Adaptivity and Optimism: An Improved Exponentiated Gradient Algorithm

Jacob Steinhardt JSTEINHARDT@CS.STANFORD.EDU
Percy Liang PLIANG@CS.STANFORD.EDU
Stanford University, 353 Serra Street, Stanford, CA 94305 USA

Abstract

We present an adaptive variant of the exponentiated gradient algorithm. Leveraging the optimistic learning framework of Rakhlin & Sridharan (2012), we obtain regret bounds that in the learning from experts setting depend on the variance and path length of the best expert, improving on results by Hazan & Kale (2008) and Chiang et al. (2012), and resolving an open problem posed by Kale (2012). Our techniques naturally extend to matrix-valued loss functions, where we present an *adaptive matrix exponentiated gradient* algorithm. To obtain the optimal regret bound in the matrix case, we generalize the Follow-the-Regularized-Leader algorithm to vector-valued payoffs, which may be of independent interest.

1. Introduction

The exponentiated gradient (EG) algorithm is a powerful tool for performing online learning in the presence of many irrelevant features (Kivinen & Warmuth, 1997; Littlestone, 1988). EG is often used in the "learning from experts" setting, in which it is also known as the weighted majority algorithm (Littlestone & Warmuth, 1989). In this setting, EG entertains regret bounds of the form

$$\text{Regret} \leq \frac{\log(n)}{\eta} + \eta \sum_{t=1}^{T} \|z_t\|_\infty^2, \tag{1}$$

where η is the step size, z_t is the vector of losses, and n is the number of experts. Such bounds (as well as slightly stronger bounds based on *local norms*) can be obtained under the mirror descent framework, a general tool that gives rise to many other online learning algorithms (see Shalev-Shwartz (2011) for a survey).

In contrast, Cesa-Bianchi et al. (2007) present a variant of this algorithm based on a multiplicative update of

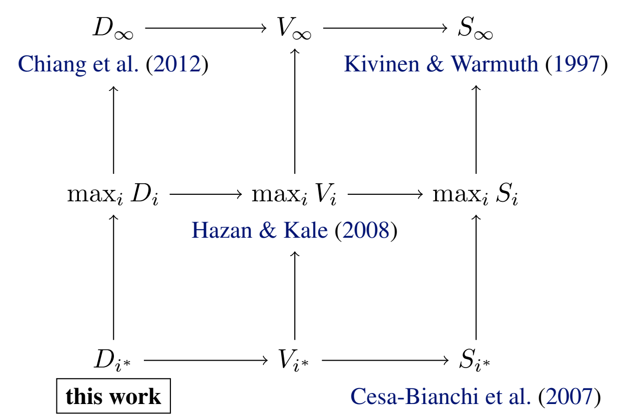

Figure 1. Summary of possible regret bounds with references to algorithms known to achieve these bounds. An arrow $A \to B$ indicates that A is a strictly better bound than B. Our algorithm simultaneously improves upon several existing results. D represents the path length, V the variance, and S the second moment; these quantities are defined formally in Section 3, Equation 24. Even in situations where D_{i^*} is $\Theta(1)$, both $\max_i D_i$ and V_{i^*} (and hence all other entries in the lattice) can be $\Theta(T)$.

$w_{t+1,i} \propto w_{t,i}(1 - \eta z_{t,i})$ rather than the usual EG update of $w_{t+1,i} \propto w_{t,i} \exp(-\eta z_{t,i})$. This algorithm cannot be cast in the mirror descent framework with a fixed regularizer, yet it achieves an improved regret bound of

$$\text{Regret} \leq \frac{\log(n)}{\eta} + \eta \sum_{t=1}^{T} z_{t,i^*}^2. \tag{2}$$

Comparing the regret bounds (2) and (1), note that (2) is in terms of the best expert i^* instead of a maximum over all experts. This latter bound can be much stronger; we show in Proposition 2.2 that there is in fact a $\Theta(\sqrt{T})$ separation of the worst-case regret in the setting where the best expert has loss identically equal to zero. Other differences between these two types of updates are discussed in Arora et al. (2012).

The fact that an algorithm achieving a better regret bound cannot be cast in the mirror descent framework is a bit unsettling. Does this mean we should abandon mirror descent

Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

as the gold standard for online learning, despite theorems asserting its optimality (Srebro et al., 2011)? We answer this question in the negative: the $(1 - \eta z_{t,i})$ update can be understood as a form of *adaptive mirror descent* (Orabona et al., 2013), where the regularizer changes in each round t in response to previously observed vectors $z_{1:t}$. We obtain a natural interpretation of the update as performing a second-order correction to the gradient.

Examining (2) more closely, we see that this corrected update should perform well when the best expert i^* incurs losses consistently close to zero; then the second term in the regret is $\sum_{t=1}^{T} z_{t,i^*}^2 \approx 0$. However, this assumption may be unrealistic, and many authors have recently considered *variance* bounds that depend only on the deviation of z_t from its average, or *path-length* bounds in terms of $z_t - z_{t-1}$ (Hazan & Kale, 2008; Chiang et al., 2012; Yang et al., 2013). Rakhlin & Sridharan (2012) present an *optimistic learning* framework that yields such bounds for any mirror descent algorithm. However, the updates in Hazan & Kale (2008) are not mirror descent updates (for any fixed regularizer), and their bounds are incomparable to the bounds obtained via optimistic learning.

In the learning from experts setting, we subsume all the previously mentioned bounds by obtaining a bound in terms of the path length of the best expert:

$$\text{Regret} \leq \frac{\log(n)}{\eta} + \eta \sum_{t=1}^{T} (z_{t,i^*} - z_{t-1,i^*})^2. \tag{3}$$

Obtaining such a bound is posed as an open problem in Kale (2012). We achieve such a regret bound (Equation 23) by applying Rakhlin's updates in the context of an adaptive mirror descent algorithm, thus obtaining an *adaptive optimistic* exponentiated gradient algorithm. When the path length is not known and η must be determined adaptively, our bounds weaken slightly but are still strong enough to answer the problem in Kale (2012), as well as to subsume all of the previously mentioned bounds in the adaptive step size setting.

Finally, we extend all these results to the matrix setting, where the learner plays a positive semidefinite matrix W_t with trace 1 (in analogy with the simplex). This setting has been extensively studied (Tsuda et al., 2005; Arora & Kale, 2007) and is important in obtaining online and approximation bounds for various combinatorial optimization problems (Arora & Kale, 2007; Hazan et al., 2012). As far as we are aware, the best known results in this setting are of the form (1). Using the machinery so far developed, all of our results extend naturally to the matrix setting. However, for the variance bound we need a new analysis tool: a variant of FTRL for *vector-valued* losses ordered relative to some cone $\mathcal{K}$.

In summary, the main contributions of this paper are:

- An interpretation of the multiplicative weights update of Cesa-Bianchi et al. (2007) as exponentiated gradient with an adaptive regularizer (Section 2).
- An improved exponentiated gradient algorithm obtaining best-known variance and path-length bounds (Section 3).
- An adaptive matrix exponentiated gradient algorithm attaining similar bounds (Section 4).
- A generalization of Follow-the-Regularized-Leader to vector-valued loss functions (Lemma 4.3).

Related work. There is a rich literature on using adaptive updates to obtain better regret bounds for online learning. A common setting is adaptive learning of a quadratic regularizer, as in the AROW (Crammer et al., 2009), AdaGrad (Duchi et al., 2011), and online preconditioning (Streeter & McMahan, 2010) algorithms. Other work includes dimension-free exponentiated gradient (Orabona, 2013), whitened perceptron (Cesa-Bianchi et al., 2005), and online adaptation of the step size (Hazan et al., 2007). The non-stationary setting was explored by Vaits et al. (2013), and McMahan & Streeter (2010) obtain regret bounds relative to a family of regularizers. More recently, many of these algorithms have been unified into a single framework by Orabona et al. (2013). To our knowledge, adaptively regularized exponentiated gradient has not been explicitly explored, though many variants on the basic multiplicative updates have been proposed (Cesa-Bianchi et al., 2007; Hazan & Kale, 2008; Chiang et al., 2012), which can be interpreted in our framework as making implicit use of an adaptive regularizer.

In addition to the variants on exponentiated gradient discussed above, Auer & Warmuth (1998) and Herbster & Warmuth (1998) have studied the case where the best expert can change over time. Finally, Sabato et al. (2012) consider a generalization of the Winnow algorithm (Littlestone, 1988), which corresponds to exponentiated gradient with a hinge-like loss, and provide a careful analysis of the regret that is more precise than the mirror descent analysis.

2. A Tale of Two Updates

Our point of departure is the two different types of multiplicative updates mentioned in the introduction. For simplicity we will consider the setting of learning from expert advice.[1] In this setting there are n experts, and the learner maintains a probability distribution $w_t \in \Delta_n$ over the experts. In each round $t = 1, \ldots, T$, the learner plays w_t, a vector $z_t \in [-1, 1]^n$ is revealed, and the learner incurs

[1] The general setting follows a nearly identical analysis and is covered in the supplementary material.

Name	Update	Prediction	Source
EG (MW1)	$\beta_{t+1} = \beta_t - \eta z_t$	$\exp(\beta_t)$	(Kivinen & Warmuth, 1997)
MW2	$\beta_{t+1,i} = \beta_{t,i} + \log(1 - \eta z_{t,i})$	$\exp(\beta_t)$	(Cesa-Bianchi et al., 2007)
Variation-MW	$\beta_{t+1,i} = \beta_{t,i} - \eta z_{t,i} - 4\eta^2(z_{t,i} - m_{t,i})^2$ $m_t = \frac{1}{t}\sum_{s=1}^{t-1} z_s$	$\exp(\beta_t)$	(Hazan & Kale, 2008)
Optimistic MW	$\beta_{t+1,i} = \beta_{t,i} - \eta z_{t,i}$	$\exp(\beta_t - \eta z_{t-1})$	(Chiang et al., 2012)
AEG-Path	$\beta_{t+1,i} = \beta_{t,i} - \eta z_{t,i} - \eta^2(z_{t,i} - z_{t-1,i})^2$	$\exp(\beta_t - \eta z_{t-1})$	this work
AMEG-Path	$B_{t+1} = B_t - \eta Z_t - \eta^2(Z_t - Z_{t-1})^2$	$\exp(B_t - \eta Z_{t-1})$	this work

Table 1. An overview of known adaptive exponentiated gradient algorithms. The AEG-Path updates incorporate components of both the Variation-MW and Optimistic MW algorithms, and are motivated by interpreting MW2 in terms of adaptive mirror descent. The AMEG-Path updates extend AEG-Path to the matrix case (which had previously only been done for MW1).

loss $w_t^\top z_t$. The learner's goal is to minimize the regret $\sup_{u \in \Delta_n} \mathrm{Regret}(u)$, where

$$\mathrm{Regret}(u) \overset{\text{def}}{=} \sum_{t=1}^{T} w_t^\top z_t - \sum_{t=1}^{T} u^\top z_t. \tag{4}$$

The learner starts by playing w_1, where $w_{1,i} = \frac{1}{n}$ for $1 \le i \le n$. On subsequent iterations, we consider two types of updates for the weight vector w_t, as shown in (MW1) and (MW2) below:

$$w_{t+1,i} \propto w_{t,i} \exp(-\eta z_{t,i}) \tag{MW1}$$

$$w_{t+1,i} \propto w_{t,i}(1 - \eta z_{t,i}), \tag{MW2}$$

where η is the step size. The regret bounds for each of (MW1) and (MW2) are well-known (see Shalev-Shwartz (2011) and Cesa-Bianchi et al. (2007) respectively) but we include them for completeness.

