

---

# Efficient Minimax Strategies for Square Loss Games

---

**Wouter M. Koolen**

Queensland University of Technology and UC Berkeley  
wouter.koolen@qut.edu.au

**Alan Malek**

University of California, Berkeley  
malek@eecs.berkeley.edu

**Peter L. Bartlett**

University of California, Berkeley and Queensland University of Technology  
peter@berkeley.edu

## Abstract

We consider online prediction problems where the loss between the prediction and the outcome is measured by the squared Euclidean distance and its generalization, the squared Mahalanobis distance. We derive the minimax solutions for the case where the prediction and action spaces are the simplex (this setup is sometimes called the Brier game) and the  $\ell_2$  ball (this setup is related to Gaussian density estimation). We show that in both cases the value of each sub-game is a quadratic function of a simple statistic of the state, with coefficients that can be efficiently computed using an explicit recurrence relation. The resulting deterministic minimax strategy and randomized maximin strategy are linear functions of the statistic.

## 1 Introduction

We are interested in general strategies for sequential prediction and decision making (a.k.a. online learning) that improve their performance with experience. Since the early days of online learning, people have formalized such learning tasks as regret games. The learner interacts with an adversarial environment with the goal of performing almost as well as the best strategy from some fixed reference set. In many cases, we have efficient algorithms with an upper bound on the regret that meets the game-theoretic lower bound (up to a small constant factor). In a few special cases, we have the exact minimax strategy, meaning that we understand the learning problem at all levels of detail. In even fewer cases we can also efficiently execute the minimax strategy. These cases serve as exemplars to guide our thinking about learning algorithms.

In this paper we add two interesting examples to the canon of efficiently computable minimax strategies. Our setup, as described in Figure 1, is as follows. The Learner and the Adversary play vectors  $\mathbf{a} \in \mathcal{A}$  and  $\mathbf{x} \in \mathcal{X}$ , upon which the Learner is penalized using the squared Euclidean distance  $\|\mathbf{a} - \mathbf{x}\|^2$  or its generalization, the squared Mahalanobis distance,

$$\|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 = (\mathbf{a} - \mathbf{x})^\top \mathbf{W}^{-1} (\mathbf{a} - \mathbf{x}),$$

parametrized by a symmetric matrix  $\mathbf{W} \succ \mathbf{0}$ . After a sequence of  $T$  such interactions, we compare the loss of the Learner to the loss of the best fixed prediction  $\mathbf{a}^* \in \mathcal{A}$ . In all our examples, this best fixed action in hindsight is the mean outcome  $\mathbf{a}^* = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t$ , regardless of  $\mathbf{W}$ . We use regret, the difference between the loss of the learner and the loss of  $\mathbf{a}^*$ , to evaluate performance. The *minimax regret* for the  $T$ -round game, also known as the value of the game, is given by

$$V := \inf_{\mathbf{a}_1} \sup_{\mathbf{x}_1} \cdots \inf_{\mathbf{a}_T} \sup_{\mathbf{x}_T} \sum_{t=1}^T \frac{1}{2} \|\mathbf{a}_t - \mathbf{x}_t\|_{\mathbf{W}}^2 - \inf_{\mathbf{a}} \sum_{t=1}^T \frac{1}{2} \|\mathbf{a} - \mathbf{x}_t\|_{\mathbf{W}}^2 \quad (1)$$

where the  $\mathbf{a}_t$  range over actions  $\mathcal{A}$  and the  $\mathbf{x}_t$  range over outcomes  $\mathcal{X}$ . The *minimax strategy* chooses the  $\mathbf{a}_t$ , given all past outcomes  $\mathbf{x}_1, \dots, \mathbf{x}_{t-1}$ , to achieve this regret. Intuitively, the minimax regret is the regret if both players play optimally while assuming the other player is doing the same.

Our first example is the Brier game, where the action and outcome spaces are the probability simplex with  $K$  outcomes. The Brier game is traditionally popular in meteorology [Bri50].

Our second example is the ball game, where the action and outcome spaces are the Euclidean norm ball, i.e.  $\mathcal{A} = \mathcal{X} = \{\mathbf{x} \in \mathbb{R}^K \mid \|\mathbf{x}\|_2 = 1\}$ . (Even though we measure loss by the squared Mahalanobis distance, we play on the standard Euclidean norm ball.) The ball game is related to Gaussian density estimation [TW00].

In each case we exhibit a strategy that can play a  $T$ -round game in  $O(TK^2)$  time. (The algorithm spends  $O(TK + K^3)$  time pre-processing the game, and then plays in  $O(K^2)$  time per round.)

Given:  $T, \mathbf{W}, \mathcal{A}, \mathcal{X}$ .  
For  $t = 1, 2, \dots, T$

- Learner chooses prediction  $\mathbf{a}_t \in \mathcal{A}$
- Adversary chooses outcome  $\mathbf{x}_t \in \mathcal{X}$
- Learner incurs loss  $\frac{1}{2}\|\mathbf{a}_t - \mathbf{x}_t\|_{\mathbf{W}}^2$ .

Figure 1: Protocol

## 2 Outline

We define our loss using the squared Mahalanobis distance, parametrized by a symmetric matrix  $\mathbf{W} \succ \mathbf{0}$ . We recover the squared Euclidean distance by choosing  $\mathbf{W} = \mathbf{I}$ . Our games will always last  $T$  rounds. For some observed data  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , the *value-to-go* for the remaining  $T - n$  rounds is given by

$$V(\mathbf{x}_1, \dots, \mathbf{x}_n) := \inf_{\mathbf{a}_{n+1}} \sup_{\mathbf{x}_{n+1}} \cdots \inf_{\mathbf{a}_T} \sup_{\mathbf{x}_T} \sum_{t=n+1}^T \frac{1}{2} \|\mathbf{a}_t - \mathbf{x}_t\|_{\mathbf{W}}^2 - \inf_{\mathbf{a}} \sum_{t=1}^T \frac{1}{2} \|\mathbf{a} - \mathbf{x}_t\|_{\mathbf{W}}^2.$$

By definition, the minimax regret (1) is  $V = V(\epsilon)$  where  $\epsilon$  is the empty sequence, and the value-to-go satisfies the recurrence

$$V(\mathbf{x}_1, \dots, \mathbf{x}_n) = \begin{cases} -\inf_{\mathbf{a}} \sum_{t=1}^T \frac{1}{2} \|\mathbf{a} - \mathbf{x}_t\|_{\mathbf{W}}^2 & \text{if } n = T, \\ \inf_{\mathbf{a}_{n+1}} \sup_{\mathbf{x}_{n+1}} \frac{1}{2} \|\mathbf{a}_{n+1} - \mathbf{x}_{n+1}\|_{\mathbf{W}}^2 + V(\mathbf{x}_1, \dots, \mathbf{x}_{n+1}) & \text{if } n < T. \end{cases} \quad (2)$$

Our analysis for the two games proceeds in a similar manner. For some past history of plays  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  of length  $n$ , we summarize the state by  $\mathbf{s} = \sum_{t=1}^n \mathbf{x}_t$  and  $\sigma^2 = \sum_{t=1}^n \mathbf{x}_t^\top \mathbf{W}^{-1} \mathbf{x}_t$ . As we will see, the value-to-go after  $n$  of  $T$  rounds can be written as  $V(\mathbf{s}, \sigma^2, n)$ ; i.e. it only depends on the past plays through  $\mathbf{s}$  and  $\sigma^2$ . More surprisingly, for each  $n$ , the value-to-go  $V(\mathbf{s}, \sigma^2, n)$  is a quadratic function of  $\mathbf{s}$  and a linear function of  $\sigma^2$  (under certain conditions on  $\mathbf{W}$ ). While it is straightforward to see that the terminal value  $V(\mathbf{s}, \sigma^2, T)$  is quadratic in the state (this is easily checked by computing the loss of the best expert and using the first case of Equation (2)), it is not at all obvious that propagating from  $V(\mathbf{s} + \mathbf{x}, \sigma^2 + \mathbf{x}^\top \mathbf{W}^{-1} \mathbf{x}, n + 1)$  to  $V(\mathbf{s}, \sigma^2, n)$ , using the second case of (2), preserves this structure.

