T \frac{\sqrt{1+P_t}}{\sqrt{t}\alpha_t} = \mathcal{O}\left(n^{1/2} \sum_{t=1}^T \left(\frac{1+P_t}{t}\right)^{1/4}\right) \dots$$

1571 Moreover, since $1+P_t \leq 1+P_T$ for any $t \in [T]$, we have:

$$1573 \sum_{t=1}^T \frac{\eta_t}{\alpha_t^2} = \mathcal{O}\left(n^{1/2}(1+P_T)^{1/4} \sum_{t=1}^T \left(\frac{1}{t}\right)^{1/4}\right) = \mathcal{O}\left(n^{1/2}(1+P_T)^{1/4} T^{3/4}\right),$$

1576 where we used in the last line that $\sum_{t=1}^T t^{-\gamma} = \mathcal{O}(T^{1-\gamma})$ for $\gamma \in (0, 1)$. Dividing by T again gives
1577 a rate of $\mathcal{O}\left(\left(\frac{1+P_T}{T}\right)^{1/4}\right)$ for the second term. Finally, we obtain by the same reasoning that
1579

$$1580 n \sum_{t \in [T]} \alpha_t \leq n^{1/2}(1+P_T)^{1/4} \sum_{t \in [T]} t^{-1/4} = \mathcal{O}\left(n^{1/2}(1+P_T)^{1/4} T^{3/4}\right),$$

1582 and dividing by $T > 0$ yields the desired rate. \square

1584 G.4 PROOF OF PROPOSITION 2

1586 **Proposition 2.** Assume **H1**, **H2** and having access to an ξ -approximate optimization oracle **O** $_{\xi}$
1587 adapted to $\left\{\sum_{i=1}^t \tilde{f}_i - \langle \sigma_t, \cdot \rangle\right\}_{t \in [T]}$. Fix $\mathcal{W} = \mathcal{W}_0$, $\mathcal{R} = \mathcal{R}_0$. Let $\{\theta_t\}_{t \in [T]}$ be the output of
1588 DF-FTPL (Algorithm 1) instantiated with learning step $\eta > 0$ and regularization coefficients $\alpha_t =$
1589 $\alpha > 0$ for any t . Then:

$$1591 1592 T^{-1} \mathbb{E}[\mathfrak{R}_T^s] \leq \tilde{\mathcal{O}}\left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi + \alpha \ln(d)\right),$$

1593 where \mathbb{E} denotes the expectation on both data and the intrinsic randomness of DF-FTPL and $\tilde{\mathcal{O}}$
1594 contains polynomial dependency in $\ln(1/\alpha), \ln(\ln(d))$.
1595

1596 *Proof.* For the sake of conciseness we use the notation \mathbb{E}_t to denote $\mathbb{E}[\cdot \mid \mathcal{H}_{t-1}]$ In what follows, we
1597 consider, for any θ , the intermediary regret: $R_T^s(\theta) := \sum_{t=1}^T F_t(\theta_t) - F_t(\theta)$, where $F_t : \theta \mapsto$
1598 $\mathbb{E}[f_t(\theta) \mid \mathcal{H}_{t-1}]$. We have, for any θ , the following decomposition of the static regret

$$1600 1601 R_T^s(\theta) = \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta_t) - \tilde{w}_t(\theta_t) \rangle] \tag{49}$$

$$1602 1603 + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) \rangle] - \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] \tag{50}$$

$$1604 1605 + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] - \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) \rangle] \tag{51}$$

1607 Then, taking the sup over Θ in Equations (49) and (51), and defining

$$1609 1610 \text{Reg}_T^S := \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) \rangle] - \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle],$$

1612 in Equation (50) yields for any θ :

$$1614 1615 R_T^s(\theta) \leq \text{Reg}_T^S + 2 \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) - \tilde{w}_t(\theta) \rangle]$$

1617 Using the fact that $\|\cdot\|_{\infty}$ is dual to $\|\cdot\|_1$:

$$1618 1619 R_T^s(\theta) \leq \text{Reg}_T^S + 2 \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t[\|\bar{g}_t(X_t)\|_{\infty} \|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1]$$

