Second-order Quantile Methods
for Experts and Combinatorial Games

Wouter M. Koolen   WOUTER.KOOLEN@QUT.EDU.AU
Queensland University of Technology, Brisbane, Australia and UC Berkeley

Tim van Erven   TIM@TIMVANERVEN.NL
Leiden University, the Netherlands

Abstract

We aim to design strategies for sequential decision making that adjust to the difficulty of the learning problem. We study this question both in the setting of prediction with expert advice, and for more general combinatorial decision tasks. We are not satisfied with just guaranteeing minimax regret rates, but we want our algorithms to perform significantly better on easy data. Two popular ways to formalize such adaptivity are second-order regret bounds and quantile bounds. The underlying notions of 'easy data', which may be paraphrased as "the learning problem has small variance" and "multiple decisions are useful", are synergetic. But even though there are sophisticated algorithms that exploit one of the two, no existing algorithm is able to adapt to both.

The difficulty in combining the two notions lies in tuning a parameter called the learning rate, whose optimal value behaves non-monotonically. We introduce a potential function for which (very surprisingly!) it is sufficient to simply put a prior on learning rates; an approach that does not work for any previous method. By choosing the right prior we construct efficient algorithms and show that they reap both benefits by proving the first bounds that are both second-order and incorporate quantiles.

Keywords: Online learning, prediction with expert advice, combinatorial prediction, easy data

1. Introduction

We study the design of adaptive algorithms for online learning (Cesa-Bianchi and Lugosi, 2006). Our work starts in the *hedge setting* (Freund and Schapire, 1997), a core instance of prediction with expert advice (Vovk, 1990, 1998; Littlestone and Warmuth, 1994) and online convex optimization (Shalev-Shwartz, 2011). Each round $t = 1, 2, \ldots$ the learner plays a probability vector $\boldsymbol{w}_t$ on K experts, the environment assigns a bounded loss to each expert in the form of a vector $\boldsymbol{\ell}_t \in [0,1]^K$, and the learner incurs loss given by the dot product $\boldsymbol{w}_t^\mathsf{T} \boldsymbol{\ell}_t$. The learner's goal is to perform almost as well as the best expert, without making any assumptions about the genesis of the losses. Specifically, the learner's performance compared to expert k is $r_t^k = \boldsymbol{w}_t^\mathsf{T} \boldsymbol{\ell}_t - \ell_t^k$, and after any number of rounds T the goal is to have small *regret* $R_T^k = \sum_{t=1}^T r_t^k$ with respect to every expert k.

The *Hedge* algorithm by Freund and Schapire (1997) ensures

$$R_T^k \ \prec \ \sqrt{T \ln K} \qquad \text{for each expert } k \qquad \qquad \text{classic} \quad (1)$$

(with $\prec$ denoting moral inequality, i.e. suppressing details inappropriate for this introduction), which is tight for adversarial (worst-case) losses (Cesa-Bianchi and Lugosi, 2006). Yet one can ask whether the worst case is also the common case, and indeed two lines of research show that

this bound can be improved greatly in various important scenarios. The first line of approaches (Cesa-Bianchi et al., 2007; Hazan and Kale, 2010; Chiang et al., 2012; Steinhardt and Liang, 2014; De Rooij et al., 2014; Gaillard et al., 2014) obtains

$$R_T^k \prec \sqrt{V_T^k \ln K} \qquad \text{for each expert } k. \qquad \text{(second order)}$$

That is, T can be reduced to V_T^k, which stands for some (there are various) kind of cumulative variance or related second-order quantity. This variance is then often shown to be small $V_T^k \ll T$ in important regimes like stochastic data (where it is typically bounded). The second line, independently in parallel, shows how to reduce the dependence on the number of experts K whenever multiple experts perform well. This is expected to occur, for example, when experts are constructed by finely discretising the parameters of a (probabilistic) model, or when learning sub-algorithms are used as experts. The resulting so-called *quantile bounds* (see Chaudhuri et al. 2009; Chernov and Vovk 2010; Luo and Schapire 2014) of the form

$$\min_{k \in \mathcal{K}} R_T^k \prec \sqrt{T\big(-\ln \pi(\mathcal{K})\big)} \qquad \text{for each subset } \mathcal{K} \text{ of experts} \qquad \text{(quantile)}$$

improve K to the reciprocal of the combined prior mass $\pi(\mathcal{K})$, at the cost of now comparing to the worst expert among $\mathcal{K}$, so intuitively the best guarantee is obtained for $\mathcal{K}$ the set of "sufficiently good" experts. (There is no requirement that the prior $\pi(k)$ is uniform, and consequently quantile bounds imply the closely related bounds with non-uniform priors, studied e.g. by Hutter and Poland 2005.) As these two types of improvement are complementary, we would like to combine them in a single algorithm. However, the mentioned two approaches are based on incompatible techniques, which until now have refused to coexist.

First Contribution We develop a new prediction strategy called *Squint* by putting priors on a parameter called the *learning rate* and on experts. For Squint, we prove the first bounds that incorporate both quantile and variance guarantees at the same time:

$$R_T^{\mathcal{K}} \prec \sqrt{V_T^{\mathcal{K}}\big(C_{\text{lr}} - \ln \pi(\mathcal{K})\big)} \qquad \text{for each subset } \mathcal{K} \text{ of experts,} \qquad (2)$$

where $R_T^{\mathcal{K}} = \mathbb{E}_{\pi(k|\mathcal{K})} R_T^k$ and $V_T^{\mathcal{K}} = \mathbb{E}_{\pi(k|\mathcal{K})} V_T^k$ denote the average (under the prior) among the reference experts $k \in \mathcal{K}$ of the regret $R_T^k = \sum_{t=1}^T r_t^k$ and the (uncentered) variance of the excess losses $V_T^k = \sum_{t=1}^T (r_t^k)^2$. The overhead C_{lr} for learning the optimal learning rate is specified below.

Some variance measures in the literature are easily interpretable functions of the data (Hazan and Kale, 2010; Chiang et al., 2012; Steinhardt and Liang, 2014), but unfortunately such measures grow linearly even when the learning problem is actually very easy in the sense that the best expert linearly outperforms all others. Other variance measures, including our $V_T^{\mathcal{K}}$, are more opaque, because in addition to the data they also depend on the algorithm (Cesa-Bianchi et al., 2007; De Rooij et al., 2014; Gaillard et al., 2014), but these measures have the advantage that variance and hence regret stop accumulating whenever the weights concentrate, as will happen when one expert is clearly the best. Furthermore, Gaillard et al. (2014) show that second-order bounds like (2) imply small regret over experts with small losses (L_T^*-bounds) and also guarantee bounded regret both in expectation and with high probability in stochastic settings with a unique best expert. Finally, if we weaken $V_T^{\mathcal{K}}$ in (2) using $(r_t^k)^2 \le |r_t^k|$, then we recover the main result from independent concurrent work by Luo and Schapire (2015, Theorem 1), which they apply to sleeping experts and tracking problems.

2

We will instantiate our scheme three times, varying the prior distribution of the learning rate, to obtain three interesting bounds. First, for the *uniform prior*, we obtain an efficient algorithm with $C_{\mathrm{lr}} = \ln V_T^{\mathcal{K}}$. Then we consider a prior that we call the *CV prior*, because it was introduced by Chernov and Vovk (2010) (to get quantile bounds), and we improve the bound to $C_{\mathrm{lr}} = \ln \ln V_T^{\mathcal{K}}$. As we consider $\ln \ln x$ to be essentially constant, this algorithm achieves our goal of combining the benefits of second-order bounds with quantiles, but unfortunately it does not have an efficient implementation. Finally, by considering the improper(!) *log-uniform prior*, we get the best of both worlds: an algorithm that is both efficient and achieves our goal with $C_{\mathrm{lr}} = \ln \ln T$. The efficient algorithms for the uniform and the log-uniform prior both perform just K operations per round, and are hence as widely applicable as vanilla Hedge.

Combinatorial games We then consider a more sophisticated setting, where instead of experts $k \in \{1, \ldots, K\}$, the elementary actions are combinatorial concepts from some class $\mathcal{C} \subseteq \{0, 1\}^K$. Many theoretically interesting and important real-world online decision problems are of this form, for example subsets (sub-problem of Principal Component Analysis), lists (or ranking), permutations (scheduling), spanning trees (communication), paths through a fixed graph (routing), etc. (see for instance Takimoto and Warmuth 2003; Kalai and Vempala 2005; Warmuth and Kuzmin 2008; Helmbold and Warmuth 2009; Cesa-Bianchi and Lugosi 2012; Warmuth et al. 2014; Audibert et al. 2014). The combinatorial structure is reflected in the loss, which decomposes into a sum of coordinate losses. That is, the loss of concept $c \in \mathcal{C}$ is $c^\mathsf{T} \ell$ for some loss vector $\ell \in [0, 1]^K$. This is natural: for example the loss of a path is the total loss of its edges. Koolen et al. (2010) develop Component Hedge (of the Mirror Descent family), with regret at most

$$R_T^{\boldsymbol{c}} \prec \sqrt{TK \mathrm{comp}(\mathcal{C})} \qquad \text{for each concept } \boldsymbol{c} \in \mathcal{C}, \tag{3}$$

where $\mathrm{comp}(\mathcal{C})$, the analog of $\ln K$ for experts, measures the complexity (entropy) of the combinatorial class $\mathcal{C}$. Luo and Schapire (2014) derive $\sqrt{T}$ quantile bounds for online convex optimization, but no second-order quantile methods were previously known for combinatorial games.

Second Contribution We extend our approach to combinatorial games and obtain the first algorithm with a second-order quantile regret bound. In the combinatorial domain, the role of the reference set of experts $\mathcal{K}$ is subsumed by a "mean concept" vector $\boldsymbol{v}$ in the convex hull $\mathrm{conv}(\mathcal{C})$ of $\mathcal{C}$, where v_k represents the probability of including coordinate k, and the regret becomes the coordinate-wise average regret $R_T^{\boldsymbol{v}} = \sum_{t,k} v_k r_t^k$. Our new predictor *Component iProd* keeps this regret below

$$R_T^{\boldsymbol{v}} \prec \sqrt{V_T^{\boldsymbol{v}} \big(\mathrm{comp}(\boldsymbol{v}) + K C_{\mathrm{lr}}\big)} \qquad \text{for each } \boldsymbol{v} \in \mathrm{conv}(\mathcal{C}). \tag{4}$$

Component iProd is a computationally efficient drop-in replacement for Component Hedge, which improves the factor TK to the averaged variance $V_T^{\boldsymbol{v}} = \sum_{t,k} v_k (r_t^k)^2$ and the complexity $\mathrm{comp}(\mathcal{C})$ of the whole class $\mathcal{C}$ to the prior entropy $\mathrm{comp}(\boldsymbol{v})$.

Even if we disregard computational efficiency, our bound (4) is not a straightforward consequence of the experts bound (2) applied with one expert for each concept, paralleling the fact that (3) does not follow from (1). The reason is that we would obtain a bound with per-concept variance $\sum_t (\sum_k v_k r_t^k)^2$ instead, which can overshoot even the straight-up worst-case bound (3) by a factor of $\sqrt{K}$ (Koolen et al. (2010) call this the *range factor problem*). To avoid this problem, our method is "collapsed" (like Component Hedge): it only maintains first and second order statistics about the K coordinates separately, not about concepts as a whole.

Related work Obtaining bounds for easy data in the experts setting is typically achieved by adaptively tuning a *learning rate*, which is a parameter found in many algorithms. Schemes for *choosing* the learning rate on-line are built by Auer et al. (2002); Cesa-Bianchi et al. (2007); Hazan and Kale (2010); De Rooij et al. (2014); Gaillard et al. (2014); Wintenberger (2014). These schemes typically choose a monotonically decreasing sequence of learning rates to prove a certain regret bound.

Other approaches try to *aggregate* multiple learning rates. The motivations and techniques here show extreme diversity, ranging from *drifting games* by Chaudhuri et al. (2009); Luo and Schapire (2014), and *defensive forecasting* by Chernov and Vovk (2010) to *minimax relaxations* by Rakhlin et al. (2013) and *budgeted timesharing* by Koolen et al. (2014). The last scheme is of note, as it does not aggregate to reproduce a bound of a certain form, but rather to compete with the optimally tuned learning rate for the Hedge algorithm.

In the related setting of statistical learning, Grünwald has been able to compete with the optimally tuned learning rate by measuring the (lack of) empirical convexity of statistical models (Grünwald, 2011), and through a method based on online-to-batch conversion (Grünwald, 2012).

Outline We introduce the Squint prediction rule for experts in Section 2. In Section 3 we motivate three choices for the prior on the learning rate, discuss the resulting algorithms and prove second-order quantile regret bounds. In Section 4 we extend Squint to combinatorial prediction tasks. We conclude with open problems in Section 5.

2. *Squint*: a Second-order Quantile Method for Experts

Let us review the expert setting protocol to fix notation. In round t the algorithm plays a probability distribution $\boldsymbol{w}_t$ on K experts and encounters loss $\boldsymbol{\ell}_t \in [0,1]^K$. The *instantaneous regret* of the algorithm compared to expert k is $r_t^k = \boldsymbol{w}_t^\mathsf{T}\boldsymbol{\ell}_t - \ell_t^k = (\boldsymbol{w}_t - \boldsymbol{e}_k)^\mathsf{T}\boldsymbol{\ell}_t$, where $\boldsymbol{e}_k$ is the unit vector in direction $k \in \{1, \ldots, K\}$. Let $R_T^k = \sum_{t=1}^T r_t^k$ be the total regret compared to expert k and let $V_T^k = \sum_{t=1}^T (r_t^k)^2$ be the cumulative uncentered variance of the instantaneous regrets.

The central building block of our approach is a *potential function* Φ that maps sequences of instantaneous regret vectors $\boldsymbol{r}_{1:T} = \langle \boldsymbol{r}_1, \ldots, \boldsymbol{r}_T \rangle$ of any length $T \geq 0$ to numbers. Potential functions are staple online learning tools (Cesa-Bianchi and Lugosi, 2006; Abernethy et al., 2014). We advance the following schema, which we call *Squint* (for *second-order quantile integral*). It consists of the potential function and associated prediction rule

$$\Phi(\boldsymbol{r}_{1:T}) = \underset{\pi(k)\gamma(\eta)}{\mathbb{E}} \left[e^{\eta R_T^k - \eta^2 V_T^k} - 1 \right], \qquad \boldsymbol{w}_{T+1} = \frac{\mathbb{E}_{\pi(k)\gamma(\eta)} \left[e^{\eta R_T^k - \eta^2 V_T^k} \eta \boldsymbol{e}_k \right]}{\mathbb{E}_{\pi(k)\gamma(\eta)} \left[e^{\eta R_T^k - \eta^2 V_T^k} \eta \right]}, \quad (5)$$

where the expectation is taken under prior distributions π on experts $k \in \{1, \ldots, K\}$ and γ on learning rates $\eta \in [0, 1/2]$ that are parameters of Squint. Although we always take these priors to be independent, our results generalize to dependent priors as well[1]. We will show in a moment that Squint ensures that the potential remains $\Phi(\boldsymbol{r}_{1:T}) \leq 0$ non-positive. Let us first investigate why non-positivity is desirable. To gain a quick-and-dirty appreciation for this, suppose that $\mathcal{K} \subseteq \{1, \ldots, K\}$ is the *reference set* of experts with good performance. Let us abbreviate their average regret and variance to $R = R_T^\mathcal{K} = \mathbb{E}_{\pi(k|\mathcal{K})} R_T^k$ and $V = V_T^\mathcal{K} = \mathbb{E}_{\pi(k|\mathcal{K})} V_T^k$. Furthermore, imagine for

---

1. Dependent priors might sometimes be useful: singling out a single expert k^* and putting prior mass $1 - 1/T$ on the pair $(k, \eta) = (k^*, 1/T)$ recovers the bound of $(\mathcal{A}, \mathcal{B})$-Prod by Sani et al. (2014).

simplicity that an oracle would tell us the optimized learning rate $\hat{\eta} = \frac{R}{2V}$ and that the prior γ would put all its mass on $\hat{\eta}$. Then non-positive potential $\Phi(\boldsymbol{r}_{1:T}) \leq 0$ implies

$$1 \;\geq\; \underset{\pi(k)}{\mathbb{E}}\left[e^{\hat{\eta}R_T^k - \hat{\eta}^2 V_T^k}\right] \;\geq\; \pi(\mathcal{K}) \underset{\pi(k|\mathcal{K})}{\mathbb{E}}\left[e^{\hat{\eta}R_T^k - \hat{\eta}^2 V_T^k}\right] \overset{\text{Jensen}}{\geq} \pi(\mathcal{K})e^{\hat{\eta}R - \hat{\eta}^2 V} \;=\; \pi(\mathcal{K})e^{\frac{R^2}{4V}}, \quad (6)$$

which immediately yields the desired variance-with-quantiles bound

$$R_T^{\mathcal{K}} \;\leq\; 2\sqrt{V_T^{\mathcal{K}}(-\ln \pi(\mathcal{K}))}. \qquad \text{(oracle learning rate)}$$

This motivates that $\Phi(\boldsymbol{r}_{1:T}) \leq 0$ is desirable, which Squint achieves by always decreasing it:

Lemma 1 *Let $\gamma(\eta)$ be supported on $[0, 1/2]$. Then Squint (5) ensures that, for any loss sequence $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_T$ in $[0, 1]^K$,*

$$\Phi(\boldsymbol{r}_{1:T}) \;\leq\; \ldots \;\leq\; \Phi(\emptyset) \;=\; 0. \qquad (7)$$

Proof The key role is played by the upper bound (Cesa-Bianchi and Lugosi, 2006, Lemma 2.4)

$$e^{x-x^2} - 1 \;\leq\; x \qquad \text{for } x \geq -1/2. \qquad (8)$$

Applying this to $\eta r_{T+1}^k \geq -1/2$, we bound the increase $\Phi(\boldsymbol{r}_{1:T+1}) - \Phi(\boldsymbol{r}_{1:T})$ of the potential by

$$\underset{\pi(k)\gamma(\eta)}{\mathbb{E}}\left[e^{\eta R_T^k - \eta^2 V_T^k}\left(e^{\eta r_{T+1}^k - (\eta r_{T+1}^k)^2} - 1\right)\right] \overset{(8)}{\leq} \underset{\pi(k)\gamma(\eta)}{\mathbb{E}}\left[e^{\eta R_T^k - \eta^2 V_T^k} \eta (\boldsymbol{w}_{T+1} - \boldsymbol{e}_k)^\mathsf{T} \boldsymbol{\ell}_{T+1}\right] = 0,$$

where the last identity holds because the algorithm's weights (5) have been chosen to satisfy it. ∎

In Section 3 we will make the proof sketch from (6) rigorous. We pursue priors $\gamma(\eta)$ that (a) pack plenty of mass close to $\hat{\eta}$, wherever it may end up; and (b) admit efficient computation of the weights $\boldsymbol{w}_{T+1}$ by means of a closed-form formula for its integrals over η. We conclude this section by putting Squint in context.

Discussion The Squint potential is a function of the vector $\sum_{t=1}^{T} \boldsymbol{r}_t$ of cumulative regrets, but also of its sum of squares, which is essential for second-order bounds. Squint is an *anytime* algorithm, i.e. it has no built-in dependence on an eventual time horizon, and its regret bounds hold at any time T of evaluation. In addition Squint is *timeless* in the sense of De Rooij et al. (2014), meaning that its predictions (current and future) are unaffected by inserting rounds with $\boldsymbol{\ell} = \boldsymbol{0}$.

The Squint potential is an average of exponentiated negative "losses" (derived from the regret) under product prior $\pi(k)\gamma(\eta)$, reminiscent of the exponential weights analysis potential. Our Squint weights could be viewed as exponential weights, but, intriguingly, for *another prior*, with $\gamma(\eta)$ replaced by $\gamma(\eta)\eta$. Mysteriously, playing the latter controls the former.

The bound (8) is hard-coded in our Squint potential function and algorithm. To instead delay this bound to the analysis, we might introduce the alternative *iProd* (for *integrated products*) scheme

$$\Phi(\boldsymbol{r}_{1:T}) = \underset{\pi(k)\gamma(\eta)}{\mathbb{E}}\left[\left(\prod_{t=1}^{T}(1 + \eta r_t^k)\right) - 1\right], \quad \boldsymbol{w}_{T+1} = \frac{\mathbb{E}_{\pi(k)\gamma(\eta)}\left[\left(\prod_{t=1}^{T}(1 + \eta r_t^k)\right)\eta \boldsymbol{e}_k\right]}{\mathbb{E}_{\pi(k)\gamma(\eta)}\left[\left(\prod_{t=1}^{T}(1 + \eta r_t^k)\right)\eta\right]}. \quad (9)$$

The iProd weights keep the iProd potential identically zero, above the Squint potential by (8), and Squint's regret bounds hence transfer to iProd. We champion Squint over the purer iProd because Squint's weights admit efficient closed form evaluation, as shown in the next section. For γ a point-mass on a fixed choice of η this advantage disappears, and iProd reduces to Modified Prod by Gaillard et al. (2014), whereas Squint becomes very similar to the OBA algorithm of Wintenberger (2014).

3. Three Choices of the Prior on Learning Rates

We will now consider different choices for the prior γ on $\eta \in [0, 1/2]$. In each case the proof of the corresponding regret bound elaborates on the argument in (6), showing that the priors place sufficient mass in a neighbourhood of the optimized learning rate $\hat{\eta}$. This might be viewed as performing a *Laplace approximation* of the integral over η, although the details vary slightly depending on the prior γ. The prior π on experts remains completely general. The proofs can be found in Appendix A.

3.1. Conjugate Prior

First we consider a conjugate prior γ with density

$$\frac{\mathrm{d}\gamma}{\mathrm{d}\eta} = \frac{e^{a\eta - b\eta^2}}{Z(a,b)} \qquad \text{where} \qquad Z(a,b) = \int_0^{1/2} e^{a\eta - b\eta^2} \mathrm{d}\eta \qquad (10)$$

for parameters $a, b \in \mathbb{R}$. The uniform prior, mentioned in the introduction, corresponds to the special case $a = b = 0$, for which $Z(a,b) = 1/2$. Abbreviating $x = a + R_T^k$ and $y = b + V_T^k$, the Squint predictions (5) then specialize to become proportional to

$$w_{T+1}^k \propto \pi(k) \int_0^{\frac{1}{2}} e^{\eta x - \eta^2 y} \eta \, \mathrm{d}\eta = \pi(k) \left(\frac{e^{\frac{x^2}{4y}} \sqrt{\pi} x \left(\mathrm{erf}\left(\frac{x}{2\sqrt{y}}\right) - \mathrm{erf}\left(\frac{x-y}{2\sqrt{y}}\right) \right)}{4y^{3/2}} + \frac{1 - e^{\frac{x}{2} - \frac{y}{4}}}{2y} \right).$$
$$(11)$$

These weights can be computed efficiently (but see Appendix B for numerically stable evaluation). For this prior, we obtain the following bound:

Theorem 2 (Conjugate Prior) *Let* $\ln_+(x) = \ln(\max\{x, 1\})$. *Then the regret of Squint* (5) *with conjugate prior* (10) *(with respect to any subset of experts $\mathcal{K}$) is bounded by*

$$R_T^{\mathcal{K}} \leq 2\sqrt{(V_T^{\mathcal{K}} + b)\left(\frac{1}{2} + \ln_+\left(\frac{Z(a,b)\sqrt{2(V_T^{\mathcal{K}} + b)}}{\pi(\mathcal{K})}\right)\right)} + 5\ln_+\left(\frac{2\sqrt{5}Z(a,b)}{\pi(\mathcal{K})}\right) - a. \quad (12)$$

The oracle tuning $a = 0$ and $b = V_T^{\mathcal{K}}$ results in $Z(a,b) \leq \frac{\sqrt{\pi}}{2\sqrt{V_T^{\mathcal{K}}}}$. Plugging this in we find that the main term in (12) becomes

$$2\sqrt{2V_T^{\mathcal{K}}\left(\frac{1}{2} + \ln_+\left(\frac{\sqrt{\pi}}{\pi(\mathcal{K})}\right)\right)},$$

which is of the form (2) that we are after, with constant overhead C_{lr} for learning the learning rate. Of course, the fact that we do not know $V_T^{\mathcal{K}}$ in advance makes this tuning impossible, and for any constant parameters a and b we get a factor of order $C_{\mathrm{lr}} = \ln V_T^{\mathcal{K}}$.

3.2. A Good Prior in Theory

The reason the conjugate prior does not achieve the optimal bound is that it does not put sufficient mass in a neighbourhood of the optimal learning rate $\hat{\eta}$ that maximizes $e^{\eta R_T^{\mathcal{K}} - \eta^2 V_T^{\mathcal{K}}}$. To see how we could address this issue, observe that we can plug $\alpha\hat{\eta}$ instead of $\hat{\eta}$ into (6) for some scaling factor

$\alpha \in (0,1)$, and still obtain the desired regret bound up to a constant factor (which depends on α). This implies that, if we could find a prior that puts a constant amount of mass on the interval $[\alpha\hat{\eta}, \hat{\eta}]$, independent of $\hat{\eta}$, then we would only pay a constant cost C_{lr} to learn the learning rate, at the price of having a slightly worse constant factor.

A prior that puts constant mass on any interval $[\alpha\hat{\eta}, \hat{\eta}]$ should have a distribution function of the form $a\ln(\eta) + b$ for some constants a and b, and hence its density should be proportional to $1/\eta$. But here we run into a problem, because, unfortunately, $1/\eta$ does not have a finite integral over $\eta \in [0, 1/2]$ and hence no such prior exists!

The solution we adopt in this section will be to adjust the density $1/\eta$ just a tiny amount so that it does integrate. Let γ have density

$$\frac{\mathrm{d}\gamma}{\mathrm{d}\eta} = \frac{\ln 2}{\eta \ln^2(\eta)}, \tag{13}$$

where $\ln^2(x) = \big(\ln(x)\big)^2$. We call this the *CV prior*, because it has previously been used to get quantile bounds by Chernov and Vovk (2010). The additional factor $1/\ln^2(\eta)$ in the prior only leads to an extra factor of $\sqrt{\ln \ln V_T^{\mathcal{K}}}$ in the bound, which we consider to be essentially a constant.

Although the motivation above suggests that we might obtain a suboptimal constant factor (depending on α), a more careful analysis shows that this does not even happen: apart from the effect of the $1/\ln^2(\eta)$ term in prior, we obtain the optimal multiplicative constant.

Theorem 3 (CV Prior) *Let* $\ln_+(x) = \ln(\max\{x, 1\})$. *Then the regret of Squint* (5) *with CV prior* (13) *(with respect to any subset of experts $\mathcal{K}$) is bounded by*

$$R_T^{\mathcal{K}} \leq \sqrt{2V_T^{\mathcal{K}} \left(1 + \sqrt{2\ln_+\left(\frac{\ln_+^2\left(\frac{2\sqrt{V_T^{\mathcal{K}}}}{2-\sqrt{2}}\right)}{\pi(\mathcal{K})\ln(2)}\right)}\right)} - 5\ln\pi(\mathcal{K}) + 4. \tag{14}$$

3.3. Improper Prior

In the last section we argued that we needed a density proportional to $1/\eta$ on $\eta \in [0, 1/2]$. Such a density would not integrate, and we studied the CV prior density instead. However, we could be bold and see what breaks if we use the improper $1/\eta$ density anyway. We should be highly suspicious though, because this density is *improper of the worst kind*: the integral $\int_0^{1/2} e^{\eta R_T - \eta^2 V_T} \frac{1}{\eta} \,\mathrm{d}\eta$ diverges no matter how many rounds of data we process (a Bayesian would say: "the posterior remains improper"). Yet it turns out that we hit no essential impossibilities: the improper prior $1/\eta$ cancels with the η present in the Squint rule (5), and the predictions are always well-defined. As we will see, we still get desirable regret bounds, but, equally important, we regain a closed-form integral for our weight computation. The Squint prediction (5) specializes to

$$w_{T+1}^k \propto \pi(k)\int_0^{1/2} e^{\eta R_T^k - \eta^2 V_T^k} \,\mathrm{d}\eta = \pi(k)\frac{\sqrt{\pi}e^{\frac{(R_T^k)^2}{4V_T^k}}\left(\mathrm{erf}\left(\frac{R_T^k}{2\sqrt{V_T^k}}\right) - \mathrm{erf}\left(\frac{R_T^k - V_T^k}{2\sqrt{V_T^k}}\right)\right)}{2\sqrt{V_T^k}}. \tag{15}$$

(We look at numerical stability in Appendix B.) This strategy provides the following guarantee:

Theorem 4 (Improper Prior) *The regret of Squint with improper prior* (15) *(with respect to any subset of experts $\mathcal{K}$) is bounded by*

$$R_T^{\mathcal{K}} \le \sqrt{2V_T^{\mathcal{K}}} \left(1 + \sqrt{2\ln\left(\frac{\frac{1}{2} + \ln(T+1)}{\pi(\mathcal{K})}\right)} \right) + 5\ln\left(1 + \frac{1 + 2\ln(T+1)}{\pi(\mathcal{K})}\right). \quad (16)$$

4. *Component iProd*: a Second-order Quantile Method for Combinatorial Games

In the combinatorial setting the elementary actions are combinatorial concepts from some class $\mathcal{C} \subseteq \{0,1\}^K$. The combinatorial structure is reflected in the loss, which decomposes into a sum of coordinate losses. That is, the loss of concept $\boldsymbol{c} \in \mathcal{C}$ is $\boldsymbol{c}^\mathsf{T}\boldsymbol{\ell}$ for some loss vector $\boldsymbol{\ell} \in [0,1]^K$. For example, the loss of a path is the total loss of its edges. We allow the learner to play a distribution p on $\mathcal{C}$ and incur the expected loss $\mathbb{E}_{p(\boldsymbol{c})}[\boldsymbol{c}^\mathsf{T}\boldsymbol{\ell}] = \mathbb{E}_{p(\boldsymbol{c})}[\boldsymbol{c}]^\mathsf{T}\boldsymbol{\ell}$. This means that the loss of p is determined by its mean, which is called the *usage* of p. We can therefore simplify the setup by having the learner play a usage vector $\boldsymbol{u} \in \mathcal{U}$, where $\mathcal{U} = \mathrm{conv}(\mathcal{C}) \subseteq [0,1]^K$ is the polytope of valid usages. The loss then becomes $\boldsymbol{u}^\mathsf{T}\boldsymbol{\ell}$.

Koolen et al. (2010) point out that the Hedge algorithm with the concepts as experts guarantees

$$R_T^{\boldsymbol{c}} \prec K\sqrt{T\mathrm{comp}(\mathcal{C})}, \qquad \text{Expanded Hedge} \quad (17)$$

upon proper tuning, where $\mathrm{comp}(\mathcal{C})$ is some appropriate notion of the complexity of the combinatorial class $\mathcal{C}$. (This is exactly (1), where the additional factor K comes from the fact that the loss of a single concept now ranges over $[0, K]$ instead of $[0, 1]$.) The computationally efficient Follow the Perturbed Leader strategy has the same bound. However, Koolen et al. (2010) show that this bound has a fundamentally suboptimal dependence on the loss range, which they call the *range factor problem*. Properly tuned, their Component Hedge algorithm (a particular instance of Mirror Descent) keeps the regret below

$$R_T^{\boldsymbol{c}} \prec \sqrt{TK\mathrm{comp}(\mathcal{C})}, \qquad \text{Component Hedge} \quad (18)$$

the improvement being due to the algorithm exploiting the sum structure of the loss. To show that this cannot be improved further, Koolen et al. (2010) exhibit matching lower bounds for a variety of combinatorial domains. Audibert et al. (2014) give an example where the upper bound (17) for Expanded Hedge is tight, so the range factor problem cannot be solved by a better analysis.

In this section we aim to develop efficient algorithms for combinatorial prediction that obtain the second-order and quantile improvements of (18), but do not suffer from the range factor problem.

It is instructive to see that our bounds (2) for Squint/iProd, when applied with a separate expert for each concept, indeed also suffer from a suboptimal loss range dependence. We find

$$R_T^{\boldsymbol{c}} \prec \sqrt{V_T^{\boldsymbol{c}}(\mathrm{comp}(\mathcal{C}) + \text{tuning cost})}, \qquad \text{Expanded Squint/iProd}$$

where $V_T^{\boldsymbol{c}} = \sum_{t=1}^T (r_T^{\boldsymbol{c}})^2 = \sum_{t=1}^T (\sum_{k=1}^K r_t^k)^2$ with $r_t^k \in [-1, 1]$ may now be as large as K^2T, whereas we know KT suffices. The reason for this is that $V_T^{\boldsymbol{c}}$ measures the variance of the concept as a whole, whereas the sum structure of the loss makes it possible to replace $V_T^{\boldsymbol{c}}$ by the sum of the variances of the components. In the analysis, this problem shows up when we apply the bound (8). To fix it, we must therefore rearrange the algorithm to be able to apply (8) once per component.

8

Algorithm 1: Component iProd. