OPT2022: 14th Annual Workshop on Optimization for Machine Learning

Clairvoyant Regret Minimization: Equivalence with Nemirovski's Conceptual Prox Method and Extension to General Convex Games

Gabriele Farina GFARINA@META.COM
FAIR, Meta AI

Christian Kroer CHRISTIAN.KROER@COLUMBIA.EDU
IEOR Department, Columbia University

Chung-Wei Lee LEECHUNG@USC.EDU
Haipeng Luo HAIPENGL@USC.EDU
Computer Science Department, University of Southern California

Abstract

A recent paper by Piliouras et al. [23, 24] introduces an uncoupled learning algorithm for normal-form games—called Clairvoyant MWU (CMWU). In this paper we show that CMWU is equivalent to the *conceptual prox method* described by Nemirovski [22]. This connection immediately shows that it is possible to extend the CMWU algorithm to any convex game, a question left open by Piliouras et al. We call the resulting algorithm—again equivalent to the conceptual prox method—*Clairvoyant OMD*. At the same time, we show that our analysis yields an improved regret bound compared to the original bound by Piliouras et al., in that the regret of CMWU scales only with the square root of the number of players, rather than the number of players themselves.

1. Introduction

A recent line of work has focused on identifying learning algorithms such that, when used by all players in a game, each player's regret grows polylogarithmically in the number of repetitions T, improving over the traditional (and unimprovable) $O(\sqrt{T})$ bounds[1] of no-regret algorithms for the more adversarial setting in which no assumption about the algorithm used by other agents is made (see, *e.g.*, [1, 7–10, 13, 14, 16, 25, 27]). By leveraging well-known connections between regret and equilibria in games (*e.g.*, [15, 17, 18, 26]), such learning algorithms can then be used as computational approaches to equilibrium finding, leading to $\tilde{O}(1/T)$ convergence to coarse correlated equilibria. This reduction from no-regret learning to equilibrium computation is largely the preferred approach in practice [2, 4, 5, 12].

In a recent work, Piliouras et al. [23] depart from this no-regret learning perspective, by introducing a new algorithm that they call *Clairvoyant MWU* (CMWU) for normal-form games. CMWU is a variant of the popular multiplicative weights updates (MWU) algorithm, where action probabilities are scaled exponentially according to their payoff at each time step. In CMWU this scaling is assumed to be done at time t with respect to a tight approximation of the payoff that the player will see *at that same time* t (hence the adjective *clairvoyant*). In CMWU, the iterates generated by the algorithm are not known to yield regret bounds, and as a decentralized protocol the CMWU dynamics require the players to coordinate on repeatedly computing a fixed point over a sequence of iterations. However, the CMWU dynamics, whether centralized or decentralized, are attractive as a method for computing a coarse correlated equilibrium (CCE).

1. For simplicity, in the introduction our $O(\cdot)$ notation hides all parameters independent of T.

Algorithm	Operations required to compute an ϵ-CCE in NFGs	Are iterates known to be no-regret?	Generalizes to convex games?	Leads to CE in NFGs?
Optimistic MWU [10]	$O\left(n\, d \log d \cdot \frac{1}{\epsilon} \log^4 \frac{1}{\epsilon}\right)$	✓	✗	✗
BM-OFTRL-LogBar [3]	$O\left(n\, \mathrm{poly}(d) \cdot \frac{1}{\epsilon} \log\left(\frac{1}{\epsilon}\right) \log\log \frac{1}{\epsilon}\right)$	✓	✗	✓
LRL-OFTRL [13]	$O\left(n\, \mathrm{poly}(d) \cdot \frac{1}{\epsilon} \log\left(\frac{1}{\epsilon}\right) \log\log \frac{1}{\epsilon}\right)$	✓	✓	✗
Clairvoyant MWU [24]	$O\left(n\, d \log(d) \cdot \frac{1}{\epsilon} \log \frac{1}{\epsilon}\right)$	✗ (but a subset is)	✗	✗
Clairvoyant OMD [this paper]	$O\left(\sqrt{n}\, d \log(d) \cdot \frac{1}{\epsilon} \log \frac{1}{\epsilon}\right)$	✗ (but a subset is)	✓	✗

Table 1: Comparison of existing learning-based methods for computing an ϵ-approximate coarse-correlated equilibrium (CCE) in a generic normal-form game (NFG) with n players, d total actions, and utilities in $[0, 1]$, for values of $\epsilon = O(1/d)$. We use the notation $\mathrm{poly}(d)$ to hide polynomial terms in d of order at least 2. We remark that BM-OFTRL-LogBar is an algorithm designed to minimize the more challenging notion of *swap regret*, and guarantees convergence to the set of correlated equilibrium (CE) in normal-form games.

In particular, they yield a very competitive $O(\log T/T)$ rate of convergence, while requiring only a single gradient computation as well as a linear-time closed-form strategy update at every iteration.

In this paper, we show that the CMWU algorithm can be viewed as an instantiation of the *conceptual prox method* (CPM), which has been studied extensively in the first-order methods literature [6, 20, 22]. Nemirovski [22] discusses the CPM as a conceptual algorithm that achieves an $O(1/T)$ rate of convergence to a solution to a monotone variational inequality (VI). Monotone VI generalizes for example the problem of computing a two-player zero-sum Nash equilibrium. He labels it a conceptual method because each step of the algorithm requires solving a non-trivial variational inequality, which means that it is not clear that the algorithm is implementable. He then shows, however, that an approximate solution to this VI can be computed in logarithmic time in the required precision, because the VI corresponds to a solution to a fixed point of a mapping which is a contraction. Finally, he goes on to show that in fact one can get the same rate by performing only two steps of the contraction. This results in the famous *mirror prox* algorithm. What we show is that the CMWU algorithm can be viewed as a specialization of CPM to the setting where the feasible set is the Cartesian product of the player's strategy spaces, and the operator used in the variational inequality is the gradient operator for each player. However, here one must depart from the CPM perspective of Nemirovski, because his paper focuses on *monotone* operators. This is because his method ultimately requires computing the average of the CPM iterates, in order to achieve convergence. For that reason, CMWU, and our generalization, is not strictly covered by Nemirovski's results. The key insight is to realize that if we are only interested in the regret of the players, then no averaging is needed, in which case one can show that the CPM method achieves constant regret, even in the case of a non-monotone operator, as is the case for the operator associated to general-sum games.