1620 Since for any $t \in [T]$, $\|\bar{g}_t(X_t)\|_\infty \leq \|\bar{g}_t(X_t)\|_2 \leq D_{\mathcal{Z}}$ by **H1**-(ii),
 1621

$$1622 \quad R_T^s(\theta) \leq \text{Reg}_T^S + 2D_{\mathcal{Z}} \sum_{t \in [T]} \mathbb{E}_{\theta} \sup_{\theta \in \Theta} \|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1$$

1624 First, notice that, by Lemma 2, and because for any t , $\alpha_t = \alpha$ we have for all t :
 1625

$$1626 \quad \mathbb{E}[\|\tilde{w}_t(\theta) - w_t^*(\theta)\|_1 \mid \mathcal{H}_{t-1}] \leq \alpha \left(1 + 2 \ln(d) C_0 \left(\ln \left(\frac{2}{\alpha} \right) + (1 - \beta) \ln^2(\alpha \ln(d)) \right) \right). \quad (52)$$

1628 Second, taking the expectation (denoted by \mathbb{E}) over $(\sigma_t)_{t \in [T]}$ and X_1, \dots, X_T on both sides gives:
 1629

$$1630 \quad \mathbb{E}[R_T^s(\theta)] \leq \mathbb{E}[\text{Reg}_T^S] + 2D_{\mathcal{Z}} \alpha T \left(1 + 2 \ln(d) C_0 \left(\ln \left(\frac{2}{\alpha} \right) + (1 - \beta) \ln^2(\alpha \ln(d)) \right) \right). \quad (53)$$

1633 To bound the first term in the right-hand-side of Equation (53). by the definition of conditional
 1634 expectation, we have:

$$1635 \quad \mathbb{E}[\text{Reg}_T^S] = \mathbb{E}_{X_1, \dots, X_T} \mathbb{E}_{\sigma_1, \dots, \sigma_T} \left[\sum_{t \in [T]} \langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) \rangle - \langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle \right]$$

1638 Then,

$$1640 \quad \mathbb{E}[\text{Reg}_T^S] \leq \mathbb{E}_{X_1, \dots, X_T} \mathbb{E}_{\sigma_1, \dots, \sigma_T} \left[\sum_{t \in [T]} \tilde{f}_t(\theta_t) - \inf_{\theta \in \Theta} \sum_{t \in [T]} \tilde{f}_t(\theta) \right].$$

1643 One recognises (up to a factor T) the left-hand side of Proposition 4 on the loss sequence $(\tilde{f}_t)_{t \in [T]}$.
 1644 Furthermore, we can use this proposition as, given our choice of \mathcal{R} , for any $t \in [T]$, \tilde{f}_t is
 1645 L -Lipschitz with $L = 5D_{\mathcal{Z}}G(4\alpha)^{-1}$ almost surely by Lemma 1 (with $\chi = 0$). We then have:
 1646

$$1647 \quad \mathbb{E}_{\sigma_1, \dots, \sigma_T} \left[\frac{1}{T} \sum_{t \in [T]} \tilde{f}_t(\theta_t) - \inf_{\theta \in \Theta} \frac{1}{T} \sum_{t \in [T]} \tilde{f}_t(\theta) \right] \leq \mathcal{O} \left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi \right). \quad (54)$$

1650 Dividing Equation (53) by T , and plugging Equation (54) gives: for all $\theta \in \Theta$:

$$1653 \quad T^{-1} \mathbb{E}[R_T^s(\theta)] \leq \mathcal{O} \left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi \right) \\ 1654 \quad + 2D_{\mathcal{Z}} \alpha \left(1 + 2 \ln(d) C_0 \left(\ln \left(\frac{2}{\alpha} \right) + (1 - \beta) \ln^2(\alpha \ln(d)) \right) \right), \\ 1655 \quad = \tilde{\mathcal{O}} \left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi + \alpha \ln(d) \right).$$

1660 Finally remark that, by the definition of the conditional expectation (thus of F_t), and because θ_t is
 1661 \mathcal{F}_{t-1} -measurable, we have for any θ :

$$1663 \quad \mathbb{E}[R_T^s(\theta)] = \mathbb{E} \left[\sum_{t=1}^T f_t(\theta_t) - f_t(\theta) \right] \\ 1664 \quad = \mathbb{E} \left[\sum_{t=1}^T f_t(\theta_t) \right] - \mathbb{E}[f_t(\theta)]$$