This compact representation of the value-function is an essential ingredient for a computationally feasible algorithm. Many minimax approaches, such as normalized maximum likelihood [Sht87], have computational complexities that scale exponentially with the time horizon. We derive a strategy that can play in constant amortized time.

Why is this interesting? We go beyond previous work in a few directions. First, we exhibit two new games that belong to the tiny class admitting computationally feasible minimax algorithms. Second, we consider the setting with squared Mahalanobis loss which allows the user intricate control over the penalization of different prediction errors. Our results clearly show how the learner should exploit this prioritization.

### 2.1 Related work

Repeated games with minimax strategies are frequently studied ([CBL06]) and, in online learning, minimax analysis has been applied to a variety of losses and repeated games; however, computa-

tionally feasible algorithms are the exception, not the rule. For example, consider log loss, first discussed in [Sht87]. While the minimax algorithm, Normalized Maximum Likelihood, is well known [CBL06], it generally requires computation that is exponential in the time horizon as one needs to aggregate over all data sequences. To our knowledge, there are two exceptions where efficient NML forecasters are possible: the multinomial case where fast Fourier transforms may be exploited [KM05], and very particular exponential families that cause NML to be a Bayesian strategy [HB12], [BGH<sup>+</sup>13]. The minimax optimal strategy is known also for: (i) the ball game with  $\mathbf{W} = \mathbf{I}$  [TW00] (our generalization to Mahalanobis  $\mathbf{W} \neq \mathbf{I}$  results in fundamentally different strategies), (ii) the ball game with  $\mathbf{W} = \mathbf{I}$  and a constraint on the player's deviation from the current empirical minimizer [ABRT08] (for which the optimal strategy is Follow-the-Leader), (iii) Lipschitz-bounded convex loss functions [ABRT08], (iv) experts with an  $L^*$  bound [AWY08], and (v) static experts with absolute loss [CBS11]. While not guaranteed to be an exhaustive list, the previous paragraph demonstrates the rarity of tractable minimax algorithms.

### 3 The Offline Problem

The regret is defined as the difference between the loss of the algorithm and the loss of the best action in hindsight. Here we calculate that action and its loss.

**Lemma 3.1.** *Suppose  $\mathcal{A} \supseteq \text{conv}(\mathcal{X})$  (this will always hold in the settings we study). For data  $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathcal{X}$ , the loss of the best action in hindsight equals*

$$\inf_{\mathbf{a} \in \mathcal{A}} \sum_{t=1}^T \frac{1}{2} \|\mathbf{a} - \mathbf{x}_t\|_{\mathbf{W}}^2 = \frac{1}{2} \left( \sum_{t=1}^T \mathbf{x}_t^\top \mathbf{W}^{-1} \mathbf{x}_t - \frac{1}{T} \left( \sum_{t=1}^T \mathbf{x}_t \right)^\top \mathbf{W}^{-1} \left( \sum_{t=1}^T \mathbf{x}_t \right) \right), \quad (3)$$

and the minimizer is the mean outcome  $\mathbf{a}^* = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t$ .

*Proof.* The unconstrained minimizer and value are obtained by equating the derivative to zero and plugging in the solution. The assumption  $\mathcal{A} \supseteq \text{conv}(\mathcal{X})$  ensures that the constraint  $\mathbf{a} \in \mathcal{A}$  is inactive.  $\square$

The best action in hindsight is curiously independent of  $\mathbf{W}$ ,  $\mathcal{A}$  and  $\mathcal{X}$ . This also shows that the follow the leader strategy that plays  $\mathbf{a}_t = \frac{1}{t-1} \sum_{s=1}^{t-1} \mathbf{x}_s$  is independent of  $\mathbf{W}$  and  $\mathcal{A}$  as well. As we shall see, the minimax strategy does not have this property.

### 4 Simplex (Brier) Game

In this section we analyze the Brier game. The action and outcome spaces are the probability simplex on  $K$  outcomes;  $\mathcal{A} = \mathcal{X} = \Delta := \{\mathbf{x} \in \mathbb{R}_+^K \mid \mathbf{1}^\top \mathbf{x} = 1\}$ . The loss is given by half the squared Mahalanobis distance,  $\frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2$ . We present a full minimax analysis of the  $T$ -round game: we calculate the game value, derive the maximin and minimax strategies, and discuss their efficient implementation.

The structure of this section is as follows. In Lemmas 4.1 and 4.2, the conclusions (value and optimizers) are obtained under the proviso that the given optimizer lies in the simplex. In our main result, Theorem 4.3, we apply these auxiliary results to our minimax analysis and argue that the maximizer indeed lies in the simplex. We immediately work from a general symmetric  $\mathbf{W} \succ \mathbf{0}$  with the following lemma.

**Lemma 4.1.** *Fix a symmetric matrix  $\mathbf{C} \succ \mathbf{0}$  and vector  $\mathbf{d}$ . The optimization problem*

$$\max_{\mathbf{p} \in \Delta} -\frac{1}{2} \mathbf{p}^\top \mathbf{C}^{-1} \mathbf{p} + \mathbf{d}^\top \mathbf{p}$$

has value  $\frac{1}{2} \left( \mathbf{d}^\top \mathbf{C} \mathbf{d} - \frac{(\mathbf{1}^\top \mathbf{C} \mathbf{d} - 1)^2}{\mathbf{1}^\top \mathbf{C} \mathbf{1}} \right) = \frac{1}{2} \left( \mathbf{d}^\top \left( \mathbf{C} - \frac{\mathbf{C} \mathbf{1} \mathbf{1}^\top \mathbf{C}}{\mathbf{1}^\top \mathbf{C} \mathbf{1}} \right) \mathbf{d} + \frac{2 \mathbf{1}^\top \mathbf{C} \mathbf{d} - 1}{\mathbf{1}^\top \mathbf{C} \mathbf{1}} \right)$  attained at optimizer

$$\mathbf{p}^* = \mathbf{C} \left( \mathbf{d} - \frac{\mathbf{1}^\top \mathbf{C} \mathbf{d} - 1}{\mathbf{1}^\top \mathbf{C} \mathbf{1}} \mathbf{1} \right) = \left( \mathbf{C} - \frac{\mathbf{C} \mathbf{1} \mathbf{1}^\top \mathbf{C}}{\mathbf{1}^\top \mathbf{C} \mathbf{1}} \right) \mathbf{d} + \frac{\mathbf{C} \mathbf{1}}{\mathbf{1}^\top \mathbf{C} \mathbf{1}}$$

provided that  $\mathbf{p}^*$  is in the simplex.