Theorem 2.1. *For any $0 < \eta \le \frac{1}{2}$ and $\|z_t\|_\infty \le 1$, the updates (MW1) and (MW2) obtain respective regret bounds of*

$$\mathrm{Regret}(u) \le \frac{\log(n)}{\eta} + \eta \sum_{i=1}^{n} \sum_{t=1}^{T} w_{t,i} z_{t,i}^2 \tag{5}$$

$$\mathrm{Regret}(u) \le \frac{\log(n)}{\eta} + \eta \sum_{i=1}^{n} u_i \sum_{t=1}^{T} z_{t,i}^2 \tag{6}$$

To understand why (6) may be a better bound than (5), suppose that the best expert has loss identically equal to zero.[2] Then the optimal u places all mass on that expert, and (6) reduces to $\frac{\log(n)}{\eta} = 2\log(n)$ for $\eta = \frac{1}{2}$.

More formally, define a sequence of losses z_t to be *quasi-realizable* if one of the experts i^* has identically zero loss and all other experts have non-negative cumulative loss, i.e. $\sum_{t=1}^{T} z_{t,i} \ge 0$. It is apparent by the preceding paragraph that (MW2) achieves asymptotically constant (as a function of T) regret for any quasi-realizable sequence. In contrast, (MW1) can suffer $\Omega(\sqrt{T})$ regret:

Proposition 2.2. *For any step size η and T, there is a quasi-realizable loss sequence $(z_t)_{t=1}^T$ and a vector $u \in \Delta_n$ such that the updates (MW1) result in $\mathrm{Regret}(u) = \Omega(\sqrt{T})$.*

The proof is given in the supplementary material, but the main idea is that (MW1) will have trouble distinguishing between an expert whose loss is always zero and an expert whose loss alternates between 1 and -1. This establishes that the apparent separation between (MW1) and (MW2) is real and not an artifact of the analysis. We remark that this separation does *not* exist when all losses are non-negative. In this case both (MW1) and (MW2) enjoy $O(1)$ regret (as a function of T).

Finally, note that (MW2) cannot be realized as mirror descent for any fixed regularizer. This is because, for any mirror descent algorithm, the prediction on round $t+1$ must be a function of $\sum_{s=1}^{t} z_s$, which is not the case for (MW2).

Adaptive mirror descent However, not all is lost, as we will obtain (MW2) in terms of an *adaptive regularizer* $\psi_t(w)$. The mirror descent predictions for an adaptive regularizer are given by

$$w_t = \nabla \psi_t^* \left(\theta_t\right), \quad \theta_t \overset{\text{def}}{=} -\eta \sum_{s=1}^{t-1} z_s, \tag{7}$$

where $\psi^*(x) \overset{\text{def}}{=} \sup_w \{w^\top x - \psi(w)\}$ is the *Fenchel conjugate* of ψ. We provide general properties of Fenchel conjugates as well as several calculations of interest in the supplementary material. See Orabona et al. (2013) for a more complete exposition on adaptive mirror descent, and Shalev-Shwartz (2011) for a general survey.

We can cast (MW2) in the adaptive mirror descent framework, as detailed in Proposition 2.3 below. As we will ex-

[2]Of course, if we knew that this was the case ahead of time, there would be far better algorithms; we use this scenario purely for illustrative purposes.

plain in the next section, these updates have a natural interpretation as "pushing the regret into the regularizer".

Proposition 2.3. *Define* $\beta_{t,i} \stackrel{\text{def}}{=} \sum_{s=1}^{t-1} \log(1 - \eta z_{s,i})$ *and let*

$$\psi_t(u) \stackrel{\text{def}}{=} \sum_{i=1}^{n} u_i \log(u_i) + u^\top(\theta_t - \beta_t). \quad (8)$$

Then adaptive mirror descent with regularizer ψ_t *corresponds exactly to the updates (MW2). The corresponding regret bound is*

$$\text{Regret}(u) \leq \frac{\psi_1^*(\theta_1) + \psi_{T+1}(u)}{\eta} \quad (9)$$

$$\leq \frac{\log(n)}{\eta} + \eta \sum_{i=1}^{n} u_i \sum_{t=1}^{T} z_{t,i}^2. \quad (10)$$

Proof. By standard properties of Fenchel conjugates, we have

$$\nabla\psi_t^*(\theta_t) = \arg\min_{w\in\Delta_n} \psi_t(w) - w^\top\theta_t \quad (11)$$

$$= \arg\min_{w\in\Delta_n} \sum_{i=1}^{n} w_i \log(w_i) - w^\top\beta_t. \quad (12)$$

From here we see that $w_{t,i} \propto \exp(\beta_{t,i}) = \prod_{s=1}^{t-1}(1-\eta z_{s,i})$, so that $w_{t,i}$ does indeed correspond to (MW2).

We omit the proof of the regret bound; it follows straightforwardly from the machinery in the next section (see Proposition 3.3). □

Proposition 2.3 says we can obtain bounds that depend on the the average squared loss z_{t,i^*}^2 of the best expert i^* (u places all its mass on i^*). But intuitively, we would like to not suffer much regret even if z_{t,i^*} is large so long as its *variation* is small. We turn to this issue in the next section.

3. Adaptive Optimistic Learning

In the previous section, we saw how to obtain regret bounds that depend on the best expert i^*, but involve the second moment. Next, we show how to use the idea of optimistic learning (Rakhlin & Sridharan, 2012) to obtain results that depend on variance or path length.

In the optimistic learning framework, we are given a sequence of "hints" m_t of what z_t might be. Then rather than choosing w_t based on the negative cumulative gradients θ_t, we choose w_t based on a preemptive update $\theta_t - \eta m_t$. The resulting regret bounds thus depend on the error in the hints ($z_t - m_t$) rather than z_t. If $m_t = 0$, we recover vanilla mirror descent; if $m_t = z_{t-1}$, we obtain path-length bounds; and if $m_t = \frac{1}{t}\sum_{s=1}^{t-1} z_s$, we obtain variance bounds. We illustrate geometrically in Figure 2 how optimistic updates can improve the regret bound.

We combine optimistic learning (Rakhlin & Sridharan, 2012) with adaptive regularization (Orabona et al., 2013) to yield Algorithm 1.

Algorithm 1 Adaptive Optimistic Mirror Descent

Given: convex regularizers ψ_t and hints m_t
Initialize $\theta_1 = 0$
for $t = 1$ **to** T **do**
  Choose $w_t = \nabla\psi_t^*(\theta_t - \eta m_t)$
  Observe z_t and suffer loss $w_t^\top z_t$
  Update $\theta_{t+1} = \theta_t - \eta z_t$
end for

The regret bound for Algorithm 1 is given in Theorem 3.1:

Theorem 3.1. *Suppose that for all t,* ψ_t *is convex and satisfies the loss-bounding property:*

$$\psi_{t+1}^*(\theta_t - \eta z_t) \leq \psi_t^*(\theta_t - \eta m_t) - \eta w_t^\top(z_t - m_t). \quad (13)$$

Then

$$\text{Regret}(u) \leq \frac{\psi_1^*(\theta_1) + \psi_{T+1}(u)}{\eta}. \quad (14)$$

Proof. The proof is a relatively straightforward combination of known results. First note that ψ_t^* is convex and that $w_t = \nabla\psi_t^*(\theta_t - \eta m_t)$. Thus, $\psi_t^*(\theta_t) \geq \psi_t^*(\theta_t - \eta m_t) + \eta w_t^\top m_t$. Then, by definition of the Fenchel conjugate together with telescoping sums, we have, for any u,

$$\begin{aligned} &u^\top\theta_{T+1} - \psi_{T+1}(u) \\ &\quad\leq \psi_{T+1}^*(\theta_{T+1}) \\ &\quad= \psi_1^*(\theta_1) + \sum_{t=1}^{T} \psi_{t+1}^*(\theta_{t+1}) - \psi_t^*(\theta_t) \\ &\quad\leq \psi_1^*(\theta_1) + \sum_{t=1}^{T} \psi_{t+1}^*(\theta_{t+1}) - \psi_t^*(\theta_t - \eta m_t) - \eta w_t^\top m_t. \end{aligned}$$

By the conditions of the theorem, the sum is termwise upper bounded by $-\eta w_t^\top z_t$ and we have

$$u^\top\theta_{T+1} + \eta\sum_{t=1}^{T} w_t^\top z_t \leq \psi_1^*(\theta_1) + \psi_{T+1}(u). \quad (15)$$

Expanding θ_{T+1} as $-\eta\sum_{t=1}^{T} z_t$ completes the proof. □

The key intuition, also spelled out by Orabona et al. (2013), is that, if we make $\psi_{t+1} - \psi_t$ large enough to "swallow the regret" on round t, then we obtain bounds that depend

Figure 2. Illustration of how optimistic updates affect the regret bound. For a fixed regularizer ψ^*, the increase in regret is bounded above by $\psi^*(\theta_{t+1}) - \psi^*(\theta_t) - \eta w_t^\top z_t$. Normally $w_t = \nabla\psi^*(\theta_t)$, so that the bound is equal to the gap between ψ^* and its tangent line, as illustrated on the left. For optimistic updates we instead take $w_t = \nabla\psi^*(\theta_t - \eta m_t)$, which replaces the tangent line by the dashed line on the right. This dashed line can be bounded by the tangent line at $\theta_t - \eta m_t$, depicted as the solid line on the right.

on the regularizer $\psi_{T+1}(u)$, rather than typical bounds that depend on Bregman divergences between θ_t and θ_{t+1}.[3]

Regularization based on corrections While Theorem 3.1 deals with general sequences of regularizers ψ_t, for our purposes we will only need to consider regularizers of a special form:

$$\psi_t(w) = \psi(w) - w^\top \left[\beta_1 - \eta^2 \sum_{s=1}^{t-1} a_s\right], \tag{16}$$

where ψ is a fixed regularizer and a_t is a sequence of *corrections*. This choice of regularizer yields the more specialized Algorithm 2, which can be interpreted as performing second-order corrections to the typical gradient updates.