1669 Then taking the infimum over θ yields:

$$1670 \quad T^{-1} \mathbb{E}[\mathfrak{R}_T^s] \leq \tilde{\mathcal{O}} \left(\eta m^2 D \frac{1}{\alpha^2} + \frac{mD}{\eta T} + \xi + \alpha \ln(d) \right).$$

1673 This concludes the proof. The second equation consists in simply plugging the proposed value of
 η, α in this bound. \square

1674 G.5 PROOF OF PROPOSITION 3
1675

1676 **Proposition 3.** Assume **H1**, **H2**, and having access to an ξ -approximate optimization oracle \mathbf{O}_ξ
1677 adapted to $\{\tilde{f}_i\}_{t \in [T]}$. Fix $\mathcal{W} = \mathcal{W}_0$, $\mathcal{R} = \mathcal{R}_0$. Let $\{\theta_t\}_{t \in [T]}$ be the output of DF-OGD (Algorithm 2)
1678 instantiated with the non-increasing sequence $(\eta_t)_{t \in [T]}$ and regularization coefficients $(\alpha_t)_{t \in [T]}$.
1679 Then:

$$1680 \begin{aligned} T^{-1}\mathbb{E}[\mathfrak{R}_T^d] &\leq \frac{D_\Theta^2 + D_\Theta\mathbb{E}[P_T]}{2T\eta_T} + \sum_{t \in [T]} \frac{25D_ZG\eta_t}{32T\alpha_t^2} \\ 1684 &\quad + 2D_Z \sum_{t \in [T]} \alpha_t [1 + 2\ln(d)C_0\{\ln(2\alpha_t^{-1}) + (1 - \beta)\ln^2(\alpha_t \ln d)\}] + \xi, \end{aligned}$$

1686 where $P_T = \sum_{t=1}^{T-1} \|\vartheta_{t+1} - \vartheta_t\|$.

1688 *Proof.* In this proof, for the sake of conciseness, we denote by \mathbb{E}_t the conditional expectation
1689 $\mathbb{E}[\cdot | \mathcal{H}_{t-1}]$.

1690 Observe that the dynamic regret can be decomposed as follows:

$$1692 \mathfrak{R}_T^d = \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta_t) - \tilde{w}_t(\theta_t) \rangle] + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) - \tilde{w}_t(\vartheta_t) \rangle] \quad (55)$$

$$1695 \quad + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\vartheta_t) \rangle] - \inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] \quad (56)$$

$$1697 \quad + \sum_{t \in [T]} \left[\inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] - \inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) \rangle] \right] \quad (57)$$

1700 First, we remark that for any t , given $\vartheta_t = \mathbf{O}(\tilde{f}_t)$, we control (56) by Jensen:

$$1702 \begin{aligned} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\vartheta_t) \rangle] - \inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta) \rangle] &\leq \mathbb{E}_t \left[\tilde{f}_t(\vartheta_t) - \inf_{\theta \in \Theta} \tilde{f}_t(\theta) \right] \\ 1704 &\leq \xi. \end{aligned}$$

1705 Then, taking the sup over Θ for each summand of the first sum of (55) (which is valid as θ_t is
1706 \mathcal{F}_{t-1} -measurable), defining by $\text{Reg}_T := \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \tilde{w}_t(\theta_t) - \tilde{w}_t(\vartheta_t) \rangle]$, and noticing for
1707 (57) that $\inf_{\theta \in \Theta} \mathbb{E}_t[f_t(\theta)] - \inf_{\theta \in \Theta} \mathbb{E}_t[\tilde{f}_t(\theta)] \leq \sup_{\theta \in \Theta} \mathbb{E}_t[f_t(\theta) - \tilde{f}_t(\theta)]$ leads to:

$$1709 \mathfrak{R}_T^d \leq \text{Reg}_T + 2 \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) - \tilde{w}_t(\theta) \rangle] + \xi T$$

1711 Using the fact that $\|\cdot\|_\infty$ is dual to $\|\cdot\|_1$:

$$1713 \mathfrak{R}_T^d \leq \text{Reg}_T + 2 \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t[\|\bar{g}_t(X_t)\|_\infty \|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1] + \xi T$$

1715 Since for any $t \in [T]$, $\|\bar{g}_t(X_t)\|_\infty \leq \|\bar{g}_t(X_t)\|_2 \leq D_Z$ almost surely by **H1**-(ii),

$$1717 \mathfrak{R}_T^d \leq \text{Reg}_T + 2D_Z \sum_{t \in [T]} \sup_{\theta \in \Theta} \mathbb{E}_t[\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1] + \xi T \quad (58)$$

1719 First, notice that, by Lemma 2, for all t :

$$1721 \sup_{\theta \in \Theta} \mathbb{E}_t[\|w_t^*(\theta) - \tilde{w}_t(\theta)\|_1] \leq \alpha_t \left(1 + 2\ln(d)C_0 \left(\ln \left(\frac{2}{\alpha_t} \right) + (1 - \beta)\ln^2(\alpha_t \ln d) \right) \right). \quad (59)$$

1723 Second, taking the expectation (denoted by \mathbb{E}) over $(u_t)_{t \in [T]}$ and X_1, \dots, X_T on both sides gives:

$$1724 \begin{aligned} \mathbb{E}[\mathfrak{R}_T^d] &\leq \mathbb{E}[\text{Reg}_T] \\ 1725 &\quad + 2D_Z \sum_{t \in [T]} \alpha_t \left(1 + 2\ln(d)C_0 \left(\ln \left(\frac{2}{\alpha_t} \right) + (1 - \beta)\ln^2(\alpha_t \ln d) \right) \right) + \xi T. \end{aligned} \quad (60)$$

1728 To bound the first term in the right-hand side of Equation (60). Remark that, by the definition of
 1729 conditional expectation:

$$1730 \quad \mathbb{E}[\text{Reg}_T] = \mathbb{E}_{X_1 \dots X_T} \mathbb{E}_{u_1, \dots, u_T} \left[\sum_{t \in [T]} \tilde{f}_t(\theta_t) - \tilde{f}_t(\vartheta_t) \right].$$

1731 One recognises the definition of \mathfrak{R}_T^d of Lemma 8. Thus, by this lemma, we know that for any
 1732 $t \in [T]$, because \tilde{f}_t is K_t -Lipschitz with $K_t = 5D_{\mathcal{Z}}G(4\alpha_t)^{-1}$ almost surely by Lemma 1, we
 1733 have:

$$1734 \quad \mathbb{E}_{u_1, \dots, u_T} \left[\sum_{t \in [T]} \tilde{f}_t(\theta_t) - \tilde{f}_t(\vartheta_t) \right] \leq \frac{1}{2\eta_T} (D_{\Theta}^2 + D_{\Theta} P_T) + \sum_{t \in [T]} \frac{25D_{\mathcal{Z}}^2 G^2 \eta_t}{32\alpha_t^2}. \quad (61)$$

1735 Plugging Equation (61) into Equation (60) and then dividing by $T > 0$ on both sides gives the
 1736 desired result. \square

1737 G.6 PROOFS OF APPENDIX F

1738 *Proof of Proposition 5.* The regret simply decomposes as follows:

$$1739 \quad \mathfrak{R}_T^d = \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \underline{w}_t(\theta_t) \rangle] - \sum_{t \in [T]} \inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) \rangle]$$

$$1740 \quad = \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), \underline{w}_t(\theta_t) - w_t^*(\theta_t) \rangle]$$

$$1741 \quad + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta_t) \rangle] - \sum_{t \in [T]} \inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) \rangle]$$

$$1742 \quad \leq \kappa + \sum_{t \in [T]} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta_t) \rangle] - \sum_{t \in [T]} \inf_{\theta \in \Theta} \mathbb{E}_t[\langle \bar{g}_t(X_t), w_t^*(\theta) \rangle],$$

1743 and the proof continues as in Theorem 2. \square

1744
 1745
 1746
 1747
 1748
 1749
 1750
 1751
 1752
 1753
 1754
 1755
 1756
 1757
 1758
 1759
 1760
 1761
 1762
 1763
 1764
 1765
 1766
 1767
 1768
 1769
 1770
 1771
 1772
 1773
 1774
 1775
 1776
 1777
 1778
 1779
 1780
 1781