*Proof.* We solve for the optimal  $\mathbf{p}^*$ . Introducing Lagrange multiplier  $\lambda$  for the constraint  $\sum_k p_k = 1$ , we need to have  $\mathbf{p} = \mathbf{C}(\mathbf{d} - \lambda \mathbf{1})$  which results in  $\lambda = \frac{\mathbf{1}^\top \mathbf{C} \mathbf{d} - 1}{\mathbf{1}^\top \mathbf{C} \mathbf{1}}$ . Thus, the maximizer equals  $\mathbf{p}^* = \mathbf{C}(\mathbf{d} - \frac{\mathbf{1}^\top \mathbf{C} \mathbf{d} - 1}{\mathbf{1}^\top \mathbf{C} \mathbf{1}} \mathbf{1})$  which produces objective value  $\frac{1}{2}(\mathbf{d} + \frac{\mathbf{1}^\top \mathbf{C} \mathbf{d} - 1}{\mathbf{1}^\top \mathbf{C} \mathbf{1}} \mathbf{1})^\top \mathbf{C}(\mathbf{d} - \frac{\mathbf{1}^\top \mathbf{C} \mathbf{d} - 1}{\mathbf{1}^\top \mathbf{C} \mathbf{1}} \mathbf{1})$ . The statement follows from simplification.  $\square$

This lemma allows us to compute the value and saddle point whenever the future payoff is quadratic.

**Lemma 4.2.** Fix symmetric matrices  $\mathbf{W} \succeq \mathbf{0}$  and  $\mathbf{A}$  such that  $\mathbf{W}^{-1} + \mathbf{A} \succeq \mathbf{0}$ , and a vector  $\mathbf{b}$ . The optimization problem

$$\min_{\mathbf{a} \in \Delta} \max_{\mathbf{x} \in \Delta} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{x}$$

achieves its value

$$\frac{1}{2} \mathbf{c}^\top \mathbf{W} \mathbf{c} - \frac{1}{2} \frac{(\mathbf{1}^\top \mathbf{W} \mathbf{c} - 1)^2}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \quad \text{where} \quad \mathbf{c} = \frac{1}{2} \text{diag}(\mathbf{W}^{-1} + \mathbf{A}) + \mathbf{b}$$

at saddle point (the maximin strategy randomizes, playing  $\mathbf{x} = \mathbf{e}_i$  with probability  $p_i^*$ )

$$\mathbf{a}^* = \mathbf{p}^* = \left( \mathbf{W} - \frac{\mathbf{W} \mathbf{1} \mathbf{1}^\top \mathbf{W}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \right) \mathbf{c} + \frac{\mathbf{W} \mathbf{1}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}}$$

provided  $\mathbf{p}^* \succeq \mathbf{0}$ .

*Proof.* The objective is convex in  $\mathbf{x}$  for each  $\mathbf{a}$  as  $\mathbf{W}^{-1} + \mathbf{A} \succeq \mathbf{0}$ , so it is maximized at a corner  $\mathbf{x} = \mathbf{e}_k$ . We apply min-max swap (see e.g. [Sio58]), properness of the loss (which implies that  $\mathbf{a}^* = \mathbf{p}^*$ ) and expand:

$$\begin{aligned} & \min_{\mathbf{a} \in \Delta} \max_{\mathbf{x} \in \Delta} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{x} \\ &= \min_{\mathbf{a} \in \Delta} \max_k \frac{1}{2} \|\mathbf{a} - \mathbf{e}_k\|_{\mathbf{W}}^2 + \frac{1}{2} \mathbf{e}_k^\top \mathbf{A} \mathbf{e}_k + \mathbf{b}^\top \mathbf{e}_k \\ &= \max_{\mathbf{p} \in \Delta} \min_{\mathbf{a} \in \Delta} \mathbb{E}_{k \sim \mathbf{p}} \left[ \frac{1}{2} \|\mathbf{a} - \mathbf{e}_k\|_{\mathbf{W}}^2 + \frac{1}{2} \mathbf{e}_k^\top \mathbf{A} \mathbf{e}_k + \mathbf{b}^\top \mathbf{e}_k \right] \\ &= \max_{\mathbf{p} \in \Delta} \mathbb{E}_{k \sim \mathbf{p}} \left[ \frac{1}{2} \|\mathbf{p} - \mathbf{e}_k\|_{\mathbf{W}}^2 + \frac{1}{2} \mathbf{e}_k^\top \mathbf{A} \mathbf{e}_k + \mathbf{b}^\top \mathbf{e}_k \right] \\ &= \max_{\mathbf{p} \in \Delta} -\frac{1}{2} \mathbf{p}^\top \mathbf{W}^{-1} \mathbf{p} + \frac{1}{2} \text{diag}(\mathbf{W}^{-1} + \mathbf{A})^\top \mathbf{p} + \mathbf{b}^\top \mathbf{p} \end{aligned}$$

The proof is completed by applying Lemma 4.1.  $\square$

#### 4.1 Minimax Analysis of the Brier Game

Next, we turn to computing  $V(\mathbf{s}, \sigma^2, n)$  as a recursion and specifying the minimax and maximin strategies. However, for the value-to-go function to retain its quadratic form, we need an alignment condition on  $\mathbf{W}$ . We say that  $\mathbf{W}$  is *aligned with the simplex* if

$$\left( \mathbf{W} - \frac{\mathbf{W} \mathbf{1} \mathbf{1}^\top \mathbf{W}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \right) \text{diag}(\mathbf{W}^{-1}) \succeq -2 \frac{\mathbf{W} \mathbf{1}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}}, \quad (4)$$

where  $\succeq$  denotes an entry-wise inequality between vectors. Note that many matrices besides  $\mathbf{I}$  satisfy this condition: for example, all symmetric  $2 \times 2$  matrices. We can now fully specify the value and strategies for the Brier game.

**Theorem 4.3.** Consider the  $T$ -round Brier game with Mahalanobis loss  $\frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2$  with  $\mathbf{W}$  satisfying the alignment condition (4). After  $n$  outcomes  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  with statistics  $\mathbf{s} = \sum_{t=1}^n \mathbf{x}_t$  and  $\sigma^2 = \sum_{t=1}^n \mathbf{x}_t^\top \mathbf{W}^{-1} \mathbf{x}_t$  the value-to-go is

$$V(\mathbf{s}, \sigma^2, n) = \frac{1}{2} \alpha_n \mathbf{s}^\top \mathbf{W}^{-1} \mathbf{s} - \frac{1}{2} \sigma^2 + \frac{1}{2} (1 - n \alpha_n) \text{diag}(\mathbf{W}^{-1})^\top \mathbf{s} + \gamma_n,$$

and the minimax and maximin strategies are given by

$$\begin{aligned} \mathbf{a}^*(\mathbf{s}, \sigma^2, n) = \mathbf{p}^*(\mathbf{s}, \sigma^2, n) &= \frac{\mathbf{W}\mathbf{1}}{\mathbf{1}^\top \mathbf{W}\mathbf{1}} + \alpha_{n+1} \left( \mathbf{s} - \frac{n\mathbf{W}\mathbf{1}}{\mathbf{1}^\top \mathbf{W}\mathbf{1}} \right) \\ &\quad + \frac{1}{2}(1 - n\alpha_{n+1}) \left( \mathbf{W} - \frac{\mathbf{W}\mathbf{1}\mathbf{1}^\top \mathbf{W}}{\mathbf{1}^\top \mathbf{W}\mathbf{1}} \right) \text{diag}(\mathbf{W}^{-1}) \end{aligned}$$

where the coefficients are defined recursively by

$$\begin{aligned} \alpha_T &= \frac{1}{T} & \gamma_T &= 0 \\ \alpha_n &= \alpha_{n+1}^2 + \alpha_{n+1} \\ \gamma_n &= \frac{(1 - n\alpha_{n+1})^2}{2} \left( \frac{1}{4} \text{diag}(\mathbf{W}^{-1})^\top \mathbf{W} \text{diag}(\mathbf{W}^{-1}) - \frac{(\frac{1}{2} \mathbf{1}^\top \mathbf{W} \text{diag}(\mathbf{W}^{-1}) - 1)^2}{\mathbf{1}^\top \mathbf{W}\mathbf{1}} \right) \\ &\quad + \gamma_{n+1}. \end{aligned}$$

*Proof.* We prove this by induction, beginning at the end of the game and working backwards in time. Assume that  $V(\mathbf{s}, \sigma^2, T)$  has the given form. Recall that the value at the end of the game is  $V(\mathbf{s}, \sigma^2, T) = -\inf_{\mathbf{a}} \sum_{t=1}^T \frac{1}{2} \|\mathbf{a} - \mathbf{x}_t\|_{\mathbf{W}}^2$  and is given by Lemma 3.1. Matching coefficients, we find  $V(\mathbf{s}, \sigma^2, T)$  corresponds to  $\alpha_T = \frac{1}{T}$  and  $\gamma_T = 0$ .