Algorithm 2 Adaptive Optimistic Mirror Descent (specialized to corrections)

Given: convex regularizer ψ, corrections a_t and hints m_t
Initialize β_1 arbitrarily
for $t = 1$ **to** T **do**
  Choose $w_t = \nabla\psi^*(\beta_t - \eta m_t)$
  Observe z_t and suffer loss $w_t^\top z_t$
  Update $\beta_{t+1} = \beta_t - \eta z_t - \eta^2 a_t$
end for

Corollary 3.2. *Suppose ψ is convex and a_t is such that $\psi^*(\beta_t - \eta z_t - \eta^2 a_t) \le \psi^*(\beta_t - \eta m_t) - \eta w_t^\top (z_t - m_t)$. Then*

$$\mathrm{Regret}(u) \le \frac{\psi^*(\beta_1) + \psi(u) - u^\top\beta_1}{\eta} + \eta u^\top \sum_{t=1}^{T} a_t. \tag{17}$$

Proof. The proof essentially consists of translating into the language of Theorem 3.1 and making use of the property that the Fenchel conjugate of $w \mapsto \psi(w) - w^\top c$ is $x \mapsto \psi^*(x + c)$.

Define $\psi_t(w) \stackrel{\text{def}}{=} \psi(w) - w^\top[\beta_1 - \eta^2 \sum_{s=1}^{t-1} a_s]$. Note that $\psi_t(w) = \psi(w) - w^\top(\beta_t - \theta_t)$ and hence $\psi_t^*(x) = \psi^*(x + (\beta_t - \theta_t))$. Then, looking at the condition of Theorem 3.1, we have $\psi_{t+1}^*(\theta_t - \eta z_t) = \psi_{t+1}^*(\theta_{t+1}) = \psi^*(\beta_{t+1})$ and $\psi_t^*(\theta_t - \eta m_t) = \psi^*(\beta_t - \eta m_t)$, so that the conditions on ψ and a_t in this corollary match those on ψ_t in Theorem 3.1. The corresponding regret bound is

$$\begin{aligned}\mathrm{Regret}(u) &\le \frac{\psi_1^*(\theta_1) + \psi_{T+1}(u)}{\eta} \\ &= \frac{\psi^*(\beta_1) + \psi(u) + u^\top[-\beta_1 + \eta^2 \sum_{t=1}^{T} a_t]}{\eta} \\ &= \frac{\psi^*(\beta_1) + \psi(u) - u^\top\beta_1}{\eta} + \eta u^\top \sum_{t=1}^{T} a_t,\end{aligned}$$

as was to be shown. □

To give some intuition for the condition in Corollary 3.2, note that $w_t = \nabla\psi^*(\beta_t - \eta m_t)$, and so $\psi^*(\beta_t - \eta z_t) \approx \psi^*(\beta_t - \eta m_t) - \eta w_t^\top(z_t - m_t)$. Since ψ^* is convex, we actually have $\psi^*(\beta_t - \eta z_t) \ge \psi^*(\beta_t - \eta m_t) - \eta w_t^\top(z_t - m_t)$, so we can view the subtraction of $\eta^2 a_t$ as a second-order correction that flips the sign of the inequality. The η^2 coefficient in front of a_t is motivated by the fact that the second-order term in the Taylor expansion of $\psi^*(\beta_t - \eta z_t)$ is of order η^2, and so for the $\eta^2 a_t$ term to cancel this out we need a_t to be of constant order.

Adaptive step size. The exposition so far assumes a fixed step size η, and the subsequent bounds we present will assume that the optimal value of η is known. In practice, it is rarely the case that we know this optimal value in advance, and it is thus necessary to choose η adaptively. We ignore this issue in the main text, but an adaptive scheme following Cesa-Bianchi et al. (2007) is provided in the supplementary material for the interested reader. We note that, for the adaptive case, our regret bound is slightly worse and corresponds to the $\max_i D_i$ entry in Figure 1.

Application to exponentiated gradient. Using the adaptive optimistic mirror descent framework, we can now ob-

[3]The typical Bregman divergence bound can be recovered by setting $\psi_{t+1}(w)$ to $\psi_t(w) + D_{\psi^*}(\theta_{t+1}\|\theta_t)$.

tain an adaptive exponentiated gradient algorithm that incorporates hints m_t. The algorithm is obtained from Algorithm 2 by setting $\psi(w) = \sum_{i=1}^n w_i \log(w_i)$ and $a_{t,i} = (z_{t,i} - m_{t,i})^2$. This choice of correction a_t makes intuitive sense, as it will downweight experts i for whom the hints $m_{t,i}$ are inaccurate.

Proposition 3.3 (Adaptive Exponentiated Gradient)**.** *Consider the updates given by $\beta_{1,i} = 0$ and $\beta_{t+1,i} = \beta_{t,i} - \eta z_{t,i} - \eta^2(z_{t,i} - m_{t,i})^2$, with prediction $w_{t,i} \propto \exp(\beta_{t,i} - \eta m_{t,i})$. Then, assuming $\|z_t\|_\infty \leq 1$, $\|m_t\|_\infty \leq 1$ and $0 < \eta \leq \frac{1}{4}$, we have for all $u \in \Delta_n$:*

$$\mathrm{Regret}(u) \leq \frac{\log(n)}{\eta} + \eta \sum_{i=1}^n u_i \sum_{t=1}^T (z_{t,i} - m_{t,i})^2. \quad (18)$$

Proof. Corollary 3.2 reduces the proof to straightforward computation. Note that, for $\psi(w) = \sum_{i=1}^n w_i \log(w_i)$ and w constrained to the simplex Δ_n, $\psi^*(\beta) = \log(\sum_{i=1}^n \exp(\beta_i))$ and $\nabla\psi^*(\beta_t - \eta m_t)$ is equal to w_t as defined in the proposition. The updates above thus correspond to Algorithm 2 and so it suffices to check that the main condition of Corollary 3.2 is satisfied with $a_{t,i} = (z_{t,i} - m_{t,i})^2$. This follows from the calculation:

$$\begin{aligned}
&\psi^*(\beta_t - \eta z_t - \eta^2 a_t) \\
&= \log(\sum_{i=1}^n \exp(\beta_{t,i} - \eta z_{t,i} - \eta^2(z_{t,i} - m_{t,i})^2)) \\
&= \log(\sum_{i=1}^n \exp(\beta_{t,i} - \eta m_{t,i}) \exp(-\eta(z_{t,i} - m_{t,i}) \\
&\qquad\qquad - \eta^2(z_{t,i} - m_{t,i})^2)) \\
&\leq \log(\sum_{i=1}^n \exp(\beta_{t,i} - \eta m_{t,i})(1 - \eta(z_{t,i} - m_{t,i}))) \\
&= \log(\sum_{i=1}^n \exp(\beta_{t,i} - \eta m_{t,i}) \\
&\qquad\qquad - \eta \sum_{i=1}^n \exp(\beta_{t,i} - \eta m_{t,i})(z_{t,i} - m_{t,i})) \\
&\leq \log(\sum_{i=1}^n \exp(\beta_{t,i} - \eta m_{t,i})) \\
&\qquad - \eta \frac{\sum_{i=1}^n \exp(\beta_{t,i} - \eta m_{t,i})(z_{t,i} - m_{t,i})}{\sum_{i=1}^n \exp(\beta_{t,i} - \eta m_{t,i})} \\
&= \psi^*(\beta_t - \eta m_t) - \eta \nabla\psi^*(\beta_t - \eta m_t)^\top (z_t - m_t).
\end{aligned}$$

The two inequalities we made use of were $\exp(-x - x^2) \leq 1 - x$ for $|x| \leq \frac{1}{2}$ and $\log(x - y) \leq \log(x) - y/x$. Having verified the condition of Corollary 3.2, we obtain a regret bound of $\frac{\psi^*(0)+\psi(u)}{\eta} + \eta \sum_{i=1}^n u^\top a_t$. Finally, we note that $\psi^*(0) = \log(n)$, $\psi(u) = \sum_{i=1}^n u_i \log(u_i) \leq 0$, and $a_{t,i} = (z_{t,i} - m_{t,i})^2$, which completes the proof. ☐

Comparison to (MW2). For $m_t = 0$ we obtain the same regret bound (6) that was obtained for the update (MW2). Interestingly, the two updates are essentially the same to second order:

$$\beta_{t+1,i} = \beta_{t,i} - \eta z_{t,i} - \eta^2 z_{t,i}^2 \quad (19)$$

$$\text{versus} \quad \beta_{t+1,i} = \beta_{t,i} + \log(1 - \eta z_{t,i}). \quad (20)$$

Since $-x - x^2 \leq \log(1 - x)$ when $|x| \leq \frac{1}{2}$, we can think of the adaptive EG updates as a second-order under-approximation to (MW2) when $m_t = 0$. The regret bound (6) for (MW2) can be obtained by a near-identical calculation to the one in Proposition 3.3.

Variance bound. By setting $m_t = \frac{1}{t}\sum_{s=1}^{t-1} z_s$, we obtain a *variance bound*

$$\mathrm{Regret} \leq \frac{\log(n)}{\eta} + \eta(2V_{i^*} + 6), \quad (21)$$

where i^* is the best expert and

$$V_i \stackrel{\text{def}}{=} \sum_{t=1}^T (z_{t,i} - \bar{z}_i)^2, \quad \bar{z} \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^T z_t \quad (22)$$

is the *variance* of expert i. This improves the result in Hazan & Kale (2008), who obtain a regret based on $\max_{i=1}^n V_i$ rather than V_{i^*}.[4]

The choice of m_t corresponds to running an auxiliary instance of Follow-the-Regularized-Leader (Shalev-Shwartz, 2011) to minimize the regret bound (18), an idea first introduced by Rakhlin & Sridharan (2012). The details are given in the supplementary material.

