Journal of Machine Learning Research 24 (2023) 1-24 Submitted 9/22; Revised 5/23; Published 10/23

A PDE approach for regret bounds under partial monitoring

Erhan Bayraktar ERHAN@UMICH.EDU
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109-1043, USA

Ibrahim Ekren IEKREN@UMICH.EDU
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109-1043, USA

Xin Zhang XIN.ZHANG@UNIVIE.AC.AT
Department of Mathematics
University of Vienna
Vienna, 1090 , Austria

Editor: Ambuj Tewari

Abstract

In this paper, we study a learning problem in which a forecaster only observes partial information. By properly rescaling the problem, we heuristically derive a limiting PDE on Wasserstein space which characterizes the asymptotic behavior of the regret of the forecaster. Using a verification type argument, we show that the problem of obtaining regret bounds and efficient algorithms can be tackled by finding appropriate smooth sub/supersolutions of this parabolic PDE.

Keywords: machine learning, expert advice framework, bandit problem, asymptotic expansion, Wasserstein derivative

1. Introduction

In this paper, we study a zero-sum game between a forecaster and an adversary. At each round, the forecaster chooses an action between $K \geq 2$ alternative actions based on his/her partial observations aiming at performing as well as the best constant strategy, while the adversary aims at maximizing the forecaster's regret. Our problem is motivated by *prediction with expert advice* and *bandit problems* (see e.g. Cesa-Bianchi and Lugosi (2006); Bubeck and Cesa-Bianchi (2012)), which are fundamental problems in online learning and sequential decision making. The main difference between prediction with expert advice and bandit problem is the information observed by the forecaster. In prediction with expert advice problems, the forecaster can monitor the outcomes of each alternative action, whereas in bandit problems, the forecaster can only observe the outcome of the action chosen. Thus, the former problem is a full information game whereas the latter is a bandit game (see e.g. Audibert and Bubeck (2010); Audibert et al. (2011)).

The most commonly used algorithm for decision making and prediction problem is the so-called *multiplicative weights algorithm*, which assigns initial weights to each expert, update these weights multiplicatively and iteratively based on their performance, and randomly choose experts according to their weights. This simple algorithm is widely used and has been proven efficient in practice. However, it cannot provide accurate regret bounds and best strategies for the forecaster. In Drenska and Kohn (2020), techniques from partial differential equations were first employed to understand asymptotic behavior of prediction of expert advice problems. Since then, it became popular and has been proven powerful in certain problems, see e.g. Drenska and Kohn (2023); Drenska and Calder (2021); Calder and Drenska (2021); Kobzar and Kohn (2022); Kobzar et al. (2020); Bayraktar et al. (2021a, 2020b,a, 2021b); Harvey et al. (2020); Zhang et al. (2022); Greenstreet et al. (2022).

In full information games, these papers rely on the fact that the difference $(X_t^i)_{i=1,\dots,N} = (G_t^i - G_t)_{i=1,\dots,N} \in \mathbb{R}^K$ between the gain G_t of the forecaster and the gain G_t^i of each action i is a natural state variable for the dynamic game between the forecaster and the adversary. Thus, the minimax regret of the forecaster satisfies a finite dimensional dynamic programming principle whose scaling limit is a parabolic partial differential equation on $\mathbb{R}^N$. For bandit games or in the presence of partial information such methodology cannot be applied. Indeed, due to partial information, the natural state variable for the dynamic programming principle is the set of probability distributions on $\mathbb{R}^N$ which encodes the distribution m_t of X_t conditional on the information of the forecaster. Thus, with partial information, the fundamental problem is to understand the dynamics of m_t and how these dynamics behave in the long-time regime.

Our main contribution consists in showing that the update of the conditional distribution between two consecutive time steps from m_t to m_{t+1} admits a scaling limit that can be described using partial differential equations in the Wasserstein space. The equations we obtain are fully nonlinear versions of the PDEs appearing in mean-field games and Mckean-Vlasov control problems, see e.g. Cardaliaguet et al. (2019); Cosso et al. (2021); Bandini et al. (2019); Bayraktar et al. (2023a) and Remark 9 for further discussions. This novel relation between the discrete-time bandit problem and the continuous-time equations comes from the fact that in the game we study, the updated measure m_{t+1} can be written as a push-forward operator on m_t, i.e. $m_{t+1} = (Id + Y_t)\sharp m_t$ where Y_t is a (random) function describing the feature learned by the forecaster on $[t, t+1]$. If the game is played T times and if we rescale the problem with its natural $\sqrt{T}$ scaling, the update of the m_t can be written as $m_{t+\frac{1}{T}} = (Id + \frac{Y_t}{\sqrt{T}})\sharp m_t$. In the long-time regime, i.e., as $T \to \infty$, by the definition of the Wasserstein derivative (see (Cardaliaguet et al., 2019, Proposition 2.3)), we obtain that for any smooth function U, we have the expansion

$$U\left(m_{t+\frac{1}{T}}\right) = U\left(m_t\right) + \frac{1}{\sqrt{T}} \int D_m U(m_t, x) Y_t(x)\, m(dx) + \mathcal{O}\left(\frac{1}{\sqrt{T}}\right).$$

Thus, the impact of the Bayesian update of the distribution m_t can be characterized in the long-time regime using the Wasserstein derivative $D_m U$. In fact, we derive a second order expansion of $U(m_{t+\frac{1}{T}})$ involving the derivatives $D_x D_m U$ and $D^2_{mm} U$ which allows us to heuristically exhibit a second order parabolic equation of type

$$\begin{aligned} 0 = {} & \partial_t U(t,m) && (1) \\ & + F\left(\int D_m U(t,m,x) m(dx), \int D_x D_m U(t,m,x) m(dx), \iint D^2_{mm} U(t,m,x,y) m(dx) m(dy)\right) \end{aligned}$$

which is expected to govern the dynamics of the prediction problem in the long-time regime. In this equation, the unknown is the function U, F is a function that can be explicitly computed from the Bayes' rule, and the derivatives are defined as in Cardaliaguet et al. (2019); Cosso et al. (2021).

The equation (1) gives simple methods to obtain algorithms and regret bounds for the long-time regime of the prediction problem with partial information. Indeed, using a verification type argument we show that the gradient $D_m\phi$ of any smooth supersolution ϕ of (1) satisfying some growth and terminal condition yields an algorithm that guarantees an upper bound for regret of order $\phi(0,\delta_0)\sqrt{T}$ where δ_0 is the Dirac mass at 0. Heuristically, the algorithm corresponding to $D_m\phi$ is the probability matching algorithm as in Drenska and Kohn (2020); Bayraktar et al. (2020a), and is a tradeoff between exploitation and exploration; see Remark 13. A similar result also holds for appropriate subsolutions.

Due to the nonlinearity on the second derivative term $D^2_{mm}U$, wellposedness of viscosity solutions for (1) is not available in the literature. Hence, the questions of establishing appropriate comparison result for viscosity solutions and obtaining the exact growth of the regret as for example in Drenska and Kohn (2020) are left for future research.

The rest of this paper is organized as follows. In Section 2, we formulate our problem and show that the value function of the game depends only on the law m_t of X_t conditional on the information of the agents. Then, using Bayes' rule, we compute explicitly the update of beliefs and prove a dynamic programming principle. In Section 3, by properly rescaling the value function and using differential calculus on the space of measures, we heuristically obtain a limiting PDE of type (1) on the Wasserstein space. In Section 4 and 5, using smooth supersolutions and subsolutions of the PDE, we construct strategies for the forecaster and the adversary, and find upper and lower bounds of expected regret.

1.1 Notations

For any positive integer K, define $[K] = \{1,\dots,K\}$, $\{\pm i\} := \{\pm i : i \in [K]\}$, and $\mathcal{S}_K$ to be the set of positive semidefiniete $K \times K$ matrices. Id stands for the identity mapping of appropriate dimension. For any $x \in \mathbb{R}^K$, denote its i-th coordinate by x^i. Let $\{e_i : i = 1,\dots,K\}$ be the canonical basis of $\mathbb{R}^K$, and for any $j \subset [K]$, denote $e_j = \sum_{i,i\in j} e_i$ and $e = \sum_{i=1}^K e_i$.

We fix $K \geq 2$ and denote by $\mathcal{P}_2(\mathbb{R}^K)$ the set of probability measures m on $\mathbb{R}^K$ such that $\int |x|^2\, m(dx) < \infty$. For any $v \in \mathbb{R}^K$, $\lambda \in \mathbb{R}$, and $m \in \mathcal{P}_2(\mathbb{R}^K)$, we define the measures $m_{\sharp v} := (Id+v)\sharp m$ and $m^{*\lambda}$ via

$$\int f(x)\, m_{\sharp v}(dx) = \int f(x+v)\, m(dx),$$
$$\int f(x)\, m^{*\lambda}(dx) = \int f(\lambda x)\, m(dx) \text{ for all } f \text{ continuous and bounded.}$$

Additionally, for any function f and $m \in \mathcal{P}_2(\mathbb{R}^K)$, we denote

$$f([m]) := \int f(x)\, m(dx).$$

2. Formulation of the problem

Our online prediction problem with partial observation can be described as a T-round game, played by a forecaster in an adversarial environment. Suppose that there are K actions. At each round t, the forecaster chooses an action $I_t \in [K]$, and independently the adversary chooses a set of winning actions $J_t \subset [K]$. The reward of action i is 1 if $i \in J_t$ and is 0 if $i \notin J_t$. Then the total gain of the forecaster G_t and the total gain G_t^i of action i evolve as

$$\begin{aligned} &G_{t+1} - G_t = \mathbf{1}_{I_t \in J_t}, \\ &G_{t+1}^i - G_t^i = \mathbf{1}_{i \in J_t}, \ \ i = 1, \dots, K. \end{aligned}$$

The goal of the forecaster is to design a robust strategy that performs as well as the best constant strategy under any adversarial environment, i.e., to minimize $\max\limits_{Adversary} \mathbb{E}[\max_i X_T^i]$, where $X_t^i := G_t^i - G_t$ is the state variable evolving as

$$X_{t+1} - X_t := e_{J_t} - \mathbf{1}_{I_t \in J_t} e \in \mathbb{R}^K.$$

Both the forecaster and the adversary are allowed to adopt randomized strategies. At each round t, they decide on distributions $b_t \in \mathcal{P}([K])$ of I_t and $a_t \in \mathcal{P}(\{0,1\}^K)$ of J_t respectively. If we allow both agents to observe the outcomes of I_t and J_t, this problem is the classical prediction with expert advice problem in the adversarial setting, see for example Cover (1966); Cesa-Bianchi and Lugosi (2006); Gravin et al. (2016); Drenska and Kohn (2020).