Now assume that  $V$  has the assumed form after  $n$  rounds. Using  $\mathbf{s}$  and  $\sigma^2$  to denote the state after  $n-1$  rounds, we can write

$$\begin{aligned} V(\mathbf{s}, \sigma^2, n-1) &= \min_{\mathbf{a} \in \Delta} \max_{\mathbf{x} \in \Delta} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} \alpha_n (\mathbf{s} + \mathbf{x})^\top \mathbf{W}^{-1} (\mathbf{s} + \mathbf{x}) \\ &\quad - \frac{1}{2} (\sigma^2 + \mathbf{x}^\top \mathbf{W}^{-1} \mathbf{x}) + \frac{1}{2} (1 - n\alpha_n) \text{diag}(\mathbf{W}^{-1})^\top (\mathbf{s} + \mathbf{x}) + \gamma_n. \end{aligned}$$

Using Lemma 4.2 to evaluate the right hand side produces a quadratic function in the state, and we can then match terms to find  $\alpha_{n-1}$  and  $\gamma_{n-1}$  and the minimax and maximin strategy. The final step is checking the  $\mathbf{p}^* \succeq 0$  condition necessary to apply Lemma 4.2, which is equivalent to  $\mathbf{W}$  being aligned with the simplex. See the appendix for a complete proof.  $\square$

This full characterization of the game allows us to derive the following minimax regret bound.

**Theorem 4.4.** *Let  $\mathbf{W}$  satisfy the alignment condition (4). The minimax regret of the  $T$ -round simplex game satisfies*

$$V \leq \frac{1 + \ln(T)}{2} \left( \frac{1}{4} \text{diag}(\mathbf{W}^{-1})^\top \mathbf{W} \text{diag}(\mathbf{W}^{-1}) - \frac{(\frac{1}{2} \mathbf{1}^\top \mathbf{W} \text{diag}(\mathbf{W}^{-1}) - 1)^2}{\mathbf{1}^\top \mathbf{W}\mathbf{1}} \right).$$

*Proof.* The regret is equal to the value of the game,  $V = V(\mathbf{0}, 0, 0) = \gamma_0$ . First observe that

$$\begin{aligned} (1 - n\alpha_{n+1})^2 &= 1 - 2n\alpha_{n+1} + n^2\alpha_{n+1}^2 \\ &= 1 - 2n\alpha_{n+1} + n^2(\alpha_n - \alpha_{n+1}) \\ &= \alpha_{n+1} + 1 - (n+1)^2\alpha_{n+1} + n^2\alpha_n. \end{aligned}$$

After summing over  $n$  the last two terms telescope, and we find

$$\gamma_0 \propto \sum_{n=0}^{T-1} (1 - n\alpha_{n+1})^2 = -T^2\alpha_T + \sum_{n=0}^{T-1} (1 + \alpha_{n+1}) = \sum_{n=1}^T \alpha_n.$$

Each  $\alpha_n$  can be bounded by  $1/n$ , as observed in [TW00, proof of Lemma 2]. In the base case  $n = T$  this holds with equality, and for  $n < T$  we have

$$\alpha_n = \alpha_{n+1}^2 + \alpha_{n+1} \leq \frac{1}{(n+1)^2} + \frac{1}{n+1} = \frac{1}{n} \frac{n(n+2)}{(n+1)^2} \leq \frac{1}{n}.$$

It follows that  $\gamma_0 \propto \sum_{n=1}^T \alpha_n \leq \sum_{n=1}^T \frac{1}{n} \leq 1 + \ln(T)$  as desired.  $\square$

## 5 Norm Ball Game

This section parallels the previous. Here, we consider the online game with Mahalanobis loss and  $\mathcal{A} = \mathcal{X} = \mathcal{O} := \{\mathbf{x} \in \mathbb{R}^K \mid \|\mathbf{x}\| \leq 1\}$ , the 2-norm Euclidian ball (not the Mahalanobis ball). We show that the value-to-go function is always quadratic in  $\mathbf{s}$  and linear in  $\sigma^2$  and derive the minimax and maximin strategies.

**Lemma 5.1.** *Fix a symmetric matrix  $\mathbf{A}$  and vector  $\mathbf{b}$  and assume  $\mathbf{A} + \mathbf{W}^{-1} \succ \mathbf{0}$ . Let  $\lambda_{\max}$  be the largest eigenvalue of  $\mathbf{W}^{-1} + \mathbf{A}$  and  $\mathbf{v}_{\max}$  the corresponding eigenvector. If  $\mathbf{b}^\top (\lambda_{\max} \mathbf{I} - \mathbf{A})^{-2} \mathbf{b} \leq 1$ , then the optimization problem*

$$\inf_{\mathbf{a} \in \mathcal{O}} \sup_{\mathbf{x} \in \mathcal{O}} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{x}^\top \mathbf{b}$$

has value  $\frac{1}{2} \mathbf{b}^\top (\lambda_{\max} \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} + \frac{1}{2} \lambda_{\max}$ , minimax strategy  $\mathbf{a}^* = (\lambda_{\max} \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}$  and a randomized maximin strategy that plays two unit length vectors, with

$$\Pr \left( \mathbf{x} = \mathbf{a}_\perp \pm \sqrt{1 - \mathbf{a}_\perp^\top \mathbf{a}_\perp} \mathbf{v}_{\max} \right) = \frac{1}{2} \pm \frac{\mathbf{a}_\parallel^\top \mathbf{v}_{\max}}{2\sqrt{1 - \mathbf{a}_\perp^\top \mathbf{a}_\perp}},$$

where  $\mathbf{a}_\perp$  and  $\mathbf{a}_\parallel$  are the components of  $\mathbf{a}^*$  perpendicular and parallel to  $\mathbf{v}_{\max}$ .