Path-length bound. For $m_t = z_{t-1}$ we obtain the algorithm AEG-Path given in Table 1 and achieve the bound

$$\mathrm{Regret} \leq \frac{\log(n)}{\eta} + \eta D_{i^*}, \quad D_i \stackrel{\text{def}}{=} \sum_{t=1}^T (z_{t,i} - z_{t-1,i})^2. \quad (23)$$

This is called a path-length bound because D_i can be thought of as the *path length* (squared) of the losses for expert i. This improves upon the algorithm and bound given in Chiang et al. (2012), where D_i is replaced with the quantity $D_\infty \stackrel{\text{def}}{=} \sum_{t=1}^T \|z_t - z_{t-1}\|_\infty^2$, which is always larger

[4] Actually, their bound is slightly better than that, but the exact bound is difficult to state concisely.

than D_{i^*}. We note that $D_{i^*} \leq 4V_{i^*} + 2$, so path-length bounds subsume variance bounds.

The path-length bound obtained above resolves a problem posed by Kale (2012), who asked whether it is possible to obtain bounds in terms of D_{i^*}.

Comparison of bounds. Recall the definitions of D_i, D_∞, and V_i, and further define V_∞, S_i, and S_∞:

$$\begin{array}{ll} D_i \overset{\text{def}}{=} \sum_{t=1}^T (z_{t,i} - z_{t-1,i})^2 & D_\infty \overset{\text{def}}{=} \sum_{t=1}^T \|z_t - z_{t-1}\|_\infty^2 \\ V_i \overset{\text{def}}{=} \sum_{t=1}^T (z_t - \bar{z}_i)^2 & V_\infty \overset{\text{def}}{=} \sum_{t=1}^T \|z_t - \bar{z}\|_\infty^2 \\ S_i \overset{\text{def}}{=} \sum_{t=1}^T z_{t,i}^2 & S_\infty \overset{\text{def}}{=} \sum_{t=1}^T \|z_t\|_\infty^2 \end{array} \tag{24}$$

Figure 1 shows the 3×3 grid of potential regret bounds, summarizing the relevant results. The original exponentiated gradient algorithm has regret in terms of S_∞, while the adaptive algorithm proposed by Cesa-Bianchi et al. (2007) obtains regret in terms of the smaller quantity S_{i^*}. Hazan & Kale (2008) obtain a bound based on $\max_{i=1}^n V_i$, and Chiang et al. (2012) obtain a bound based on D_∞. All three of these latter bounds are incomparable, but our AEG-Path algorithm obtains a bound in terms of D_{i^*}, which is strictly better than all of the above. We note that in some cases, slightly better bounds can be obtained in terms of the behavior of the learner (see e.g. Section 1.2 of Hazan & Kale (2008)), but we omit these results for brevity and because the behavior of the learner is not known ahead of time.

4. Extension to Matrices

We now extend our results to the matrix setting, where the learner chooses a positive semidefinite matrix W with $\mathrm{tr}(W) = 1$. The flexibility of Corollary 3.2 makes the extension to this case straightforward; essentially the only change is replacing the regularizer $\sum_{i=1}^n w_i \log(w_i)$ with $\mathrm{tr}(W \log(W)) = \sum_{i=1}^n \lambda_i \log(\lambda_i)$, where $(\lambda_i)_{i=1}^n$ are the eigenvalues of W.

Setup. On each round the learner chooses a matrix W_t with $W_t \succeq 0$ and $\mathrm{tr}(W_t) = 1$, and a matrix of losses Z_t is revealed; Z_t is assumed to be symmetric and to satisfy $\|Z_t\|_{\mathrm{op}} \leq 1$, where $\|\cdot\|_{\mathrm{op}}$ is the operator norm (maximum singular value). The loss in round t is $\mathrm{tr}(W_t Z_t)$. Note that we can embed the vector setting in the matrix setting via $w_t \mapsto \mathrm{diag}(w_t)$, $z_t \mapsto \mathrm{diag}(z_t)$, where $\mathrm{diag}(v)$ is the diagonal matrix V with $V_{ii} = v_i$.

To give some intuition, the constraint that $\mathrm{tr}(W) = 1$ means that W can be written as a convex combination $\sum_{i=1}^n p_i v_i v_i^\top$ of unit vectors. The inner product $\mathrm{tr}(WZ)$ can then be written as $\sum_{i=1}^n p_i \cdot (v_i^\top Z v_i)$. Thus an equivalent game would be for the learner to (stochastically) pick a vector v and receive payoff $v^\top Z v$. Here the stochasticity of the choices is crucial because $v^\top Z v$ is not convex (since Z can have negative eigenvalues). See Warmuth & Kuzmin (2006) for more on this interpretation.

We start by extending the adaptive EG algorithm (Proposition 3.3) to the matrix setting:

Proposition 4.1 (Adaptive matrix exponentiated gradient). *For any sequence of matrices M_t, consider the updates given by $B_1 = 0$ and $B_{t+1} = B_t - \eta Z_t - \eta^2 (Z_t - M_t)^2$, with prediction $W_t = \frac{\exp(B_t - \eta M_t)}{\mathrm{tr}(\exp(B_t - \eta M_t))}$. For $0 < \eta \leq \frac{1}{4}$, $\|Z_t\|_{\mathrm{op}} \leq 1$, and $\|M_t\|_{\mathrm{op}} \leq 1$, we have*

$$\mathrm{Regret}(U) \leq \frac{\log(n)}{\eta} + \eta \sum_{i=1}^n \mathrm{tr}(U(Z_t - M_t)^2) \tag{25}$$

for all $U \succeq 0$ with $\mathrm{tr}(U) = 1$.

The main additional tool we need is the Golden-Thompson inequality $\mathrm{tr}(\exp(A+B)) \leq \mathrm{tr}(\exp(A)\exp(B))$ (Golden, 1965; Thompson, 1965). Otherwise, the proof proceeds as in Proposition 3.3, so we leave the details for the supplementary material.

Path-length and variance bounds. By setting M_t to Z_{t-1} as before, we obtain the algorithm AMEG-Path in Table 1 and achieve the following *path-length bound*:

$$\mathrm{Regret}(U) \leq \frac{\log(n)}{\eta} + \eta \sum_{t=1}^T \mathrm{tr}(U(Z_t - Z_{t-1})^2). \tag{26}$$

We now turn our attention to the variance bound. The path length bound already implies a variance bound, but deriving a variance bound directly provides additional insight as well as better constants. Mimicking Rakhlin & Sridharan (2012), we would like to set M_t to $\frac{1}{t}\sum_{s=1}^{t-1} Z_s$ and then interpret this choice of M_t as playing Follow-the-Regularized-Leader (FTRL) to minimize the sum in (25). In previous applications this has been straightforward, but here, due to the adaptivity of the regularizer, the sum (25) is a function of U, which is not known in advance. We address this issue with Lemmas 4.2 and 4.3 below. Lemma 4.3 establishes that there is an optimal value M^* for M_t that is independent of U. Lemma 4.3 provides a way of attaining the optimum; the lemma is fairly general and may be useful in obtaining variance bounds for other adaptive regularizers.

Lemma 4.2. *For any $\delta \geq 0$, define $M^* \overset{\text{def}}{=} \frac{1}{T+\delta}\sum_{t=1}^T Z_t$. Then, for any symmetric matrix M', we have*

$$\delta(M^*)^2 + \sum_{t=1}^T (Z_t - M^*)^2 \preceq \delta(M')^2 + \sum_{t=1}^T (Z_t - M')^2.$$

The proof is in the supplementary material. We remark that the proof is almost purely algebraic, and only relies on the property that $D^2 \succeq 0$ for any symmetric matrix D.

Setting δ to 0, we see that $\bar{Z} = \frac{1}{T}\sum_{t=1}^{T} Z_t$ is the optimal (fixed) value of M_t for any $U \succeq 0$. We now have a target value $\bar{Z}$ for the M_t, but we cannot simply apply the standard FTRL Lemma, since we need a result of the form

$$\sum_{t=1}^{T}(Z_t - M_t)^2 \preceq \sum_{t=1}^{T}(Z_t - \bar{Z})^2 + \alpha I, \tag{27}$$

which cannot be straightforwardly expressed as a regret bound (the αI term is meant to be the matrix equivalent of a small constant α). We deal with this by deriving a generalization of the FTRL algorithm, which we call FTRL-$\mathcal{K}$. This algorithm has vector-valued losses and obtains regret relative to a partial ordering defined by a cone $\mathcal{K}$.[5]

An important notion is that of a *global minimizer*. For a function $f : \mathcal{X} \to V$ where V is a vector space and a cone $\mathcal{K} \subset V$, we say that x is a global minimizer of f relative to $\mathcal{K}$ if $f(x) \leq_{\mathcal{K}} f(y)$ for all $y \in \mathcal{X}$; that is, $x + \mathcal{K}$ contains the image of f. Intuitively, $\mathcal{K}$ must contain all the directions in which f can vary relative to $f(x)$.

Lemma 4.3 (FTRL-$\mathcal{K}$)**.** *Suppose that for all $1 \leq t \leq T+1$, there exists a global minimizer M_t of $\psi(M) + \sum_{s=1}^{t-1} f_s(M)$. Then for all M,*

$$\sum_{t=1}^{T} f_t(M_t) - f_t(M) \leq_{\mathcal{K}} \psi(M) - \psi(M_1) + \sum_{t=1}^{T} f_t(M_t) - f_t(M_{t+1}). \tag{28}$$

Taking $\psi(M) = M^2$, $f_t(M) = (Z_t - M)^2$, and $\mathcal{K}$ the cone of PSD matrices, we obtain the following corollary:

Corollary 4.4. *Suppose that we choose $M_t = \frac{1}{t}\sum_{s=1}^{t-1} Z_s$. Then, assuming $\|Z_t\|_{\mathrm{op}} \leq 1$ for all t, we have*

$$\sum_{t=1}^{T}(Z_t - M_t)^2 \preceq 2\sum_{t=1}^{T}(Z_t - \bar{Z})^2 + 6I, \tag{29}$$

for $\bar{Z} \stackrel{\mathrm{def}}{=} \frac{1}{T}\sum_{t=1}^{T} Z_t$.