Let us now describe information observed by the forecaster and his/her admissible strategies in the partial information problem we aim to study. At initial time $t = 0$, both the adversary and the forecaster get informed of the distribution m_0 of X_0. For any $t \geq 0$, the random variable

$$Y_t := \mathbf{1}_{I_t \in J_t} I_t - \mathbf{1}_{I_t \notin J_t} I_t \in \{\pm i\}$$

indicates whether the forecaster makes a good decision or not. Both players can observe the law of adversary's control a_{t-1} and the indicator y_{t-1}. Their accumulated information is given by

$$h_t := (m_0, a_0, y_0 \dots, a_{t-1}, y_{t-1}) \in \mathcal{H}_t, \quad (h_0 := m_0 \in \mathcal{H}_0),$$

where $\mathcal{H}_t := \mathcal{P}(\mathbb{R}^K) \times \left(\mathcal{P}(\{0,1\}^K) \times \{\pm i\}\right)^t$. The strategies of the forecaster and the adversary are measurable functions $\beta_t : \mathcal{H}_t \to \mathcal{P}([K])$ and $\alpha_t : \mathcal{H}_t \to \mathcal{P}(\{0,1\}^K)$ respectively. Note that α_t, β_t denote functions of $\mathcal{H}_t$, while a_t, b_t denote the output of α_t, β_t respectively. Define $\mathcal{A}$ to be the set of all possible strategies $\alpha := (\alpha_0, \alpha_1, \dots, \alpha_{T-1})$, and $\mathcal{B}$ similarly.

Suppose this game starts from time t with an initial distribution $m \in \mathcal{P}(\mathbb{R}^K)$. Then given any strategies $\alpha \in \mathcal{A}$, $\beta \in \mathcal{B}$, the regret for the forecaster is given by

$$\gamma_T(t, m, \alpha, \beta) := \mathbb{E}^{m,\alpha,\beta}[\max_i X_T^i \mid X_t \sim m].$$

From the perspective of the forecaster, we aim at solving a minimax problem

$$v_T(t, m) := \inf_{\beta \in \mathcal{B}} \sup_{\alpha \in \mathcal{A}} \gamma_T(t, m, \alpha, \beta), \tag{2}$$

and we denote this two player game by $\Gamma_T(t, m)$. Note that on the state space $\mathcal{P}(\mathbb{R}^K)$ this game is of complete information. As we will see in the next subsection, the information structure allows both agents to compute the value function iteratively backward in time. Moreover, our final payoff is regret $\mathbb{E}^{m,\alpha,\beta}[\max_i X_T^i \mid X_t \sim m]$ instead of pseudo-regret $\max_i \mathbb{E}^{m,\alpha,\beta}[X_T^i \mid X_t \sim m]$; see for example Bubeck and Cesa-Bianchi (2012). Since the final payoffs can be written as functions of $\mathcal{L}(X_T)$ in both cases, our PDE method can also be applied to pseudo-regret analysis.

Remark 1 *(i) Our formulation is motivated by the classical bandit problems; see for example Cesa-Bianchi and Lugosi (2006); Bubeck and Cesa-Bianchi (2012). Similar to the bandit problems, the forecaster observes partial information Y_t instead of X_t, and both adversary and forecaster choose their strategies to be played at each round. In our framework both players observe Y_t, a_t and hence they can compute the same update of m_t. In the adversarial bandit problem, the forecaster cannot observe the randomization a_t of the adversary and therefore cannot update m_t. This additional uncertainty in future dynamics of m_t is the main difference between our framework and the classical adversarial bandit problems.*

Bayesian approaches have also been widely used in stochastic multi-armed problems, see for example Agrawal and Goyal (2013) and the references therein. In such problems, reward functions are i.i.d. with some fixed unknown distribution, and the belief of the forecaster on the reward functions are updated at each round by Bayes' rules. However in our framework, distributions of reward functions are chosen by the adversary so as to maximize the regret of forecaster.

(ii)

Gangs of bandit problems have applications in online recommendation systems; see for example Cesa-Bianchi et al. (2013), and Herbster et al. (2021) which considers this in the adversarial setup. In these problems, recommendation systems serve content to a group of users by taking advantage of underlying network of social relationships among them. The system makes the same recommendation to users similar to each other. We consider a simplified framework where all the users are of the same type and their feedbacks J_t are sampled from the same distribution a_t. Assume that the privacy cookie allows the system to collect statistical data of J_t without knowing the identity of users. In this way, the system learns the distribution a_t of J_t, and it aims to make recommendations in a robust manner.

(iii) In our context, since the forecaster learns a_t, he/she can update m_t via Bayes' rule. This update is impossible in the classical bandit problems. An interesting question that is left for future research is to extend our PDE tools to classical bandit problems.

2.1 Dynamic programming principle

In this subsection, we establish the dynamic programming principle for the game (2), and reduce controls α, β to functions of conditional distribution of the state X. Let us first compute the distribution of X, i.e., belief, given prior information. Suppose the current distribution is m and X is a random variable with distribution m. We denote ΔX the change of X between two rounds. The players choose strategies $a \in \mathcal{P}(\{0,1\}^K)$ and $b \in \mathcal{P}([K])$ respectively, and receive signal $y \in \{\pm i\}$. We denote by $\mathcal{L}^{a,b}$ the distribution of a random variable and by $\mathbb{P}^{a,b}$ the probability of an event given the strategies of the agents. We omit the superscripts a or b if this dependence is clear from the context. We also denote by

$$l(m,a,y) := \mathcal{L}^a(X + \Delta X | X \sim m, Y = y) \in \mathcal{P}(\mathbb{R}^K)$$

the Bayesian update of the distribution.

We will compute the explicit formula of $l(m,a,y)$ in the next Lemma. For any $a \in \mathcal{P}(\{0,1\}^K)$, $b \in \mathcal{P}([K])$, denote

$$\hat{a}(i) := \sum_{j:i\in j} a(j), \quad \hat{a}(-i) := \sum_{j:i\notin j} a(j), \quad \forall\, i \in [K].$$

Proposition 2 *Given $a \in \mathcal{P}(\{0,1\}^K)$ and the distribution $m \in \mathcal{P}(\mathbb{R}^K)$, we have that*

$$l(m,a,i) = \sum_{j:i\in j} \left(\frac{a(j)}{\hat{a}(i)} m\right)_{\sharp - e_{j^c}}, \quad l(m,a,-i) = \sum_{j:i\notin j} \left(\frac{a(j)}{\hat{a}(-i)} m\right)_{\sharp e_j}.$$

We make the convention in these expressions that $l(m,a,y) = \delta_{\mathbf{0}} \in \mathcal{P}(\mathbb{R}^K)$ whenever $\hat{a}(y) = 0$.

Proof For $j \subset [K]$ and $i \in [K]$, it can be easily verified that

$$\begin{aligned}
\mathbb{P}(\Delta X = e_j, Y = -i, X \in dx) &= \mathbf{1}_{i\notin j} a(j) b(i) m(dx)\\
\mathbb{P}(\Delta X = -e_{j^c}, Y = i, X \in dx) &= \mathbf{1}_{i\in j} a(j) b(i) m(dx)\\
\mathbb{P}(Y = i) &= b(i) \sum_{k:i\in k} a(k)\\
\mathbb{P}(Y = -i) &= b(i) \sum_{k:i\notin k} a(k)\\
\mathbb{P}(\Delta X = e_j, X \in dx \,|\, Y = -i) &= \frac{\mathbf{1}_{i\notin j} a(j) m(dx)}{\sum_{k:i\notin k} a(k)}\\
\mathbb{P}(\Delta X = -e_{j^c}, X \in dx \,|\, Y = i) &= \frac{\mathbf{1}_{i\in j} a(j) m(dx)}{\sum_{k:i\in k} a(k)}.
\end{aligned}$$

Therefore, conditioning on Y, the distribution of $X + \Delta X$ is given by

$$\begin{aligned}
\mathbb{P}\left[(X + \Delta X) \in dx \,|\, Y = i\right] &= \sum_{j:i\in j} \mathbb{P}\left[X \in d(x + e_{j^c}), \Delta X = -e_{j^c} \,|\, Y = i\right]\\
&= \sum_{j:i\in j} \frac{a(j) m(d(x + e_{j^c}))}{\hat{a}(i)},\\
&= \sum_{j:i\in j} \left(\frac{a(j)}{\hat{a}(i)} m\right)_{\sharp - e_{j^c}} (dx),
\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\left[(X+\Delta X)\in dx\,|\,Y=-i\right] &= \sum_{j:i\notin j}\mathbb{P}\left[X\in d(x-e_j),\Delta X=e_j\,|\,Y=-i\right]\\ &= \sum_{j:i\notin j}\frac{a(j)m(d(x-e_j))}{\hat{a}(-i)}\\ &= \sum_{j:i\notin j}\left(\frac{a(j)}{\hat{a}(-i)}m\right)_{\sharp e_j}(dx).\end{aligned}$$

■

The following theorem proves a dynamic programming principle showing that one can solve (2) with a backward induction.