*Proof.* As the objective is convex, the inner optimum must be on the boundary and hence will be at a unit vector  $\mathbf{x}$ . Introduce a Lagrange multiplier  $\lambda$  for  $\mathbf{x}^\top \mathbf{x} \leq 1$  to get the Lagrangian

$$\inf_{\mathbf{a} \in \mathcal{O}} \inf_{\lambda \geq 0} \sup_{\mathbf{x}} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{x}^\top \mathbf{b} + \lambda \frac{1}{2} (1 - \mathbf{x}^\top \mathbf{x}).$$

This is concave in  $\mathbf{x}$  if  $\mathbf{W}^{-1} + \mathbf{A} - \lambda \mathbf{I} \preceq \mathbf{0}$ , that is,  $\lambda_{\max} \leq \lambda$ . Differentiating yields the optimizer  $\mathbf{x}^* = (\mathbf{W}^{-1} + \mathbf{A} - \lambda \mathbf{I})^{-1} (\mathbf{W}^{-1} \mathbf{a} - \mathbf{b})$ , which leaves us with an optimization in only  $\mathbf{a}$  and  $\lambda$ :

$$\inf_{\mathbf{a} \in \mathcal{O}} \inf_{\lambda \geq \lambda_{\max}} \frac{1}{2} \mathbf{a}^\top \mathbf{W}^{-1} \mathbf{a} - \frac{1}{2} (\mathbf{W}^{-1} \mathbf{a} - \mathbf{b})^\top (\mathbf{W}^{-1} + \mathbf{A} - \lambda \mathbf{I})^{-1} (\mathbf{W}^{-1} \mathbf{a} - \mathbf{b}) + \frac{1}{2} \lambda.$$

Since the infimums are over closed sets, we can exchange their order. Unconstrained optimization of  $\mathbf{a}$  results in  $\mathbf{a}^* = (\lambda \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}$ . Evaluating the objective at  $\mathbf{a}^*$  and using  $\mathbf{W}^{-1} \mathbf{a}^* - \mathbf{b} = \mathbf{W}^{-1} (\lambda \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} - \mathbf{b} = (\mathbf{W}^{-1} + \mathbf{A} - \lambda \mathbf{I})^{-1} (\lambda \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}$  results in

$$\inf_{\lambda \geq \lambda_{\max}} \frac{1}{2} \mathbf{b}^\top (\lambda \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} + \frac{1}{2} \lambda = \inf_{\lambda \geq \lambda_{\max}} \frac{1}{2} \left( \sum_i \frac{(\mathbf{u}_i^\top \mathbf{b})^2}{\lambda - \lambda_i} + \lambda \right),$$

using the spectral decomposition  $\mathbf{A} = \sum_i \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$ . For  $\lambda \geq \lambda_{\max}$ , we have  $\lambda \geq \lambda_i$ . Taking derivatives, provided  $\mathbf{b}^\top (\lambda_{\max} \mathbf{I} - \mathbf{A})^{-2} \mathbf{b} \leq 1$ , this function is increasing in  $\lambda \geq \lambda_{\max}$ , and so obtains its infimum at  $\lambda_{\max}$ . Thus, when the assumed inequality is satisfied, the  $\mathbf{a}^*$  is minimax for the given  $\mathbf{x}^*$ .

To obtain the maximin strategy, we can take the usual convexification where the Adversary plays distributions  $P$  over the unit sphere. This allows us to swap the infimum and supremum (see e.g. Sion's minimax theorem[Sio58]) and obtain an equivalent optimization problem. We then see that the objective only depends on the mean  $\boldsymbol{\mu} = \mathbb{E} \mathbf{x}$  and second moment  $\mathbf{D} = \mathbb{E} \mathbf{x} \mathbf{x}^\top$  of the distribution  $P$ . The characterization in [KNW13, Theorem 2.1] tells us that  $\boldsymbol{\mu}, \mathbf{D}$  are the first two moments of a distribution on units iff  $\text{tr}(\mathbf{D}) = 1$  and  $\mathbf{D} \succeq \boldsymbol{\mu} \boldsymbol{\mu}^\top$ . Then, our usual min-max swap yields

$$\begin{aligned} V &= \sup_P \inf_{\mathbf{a} \in \mathcal{O}} \mathbb{E}_{\mathbf{x} \sim P} \left[ \frac{1}{2} \mathbf{a}^\top \mathbf{W}^{-1} \mathbf{a} - \mathbf{a}^\top \mathbf{W}^{-1} \mathbf{x} + \frac{1}{2} \mathbf{x}^\top \mathbf{W}^{-1} \mathbf{x} + \frac{1}{2} \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{x} \right] \\ &= \sup_{\boldsymbol{\mu}, \mathbf{D}} \inf_{\mathbf{a} \in \mathcal{O}} \frac{1}{2} \mathbf{a}^\top \mathbf{W}^{-1} \mathbf{a} - \mathbf{a}^\top \mathbf{W}^{-1} \boldsymbol{\mu} + \frac{1}{2} \text{tr}((\mathbf{W}^{-1} + \mathbf{A}) \mathbf{D}) + \mathbf{b}^\top \boldsymbol{\mu} \\ &= \sup_{\boldsymbol{\mu}, \mathbf{D}} -\frac{1}{2} \boldsymbol{\mu}^\top \mathbf{W}^{-1} \boldsymbol{\mu} + \frac{1}{2} \text{tr}((\mathbf{W}^{-1} + \mathbf{A}) \mathbf{D}) + \mathbf{b}^\top \boldsymbol{\mu} \\ &= -\frac{1}{2} \mathbf{a}^{*\top} \mathbf{W}^{-1} \mathbf{a}^* + \mathbf{b}^\top \mathbf{a}^* + \sup_{\substack{\mathbf{D} \succeq \mathbf{a}^* \mathbf{a}^{*\top} \\ \text{tr}(\mathbf{D})=1}} \frac{1}{2} \text{tr}((\mathbf{W}^{-1} + \mathbf{A}) \mathbf{D}) \end{aligned}$$

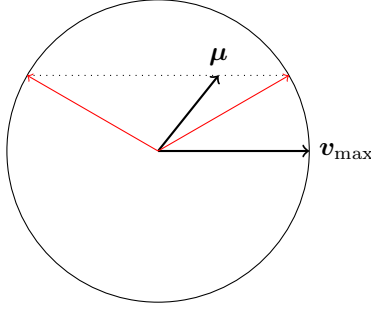


Figure 2: Illustration of the maximin distribution from Lemma 5.1. The mixture of red unit vectors with mean  $\boldsymbol{\mu}$  has second moment  $\boldsymbol{D} = \boldsymbol{\mu}\boldsymbol{\mu}^\top + (1 - \boldsymbol{\mu}^\top\boldsymbol{\mu})\boldsymbol{v}_{\max}\boldsymbol{v}_{\max}^\top$ .

where the second equality uses  $\boldsymbol{a} = \boldsymbol{\mu}$  and the third used the saddle point condition  $\boldsymbol{\mu}^* = \boldsymbol{a}^*$ . The matrix  $\boldsymbol{D}$  with constraint  $\text{tr}(\boldsymbol{D}) = 1$  now seeks to align with the largest eigenvector of  $\boldsymbol{W}^{-1} + \boldsymbol{A}$  but it also has to respect the constraint  $\boldsymbol{D} \succeq \boldsymbol{a}^*\boldsymbol{a}^{*\top}$ . We now re-parameterize by  $\boldsymbol{C} = \boldsymbol{D} - \boldsymbol{a}^*\boldsymbol{a}^{*\top}$ . We then need to find

$$\sup_{\substack{\boldsymbol{C} \succeq \mathbf{0} \\ \text{tr}(\boldsymbol{C}) = 1 - \boldsymbol{a}^{*\top}\boldsymbol{a}^*}} \frac{1}{2} \text{tr}((\boldsymbol{W}^{-1} + \boldsymbol{A})\boldsymbol{C}).$$

By linearity of the objective the maximizer is of rank 1, and hence this is a (scaled) maximum eigenvalue problem, with solution given by  $\boldsymbol{C}^* = (1 - \boldsymbol{a}^{*\top}\boldsymbol{a}^*)\boldsymbol{v}_{\max}\boldsymbol{v}_{\max}^\top$ , so that  $\boldsymbol{D}^* = \boldsymbol{a}^*\boldsymbol{a}^{*\top} + (1 - \boldsymbol{a}^{*\top}\boldsymbol{a}^*)\boldsymbol{v}_{\max}\boldsymbol{v}_{\max}^\top$ . This essentially reduces finding  $\boldsymbol{P}$  to a 2-dimensional problem, which can be solved in closed form [KNW13, Lemma 4.1]. It is easy to verify that the mixture in the theorem has the desired mean  $\boldsymbol{a}^*$  and second moment  $\boldsymbol{D}^*$ . See Figure 2 for the geometrical intuition.  $\square$

Notice that both the minimax and maximin strategies only depend on  $\boldsymbol{W}$  through  $\lambda_{\max}$  and  $\boldsymbol{v}_{\max}$ .

## 5.1 Minimax Analysis of the Ball Game

With the above lemma, we can compute the value and strategies for the ball game in an analogous way to Theorem 4.3. Again, we find that the value function at the end of the game is quadratic in the state, and, surprisingly, remains quadratic under the backwards induction.

**Theorem 5.2.** *Consider the  $T$ -round ball game with loss  $\frac{1}{2}\|\boldsymbol{a} - \boldsymbol{x}\|_{\boldsymbol{W}}^2$ . After  $n$  rounds, the value-to-go for a state with statistics  $\boldsymbol{s} = \sum_{t=1}^n \boldsymbol{x}_t$  and  $\sigma^2 = \sum_{t=1}^n \boldsymbol{x}_t^\top \boldsymbol{W}^{-1} \boldsymbol{x}_t$  is*

$$V(\boldsymbol{s}, \sigma^2, n) = \frac{1}{2} \boldsymbol{s}^\top \boldsymbol{A}_n \boldsymbol{s} - \frac{1}{2} \sigma^2 + \gamma_n.$$

The minimax strategy plays

$$\boldsymbol{a}^*(\boldsymbol{s}, \sigma^2, n) = (\lambda_{\max} \boldsymbol{I} - (\boldsymbol{A}_{n+1} - \boldsymbol{W}^{-1}))^{-1} \boldsymbol{A}_{n+1} \boldsymbol{s}$$

and the maximin strategy plays two unit length vectors with

$$\Pr\left(\boldsymbol{x} = \boldsymbol{a}_\perp \pm \sqrt{1 - \boldsymbol{a}_\perp^\top \boldsymbol{a}_\perp} \boldsymbol{v}_{\max}\right) = \frac{1}{2} \pm \frac{\boldsymbol{a}_\parallel^\top \boldsymbol{v}_{\max}}{2\sqrt{1 - \boldsymbol{a}_\perp^\top \boldsymbol{a}_\perp}},$$

where  $\lambda_{\max}$  and  $\boldsymbol{v}_{\max}$  correspond to the largest eigenvalue of  $\boldsymbol{A}_{n+1}$  and  $\boldsymbol{a}_\perp$  and  $\boldsymbol{a}_\parallel$  are the components of  $\boldsymbol{a}^*$  perpendicular and parallel to  $\boldsymbol{v}_{\max}$ . The coefficients  $\boldsymbol{A}_n$  and  $\gamma_n$  are determined recursively by base case  $\boldsymbol{A}_T = \frac{1}{T} \boldsymbol{W}^{-1}$  and  $\gamma_T = 0$  and recursion

$$\begin{aligned} \boldsymbol{A}_n &= \boldsymbol{A}_{n+1} (\boldsymbol{W}^{-1} + \lambda_{\max} \boldsymbol{I} - \boldsymbol{A}_{n+1})^{-1} \boldsymbol{A}_{n+1} + \boldsymbol{A}_{n+1} \\ \gamma_n &= \frac{1}{2} \lambda_{\max} + \gamma_{n+1}. \end{aligned}$$

*Proof outline.* The proof is by induction on the number  $n$  of rounds played. In the base case  $n = T$  we find (see (3))  $\mathbf{A}_T = \frac{1}{T} \mathbf{W}^{-1}$  and  $\gamma_T = 0$ . For the the induction step, we need to calculate

$$V(\mathbf{s}, \sigma^2, n) = \inf_{\mathbf{a} \in \mathcal{O}} \sup_{\mathbf{x} \in \mathcal{O}} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + V(\mathbf{s} + \mathbf{x}, \sigma^2 + \mathbf{x}^\top \mathbf{W}^{-1} \mathbf{x}, n + 1).$$

Using the induction hypothesis, we expand the right-hand-side to

$$\inf_{\mathbf{a} \in \mathcal{O}} \sup_{\mathbf{x} \in \mathcal{O}} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} (\mathbf{s} + \mathbf{x})^\top \mathbf{A}_{n+1} (\mathbf{s} + \mathbf{x}) - \frac{1}{2} (\sigma^2 + \mathbf{x}^\top \mathbf{W}^{-1} \mathbf{x}) + \gamma_{n+1}.$$

which we can evaluate by applying Lemma 5.1 with  $\mathbf{A} = \mathbf{A}_{n+1} - \mathbf{W}^{-1}$  and  $\mathbf{b} = \mathbf{A}_{n+1} \mathbf{s}$ . Collecting terms and matching with  $V(\mathbf{s}, \sigma^2, n) = \frac{1}{2} \mathbf{s}^\top \mathbf{A}_n \mathbf{s} - \frac{1}{2} \sigma^2 + \gamma_n$  yields the recursion for  $\mathbf{A}_n$  and  $\gamma_n$  as well as the given minimax and maximin strategies. As before, much of the algebra has been moved to the appendix.  $\square$

**Understanding the eigenvalues of  $\mathbf{A}_n$**  As we have seen from the  $\mathbf{A}_n$  recursion, the eigensystem is always the same as that of  $\mathbf{W}^{-1}$ . Thus, we can characterize the minimax strategy completely by its effect on the eigenvalues of  $\mathbf{W}^{-1}$ . Denote the eigenvalues of  $\mathbf{A}_n$  and  $\mathbf{W}^{-1}$  to be  $\lambda_n^i$  and  $\nu_i$ , respectively, with  $\lambda_n^1$  corresponding to the largest eigenvalue. The eigenvalues follow:

$$\lambda_n^i = \frac{(\lambda_{n+1}^i)^2}{\nu_i + \lambda_{n+1}^1 - \lambda_{n+1}^i} + \lambda_{n+1}^i = \frac{\lambda_{n+1}^i (\nu_i + \lambda_{n+1}^1)}{\nu_i + \lambda_{n+1}^1 - \lambda_{n+1}^i},$$

which leaves the order of the  $\lambda_n^i$  unchanged. The largest eigenvalue  $\lambda_n^1$  satisfies the recurrence  $\lambda_n^1 / \nu_1 = 1/T$  and  $\lambda_n^1 / \nu_1 = (\lambda_{n+1}^1 / \nu_1)^2 + \lambda_{n+1}^1 / \nu_1$ , which, remarkably, is the same recurrence for the  $\alpha_n$  parameter in the Brier game, i.e.  $\lambda_n^{\max} = \alpha_n \nu_{\max}$ .

This observation is the key to analyzing the minimax regret.

**Theorem 5.3.** *The minimax regret of the  $T$ -round ball game satisfies  $\gamma_0 = \frac{1 + \ln(T)}{2} \lambda_{\max}(\mathbf{W}^{-1})$*

*Proof.* We have  $V = V(\mathbf{0}, 0, 0) = \gamma_0 = (1/2) \sum_{n=1}^T \lambda_n^{\max} = (1/2) \lambda_{\max}(\mathbf{W}^{-1}) \sum_{n=1}^T \alpha_n$ . The proof of Theorem 4.4 gives the bound on  $\sum_{n=1}^T \alpha_n$ .  $\square$

Taking stock, we find that the minimax regrets of the Brier game (Theorems 4.3) and ball game (Theorems 5.2) have identical dependence on the horizon  $T$  but differ in a complexity factor arising from the interaction of the action space and the loss matrix  $\mathbf{W}$ .