Both proofs can be found in the supplementary material. Combining Proposition 4.1 with Corollary 4.4 gives the desired variance bound:

Corollary 4.5. *For $0 < \eta \leq \frac{1}{4}$ and $\|Z_t\|_{\mathrm{op}} \leq 1$, setting $M_t = \frac{1}{t}\sum_{s=1}^{t-1} Z_s$ achieves a bound of*

$$\mathrm{Regret}(U) \leq \frac{\log(n)}{\eta} + \eta\left[2\sum_{t=1}^{T}\mathrm{tr}(U(Z_t - \bar{Z})^2) + 6\right].$$

We remark that by optimizing the proof of Corollary 4.4, we can replace the constants 2 with $1+\epsilon$ for any $\epsilon > 0$.

[5] Recall that for a cone $\mathcal{K}$ satisfying $\mathcal{K} \cap (-\mathcal{K}) = \{0\}$, we define the partial order $x \leq_{\mathcal{K}} y$ iff $y - x \in \mathcal{K}$. Common choices of $\mathcal{K}$ are the positive orthant and the positive semidefinite cone.

5. Discussion

We have presented an adaptive exponentiated gradient algorithm, which attains regret bounded by the variance and path length of the best expert in hindsight. To achieve these bounds, we relied on the synergy of adaptivity and optimism, allowing us to use "hints" for immediate prediction, and adaptively performing a second-order correction to the gradient updates based on the accuracy of the hints. A remaining open problem is to adaptively tune the step size to achieve asymptotically optimal regret.

Recently, Duchi et al. (2011) proposed AdaGrad, an adaptive *subgradient* algorithm. A major difference is that they update their regularizer by a large multiplicative amount in each round, whereas our regularizer changes by a small additive second-order term $\eta^2 u_t$. We also obtain different regret bounds; at a high level, AdaGrad can be expected to perform well when the optimal predictor is dense but the gradient updates are sparse. In contrast, our algorithm will perform well when the optimal predictor is sparse but the gradient updates are dense.

Our FTRL-$\mathcal{K}$ lemma (Lemma 4.3) is closely related to Blackwell approachability (Blackwell, 1956); see Perchet (2013) for a recent survey. As far as we can tell, the conditions in Lemma 4.3 are not equivalent to Blackwell approachability; they are (intuitively) stronger but have the advantage of offering a potentially tighter analysis, as in Corollary 4.4. Abernethy et al. (2011) recently provided a very elegant connection between Blackwell approachability and regret minimization; our algorithm is, however, different from theirs. We note that the global minimizer criterion is essentially a lower bound on the curvature of the cumulative regularized loss near its optimum. We could thus imagine adding to the regularizer term until the criterion held, if necessary.

Finally, we think the general idea of "pushing the regret into the regularizer", as in Theorem 3.1 and in earlier work (Orabona, 2013; Orabona et al., 2013), is quite interesting, as it allows us to obtain regret bounds in terms of the best expert rather than the learner. It should be the case that any time our regret involves a sum $\sum_{t=1}^{T}\|z_t - m_t\|_{w_t}^2$, where $\|\cdot\|_{w_t}$ is a local norm, we can instead obtain a bound on $\mathrm{Regret}(u)$ involving $\sum_{t=1}^{T}\|z_t - m_t\|_u^2$, as long as ψ^* is well-behaved (perhaps having a bounded third derivative). Precisely characterizing these conditions, and obtaining such local norm results for cases beyond the entropy and von-Neumann (matrix) entropy, is an interesting direction of future work.

Acknowledgments. We thank Paul Christiano and Jonathan Huggins for helpful discussions, as well as the anonymous reviewers for providing a thorough review of the paper and several helpful comments. The first author was supported by the Hertz Foundation.

References

Abernethy, Jacob, Bartlett, Peter L, and Hazan, Elad. Blackwell approachability and no-regret learning are equivalent. *JMLR: Workshop and Conference Proceedings (COLT)*, 19:27–46, 2011.

Arora, Sanjeev and Kale, Satyen. A combinatorial, primal-dual approach to semidefinite programs. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pp. 227–236. ACM, 2007.

Arora, Sanjeev, Hazan, Elad, and Kale, Satyen. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

Auer, Peter and Warmuth, Manfred K. Tracking the best disjunction. *Machine Learning*, 32(2):127–150, 1998.

Blackwell, David. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

Cesa-Bianchi, Nicolò, Conconi, Alex, and Gentile, Claudio. A second-order perceptron algorithm. *SIAM Journal on Computing*, 34(3):640–668, 2005.

Cesa-Bianchi, Nicolo, Mansour, Yishay, and Stoltz, Gilles. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2-3):321–352, 2007.

Chiang, Chao-Kai, Yang, Tianbao, Lee, Chia-Jung, Mahdavi, Mehrdad, Lu, Chi-Jen, Jin, Rong, and Zhu, Shenghuo. Online optimization with gradual variations. *Journal of Machine Learning Research*, 2012.

Crammer, Koby, Kulesza, Alex, and Dredze, Mark. Adaptive regularization of weight vectors. *Machine Learning*, pp. 1–33, 2009.

Duchi, John, Hazan, Elad, and Singer, Yoram. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research*, pp. 2121–2159, 2011.

Golden, Sidney. Lower bounds for the helmholtz function. *Physical Review*, 137(4B):B1127, 1965.

Hazan, Elad. The convex optimization approach to regret minimization. *Optimization for machine learning*, pp. 287, 2011.

Hazan, Elad and Kale, Satyen. Extracting certainty from uncertainty: Regret bounded by variation in costs. In *Proceedings of the Twenty First Annual Conference on Computational Learning Theory*, 2008.

Hazan, Elad, Rakhlin, Alexander, and Bartlett, Peter L. Adaptive online gradient descent. In *Advances in Neural Information Processing Systems*, pp. 65–72, 2007.

Hazan, Elad, Kale, Satyen, and Shalev-Shwartz, Shai. Near-optimal algorithms for online matrix prediction. *arXiv preprint arXiv:1204.0136*, 2012.

Herbster, Mark and Warmuth, Manfred K. Tracking the best expert. *Machine Learning*, 32(2):151–178, 1998.

Kale, Satyen. Commentary on "online optimization with gradual variations". *Journal of Machine Learning Research*, pp. 6–24, 2012.

Kivinen, Jyrki and Warmuth, Manfred K. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

Littlestone, Nick. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine learning*, 2(4):285–318, 1988.

Littlestone, Nick and Warmuth, Manfred K. The weighted majority algorithm. In *Foundations of Computer Science, 30th Annual Symposium on*, pp. 256–261. IEEE, 1989.

McMahan, H Brendan and Streeter, Matthew. Adaptive bound optimization for online convex optimization. *arXiv preprint arXiv:1002.4908*, 2010.

Orabona, Francesco. Dimension-free exponentiated gradient. In *Advances in Neural Information Processing Systems*, pp. 1806–1814, 2013.

Orabona, Francesco, Crammer, Koby, and Cesa-Bianchi, Nicolo. A generalized online mirror descent with applications to classification and regression. *arXiv preprint arXiv:1304.2994*, 2013.

Perchet, Vianney. Approachability, regret and calibration; implications and equivalences. *arXiv preprint arXiv:1301.2663*, 2013.

Rakhlin, Alexander and Sridharan, Karthik. Online learning with predictable sequences. *arXiv preprint arXiv:1208.3728*, 2012.

Sabato, Sivan, Shalev-Shwartz, Shai, Srebro, Nathan, Hsu, Daniel, and Zhang, Tong. Learning sparse low-threshold linear classifiers. *arXiv preprint arXiv:1212.3276*, 2012.

Shalev-Shwartz, Shai. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2): 107–194, 2011.

Srebro, Nati, Sridharan, Karthik, and Tewari, Ambuj. On the universality of online mirror descent. In *Advances in Neural Information Processing Systems*, pp. 2645–2653, 2011.

Streeter, Matthew and McMahan, H Brendan. Less regret via online conditioning. *arXiv preprint arXiv:1002.4862*, 2010.

Thompson, Colin J. Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6:1812, 1965.

Tsuda, Koji, Rätsch, Gunnar, and Warmuth, Manfred K. Matrix exponentiated gradient updates for on-line learning and bregman projection. In *Journal of Machine Learning Research*, pp. 995–1018, 2005.

Vaits, Nina, Moroshko, Edward, and Crammer, Koby. Second-order non-stationary online learning for regression. *arXiv preprint arXiv:1303.0140*, 2013.

Warmuth, Manfred K and Kuzmin, Dima. Online variance minimization. In *Learning Theory*, pp. 514–528. Springer, 2006.

Yang, Tianbao, Mahdavi, Mehrdad, Jin, Rong, and Zhu, Shenghuo. Regret bounded by gradual variation for online convex optimization. *Machine Learning*, pp. 1–41, 2013.

Appendix

This appendix contains several pieces of exposition that were removed from the main text due to space constraints. First, in Appendix A, we provide several properties of Fenchel conjugates, which we hope will serve as a useful reference. In Appendix B, we provide proofs for results that were stated without proof in the main text. In Appendix C, we re-prove one of these results in the vector case for the convenience of readers who do not wish to read through matrix manipulations. In Appendix D, we generalize the exponentiated gradient results from the simplex case to the unconstrained case. Finally, in Appendix E, we show how to adaptively control the step size η in our algorithms to obtain regret almost as good as if the optimal η were known in advance.

A. Properties of Fenchel Conjugates

Throughout this paper we make extensive use of properties of Fenchel conjugates. We provide them here for reference. In all cases we assume that ψ is a convex function. We assume that the argument w to ψ is constrained to lie in some convex set S.

A.1. General Properties

Definition. The Fenchel conjugate $\psi^*(\beta)$ of a function $\psi(w)$ is defined as $\psi^*(\beta) \stackrel{\text{def}}{=} \sup_{w \in S} w^\top \beta - \psi(w)$.

Gradient. Let w be the maximizing vector in the preceding definition. Then w is a subgradient of ψ^* at β. If ψ^* is differentiable then

$$\nabla \psi^*(\beta) = \arg\max_{w \in S} w^\top \beta - \psi(w), \tag{30}$$

and in particular $\nabla \psi^*(\beta) \in S$ for all β.