Theorem 3 *For any distribution $m\in\mathcal{P}(\mathbb{R}^K)$ and $T\in\mathbb{N}$ we have that*

$$\begin{aligned}v_T(t,m)= \inf_{b\in\mathcal{P}([K])}\sup_{a\in\mathcal{P}(\{0,1\}^K)}&\left(\sum_{i=1}^K b(i)\hat{a}(i)v_T(t+1,l(m,a,i))\right.\\ &\left.+\sum_{i=1}^K b(i)\hat{a}(-i)v_T(t+1,l(m,a,-i))\right),\end{aligned}\tag{3}$$

where $b(i)\hat{a}(i)$, $b(i)\hat{a}(-i)$ represent the probability of receiving signal i, $-i$ respectively, and $l(m,a,\pm i)$ is the update of beliefs.

Proof The equation (3) holds trivially for $t=T-1$. Suppose it is true for $t+1$. Let us prove it for t. Denote by v the value of the right hand side of (3). For any $\alpha\in\mathcal{A}$ and $\beta\in\mathcal{B}$, denote $\alpha_{t+1:T}=\{\alpha_{t+1},\dots,\alpha_{T-1}\}$, $\beta_{t+1:T}=\{\beta_{t+1},\dots,\beta_{T-1}\}$. It is clear that

$$\begin{aligned}\gamma_T(t,m,\alpha,\beta)=&\sum_{i=1}^K\sum_{k:i\in k}\beta_t(i)\alpha_t(k)\gamma_T(t+1,l(m,\alpha_t,i),\alpha_{t+1:T},\beta_{t+1:T})\\ &+\sum_{i=1}^K\sum_{k:i\notin k}\beta_t(i)\alpha_t(k)\gamma_T(t+1,l(m,\alpha_t,-i),\alpha_{t+1:T},\beta_{t+1:T}),\end{aligned}\tag{4}$$

where $l(m,\alpha_t,\pm i)$ is the conditional distribution of X_{t+1}. For the game $\gamma_T(t+1,l(m,\alpha_t,\pm i))$, due to our induction hypothesis, the value of this game exists and is just $v_T(t+1,l(m,\alpha_t,\pm i))$. Taking supremum over α on both sides of (4), it can be easily seen that

$$\begin{aligned}\sup_\alpha\gamma_T(t,m,\alpha,\beta)\geq&\sup_{\alpha_t}\left(\sum_{i=1}^K\sum_{k:i\in k}\beta_t(i)\alpha_t(k)v_T(t+1,l(m,\alpha_t,i))\right.\\ &\left.+\sum_{i=1}^K\sum_{k:i\notin k}\beta_t(i)\alpha_t(k)v_T(t+1,l(m,\alpha_t,-i))\right).\end{aligned}$$

Taking infimum over β, we conclude that $v_T(c,m) \geq v$.

Then we prove that for any $\epsilon > 0$, there exists a robust strategy β^* of the forecaster such that

$$\sup_{\alpha} \gamma_T(t,m,\alpha,\beta^*) < v + 2\epsilon. \tag{5}$$

Take $\beta_t^* \in \mathcal{P}([K])$ with the property that

$$\begin{aligned} v + \epsilon > \sup_{a \in \mathcal{P}(\{0,1\}^K)} \Bigg(& \sum_{i=1}^{K} \sum_{k: i \in k} \beta_t^*(i) a(k) v_T(t+1, l(m,a,i)) \\ & + \sum_{i=1}^{K} \sum_{k: i \notin k} \beta_t^*(i) a(k) v_T(t+1, l(m,a,-i)) \Bigg). \end{aligned}$$

By induction hypothesis, for any belief $l(m,a,\pm i)$, the forecaster can choose a strategy $\beta_{t+1:T}^*$ such that

$$v_T(t+1, l(m,a,\pm i)) + \epsilon > \sup_{\alpha_{t+1:T}} \gamma_T(t+1, l(m,a,\pm i), \alpha_{t+1:T}, \beta_{t+1:T}^*).$$

Taking $\beta^* = (\beta_t^*, \beta_{t+1:T}^*)$, clearly it is measurable and satisfies (5). ■

3. Heuristic expansion of the rescaled value function

Let us define the rescaled value functions

$$u^T(s,m) := \frac{1}{\sqrt{T}} v_T\left(\lceil sT \rceil, m^{*\sqrt{T}}\right),$$

and equivalently

$$v_T(\lceil sT \rceil, m) = \sqrt{T} u^T\left(s, m^{*\sqrt{T^{-1}}}\right).$$

For any $a \in \mathcal{P}(\{0,1\}^K)$ and belief $m \in \mathcal{P}(\mathbb{Z}^K)$, denote

$$A^{a,m}_{i,\sqrt{T}} = \left(\sum_{j: i \in j} \left(\frac{a(j)}{\hat{a}(i)} m^{*\sqrt{T}} \right)_{\sharp - e_{j^c}} \right)^{*\frac{1}{\sqrt{T}}}, \quad A^{a,m}_{-i,\sqrt{T}} = \left(\sum_{j: i \notin j} \left(\frac{a(j)}{\hat{a}(-i)} m^{*\sqrt{T}} \right)_{\sharp e_j} \right)^{*\frac{1}{\sqrt{T}}}. \tag{6}$$

Then due to (3), it holds that

$$\begin{aligned} u^T\left(s - \frac{1}{T}, m\right) = \inf_{b \in \mathcal{P}([K])} \sup_{a \in \mathcal{P}(\{0,1\}^K)} \Bigg(& \sum_i b(i) \hat{a}(i) u^T\left(s, A^{a,m}_{i,\sqrt{T}}\right) \\ & + \sum_i b(i) \hat{a}(-i) u^T\left(s, A^{a,m}_{-i,\sqrt{T}}\right) \Bigg), \end{aligned} \tag{7}$$

with the terminal condition

$$u^T(1,m) = \int_{x\in\mathbb{R}^K} \max_i x^i \, m(dx).$$

Now we want to derive a limit for (7) as $T \to \infty$. This derivation requires us to take derivatives in the direction of $A^{a,m}_{i,\sqrt{T}} - m$ and $A^{a,m}_{-i,\sqrt{T}} - m$ in the Wasserstein space. Let us introduce the differentiability of functions over the Wasserstein space as defined in Cardaliaguet et al. (2019); Carmona et al. (2018).

A function $u : \mathcal{P}_2(\mathbb{R}^K) \mapsto \mathbb{R}$ is said to be Fréchet differentiable if there exists a continuous function

$$\frac{\delta u}{\delta m} : \mathcal{P}_2(\mathbb{R}^K) \times \mathbb{R}^K \mapsto \mathbb{R}$$

so that for all $(m, m') \in \mathcal{P}_2(\mathbb{R}^K)$, we have that

$$\lim_{h\to 0} \frac{u(m + h(m'-m)) - u(m)}{h} = \int \frac{\delta u}{\delta m}(m,x)\,(m'-m)(dx).$$

Whenever $\frac{\delta u}{\delta m}$ is differentiable in x, we also define

$$D_m u(m,x) = D_x \frac{\delta u}{\delta m}(m,x) \in \mathbb{R}^K.$$

We define $D_x D_m u(m,x)$ to be the derivative of $x \mapsto D_m u(m,x)$ in x, and $D^2_{mm} u(m,x,y)$ to be the derivative of $m \mapsto D_m u(m,x)$ in m as above; see Cardaliaguet et al. (2019); Chow and Gangbo (2019).

Definition 4 *A function $u : \mathcal{P}_2(\mathbb{R}^K) \to \mathbb{R}$ is said to be $\mathcal{C}^1$ if $D_m u(m,x)$ is continuous and has at most quadratic growth in x, i.e.,*

$$|D_m u(m,x)| \le C(1 + |x|^2).$$

It is said to be $\mathcal{C}^2$ if $D_x D_m u(m,x)$ and $D^2_{mm} u(m,x,y)$ are continuous, and have at most quadratic growth in x and (x,y) respectively.

It is shown in (Cardaliaguet et al., 2019, Proposition 2.3) that $D_m u$ can be understood as a derivative of u along push-forward directions, meaning that for all Borel measurable bounded vector field $\phi : \mathbb{R}^K \mapsto \mathbb{R}^K$ we have

$$\lim_{h\to 0} \frac{u((Id + h\phi)_\sharp m) - u(m)}{h} = \int D_m u(m,x)\phi(x)\, m(dx).$$

Due to the expression of $A^{a,m}_{i,\sqrt{T}}$ and $A^{a,m}_{-i,\sqrt{T}}$, we need to take derivatives in the directions $\left(Id + \frac{e_j}{T}\right)$ which are constant vector fields. However, the presence of terms $\frac{a(j)}{\hat{a}(i)}\sharp m$ in (6) is a randomization among the directions of the vector fields. The following Proposition shows that at the leading order, we can simplify these perturbations by averaging over these different vector fields. We recall the notational convention that for all $m' \in \mathcal{P}_2(\mathbb{R}^K)$

$$D_m u(m,[m']) = \int D_m u(m,x)\, m'(dx) \in \mathbb{R}^K.$$

Proposition 5 *Suppose $u \in \mathcal{C}^1(\mathcal{P}(\mathbb{R}^K);\mathbb{R})$. Then for all $a \in \mathcal{P}(\{0,1\}^K)$ and $i \in [K]$, we have that*

$$\lim_{T\to\infty} \sqrt{T}\left(u(A^{a,m}_{i,\sqrt{T}}) - u(m)\right) = -\mathcal{V}^\top_{a,i} D_m u\,(m,[m])$$
$$\lim_{T\to\infty} \sqrt{T}\left(u(A^{a,m}_{-i,\sqrt{T}}) - u(m)\right) = \mathcal{V}^\top_{a,-i} D_m u\,(m,[m])\,,$$

where

$$\mathcal{V}_{a,i} := \sum_{j:i\in j} \frac{a(j)}{\hat{a}(i)} e_{j^c} \in \mathbb{R}^K, \quad \mathcal{V}_{a,-i} := \sum_{j:i\notin j} \frac{a(j)}{\hat{a}(-i)} e_j \in \mathbb{R}^K.$$

Remark 6 *Note that $-\mathcal{V}_{a,i} \in \mathbb{R}^K$ (resp. $\mathcal{V}_{a,-i} \in \mathbb{R}^K$) represents the increase in the expectation of X_t given the information that $Y = i$ (resp. $Y = -i$) and the adversary's strategy a.*