Using our perspective on CPM as a method for computing a sequence of strategies with low regret, we show that it is possible to generalize the CMWU method, which only applies to normal-form games, to games with arbitrary convex and compact decision sets, and utility functions that are concave with bounded gradients that are also Lipschitz continuous. This answers an open problem of Piliouras et al. [23], where they ask whether CMWU could be generalized to exactly such a setting. We call the resulting algorithm *Clairvoyant Online Mirror Descent* (COMD). We stress that while this particular result for the algorithm is new, the algorithm is really an instantiation of the CPM of Nemirovski [22]. By appealing to the contraction argument of [22], we also show that it is possible to improve the regret bound of CMWU in the case of normal-form games: we improve the dependence on the number of players n from being linear to only $\sqrt{n}$. Finally, we go on to develop concrete bounds for the COMD method in the case where only a finite number of steps of the contraction are performed. We show that in this case one can perform $O(\log t)$ steps of the contraction at iteration t, while retaining the guarantee of constant regret. An immediate consequence of our result is that by instantiating this approximate variant of COMD with a dilated entropy regularizer [12, 19, 21], we get the first algorithm that has an $O(\log T/T)$ rate of convergence to a normal-form coarse correlated equilibrium for extensive-form games, while requiring a linear-time update at every iteration.

2. Setting

2.1. Convex Games and Variational Inequality

We let $[\![n]\!] \coloneqq \{1, 2, \ldots, n\}$ be a set of players, with $n \in \mathbb{N} \coloneqq \{1, 2, \ldots\}$. In this paper, we operate on *convex games*, whereby each player $i \in [\![n]\!]$ has a nonempty convex and compact set of strategies $\mathcal{X}_i \subseteq \mathbb{R}^{d_i}$. For a *joint strategy profile* $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \bigtimes_{j=1}^{n} \mathcal{X}_j$, the reward of player i is given by a differentiable concave utility function $u_i : \bigtimes_{j=1}^{n} \mathcal{X}_j \to \mathbb{R}$, subject to the following standard assumptions:

1. (Concavity) $u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$ is *concave* in $\boldsymbol{x}_i$ for any $\boldsymbol{x}_{-i} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_n) \in \bigtimes_{j \neq i} \mathcal{X}_j$;
2. (Bounded gradients) for any $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \bigtimes_{j=1}^{n} \mathcal{X}_j$, $\nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is bounded,
3. (L-smoothness) The gradient $\nabla_{\boldsymbol{x}_i} u_i$ is Lipschitz-continuous.

In the rest of the paper, we will often find it beneficial to view strategy updates in the game as *global*, that is, operating on all players at the same time rather than each player individually. For that reason, we now introduce notation to operate on the Cartesian product of all strategy spaces. First, we denote the sum of dimensions of the strategy spaces of the players with the letter $d \coloneqq d_1 + \cdots + d_n$. The joint strategy space of the game is $\mathcal{Z} \coloneqq \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$, and we will consistently denote elements in $\mathcal{Z}$ using the letter $\boldsymbol{z}$ or variants thereof. Given a vector $\boldsymbol{z} \in \mathbb{R}^d$, we will denote as $\boldsymbol{z}_i \in \mathbb{R}^{d_i}$ the portion of the vector belonging to player i, that is, we let $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$ be the (unique) vectors such that $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$.

In this global notation over the game, a key quantity associated with the game is the vector-valued function mapping strategies to payoff gradients for all players, that is,

$$F : \mathcal{Z} \to \mathbb{R}^d, \qquad F(\boldsymbol{z}) \coloneqq \begin{pmatrix} -\nabla_{\boldsymbol{x}_1} u_1(\boldsymbol{z}) \\ \vdots \\ -\nabla_{\boldsymbol{x}_n} u_n(\boldsymbol{z}) \end{pmatrix}.$$

When viewed through the global lenses of the function F, the properties listed above imply the following:

Assumption 1 *For an appropriate primal-dual norm pair $(\|\cdot\|, \|\cdot\|_*)$ over $\mathbb{R}^d$, the game operator $F : \mathcal{Z} \to \mathbb{R}^d$ satisfies:*

- *(Boundedness) $\|F(\boldsymbol{z})\|_* \leq B$ for all $\boldsymbol{z} \in \mathcal{Z}$.*
- *(Lipschitz continuity) There exists $L > 0$ such that $\|F(\boldsymbol{z}) - F(\boldsymbol{z}')\|_* \leq L\|\boldsymbol{z} - \boldsymbol{z}'\|$.*

The variational inequality associated with the operator F, that is, the problem of finding $\boldsymbol{z} \in \mathcal{Z}$ such that

$$\langle F(\boldsymbol{z}), \boldsymbol{z}' - \boldsymbol{z} \rangle \geq 0 \qquad \forall \boldsymbol{z}' \in \mathcal{Z} \tag{1}$$

is exactly equivalent to the problem of computing a Nash equilibrium of the game (see, *e.g.*, [11, Proposition 1.4.2]). We remark that generally F is *not* a monotone operator, that is, there might exist $\boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}$ such that $\langle F(\boldsymbol{z}) - F(\boldsymbol{z}'), \boldsymbol{z} - \boldsymbol{z}' \rangle \not\geq 0$. Nonetheless, in this paper we will be concerned with applying the conceptual prox method, which was designed for monotone operators, to (1).

2.2. Proximal Setup

For each player $i \in [\![n]\!]$, we assume that a strongly convex regularizer $\varphi_i : \mathcal{Z} \to \mathbb{R}$ has been chosen. Each regularizer φ_i induces a generalized notion of distance—called *Bregman divergence*—over $\mathcal{X}_i$, defined as

$$D_i(\cdot \,\|\, \cdot) : \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}_{\geq 0}, \qquad D_i(\boldsymbol{x} \,\|\, \boldsymbol{x}') \coloneqq \varphi_i(\boldsymbol{x}) - \varphi_i(\boldsymbol{x}') - \langle \nabla \varphi_i(\boldsymbol{x}'), \boldsymbol{x} - \boldsymbol{x}' \rangle.$$

We combine the regularizer φ_i for each player's strategy space into a *global*, composite regularizer

$$\varphi : \mathcal{Z} \to \mathbb{R}, \qquad \varphi : \boldsymbol{z} \mapsto \varphi_i(\boldsymbol{z}_i) + \cdots + \varphi_n(\boldsymbol{z}_n).$$

Correspondingly, the Bregman divergence induced by φ is the function $D(\cdot \,\|\, \cdot) : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}_{\geq 0}$,

$$D(\boldsymbol{z} \,\|\, \boldsymbol{z}') \coloneqq \varphi(\boldsymbol{x}) - \varphi(\boldsymbol{z}') - \langle \nabla \varphi(\boldsymbol{z}'), \boldsymbol{z} - \boldsymbol{z}' \rangle = D_1(\boldsymbol{z}_1 \,\|\, \boldsymbol{z}_1') + \cdots + D_n(\boldsymbol{z}_n \,\|\, \boldsymbol{z}_n').$$

As long as each φ_i is strongly convex, then so is φ. Specifically, in the rest of the paper we operate under the following assumption.

Assumption 2 *The per-player regularizers φ_i are chosen so that the global regularizer $\varphi : \mathcal{Z} \to \mathbb{R}_{\geq 0}$ is 1-strongly convex with respect to the norm $\|\cdot\| : \mathcal{Z} \to \mathbb{R}$ for which Assumption 1 holds.*

With that, we are able to define the *prox operator*, which we define for the global space $\mathcal{Z}$ and global regularizer φ. Given a *center* $\boldsymbol{z} \in \mathcal{Z}$ and a gradient $\boldsymbol{g} \in \mathbb{R}^d$, the prox operator $\Pi_{\boldsymbol{z}}(\boldsymbol{g})$ generalizes the notion of gradient step away from $\boldsymbol{z}$ in the direction of $-\boldsymbol{g}$, and is defined as follows.

Definition 1 (Prox operator) *The prox operator associated with φ is defined as*

$$\Pi_{\boldsymbol{z}}(\boldsymbol{g}) = \operatorname*{arg\,min}_{\hat{\boldsymbol{z}} \in \mathcal{Z}} \left\{ \boldsymbol{g}^\top \hat{\boldsymbol{z}} + D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}) \right\} = \operatorname*{arg\,min}_{\hat{\boldsymbol{z}} \in \mathcal{Z}} \{ \langle \boldsymbol{g} - \nabla \varphi(\boldsymbol{z}), \hat{\boldsymbol{z}} \rangle + \varphi(\hat{\boldsymbol{z}}) \} \tag{2}$$

for any center $\boldsymbol{z} \in \mathcal{Z}$ and a vector $\boldsymbol{g} \in \mathbb{R}^d$.

We recall standard properties of the prox operator in Appendix B.

3. Conceptual Prox Method

The key observation underpinning the *conceptual prox method* is the following straightforward lemma (all proofs are in the Appendix).

Lemma 2 *Let $t \in \mathbb{N}$ and $\boldsymbol{z}^{(t-1)} \in \mathcal{Z}$ be arbitrary. If the point $\boldsymbol{z}^{(t)}$ satisfies the fixed point equation $\boldsymbol{z}^{(t)} = \Pi_{\boldsymbol{z}^{(t-1)}}\big(\eta F(\boldsymbol{z}^{(t)})\big)$, then*

$$\eta \, \langle F(\boldsymbol{z}^{(t)}), \hat{\boldsymbol{z}} - \boldsymbol{z}^{(t)} \rangle \geq D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{(t)}) - D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{(t-1)}) + D(\boldsymbol{z}^{(t)} \,\|\, \boldsymbol{z}^{(t-1)}) \qquad \forall \hat{\boldsymbol{z}} \in \mathcal{Z}, \tag{3}$$

and in particular, for all players i,

$$\eta \, \langle \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{z}^{(t)}), \hat{\boldsymbol{z}}_i - \boldsymbol{z}_i^{(t)} \rangle \leq -D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(t)}) + D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(t-1)}) - D_i(\boldsymbol{z}_i^{(t)} \,\|\, \boldsymbol{z}_i^{(t-1)}) \qquad \forall \hat{\boldsymbol{z}}_i \in \mathcal{X}_i. \tag{4}$$

By noting that the right-hand side of (4) is telescopic, Theorem 2 immediately implies the following.

Corollary 3 (Constant per-player regret) *Let $\boldsymbol{z}^{(0)} \in \mathcal{Z}$ be arbitrary, and suppose that recursively $\boldsymbol{z}^{(t)} \in \mathcal{Z}$ at all times $t \in \mathbb{N}$ satisfies the following fixed-point equation*

$$\boldsymbol{z}^{(t)} = \Pi_{\boldsymbol{z}^{(t-1)}}\Big(\eta F(\boldsymbol{z}^{(t)})\Big). \tag{$\star$}$$

Then, at all times $T \in \mathbb{N}$, the per-player regret is upper bounded as

$$\mathrm{Reg}_i^T := \max_{\hat{\boldsymbol{z}}_i \in \mathcal{X}_i} \sum_{t=1}^{T} \langle \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{z}^{(t)}), \hat{\boldsymbol{z}}_i - \boldsymbol{z}_i^{(t)} \rangle \le \frac{1}{\eta} \max_{\hat{\boldsymbol{z}}_i \in \mathcal{X}_i} \Big\{ D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(0)}) \Big\}.$$

The algorithm defined in Theorem 3 is called the *conceptual prox method (CPM)* (see also Nemirovski [22]). Theorem 3 shows that the per-player regret cumulated up to any time T by the fixed-point iterates $\boldsymbol{z}^{(t)} = \Pi_{\boldsymbol{z}^{(t-1)}}\big(F(\boldsymbol{z}^{(t)})\big)$ produced by the CPM is bounded by the range of the divergence $D_i(\cdot \,\|\, \boldsymbol{z}_i^{(0)})$, a quantity independent of time. For example, when $\mathcal{X}_i = \Delta^m$ is the m-simplex, $\boldsymbol{z}_i^{(0)}$ is the uniform strategy, and φ_i is negative entropy, then $\max_{\boldsymbol{x} \in \mathcal{X}_i} D_i(\boldsymbol{x} \,\|\, \boldsymbol{z}_i^{(0)}) = \log m$.

Existence and Computation of Fixed-Point Solutions At this stage, it is perhaps unclear why the fixed points ($\star$) exist and how one can compute them. The key lies in the following observation, which dates back to at least the work of Nemirovski [22]:

Observation 1 *At all times t the map $\boldsymbol{w} \mapsto \Pi_{\boldsymbol{z}^{(t-1)}}(\eta F(\boldsymbol{w}))$ is ηL-Lipschitz continuous. Hence, as long as $\eta < 1/L$, the above function is a contraction, and the fixed point is therefore unique. Consequently, convergence to an ϵ-fixed point can be achieved via a number of iterations that scales proportionally to $\log(1/\epsilon)$.*

This is straightforward: the proximal operator itself is 1-Lipschitz continuous (Theorem 7), and ηF is ηL-Lipschitz continuous, so their composition is ηL-Lipschitz continuous. So, at least *approximate* fixed-point solutions are easy to compute. In the next section we quantify the error introduced by the approximation in the fixed-point solution.

4. Conceptual Prox Method with Approximate Fixed Points

A more refined analysis of the argument employed in Theorem 3 takes into account error in the computation of the fixed-point solutions ($\star$). We start by relaxing Theorem 2.

Lemma 4 *Let $t \in \mathbb{N}$ and $\boldsymbol{z}^{(t-1)} \in \mathcal{Z}$ be arbitrary. Let $\boldsymbol{w}^{(t)} \in \mathcal{Z}$ be an approximate fixed point, in the sense that for some $\epsilon^{(t)} \ge 0$,*

$$\Big\| \boldsymbol{w}^{(t)} - \Pi_{\boldsymbol{z}^{(t-1)}}\Big(\eta F(\boldsymbol{w}^{(t)})\Big) \Big\| \le \epsilon^{(t)}. \tag{5}$$

Then, the point $\boldsymbol{z}^{(t)} := \Pi_{\boldsymbol{z}^{(t-1)}}\big(\eta F(\boldsymbol{w}^{(t)})\big)$ satisfies

$$\eta \,\langle F(\boldsymbol{w}^{(t)}), \hat{\boldsymbol{z}} - \boldsymbol{w}^{(t)} \rangle \ge D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{(t)}) - D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{(t-1)}) + D(\boldsymbol{z}^{(t)} \,\|\, \boldsymbol{z}^{(t-1)}) - \eta B \epsilon^{(t)} \qquad \forall \hat{\boldsymbol{z}} \in \mathcal{Z}, \tag{6}$$

and in particular, for all players i,

$$\eta \,\langle \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{w}^{(t)}), \hat{\boldsymbol{z}}_i - \boldsymbol{w}_i^{(t)} \rangle \le -D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(t)}) + D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(t-1)}) - D_i(\boldsymbol{z}_i^{(t)} \,\|\, \boldsymbol{z}_i^{(t-1)}) + \eta B \epsilon^{(t)} \qquad \forall \hat{\boldsymbol{z}}_i \in \mathcal{X}_i.$$

Repeating the analysis we already carried out for Theorem 3, this time using Theorem 4, we obtain the following.

Corollary 5 *Let $\boldsymbol{z}^{(0)} \in \mathcal{Z}$ be arbitrary, and recursively let $\boldsymbol{z}^{(t)}, \boldsymbol{w}^{(t)} \in \mathcal{Z}$ at all times $t \in \mathbb{N}$ be such that*

$$\left\| \boldsymbol{w}^{(t)} - \Pi_{\boldsymbol{z}^{(t-1)}}\Big(\eta F(\boldsymbol{w}^{(t)})\Big) \right\| \le \epsilon^{(t)}, \qquad \boldsymbol{z}^{(t)} := \Pi_{\boldsymbol{z}^{(t-1)}}\Big(\eta F(\boldsymbol{w}^{(t)})\Big). \tag{$\star\star$}$$

Then, at all times $T \in \mathbb{N}$, the per-player regret associated with iterates $\boldsymbol{w}^{(t)}$ is upper bounded as

$$\mathrm{Reg}_i^T := \max_{\hat{\boldsymbol{z}}_i \in \mathcal{X}_i} \sum_{t=1}^{T} \langle \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{w}^{(t)}), \hat{\boldsymbol{z}}_i - \boldsymbol{w}_i^{(t)} \rangle \le \frac{1}{\eta} \max_{\hat{\boldsymbol{z}}_i \in \mathcal{X}_i} \Big\{ D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(0)}) \Big\} + B \sum_{t=1}^{T} \epsilon^{(t)}.$$

Observation 2 *When the choice $\epsilon^{(t)} = 1/t^2$ is used, then the sum of errors $\sum_{t=1}^{T} \epsilon^{(t)}$ is an additive constant bounded by 2, and therefore does not affect the constant per-player regret guarantees, while at the same time requiring $O(\log t)$ fixed-point iterations per iteration of the learning algorithm.*

4.1. Centralized Implementation

By combining Theorem 5 together with the concrete choice of errors $\epsilon^{(t)}$ given in Observation 2 and fixed-point iterations on the map $\Pi_{\boldsymbol{z}^{(t-1)}} \circ \eta F$, which is a contraction for $\eta \le 1/(2L)$ (Observation 1), we obtain Algorithm 1, whose properties summarized in Theorem 6 follow directly from the preceding discussion.