## 6 Conclusion

In this paper, we have presented two games that, unexpectedly, have computationally efficient minimax strategies. While the structure of the squared Mahalanobis distance is important, it is the interplay between the loss and the constraint set that allows efficient calculation of the backwards induction, value-to-go, and achieving strategies. For example, the squared Mahalanobis game with  $\ell_1$  ball action spaces does not admit a quadratic value-to-go unless  $\mathbf{W} = \mathbf{I}$ .

We emphasize the low computational cost of this method despite the exponential blow-up in state space size. In the Brier game, the  $\alpha_n$  coefficients need to be precomputed, which can be done in  $O(T)$  time. Similarly, computation of the eigenvalues of the  $\mathbf{A}_n$  coefficients for the ball game can be done in  $O(TK + K^3)$  time. Then, at each iteration of the algorithm, only matrix-vector multiplications between the current state and the precomputed parameters are required. Hence, playing either  $T$  round game requires  $O(TK^2)$  time. Unfortunately, as is the case with most minimax algorithms, the time horizon must be known in advance.

There are many different future directions. We are currently pursuing a characterization of action spaces that permit quadratic value functions under squared Mahalanobis loss, and investigating connections between losses and families of value functions closed under backwards induction. There is some notion of conjugacy between losses, value-to-go functions, and action spaces, but a generalization seems difficult: the Brier game and ball game worked out for seemingly very different reasons.



## References

- [ABRT08] Jacob Abernethy, Peter L. Bartlett, Alexander Rakhlin, and Ambuj Tewari. Optimal strategies and minimax lower bounds for online convex games. In Servedio and Zhang [SZ08], pages 415–423.
- [AWY08] Jacob Abernethy, Manfred K. Warmuth, and Joel Yellin. When random play is optimal against an adversary. In Servedio and Zhang [SZ08], pages 437–446.
- [BGH<sup>+</sup>13] Peter L. Bartlett, Peter Grunwald, Peter Harremoës, Fares Hedayati, and Wojciech Kotłowski. Horizon-independent optimal prediction with log-loss in exponential families. *CoRR*, abs/1305.4324, 2013.
- [Bri50] Glenn W Brier. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.
- [CBL06] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.
- [CBS11] Nicolò Cesa-Bianchi and Ohad Shamir. Efficient online learning via randomized rounding. In J. Shawe-Taylor, R.S. Zemel, P. Bartlett, F.C.N. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 343–351, 2011.
- [HB12] Fares Hedayati and Peter L. Bartlett. Exchangeability characterizes optimality of sequential normalized maximum likelihood and bayesian prediction with jeffreys prior. In *International Conference on Artificial Intelligence and Statistics*, pages 504–510, 2012.
- [KM05] Petri Kontkanen and Petri Myllymäki. A fast normalized maximum likelihood algorithm for multinomial data. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI-05)*, pages 1613–1616, 2005.
- [KNW13] Wouter M. Koolen, Jiazhong Nie, and Manfred K. Warmuth. Learning a set of directions. In Shai Shalev-Shwartz and Ingo Steinwart, editors, *Proceedings of the 26th Annual Conference on Learning Theory (COLT)*, June 2013.
- [Sht87] Yurii Mikhailovich Shtar’kov. Universal sequential coding of single messages. *Problemy Peredachi Informatsii*, 23(3):3–17, 1987.
- [Sio58] Maurice Sion. On general minimax theorems. *Pacific J. Math.*, 8(1):171–176, 1958.
- [SZ08] Rocco A. Servedio and Tong Zhang, editors. *21st Annual Conference on Learning Theory - COLT 2008, Helsinki, Finland, July 9-12, 2008*. Omnipress, 2008.
- [TW00] Eiji Takimoto and Manfred K. Warmuth. The minimax strategy for Gaussian density estimation. In *13th COLT*, pages 100–106, 2000.

## A Proofs

*Proof of Theorem 4.3.* We prove by induction. Recall that  $V(\mathbf{s}, \sigma^2, T)$ , the value at the end of the game, already has

$$V(\mathbf{s}, \sigma^2, T) = \frac{1}{2T} \mathbf{s}^\top \mathbf{W}^{-1} \mathbf{s} - \frac{1}{2} \sigma^2,$$

corresponding to  $\alpha_T = \frac{1}{T}$  and  $\gamma_T = 0$ . Adding a constant to the objective in Lemma 4.2 only adds the same offset to the value without changing the saddle point. Now, assume the induction hypothesis for  $n \leq T$  rounds. Using the abbreviation  $\mathbf{d} = \text{diag}(\mathbf{W}^{-1})$ , we then need to evaluate

$$\begin{aligned} V(\mathbf{s}, \sigma^2, n-1) &= \min_{\mathbf{a} \in \Delta} \max_{\mathbf{x} \in \Delta} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} \alpha_n (\mathbf{s} + \mathbf{x})^\top \mathbf{W}^{-1} (\mathbf{s} + \mathbf{x}) \\ &\quad - \frac{1}{2} (\sigma^2 + \mathbf{x}^\top \mathbf{W}^{-1} \mathbf{x}) + \frac{1}{2} (1 - n\alpha_n) \mathbf{d}^\top (\mathbf{s} + \mathbf{x}) + \gamma_n. \\ &= \min_{\mathbf{a} \in \Delta} \max_{\mathbf{x} \in \Delta} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} \alpha_n \mathbf{x}^\top \mathbf{W}^{-1} \mathbf{x} - \frac{1}{2} \mathbf{x}^\top \mathbf{W}^{-1} \mathbf{x} \\ &\quad + \frac{1}{2} (1 - n\alpha_n) \mathbf{d}^\top \mathbf{x} + \alpha_n \mathbf{s}^\top \mathbf{W}^{-1} \mathbf{x} + \frac{1}{2} \alpha_n \mathbf{s}^\top \mathbf{W}^{-1} \mathbf{s} \\ &\quad + \frac{1}{2} (1 - n\alpha_n) \mathbf{d}^\top \mathbf{s} - \frac{1}{2} \sigma^2 + \gamma_n. \end{aligned}$$

Applying Lemma 4.2 with  $\mathbf{A} = (\alpha_n - 1) \mathbf{W}^{-1}$  and  $\mathbf{b} = \frac{1}{2} (1 - n\alpha_n) \mathbf{d} + \alpha_n \mathbf{W}^{-1} \mathbf{s}$  produces the value (where all time indices are  $n$ )

$$\begin{aligned} V(\mathbf{s}, \sigma^2, n-1) &= \frac{1}{2} (\alpha^2 + \alpha) \mathbf{s}^\top \mathbf{W}^{-1} \mathbf{s} + \left( \alpha \frac{1 - (n-1)\alpha}{2} + \frac{1}{2} (1 - n\alpha) \right) \mathbf{d}^\top \mathbf{s} \\ &\quad + \frac{1}{2} \left( \frac{1 - (n-1)\alpha}{2} \right)^2 \mathbf{d}^\top \mathbf{W} \mathbf{d} \\ &\quad - \frac{1}{2} \frac{\left( \frac{1 - (n-1)\alpha}{2} \mathbf{1}^\top \mathbf{W} \mathbf{d} + \alpha \mathbf{1}^\top \mathbf{s} - 1 \right)^2}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \\ &\quad - \frac{1}{2} \sigma^2 + \gamma \end{aligned}$$

and the strategy

$$\mathbf{p}^* = \left( \mathbf{W} - \frac{\mathbf{W} \mathbf{1} \mathbf{1}^\top \mathbf{W}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \right) \left( \frac{1}{2} (1 - (n-1)\alpha_n) \mathbf{d} + \alpha_n \mathbf{W}^{-1} \mathbf{s} \right) + \frac{\mathbf{W} \mathbf{1}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}}.$$

Using the fact that  $\mathbf{1}^\top \mathbf{s} = n - 1$ , the value can be further simplified to

$$\begin{aligned} V(\mathbf{s}, \sigma^2, n-1) &= \frac{1}{2} (\alpha^2 + \alpha) \mathbf{s}^\top \mathbf{W}^{-1} \mathbf{s} + \frac{1}{2} (1 - (n-1)(\alpha^2 + \alpha)) \mathbf{d}^\top \mathbf{s} \\ &\quad + (1 - (n-1)\alpha)^2 \left( \frac{1}{8} \mathbf{d}^\top \mathbf{W} \mathbf{d} - \frac{1}{2} \frac{(\frac{1}{2} \mathbf{1}^\top \mathbf{W} \mathbf{d} - 1)^2}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \right) - \frac{1}{2} \sigma^2 + \gamma. \end{aligned}$$

This establishes the induction step upon setting

$$\begin{aligned} \alpha_n &= \alpha_{n+1}^2 + \alpha_{n+1}, \\ \gamma_n &= \gamma_{n+1} + (1 - n\alpha_{n+1})^2 \left( \frac{1}{8} \mathbf{d}^\top \mathbf{W} \mathbf{d} - \frac{1}{2} \frac{(\frac{1}{2} \mathbf{1}^\top \mathbf{W} \mathbf{d} - 1)^2}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \right). \end{aligned}$$

Finally, we need to verify that the strategies above respect the constraints of the game, i.e. that  $\mathbf{a}^*$  and  $\mathbf{p}^*$  are in the simplex. Otherwise, the above calculations do not correspond to the minimax strategy. We need to show for all  $\mathbf{s} \geq 0$  with sum  $n - 1$  that

$$\alpha_n \mathbf{s} + \frac{1}{2} (1 - (n-1)\alpha_n) \mathbf{W} \mathbf{d} \geq \left( \alpha_n (n-1) + \frac{1}{2} (1 - (n-1)\alpha_n) (\mathbf{1}^\top \mathbf{W} \mathbf{d}) - 1 \right) \frac{\mathbf{W} \mathbf{1}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}}.$$

As this needs to hold for all  $\mathbf{s} \geq 0$ , we need in fact that component-wise

$$\frac{1}{2}(1 - (n-1)\alpha_n) \mathbf{W} \mathbf{d} \geq \left( \alpha_n(n-1) + \frac{1}{2}(1 - (n-1)\alpha_n)(\mathbf{1}^\top \mathbf{W} \mathbf{d}) - 1 \right) \frac{\mathbf{W} \mathbf{1}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}},$$

that is

$$\left( \mathbf{W} - \frac{\mathbf{W} \mathbf{1} \mathbf{1}^\top \mathbf{W}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \right) \text{diag}(\mathbf{W}^{-1}) \succeq \frac{2((n-1)\alpha_n - 1)}{1 - (n-1)\alpha_n} \frac{\mathbf{W} \mathbf{1}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} = -2 \frac{\mathbf{W} \mathbf{1}}{\mathbf{1}^\top \mathbf{W} \mathbf{1}}.$$

□

*Proof of Theorem 5.2.* Recall that we needed to calculate

$$V(\mathbf{s}, \sigma^2, n) = \inf_{\mathbf{a}} \sup_{\mathbf{x}} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} (\mathbf{s} + \mathbf{x})^\top \mathbf{A}_{n+1} (\mathbf{s} + \mathbf{x}) - \frac{1}{2} (\sigma^2 + \mathbf{x}^\top \mathbf{W}^{-1} \mathbf{x}) + \gamma_{n+1}$$

which may be reorganized to

$$\frac{1}{2} \mathbf{s}^\top \mathbf{A}_{n+1} \mathbf{s} - \frac{1}{2} \sigma^2 + \gamma_{n+1} + \inf_{\mathbf{a}} \sup_{\mathbf{x}} \frac{1}{2} \|\mathbf{a} - \mathbf{x}\|_{\mathbf{W}}^2 + \frac{1}{2} \mathbf{x}^\top (\mathbf{A}_{n+1} - \mathbf{W}^{-1}) \mathbf{x} + \mathbf{s}^\top \mathbf{A}_{n+1} \mathbf{x}.$$

We now apply Lemma 5.1 with  $\mathbf{A} = \mathbf{A}_{n+1} - \mathbf{W}^{-1}$  and  $\mathbf{b} = \mathbf{A}_{n+1} \mathbf{s}$ . Let  $\lambda_{n+1}^{\max}$  be the largest eigenvalue of  $\mathbf{A}_{n+1} - \mathbf{W}^{-1} + \mathbf{W}^{-1} = \mathbf{A}_{n+1}$ . If we have

$$1 \geq \mathbf{s}^\top \mathbf{A}_{n+1} (\mathbf{W}^{-1} + \lambda_{n+1}^{\max} \mathbf{I} - \mathbf{A}_{n+1})^{-2} \mathbf{A}_{n+1} \mathbf{s}, \quad (5)$$

then the value equals

$$\frac{1}{2} \mathbf{s}^\top \mathbf{A}_{n+1} \mathbf{s} - \frac{1}{2} \sigma^2 + \gamma_{n+1} + \frac{1}{2} \lambda_{n+1}^{\max} + \frac{1}{2} \mathbf{s}^\top \mathbf{A}_{n+1} (\mathbf{W}^{-1} + \lambda_{n+1}^{\max} \mathbf{I} - \mathbf{A}_{n+1})^{-1} \mathbf{A}_{n+1} \mathbf{s}.$$

So if we assert that  $V(\mathbf{s}, \sigma^2, n) = \frac{1}{2} \mathbf{s}^\top \mathbf{A}_n \mathbf{s} - \frac{1}{2} \sigma^2 + \gamma_n$ , we find that we must have the correspondence:

$$\begin{aligned} \mathbf{A}_n &= \mathbf{A}_{n+1} (\mathbf{W}^{-1} + \lambda_{n+1}^{\max} \mathbf{I} - \mathbf{A}_{n+1})^{-1} \mathbf{A}_{n+1} + \mathbf{A}_{n+1}, \\ \gamma_n &= \frac{1}{2} \lambda_{n+1}^{\max} + \gamma_{n+1}. \end{aligned}$$

The final piece is to check Equation (5). To see this, let  $\nu$  denote the eigenvalues of  $\mathbf{W}^{-1}$ . Observe that the largest eigenvalue of  $\mathbf{A}_{n+1} (\mathbf{W}^{-1} + \lambda_{n+1}^{\max} \mathbf{I} - \mathbf{A}_{n+1})^{-2} \mathbf{A}_{n+1}$  is  $(\lambda_{n+1}^{\max} / \nu_{\max})^2$ , because the mapping

$$\lambda \mapsto \frac{\lambda^2}{(\nu + \lambda_{n+1}^{\max} - \lambda)^2}$$

is monotonic in  $\lambda \leq \lambda_{n+1}^{\max}$ . The proof now follows by combining the recurrence for the largest eigenvalue from Section 5.1 with the bound on  $\alpha_n$  from the proof of Theorem 4.4:

$$\mathbf{s}^\top \mathbf{A}_{n+1} (\mathbf{W}^{-1} + \lambda_{n+1}^{\max} \mathbf{I} - \mathbf{A}_{n+1})^{-2} \mathbf{A}_{n+1} \mathbf{s} \leq \left( \frac{\lambda_{n+1}^{\max}}{\nu_{\max}} \right)^2 \|\mathbf{s}\|^2 \leq \alpha_{n+1}^2 n^2 \leq \frac{n^2}{(n+1)^2} < 1.$$

□