Proof Let us only compute the derivative in the direction of $A^{a,m}_{i,\sqrt{T}} - m$. By the definition of $\frac{\delta u}{\delta m}$, denoting $\widetilde{A}_{s,\sqrt{T},m} = m + s(A^{a,m}_{i,\sqrt{T}} - m)$ we have that

$$\begin{aligned}
&\sqrt{T}(u(A^{a,m}_{i,\sqrt{T}}) - u(m)) \\
&= \sqrt{T}\int_0^1 \int \frac{\delta u}{\delta m}\left(\widetilde{A}_{s,\sqrt{T},m}, x\right)\,(A^{a,m}_{i,\sqrt{T}} - m)(dx)\,ds \\
&= \sum_{j:i\in j} \frac{a(j)}{\hat{a}(i)} \sqrt{T}\int_0^1 \int \frac{\delta u}{\delta m}\left(\widetilde{A}_{s,\sqrt{T},m}, x - \frac{e_{j^c}}{\sqrt{T}}\right) - \frac{\delta u}{\delta m}\left(\widetilde{A}_{s,\sqrt{T},m}, x\right)\,m(dx)\,ds,
\end{aligned}$$

and thus

$$\lim_{T\to\infty} \sqrt{T}\left(u(A^{a,m}_{i,\sqrt{T}}) - u(m)\right) = -\sum_{j:i\in j} \frac{a(j)}{\hat{a}(i)} \int e^\top_{j^c} D_m u\,(m,x)\;m(dx).$$

■

We can now give the second order expansion along $T \mapsto u(A^{a,m}_{y,\sqrt{T}})$ for all $y = \pm i$.

Proposition 7 *Suppose $u \in C^2(\mathcal{P}(\mathbb{R}^K);\mathbb{R})$. Then we have that*

$$\begin{aligned}
&\lim_{T\to\infty} T\left(u(A^{a,m}_{i,\sqrt{T}}) - u(m) + \frac{1}{\sqrt{T}}\mathcal{V}^\top_{a,i} D_m u\,(m,[m])\right) \\
&= \frac{1}{2}\sum_{j:i\in j} \frac{a(j)}{\hat{a}(i)} e^\top_{j^c} D_x D_m u\,(m,[m])\,e_{j^c} \qquad (8) \\
&\quad + \frac{1}{2}\sum_{k,j:i\in k, i\in j} \frac{a(j)}{\hat{a}(i)}\frac{a(k)}{\hat{a}(i)} e^\top_{j^c} D^2_{mm} u\,(m,[m],[m])\,e_{k^c},
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{T\to\infty} T\left(u(A^{a,m}_{-i,\sqrt{T}}) - u(m) - \frac{1}{\sqrt{T}}\mathcal{V}^{\top}_{a,-i} D_m u\left(m,[m]\right)\right)\\
&= \frac{1}{2}\sum_{j:i\notin j}\frac{a(j)}{\hat{a}(-i)} e_j^{\top} D_x D_m u\left(m,[m]\right) e_j \qquad (9)\\
&+ \frac{1}{2}\sum_{k,j:i\notin k, i\notin j}\frac{a(j)}{\hat{a}(-i)}\frac{a(k)}{\hat{a}(-i)} e_j^{\top} D^2_{mm} u\left(m,[m],[m]\right) e_k
\end{aligned}$$

Remark 8 *The Propositions 5 and 7 show that, at the leading orders, the impact of the scaled update $A^{a,m}_{y,\sqrt{T}}$ of m on a smooth function u can be characterized by multiplication of $D_m u$, $D_x D_m u$, and $D^2_{mm} u$ with some matrices depending only on a.*

Proof Using the (Cardaliaguet et al., 2019, Equality (25)), we have

$$\begin{aligned}
&T\left(u(A^{a,m}_{i,\sqrt{T}}) - u(m) + \frac{1}{\sqrt{T}}\sum_{j:i\in j}\frac{a(j)}{\hat{a}(i)}\int e_{j^c}^{\top} D_m u\left(m,x\right) dm(x)\right)\\
&= \sum_{j:i\in j}\frac{a(j)}{\hat{a}(i)} T\int_0^1\int \frac{\delta u}{\delta m}\left(\widetilde{A}_{s,\sqrt{T},m}, x - \frac{e_{j^c}}{\sqrt{T}}\right) - \frac{\delta u}{\delta m}\left(\widetilde{A}_{s,\sqrt{T},m}, x\right)\\
&\quad + \frac{e_{j^c}^{\top}}{\sqrt{T}} D_x \frac{\delta u}{\delta m} u\left(m,x\right) dm(x) ds.
\end{aligned}$$

Let us compute the limit of integrand on the right hand side. By Taylor expansion on x, it can ben seen that

$$\begin{aligned}
&T\left(\frac{\delta u}{\delta m}\left(\widetilde{A}_{s,\sqrt{T},m}, x - \frac{e_{j^c}}{\sqrt{T}}\right) - \frac{\delta u}{\delta m}\left(\widetilde{A}_{s,\sqrt{T},m}, x\right) + \frac{e_{j^c}^{\top}}{\sqrt{T}} D_x \frac{\delta u}{\delta m} u\left(m,x\right)\right)\\
&= \frac{1}{2} e_{j^c}^{\top} D_x^2 \frac{\delta u}{\delta m}\left(\widetilde{A}_{s,\sqrt{T},m}, \widetilde{x}_T\right) e_{j^c} - \sqrt{T} e_{j^c}^{\top}\left(D_x \frac{\delta u}{\delta m}\left(\widetilde{A}_{s,T,m}, x\right) - D_x \frac{\delta u}{\delta m} u\left(m,x\right)\right)
\end{aligned}$$

where $\widetilde{x}_T$ is some point on the line segment joining x and $x - \frac{e_{jc}}{\sqrt{T}}$. Denoting $\widetilde{A}_{r,s,\sqrt{T},m} = m + r(\widetilde{A}_{s,\sqrt{T},m} - m)$, the right hand side of the above equation equals to

$$\begin{aligned}
&\frac{1}{2} e_{j^c}^{\top} D_x^2 \frac{\delta u}{\delta m}\left(\widetilde{A}_{s,\sqrt{T},m}, \widetilde{x}_T\right) e_{j^c} - s\sum_{k:i\in k}\frac{a(k)}{\hat{a}(i)}\int_0^1\int\\
&\sqrt{T}\left(e_{j^c}^{\top} D_x \frac{\delta^2 u}{\delta m^2}\left(\widetilde{A}_{r,s,\sqrt{T},m}, x, x' - \frac{e_{k^c}}{\sqrt{T}}\right) - e_{j^c}^{\top} D_x \frac{\delta^2 u}{\delta m^2}\left(\widetilde{A}_{r,s,\sqrt{T},m}, x, x'\right)\right) dm(x')dx.
\end{aligned}$$

Letting $T \to \infty$, it converges to

$$\frac{1}{2} e_{j^c}^{\top} D_x^2 \frac{\delta u}{\delta m}\left(m,x\right) e_{j^c} + s\sum_{k:i\in k}\frac{a(k)}{\hat{a}(i)} e_{j^c}^{\top}\int D^2_{x,x'}\frac{\delta^2 u}{\delta m^2} u\left(m,x,x'\right) e_{k^c}\, dm(x'),$$

and hence we obtain (8) by integrating over x. Similar computation yields to (9). ■

We now use (7) to obtain a formal asymptotics for u^T as $T \to \infty$. Assuming u^T converges to a $\mathcal{C}^2$ function $u : [0,1] \times \mathcal{P}(\mathbb{R}^K) \to \mathbb{R}$, the dynamic programming principle yields to

$$\begin{aligned} 0 = \inf_{b \in \mathcal{P}([K])} \sup_{a \in \mathcal{P}(\{0,1\}^K)} \sum_i b(i)\hat{a}(i)T \left(u^T \left(s, A^{a,m}_{i,\sqrt{T}} \right) - u^T \left(s - \frac{1}{T}, m \right) \right) \\ + b(i)\hat{a}(-i)T \left(u^T \left(s, A^{a,m}_{-i,\sqrt{T}} \right) - u^T \left(s - \frac{1}{T}, m \right) \right). \end{aligned}$$

Using Proposition 5 and 7 for large enough T, we obtain that

$$\begin{aligned} \mathcal{O}(1) &= \partial_t u(t,m) + \inf_{b \in \mathcal{P}([K])} \sup_{a \in \mathcal{P}(\{0,1\}^K)} \sqrt{T} \sum_i b(i) \left(\hat{a}(-i)\mathcal{V}_{a,-i} - \hat{a}(i)\mathcal{V}_{a,i} \right)^\top D_m u\left(t,m,[m]\right) \\ &+ \frac{1}{2} b(i)\hat{a}(i) \left(\mathcal{V}_{a,i}^\top D^2_{mm} u\left(t,m,[m],[m]\right) \mathcal{V}_{a,i} + \sum_{j : i \in j} \frac{a(j)}{\hat{a}(i)} e_{j^c}^\top D_x D_m u\left(t,m,[m]\right) e_{j^c} \right) \\ &+ \frac{1}{2} b(i)\hat{a}(-i) \left(\mathcal{V}_{a,-i}^\top D^2_{mm} u\left(t,m,[m],[m]\right) \mathcal{V}_{a,-i} + \sum_{j : i \notin j} \frac{a(j)}{\hat{a}(-i)} e_j^\top D_x D_m u\left(t,m,[m]\right) e_j \right). \end{aligned} \tag{10}$$

Notice that $A^{a,m_{\sharp\epsilon\mathbf{1}}}_{y,\sqrt{T}} = \left(A^{a,m}_{y,\sqrt{T}} \right)_{\sharp\epsilon\mathbf{1}}$ for any $y \in \{\pm i\}$, and the final condition satisfies $u^T(1, m_{\sharp\epsilon\mathbf{1}}) = u^T(1,m) + \epsilon$. Therefore by backward induction, we have $u^T(t, m_{\sharp\epsilon\mathbf{1}}) = u^T(t,m) + \epsilon$ for any $t \in [0,1]$, and also in its limit as $T \to \infty$

$$u(t, m_{\sharp\epsilon\mathbf{1}}) = u(t,m) + \epsilon.$$

Thus, thanks to (Cardaliaguet et al., 2019, Proposition 2.3), we have that

$$\mathbf{1}^\top D_m u(t,m,[m]) = 1.$$

Additionally, each component of $D_m u(t,m,[m])$ is clearly non-negative, which implies that $D_m u(t,m,[m]) \in \mathbb{R}^K$ is simplex valued. Denoting $u_i(t,m)$ the ith component of $D_m u\left(t,m,[m]\right)$, we have that

$$\begin{aligned} &\sum_i b(i) \left(\sum_{j : i \notin j} a(j) e_j^\top - \sum_{j : i \in j} a(j) e_{j^c}^\top \right) D_m u\left(t,m,[m]\right) \\ &= \sum_i b(i) \left(\sum_j a(j) e_j^\top - \sum_{j : i \in j} a(j) \mathbf{1}^\top \right) D_m u\left(t,m,[m]\right) \\ &= \sum_j a(j) \sum_{i \in j} u_i(t,m) - \sum_i b(i) \sum_{j : i \in j} a(j) = \sum_i (u_i(t,m) - b(i)) \sum_{j : i \in j} a(j). \end{aligned} \tag{11}$$

Thus, in order to have the equality (10), the coefficients of the $\sqrt{T}$ term must be zero, i.e.,

$$\begin{aligned} 0 &= \inf_{b\in\mathcal{P}([K])} \sup_{a\in\mathcal{P}(\{0,1\}^K)} \sum_i b(i) \left(\hat{a}(-i)\mathcal{V}_{a,-i} - \hat{a}(i)\mathcal{V}_{a,i}\right)^\top D_m u\left(t,m,[m]\right) \\ &= \inf_{b\in\mathcal{P}([K])} \sup_{a\in\mathcal{P}(\{0,1\}^K)} \sum_i (u_i(t,m) - b(i)) \sum_{j:i\in j} a(j). \end{aligned}$$

Otherwise, the first order term explodes. Therefore the forecaster is forced to choose the strategy $b = D_m u\left(t,m,[m]\right)$, and we obtain the PDE

$$\begin{aligned} 0 = {} & \partial_t u(t,m) + \sup_{a\in\mathcal{P}(\{0,1\}^K)} \sum_i \\ & + \frac{1}{2} u_i(t,m)\hat{a}(i) \left(\mathcal{V}_{a,i}^\top D^2_{mm} u\left(t,m,[m],[m]\right) \mathcal{V}_{a,i} + \sum_{j:i\in j} \frac{a(j)}{\hat{a}(i)} e_{j^c}^\top D_x D_m u\left(t,m,[m]\right) e_{j^c} \right) \\ & + \frac{1}{2} u_i(t,m)\hat{a}(-i) \left(\mathcal{V}_{a,-i}^\top D^2_{mm} u\left(t,m,[m],[m]\right) \mathcal{V}_{a,-i} + \sum_{j:i\notin j} \frac{a(j)}{\hat{a}(-i)} e_{j}^\top D_x D_m u\left(t,m,[m]\right) e_{j} \right). \end{aligned} \tag{12}$$

Remark 9 *(i) We say $a \in \mathcal{P}(\{0,1\}^K)$ is a balanced strategy if $\sum_{j:i\in j} a(j)$ is independent of i, and denote by $\mathcal{E}$ the set of all balanced strategies. According to* (11)*, if we restrict a in* (10) *to be balanced, the first order term vanishes for any $b \in \mathcal{P}([K])$.*

(ii)The standard tool to show the convergence of u^T to the solution of (12) *is to use the stability and comparison of viscosity solutions, see for example Drenska and Kohn (2023); Barles and Souganidis (1991) in the finite dimensional cases. However, a comparison result for viscosity solution of PDEs on the Wasserstein space is not available in the literature in the generality we need. The viscosity theory of first order PDEs on the Wasserstein space has been studied in Burzoni et al. (2020); Cosso et al. (2021); Mete Soner and Yan (2022), and second order PDEs on Wasserstein space is more challenging due to the lack of Ishii's lemma. Bandini et al. (2019) studies a second order PDE associated with a stochastic filtering problem, and by lifting the equation to a Hilbert space they obtained the well-posedness. However, the relation between the lifted PDE and the original one is unclear. In Cox et al. (2021), the authors proved the uniqueness of a second order PDE associated with a control problem under a very specific definition of viscosity solution which might not enjoy stability results needed to for the convergence problems we aim to study. Our second order PDE* (12) *is nonlinear, degenerate, and is different from the PDEs that appeared in Bandini et al. (2019); Cox et al. (2021).*

(iii) Because the second derivative terms $D^2_{mm}u$ and $D_x D_m u$ are expected to explode as $t \to 1$, the generator of (12) *is expected to become discontinuous as $t \to 1$. Thus, it is more*

convenient to use the equation

$$\begin{aligned}0 = \partial_t u(t,m) + &\sup_{i,a\in\mathcal{P}(\{0,1\}^K)} \qquad (13)\\ &+ \frac{1}{2}\hat{a}(i)\left(\mathcal{V}_{a,i}^\top D^2_{mm}u\,(t,m,[m],[m])\,\mathcal{V}_{a,i} + \sum_{j:i\in j}\frac{a(j)}{\hat{a}(i)}e_{j^c}^\top D_x D_m u\,(t,m,[m])\,e_{j^c}\right)\\ &+ \frac{1}{2}\hat{a}(-i)\left(\mathcal{V}_{a,-i}^\top D^2_{mm}u\,(t,m,[m],[m])\,\mathcal{V}_{a,-i} + \sum_{j:i\notin j}\frac{a(j)}{\hat{a}(-i)}e_j^\top D_x D_m u\,(t,m,[m])\,e_j\right)\end{aligned}$$

to obtain regret bounds. Indeed, any supersolution of (13) *is clearly a supersolution of* (12) *and the generator of* (13) *is Lipschitz continuous on the derivatives of* u*. Thus, one can expect a simpler proof of comparison of viscosity solutions.*

4. Upper bound by smooth supersolution of the PDE

In this part, we design robust strategies of the forecaster using smooth supersolutions of (12). Note that (12) becomes simpler if $D^2_{mm}u = 0$. This is the case if u is linear in m. The following Lemma uses this idea to generate simple supersolutions to (12).

Lemma 10 *Suppose* ϕ *is a classical solution of*

$$\begin{aligned}&0 \geq \partial_t\phi(t,x) + \frac{1}{2}\sup_{i,a\in\mathcal{P}(\{0,1\}^K)} Tr\left(D^2_{xx}\phi\,(t,x)\left(\sum_j a(j)\left(\mathbf{1}_{i\in j}e_{j^c}e_{j^c}^\top + \mathbf{1}_{i\notin j}e_j e_j^\top\right)\right)\right) \qquad (14)\\ &\phi(1,x) \geq \max_i x^i, \;\; \phi(t, x+\lambda\mathbf{1}) = \phi(t,x) + \lambda.\end{aligned}$$

Then, the function $\Phi : [0,1]\times\mathcal{P}_2(\mathbb{R}^K) \mapsto \mathbb{R}$ *defined by*

$$\Phi(t,m) = \phi(t,[m]) := \int \phi(t,x)\,m(dx)$$

is a smooth supersolution to (12) *with*

$$D_m\Phi\,(t,m,x) = D_x\phi(t,x),\;\; D_xD_m\Phi\,(t,m,x) = D^2_{xx}\phi(t,x),\;\; D^2_{mm}\Phi\,(t,m,x,y) = 0. \qquad (15)$$

Proof Using (15) which can be easily verified, together with the supersolution property of ϕ we have that

$$0 \geq \partial_t \phi(t,[m]) + \frac{1}{2}\int \sup_{i,a\in\mathcal{P}(\{0,1\}^K)} Tr\left(D^2_{xx}\phi\,(t,x)\left(\sum_j a(j)\left(\mathbf{1}_{i\in j}e_{j^c}e_{j^c}^\top + \mathbf{1}_{i\notin j}e_je_j^\top\right)\right)\right)dm(x)$$

$$\geq \partial_t \phi(t,[m]) + \frac{1}{2}\sup_{i,a\in\mathcal{P}(\{0,1\}^K)} Tr\left(D^2_{xx}\phi\,(t,[m])\left(\sum_j a(j)\left(\mathbf{1}_{i\in j}e_{j^c}e_{j^c}^\top + \mathbf{1}_{i\notin j}e_je_j^\top\right)\right)\right)$$

$$\geq \partial_t \Phi(t,[m]) + \frac{1}{2}\sup_{i,a\in\mathcal{P}(\{0,1\}^K)} Tr\left(D_xD_m\Phi\,(t,m,[m])\left(\sum_j a(j)\left(\mathbf{1}_{i\in j}e_{j^c}e_{j^c}^\top + \mathbf{1}_{i\notin j}e_je_j^\top\right)\right)\right)$$

$$\geq \partial_t \Phi(t,[m]) + \frac{1}{2}\sup_{a\in\mathcal{P}(\{0,1\}^K)}\sum_i$$

$$\Phi_i(t,m)\,Tr\left(D_xD_m\Phi\,(t,m,[m])\left(\sum_j a(j)\left(\mathbf{1}_{i\in j}e_{j^c}e_{j^c}^\top + \mathbf{1}_{i\notin j}e_je_j^\top\right)\right)\right),$$

where $\Phi_i(t,m)$ denotes the i-th coordinate of $D_m\Phi(t,m,[m])$. This proves the supersolution property we want. ■

Remark 11 *It can be easily verified that smooth supersolutions of*

$$0 = \partial_t \phi(t,x) + \frac{1}{2}\sup_{a\in\mathcal{P}(\{0,1\}^K)}\sum_i$$

$$\partial_{x^i}\phi(t,x)\,Tr\left(D^2_{xx}\phi\,(t,x)\left(\sum_j a(j)\left(\mathbf{1}_{i\in j}e_{j^c}e_{j^c}^\top + \mathbf{1}_{i\notin j}e_je_j^\top\right)\right)\right)$$

cannot generate supersolutions of (12) *simply by integrating x over m. Note that equation* (14) *is degenerate, and it is not clear whether a classical solutions exists. We will construct a smooth supersolution using heat potential in Example 1.*

We now show how we can use the Lemma 10 to obtain regret bounds. Fix a large time horizon T. Denote $\widetilde{m} := m^{*\frac{1}{\sqrt{T}}}$, $t_n = \frac{n}{T}$, where n denotes the current step. For any smooth supersolotuion ϕ of (14), we define a strategy of the forecaster

$$(\beta_0^*, \cdots, \beta_{T-1}^*)$$

via

$$\beta_n^*(m) := D_m\Phi\left(t_n, \tilde{m}, [\widetilde{m}]\right). \tag{16}$$

Suppose that the initial belief is m_0, and denote random belief as $(m_n)_{n=1,\ldots,T}$. Then it is clear that

$$v_T(m_0) \leq \sup_\alpha \mathbb{E}^{\beta^*,\alpha}\left[f([m_T])\right],$$

where f is the terminal condition $f(x) := \max_i x^i$. The following Proposition provides assumptions for such a methodology to yield to regret bounds.

Proposition 12 *Suppose ϕ is a classical solution of* (14) *and*

$$|\partial^2_{tt}\phi(t,x)| \le \frac{C}{(1-t)^{3/2}}, \quad |\partial^3_{xxx}\phi(t,x)| + |\partial^2_{tx}\phi(t,x)| \le \frac{C}{1-t}, \quad \forall x \in \mathbb{R}^K, \tag{17}$$

for some positive constant C. Then the strategy β^* of the forecaster defined in (16) *yields regret bounded above by $\sqrt{T}\phi(0,[\widetilde{m}_0])$ asymptotically.*

Proof Our goal is to show that $\lim\limits_{T\to\infty} \frac{1}{\sqrt{T}} \sup_\alpha \mathbb{E}^{\beta^*,\alpha}\left[f([m_T])\right] - \phi\left(0,[\widetilde{m}_0]\right) \le 0$. First we rewrite the difference as a telescopic sum

$$\begin{aligned}&\frac{1}{\sqrt{T}} \sup_\alpha \mathbb{E}^{\beta^*,\alpha}\left[f([m_T])\right] - \phi\left(0,[\widetilde{m}_0]\right) = \sup_\alpha \mathbb{E}^{\beta^*,\alpha}\left[f([\widetilde{m}_T])\right] - \phi\left(0,[\widetilde{m}_0]\right)\\ &= \sup_\alpha \sum_{n=0}^{T-1} \left(\mathbb{E}^{\beta^*,\alpha}\left[\phi(t_{n+1},[\widetilde{m}_{n+1}])\right] - \mathbb{E}^{\beta^*,\alpha}\left[\phi(t_n,[\widetilde{m}_n])\right]\right).\end{aligned}$$

Conditioning on $\widetilde{m}_n = m$, we have that

$$\begin{aligned}&\mathbb{E}^{\beta^*,a}\left[\phi(t_{n+1},[\widetilde{m}_{n+1}]) - \phi(t_n,[\widetilde{m}_n]) \,|\, \widetilde{m}_n = m\right] \\ &= \sum_i \beta^*_n(m_n)(i)\left(\hat{a}(i)\phi\left(t_{n+1},\left[A^{a,m}_{i,\sqrt{T}}\right]\right) + \hat{a}(-i)\phi\left(t_{n+1},\left[A^{a,m}_{-i,\sqrt{T}}\right]\right)\right) - \phi(t_n,[m]).\end{aligned}\tag{18}$$

Using the linear structure of $\phi(t,[m])$, it can be seen that

$$\begin{aligned}&\phi\left(t_{n+1},\left[A^{a,m}_{i,\sqrt{T}}\right]\right) - \phi(t_n,[m])\\ &= \sum_{j:i\in j} \frac{a(j)}{\hat{a}(i)} \int \left(\phi\left(t_{n+1}, x - \frac{e_{j^c}}{\sqrt{T}}\right) - \phi(t_n,x)\right) m(dx).\end{aligned}\tag{19}$$

For any $i \in j \subset [K]$, we have the equality

$$\begin{aligned}&\phi\left(t_{n+1}, x - \frac{e_{j^c}}{\sqrt{T}}\right) - \phi(t_n,x)\\ &= -\frac{e_{j^c}}{\sqrt{T}}\partial_x\phi(t_n,x) + \frac{1}{T}\left(\partial_t\phi(t_n,x) + \frac{1}{2}e^\top_{j^c}\partial^2_{xx}\phi(t_n,x)e_{j^c}\right)\\ &\quad + \frac{1}{T}\int_0^1 \partial_t\phi\left(t_n + \frac{s}{T}, x - \frac{se_{j^c}}{\sqrt{T}}\right) - \partial_t\phi(t_n,x)\,ds\\ &\quad - \frac{e_{j^c}}{\sqrt{T}}\int_0^1 \partial_x\phi\left(t_n + \frac{s}{T}, x - \frac{se_{j^c}}{\sqrt{T}}\right) - \partial_x\phi\left(t_n, x - \frac{se_{j^c}}{\sqrt{T}}\right)ds\\ &\quad + \frac{1}{T}\int_0^1 (1-s)e^\top_{j^c}\left(\partial^2_{xx}\phi\left(t_n, x - \frac{se_{j^c}}{\sqrt{T}}\right) - \partial^2_{xx}\phi(t_n,x)\right)e_{j^c}\,ds.\end{aligned}\tag{20}$$

Using our assumption (17), we can estimate the last three terms in the equation above

$$\begin{aligned}
&\left|\frac{1}{T}\int_0^1 \partial_t\phi\left(t_n+\frac{s}{T}, x-\frac{se_{j^c}}{\sqrt{T}}\right)-\partial_t\phi\left(t_n,x\right)ds\right| \le C\int_0^{1/T}\frac{\frac{1}{T}-s}{(1-t_n-s)^{3/2}}ds\\
&\left|\frac{e_{j^c}}{\sqrt{T}}\int_0^1 \partial_x\phi\left(t_n+\frac{s}{T}, x-\frac{se_{j^c}}{\sqrt{T}}\right)-\partial_x\phi\left(t_n, x-\frac{se_{j^c}}{\sqrt{T}}\right)ds\right| \le C\sqrt{T}\int_0^{1/T}\frac{\frac{1}{T}-s}{1-t_n-s}ds\\
&\left|\frac{1}{T}\int_0^1 (1-s)e_{j^c}^\top\left(\partial_{xx}^2\phi\left(t_n, x-\frac{se_{j^c}}{\sqrt{T}}\right)-\partial^2_{xx}\phi(t_n,x)\right)e_{j^c}\,ds\right| \le \frac{C}{T^{3/2}(1-t_n)}.
\end{aligned}$$

Let us define

$$O(T,n):=C\left(\int_0^{1/T}\frac{\frac{1}{T}-s}{(1-t_n-s)^{3/2}}ds+\sqrt{T}\int_0^{1/T}\frac{\frac{1}{T}-s}{1-t_n-s}ds+\frac{1}{T^{3/2}(1-t_n)}\right). \tag{21}$$

Now plugging (19) and (20) into (18), we obtain that

$$\begin{aligned}
&\mathbb{E}^{\beta^*,a}\left[\phi(t_{n+1},[\widetilde{m}_{n+1}])-\phi(t_n,[\widetilde{m}_n])\,|\,\widetilde{m}_n=m\right]\\
&\le\frac{1}{\sqrt{T}}\sum_i\beta_n^*(m_n)(i)\left(\sum_{j,i\notin j}a(j)e_j^\top-\sum_{j:i\in j}a(j)e_{j^c}^\top\right)\partial_x\phi(t_n,[m])+\frac{1}{T}\times\\
&\left(\partial_t\phi(t_n,[m])+\frac{1}{2}\sum_i\beta_n^*(m_n)(i)Tr\left(D^2_{xx}\phi\left(t_n,[m]\right)\left(\sum_j a(j)\left(\mathbf{1}_{i\in j}e_{j^c}e_{j^c}^\top+\mathbf{1}_{i\notin j}e_je_j^\top\right)\right)\right)\right)\\
&+O(T,n).
\end{aligned}\tag{22}$$

The first term on the right hand side vanishes due to our choice of β^*, the second term is non-positive due to the supersolution property of ϕ, and thus we obtain that

$$\mathbb{E}^{\beta^*,a}\left[\phi(t_{n+1},[\widetilde{m}_{n+1}])-\phi(t_n,[\widetilde{m}_n])\,|\,\widetilde{m}_n=m\right]\le O(T,n).$$

Summing up from $n=0$ to $T-1$, taking supremum over $\alpha\in\mathcal{A}$, and letting $T\to\infty$, we conclude that

$$\lim_{T\to\infty}\frac{1}{\sqrt{T}}\sup_\alpha\mathbb{E}^{\beta^*,\alpha}\left[f([m_T])\right]-\phi\left(0,[\widetilde{m}_0]\right)\le\lim_{T\to\infty}\sum_{n=0}^{T-1}O(T,n)=0.$$

■

Remark 13 *For any fixed terminal T, suppose α^*,β^* are optimal strategies of the adversary and forecaster that yield the value function $v_T(t,m)$. If v_T is regular in m, we denote by $\partial_{x^j}v_T(t,m)$ the j-th coordinate of $D_mv_T(t,m,[m])$. Then according to the definition of*

Wasserstein derivative, we have

$$\begin{aligned}
\partial_{x^j} v_T(t,m) &= \lim_{\epsilon\to 0} \frac{v_T(t,(id+\epsilon e_j)_\# m) - v_T(t,m)}{\epsilon} \\
&= \lim_{\epsilon\to 0} \frac{\mathbb{E}^{\alpha^*,\beta^*}[\max_i X_T^i \,|\, X_t \sim (id+\epsilon e_j)_\# m] - \mathbb{E}^{\alpha^*,\beta^*}[\max_i X_T^i \,|\, X_t \sim m]}{\epsilon} \\
&\approx \lim_{\epsilon\to 0} \frac{\mathbb{E}^{\alpha^*,\beta^*}[\max_i (X_T^i + \epsilon \mathbf{1}_{i=j}) \,|\, X_t \sim m] - \mathbb{E}^{\alpha^*,\beta^*}[\max_i X_T^i \,|\, X_t \sim m]}{\epsilon} \\
&\approx \mathbb{E}^{\alpha^*,\beta^*}\left[\mathbf{1}_{\{X_T^j \geq X_T^i,\, i=1,\dots,K\}} \,|\, X_t \sim m\right],
\end{aligned}$$

which is approximately the probability that the j-th action finishes as the optimal constant strategy. Therefore, the strategy (16) *can be understood as the probability matching algorithm in the limit.*

In the following example, we provide a smooth supersolution using heat potentials. One may also construct supersolutions using other potentials as in Kobzar et al. (2020).

Example 1 *Let us take ϕ to be the smooth solution of the following heat equation*

$$\begin{cases} \partial_t \phi + \frac{1}{2}\Delta\phi = 0 & on\ \mathbb{R}^K \times [0,1) \\ \phi(1,x) = f(x) & on\ \mathbb{R}^K \times \{1\}\ . \end{cases}$$

It can be easily verified as in (Bayraktar et al., 2021a, Proposition 19) that ϕ satisfies (17)*. According to (Kobzar et al., 2020, Appendix F.1), we know that $D^2_{x^l x^k}\phi(t,x) > 0$ if $l = k$ and $D^2_{x^l x^k}\phi(t,x) < 0$ if $l \neq k$. Therefore for any $i \in [K]$ and $j \subset [K]$, we have that*

$$\frac{1}{2} Tr\left(D^2_{xx}\phi(t,x)\left(\mathbf{1}_{i\in j} e_{j^c} e_{j^c}^\top + \mathbf{1}_{i\notin j} e_j e_j^\top\right)\right) \leq \frac{1}{2}\Delta\phi(t,x),$$

and hence

$$\frac{1}{2}\sup_{i,a\in\mathcal{P}(\{0,1\}^K)} Tr\left(D^2_{xx}\phi\,(t,x)\left(\sum_j a(j)\left(\mathbf{1}_{i\in j} e_{j^c} e_{j^c}^\top + \mathbf{1}_{i\notin j} e_j e_j^\top\right)\right)\right) \leq \frac{1}{2}\Delta\phi(t,x).$$

Thus ϕ is a smooth supersolution of (14) *which satisfies* (17) *according to Kobzar et al. (2020). By Feynman-Kac formula, we have $\phi(0,x) = \mathbb{E}_x[f(N^1,N^2,\dots,N^K)]$ where N^i is a standard normal. Supposing $x = (0,\dots,0)$, then by Jensen's inequality we have that for any $t \geq 0$*

$$e^{t\mathbb{E}[f(N^1,\dots,N^K)]} \leq \mathbb{E}[e^{tf(N^1,\dots,N^K)}] \leq K\mathbb{E}[e^{tN^1}] = Ke^{t^2/2},$$

and hence $\mathbb{E}[f(N^1,\dots,N^K)] \leq \frac{\log K}{t} + \frac{t}{2}$. Choosing $t = \sqrt{2\log K}$, we obtain that $\phi(0,0) \leq \sqrt{2\log K}$. Therefore, when initial belief is δ_0, in our game where both agents have partial information, the asymptotic regret is bounded above by $\sqrt{2T\log K}$. It is smaller than the expected regret $5.15\sqrt{TK\log K} + \sqrt{\frac{TK}{\log K}}$ in the case of adversarial bandit where both agents only observe Y_t (Bubeck and Cesa-Bianchi, 2012, Theorem 3.4). The regret bound we obtain is two times larger than the performance of multiplicative weight algorithms obtained in Gravin et al. (2017).

Remark 14 *Our main contribution in terms of regret bound is to extend the PDE based methodology of Kobzar et al. (2020) to the version of bandit problems we study. In Lemma 10, this bound is obtained by considering a functional linear in m in the sense that $\Phi(t,m) = \int \phi(t,x)m(dx)$. Similar to Kobzar et al. (2020), the PDE tools are expected to yield sharper bounds by considering more sophisticated supersolutions to* (14).

For example, any solution of

$$\begin{aligned} 0 = \partial_t u(t,m) + &\sup_{i,a\in\mathcal{P}(\{0,1\}^K)} \\ &+ \frac{1}{2}\left(\mathcal{V}_{a,i}^\top D^2_{mm}u\,(t,m,[m],[m])\,\mathcal{V}_{a,i} + \sum_{j:i\in j}\frac{a(j)}{\hat{a}(i)}e_{j^c}^\top D_x D_m u\,(t,m,[m])\,e_{j^c}\right) \\ &+ \frac{1}{2}\left(\mathcal{V}_{a,-i}^\top D^2_{mm}u\,(t,m,[m],[m])\,\mathcal{V}_{a,-i} + \sum_{j:i\notin j}\frac{a(j)}{\hat{a}(-i)}e_{j}^\top D_x D_m u\,(t,m,[m])\,e_{j}\right) \end{aligned} \tag{23}$$

is a supersolution of (14)*. For all $i \in [K]$ and $a \in \mathcal{P}(\{0,1\}^K)$, we can define the symmetric matrices*

$$\begin{aligned} \Sigma(i,a) &= \Big(a(e_i)\mathcal{V}_{a,e_i}\mathcal{V}_{a,e_i}^\top + a(-e_i)\mathcal{V}_{a,-e_i}\mathcal{V}_{a,-e_i}^\top\Big) \\ \tilde{\Sigma}(i,a) &= \left(a(e_i)\sum_{j:i\in j}\frac{a(j)}{a(e_i)}e_{j^c}^\top e_{j^c} + a(-e_i)\sum_{j:i\notin j}\frac{a(j)}{a(-e_i)}e_j^\top e_j\right) - \Sigma(i,a). \end{aligned}$$

By computing $v^\top\Sigma(i,a)v$ and $v^\top\tilde{\Sigma}(i,a)v$ for $v \in \mathbb{R}^K$, one can show that these matrices are non-negative. Thus, (23) *can be written as the Hamilton-Jacobi-Bellman equation*

$$0 = \partial_t u(t,m) + \frac{1}{2}\sup_{i,a\in\mathcal{P}(\{0,1\}^K)} Tr\left(\mathcal{H}u\,(t,m)\,\Sigma\,(i,a) + D_x D_m u\,(t,m,[m])\,\tilde{\Sigma}\,(i,a)\right) \tag{24}$$

where in line with Chow and Gangbo (2019), the term

$$\mathcal{H}U(t,m) := \int D_x D_m U(t,m,x)m(dx) + \int\int D^2_{mm}U(t,m,x,y)m(dx)m(dy)$$

is the so-called the Wasserstien Hessian of $U(t,\cdot)$. A simple computation shows that the value function corresponding to a controlled version of (Chow and Gangbo, 2019, Equation (1.8)) would yield to a viscosity solution to (24)*; see (Chow and Gangbo, 2019, Remark 3.5). Then, this value function can be used as a supersolution of* (14) *(which would indeed depend nonlinearly on m). However such a methodology requires a comparison result for viscosity solutions of* (24) *(or smoothness of the value function) to obtain regret bounds. This comparison result has been obtained in Bayraktar et al. (2023a,b). The convergence of discrete time value functions and computation of improved regret bounds are being addressed by the authors in an ongoing project.*

5. Lower bound by smooth subsolution of the PDE

As in the last section, we construct strategies for the adversary using smooth subsolutions of (12). Recall that $\mathcal{E}$ is the set of balanced strategies defined in Remark 9. The proof of following lemma is almost the same as Lemma 10 and thus we omit it.

Lemma 15 *Let ϕ be a smooth solution of*

$$0 \leq \partial_t \phi(t,x) + \frac{1}{2} \inf_i Tr\left(D^2_{xx}\phi\left(t,x\right)\left(\sum_j a_t(j)\left(\mathbf{1}_{i\in j} e_{j^c} e_{j^c}^\top + \mathbf{1}_{i\notin j} e_j e_j^\top\right)\right)\right)$$
$$\phi(1,x) \leq \max_i x^i,\ \ \phi(t, x+\lambda\mathbf{1}) = \phi(t,x)+\lambda,$$

where $a_t \in \mathcal{E}$, $t\in[0,1], m \in \mathcal{P}(\mathbb{R}^K)$ are balanced strategies. Then, the function $\Phi : [0,1]\times \mathcal{P}_2(\mathbb{R}^K) \mapsto \mathbb{R}$ defined by

$$\Phi(t,m) = \phi(t,[m]) = \int \phi(t,x)\, m(dx)$$

is a smooth subsolution to (12)*.*

Remark 16 *Note that in Lemma 15, the choice of balanced strategies a_t only depends on time t.*

Given balanced strategies $(a_t)_{t\in[0,1]}$ and subsolution ϕ as in Lemma 15, we construct strategies for the adversary in the original game (2). For a large time horizon T. Let us denote $t_n = \frac{n}{T}$, where n is the current step. We define a strategy α^* of the adversary via

$$\alpha^*_n = a_{t_n}, \quad n = 0,\dots,T-1.$$

Proposition 17 *Suppose $(a_t)_{t\in[0,1]}$, ϕ are balanced strategies and classical solutions as in Lemma 15 that satisfies* (17)*. Let m_0 be the initial belief. Then the strategy α^* of the adversary defined yields regret bounded below by $\sqrt{T}\phi\left(0,\left[m_0^{*\frac{1}{\sqrt{T}}}\right]\right)$ asymptotically.*

Proof The argument is almost the same as that of Proposition 12. Just notice that (22) now becomes

$$\begin{aligned}&\mathbb{E}^{b,\alpha^*}\left[\phi(t_{n+1},[\widetilde{m}_{n+1}]) - \phi(t_n,[\widetilde{m}_n])\,|\,\widetilde{m}_n = m\right]\\ &\geq \frac{1}{\sqrt{T}}\sum_i b(i)\left(\sum_{j:i\notin j}\alpha^*_n(j)e_j^\top - \sum_{j:i\in j}\alpha^*_n(j)e_{j^c}^\top\right)\partial_x\phi(t_n,[m]) + \frac{1}{T}\times\\ &\left(\partial_t\phi(t_n,[m]) + \frac{1}{2}\sum_i b(i) Tr\left(D^2_{xx}\phi\left(t_n,[m]\right)\left(\sum_j \alpha^*_n(j)\left(\mathbf{1}_{i\in j}e_{j^c}e_{j^c}^\top + \mathbf{1}_{i\notin j}e_je_j^\top\right)\right)\right)\right)\\ &+ O(T,n),\end{aligned}$$

where $O(T,n)$ is defined in (21). The first order term on the right hand vanishes since α_n^* is balanced, and second order term is nonnegative due to the subsolution property of ϕ. Thus we obtain that $\mathbb{E}^{b,\alpha^*}\left[\phi(t_{n+1},[\widetilde{m}_{n+1}]) - \phi(t_n,[\widetilde{m}_n]) \,|\, \widetilde{m}_n = m\right] \geq O(T,n)$. Then summing up from $n=0$ to $T-1$, taking infimum over $\beta \in \mathcal{B}$, and letting $T \to \infty$, we conclude our result.

■

Example 2 *Let us take a_t to be the uniformly distribution over $\{0,1\}^K$ for each $t \in [0,1]$. Then it can be easily verified that*

$$\sum_j a_t(j)\left(\mathbf{1}_{i\in j} e_{j^c} e_{j^c}^\top + \mathbf{1}_{i\notin j} e_j e_j^\top\right) = \frac{1}{4}\left(ee^\top + I_K - e_i e^\top - ee_i^\top\right), \quad \forall\, i \in [K].$$

where I_K stands for the identity matrix of dimension $K \times K$.

Let us take ϕ to be the smooth solution of the following heat equation

$$\begin{cases} \partial_t \phi + \frac{1}{8}\Delta\phi = 0 & \text{on } \mathbb{R}^K \times [0,1) \\ \phi(1,x) = f(x) & \text{on } \mathbb{R}^K \times \{1\}\ . \end{cases}$$

Thanks to the translation invariance property of ϕ, i.e., $\phi(t, x + c\mathbf{1}) = \phi(t,x) + c$, we have $D^2_{xx}\phi(t,x)\,e = 0$ for all $t \in [0,1)$. Then, it can be easily seen that such ϕ and $(a_t)_{t\in[0,1]}$ satisfy all the assumptions in Proposition 17. Therefore when initial belief is δ_0, the asymptotic asymptotic regret is bounded below by $\sqrt{T}\phi(0,0)$. By Feynman-Kac formula, $\phi(0,0) = \mathbb{E}[f(N^1,\dots,N^K)] = \mathbb{E}[\max_i N^i]$ where N^i is independently gaussian distributed with mean 0 and variance $1/4$ for each $i = 1,\dots,K$. Then according to (Orabona and Pal, 2015, Theorem 3), we obtain a lower bound $\phi(0,0) = \mathbb{E}[\max_i N^i] \geq 0.065\sqrt{\log K} - 0.35$.

Acknowledgments

E. Bayraktar is partially supported by the NSF grant DMS-2106556 and by the Susan M. Smith chair. I. Ekren is supported in part by the NSF Grant DMS-2007826

References

Shipra Agrawal and Navin Goyal. Further optimal regret bounds for thompson sampling. In Carlos M. Carvalho and Pradeep Ravikumar, editors, *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, volume 31, pages 99–107, 2013.

Jean-Yves Audibert and Sébastien Bubeck. Regret bounds and minimax policies under partial monitoring. *The Journal of Machine Learning Research*, 11:2785–2836, 2010.

Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Minimax policies for combinatorial prediction games. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 107–132. JMLR Workshop and Conference Proceedings, 2011.

Elena Bandini, Andrea Cosso, Marco Fuhrman, and Huyên Pham. Randomized filtering and Bellman equation in Wasserstein space for partial observation control problem. *Stochastic Process. Appl.*, 129(2):674–711, 2019.

Guy Barles and Panagiotis E Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic analysis*, 4(3):271–283, 1991.

Erhan Bayraktar, Ibrahim Ekren, and Xin Zhang. Finite-time 4-expert prediction problem. *Comm. Partial Differential Equations*, 45(7):714–757, 2020a.

Erhan Bayraktar, Ibrahim Ekren, and Yili Zhang. On the asymptotic optimality of the comb strategy for prediction with expert advice. *Ann. Appl. Probab.*, 30(6):2517–2546, 2020b.

Erhan Bayraktar, Ibrahim Ekren, and Xin Zhang. Prediction against a limited adversary. *J. Mach. Learn. Res.*, 22(1), 2021a.

Erhan Bayraktar, H. Vincent Poor, and Xin Zhang. Malicious experts versus the multiplicative weights algorithm in online prediction. *IEEE Transactions on Information Theory*, 67(1):559–565, 2021b.

Erhan Bayraktar, Ibrahim Ekren, and Xin Zhang. Comparison of viscosity solutions for a class of second order PDEs on the Wasserstein space. *arXiv:2309.05040*, 2023a.

Erhan Bayraktar, Ibrahim Ekren, and Xin Zhang. A smooth variational principle on Wasserstein space. *Proc. Amer. Math. Soc.*, 151(9):4089–4098, 2023b.

Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Matteo Burzoni, Vincenzo Ignazio, A Max Reppen, and H Mete Soner. Viscosity solutions for controlled mckean–vlasov jump-diffusions. *SIAM Journal on Control and Optimization*, 58(3):1676–1699, 2020.

Jeff Calder and Nadejda Drenska. Asymptotically optimal strategies for online prediction with history-dependent experts. *Journal of Fourier Analysis and Applications*, (27):1–20, August 2021.

Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. *The Master Equation and the Convergence Problem in Mean Field Games:(AMS-201)*, volume 201. Princeton University Press, 2019.

René Carmona, François Delarue, et al. *Probabilistic theory of mean field games with applications I-II.* Springer, 2018.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games.* Cambridge university press, 2006.

Nicolò Cesa-Bianchi, Claudio Gentile, and Giovanni Zappella. A gang of bandits. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26, 2013.

Yat Tin Chow and Wilfrid Gangbo. A partial laplacian as an infinitesimal generator on the wasserstein space. *Journal of Differential Equations*, 267(10):6065–6117, 2019.

Andrea Cosso, Fausto Gozzi, Idris Kharroubi, Huyên Pham, and Mauro Rosestolato. Master bellman equation in the wasserstein space: Uniqueness of viscosity solutions. *arXiv:2107.10535*, 2021.

Thomas M Cover. Behavior of sequential predictors of binary sequences. Technical report, STANFORD UNIV CALIF STANFORD ELECTRONICS LABS, 1966.

Alexander M. G. Cox, Sigrid Källblad, Martin Larsson, and Sara Svaluto-Ferro. Controlled Measure-Valued Martingales: a Viscosity Solution Approach. *arXiv:2109.00064*, 2021.

Nadejda Drenska and Jeff Calder. Online Prediction With History-Dependent Experts: The General Case. *To appear in Communications on Pure and Applied Mathematics*, 2021.

Nadejda Drenska and Robert V. Kohn. Prediction with expert advice: a PDE perspective. *J. Nonlinear Sci.*, 30(1):137–173, 2020.

Nadejda Drenska and Robert V. Kohn. A PDE approach to the prediction of a binary sequence with advice from two history-dependent experts. *Comm. Pure Appl. Math.*, 76 (4):843–897, 2023.

Nick Gravin, Yuval Peres, and Balasubramanian Sivan. Towards optimal algorithms for prediction with expert advice. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 528–547, USA, 2016. Society for Industrial and Applied Mathematics.

Nick Gravin, Yuval Peres, and Balasubramanian Sivan. Tight lower bounds for multiplicative weights algorithmic families. In *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

Laura Greenstreet, Nicholas J. A. Harvey, and Victor Sanches Portella. Efficient and optimal fixed-time regret with two experts. In Sanjoy Dasgupta and Nika Haghtalab, editors, *Proceedings of The 33rd International Conference on Algorithmic Learning Theory*, volume 167 of *Proceedings of Machine Learning Research*, pages 436–464. PMLR, 29 Mar–01 Apr 2022.

Nicholas J. A. Harvey, Christopher Liaw, Edwin A. Perkins, and Sikander Randhawa. Optimal anytime regret for two experts. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1404–1415, 2020.

Mark Herbster, Stephen Pasteris, Fabio Vitale, and Massimiliano Pontil. A gang of adversarial bandits. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman

Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 2265–2279, 2021.

Vladimir A. Kobzar and Robert V. Kohn. A PDE-Based Analysis of the Symmetric Two-Armed Bernoulli Bandit. *arXiv:2202.05767*, 2022.

Vladimir A. Kobzar, Robert V. Kohn, and Zhilei Wang. New potential-based bounds for prediction with expert advice. volume 125 of *Proceedings of Machine Learning Research*, pages 2370–2405. PMLR, 09–12 Jul 2020.

H. Mete Soner and Qinxin Yan. Viscosity Solutions for McKean-Vlasov Control on a torus. *arXiv:2212.11053*, 2022.

Francesco Orabona and David Pal. Optimal Non-Asymptotic Lower Bound on the Minimax Regret of Learning with Expert Advice. *arXiv:1511.02176*, 2015.

Zhiyu Zhang, Ashok Cutkosky, and Ioannis Paschalidis. PDE-Based Optimal Strategy for Unconstrained Online Learning. *arXiv:2201.07877*, 2022.