Proceedings of Machine Learning Research vol 75:1–26, 2018

The Many Faces of Exponential Weights in Online Learning

Dirk van der Hoeven DIRKVDERHOEVEN@GMAIL.COM
and **Tim van Erven** TIM@TIMVANERVEN.NL
Statistics Department, Leiden University, the Netherlands

Wojciech Kotłowski WKOTLOWSKI@CS.PUT.POZNAN.PL
Institute of Computing Science, Poznań University of Technology, Poland

Abstract

A standard introduction to online learning might place Online Gradient Descent at its center and then proceed to develop generalizations and extensions like Online Mirror Descent and second-order methods. Here we explore the alternative approach of putting Exponential Weights (EW) first. We show that many standard methods and their regret bounds then follow as a special case by plugging in suitable surrogate losses and playing the EW posterior mean. For instance, we easily recover Online Gradient Descent by using EW with a Gaussian prior on linearized losses, and, more generally, all instances of Online Mirror Descent based on regular Bregman divergences also correspond to EW with a prior that depends on the mirror map. Furthermore, appropriate quadratic surrogate losses naturally give rise to Online Gradient Descent for strongly convex losses and to Online Newton Step. We further interpret several recent adaptive methods (iProd, Squint, and a variation of Coin Betting for experts) as a series of closely related reductions to exp-concave surrogate losses that are then handled by Exponential Weights. Finally, a benefit of our EW interpretation is that it opens up the possibility of sampling from the EW posterior distribution instead of playing the mean. As already observed by Bubeck and Eldan, this recovers the best-known rate in Online Bandit Linear Optimization.

1. Introduction

Exponential Weights (EW) (Vovk, 1990; Littlestone and Warmuth, 1994) is a method for keeping track of uncertainty about the best action in sequential prediction tasks. It is most commonly considered for a finite number of actions in the prediction with expert advice setting, where each of the actions corresponds to following the advice of one of a finite number of experts, and in this context it is asymptotically minimax optimal (Cesa-Bianchi and Lugosi, 2006, Section 2.2). However, in the present work we mostly consider EW on continuous action spaces in the more general setting of Online Convex Optimization (Hazan, 2016), where we show that surprisingly many standard methods turn out to be special cases of EW.

EW keeps track of a probability distribution over actions that is updated in each round of the prediction task by multiplying the probability of each action by a factor that is exponentially decreasing in the action's error or *loss* in that round, and renormalizing. This type of update is quite flexible: by assigning appropriate surrogate losses to the actions, it covers any kind of multiplicative probability updates, including, for instance, those of the Prod algorithm (Cesa-Bianchi et al., 2007). For best performance, losses often need to be scaled by a positive parameter called the learning rate, and the algorithm may also be biased towards particular actions by the choice of its initial distribution, which is called the prior. For continuous sets of actions, efficient implementations of EW

are often restricted to conjugate priors for which the EW distribution can be analytically computed, but sampling approximations based on random walks can also provide appealing trade-offs between computational complexity and prediction accuracy, even for a single random walk step per round (Narayanan and Rakhlin, 2017; Kalai and Vempala, 2002).

The usual presentation of Online Convex Optimization would introduce EW as a special case of Mirror Descent (MD) or Follow-the-Regularized-Leader (FTRL) with the Kullback-Leibler divergence as the regularizer. However, here we turn this view on its head and show that all instances of MD based on regular Bregman divergences (Banerjee et al., 2005) in fact correspond to EW on a continuous set of actions (Section 3.3). In particular, Gradient Descent (GD) comes from using a Gaussian prior on linearized losses (Section 3.2), which is striking because GD has been contrasted with the Exponentiated Gradient Plus-Minus algorithm (Kivinen and Warmuth, 1997) that is readily seen to be an instance of EW (Section 3.1). In addition, the unnormalized relative entropy regularizer (Helmbold and Warmuth, 2009), which is normally considered a generalization of EW, turns out to be a special case of EW as well for a multivariate Poisson prior (Section 3.3). Furthermore, in Section 4 we show that running EW on suitable quadratic approximations of the losses recovers Gradient Descent for strongly convex losses (Hazan et al., 2007) and, as already observed by Van Erven and Koolen (2016), Online Newton Step (Hazan et al., 2007). The Vovk-Azoury-Warmuth forecaster would also be an example of running EW on quadratic losses, but we refer to (Vovk, 2001) for its analysis, which requires a generalized proof technique (see also the discussion by Orabona et al. (2015)). We do consider the recent adaptive iProd, Squint and Coin Betting methods of Koolen and Van Erven (2015); Orabona and Pál (2016), which learn the optimal learning rate for prediction with expert advice, and show that these may also be viewed as running EW after a reduction of the original prediction task to various closely related surrogate tasks in which the learning rate is just one of the parameters that does not need to be treated specially (Section 5). Finally, in the context of Bandit Linear Optimization, the SCRiBLe method (Abernethy et al., 2008) may be viewed as an approximation to EW, and an application of EW outlined by Bubeck and Eldan (2015) achieves the best-known rate (we provide the technical details they omit in Section 6).

Related Work The diverse applications of EW on a finite number of actions range, for instance, from boosting (Freund and Schapire, 1997) to differential privacy (Dwork and Roth, 2014) to multi-armed bandits (Auer et al., 2002), and many algorithms in computer science can be viewed as special cases of EW (Arora et al., 2012). EW has also been considered for continuous sets of actions, often in the context of universal coding in information theory, where the goal is to sequentially compress a sequence of symbols. In this case, actions parametrize a set of probability distributions and the loss of an action is the logarithmic loss for the corresponding probability distribution on the symbol that is being compressed (Cesa-Bianchi and Lugosi, 2006, Chapter 9). EW (with learning rate 1) then simplifies to Bayesian probability updating. The choice of prior has received much attention in this literature, with Jeffreys' prior being shown to be asymptotically minimax optimal for exponential families with parameters restricted to suitable bounded sets (Grünwald, 2007, Chapter 8). Without parameter restrictions, Jeffreys' prior is still minimax optimal up to constants for the Bernoulli and multinomial models (Krichevsky and Trofimov, 1981; Xie and Barron, 2000). Several applications to other losses are also closely related to the log loss: Online Ridge Regression corresponds to EW on the squared loss, which matches the log loss for Gaussian distributions; and Cover's method for portfolio selection (Cover, 1991), which is EW on Cover's loss, may be interpreted as learning a mixture model under the log loss (Orseau et al., 2017). In general, continuous EW is not restricted

Input: a convex set of distributions $\mathcal{P}$ over $\boldsymbol{w}$, a prior $P_1 \in \mathcal{P}$ and learning rates $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_T > 0$	
Lazy Exponential Weights	Greedy Exponential Weights
$\tilde{P}_{t+1} = \arg\min_P \ \mathop{\mathbb{E}}_P \left[\sum_{s=1}^t f_s(\boldsymbol{w})\right] + \frac{1}{\eta_t} \mathrm{KL}(P\|P_1)$	$\tilde{P}_{t+1} = \arg\min_P \ \mathop{\mathbb{E}}_P[f_t(\boldsymbol{w})] + \frac{1}{\eta_t} \mathrm{KL}(P\|P_t)$
$P_{t+1} = \arg\min_{P\in\mathcal{P}} \ \mathrm{KL}(P\|\tilde{P}_{t+1})$	$P_{t+1} = \arg\min_{P\in\mathcal{P}} \ \mathrm{KL}(P\|\tilde{P}_{t+1})$

Figure 1: The lazy and greedy versions of Exponential Weights

to the log loss, however, and has been considered e.g. for general convex losses (Dick et al., 2014) or as a computationally inefficient gold standard for exp-concave losses (Hazan et al., 2007).

2. Exponential Weights

In Online Convex Optimization (OCO) (Shalev-Shwartz, 2011; Hazan, 2016) a learner repeatedly chooses actions $\boldsymbol{w}_t$ from a convex set $\mathcal{W} \subseteq \mathbb{R}^d$ during rounds $t = 1, \ldots, T$, and suffers losses $f_t(\boldsymbol{w}_t)$, where $f_t : \mathcal{W} \to \mathbb{R}$ is a convex function. The learner's goal is to achieve small *regret* $\mathcal{R}_T(\boldsymbol{u}) = \sum_{t=1}^T f_t(\boldsymbol{w}_t) - \sum_{t=1}^T f_t(\boldsymbol{u})$ with respect to any comparator action $\boldsymbol{u} \in \mathcal{W}$, which measures the difference between the cumulative loss of the learner and the cumulative loss it could have achieved by playing the oracle action $\boldsymbol{u}$ from the start. We will assume the domain of the losses f_t is extended from $\mathcal{W}$ to $\mathbb{R}^d$ with convexity of f_t being preserved. This comes without loss of generality as one can always set $f_t(\boldsymbol{w}) = \infty$ outside $\mathcal{W}$, but we will use more natural and straightforward extensions throughout the paper (e.g. when the f_t are linear or quadratic functions).

The central topic of this work is the Exponential Weights (EW) algorithm, which keeps track of uncertainty over actions expressed by a distribution P_t and comes in the two flavors shown in Figure 1 (our naming follows Zinkevich (2003)), where we let $\mathrm{KL}(P\|Q) = \mathbb{E}_P\left[\ln \frac{\mathrm{d}P}{\mathrm{d}Q}\right]$ denote the Kullback-Leibler (KL) divergence between distributions P and Q. The algorithm gets its name from the distributions $\tilde{P}_t$, whose densities have the following exponential forms:

$$\mathrm{d}\tilde{P}_{t+1}(\boldsymbol{w}) = \frac{e^{-\eta_t \sum_{s=1}^t f_s(\boldsymbol{w})}\,\mathrm{d}P_1(\boldsymbol{w})}{\int e^{-\eta_t \sum_{s=1}^t f_s(\boldsymbol{w})}\,\mathrm{d}P_1(\boldsymbol{w})} \qquad \text{(lazy EW)} \qquad (1)$$

$$\mathrm{d}\tilde{P}_{t+1}(\boldsymbol{w}) = \frac{e^{-\eta_t f_t(\boldsymbol{w})}\,\mathrm{d}P_t(\boldsymbol{w})}{\int e^{-\eta_t f_t(\boldsymbol{w})}\,\mathrm{d}P_t(\boldsymbol{w})} \qquad \text{(greedy EW).} \qquad (2)$$

In the case that $\mathcal{P}$ contains all possible distributions over $\mathbb{R}^d$ (for which the projection step becomes void) and the *learning rates* η_t are constant $\eta_1 = \cdots = \eta_T = \eta$, both versions of EW are equivalent. In general they differ, and enjoy the following regret bounds with respect to a potentially randomized comparator drawn from a comparator distribution Q, which follow from a standard MD analysis (Hazan, 2016) and a reformulation of the standard FTRL analysis that works for distributions P_t on continuous spaces, which cannot be expressed as the finite-dimensional vectors that are usually assumed (the proof details are in Appendix A):

Lemma 1 (EW Regret) *Suppose that $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_T > 0$, and that the minima that define $\tilde{P}_t$ and P_t are uniquely achieved. Let $Q \in \mathcal{P}$ be any comparator distribution such that $\mathrm{KL}(Q\|\tilde{P}_t) < \infty$*

for all t, let $\{\boldsymbol{w}_t \in \mathcal{W}\}_{t=1}^T$ be the actions of any learner, and define $\eta_0 \stackrel{\text{def}}{=} \eta_1$. Then EW satisfies

$$\mathop{\mathbb{E}}_{\boldsymbol{u}\sim Q}[\mathcal{R}(\boldsymbol{u})] \le \frac{1}{\eta_T}\mathrm{KL}(Q\|P_1) + \sum_{t=1}^T \underbrace{\left\{ f_t(\boldsymbol{w}_t) + \frac{1}{\eta_{t-1}} \ln \mathop{\mathbb{E}}_{P_t(\boldsymbol{w})}\left[e^{-\eta_{t-1} f_t(\boldsymbol{w})}\right] \right\}}_{\text{“mixability gap”}} \quad \text{(lazy EW)} \quad (3)$$

$$\mathop{\mathbb{E}}_{\boldsymbol{u}\sim Q}[\mathcal{R}(\boldsymbol{u})] \le \frac{1}{\eta_1}\mathrm{KL}(Q\|P_1) + \left(\frac{1}{\eta_T} - \frac{1}{\eta_1}\right) \max_{t=2,\ldots,T} \mathrm{KL}(Q\|P_t)$$
$$+ \sum_{t=1}^T \underbrace{\left\{ f_t(\boldsymbol{w}_t) + \frac{1}{\eta_t} \ln \mathop{\mathbb{E}}_{P_t(\boldsymbol{w})}\left[e^{-\eta_t f_t(\boldsymbol{w})}\right] \right\}}_{\text{“mixability gap”}} \quad \text{(greedy EW).} \quad (4)$$

While the predictions $\boldsymbol{w}_t$ in Lemma 1 are arbitrary actions from $\mathcal{W}$, one always chooses $\boldsymbol{w}_t$ to be some function of P_t. A general mapping from P_t to $\boldsymbol{w}_t$ is called a *substitution function* (Vovk, 2001) and is usually designed to give the best bound on the mixability gap in trial t. Throughout the paper, we will use the mean $\boldsymbol{w}_t = \mathbb{E}_{P_t}[\boldsymbol{w}]$ as our substitution function, which is a typical choice, although alternatives may be better in specific cases (Vovk, 2001). To ensure that $\boldsymbol{w}_t \in \mathcal{W}$, we will also generally assume that $\mathcal{P} = \{P : \mathbb{E}_P[\boldsymbol{w}] \in \mathcal{W}\}$, which is convex.

Bounding the mixability gap is a crucial part of the regret analysis of EW (Vovk, 2001; De Rooij et al., 2014). In the special case that the losses are *α-exp-concave* for $\alpha > 0$ (i.e. if $e^{-\alpha f(\boldsymbol{w})}$ is concave), the mixability gap for $\eta_t \le \alpha$ is at most 0. This happens in the following example.

Example 1 (The Krichevsky-Trofimov Estimator) *Let $\mathcal{W} = [0,1]$ and let the loss function be the log loss: $f_t(w) = -x_t \ln(w) - (1-x_t)\ln(1-w)$, where $x_t \in \{0,1\}$. A standard algorithm in this case is the Krichevsky-Trofimov forecaster $w_t = (\sum_{s=1}^{t-1} x_s + \frac{1}{2})/t$ (Cesa-Bianchi and Lugosi, 2006, Chapter 9), which is is well known to be the mean $w_t = \mathbb{E}_{P_t}[w]$ of non-projected EW with a $\beta(\frac{1}{2},\frac{1}{2})$ prior and a fixed learning rate $\eta_t = 1$. For the log loss, the mixability gap is 0. To bound the remaining terms in Lemma 1, we choose $Q = P_{T+1}$, which gives:*

$$\sum_{t=1}^T f_t(w_t) \le \mathop{\mathbb{E}}_{P_{T+1}(w)}\left[\sum_{t=1}^T f_t(w)\right] + \mathrm{KL}(P_{T+1}\|P_1) = -\ln \mathop{\mathbb{E}}_{P_1(w)}[w^{\sum_{t=1}^T x_t}(1-w)^{T-\sum_{t=1}^T x_t}]$$
$$\le -\ln \max_w \left\{ w^{\sum_{t=1}^T x_t}(1-w)^{T-\sum_{t=1}^T x_t} \right\} + \ln(2\sqrt{T}) = \min_w \sum_{t=1}^T f_t(w) + \ln(2\sqrt{T}),$$

where the last inequality holds by (Cesa-Bianchi and Lugosi, 2006, Lemma 9.3).

For most regret bounds derived from Lemma 1 the structure of the proof remains the same: we need both a bound on the mixability gap, and a choice for Q for which the expected loss under Q together with $\mathrm{KL}(Q\|P_1)$ can be related to the loss of a deterministic comparator.

3. Linearized Losses

A standard approach in OCO is to lower-bound the convex losses f_t by their tangent at $\boldsymbol{w}_t$, which leads to the following upper bound on the regret in terms of the linearized surrogate losses $\ell_t(\boldsymbol{w}) =$

$\langle \boldsymbol{w}, \boldsymbol{g}_t \rangle$, where $\boldsymbol{g}_t = \nabla f_t(\boldsymbol{w}_t) = (g_{t,1}, \dots, g_{t,d})^\intercal$ is the gradient at $\boldsymbol{w}_t$:

$$\sum_{t=1}^{T} \big(f_t(\boldsymbol{w}_t) - f_t(\boldsymbol{u})\big) \le \sum_{t=1}^{T} \big(\ell_t(\boldsymbol{w}_t) - \ell_t(\boldsymbol{u})\big). \tag{5}$$

3.1. Exponentiated Gradient Plus-Minus as Exponential Weights

The Exponentiated Gradient Plus-Minus (EG$^\pm$) algorithm (Kivinen and Warmuth, 1997) starts with weight vectors $\boldsymbol{w}_t^- = \boldsymbol{w}_t^+ = (1/d, \dots, 1/d) \in \mathbb{R}^d$, which are updated according to

$$w_{t+1,i}^+ = \frac{w_{t,i}^+ e^{-\eta_t \langle \boldsymbol{e}_i, \boldsymbol{g}_t \rangle}}{\sum_{j=1}^d \big(w_{t,j}^+ e^{-\eta_t \langle \boldsymbol{e}_j, \boldsymbol{g}_t \rangle} + w_{t,j}^- e^{\eta_t \langle \boldsymbol{e}_j, \boldsymbol{g}_t \rangle}\big)}, \quad w_{t+1,i}^- = \frac{w_{t,i}^- e^{\eta_t \langle \boldsymbol{e}_i, \boldsymbol{g}_t \rangle}}{\sum_{j=1}^d \big(w_{t,j}^+ e^{-\eta_t \langle \boldsymbol{e}_j \boldsymbol{g}_t \rangle} + w_{t,j}^- e^{\eta_t \langle \boldsymbol{e}_j, \boldsymbol{g}_t \rangle}\big)},$$

and predicts by $\boldsymbol{w}_t \in \{\boldsymbol{w} : \|\boldsymbol{w}\|_1 \le 1\}$ with components $w_{t,i} = w_{t,i}^+ - w_{t,i}^-$.

This is readily seen to be the mean $\boldsymbol{w}_t = \mathbb{E}_{P_t}[\boldsymbol{w}]$ of EW (without projections) on the linearized losses (5) with a discrete uniform prior P_1 on the standard basis vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_d$, which form the corners of the probability simplex, and their negations $-\boldsymbol{e}_1, \dots, -\boldsymbol{e}_d$. The regular Exponentiated Gradient algorithm is recovered by initializing $\boldsymbol{w}_1^- = (0, \dots, 0)$, which corresponds to placing prior mass only on $\boldsymbol{e}_1, \dots, \boldsymbol{e}_d$. Kivinen and Warmuth (1997) also extend the algorithm to scale up the domain by a factor $M > 0$, which corresponds to a discrete prior on $M\boldsymbol{e}_1, \dots, M\boldsymbol{e}_d$ for EG and also on $-M\boldsymbol{e}_1, \dots, -M\boldsymbol{e}_d$ for EG$^\pm$. Hence we may analyze these methods using Lemma 1, which leads to the following regret bound for EG$^\pm$ (see Appendix B):

Theorem 2 (EG$^\pm$ as EW) *Suppose $\|\boldsymbol{g}_t\|_\infty \le G$ for all t. Then the regret of EG$^\pm$ for scale factor $M > 0$ and constant learning rate $\eta_t = \sqrt{\frac{2\ln(2d)}{TM^2G^2}}$ satisfies*

$$\mathcal{R}_T(\boldsymbol{u}) \le GM\sqrt{2T\ln(2d)} \qquad \text{for all } \boldsymbol{u} \text{ such that } \|\boldsymbol{u}\|_1 \le M.$$

3.2. Gradient Descent as Exponential Weights

The prior of EG$^\pm$ is adapted to comparators $\boldsymbol{u}$ with small L_1-norm. How do we change the prior to favor comparators with small L_2-norm? A natural and computationally efficient choice is to use a Gaussian prior $P_1 = \mathcal{N}(\boldsymbol{w}_1, \sigma^2 \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix. Then it turns out that all EW distributions P_t are Gaussian with the Gradient Descent (GD) predictions as their means:

Theorem 3 (Gradient Descent as EW) *Let $\mathcal{P} = \{P : \mathbb{E}_P[\boldsymbol{w}] \in \mathcal{W}\}$. Then, for Gaussian prior $P_1(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}_1, \sigma^2 \boldsymbol{I})$, lazy and greedy EW with learning rates η_t on the linearized losses (5) yield Gaussian distributions $\tilde{P}_t = \mathcal{N}(\tilde{\boldsymbol{w}}_t, \sigma^2 \boldsymbol{I})$ and $P_t = \mathcal{N}(\boldsymbol{w}_t, \sigma^2 \boldsymbol{I})$ with the same covariance as the prior. The means $\tilde{\boldsymbol{w}}_t$ and $\boldsymbol{w}_t$ coincide with lazy and greedy GD (Figure 2), except that the learning rates in GD are scaled to $\sigma^2\eta_t$ by the prior variance σ^2. Moreover, Lemma 1 directly implies:*

$$\mathcal{R}_T(\boldsymbol{u}) \le \frac{\|\boldsymbol{u} - \boldsymbol{w}_1\|_2^2}{2\sigma^2\eta_T} + \frac{\sigma^2}{2}\sum_{t=1}^{T} \eta_{t-1}\|\boldsymbol{g}_t\|_2^2 \qquad \textit{(lazy GD)}$$

$$\mathcal{R}_T(\boldsymbol{u}) \le \frac{\max_t \|\boldsymbol{u} - \boldsymbol{w}_t\|_2^2}{2\sigma^2\eta_T} + \frac{\sigma^2}{2}\sum_{t=1}^{T} \eta_t\|\boldsymbol{g}_t\|_2^2 \qquad \textit{(greedy GD).}$$

Input: Convex set $\mathcal{W}$ and learning rates $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_T > 0$	
Lazy Gradient Descent	Greedy Gradient Descent
$\tilde{\boldsymbol{w}}_{t+1} = \boldsymbol{w}_1 - \eta_t \sum_{s=1}^{t} \boldsymbol{g}_s$	$\tilde{\boldsymbol{w}}_{t+1} = \boldsymbol{w}_t - \eta_t \boldsymbol{g}_t$
$\boldsymbol{w}_{t+1} = \operatorname*{arg\,min}_{\boldsymbol{w} \in \mathcal{W}} \frac{1}{2}\|\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1}\|_2^2$	$\boldsymbol{w}_{t+1} = \operatorname*{arg\,min}_{\boldsymbol{w} \in \mathcal{W}} \frac{1}{2}\|\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1}\|_2^2$

Figure 2: The lazy and greedy versions of Gradient Descent

We note that in this case the parametrization of EW is redundant, because changing the prior variance σ^2 has the same effect on the predictions $\boldsymbol{w}_t$ and the regret bounds as scaling all η_t.

Proof $\tilde{P}_t = \mathcal{N}(\tilde{\boldsymbol{w}}_t, \sigma^2 \boldsymbol{I})$ may be verified analytically from (1) and (2). The fact that the projections P_t onto $\mathcal{P}$ preserve Gaussianity with the same covariance matrix is a property of projecting a member of an exponential family onto a set of distributions defined by a convex constraint on their means. (This follows from Lemma 11 in Appendix C or see (Van Erven and Koolen, 2016, Lemma 9) for the Gaussian case.) The regret bounds follow by taking $Q = \mathcal{N}(\boldsymbol{u}, \sigma^2 \boldsymbol{I})$, for which $\mathrm{KL}(Q\|P_t) = \frac{1}{2\sigma^2}\|\boldsymbol{u} - \boldsymbol{w}_t\|_2^2$, and evaluating the mixability gap in closed form. ■

3.3. Mirror Descent and FTRL as EW

The fact that Gradient Descent is an instance of EW raises the question of whether other instances of MD or FTRL are special cases of EW as well. Let $F^*(\boldsymbol{w}) = \sup_{\boldsymbol{\theta}} \langle \boldsymbol{w}, \boldsymbol{\theta} \rangle - F(\boldsymbol{\theta})$ denote the convex conjugate of F, and let $B_{F^*}(\boldsymbol{u}\|\boldsymbol{w}) = F^*(\boldsymbol{u}) - F^*(\boldsymbol{w}) - \nabla F^*(\boldsymbol{w})^\intercal(\boldsymbol{u} - \boldsymbol{w})$ denote the corresponding Bregman divergence. Then MD and FTRL are defined in Figure 3 for Legendre functions $F(\boldsymbol{\theta})$ on $\mathbb{R}^d$ (Cesa-Bianchi and Lugosi, 2006). We consider exponential families that take the form $\mathcal{E} = \{P_{\boldsymbol{\theta}} \mid \mathrm{d}P_{\boldsymbol{\theta}}(\boldsymbol{w}) = e^{\langle \boldsymbol{\theta}, \boldsymbol{w} \rangle - F(\boldsymbol{\theta})} \mathrm{d}K(\boldsymbol{w}), \boldsymbol{\theta} \in \Theta\}$ for a nonnegative *carrier measure* K, cumulant generating function $F(\boldsymbol{\theta}) = \ln \int e^{\langle \boldsymbol{\theta}, \boldsymbol{w} \rangle} \mathrm{d}K(\boldsymbol{w})$ and parameter space $\Theta = \{\boldsymbol{\theta} \mid F(\boldsymbol{\theta}) < \infty\} \subset \mathbb{R}^d$. These are called *regular* if Θ is an open set. We then start with the following relation between MD and EW, which is proved in Appendix C:

Theorem 4 (Mirror Descent as EW) *Suppose F is the cumulant generating function of a regular exponential family $\mathcal{E}$. Then the lazy and greedy versions of MD predict with the means $\boldsymbol{w}_t = \mathbb{E}_{P_t}[\boldsymbol{w}]$ of lazy and greedy EW on the linearized losses (5) with the same η_t, prior $P_{\boldsymbol{\theta}_1}$ for $\boldsymbol{\theta}_1 = \nabla F^*(\boldsymbol{w}_1)$ and $\mathcal{P} = \{P : \mathbb{E}_P[\boldsymbol{w}] \in \mathcal{W}\}$.*

To answer our question, we therefore need to know whether, for any Legendre function F^*, the convex conjugate $(F^*)^* = F$ corresponds to the cumulant generating function of some exponential family, which means we need to find a corresponding carrier K. Nonconstructive existence of such K has been studied by Banerjee et al. (2005, Theorem 6), who show that there is in fact a bijection between *regular* Bregman divergences and regular exponential families, where regular Bregman divergences based on F^* are defined to be those for which $e^{F(\boldsymbol{\theta})}$ is a continuous, exponentially convex[1] function such that $\Theta = \{\boldsymbol{\theta} \mid F(\boldsymbol{\theta}) < \infty\}$ is open and F is strictly convex.

There is no easy general procedure to construct the corresponding carrier K for a given Legendre function F^*. However, for the Gradient Descent example from Section 3.2 we see that $F^*(\boldsymbol{w}) = \frac{1}{2\sigma^2}\|\boldsymbol{w}\|_2^2$ is the convex conjugate of the cumulant generating function for $K(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. We also give another example:

1. Exponentially convex in the sense of Banerjee et al. (2005, Definition 7).

Input: Legendre function F, convex set $\mathcal{W}$, and learning rates $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_T > 0$	
FTRL / Lazy Mirror Descent	Greedy Mirror Descent
$\tilde{\boldsymbol{w}}_{t+1} = \arg\min_{\boldsymbol{w}} \sum_{s=1}^{t} \langle \boldsymbol{w}, \boldsymbol{g}_s \rangle + \frac{1}{\eta_t} B_{F^*}(\boldsymbol{w} \| \boldsymbol{w}_1)$	$\tilde{\boldsymbol{w}}_{t+1} = \arg\min_{\boldsymbol{w}} \langle \boldsymbol{w}, \boldsymbol{g}_t \rangle + \frac{1}{\eta_t} B_{F^*}(\boldsymbol{w} \| \boldsymbol{w}_t)$
$\boldsymbol{w}_{t+1} = \arg\min_{\boldsymbol{w} \in \mathcal{W}} B_{F^*}(\boldsymbol{w} \| \tilde{\boldsymbol{w}}_{t+1})$,	$\boldsymbol{w}_{t+1} = \arg\min_{\boldsymbol{w} \in \mathcal{W}} B_{F^*}(\boldsymbol{w} \| \tilde{\boldsymbol{w}}_{t+1})$.

Figure 3: The lazy and greedy versions of Mirror Descent. Lazy MD is usually called FTRL.

Input: Convex set $\mathcal{W}$ and learning rate $\eta > 0$	
Lazy EW Gaussian prior quadratic loss	Greedy EW Gaussian prior quadratic loss
$\Sigma_{t+1}^{-1} = \Sigma_t^{-1} + \eta \boldsymbol{M}_t$	$\Sigma_{t+1}^{-1} = \Sigma_t^{-1} + \eta \boldsymbol{M}_t$
$\tilde{\boldsymbol{w}}_{t+1} = \tilde{\boldsymbol{w}}_t - \eta \Sigma_{t+1} \boldsymbol{g}_t$	$\tilde{\boldsymbol{w}}_{t+1} = \boldsymbol{w}_t - \eta \Sigma_{t+1} \boldsymbol{g}_t$
$\boldsymbol{w}_{t+1} = \arg\min_{\boldsymbol{w} \in \mathcal{W}} (\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1})^\intercal \Sigma_{t+1}^{-1} (\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1})$	$\boldsymbol{w}_{t+1} = \arg\min_{\boldsymbol{w} \in \mathcal{W}} (\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1})^\intercal \Sigma_{t+1}^{-1} (\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1})$

Figure 4: The means and covariances of both versions of Exponential Weights with a multivariate normal prior and a constant learning rate η run on the quadratic surrogate loss (6)

Example 2 (Unnormalized Relative Entropy) *Consider MD with regularization based on the* unnormalized relative entropy $B_{F^*}(\boldsymbol{w} \| \boldsymbol{u}) = \sum_{i=1}^{d} (w_i \ln \frac{w_i}{u_i} - w_i + u_i)$ *for* $\boldsymbol{w}, \boldsymbol{u} \in \mathbb{R}_+^d$*, which is the Bregman divergence generated by* $F^*(\boldsymbol{w}) = \sum_{i=1}^{d} w_i(\ln(w_i) - 1)$ *(Cesa-Bianchi and Lugosi, 2006). We have* $F(\boldsymbol{\theta}) = \sum_{i=1}^{d} e^{\theta_i}$*. Interestingly, the exponential family with this cumulant generating function is the set of Poisson distributions, extended i.i.d. to* d *dimensions. To see this for* $d = 1$*, note that if we start with the usual parametrization of Poisson, we have*

$$P_\lambda(w) = e^{-\lambda} \frac{\lambda^w}{w!} = \frac{1}{w!} e^{-\lambda + w \ln \lambda} \qquad \text{on } w \in \{0, 1, 2, \ldots\},$$

for which the natural parameter is $\theta = \ln \lambda$ *and we see that the cumulant generating function is* $F(\theta) = \lambda = e^\theta$*. Thus, EW with the product prior* $P_1(\boldsymbol{w}) = \prod_{i=1}^{d} P_{\lambda_i}(w_i)$ *corresponds to MD with unnormalized relative entropy, where we need to set* $(\lambda_1, \ldots, \lambda_d) = \exp(\boldsymbol{\theta}_1) = \exp(\nabla F^*(\boldsymbol{w}_1)) = \boldsymbol{w}_1$ *to match the starting point of MD:* $\mathbb{E}_{P_1}[\boldsymbol{w}] = \boldsymbol{w}_1$*. Note that in this case the EW distributions* P_t *are discrete.*

4. Quadratic Losses

In this section we assume that the losses f_t satisfy quadratic lower bounds:

$$f_t(\boldsymbol{w}) - f_t(\boldsymbol{w}_t) \geq \langle \boldsymbol{w} - \boldsymbol{w}_t, \boldsymbol{g}_t \rangle + \frac{1}{2} (\boldsymbol{w} - \boldsymbol{w}_t)^\intercal \boldsymbol{M}_t (\boldsymbol{w} - \boldsymbol{w}_t) =: \ell_t(\boldsymbol{w}), \tag{6}$$

where $\boldsymbol{M}_t$ is a positive semi-definite matrix. Generalizing the results from Section 3, EW with Gaussian prior on the surrogate loss ℓ_t yields explicitly computable Gaussian distributions P_t (see also Van Erven and Koolen, 2016; Koolen, 2016):

Theorem 5 *Let* $P_1 = \mathcal{N}(\boldsymbol{w}_1, \Sigma_1)$*. Both versions of the Exponential Weights algorithm, run on* ℓ_t *with learning rate* η *and* $\mathcal{P} = \{P : \mathbb{E}_P[\boldsymbol{w}] \in \mathcal{W}\}$*, yield a multivariate normal distribution* $P_{t+1} =$

$\mathcal{N}(\boldsymbol{w}_{t+1}, \Sigma_{t+1})$ with mean and covariance matrix given in Figure 4. Furthermore, Lemma 1 implies that for all $\boldsymbol{u} \in \mathbb{R}^d$ both versions of EW satisfy:

$$\mathcal{R}_T(\boldsymbol{u}) \leq \frac{1}{2\eta}(\boldsymbol{w}_1 - \boldsymbol{u})^\intercal \Sigma_1^{-1}(\boldsymbol{w}_1 - \boldsymbol{u}) + \frac{\eta}{2}\sum_{t=1}^{T} \boldsymbol{g}_t^\intercal \Sigma_{t+1} \boldsymbol{g}_t. \tag{7}$$

The proof of Theorem 5 in Appendix D.1 is a straightforward generalization of Theorem 3 for constant learning rate $\eta_t = \eta$, which is recovered with $\boldsymbol{M}_t = \boldsymbol{0}$. Like in Theorem 3, the parametrization by η and σ^2 is redundant in that only the product $\eta\sigma^2$ affects the predictions $\boldsymbol{w}_t$ or the bound (7).

4.1. Gradient Descent: Quadratic Approximation of Strongly Convex Losses

For α-strongly convex loss functions, (6) holds with $\boldsymbol{M}_t = \alpha \boldsymbol{I}$. The standard approach for these loss functions is to use greedy Gradient Descent with a time-varying learning rate $\eta_t = 1/(\alpha t)$ (Hazan et al., 2007). Interestingly, greedy GD with the closely related choice $\eta_t = 1/(\frac{1}{\eta\sigma^2} + \alpha t)$ turns out to be a special case of greedy EW with *fixed* learning rate η and prior $P_1 = \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. Applying Theorem 5 results in the following corollary, proved in Appendix D.2:

Corollary 6 *Suppose $\|\boldsymbol{u}\|_2 \leq D$ and $\|\boldsymbol{g}_t\|_2 \leq G$. Then the regret of both versions of the Exponential Weights algorithm with prior $\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ and constant learning rate η, run on the surrogate loss (6) with $\boldsymbol{M}_t = \alpha \boldsymbol{I}$, satisfies:*

$$\mathcal{R}_T(\boldsymbol{u}) \leq \frac{G^2}{2\alpha} \ln\left(\frac{\frac{1}{\eta\sigma^2} + \alpha T}{\frac{1}{\eta\sigma^2} + \alpha}\right) + \frac{G^2}{\frac{2}{\eta\sigma^2} + 2\alpha} + \frac{D^2}{2\eta\sigma^2}.$$

The standard learning rate and corresponding regret bound for GD (Hazan et al., 2007) correspond to the limiting case $\eta\sigma^2 \to \infty$. Formally speaking, this case is not covered here, but for $\eta \to \infty$ EW reduces to Follow-the-Leader (on the surrogate loss (6)), and taking $\sigma^2 \to \infty$ would lead to EW with an *improper prior*, which becomes a proper EW posterior P_2 after one round.

4.2. Online Newton Step: Quadratic Approximation of Exp-concave Losses

For α-exp-concave loss functions, (6) holds with $\boldsymbol{M}_t = \beta \boldsymbol{g}_t \boldsymbol{g}_t^\intercal$, where $\beta = \frac{1}{2}\min\{\frac{1}{4GB}, \alpha\}$, assuming $\|\boldsymbol{g}_t\|_2 \leq G$ and $B = \max_{\boldsymbol{w},\boldsymbol{u}\in\mathcal{W}} \|\boldsymbol{w} - \boldsymbol{u}\|_2$ (Hazan et al., 2007, Lemma 3). Running Exponential Weights on $\ell_t(\boldsymbol{w})$ with prior $\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ leads to the Online Newton Step algorithm (Hazan et al., 2007) with the following regret bound, shown in Appendix D.3:

Corollary 7 *Suppose $\|\boldsymbol{u}\|_2 \leq D$ and $\|\boldsymbol{g}_t\|_2 \leq G$. Then the regret of both versions of the Exponential Weights algorithm with prior $\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ and learning rate η, run on the surrogate loss (6) with $\boldsymbol{M}_t = \beta \boldsymbol{g}_t \boldsymbol{g}_t^\intercal$, satisfies:*

$$\mathcal{R}_T(\boldsymbol{u}) \leq \frac{d}{2\beta} \ln\left(1 + \frac{\eta\sigma^2 \beta G^2 T}{d}\right) + \frac{D^2}{2\eta\sigma^2}. \tag{8}$$

The results of Hazan et al. (2007) correspond to setting $\eta\sigma^2 = \beta D^2$, together with some simplifying upper bounds on (8).

5. Adaptivity by Reduction to Exponential Weights

In this section we show how several recent adaptive methods in the prediction with experts setting — namely iProd (Koolen and Van Erven, 2015), Squint (Koolen and Van Erven, 2015) and a variation of Coin Betting for experts (Orabona and Pál, 2016) –, whose original analyses seem unrelated at first sight, can all be viewed as applying exponential weights after reductions of the original OCO task to various closely related surrogate OCO tasks. The known regret bounds for these methods are also recovered from the reductions upon plugging in regret bounds for EW in the surrogate tasks.

5.1. Reduction for iProd

The experts setting consists of linear losses $f_t(\boldsymbol{w}) = \langle \boldsymbol{w}, \boldsymbol{g}_t \rangle$ over the simplex $\mathcal{W} = \{\boldsymbol{w} : w_i \geq 0, \sum_{i=1}^d w_i = 1\}$, with $g_{t,i} \in [0,1]$. The instantaneous regret in round t with respect to expert i is $r_t(i) = f_t(\boldsymbol{w}_t) - f_t(\boldsymbol{e}_i)$ and $\mathcal{R}_T(i) = \sum_{t=1}^T r_t(i)$ is the total regret. iProd achieves a second-order regret bound in terms of the data-dependent quantity $\mathcal{V}_T(i) = \sum_{t=1}^T r_t(i)^2$, which is much smaller than the worst-case regret in many common cases (Koolen et al., 2016).

In the surrogate OCO task for iProd, predictions take the form of joint distributions P_t on (η, i) for $\eta \in [0,1]$ and $i \in \{1, \dots, d\}$. These map back to predictions in the original task via

$$\boldsymbol{w}_t = \frac{\mathbb{E}_{P_t}[\eta \boldsymbol{e}_i]}{\mathbb{E}_{P_t}[\eta]}, \tag{9}$$

which is like the marginal mean of P_t on experts, except that it is *tilted* to favor larger η. The surrogate loss in the surrogate task is

$$\ell_t(\eta, i) = -\ln\left(1 + \eta r_t(i)\right), \tag{10}$$

and our aim will be to achieve small *mix-regret* with respect to any comparator distribution Q on (η, i), which we define as $S(Q) = \sum_{t=1}^T - \ln \mathbb{E}_{P_t}\left[e^{-\ell_t(\eta,i)}\right] - \mathbb{E}_Q\left[\sum_{t=1}^T \ell_t(\eta, i)\right]$. The mix-regret allows exponential mixing of predictions according to P_t just like for exp-concave losses, so there is no mixability gap to pay. Exponential weights with constant learning rate 1 on the losses ℓ_t therefore achieves $S(Q) \leq \mathrm{KL}(Q\|P_1)$ for any Q.[2] The resulting predictions $\boldsymbol{w}_t$ are those of the iProd algorithm. As shown in Appendix E.1, they achieve the following regret bound, which depends on the surrogate regret of EW:

Theorem 8 (iProd Reduction to EW) *Restrict the domain for η to $[0, \frac{1}{2}]$. Then any choice of P_t in the surrogate OCO task defined above induces regret bounded by*

$$\mathop{\mathbb{E}}_{Q}[\eta] \sum_{t=1}^T f_t(\boldsymbol{w}_t) - \mathop{\mathbb{E}}_{Q}\left[\eta \sum_{t=1}^T f_t(\boldsymbol{e}_i)\right] \leq \mathop{\mathbb{E}}_{Q}\left[\eta^2 \mathcal{V}_T(i)\right] + S(Q) \qquad \text{for any } Q \text{ on } (\eta, i) \tag{11}$$

in the original prediction with expert advice task.

In particular, if we use EW in the surrogate OCO task with learning rate 1 and any product prior $P_1 = \gamma \times \pi$ for γ a distribution on $\eta \in [0, \frac{1}{2}]$ and π a distribution on i, and we take as comparator $Q = \gamma(\eta \mid \eta \in [\hat{\eta}/2, \hat{\eta}]) \times \hat{\pi}$ for any $\hat{\eta} \in [0, \frac{1}{2}]$ and distribution $\hat{\pi}$ on i that can both depend on all the losses, then

$$\mathop{\mathbb{E}}_{\hat{\pi}}\left[\mathcal{R}_T(i)\right] \leq 2\hat{\eta} \mathop{\mathbb{E}}_{\hat{\pi}}[\mathcal{V}_T(i)] + \frac{2}{\hat{\eta}}\Big(\mathrm{KL}(\hat{\pi}\|\pi) - \ln \gamma([\hat{\eta}/2, \hat{\eta}])\Big). \tag{12}$$

2. This follows e.g. from Lemma 1 by subtracting $\sum_t f_t(\boldsymbol{w}_t)$ on both sides of (3) and rearranging.

Crucially, the algorithm does not need to know $\hat{\eta}$ in advance, but (12) still holds for all $\hat{\eta}$ simultaneously. To minimize (12) in $\hat{\eta}$ we can restrict ourselves to $\hat{\eta} \geq 1/\sqrt{T}$ without loss of generality, so that a prior density $\mathrm{d}\gamma(\eta)/\mathrm{d}\eta \propto 1/\eta$ on $[1/\sqrt{T}, 1/2]$ achieves $-\ln \gamma([\hat{\eta}/2, \hat{\eta}]) = O(\ln \ln T)$. After optimizing $\hat{\eta}$, this leads to an adaptive regret bound of

$$\mathop{\mathbb{E}}_{\hat{\pi}}\left[\mathcal{R}_T(i)\right] = O\left(\sqrt{\mathop{\mathbb{E}}_{\hat{\pi}}[\mathcal{V}_T(i)]\Big(\mathrm{KL}(\hat{\pi}\|\pi) + \ln \ln T\Big)}\right) \qquad \text{for all } \hat{\pi}, \tag{13}$$

which recovers the results of Koolen and Van Erven (2015) (see also (Koolen, 2015)).

5.2. Reduction for Squint

Running EW with a continuous prior on η for the iProd surrogate losses from (10) requires evaluating a t-degree polynomial in η in every round, and therefore leads to $O(T^2)$ total running time. This may be reduced to $O(T \ln T)$ by using a prior γ on an exponentially spaced grid of η (as in MetaGrad (Van Erven and Koolen, 2016)), but in the experts setting even the extra $\ln T$ factor in run time can be avoided. This is possible by moving the 'prod bound' that occurs in the proof of Theorem 8, from the analysis into the algorithm by replacing the surrogate loss from (10) by the slightly larger surrogate loss

$$\ell_t(\eta, i) = -\eta r_t(i) + \eta^2 r_t(i)^2, \tag{14}$$

which turns iProd into Squint. Because this surrogate is quadratic in η, it becomes possible to run EW in the resulting surrogate OCO task and evaluate the resulting integrals over η in closed form for suitable choices of the prior on η, so that Squint has $O(T)$ run time (see Koolen and Van Erven (2015) for a detailed discussion of the choice of prior). Moreover, as shown in Appendix E.2, it satisfies exactly the same guarantees as iProd.

5.3. Reduction for Coin Betting

If we are willing to give up on second-order bounds, but still want to learn η, then there is another way to obtain an algorithm with $O(T)$ run time by bounding the iProd surrogate loss, which leads to a variant of the Coin Betting algorithm for experts of Orabona and Pál (2016). Our presentation and analysis are very different from (Orabona and Pál, 2016), but we obtain exactly the same regret bound for essentially the same algorithm, and we can explain some design choices that required clever insights by Orabona and Pál (2016), as natural consequences of running EW in the surrogate OCO task that we end up with.

The idea is to split the learning of $\eta \in [0, 1]$ and i into separate steps: for each i, we restrict $P_t(\eta \mid i)$ to be a point mass on some η_t^i, and we will choose η_t^i to achieve small regret for the surrogate loss

$$\ell_t^i(\eta) = -\frac{1 + r_t(i)}{2} \ln \frac{1 + \eta}{2} - \frac{1 - r_t(i)}{2} \ln \frac{1 - \eta}{2} - \ln 2,$$

which upper bounds (10) by convexity of the negative logarithm. We then plug in the choices of η_t^i in (10) and learn i for the resulting surrogate losses $\tilde{\ell}_t(i) = -\ln(1 + \eta_t^i r_t(i))$. For $\eta \in [0, 1]$ and $\hat{\pi}$ a distribution on i, let

$$S_T^i(\eta) = \sum_{t=1}^T \ell_t^i(\eta_t^i) - \sum_{t=1}^T \ell_t^i(\eta), \qquad \tilde{S}_T(\hat{\pi}) = \sum_{t=1}^T -\ln \mathop{\mathbb{E}}_{i \sim P_t}\left[e^{-\tilde{\ell}_t(i)}\right] - \mathop{\mathbb{E}}_{\hat{\pi}}\left[\sum_{t=1}^T \tilde{\ell}_t(i)\right]$$

be the mix-regret in the two surrogate OCO tasks. (Notice that in S_T^i the mix-regret has collapsed to the ordinary regret, because we are restricting ourselves to play point masses on η.) Also let $\mathcal{R}_T^+(i) = \max\{\mathcal{R}_T(i), 0\}$ be the nonnegative part of the regret, and define $\mathrm{B}(x\|y) = x \ln \frac{x}{y} + (1-x) \ln \frac{1-x}{1-y}$ to be the Kullback-Leibler divergence between two Bernoulli distributions, which satisfies $\mathrm{B}(x\|y) \geq 2(x-y)^2$ by Pinsker's inequality. Then this reduction gives the following regret bound, proved in Appendix E.3:

Theorem 9 (Coin Betting Reduction to EW) *Any choice of distributions P_t on i and learning rates η_t^i in the surrogate OCO task defined above induces regret bounded by*

$$\mathop{\mathbb{E}}_{\hat{\pi}}\left[\mathrm{B}\left(\tfrac{1}{2} + \tfrac{\mathcal{R}_T^+(i)}{2T} \| \tfrac{1}{2}\right)\right] \leq \tfrac{1}{T}\left(\mathop{\mathbb{E}}_{\hat{\pi}}\left[S_T^i\left(\tfrac{\mathcal{R}_T^+(i)}{T}\right)\right] + \tilde{S}_T(\hat{\pi})\right) \qquad \textit{for any } \hat{\pi} \textit{ on } i \tag{15}$$

in the original prediction with expert advice task.

In particular, if we use EW with learning rate 1 and prior π on i for the losses $\tilde{\ell}_t$, and for the losses ℓ_t^i we let η_t^i be the mean of lazy EW with learning rate 1 and with prior on $\eta \in [-1, +1]$ such that $\frac{1+\eta}{2}$ has a beta-distribution $\beta(a, a)$ with $a = \frac{T}{4} + \frac{1}{2}$ and with projections onto $\mathcal{P} = \{P \mid \mathbb{E}_P[\eta] \in [0, 1]\}$, then

$$\mathop{\mathbb{E}}_{\hat{\pi}}\left[\mathcal{R}_T(i)\right] \leq \sqrt{3T\left(\mathrm{KL}(\hat{\pi}\|\pi) + 3\right)} \qquad \textit{for any } \hat{\pi} \textit{ on } i. \tag{16}$$

Compared to (13), (16) avoids a $\ln \ln T$ term, but it has lost the benefits of the second-order factor $\mathbb{E}_{\hat{\pi}}[\mathcal{V}_T(i)] \leq T$. This may be explained by its upper bound $\ell_t^i(\eta) \geq \ell_t(\eta, i)$, which is tight only in the extreme case that $r_t(i) \in \{-1, +1\}$.

The Resulting Coin Betting Algorithm EW on the losses ℓ_t^i with the (conjugate) $\beta(a, a)$ prior is a generalization of the Krichevsky-Trofimov estimator (see Example 1) and its mean has the closed form $\frac{\mathcal{R}_{t-1}(i)}{t-1+2a}$. Lazily projecting onto $\mathcal{P}$ then simply amounts to clipping at 0 (by convexity of KL-divergence in its first argument, which implies that the constraint $\mathbb{E}_P[\eta] \geq 0$ will be satisfied with equality when we project from a distribution with negative mean). This means that $\eta_t^i = \max\left\{\frac{\mathcal{R}_{t-1}(i)}{t-1+2a}, 0\right\}$. By (9) the Coin Betting algorithm from the theorem predicts with weights $w_{t,i}$ obtained by normalizing the unnormalized weights $\tilde{w}_{t,i} = \tilde{p}_t(i)\eta_t^i$, where $\tilde{p}_t(i)$ is the unnormalized probability $P_t(i)$ of EW on the losses $\tilde{\ell}_t$, which recursively satisfies

$$\tilde{p}_t(i) := \pi(i) \prod_{s=1}^{t-1} (1 + \eta_s^i r_s(i)) = \tilde{p}_{t-1}(i) + \tilde{w}_{t-1,i} r_{t-1}(i) = \ldots = \pi(i) + \sum_{s=1}^{t-1} \tilde{w}_{s,i} r_s(i).$$

Interestingly, Orabona and Pál (2016) interpret the unnormalized EW probabilities $\tilde{p}_t(i)$ as the *Wealth* for expert i that is achieved by a gambler.

The interpretation in Theorem 9 explains three design choices by Orabona and Pál (2016): first, their choice of potential function, which naturally arises in our proof when we bound the regret $S_T^i(\mathcal{R}_T^+(i)/T)$ for EW using Lemma 1. Second, the choice for a, which in the original analysis comes from defining a shifted potential function, is simply specifying a prior with most mass in a region of order $1/\sqrt{T}$ around $\eta = 0$. And, third, the clipping of the unnormalized weights $\tilde{w}_{t,i}$ to 0 when $\mathcal{R}_{t-1}(i) < 0$, which in our presentation happens automatically because the learning rate η_t^i is projected to be 0 if it would otherwise become negative. Defining a prior on positive learning rates

directly would be possible in theory, but not with a conjugate prior, so the computational efficiency of the algorithm is made possible by the projections.

There is one slight difference between the algorithm we obtain here and the original Coin Betting algorithm of Orabona and Pál (2016): in the original method the instantaneous regrets are clipped to $\max\{r_t(i), 0\}$ when $\mathcal{R}_{t-1}(i) < 0$, which our method does not do. Apparently there is some amount of freedom in the design of this type of algorithm.

6. Online Linear Optimization with Bandit Feedback

A benefit of the EW interpretation of MD is that it opens up the possibility of sampling from the EW posterior distribution instead of playing the mean. Here we show how this option can be leveraged to obtain an algorithm for online linear optimization with bandit feedback (Dani et al., 2007; Abernethy et al., 2008), which recovers the best known rate $O(d\sqrt{T \ln T})$. A proof of this fact has already been outlined by Bubeck and Eldan (2015), but here we fill in the technical details.

The linear bandit setting consists of linear losses $f_t(\boldsymbol{w}) = \langle \boldsymbol{w}, \boldsymbol{g}_t \rangle \in [-1, +1]$, but instead of seeing the vectors $\boldsymbol{g}_t$ we only observe $f_t(\boldsymbol{w}_t)$ for the algorithm's choice $\boldsymbol{w}_t$. The algorithm can randomize its choice $\boldsymbol{w}_t$, and $\boldsymbol{g}_t$ is fixed before the outcome of this randomization. The goal is to minimize the expected regret $\mathbb{E}[\mathcal{R}_T(\boldsymbol{u})]$, where the expectation is with respect to the algorithm's randomness.

We consider the EW algorithm with fixed learning rate η and uniform prior distribution P_1 over $\mathcal{W}$. In each round t, after observing $f_t(\boldsymbol{w}_t) = \langle \boldsymbol{w}_t, \boldsymbol{g}_t \rangle$, the algorithm constructs a random, unbiased estimate $\tilde{\boldsymbol{g}}_t$ of the loss vector $\boldsymbol{g}_t$ and uses this estimate to update P_t to P_{t+1}. It is easy to verify that, for each t, P_t is a member of the exponential family with cumulant generating function $F(\boldsymbol{\theta}) = \ln \int_{\mathcal{W}} e^{\langle \boldsymbol{w}, \boldsymbol{\theta} \rangle} \, \mathrm{d}\boldsymbol{w}$. At trial t, the algorithm samples $\boldsymbol{w}_t \sim Q_t$, where $Q_t = (1-\gamma)P_t + \gamma R$ is a mixture of the EW distribution P_t and a fixed "exploration" distribution R, chosen to be *John's exploration* (Bubeck et al., 2012). Using that the convex conjugate of F is a universal $O(d)$-self concordant barrier on $\mathcal{W}$ (Bubeck and Eldan, 2015), it can be shown that, when η and γ are appropriately chosen, this algorithm achieves expected regret of order $O(d\sqrt{T \ln T})$ (see Appendix F).

It is interesting to compare with the *SCRiBLe* algorithm (Abernethy et al., 2012), which replaces EW by MD. By the results of Section 3.3, this is an essentially equivalent approach, except that SCRiBLe employs a sampling strategy based on the spectrum of the Hessian of F^*, without reference to the EW distribution, and achieves a regret bound that is suboptimal in d. This shows that the EW interpretation of MD is clearly beneficial in the bandit setting.

7. Discussion

We conclude with several remarks: first, we point out that there may be computational reasons to avoid defining the prior directly on the domain $\mathcal{W}$ of interest: as shown for instance in Sections 3.2 and 4, defining a Gaussian prior on all of $\mathbb{R}^d$ and then projecting the mean onto $\mathcal{W}$ can be computationally more efficient. In the context of sampling from the EW distribution, discussed in Section 6, this might also make sense if we project onto the alternative (smaller) set of distributions $\mathcal{P} = \{P \mid P(\mathcal{W}) = 1\} \subset \{P \mid \mathbb{E}_P[\boldsymbol{w}] \in \mathcal{W}\}$ that are supported on $\mathcal{W}$, which amounts to conditioning on $\mathcal{W}$. Second, there seems to be a discrepancy between the body of work for the log loss cited in the introduction, which strongly suggests using Jeffreys' prior, and the uniform prior suggested in Section 6 in the context of the universal barrier.

Acknowledgments

The authors would like to thank Wouter Koolen for extensive discussions underlying Theorems 3, 5, 8 and 12. A precursor to Theorem 4 previously appeared in Van der Hoeven's master's thesis (Van der Hoeven, 2016). He was supported by the Netherlands Organization for Scientific Research (NWO grant TOP2EW.15.211). Kotłowski was supported by the Polish National Science Centre (grant no. 2016/22/E/ST6/00299).

References

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the 21th Annual Conference on Learning Theory (COLT)*, pages 263–274, 2008.

Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Interior-point methods for full-information and bandit online learning. *IEEE Trans. Information Theory*, 58(7):4164–4175, 2012.

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

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multi-armed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

Arindam Banerjee, Srujana Merugu, Inderjit S. Dhillon, and Joydeep Ghosh. Clustering with Bregman divergences. *The Journal of Machine Learning Research*, 6:1705–1749, 2005.

Sébastien Bubeck and Ronen Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. In *Proceedings of the 28th Annual Conference on Learning Theory (COLT)*, pages 279–279, 2015.

Sébastien Bubeck, Nicolò Cesa-Bianchi, and Sham M. Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, pages 41.1–41.14, 2012.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge university press, 2006.

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

Thomas M. Cover. Universal portfolios. *Mathematical Finance*, 1:1–29, 1991.

Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

Varsha Dani, Thomas Hayes, and Sham Kakade. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems 20 (NIPS)*, pages 345–352, 2007.

Travis Dick, András György, and Csaba Szepesvári. Online learning in Markov decision processes with changing cost sequences. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 512–520, 2014.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Tim van Erven and Wouter M. Koolen. Metagrad: Multiple learning rates in online learning. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pages 3666–3674, 2016.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

Peter D. Grünwald. *The minimum description length principle*. MIT press, 2007.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

David P. Helmbold and Manfred K. Warmuth. Learning permutations with Exponential Weights. *Journal of Machine Learning Research*, 10:1705–1736, 2009.

Dirk van der Hoeven. Is Mirror Descent a special case of Exponential Weights? Master's thesis, Leiden University, The Netherlands, 2016. Available from http://pub.math.leidenuniv.nl/~hoevendvander/.

Shunsuke Ihara. *Information Theory for Continuous Systems*, volume 2. World Scientific, 1993.

Adam Kalai and Santosh Vempala. Efficient algorithms for universal portfolios. *Journal of Machine Learning Research*, 3(Nov):423–440, 2002.

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

Wouter M. Koolen. The relative entropy bound for Squint. *Blog August 13: http://blog.wouterkoolen.info/Squint_PAC/post.html*, 2015.

Wouter M. Koolen. Exploiting curvature using Exponential Weights. *Blog September 6: http://blog.wouterkoolen.info/EW4Quadratic/post.html*, 2016.

Wouter M. Koolen and Tim van Erven. Second-order quantile methods for experts and combinatorial games. In *Proceedings of The 28th Conference on Learning Theory (COLT)*, pages 1155–1175, 2015.

Wouter M. Koolen, Peter Grünwald, and Tim van Erven. Combining adversarial guarantees and stochastic fast rates in online learning. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pages 4457–4465, 2016.

Raphail Krichevsky and Victor Trofimov. The performance of universal encoding. *IEEE Transactions on Information Theory*, 27(2):199–207, 1981.

Nick Littlestone and Manfred K. Warmuth. The Weighted Majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

Hariharan Narayanan and Alexander Rakhlin. Efficient sampling from time-varying log-concave distributions. *The Journal of Machine Learning Research*, 18(1):4017–4045, 2017.

Frank Nielsen and Richard Nock. Entropies and cross-entropies of exponential families. In *17th IEEE International Conference on Image Processing (ICIP)*, pages 3621–3624. IEEE, 2010.

Francesco Orabona and Dávid Pál. Coin betting and parameter-free online learning. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pages 577–585, 2016.

Francesco Orabona, Koby Crammer, and Nicolò Cesa-Bianchi. A generalized online mirror descent with applications to classification and regression. *Machine Learning*, 99(3):411–435, 2015.

Laurent Orseau, Tor Lattimore, and Shane Legg. Soft-Bayes: Prod for mixtures of experts with log-loss. In *International Conference on Algorithmic Learning Theory 28 (ALT)*, pages 372–399, 2017.

Steven de Rooij, Tim van Erven, Peter D. Grünwald, and Wouter M. Koolen. Follow the Leader if you can, Hedge if you must. *Journal of Machine Learning Research*, 15:1281–1316, 2014.

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

Volodimir G. Vovk. Aggregating strategies. In *Proceedings of the 3rd Annual Conference on Learning Theory (COLT)*, pages 371–383, 1990.

Volodimir G. Vovk. Competitive on-line statistics. *International Statistical Review*, 69(2):213–248, 2001.

Qun Xie and Andrew R. Barron. Asymptotic minimax regret for data compression, gambling, and prediction. *IEEE Transactions on Information Theory*, 46(2):431–445, 2000.

Martin Zinkevich. Online convex programming and generalized infinitesimal Gradient Ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, pages 928–936, 2003.

Appendix A. Proof of Lemma 1 from Section 2

In the following we make use of the generalized Pythagorean inequality for Kullback-Leibler divergence (Csiszár, 1975): for $P_t = \arg\min_{P \in \mathcal{P}} \ \mathrm{KL}(P \| \tilde{P}_t)$ and any $Q \in \mathcal{P}$:

$$\mathrm{KL}(Q \| \tilde{P}_t) \geq \mathrm{KL}(Q \| P_t) + \mathrm{KL}(P_t \| \tilde{P}_t). \tag{17}$$

For greedy EW we have

$$\begin{aligned}\frac{1}{\eta_t}\big(\mathrm{KL}(Q\|P_t)-\mathrm{KL}(Q\|P_{t+1})\big) &\geq \frac{1}{\eta_t}\Big(\mathrm{KL}(Q\|P_t)-\mathrm{KL}(Q\|\tilde{P}_{t+1})\Big) && \text{(from (17))}\\ &= \mathop{\mathbb{E}}_{Q}[f_t(\boldsymbol{w})]-\frac{1}{\eta_t}\ln\mathop{\mathbb{E}}_{P_t}\left[e^{-\eta_t f_t(\boldsymbol{w})}\right] && \text{(from (2))}\end{aligned}$$

in any trial t. Summing over trials gives:

$$\begin{aligned}\sum_{t=1}^{T}\mathop{\mathbb{E}}_{Q}[f_t(\boldsymbol{w})]-\frac{1}{\eta_t}\ln\mathop{\mathbb{E}}_{P_t}\left[e^{-\eta_t f_t(\boldsymbol{w})}\right] &\leq \sum_{t=1}^{T}\frac{1}{\eta_t}\big(\mathrm{KL}(Q\|P_t)-\mathrm{KL}(Q\|P_{t+1})\big)\\ &= \frac{1}{\eta_1}\mathrm{KL}(Q\|P_1)-\frac{1}{\eta_T}\mathrm{KL}(Q\|P_{T+1})+\sum_{t=2}^{T}\mathrm{KL}(Q\|P_t)\left(\frac{1}{\eta_t}-\frac{1}{\eta_{t-1}}\right)\\ &\leq \frac{1}{\eta_1}\mathrm{KL}(Q\|P_1)+\max_{t=2,\ldots,T}\mathrm{KL}(Q\|P_t)\left(\frac{1}{\eta_T}-\frac{1}{\eta_1}\right).\end{aligned}$$

Rearranging the terms and adding $\sum_{t=1}^{T} f_t(\boldsymbol{w}_t)$ on both sides results in (4).

We now proceed with the proof of lazy EW, starting from:

$$\begin{aligned}-\frac{1}{\eta_{t-1}}\ln\mathop{\mathbb{E}}_{P_t}[e^{-\eta_{t-1}f_t(\boldsymbol{w})}] &= \min_{P}\left\{\mathop{\mathbb{E}}_{P}[f_t(\boldsymbol{w})]+\frac{1}{\eta_{t-1}}\mathrm{KL}(P\|P_t)\right\}\\ &\leq \mathop{\mathbb{E}}_{P_{t+1}}[f_t(\boldsymbol{w})]+\frac{1}{\eta_{t-1}}\mathrm{KL}(P_{t+1}\|P_t)\\ &\leq \mathop{\mathbb{E}}_{P_{t+1}}[f_t(\boldsymbol{w})]+\frac{1}{\eta_{t-1}}\mathrm{KL}(P_{t+1}\|\tilde{P}_t)-\frac{1}{\eta_{t-1}}\mathrm{KL}(P_t\|\tilde{P}_t), \qquad (18)\end{aligned}$$

where the last inequality is from the Pythagorean inequality (17) applied with $Q=P_{t+1}$. By (1):

$$\ln\frac{\mathrm{d}\tilde{P}_t(\boldsymbol{w})}{\mathrm{d}P_1(\boldsymbol{w})}=-\eta_{t-1}\sum_{s=1}^{t-1}f_s(\boldsymbol{w})-\ln\mathop{\mathbb{E}}_{P_1}\left[e^{-\eta_{t-1}\sum_{s=1}^{t-1}f_s(\boldsymbol{w})}\right],$$

which gives:

$$\begin{aligned}\frac{1}{\eta_{t-1}}\mathrm{KL}(P_{t+1}\|\tilde{P}_t)-\frac{1}{\eta_{t-1}}\mathrm{KL}(P_t\|\tilde{P}_t) &= \frac{1}{\eta_{t-1}}\mathrm{KL}(P_{t+1}\|P_1)-\frac{1}{\eta_{t-1}}\mathrm{KL}(P_t\|P_1)\\ &\quad+\mathop{\mathbb{E}}_{P_{t+1}}\left[\sum_{s=1}^{t-1}f_s(\boldsymbol{w})\right]-\mathop{\mathbb{E}}_{P_t}\left[\sum_{s=1}^{t-1}f_s(\boldsymbol{w})\right].\end{aligned}$$

Plugging this into (18) and using $\eta_t\leq\eta_{t-1}$ results in:

$$\begin{aligned}-\frac{1}{\eta_{t-1}}\ln\mathop{\mathbb{E}}_{P_t}[e^{-\eta_{t-1}f_t(\boldsymbol{w})}] &\leq \frac{1}{\eta_t}\mathrm{KL}(P_{t+1}\|P_1)-\frac{1}{\eta_{t-1}}\mathrm{KL}(P_t\|P_1)\\ &\quad+\mathop{\mathbb{E}}_{P_{t+1}}\left[\sum_{s=1}^{t}f_s(\boldsymbol{w})\right]-\mathop{\mathbb{E}}_{P_t}\left[\sum_{s=1}^{t-1}f_s(\boldsymbol{w})\right].\end{aligned}$$

Summing over trials makes the terms on the right-hand side telescope and gives:

$$\begin{aligned}\sum_{t=1}^{T} -\frac{1}{\eta_{t-1}} \ln \mathop{\mathbb{E}}_{P_t}[e^{-\eta_{t-1} f_t(\boldsymbol{w})}] &\leq \frac{1}{\eta_T} \mathrm{KL}(P_{T+1} \| P_1) + \mathop{\mathbb{E}}_{P_{T+1}} \left[\sum_{t=1}^{T} f_t(\boldsymbol{w}) \right] \\ &= \min_{P \in \mathcal{P}} \left\{ \mathop{\mathbb{E}}_{P} \left[\sum_{t=1}^{T} f_t(\boldsymbol{w}) \right] + \frac{1}{\eta_T} \mathrm{KL}(P \| P_1) \right\} \\ &\leq \mathop{\mathbb{E}}_{Q} \left[\sum_{t=1}^{T} f_t(\boldsymbol{w}) \right] + \frac{1}{\eta_T} \mathrm{KL}(Q \| P_1),\end{aligned}$$

where the equality expresses an equivalent way to define lazy EW. Rearranging the terms and adding $\sum_{t=1}^{T} f_t(\boldsymbol{w}_t)$ on both sides results in (3).

Appendix B. Proof of Theorem 2

Proof Rather than scaling canonical vectors $\boldsymbol{e}_i$, $i = 1, \ldots, d$ and the comparator $\boldsymbol{u}$ by M, we scale the loss vectors by defining $\boldsymbol{g}'_t = M\boldsymbol{g}_t$, so that the losses remain the same: $\langle \boldsymbol{e}_i, \boldsymbol{g}'_t \rangle = \langle M\boldsymbol{e}_i, \boldsymbol{g}_t \rangle$ for all i and all t. Let $\boldsymbol{w}_1 = (\boldsymbol{w}_1^+, \boldsymbol{w}_1^-)$, and let $\boldsymbol{w}_t^+$, $\boldsymbol{w}_t^-$ be the result of running EG plus-minus on $\boldsymbol{g}'_t$. For any $\boldsymbol{u}$ with $\sum_{i=1}^{2d} u_i = 1$ and $u_i \geq 0$ invoking Lemma 1 gives:

$$\begin{aligned}\sum_{t=1}^{T} \langle \boldsymbol{w}_t - \boldsymbol{u}, \boldsymbol{g}'_t \rangle \leq \frac{1}{\eta} \mathrm{KL}(\boldsymbol{u} \| \boldsymbol{w}_1) \\ + \sum_{t=1}^{T} \langle \boldsymbol{w}_t^+, \boldsymbol{g}'_t \rangle - \langle \boldsymbol{w}_t^-, \boldsymbol{g}'_t \rangle + \frac{1}{\eta} \ln \Big(\sum_{i=1}^{d} (w_{t,i}^+ e^{-\eta_t \langle \boldsymbol{e}_i, \boldsymbol{g}'_t \rangle} + w_{t,i}^- e^{\eta_t \langle \boldsymbol{e}_i, \boldsymbol{g}'_t \rangle}) \Big). \quad (19)\end{aligned}$$

The first term on the right-hand side of (19) can be bounded by: $\max_{\boldsymbol{u}: \sum_{i=1}^{2d} u_i = 1,\, u_i \geq 0} \mathrm{KL}(\boldsymbol{u} \| \boldsymbol{w}_1) = \ln(2d)$. To bound the second term on the right-hand side of (19), we make use of Hoeffding's Lemma (Cesa-Bianchi and Lugosi, 2006, Lemma A.1), which together with $|\langle \boldsymbol{e}_i, \boldsymbol{g}'_t \rangle| \leq MG$ gives:

$$\sum_{t=1}^{T} \langle \boldsymbol{w}_t^+, \boldsymbol{g}'_t \rangle - \langle \boldsymbol{w}_t^-, \boldsymbol{g}'_t \rangle + \frac{1}{\eta} \ln \Big(\sum_{i=1}^{d} (w_{t,i}^+ e^{-\eta_t \langle \boldsymbol{e}_i, \boldsymbol{g}'_t \rangle} + w_{t,i}^- e^{\eta_t \langle \boldsymbol{e}_i, \boldsymbol{g}'_t \rangle}) \Big) \leq \frac{\eta M^2 G^2}{2}.$$

Summing over trials results in a bound on the regret:

$$\sum_{t=1}^{T} \langle \boldsymbol{w}_t - \boldsymbol{u}, \boldsymbol{g}'_t \rangle \leq \frac{\ln(2d)}{\eta} + \eta \frac{T M^2 G^2}{2}.$$

Plugging in the optimal $\eta = \sqrt{\frac{2 \ln(2d)}{T M^2 G^2}}$ yields the desired result. ■

Appendix C. Proof of Theorem 4

Before proving the theorem, we need two lemmas:

Lemma 10 (Banerjee et al. (2005); Nielsen and Nock (2010)) *The KL divergence between two members, P and Q, of the same regular exponential family $\mathcal{E}$ with cumulant generating function F can be expressed by the Bregman divergence between their natural parameters, $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_Q$, or their expectation parameters, $\boldsymbol{\mu}_P$ and $\boldsymbol{\mu}_Q$. The first Bregman divergence is generated by the cumulant generating function F and the second Bregman divergence is generated by the convex conjugate of the cumulant generating function F^*:*

$$\mathrm{KL}(P\|Q) = B_F(\boldsymbol{\theta}_Q\|\boldsymbol{\theta}_P) = B_{F^*}(\boldsymbol{\mu}_P\|\boldsymbol{\mu}_Q).$$

Lemma 11 *(Ihara, 1993, Theorem 3.1.4) Let $\boldsymbol{\mu}$ be arbitrary and define $\mathcal{P} = \{P : \mathbb{E}_P[\boldsymbol{w}] = \boldsymbol{\mu}\}$. Then, for any member Q of an exponential family $\mathcal{E}$,*

$$\min_{P\in\mathcal{P}} \mathrm{KL}(P\|Q)$$

is achieved by $P \in \mathcal{E}$ such that $\mathbb{E}_P[\boldsymbol{w}] = \boldsymbol{\mu}$, provided such a P exists.

Proof [of Theorem 4] Let $\boldsymbol{w}_t$ be the weights produced by the greedy version of MD. Then

$$\begin{aligned}\min_{P\in\mathcal{P}} \left\{\mathop{\mathbb{E}}_{P}[\langle \boldsymbol{w}, \boldsymbol{g}_t\rangle] + \frac{1}{\eta_t}\mathrm{KL}(P\|P_t)\right\} &= \min_{\boldsymbol{\mu}\in\mathcal{W}} \min_{P\,:\,\mathbb{E}_P[\boldsymbol{w}]=\boldsymbol{\mu}} \left\{\mathop{\mathbb{E}}_{P}[\langle \boldsymbol{w}, \boldsymbol{g}_t\rangle] + \frac{1}{\eta_t}\mathrm{KL}(P\|P_t)\right\} \\ &= \min_{\boldsymbol{\mu}\in\mathcal{W}} \min_{P\in\mathcal{E}\,:\,\mathbb{E}_P[\boldsymbol{w}]=\boldsymbol{\mu}} \left\{\langle \boldsymbol{\mu}, \boldsymbol{g}_t\rangle + \frac{1}{\eta_t}\mathrm{KL}(P\|P_t)\right\},\end{aligned}$$

where in the second step we can restrict to minimization over $\mathcal{E}$ by Lemma 11. Introducing the short-hand notation $\boldsymbol{\mu}_P = \mathbb{E}_P[\boldsymbol{w}]$, we thus get for the greedy version of EW:

$$P_{t+1} = \mathop{\arg\min}_{P\in\mathcal{E}:\boldsymbol{\mu}_P\in\mathcal{W}} \left\{\langle \boldsymbol{\mu}_P, \boldsymbol{g}_t\rangle + \frac{1}{\eta_t}\mathrm{KL}(P\|P_t)\right\} = \mathop{\arg\min}_{P\in\mathcal{E}:\boldsymbol{\mu}_P\in\mathcal{W}} \left\{\langle \boldsymbol{\mu}_P, \boldsymbol{g}_t\rangle + \frac{1}{\eta_t}B_{F^*}(\boldsymbol{\mu}_P\|\boldsymbol{\mu}_{P_t})\right\},$$

where we used Lemma 10. But the last expression coincides with the definition of the greedy MD weight update, and since it applies to all t, we have $\boldsymbol{\mu}_{P_{t+1}} = \boldsymbol{w}_{t+1}$ for all t, provided $\boldsymbol{\mu}_{P_1} = \boldsymbol{w}_1$ (which holds by assumption). An analogous argument can be made to show the equivalence of the lazy versions of MD and EW. ■

Appendix D. Proofs for Section 4

D.1. Proof of Theorem 5

Proof $\tilde{P}_t = \mathcal{N}(\tilde{\boldsymbol{w}}_t, \Sigma_t)$ may be verified analytically from (1) and (2). The fact that projections P_t onto $\mathcal{P}$ preserve Gaussianity with the same covariance matrix follows from Lemma 9 in Van Erven and Koolen (2016). Lemma 1 gives a bound on the regret w.r.t. randomized forecaster $Q = \mathcal{N}(\boldsymbol{u}, \Sigma_Q)$:

$$\sum_{t=1}^T f_t(\boldsymbol{w}_t) - \sum_{t=1}^T \mathop{\mathbb{E}}_{Q}[f_t(\boldsymbol{w})] \le \frac{1}{\eta}\mathrm{KL}(Q\|P_1) + \sum_{t=1}^T f_t(\boldsymbol{w}_t) + \frac{1}{\eta}\ln \mathop{\mathbb{E}}_{P_t}\left[e^{-\eta f_t(\boldsymbol{w})}\right].$$

The KL divergence between two Gaussians is given by (Ihara, 1993, Theorem 1.8.2):

$$\mathrm{KL}(Q\|P_1) = \frac{1}{2}(\ln\left(\frac{\det(\Sigma_Q)}{\det(\Sigma_1)}\right) + \mathrm{Tr}(\Sigma_Q\Sigma_1^{-1}) + (\boldsymbol{u}-\boldsymbol{w}_1)^\intercal\Sigma_1^{-1}(\boldsymbol{u}-\boldsymbol{w}_1) - d.$$

The mixability gap can be evaluated in closed form by calculating the Gaussian integral:

$$\ln \mathop{\mathbb{E}}_{P_t}\left[e^{\eta(f_t(\boldsymbol{w}_t)-f_t(\boldsymbol{w}))}\right] = \frac{\eta^2}{2}\boldsymbol{g}_t^\intercal\Sigma_{t+1}\boldsymbol{g}_t - \frac{1}{2}\ln\left(\frac{\det(\Sigma_t)}{\det(\Sigma_{t+1})}\right).$$

Also, the expectation of the instantaneous regret can be computed exactly:

$$f_t(\boldsymbol{w}_t) - \mathop{\mathbb{E}}_{Q}[f_t(\boldsymbol{w})] = f_t(\boldsymbol{w}_t) - f_t(\boldsymbol{u}) - \frac{1}{2}\mathrm{Tr}(\Sigma_Q\boldsymbol{M}_t).$$

Summing the above over the trials, we get the following upper bound on the regret:

$$\sum_{t=1}^T f_t(\boldsymbol{w}_t) - \sum_{t=1}^T f_t(\boldsymbol{u}) \le \frac{\ln\left(\frac{\det(\Sigma_{T+1})}{\det(\Sigma_Q)}\right) + \mathrm{Tr}(\Sigma_Q\Sigma_{T+1}^{-1}) - d + (\boldsymbol{w}_1-\boldsymbol{u})^\intercal\Sigma_1^{-1}(\boldsymbol{w}_1-\boldsymbol{u})}{2\eta} + \eta\sum_{t=1}^T \boldsymbol{g}_t^\intercal\Sigma_{t+1}\boldsymbol{g}_t,$$

which holds for all Σ_Q. By plugging in the optimal value $\Sigma_Q = \Sigma_{T+1}$, the bound simplifies to:

$$\sum_{t=1}^T f_t(\boldsymbol{w}_t) - \sum_{t=1}^T f_t(\boldsymbol{u}) \le \frac{1}{2\eta}(\boldsymbol{w}_1-\boldsymbol{u})^\intercal\Sigma_1^{-1}(\boldsymbol{w}_1-\boldsymbol{u}) + \frac{\eta}{2}\sum_{t=1}^T \boldsymbol{g}_t^\intercal\Sigma_{t+1}\boldsymbol{g}_t,$$

which concludes the proof. ■

D.2. Proof of Corollary 6

Proof Using Theorem 5 gives:

$$\begin{aligned}
\sum_{t=1}^T f_t(\boldsymbol{w}_t) - \sum_{t=1}^T f_t(\boldsymbol{u}) &\le \frac{1}{2\eta\sigma^2}\|\boldsymbol{u}\|_2^2 + \frac{\eta}{2}\sum_{t=1}^T \frac{1}{\frac{1}{\sigma^2}+\alpha\eta t}\|\boldsymbol{g}_t\|_2^2 \\
&\le \frac{1}{2\eta\sigma^2}D^2 + \frac{\eta}{2}G^2\sum_{t=1}^T \frac{1}{\frac{1}{\sigma^2}+\alpha\eta t} \\
&\le \frac{1}{2\eta\sigma^2}D^2 + \frac{\eta G^2}{2(\frac{1}{\sigma^2}+\alpha\eta)} + \frac{\eta}{2}G^2\int_1^T \frac{1}{\frac{1}{\sigma^2}+\alpha\eta t}dt \\
&= \frac{1}{2\eta\sigma^2}D^2 + \frac{G^2}{2(\frac{1}{\eta\sigma^2}+\alpha)} + \frac{G^2}{2\alpha}\big(\ln(\tfrac{1}{\eta\sigma^2}+\alpha T) - \ln(\tfrac{1}{\eta\sigma^2}+\alpha)\big),
\end{aligned}$$

which was to be shown. ■

D.3. Proof of Corollary 7

Proof Using Theorem 5 gives:

$$\mathcal{R}_T(\boldsymbol{u}) \leq \frac{D^2}{2\eta\sigma^2} + \frac{\eta}{2}\sum_{t=1}^{T} \boldsymbol{g}_t^\intercal \Sigma_{t+1} \boldsymbol{g}_t. \tag{20}$$

We start by bounding the second term on the right-hand side of (20). Using Lemma 11.11 from Cesa-Bianchi and Lugosi (2006) and the basic inequality $1 - x \leq -\ln x$, we bound:

$$\eta\beta \boldsymbol{g}_t^\intercal \Sigma_{t+1} \boldsymbol{g}_t = 1 - \frac{\det(\Sigma_t^{-1})}{\det(\Sigma_{t+1}^{-1})} \leq \ln \frac{\det(\Sigma_{t+1}^{-1})}{\det(\Sigma_t^{-1})},$$

which after summing over trials gives:

$$\begin{aligned}\sum_{t=1}^{T} \eta\beta \boldsymbol{g}_t^\intercal \Sigma_{t+1} \boldsymbol{g}_t &\leq \ln \frac{\det(\Sigma_{T+1}^{-1})}{\det(\Sigma_1^{-1})} = \ln\det\left(\boldsymbol{I} + \eta\sigma^2\beta \sum_{t=1}^{T} \boldsymbol{g}_t \boldsymbol{g}_t^\intercal\right) \\ &= \sum_{i=1}^{d} \ln(1+\lambda_i) \leq d \ln\left(1 + \frac{\eta\sigma^2\beta G^2 T}{d}\right),\end{aligned}$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of $\eta\sigma^2\beta \sum_{t=1}^{T} \boldsymbol{g}_t \boldsymbol{g}_t^\intercal$, and the last inequality follows by maximizing under the constraint that $\sum_i \lambda_i = \operatorname{Tr}(\eta\sigma^2\beta \sum_{t=1}^{T} \boldsymbol{g}_t \boldsymbol{g}_t^\intercal) \leq \sigma^2\eta\beta G^2 T$. As discussed by Cesa-Bianchi and Lugosi (2006, proof and discussion of Theorem 11.7), the maximum is achieved when $\lambda_i = \sigma^2\eta\beta G^2 T/d$ for all i.

All together we find:

$$\mathcal{R}_T(\boldsymbol{u}) \leq \frac{D^2}{2\eta\sigma^2} + \frac{d}{2\beta} \ln\left(1 + \frac{\eta\sigma^2\beta G^2 T}{d}\right),$$

which was to be shown. ■

Appendix E. Proofs for Section 5

E.1. Proof of Theorem 8

Abbreviate $m_t(P) = -\ln \mathbb{E}_P\left[e^{-\ell_t(\eta,i)}\right]$ and define the potential $\Phi_T = e^{-\sum_{t=1}^{T} m_t(P_t)}$. Then $\Phi_T = \Phi_{T-1} = \cdots = \Phi_0 = 1$ since

$$\Phi_T - \Phi_{T-1} = e^{-\sum_{t=1}^{T-1} m_t(P_t)} \mathop{\mathbb{E}}_{P_T}\left[\eta r_T(i)\right] = 0,$$

where the last identity holds for any loss vector $\boldsymbol{g}_t$ by the definition of $\boldsymbol{w}_T$. For any comparator Q on (η, i), it follows that

$$0 = \sum_{t=1}^{T} m_t(P_t) = \sum_{t=1}^{T} \mathop{\mathbb{E}}_{Q}[\ell_t(\eta,i)] + S(Q) \leq \sum_{t=1}^{T} \mathop{\mathbb{E}}_{Q}[-\eta r_t(i) + \eta^2 r_t(i)^2] + S(Q),$$

where the last inequality is an application of the 'prod-bound' $-\ln(1+x) \leq -x + x^2$ with $x = \eta r_t(i)$, which holds for any $x \geq -\frac{1}{2}$ (Cesa-Bianchi et al., 2007, Lemma 1). The result (11) is a direct consequence, and (12) follows upon bounding $\mathbb{E}_Q[\eta] \geq \hat{\eta}/2$ and $\mathbb{E}_Q[\eta^2] \leq \hat{\eta}^2$ and plugging in that $S(Q) \leq \mathrm{KL}(Q\|P_1) = \mathrm{KL}(\hat{\pi}\|\pi) - \ln \gamma([\hat{\eta}/2, \hat{\eta}])$ for EW.

E.2. Proof of Theorem 12

Theorem 12 (Squint Reduction to EW) *The exact same statement as in Theorem 8 also holds when we replace the surrogate loss (10) by (14).*

Thus (13) also holds, and we recover the results of (Koolen and Van Erven, 2015) for Squint.

Remark 13 *The Metagrad algorithm (Van Erven and Koolen, 2016) is similar to Squint on a continuous set of experts indexed by $\boldsymbol{w} \in \mathbb{R}^d$ with losses $f_t(\boldsymbol{w}) = \boldsymbol{w}^\intercal \boldsymbol{g}_t$, and the analysis of Theorem 12 can be extended to handle this case.*

Proof Let $m_t(P)$ and Φ_T be as in the proof of Theorem 8, but for the new surrogate loss (14). Then $\Phi_T \leq \Phi_{T-1} \leq \ldots \leq \Phi_0 = 1$, because

$$\Phi_T - \Phi_{T-1} = e^{-\sum_{t=1}^{T-1} m_t(P_t)} \left(\mathop{\mathbb{E}}_{P_T} \left[e^{-f_t(\eta,i)} \right] - 1 \right) \leq e^{-\sum_{t=1}^{T-1} m_t(P_t)} \mathop{\mathbb{E}}_{P_T} \left[\eta r_T(i) \right] = 0,$$

where the inequality follows from the 'prod bound' (see the proof of Theorem 8) and the final equality is again by definition of $\boldsymbol{w}_T$. For any Q, it follows that

$$0 \leq \sum_{t=1}^{T} m_t(P_t) = \sum_{t=1}^{T} \mathop{\mathbb{E}}_{Q}[\ell_t(\eta, i)] + S(Q) = \sum_{t=1}^{T} \mathop{\mathbb{E}}_{Q}[-\eta r_t(i) + \eta^2 r_t(i)^2] + S(Q),$$

which implies that (11) also holds for Squint. Since (12) is a corollary, it also follows directly. ■

E.3. Proof of Theorem 9

The proof of Theorem 9 follows the same general steps as the proofs for Theorems 8 and 12. However, bounding the mix-regret $S_T^i(\eta)$ using a similar analysis as for the Krichevsky-Trofimov estimator from Example 1 would lead to an extra $\ln T$ factor in the regret. This is avoided using a more delicate analysis that holds specifically for the regret with respect to $\eta = \mathcal{R}_T^+(i)/T$, which requires a technical analytic inequality by Orabona and Pál (2016, Lemma 16).

Proof For ℓ_t as in (10), let $m_t = -\ln \mathbb{E}_{i \sim P_t} \left[e^{-\ell_t(\eta_t^i, i)} \right]$. Then, by the same argument as in the proof of Theorem 8, $\Phi_T = e^{-\sum_{t=1}^{T} m_t} = 1$. For any distribution $\hat{\pi}$ on i and any $\hat{\eta}^i \in [0,1]$, we therefore have

$$\begin{aligned} 0 = \sum_{t=1}^{T} m_t &= \mathop{\mathbb{E}}_{\hat{\pi}} \left[\sum_{t=1}^{T} \ell_t(\eta_t^i, i) \right] + \tilde{S}_T(\hat{\pi}) \leq \mathop{\mathbb{E}}_{\hat{\pi}} \left[\sum_{t=1}^{T} \ell_t^i(\eta_t^i) \right] + \tilde{S}_T(\hat{\pi}) \\ &= \mathop{\mathbb{E}}_{\hat{\pi}} \left[\sum_{t=1}^{T} \ell_t^i(\hat{\eta}^i) + S_T^i(\hat{\eta}^i) \right] + \tilde{S}_T(\hat{\pi}). \end{aligned} \tag{21}$$

The minimizer of $\sum_{t=1}^{T} \ell_t^i(\eta)$ over $\eta \in [0,1]$ is $\hat{\eta}^i = \mathcal{R}_T^+(i)/T$. Plugging this in, we find that

$$\sum_{t=1}^{T} \ell_t^i(\hat{\eta}^i) = -T\,\mathrm{B}(\tfrac{1}{2} + \tfrac{\mathcal{R}_T^+(i)}{2T} \| \tfrac{1}{2}). \tag{22}$$

Substituting (22) in (21) and reorganizing we obtain (15).

If we specialize to EW, then $\tilde{S}_T(\hat{\pi}) \le \mathrm{KL}(\hat{\pi}\|\pi)$ by the same argument as for iProd. In addition, to bound $S_T^i(\hat{\eta}^i)$, let $\tilde{\beta}(x,y)$ be the distribution on $\eta \in [-1,+1]$ such that $(1+\eta)/2$ has a $\beta(x,y)$ distribution. Then Lemma 1 and the observation that the mixability gap is at most 0 because ℓ_t^i is 1-exp-concave, together imply that

$$S_T^i(\hat{\eta}^i) \le \min_{Q\in\mathcal{P}} \Big\{ \underbrace{\mathop{\mathbb{E}}_{\eta\sim Q}\big[\sum_{t=1}^{T} \ell_t^i(\eta)\big] + \mathrm{KL}(Q\|\tilde{\beta}(a,a))}_{A(Q,i)} \Big\} - \underbrace{\sum_{t=1}^{T} \ell_t^i(\hat{\eta}^i)}_{B(i)}.$$

We first rewrite $B(i)$ using (22). Then it remains to bound the term with $A(Q,i)$ in expectation under $\hat{\pi}$. To this end we may assume that $\mathcal{R}_T(\hat{\pi}) := \mathbb{E}_{\hat{\pi}}[\mathcal{R}_T(i)] \ge 0$ without loss of generality (otherwise (16) holds trivially). Hence

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\sim\hat{\pi}}\big[\min_{Q\in\mathcal{P}} A(Q,i)\big] \le \min_{Q\in\mathcal{P}} \mathop{\mathbb{E}}_{i\sim\hat{\pi}}\big[A(Q,i)\big] \\
&\quad= \min_{Q\in\mathcal{P}} \left\{ \mathop{\mathbb{E}}_{\eta\sim Q}\Big[-\frac{T+\mathcal{R}_T(\hat{\pi})}{2} \ln\frac{1+\eta}{2} - \frac{T-\mathcal{R}_T(\hat{\pi})}{2} \ln\frac{1-\eta}{2} - T\ln 2\Big] + \mathrm{KL}(Q\|\tilde{\beta}(a,a)) \right\} \\
&\quad= -\ln\left(2^T \mathop{\mathbb{E}}_{X\sim\beta(a,a)}\Big[X^{\frac{T+\mathcal{R}_T(\hat{\pi})}{2}} (1-X)^{\frac{T-\mathcal{R}_T(\hat{\pi})}{2}} \Big] \right) \\
&\quad= -\ln\left(\frac{2^T \Gamma(2a) \Gamma\big(\frac{T+\mathcal{R}_T(\hat{\pi})}{2} + a\big) \Gamma\big(\frac{T-\mathcal{R}_T(\hat{\pi})}{2} + a\big)}{\Gamma(a)^2 \Gamma(T+2a)} \right) \\
&\quad\le \frac{-\mathcal{R}_T(\hat{\pi})^2}{2T+4a-2} + \frac{1}{2}\ln\frac{T+2a-1}{2a} + \ln(e\sqrt{\pi}),
\end{aligned}$$

where we have plugged in the minimizing $Q = \tilde{\beta}(\frac{T+\mathcal{R}_T(\hat{\pi})}{2} + a, \frac{T-\mathcal{R}_T(\hat{\pi})}{2} + a)$, which has nonnegative mean under our assumption that $\mathcal{R}_T(\hat{\pi}) \ge 0$, and where the last inequality holds by (Orabona and Pál, 2016, Lemma 16), which applies for $a \ge 1/2$, $\mathcal{R}_T(\hat{\pi}) \in [-T, T]$ and $T \ge 1$.

With these regret bounds for EW, (15) specializes to

$$\mathcal{R}_T(\hat{\pi}) \le \sqrt{(2T+4a-2)\left(\frac{1}{2}\ln\frac{T+2a-1}{2a} + \ln(e\sqrt{\pi}) + \mathrm{KL}(\hat{\pi}\|\pi)\right)}.$$

The result so far holds for any $a \ge \frac{1}{2}$. Plugging in the choice $a = \frac{T}{4} + \frac{1}{2}$, suggested by Orabona and Pál (2016), and using $\frac{1}{2}\ln\frac{3T}{T+2} + \ln(e\sqrt{\pi}) \le 3$ completes the proof. ■

Appendix F. Analysis of the Algorithm from Section 6

Let $\mathcal{W} \subset \mathbb{R}^d$ be a compact convex set. Following Bubeck et al. (2012), we assume without loss of generality that $\mathcal{W}$ is full rank, meaning that the linear combinations of $\mathcal{W}$ span $\mathbb{R}^d$ (otherwise we can express the elements of $\mathcal{W}$ in a lower dimensional space).

At trials $t = 1, 2, \ldots, T$, the algorithm plays with a randomized choice $\boldsymbol{w}_t \in \mathcal{W}$, the adversary chooses an unobserved loss vector $\boldsymbol{g}_t$, which is not allowed to depend on the realization of $\boldsymbol{w}_t$, and the learner suffers and observes bounded loss $\langle \boldsymbol{w}_t, \boldsymbol{g}_t \rangle$. The goal is to minimize the expected regret: $\mathbb{E}[\mathcal{R}_T(\boldsymbol{u})] = \mathbb{E}\left[\sum_{t=1}^T \langle \boldsymbol{w}_t - \boldsymbol{u}, \boldsymbol{g}_t \rangle\right]$ for any choice of the comparator $\boldsymbol{u} \in \mathcal{W}$. We consider EW with a fixed learning rate η and a prior distribution P_1 that is uniform over $\mathcal{W}$. At each trial t, after observing the loss $\langle \boldsymbol{w}_t, \boldsymbol{g}_t \rangle$, the algorithm constructs a random, unbiased estimate $\tilde{\boldsymbol{g}}_t$ of the loss vector $\boldsymbol{g}_t$ (described below), and uses this estimate to update the posterior. Since the projection step can be dropped (as P_1 is supported on $\mathcal{W}$), the greedy and lazy versions of EW coincide and the posterior is given by $\mathrm{d}P_t(\boldsymbol{w}) \propto \exp(-\eta \sum_{s=1}^{t-1} \langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_s \rangle) \mathrm{d}\boldsymbol{w}$ for all $\boldsymbol{w} \in \mathcal{W}$. Defining $\boldsymbol{\theta}_t = -\eta \sum_{s=1}^{t-1} \tilde{\boldsymbol{g}}_s$ (with $\boldsymbol{\theta}_1 = \boldsymbol{0}$), we can concisely write:

$$\mathrm{d}P_{t+1}(\boldsymbol{w}) = e^{\langle \boldsymbol{w}, \boldsymbol{\theta}_t \rangle - F(\boldsymbol{\theta}_t)} \mathrm{d}\boldsymbol{w} \quad \forall \boldsymbol{w} \in \mathcal{W}, \qquad \text{where } F(\boldsymbol{\theta}) = \ln \int_{\mathcal{W}} e^{\langle \boldsymbol{w}, \boldsymbol{\theta} \rangle} \, \mathrm{d}\boldsymbol{w}$$

is the cumulant generating function. At trial t, the EW algorithm samples $\boldsymbol{w}_t \sim Q_t$, where $Q_t = (1-\gamma)P_t + \gamma R$ for $\gamma \in (0,1)$ is a mixture of the posterior P_t and a fixed "exploration" distribution R. The exploration distribution is chosen to be *John's exploration*, defined as follows (Bubeck et al., 2012). Let $\mathcal{K}$ be the ellipsoid of minimal volume *enclosing* $\mathcal{W}$:

$$\mathcal{K} = \{\boldsymbol{w} \in \mathbb{R}^d \colon (\boldsymbol{w} - \boldsymbol{w}_0)^\intercal \boldsymbol{H}^{-1} (\boldsymbol{w} - \boldsymbol{w}_0) \le 1\} \tag{23}$$

for some positive definite matrix $\boldsymbol{H}$ and $\boldsymbol{w}_0 \in \mathbb{R}^d$. In what follows we assume without loss of generality that $\mathcal{W}$ is centered in the sense that $\boldsymbol{w}_0 = \boldsymbol{0}$ (otherwise all $\boldsymbol{w} \in \mathcal{W}$ need to be shifted by $\boldsymbol{w}_0$). Bubeck et al. (2012) show that one can choose $M \le d(d+1)/2 + 1$ contact points $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M \in \mathcal{K} \cap \mathcal{W}$, and a distribution R over these points that satisfies:

$$\mathop{\mathbb{E}}_{\boldsymbol{w} \sim R}[\boldsymbol{w}\boldsymbol{w}^\intercal] = \frac{1}{d}\boldsymbol{H}. \tag{24}$$

The estimate $\tilde{\boldsymbol{g}}_t$ is constructed based on the observed loss $\langle \boldsymbol{w}_t, \boldsymbol{x}_t \rangle$, by:

$$\tilde{\boldsymbol{g}}_t = \langle \boldsymbol{w}_t, \boldsymbol{g}_t \rangle \left(\mathop{\mathbb{E}}_{Q_t}[\boldsymbol{w}\boldsymbol{w}^\intercal] \right)^{-1} \boldsymbol{w}_t.$$

We now show the following regret bound for the resulting algorithm:

Theorem 14 *Assume the losses are bounded: $|\langle \boldsymbol{w}, \boldsymbol{g}_t \rangle| \le 1$ for all $\boldsymbol{w} \in \mathcal{W}$ and all t. Let $\eta = \sqrt{\frac{\nu \ln T}{3dT}}$, where $\nu = O(d)$ is the self-concordant barrier parameter of F^*, and let $\gamma = \eta d$. Then the expected regret for the EW algorithm described above is bounded by*

$$\mathbb{E}[\mathcal{R}_T(\boldsymbol{u})] \le 2\sqrt{3\nu d T \ln T} + 2 = O(d\sqrt{T \ln T}).$$

Proof We first verify that the estimate $\tilde{\boldsymbol{g}}_t$ of $\boldsymbol{g}_t$ is unbiased:

$$\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}[\tilde{\boldsymbol{g}}_t] = \mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}\left[\left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^\intercal]\right)^{-1}\boldsymbol{w}_t\langle \boldsymbol{w}_t, \boldsymbol{g}_t\rangle\right] = \left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^\intercal]\right)^{-1}\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}[\boldsymbol{w}_t\boldsymbol{w}_t^\intercal]\,\boldsymbol{g}_t = \boldsymbol{g}_t.$$

Furthermore, due to the inclusion of the exploration distribution R, we have:

$$\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^\intercal] = (1-\gamma)\mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}[\boldsymbol{w}\boldsymbol{w}^\intercal] + \gamma\mathop{\mathbb{E}}_{\boldsymbol{w}\sim R}[\boldsymbol{w}\boldsymbol{w}^\intercal] \succeq \frac{\gamma}{d}\boldsymbol{H},$$

(where $\boldsymbol{A}\succeq\boldsymbol{B}$ means $\boldsymbol{A}-\boldsymbol{B}$ is positive semidefinite), and hence for any $\boldsymbol{u}\in\mathcal{W}$:

$$\left\langle \boldsymbol{u}, \left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^\intercal]\right)^{-1}\boldsymbol{u}\right\rangle \le \left\langle \boldsymbol{u}, \frac{d}{\gamma}\boldsymbol{H}^{-1}\boldsymbol{u}\right\rangle \le \frac{d}{\gamma}, \tag{25}$$

where the last inequality is from the fact that $\mathcal{W}\subseteq\mathcal{K}$ and from the definition of $\mathcal{K}$ in (23). This, however, implies that the linear losses induced by $\tilde{\boldsymbol{g}}_t$ are bounded for any $\boldsymbol{u}\in\mathcal{W}$:

$$\begin{aligned}\langle \boldsymbol{u}, \tilde{\boldsymbol{g}}_t\rangle &= \langle \boldsymbol{w}_t, \boldsymbol{g}_t\rangle \left\langle \boldsymbol{u}, \left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^\intercal]\right)^{-1}\boldsymbol{w}_t\right\rangle \\ &\le |\langle \boldsymbol{w}_t, \boldsymbol{g}_t\rangle| \left\langle \boldsymbol{w}_t, \left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^\intercal]\right)^{-1}\boldsymbol{w}_t\right\rangle^{1/2} \left\langle \boldsymbol{u}, \left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^\intercal]\right)^{-1}\boldsymbol{u}\right\rangle^{1/2} \le \frac{d}{\gamma},\end{aligned} \tag{26}$$

where the first inequality is from the Cauchy-Schwarz inequality (for positive semidefinite $\boldsymbol{A}$, $\boldsymbol{x}^\intercal\boldsymbol{A}\boldsymbol{y}\le(\boldsymbol{x}^\intercal\boldsymbol{A}\boldsymbol{x})^{1/2}(\boldsymbol{y}^\intercal\boldsymbol{A}\boldsymbol{y})^{1/2}$), while the second inequality is due to assumption $|\langle \boldsymbol{w}, \boldsymbol{g}_t\rangle|\le 1$ and due to (25) applied twice (first to $\boldsymbol{u}$ and then to $\boldsymbol{w}_t$).

Let $\boldsymbol{\mu}_t$ be the mean value of P_t: $\boldsymbol{\mu}_t = \mathbb{E}_{P_t}[\boldsymbol{w}]$. As a general property of exponential families or as a consequence of Theorem 4, we have $\boldsymbol{\mu}_t = \nabla F(\boldsymbol{\theta}_t)$, and $\boldsymbol{\mu}_t$ and $\boldsymbol{\theta}_t$ are conjugate parameters of the exponential family. Let us fix a comparator $\boldsymbol{u}\in\mathcal{W}$ and define $P_{\boldsymbol{u}}$ to be the member of the exponential family with cumulant generating function F that has mean value $\boldsymbol{u}$: $\mathbb{E}_{\boldsymbol{w}\sim P_{\boldsymbol{u}}}[\boldsymbol{w}] = \boldsymbol{u}$. We now apply Lemma 1 for the EW algorithm on the sequence of linear losses induced by $\tilde{\boldsymbol{g}}_1,\dots,\tilde{\boldsymbol{g}}_T$ to get:

$$\begin{aligned}\sum_{t=1}^T\langle \boldsymbol{\mu}_t - \boldsymbol{u}, \tilde{\boldsymbol{g}}_t\rangle &= \sum_{t=1}^T\langle \boldsymbol{\mu}_t, \tilde{\boldsymbol{g}}_t\rangle - \sum_{t=1}^T\mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_{\boldsymbol{u}}}[\langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_t\rangle] \\ &\le \frac{1}{\eta}\mathrm{KL}(P_{\boldsymbol{u}}\|P_1) + \sum_{t=1}^T\langle \boldsymbol{\mu}_t, \tilde{\boldsymbol{g}}_t\rangle + \frac{1}{\eta}\ln\mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}\left[e^{-\eta\langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_t\rangle}\right]\end{aligned}$$

(note that in this section we use $\boldsymbol{\mu}_t$ to denote the mean of P_t, while $\boldsymbol{w}_t$ is reserved for the randomized action at trial t sampled from Q_t). Since $P_{\boldsymbol{u}}$ and P_1 are members of the same exponential family, the KL-term can be re-expressed using Lemma 10:

$$\mathrm{KL}(P_{\boldsymbol{u}}\|P_1) = D_{F^*}(\boldsymbol{u}\|\boldsymbol{\mu}_1) = F^*(\boldsymbol{u}) - F^*(\boldsymbol{\mu}_1) - \underbrace{\nabla F^*(\boldsymbol{\mu}_1)^\intercal}_{\boldsymbol{0}}(\boldsymbol{\mu}-\boldsymbol{\mu}_1) = F^*(\boldsymbol{u}) - F^*(\boldsymbol{\mu}_1),$$

where we used the fact that $\boldsymbol{\mu}_1$ has conjugate parameter $\boldsymbol{\theta}_1 = \boldsymbol{0}$, and thus $\nabla F^*(\boldsymbol{\mu}_1) = \boldsymbol{\theta}_1 = \boldsymbol{0}$. To bound the mixability gap, we will now use that by assumption $\eta = \frac{\gamma}{d}$, so that by (26) we have

$|\eta\langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_t\rangle| \leq 1$ for any $\boldsymbol{w} \in \mathcal{W}$. Using the fact that $e^{-s} \leq 1 - s + s^2$ holds for $s \geq -1$, and combining with $\ln(1+x) \leq x$ gives:

$$\begin{aligned}\langle \boldsymbol{\mu}_t, \tilde{\boldsymbol{g}}_t\rangle + \frac{1}{\eta}\ln \mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}\left[e^{-\eta\langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_t\rangle}\right] &\leq \langle \boldsymbol{\mu}_t, \tilde{\boldsymbol{g}}_t\rangle + \frac{1}{\eta}\ln\left(1 + \mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}\left[-\eta\langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_t\rangle + \eta^2\langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_t\rangle^2\right]\right)\\ &\leq \underbrace{\langle \boldsymbol{\mu}_t, \tilde{\boldsymbol{g}}_t\rangle - \mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}[\langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_t\rangle]}_{=0} + \eta \mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}\left[\langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_t\rangle^2\right]\\ &= \eta \tilde{\boldsymbol{g}}_t^{\intercal} \mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}]\,\tilde{\boldsymbol{g}}_t.\end{aligned}$$

Combining the bounds on the KL-term and the mixability gap gives:

$$\sum_{t=1}^{T}\langle \boldsymbol{\mu}_t - \boldsymbol{u}, \tilde{\boldsymbol{g}}_t\rangle \leq \frac{F^*(\boldsymbol{u}) - F^*(\boldsymbol{\mu}_1)}{\eta} + \eta\sum_{t=1}^{T}\tilde{\boldsymbol{g}}_t^{\intercal}\mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}]\,\tilde{\boldsymbol{g}}_t. \tag{27}$$

We can use this result to bound the regret of the original algorithm in the following way. First, note that:

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}[\langle \boldsymbol{w}_t - \boldsymbol{u}, \boldsymbol{g}_t\rangle] &= \gamma\big\langle \mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim R}[\boldsymbol{w}_t] - \boldsymbol{u}, \boldsymbol{g}_t\big\rangle + (1-\gamma)\big\langle \mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim P_t}[\boldsymbol{w}_t] - \boldsymbol{u}, \boldsymbol{g}_t\big\rangle\\ &\leq 2\gamma + (1-\gamma)\big\langle \boldsymbol{\mu}_t - \boldsymbol{u}, \boldsymbol{g}_t\big\rangle = 2\gamma + (1-\gamma)\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}\left[\langle \boldsymbol{\mu}_t - \boldsymbol{u}, \tilde{\boldsymbol{g}}_t\rangle\right],\end{aligned}$$

where the random quantity in the last expectation is $\tilde{\boldsymbol{g}}_t$, because it depends on $\boldsymbol{w}_t$. Therefore:

$$\begin{aligned}\sum_{t=1}^{T}\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}[\langle \boldsymbol{w}_t - \boldsymbol{u}, \boldsymbol{g}_t\rangle] &\leq 2\gamma T + (1-\gamma)\sum_{t=1}^{T}\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}[\langle \boldsymbol{\mu}_t - \boldsymbol{u}, \tilde{\boldsymbol{g}}_t\rangle]\\ &\leq 2\gamma T + \frac{F^*(\boldsymbol{u}) - F^*(\boldsymbol{\mu}_1)}{\eta} + \eta(1-\gamma)\sum_{t=1}^{T}\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}\left[\tilde{\boldsymbol{g}}_t^{\intercal}\mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}]\,\tilde{\boldsymbol{g}}_t\right]\\ &\leq 2\gamma T + \frac{F^*(\boldsymbol{u}) - F^*(\boldsymbol{\mu}_1)}{\eta} + \eta\sum_{t=1}^{T}\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}\left[\tilde{\boldsymbol{g}}_t^{\intercal}\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}]\,\tilde{\boldsymbol{g}}_t\right], \quad (28)\end{aligned}$$

where the second inequality is from (27), while the last inequality is due to:

$$\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}] = (1-\gamma)\mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}] + \gamma\mathop{\mathbb{E}}_{\boldsymbol{w}\sim R}[\boldsymbol{w}\boldsymbol{w}^{\intercal}] \succeq (1-\gamma)\mathop{\mathbb{E}}_{\boldsymbol{w}\sim P_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}].$$

Using the definition of $\tilde{\boldsymbol{g}}_t$ and $\langle \boldsymbol{w}_t, \boldsymbol{g}_t\rangle^2 \leq 1$, we further bound:

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}\left[\tilde{\boldsymbol{g}}_t^{\intercal}\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}]\,\tilde{\boldsymbol{g}}_t\right] &\leq \mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}\left[\boldsymbol{w}_t^{\intercal}\left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}]\right)^{-1}\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}]\left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}]\right)^{-1}\boldsymbol{w}_t\right]\\ &= \sum_{t=1}^{T}\mathop{\mathbb{E}}_{\boldsymbol{w}_t\sim Q_t}\left[\operatorname{Tr}\left(\left(\mathop{\mathbb{E}}_{\boldsymbol{w}\sim Q_t}[\boldsymbol{w}\boldsymbol{w}^{\intercal}]\right)^{-1}\boldsymbol{w}_t\boldsymbol{w}_t^{\intercal}\right)\right]\\ &= \sum_{t=1}^{T}\operatorname{Tr}(\boldsymbol{I}) = Td.\end{aligned}$$

Plugging the above into (28) and taking expectation with respect to the randomness of the algorithm results in the following bound on the expected regret:

$$\mathbb{E}[\mathcal{R}_T(\boldsymbol{u})] = \mathbb{E}\left[\sum_{t=1}^{T} \mathop{\mathbb{E}}_{\boldsymbol{w}_t \sim Q_t}[\langle \boldsymbol{w}_t - \boldsymbol{u}, \boldsymbol{g}_t \rangle]\right] \leq 2\gamma T + \frac{F^*(\boldsymbol{u}) - F^*(\boldsymbol{\mu}_1)}{\eta} + \eta T d.$$

What is left to bound is $F^*(\boldsymbol{u}) - F^*(\boldsymbol{\mu}_1)$. To this end, define the Minkowski function (Abernethy et al., 2012) on $\mathcal{W}$ as:

$$\pi_{\boldsymbol{\mu}}(\boldsymbol{w}) = \inf\{t \geq 0 \colon \boldsymbol{\mu} + t^{-1}(\boldsymbol{w} - \boldsymbol{\mu}) \in \mathcal{W}\}.$$

Bubeck and Eldan (2015) show that F^* is a ν-self concordant barrier on $\mathcal{W}$ with $\nu = O(d)$. Using this property and Theorem 2.2 from Abernethy et al. (2012) we get:

$$F^*(\boldsymbol{u}) - F^*(\boldsymbol{\mu}_1) \leq \nu \ln\left(\frac{1}{1 - \pi_{\boldsymbol{\mu}_1}(\boldsymbol{u})}\right).$$

If $\boldsymbol{u}$ is such that $\pi_{\boldsymbol{\mu}_1}(\boldsymbol{u}) \leq 1 - \frac{1}{T}$, then $F^*(\boldsymbol{u}) - F^*(\boldsymbol{\mu}_1) \leq \nu \ln T$. On the other hand, if $\pi_{\boldsymbol{\mu}_1}(\boldsymbol{u}) \leq 1 - \frac{1}{T}$, we define a new comparator $\boldsymbol{u}' = (1 - \frac{1}{T})\boldsymbol{u} + \frac{1}{T}\boldsymbol{\mu}_1$, for which $\pi_{\boldsymbol{\mu}_1}(\boldsymbol{u}') \leq 1 - \frac{1}{T}$ (Abernethy et al., 2012), and use the regret bound above for $\boldsymbol{u}'$ to get:

$$\begin{aligned}\mathbb{E}[\mathcal{R}_T(\boldsymbol{u})] &= \mathbb{E}[\mathcal{R}_T(\boldsymbol{u}')] + \sum_{t=1}^{T}\langle \boldsymbol{u}' - \boldsymbol{u}, \boldsymbol{g}_t\rangle = \mathbb{E}[\mathcal{R}_T(\boldsymbol{u}')] + \frac{1}{T}\sum_{t=1}^{T}\langle \boldsymbol{\mu}_1 - \boldsymbol{u}, \boldsymbol{g}_t\rangle \\ &\leq 2\gamma T + \frac{F^*(\boldsymbol{u}') - F^*(\boldsymbol{\mu}_1)}{\eta} + \eta T d + 2 \leq 2\gamma T + \frac{\nu \ln T}{\eta} + \eta T d + 2.\end{aligned}$$

Recalling that $\gamma = \eta d$ and tuning $\eta = \sqrt{\frac{\nu \ln T}{3dT}}$ gives the claimed bound. ■