Algorithm 1: Conceptual prox method with approximate fixed points (*centralized* implementation)

Data: $\boldsymbol{z}^{(0)} \in \mathcal{Z}$ initial point, $0 < \eta \le 1/(2L)$ learning rate, $\epsilon^{(t)}$ desired fixed-point approximation error

```
for t = 1, 2, ... do
    w^(t) ← z^(t-1)
    while ‖ w^(t) − Π_{z^(t-1)}( ηF(w^(t)) ) ‖ > ε^(t) do
        w^(t) ← Π_{z^(t-1)}( ηF(w^(t)) )          [▷ Fixed-point iteration]
    z^(t) ← Π_{z^(t-1)}( ηF(w^(t)) )
```

Theorem 6 *At all times $t = 1, 2, \dots$ in Algorithm 1,*

1. *The internal* **while** *loop runs for at most $\log_2(\max_{\boldsymbol{z},\boldsymbol{z}' \in \mathcal{Z}} \|\boldsymbol{z} - \boldsymbol{z}'\|) + \log_2 \frac{1}{\epsilon^{(t)}}$ iterations.*
2. *The iterates $\boldsymbol{z}^{(t)}$ produced by the algorithm achieve regret upper bounded by*

$$\mathrm{Reg}_i^t := \max_{\hat{\boldsymbol{z}}_i \in \mathcal{X}_i} \sum_{\tau=1}^{t} \langle \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{w}^{(\tau)}), \hat{\boldsymbol{z}}_i - \boldsymbol{w}_i^{(\tau)} \rangle \le \frac{1}{\eta} \max_{\hat{\boldsymbol{z}}_i \in \mathcal{X}_i} \Big\{ D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}^{(0)}) \Big\} + B \sum_{\tau=1}^{t} \epsilon^{(\tau)}$$

for each player i. Correspondingly, the average product distribution of play $\frac{1}{t} \sum_{\tau=1}^{t} \boldsymbol{z}_1^{(\tau)} \otimes \dots \otimes \boldsymbol{z}_n^{(\tau)}$ is a κ-coarse correlated equilibrium of the game, with

$$\kappa = \frac{\mathrm{Reg}_i^t}{t} \le \frac{1}{\eta\, t} \max_{\hat{\boldsymbol{z}}_i \in \mathcal{X}_i} \Big\{ D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}^{(0)}) \Big\} + \frac{B}{t} \sum_{\tau=1}^{t} \epsilon^{(\tau)}$$

Once again, we remark that the choice $\epsilon^{(t)} = 1/t^2$ for all t is natural, and leads to constant per-player regret, as well as convergence to a coarse correlated equilibrium of the convex game at the rate $1/t$. Alternatively, if the total number of iterations T was known in advance, the choice $\epsilon^{(t)} = 1/T$ would lead to a similar result.

We refer to Algorithm 1 as a *centralized* implementation because it operates directly on the product space $\mathcal{Z}$. We believe that this is the natural setting in which to analyze the algorithm. However, as we lay out in the next section, the individual steps that make Algorithm 1 can be expanded to have a learning dynamic flavor for each player, with some important caveats.

4.2. Decentralized Implementation: Clairvoyant OMD

In this section we show that Algorithm 1 can be implemented in the form of a decentralized learning algorithm where each player independently updates their strategy upon observing the gradient of their utility. The key is in the observation that, given the definition of $\varphi = \varphi_1 + \cdots + \varphi_n$, the proximal operator on Line 4 decomposes as

$$\Pi_{\boldsymbol{z}^{(t-1)}}\left(\eta F(\boldsymbol{w}^{(t)})\right) = \begin{pmatrix} \underset{\boldsymbol{x}_1 \in \mathcal{X}_1}{\arg\min}\left\{-\eta\langle \nabla_{\boldsymbol{x}_1} u_1(\boldsymbol{w}^{(t)}), \boldsymbol{x}_1\rangle + D_1(\boldsymbol{x}_1 \,\|\, \boldsymbol{z}_1^{(t-1)})\right\} \\ \vdots \\ \underset{\boldsymbol{x}_n \in \mathcal{X}_n}{\arg\min}\left\{-\eta\langle \nabla_{\boldsymbol{x}_n} u_n(\boldsymbol{w}^{(t)}), \boldsymbol{x}_n\rangle + D_n(\boldsymbol{x}_n \,\|\, \boldsymbol{z}_n^{(t-1)})\right\} \end{pmatrix},$$

which corresponds to an OMD update (with linearized losses) for each player. Then, by fixing the number of fixed-point iterations (that is, repetitions of the **while** loop) to the quantity

$$N^{(t)} \coloneqq 1 + \log_2\left(\max_{\boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}} \|\boldsymbol{z} - \boldsymbol{z}'\|\right) + \log_2 \frac{1}{\epsilon^{(t)}},$$

it becomes guaranteed that the approximation error of the approximate fixed-point generated after $N^{(t)} - 1$ iterations is less than $\epsilon^{(t)}$, and thus we can rewrite Algorithm 1 as in Algorithm 2.

As Algorithm 2 is amounts to an alternative implementation of Theorem 6, the guarantees of Theorem 6 apply to Algorithm 2 as well. We also remark that in the latter implementation, each player $i \in [\![n]\!]$ crucially updates their approximate fixed-point strategy $\boldsymbol{w}_i^{(t')}$ using as gradient vector the gradient of their utility evaluated in the strategy profile $(\boldsymbol{w}_1^{(t'-1)}, \ldots, \boldsymbol{w}_n^{(t'-1)})$. This is perfectly consistent with the framework of learning in games, where each player updates their strategy based on their gradient at the current strategy profile.

One caveat with this implementation is that only the iterates $\boldsymbol{z}^{(t)}$ (and not the approximate fixed-point iterates $\boldsymbol{w}^{(t')}$) are guaranteed to cumulate low regret. In other words, in the learning-in-games interpretation, only the subsequence of iterates $\{\boldsymbol{w}^{(N^{(1)})}, \boldsymbol{w}^{(N^{(1)}+N^{(2)})}, \ldots\}$ is guaranteed to have low regret. Therefore, when interpreted as a learning algorithm, Algorithm 2 provides uncoupled learning dynamics, but is not a *no-regret* algorithm in the classic sense. Nonetheless, a predictable subsequence of iterates guarantees constant regret, and therefore the algorithm can be used to extract an approximate coarse correlated equilibrium.

The name *Clairvoyant OMD* was chosen to explicitly point out that the algorithm is a generalization, to general convex games, of the Clairvoyant MWU algorithm recently introduced by Piliouras et al. [23] in the special case of normal-form games and for certain specific choices of regularizers—see also Appendix C. We remark also that while Clairvoyant MWU was originally introduced as a *centralized* algorithm, the latter decentralized interpretation was preferred in the later revision of the paper by the same authors [24].

Algorithm 2: Decentralized, per-player implementation (Clairvoyant OMD)

Data: $\boldsymbol{z}_i^{(0)} \in \mathcal{X}_i$ initial strategy for each player, $0 < \eta \leq 1/(2L)$ learning rate, $\epsilon^{(t)}$ desired fixed-point approximation error

$t' \leftarrow 0$

$\boldsymbol{w}^{(t')} \leftarrow \boldsymbol{z}^{(0)}$

for $t = 1, 2, \dots$ **do**

  [▷ Begin unrolling of **while** loop on Line 4 of Algorithm 1]

  for $k = 1, \dots, N^{(t)}$ **do**

    for *each player* $i \in [\![n]\!]$*, in parallel* **do**

      $t' \leftarrow t' + 1$

      $\boldsymbol{w}_i^{(t')} \leftarrow \operatorname*{arg\,min}_{\boldsymbol{x}_i \in \mathcal{X}_i} \left\{ -\eta \langle \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{w}_1^{(t'-1)}, \dots, \boldsymbol{w}_n^{(t'-1)}), \boldsymbol{x}_i \rangle + D_i(\boldsymbol{x}_i \,\|\, \boldsymbol{z}_i^{(t-1)}) \right\}$

  [▷ End unrolling of **while** loop on Line 4 of Algorithm 1]

  $\boldsymbol{z}^{(t)} \leftarrow \boldsymbol{w}^{(t')}$

In Appendix C we discuss how the previous discussion informs an improved analysis of the Clairvoyant MWU algorithm for normal-form games.

References

[1] Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. Near-optimal no-regret learning for correlated equilibria in multi-player general-sum games. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 736–749. ACM, 2022.

[2] Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Andrea Celli, and Tuomas Sandholm. Faster no-regret learning dynamics for extensive-form correlated and coarse correlated equilibria. In *EC '22: The 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, July 11 - 15, 2022*, pages 915–916. ACM, 2022. doi: 10.1145/3490486.3538288. URL `https://doi.org/10.1145/3490486.3538288`.

[3] Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Tuomas Sandholm. Uncoupled learning dynamics with $O(\log T)$ swap regret in multiplayer games, 2022.

[4] Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218), January 2015.

[5] Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, page eaao1733, Dec. 2017.

[6] Gong Chen and Marc Teboulle. Convergence analysis of a proximal-like minimization algorithm using bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

[7] Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

[8] Constantinos Daskalakis and Noah Golowich. Fast rates for nonparametric online learning: from realizability to learning in games. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 846–859. ACM, 2022.

[9] Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011.

[10] Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021*, pages 27604–27616, 2021.

[11] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

[12] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Better regularization for sequential decision spaces: Fast convergence rates for Nash, correlated, and team equilibria. In *ACM Conference on Economics and Computation*, 2021.

[13] Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Near-optimal no-regret learning for general convex games, 2022. URL `https://arxiv.org/abs/2206.08742`.

[14] Gabriele Farina, Chung-Wei Lee, Haipeng Luo, and Christian Kroer. Kernelized multiplicative weights for 0/1-polyhedral games: Bridging the gap between learning in extensive-form and normal-form games. *The 39th International Conference on Machine Learning (ICML)*, 2022.

[15] Dean Foster and Rakesh Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55, 1997.

[16] Dylan J. Foster, Zhiyuan Li, Thodoris Lykouris, Karthik Sridharan, and Éva Tardos. Learning in games: Robustness of fast convergence. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, pages 4727–4735, 2016.

[17] Yoav Freund and Robert Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

[18] Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

[19] Samid Hoda, Andrew Gilpin, Javier Peña, and Tuomas Sandholm. Smoothing techniques for computing Nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2), 2010.

[20] Krzysztof C Kiwiel. Proximal minimization methods with generalized bregman functions. *SIAM journal on control and optimization*, 35(4):1142–1168, 1997.

[21] Christian Kroer, Kevin Waugh, Fatma Kılınç-Karzan, and Tuomas Sandholm. Faster algorithms for extensive-form game solving via improved smoothing functions. *Mathematical Programming*, 2020.

[22] Arkadi Nemirovski. Prox-method with rate of convergence O(1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1), 2004.

[23] Georgios Piliouras, Ryann Sim, and Stratis Skoulakis. Optimal no-regret learning in general games: Bounded regret with unbounded step-sizes via Clairvoyant MWU, 2021. URL `https://arxiv.org/abs/2111.14737v1`.

[24] Georgios Piliouras, Ryann Sim, and Stratis Skoulakis. Beyond time-average convergence: Near-optimal uncoupled online learning via Clairvoyant Multiplicative Weights Update, 2022. URL `https://arxiv.org/abs/2111.14737v4`.

[25] Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pages 3066–3074, 2013.

[26] Tim Roughgarden. Intrinsic robustness of the price of anarchy. *J. ACM*, 62(5):32:1–32:42, 2015.

[27] Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, pages 2989–2997, 2015.

Appendix A. Example: Normal-Form Games

A notable class of convex games are *normal-form* games. In normal-form games, each player $i \in [\![n]\!]$ has a finite set of actions of cardinality d_i. For any possible combination of actions of the players, each player receives a utility, which we assume to be in the range $[-V, V]$. Players are free to select any distribution over their actions as their strategy, that is, a point in the probability simplex

$$\mathcal{X}_i = \Delta^{d_i} \coloneqq \{\boldsymbol{x} \in \mathbb{R}^{d_i}_{\ge 0} : \mathbf{1}^\top \boldsymbol{x} = 1\}.$$

Each player's utility as a function of the choice of distributions is the expected utility corresponding to actions sampled for those distributions.

We use the following primal-dual norm pair $(\|\cdot\|_\Delta, \|\cdot\|_{*\Delta})$ on the Cartesian product space $\mathbb{R}^d$

$$\begin{aligned} \|\cdot\|_\Delta : \mathbb{R}^d \to \mathbb{R}_{\ge 0}, \qquad & \boldsymbol{z} \mapsto \sqrt{\|\boldsymbol{z}_1\|_1^2 + \cdots + \|\boldsymbol{z}_n\|_1^2}, \\ \|\cdot\|_{*\Delta} : \mathbb{R}^d \to \mathbb{R}_{\ge 0}, \qquad & \boldsymbol{z} \mapsto \sqrt{\|\boldsymbol{z}_1\|_\infty^2 + \cdots + \|\boldsymbol{z}_n\|_\infty^2}. \end{aligned}$$

For this choice of norms, it is immediate to check that the operator F satisfies Assumption 1 for the choice $B \coloneqq \sqrt{n}\,V$ and $L \coloneqq \sqrt{n}\,V$.

A standard choice of regularization for normal-form games is *negative entropy*, that is, the regularizer

$$\varphi_i : \Delta^{d_i} \ni \boldsymbol{x} \mapsto \sum_{j=1}^{d_i} \boldsymbol{x}[j] \log \boldsymbol{x}[j]$$

for all player $i \in [\![n]\!]$. Negative entropy is 1-strongly convex with respect to the ℓ_1 norm, and therefore the composite regularizer $\varphi = \varphi_1 + \cdots + \varphi_n$ is 1-strongly convex with respect to the norm $\|\cdot\|_\Delta$ introduced above, in compliance with Assumption 2. Furthermore, we remark that for any player i, the uniform strategy $\boldsymbol{c}_i \coloneqq (1/d_i, \ldots, 1/d_i)$ satisfies

$$D_i(\boldsymbol{x} \,\|\, \boldsymbol{c}_i) \le \log d_i \qquad \forall \boldsymbol{x} \in \mathcal{X}_i = \Delta^{d_i}. \tag{7}$$

Appendix B. Known Properties of Prox Operators

The following properties of proximal operators are standard in the literature (see, *e.g.*, Nemirovski [22]).

Lemma 7 (The prox operator is Lipschitz continuous) *Given any center $\boldsymbol{z} \in \mathcal{Z}$, the prox operator is Lipschitz continuous with constant 1:*

$$\|\Pi_{\boldsymbol{z}}(\boldsymbol{g}) - \Pi_{\boldsymbol{z}}(\boldsymbol{g}')\| \le \|\boldsymbol{g} - \boldsymbol{g}'\|_* \qquad \forall\, \boldsymbol{g}, \boldsymbol{g}' \in \mathbb{R}^d.$$

We remark that the above inequality uses the primal-dual norm pair for which φ is 1-strongly convex, which is the same primal-dual norm pair for which Assumption 1 holds, as per Assumption 2.

Lemma 8 $\Pi_{\boldsymbol{z}}(\boldsymbol{g}) = \boldsymbol{z}^*$ *if and only if* $\langle \boldsymbol{g} - \nabla\varphi(\boldsymbol{z}) + \nabla\varphi(\boldsymbol{z}^*), \hat{\boldsymbol{z}} - \boldsymbol{z}^*\rangle \ge 0$ *for all* $\hat{\boldsymbol{z}} \in \mathcal{Z}$.

Theorem 8 simply states the (necessary and sufficient) first-order optimality condition for the prox operator problem defined in (2). It immediately implies the following.

Corollary 9 *For all $\boldsymbol{z} \in \mathcal{Z}$ and $\boldsymbol{g}$, one has*

$$D(\hat{\boldsymbol{z}} \,\|\, \Pi_{\boldsymbol{z}}(\boldsymbol{g})) - D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}) + D(\Pi_{\boldsymbol{z}}(\boldsymbol{g}) \,\|\, \boldsymbol{z}) \le \langle \boldsymbol{g}, \hat{\boldsymbol{z}} - \Pi_{\boldsymbol{z}}(\boldsymbol{g})\rangle \qquad \forall \hat{\boldsymbol{z}} \in \mathcal{Z}.$$

In Appendix A we discuss the proximal setup in normal-form games, an important special case of convex games.

Appendix C. The Special Case of Clairvoyant MWU

In the special case of normal-form games where each player's simplex strategy space Δ^{d_i} has been equipped with the negative entropy regularizer (Appendix A), the OMD-like update step on Line 7 of Algorithm 2 has the closed-form solution

$$\boldsymbol{w}_i^{(k)}[a] \propto \boldsymbol{z}_i^{(t-1)}[a] \cdot \exp\Big\{\eta \cdot \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{w}_1^{(k-1)}, \dots, \boldsymbol{w}_n^{(k-1)})[a]\Big\}$$

By plugging the above closed formula into Algorithm 1 and Algorithm 2, we recover the centralized and decentralized versions of the Clairvoyant MWU algorithm introduced by Piliouras et al. [23, 24].

As already mentioned in Appendix A, in that setting the game operator F is upper bounded (with respect to the dual norm $\|\cdot\|_{*\Delta}$) by B and is L-Lipschitz continuous with respect to the $(\|\cdot\|_\Delta, \|\cdot\|_{*\Delta})$ norm pair, where $B = L = \sqrt{n}\,V$ and V is the maximum absolute utility for any player in the game. Furthermore, negative entropy is 1-strongly convex with respect to $\|\cdot\|_\Delta$. Hence, by choosing

$$\eta := \frac{1}{2\sqrt{n}\,V}, \qquad \epsilon^{(t)} := \frac{1}{t^2},$$

from Theorem 6 we recover a per-player regret of

$$\mathrm{Reg}_i^T \le 2\sqrt{n}\,V(1 + \log(d_i)) = O(\sqrt{n}V\log(d_i))$$

for the iterates $\boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(T)}$ produced by Algorithms 1 and 2. The number of intermediate iterates $\boldsymbol{w}^{(t')}$ produced by Algorithm 2 in this case is

$$\sum_{t=1}^T N^{(t)} = \sum_{t=1}^T \left(1 + \log_2 \frac{1}{\epsilon^{(t)}} + \log_2 \max_{\boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}} \|\boldsymbol{z} - \boldsymbol{z}'\|_\Delta\right) = O(T\log T + T\log n),$$

where the last step uses the fact that

$$\max_{\boldsymbol{z},\boldsymbol{z}'\in\mathcal{Z}} \|\boldsymbol{z}-\boldsymbol{z}'\|_\Delta = \sqrt{\sum_{i=1}^{n} \max_{\boldsymbol{x},\boldsymbol{x}'\in\Delta^{d_i}} \|\boldsymbol{x}-\boldsymbol{x}'\|_1^2} \le 2\sqrt{n}.$$

This shows that from $O(T\log T + T\log n)$ iterates of $\boldsymbol{w}^{(t')}$ it is possible to extract a subsequence of T iterates (those corresponding to $\boldsymbol{w}^{(N^{(1)})}, \boldsymbol{w}^{(N^{(1)}+N^{(2)})}$, and so on), that cumulate $O(\sqrt{n}V\log(d_i))$ regret. This regret bound refines that of Piliouras et al. [24] in the dependence on the number of players ($\sqrt{n}$ rather than n).

Appendix D. Proofs

Proof of Theorem 9 Expanding the definition of the Bregman divergence, the statement is equivalent to

$$\langle \nabla\varphi(\boldsymbol{z}) - \nabla\varphi(\Pi_{\boldsymbol{z}}(\boldsymbol{g})), \hat{\boldsymbol{z}} - \Pi_{\boldsymbol{z}}(\boldsymbol{g})\rangle \le \langle \boldsymbol{g}, \hat{\boldsymbol{z}} - \Pi_{\boldsymbol{z}}(\boldsymbol{g})\rangle,$$

which in turn is equivalent to

$$\langle \boldsymbol{g} - \nabla\varphi(\boldsymbol{z}) + \nabla\varphi(\Pi_{\boldsymbol{z}}(\boldsymbol{g})), \hat{\boldsymbol{z}} - \Pi_{\boldsymbol{z}}(\boldsymbol{g})\rangle \ge 0.$$

Applying Theorem 8 yields the statement. ■

Proof of Theorem 3 The statement follows immediately from summing (4) for $t = 1, \dots, T$, and noticing that the terms $-D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(t)}) + D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(t-1)})$ telescope:

$$\begin{aligned}\eta\sum_{t=1}^{T}\langle \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{z}^{(t)}), \hat{\boldsymbol{z}}_i - \boldsymbol{z}_i^{(t)}\rangle &\overset{(4)}{\le} \sum_{t=1}^{T}\Big(-D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(t)}) + D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(t-1)}) - D_i(\boldsymbol{z}_i^{(t)} \,\|\, \boldsymbol{z}_i^{(t-1)})\Big)\\ &= D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(0)}) - D_i(\hat{\boldsymbol{z}}_i \,\|\, \boldsymbol{z}_i^{(T)}) - \sum_{t=1}^{T} D_i(\boldsymbol{z}_i^{(t)} \,\|\, \boldsymbol{z}_i^{(t-1)})\end{aligned}$$

for all $\hat{\boldsymbol{z}}_i \in \mathcal{X}_i$. Dividing by η and taking a maximum over $\hat{\boldsymbol{z}}_i \in \mathcal{X}_i$ yields first inequality in the statement. The second inequality follows immediately by using the fact that divergences are always nonnegative. ■

Proof of Theorem 4 The proof of the second part of the statement is identical to that of Theorem 2. Hence, we focus on proving (6). Again, we start from Theorem 9, this time applied with $\boldsymbol{g} = \eta F(\boldsymbol{w}^{(t)})$ and $\boldsymbol{z} = \boldsymbol{z}^{(t-1)}$:

$$\begin{aligned}D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{(t)}) - D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{(t-1)}) + D(\boldsymbol{z}^{(t)} \,\|\, \boldsymbol{z}^{(t-1)}) &\le \eta\,\langle F(\boldsymbol{w}^{(t)}), \hat{\boldsymbol{z}} - \boldsymbol{z}^{(t)}\rangle\\ &= \eta\,\langle F(\boldsymbol{w}^{(t)}), \hat{\boldsymbol{z}} - \boldsymbol{w}^{(t)}\rangle + \eta\,\langle F(\boldsymbol{w}^{(t)}), \boldsymbol{w}^{(t)} - \boldsymbol{z}^{(t)}\rangle\\ &\le \eta\,\langle F(\boldsymbol{w}^{(t)}), \hat{\boldsymbol{z}} - \boldsymbol{w}^{(t)}\rangle + \eta\|F(\boldsymbol{w}^{(t)})\|_*\|\boldsymbol{w}^{(t)} - \boldsymbol{z}^{(t)}\|\\ &\le \eta\,\langle F(\boldsymbol{w}^{(t)}), \hat{\boldsymbol{z}} - \boldsymbol{w}^{(t)}\rangle + \eta B\epsilon^{(t)},\end{aligned}$$

where the second inequality follows from the definition of the dual norm, and the last inequality from using the definition of B, introduced in Assumption 1, and (5). Rearranging yields (6). ■

Proof of Theorem 2 Take any $\hat{\boldsymbol{z}} \in \mathcal{Z}$ and use Theorem 9 with $\boldsymbol{g} = \eta F(\boldsymbol{z}^{(t)})$ and $\boldsymbol{z} = \boldsymbol{z}^{(t-1)}$:

$$D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{(t)}) - D(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{(t-1)}) + D(\boldsymbol{z}^{(t)} \,\|\, \boldsymbol{z}^{(t-1)}) \leq \eta \, \langle F(\boldsymbol{z}^{(t)}), \hat{\boldsymbol{z}} - \boldsymbol{z}^{(t)} \rangle,$$

which is exactly (3). Since the inequality holds for any $\hat{\boldsymbol{z}} \in \mathcal{Z}$, it holds in particular for any vector of the form $\hat{\boldsymbol{z}} = (\boldsymbol{z}_1^{(t)}, \ldots, \boldsymbol{z}_{i-1}^{(t)}, \hat{\boldsymbol{z}}_i, \boldsymbol{z}_{i+1}^{(t)}, \ldots, \boldsymbol{z}_n^{(t)}) \in \mathcal{Z}$. Substituting this particular choice into (3) and expanding the definitions yields (4). ■