Translations. For any vector c, define $\psi_c(w)$ to be $\psi(w) - w^\top c$. Then $\psi_c^*(\beta) = \psi^*(\beta + c)$.

A.2. Calculations (vector case)

Simplex. Let $S = \Delta_n$ and $\psi(w) = \sum_{i=1} w_i \log(w_i)$. This choice of ψ is also called the negative entropy (as well as, somewhat confusingly, an entropic regularizer). Then we have $\psi^*(\beta) = \log(\sum_{i=1}^n \exp(\beta_i))$ and $\nabla \psi^*(\beta)_i = \frac{\exp(\beta_i)}{\sum_{j=1}^n \exp(\beta_j)}$.

To see the latter, we note that applying the KKT conditions to $w^\top \beta - \psi(w)$ implies that the maximizer (and hence the gradient $\nabla \psi^*(\beta)$) satisfies $\beta_i = \log(w_i) + 1 + \lambda$ for some scalar λ, hence $w_i \propto \exp(\beta_i)$, and so $\nabla \psi^*(\beta_i) = w_i = \frac{\exp(\beta_i)}{\sum_j \exp(\beta_j)}$. Computing $\psi^*(\beta)$ now only involves evaluating $w^\top \beta - \psi(w)$ at its maximizing value, yielding (where we define Z_β as $\sum_j \exp(\beta_j)$)

$$\psi^*(\beta) = \sum_{i=1}^n \left[\beta_i \frac{\exp(\beta_i)}{Z_\beta} - \log(\exp(\beta_i)/Z_\beta) \frac{\exp(\beta_i)}{Z_\beta}\right] \tag{31}$$

$$= \sum_{i=1}^n \frac{\exp(\beta_i)}{Z_\beta} \log(Z_\beta) \tag{32}$$

$$= \log(Z_\beta), \tag{33}$$

which completes the calculation.

Non-negative orthant. If instead S is the non-negative orthant and we now take $\psi(w) = \sum_{i=1}^n w_i \log(w_i)$, we will have $\psi^*(\beta) = \sum_{i=1}^n \exp(\beta_i)$ and $\nabla \psi^*(\beta)_i = \exp(\beta_i)$.

To see this, again apply the KKT conditions to $w^\top \beta - \psi(w)$, which imply that the maximizing value of w satisfies

$\beta_i = \log(w_i)$, and hence $\nabla\psi^*(\beta)_i = w_i = \exp(\beta_i)$. Evaluating $w^\top \beta - \psi(w)$ at this point yields

$$\psi^*(\beta) = \sum_{i=1}^{n} [\beta_i \exp(\beta_i) - \beta_i \exp(\beta_i) + \exp(\beta_i)] \tag{34}$$

$$= \sum_{i=1}^{n} \exp(\beta_i), \tag{35}$$

thus completing the calculation.

A.3. Calculations (matrix case)

Trace constrained. Let $S = \{W \mid W \succeq 0, \mathrm{tr}(W) = 1\}$ and let $\psi(W) = \mathrm{tr}(W \log(W))$. This choice of ψ is called the von-Neumann entropy. We have $\psi^*(B) = \log(\mathrm{tr}(\exp(B)))$ and $\nabla\psi^*(B) = \frac{\exp(B)}{\mathrm{tr}(\exp(B))}$. Note that in this case $\psi^*(B)$ is defined as $\sup_{W \in S} \mathrm{tr}(WB) - \mathrm{tr}(W \log(W))$.

To calculate $\nabla\psi^*$, note that the KKT conditions yield $B = \log(W) + (1 + \lambda)I$ for the maximizing value of W. Thus $W \propto \exp(B)$ and hence $\nabla\psi^*(B) = W = \frac{\exp(B)}{\mathrm{tr}(\exp(B))}$. Defining Z_B to be $\mathrm{tr}(\exp(B))$ and plugging back in yields

$$\psi^*(B) = \mathrm{tr}(B \exp(B))/Z_B - \mathrm{tr}(\exp(B)[B - \log(Z_B)I])/Z_B \tag{36}$$

$$= \mathrm{tr}(\exp(B)) \log(Z_B)/Z_B \tag{37}$$

$$= \log(Z_B), \tag{38}$$

which completes the calculations for the trace-constrained case.

Trace unconstrained. Let $S = \{W \mid W \succeq 0\}$ and let $\psi(W) = \mathrm{tr}(W \log(W) - W)$. We have $\psi^*(B) = \mathrm{tr}(\exp(B))$ and $\nabla\psi^*(B) = \exp(B)$.

To calculate $\nabla\psi^*$, note that the KKT conditions yield $B = \log(W)$ and hence $\nabla\psi^*(B) = W = \exp(B)$. Plugging back in to ψ^* yields

$$\psi^*(B) = \mathrm{tr}(B \exp(B)) - \mathrm{tr}(\exp(B) \log(\exp(B)) - \exp(B)) \tag{39}$$

$$= \mathrm{tr}(\exp(B)), \tag{40}$$

which completes the calculations for the unconstrained case.

B. Deferred Proofs

In this section we prove all results stated in the main text that were deferred to the supplementary material.

Proof of Proposition 2.2. We will construct two sequences $(z_t)_{t=1}^T$ such that exponentiated gradient with any fixed step size η will perform poorly ($\Omega(\sqrt{T})$) on at least one of them. Our constructed sequences will involve $n = 2$ experts. In both sequences, the first expert has $z_{t,1} = 0$ for all t, and $z_{t,2}$ will satisfy $\sum_{t=1}^T z_{t,2} \geq 0$ to ensure quasi-realizability.

Sequence 1. The second expert has loss $z_{t,2} = (-1)^{t-1}$. Then $\sum_{t=1}^T z_{t,2}$ is either 0 or 1 depending on the parity of T, and in particular is non-negative. On odd-numbered rounds, $w_t = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}^\top$, and on even-numbered rounds, $w_t = \begin{bmatrix} \frac{1}{1+\exp(-\eta)} & \frac{1}{1+\exp(\eta)} \end{bmatrix}^\top$. Assume that $\eta \leq 1$. The total loss (and hence regret) of the learner is then at least

$$\sum_{k=1}^{\lfloor \frac{T}{2} \rfloor} \frac{1}{2} - \frac{1}{1 + \exp(\eta)} = \left\lfloor \frac{T}{2} \right\rfloor \left(\frac{1}{2} - \frac{1}{1 + \exp(\eta)} \right) \tag{41}$$

$$\geq \left\lfloor \frac{T}{2} \right\rfloor \left(\frac{1}{2} - \frac{1}{2 + 2\eta} \right) \tag{42}$$

$$\geq \frac{1}{4} \left\lfloor \frac{T}{2} \right\rfloor \eta. \tag{43}$$

So, for any $\eta \leq 1$, there is a quasi-realizable sequence with regret at least $\frac{1}{4}\left\lfloor\frac{T}{2}\right\rfloor\eta$. Since (41) can be seen to be an increasing function of η, we have a lower bound of $\frac{1}{4}\left\lfloor\frac{T}{2}\right\rfloor\min(\eta, 1)$. The point is that for large η, the learner will pay heavily because it switches around too much.

Sequence 2. On the other hand, we consider the sequence given by $z_{t,2} = 1$ for all t. Then $w_{t,2} = \frac{1}{1+\exp((t-1)\eta)}$, which for $t \leq \left\lceil\frac{1}{\eta}\right\rceil$ is at least $\frac{1}{1+e}$. Therefore, the regret of the learner on this sequence is at least $\frac{1}{1+e}\min\left(T, \frac{1}{\eta}\right)$. The point is that for small η, the learner will pay heavily because it can't decrease the weight on expert 2 fast enough.

Combining these together, we see that the first sequence inflicts a regret of $\Omega(\sqrt{T})$ whenever $\eta \geq 1/\sqrt{T}$, whereas the second sequence inflicts a regret of $\Omega(1/\sqrt{T})$ whenever $\eta \leq 1/\sqrt{T}$. Since one of these two conditions on η must always be satisfied, one of these sequences will always inflict regret $\Omega(1/\sqrt{T})$, thus proving the proposition. $\square$

Proof of Proposition 4.1. As noted in the main text, the proof parallels Proposition 3.3, with the main new tool being the Golden-Thompson inequality, which says that $\mathrm{tr}(\exp(A+B)) \leq \mathrm{tr}(\exp(A)\exp(B))$ (Golden, 1965; Thompson, 1965).

When $\psi(W) = \mathrm{tr}(W\log(W))$ and W is constrained to have trace 1, we have $\psi^*(B) = \log(\mathrm{tr}(\exp(B)))$ and $\nabla\psi^*(B) = \frac{\exp(B)}{\mathrm{tr}(\exp(B))}$, so that $\nabla\psi^*(B_t - \eta M_t)$ matches W_t as given in the proposition. So, again, we are performing an instance of Algorithm 2 and it suffices to check that the condition of Corollary 3.2 is satisfied for $A_t = (Z_t - M_t)^2$. To do so, we use the Golden-Thompson inequality together with the fact that $-X - X^2 \preceq \log(I - X)$ for $-\frac{1}{2}I \preceq X \preceq \frac{1}{2}I$. We have

$$\begin{aligned}
&\psi^*(B_t - \eta Z_t - \eta^2 A_t)\\
&= \log(\mathrm{tr}(\exp(B_t - \eta Z_t - \eta^2(Z_t - M_t)^2)))\\
&\leq \log(\mathrm{tr}(\exp(B_t - \eta M_t)\exp(-\eta(Z_t - M_t) - \eta^2(Z_t - M_t)^2)))\\
&\leq \log(\mathrm{tr}(\exp(B_t - \eta M_t)(I - \eta(Z_t - M_t))))\\
&= \log(\mathrm{tr}(\exp(B_t - \eta M_t)) - \eta\,\mathrm{tr}(\exp(B_t - \eta M_t)(Z_t - M_t)))\\
&\leq \log(\mathrm{tr}(\exp(B_t - \eta M_t))) - \eta\frac{\mathrm{tr}(\exp(B_t - \eta M_t)(Z_t - M_t))}{\mathrm{tr}(\exp(B_t - \eta M_t))}\\
&= \psi^*(B_t - \eta M_t) - \eta\langle\nabla\psi^*(B_t - \eta M_t), Z_t - M_t\rangle.
\end{aligned}$$

This verifies the condition of Corollary 3.2, so that we have a regret bound of $\frac{\psi^*(0)+\psi(U)}{\eta} + \eta\sum_{t=1}^T \mathrm{tr}(UA_t)$. Finally, noting that $\psi^*(0) = \log(n)$, $\psi(U) = \mathrm{tr}(U\log(U)) \leq 0$, and $A_t = (Z_t - M_t)^2$ completes the proof. $\square$

Proof of Lemma 4.2. Write $M' = M^* + D$. Then we have

$$\delta(M')^2 + \sum_{t=1}^T (Z_t - M')^2 \tag{44}$$

$$= \delta(M^* + D)^2 + \sum_{t=1}^T (Z_t - M^* - D)^2 \tag{45}$$

$$= \delta(M^*)^2 + \sum_{t=1}^T (Z_t - M^*)^2 + \left[\delta M^* + \sum_{t=1}^T (M^* - Z_t)\right] D + D\left[\delta M^* + \sum_{t=1}^T (M^* - Z_t)\right] + (T+\delta)D^2 \tag{46}$$

$$= \delta(M^*)^2 + \sum_{t=1}^T (Z_t - M^*)^2 + (T+\delta)D^2 \tag{47}$$

$$\succeq \delta(M^*)^2 + \sum_{t=1}^T (Z_t - M^*)^2, \tag{48}$$

which completes the lemma. $\square$

Proof of Lemma 4.3. The proof is structurally identical to the vector case (see Hazan (2011) for a proof of the vector case). We will prove the lemma by induction on T. Note that the lemma is equivalent to showing that

$$\psi(M_1) + \sum_{t=1}^{T} f_t(M_{t+1}) \leq_{\mathcal{K}} \psi(M) + \sum_{t=1}^{T} f_t(M) \tag{49}$$

for all M. In the base case $T = 0$, we have

$$\psi(M_1) \leq_{\mathcal{K}} \psi(M), \tag{50}$$

which follows from the fact that M_1 is a global minimizer of ψ and hence $\psi(M_1) \leq_{\mathcal{K}} \psi(M)$ for all M. For the inductive step, suppose that

$$\sum_{t=1}^{T-1} f_t(M_{t+1}) \leq_{\mathcal{K}} \psi(M) + \sum_{t=1}^{T-1} f_t(M) \tag{51}$$

for all M, and invoke this for the particular choice $M = M_{T+1}$. Then we have

$$\psi(M_1) + \sum_{t=1}^{T} f_t(M_{t+1}) = \psi(M_1) + \left[\sum_{t=1}^{T-1} f_t(M_{t+1})\right] + f_T(M_{T+1}) \tag{52}$$

$$\leq_{\mathcal{K}} \psi(M_{T+1}) + \left[\sum_{t=1}^{T-1} f_t(M_{T+1})\right] + f_T(M_{T+1}) \tag{53}$$

$$= \psi(M_{T+1}) + \sum_{t=1}^{T} f_t(M_{T+1}) \tag{54}$$

$$\leq_{\mathcal{K}} \psi(M) + \sum_{t=1}^{T} f_t(M) \tag{55}$$

for all M, where we use the fact that M_{T+1} is a global minimizer of $\psi(M) + \sum_{t=1}^{T} f_t(M)$ for the last inequality. This completes the induction and hence the proof. $\square$

Proof of Corollary 4.4. The key tool is the *matrix Young's inequality*: $AB + BA \preceq \frac{1}{\gamma}A^2 + \gamma B^2$ for all symmetric A, B and all $\gamma > 0$. (This follows immediately upon expanding $(A/\sqrt{\gamma} - \sqrt{\gamma}B)^2 \succeq 0$.) We then note that, by Lemma 4.2, M_t obeys Lemma 4.3 with $\psi(M) = M^2$, $f_t(M) = (M - Z_t)^2$, and $\mathcal{K}$ the cone of positive semidefinite matrices. Therefore:

$$\sum_{t=1}^{T} (Z_t - M_t)^2 - (Z_t - \bar{Z})^2 \preceq \bar{Z}^2 + \sum_{t=1}^{T} (Z_t - M_t)^2 - (Z_t - M_{t+1})^2 \tag{56}$$

$$= \bar{Z}^2 + \sum_{t=1}^{T} \left[Z_t(M_{t+1} - M_t) + (M_{t+1} - M_t)Z_t + M_t^2 - M_{t+1}^2\right] \tag{57}$$

$$= \bar{Z}^2 + M_1^2 - M_{T+1}^2 + \sum_{t=1}^{T} \left[Z_t(M_{t+1} - M_t) + (M_{t+1} - M_t)Z_t\right] \tag{58}$$

$$= \bar{Z}^2 + M_1^2 - M_{T+1}^2 + \sum_{t=1}^{T} \frac{1}{t+1}\left[Z_t(Z_t - M_t) + (Z_t - M_t)Z_t\right] \tag{59}$$

(since $M_{t+1} = \frac{1}{t+1}Z_t + \frac{t}{t+1}M_t$)

$$\preceq I + \sum_{t=1}^{T} \frac{Z_t^2}{\gamma(t+1)^2} + \gamma(Z_t - M_t)^2 \tag{60}$$

$$\preceq I + \frac{I}{\gamma} + \gamma \sum_{t=1}^{T} (Z_t - M_t)^2. \tag{61}$$

(For the second-to-last inequality, note that $M_1 = 0$ and hence $M_1^2 - M_{T+1}^2 \preceq 0$.) Re-arranging yields

$$\sum_{t=1}^{T} (Z_t - M_t)^2 \leq \frac{1}{1-\gamma} \left(\frac{1+\gamma}{\gamma} I + \sum_{t=1}^{T} (Z_t - \bar{Z})^2 \right). \tag{62}$$

Setting γ to $\frac{1}{2}$ gives the desired result. Note that by instead setting γ to $\frac{\epsilon}{2}$, we can replace the constants 2 and 6 by $1+\epsilon$ and $\frac{6}{\epsilon}$ for any $\epsilon \leq 1$. □

C. Improved Variance Bound

We claimed in Section 3 that we could obtain a regret bound in terms of $2V_i + 6$ by using the optimistic prediction based on $m_t = \frac{1}{t}\sum_{s=1}^{t-1} z_s$. The following proposition establishes this. Its proof is essentially the same as that of Corollary 4.5, and in fact is implied by Corollary 4.5. The only purpose of this section is to keep proofs accessible to readers who prefer not to read through algebraic manipulations of matrices.

Proposition C.1. *Suppose that we choose $m_{t,i} = \frac{1}{t}\sum_{s=1}^{t-1} z_{s,i}$ and that $\|z_s\|_\infty \leq 1$. Then for all i and all $0 < \epsilon \leq 1$ we have*

$$\sum_{t=1}^{T} (z_{t,i} - m_{t,i})^2 \leq 2 \sum_{t=1}^{T} (z_{t,i} - m_i^*)^2 + 6. \tag{63}$$

Proof. Note that $m_{t,i}$ is the minimizer of $m_i^2 + \sum_{s=1}^{t-1} (z_{s,i} - m_i)^2$. Therefore, by the FTRL Lemma (Hazan, 2011), we have

$$\sum_{t=1}^{T} (z_{t,i} - m_{t,i})^2 - (z_{t,i} - m_i^*)^2 \leq (m_i^*)^2 + \sum_{t=1}^{T} (z_{t,i} - m_{t,i})^2 - (z_{t,i} - m_{t+1,i})^2 \tag{64}$$

$$= (m_i^*)^2 + \sum_{t=1}^{T} 2z_{t,i}(m_{t+1,i} - m_{t,i}) + m_{t,i}^2 - m_{t+1,i}^2 \tag{65}$$

$$= (m_i^*)^2 + m_{1,i}^2 - m_{T+1,i}^2 + \sum_{t=1}^{T} \frac{2}{t+1} z_{t,i}(z_{t,i} - m_{t,i}) \tag{66}$$

$$\leq 1 + \sum_{t=1}^{T} \frac{z_{t,i}^2}{\gamma(t+1)^2} + \gamma(z_{t,i} - m_{t,i})^2 \tag{67}$$

$$\leq 1 + \frac{1}{\gamma} + \gamma \sum_{t=1}^{T} (z_{t,i} - m_{t,i})^2. \tag{68}$$

Re-arranging yields

$$\sum_{t=1}^{T} (z_{t,i} - m_{t,i})^2 \leq \frac{1}{1-\gamma} \left(\frac{1+\gamma}{\gamma} + \sum_{t=1}^{T} (z_{t,i} - m_i^*)^2 \right). \tag{69}$$

Setting γ to $\frac{1}{2}$ then yields the desired result. □

D. Bounds for Exponentiated Gradient in the Unconstrained Case

The main text contained an analysis of adaptive versions of the exponentiated gradient and matrix exponentiated gradient algorithms. However, this analysis was for the case that the weights were constrained to the simplex (or that $\mathrm{tr}(W) = 1$ in the case of matrices). In Section 2 we promised to include an analysis of these algorithms in the unconstrained case, and we do so here. Note that this "unconstrained case" still has the constraint $w \geq 0$ (or $W \succeq 0$ for matrices), although this is not a serious limitation since we can split w into its positive and negative components (see Kivinen & Warmuth (1997) for details).

The updates and proofs are almost identical. The major difference is in the initialization, where to obtain good bounds we need to initialize $\beta_{1,i}$ to $-\log(n)$ rather than 0 (in the matrix case, we need to initialize B_1 to $-\log(n)I$). The complete algorithms are shown below:

Exponentiated Gradient:

$$\begin{aligned} \beta_{1,i} &= -\log(n) \\ w_{t,i} &= \exp(\beta_{t,i} - \eta m_{t,i}) \\ \beta_{t+1,i} &= \beta_{t,i} - \eta z_{t,i} - \eta^2 (z_{t,i} - m_{t,i})^2 \end{aligned} \tag{70}$$

Matrix Exponentiated Gradient:

$$\begin{aligned} B_1 &= -\log(n)I \\ W_t &= \exp(B_t - \eta M_t) \\ B_{t+1} &= B_t - \eta Z_t - \eta^2 (Z_t - M_t)^2 \end{aligned} \tag{71}$$

We have the following regret bounds in the vector and matrix cases:

Proposition D.1. *For $\|z_t\|_\infty \leq 1$, $\|m_t\|_\infty \leq 1$, and $0 < \eta \leq \frac{1}{4}$, the unconstrained exponentiated gradient updates (70) achieve the bound*

$$\mathrm{Regret}(u) \leq \frac{1 + (\log(n) - 1)\|u\|_1 + \sum_{i=1}^n u_i \log(u_i)}{\eta} + \eta \sum_{i=1}^n u_i \sum_{t=1}^T (z_{t,i} - m_{t,i})^2. \tag{72}$$

Proposition D.2. *For $\|Z_t\|_{\mathrm{op}} \leq 1$, $\|M_t\|_{\mathrm{op}} \leq 1$, and $0 < \eta \leq \frac{1}{4}$, the unconstrained matrix exponentiated gradient updates (71) achieve the bound*

$$\mathrm{Regret}(U) \leq \frac{1 + (\log(n) - 1)\,\mathrm{tr}(U) + \mathrm{tr}(U \log(U))}{\eta} + \eta \sum_{t=1}^T \mathrm{tr}(U(Z_t - M_t)^2). \tag{73}$$

The proofs are basically identical to the proofs of Propositions 3.3 and 4.1, but we include them for completeness.

Proof of Proposition D.1. We note that, for $\psi(w) = \sum_{i=1}^n w_i \log(w_i) - w_i$ and w constrained to be non-negative, $\psi^*(\beta) = \sum_{i=1}^n \exp(\beta_i)$ and $\nabla \psi^*(\beta_t - \eta m_t)$ is equal to w_t as defined in the proposition. It therefore suffices to check that the condition of Corollary 3.2 is satisfied with $a_{t,i} = (z_{t,i} - m_{t,i})^2$. We have

$$\psi^*(\beta_t - \eta z_t - \eta^2 a_t) = \sum_{i=1}^n \exp(\beta_{t,i} - \eta z_{t,i} - \eta^2 (z_{t,i} - m_{t,i})^2) \tag{74}$$

$$= \sum_{i=1}^n \exp(\beta_{t,i} - \eta m_{t,i}) \exp(-\eta(z_{t,i} - m_{t,i}) - \eta^2 (z_{t,i} - m_{t,i})^2) \tag{75}$$

$$\leq \sum_{i=1}^n \exp(\beta_{t,i} - \eta m_{t,i})(1 - \eta(z_{t,i} - m_{t,i})) \tag{76}$$

$$= \sum_{i=1}^{n} \exp(\beta_{t,i} - \eta m_{t,i}) - \eta \sum_{i=1}^{n} \exp(\beta_{t,i} - \eta m_{t,i})(z_{t,i} - m_{t,i}) \tag{77}$$

$$= \psi^*(\beta_t - \eta m_t) - \eta \nabla\psi^*(\beta_t - \eta m_t)^\top (z_t - m_t). \tag{78}$$

The one inequality we made use of was $\exp(-x - x^2) \leq 1 - x$ for $|x| < \frac{1}{2}$. This verifies the condition of Corollary 3.2, yielding a regret bound of $\frac{\psi^*(\beta_1)+\psi(u)-u^\top\beta_1}{\eta} + \eta \sum_{i=1}^{n} u^\top a_t$. Finally, we note that $\psi^*(\beta_1) = \sum_{i=1}^{n} \exp(-\log(n)) = 1$, $\psi(u) - u^\top \beta_1 = \sum_{i=1}^{n} u_i \log(u_i) + (\log(n) - 1)u_i$, and $a_{t,i} = (z_{t,i} - m_{t,i})^2$, which completes the proof. □

Proof of Proposition D.2. When $\psi(W) = \mathrm{tr}(W \log(W) - W)$ and W is constrained to be positive semidefinite, we have $\psi^*(B) = \log(\mathrm{tr}(\exp(B)))$ and $\nabla\psi^*(B) = \exp(B)$, so that $\nabla\psi^*(B_t - \eta M_t)$ matches W_t as given in the proposition. So, again, it suffices to check that the condition of Corollary 3.2 is satisfied for $A_t = (Z_t - M_t)^2$. To do so, we need to make use of the Golden-Thompson inequality $\mathrm{tr}(\exp(A + B)) \preceq \mathrm{tr}(\exp(A)\exp(B))$ (Golden, 1965; Thompson, 1965), together with the fact that $-X - X^2 \preceq \log(I - X)$ for $-\frac{1}{2}I \preceq X \preceq \frac{1}{2}I$. We then have

$$\psi^*(B_t - \eta Z_t - \eta^2 A_t) = \mathrm{tr}(\exp(B_t - \eta Z_t - \eta^2 (Z_t - M_t)^2)) \tag{79}$$

$$\leq \mathrm{tr}(\exp(B_t - \eta M_t) \exp(-\eta(Z_t - M_t) - \eta^2 (Z_t - M_t)^2)) \tag{80}$$

$$\leq \mathrm{tr}(\exp(B_t - \eta M_t)(I - \eta(Z_t - M_t))) \tag{81}$$

$$= \mathrm{tr}(\exp(B_t - \eta M_t)) - \eta\, \mathrm{tr}(\exp(B_t - \eta M_t)(Z_t - M_t)) \tag{82}$$

$$= \psi^*(B_t - \eta M_t) - \eta \langle \nabla\psi^*(B_t - \eta M_t), Z_t - M_t \rangle. \tag{83}$$

This verifies the condition of Corollary 3.2, so that we have a regret bound of $\frac{\psi^*(B_1)+\psi(U)-\mathrm{tr}(B_1 U)}{\eta} + \eta \sum_{t=1}^{T} \mathrm{tr}(U A_t)$. Finally, noting that $\psi^*(B_1) = \mathrm{tr}(\frac{1}{n} I) = 1$, $\psi(U) - \mathrm{tr}(B_1 U) = \mathrm{tr}(U \log(U)) + (\log(n) - 1)\,\mathrm{tr}(U)$, and $A_t = (Z_t - M_t)^2$ completes the proof. □

E. Adaptive Step Size

In this section we show how to obtain an adaptive version of Algorithm 2, which relies on the standard doubling trick. The adaptive algorithm is given as Algorithm 3. The regret bound of this procedure when applied to learning from experts is worse than in the non-adaptive case, depending (in the language of Figure 1 and (24)) on $\max_i D_i$ rather than D_{i^*} (in other words, the maximum path length of any expert rather than the path length of the best expert).

The algorithm basically calls Algorithm 2 repeatedly with different step sizes, halving the step size every time the regret exceeds a certain bound. For this algorithm we require a bound B on the inner product term $u^\top z_t$ and a bound C on the regularizer term in the regret bound. Cesa-Bianchi et al. (2007) proposed an adaptive step size scheme in the learning from experts setting that does not require knowledge of B. It would be interesting to apply the same ideas here, but we have not tried to do so, although the exposition given below follows Section 3.1 of the same paper.

The regret of Algorithm 3 is bounded in the following theorem:

Theorem E.1. *Let $u_t \in \arg\min_u u^\top \sum_{s=1}^{t} z_s$ and let $Q_t = u_t^\top \sum_{s=1}^{t} a_s$. Let $Q = \max\left(B, \max_{t=1}^{T} Q_t\right)$. Then the regret of Algorithm 3 is bounded as*

$$\mathrm{Regret} \leq B \left\lceil 1 + \log\left(\frac{Q}{B}\right) \right\rceil + 10\sqrt{CQ}. \tag{84}$$

Proof. First note η is monotonically non-increasing across rounds, and decays by a factor of 2 every time it changes. We can group the rounds based on what value of η was used in that round; in this way, Algorithm 3 is equivalent to running several sub-algorithms, each of which is an instance of Algorithm 2. The total regret is then bounded above by the sum of the regrets of these individual algorithms.

Now consider the rounds when η is equal to $2^{-j}\sqrt{\frac{C}{B}}$. Let t_j be the final such round. By construction, we must have $u_{t-1}^\top \sum_{s=1}^{t-1} a_s \leq 4^{j+1} B$, or else we would have already decreased η by the next factor of 2. Let Regret_j denote the regret

Algorithm 3 Adaptive Step Size Mirror Descent

Given: convex regularizer ψ, corrections a_t, hints m_t, and β
Let B be any bound on $\max_{t=1}^{T} u^\top z_t$
Let C be any upper bound on $\psi^*(\beta_1) + \psi(u) - u^\top \beta_1$
$Q, \eta, t \leftarrow B, \sqrt{\frac{C}{B}}, 1$
while there are rounds remaining **do**
  $\beta_t \leftarrow \beta$
  while $\sqrt{\frac{C}{Q}} \geq \frac{\eta}{2}$ **do**
    Choose $w_t = \nabla\psi^*(\beta_t - \eta_t m_t)$
    Observe z_t and suffer loss $w_t^\top z_t$
    Update $\beta_{t+1} = \beta_t - \eta_t z_t - \eta_t^2 a_t$
    Let $u_t \in \arg\min_u u^\top \sum_{s=1}^{t} z_s$
    $Q \leftarrow \max(Q, u_t^\top \sum_{s=1}^{t} a_s)$
    $t \leftarrow t + 1$
  end while
  $\eta \leftarrow \frac{\eta}{2}$
end while

of the sub-algorithm on this set of rounds. Note that it is bounded above by B plus the regret on all but the last of these rounds. Then we have

$$\begin{aligned} \mathrm{Regret}_j &\leq B + \frac{C}{\eta} + \eta u_{t-1}^\top \sum_{s=1}^{t-1} a_s && (85) \\ &= B + 2^j\sqrt{CB} + 2^{-j}4^{j+1}\sqrt{CB} && (86) \\ &= B + 5 \cdot 2^j \sqrt{CB} && (87) \\ &\leq B + 5\sqrt{CQ_{t_j}}. && (88) \end{aligned}$$

Note that $\sqrt{Q_{t_j}} \geq 2\sqrt{Q_{t_{j-1}}}$ by construction. Then we have

$$\begin{aligned} \mathrm{Regret} &\leq \sum_j \mathrm{Regret}_j && (89) \\ &\leq \sum_j B + 5\sqrt{CQ_{t_j}} && (90) \\ &\leq B\left\lceil 1 + \log\left(\frac{Q}{B}\right)\right\rceil + 10\sqrt{CQ}, && (91) \end{aligned}$$

as was to be shown. $\square$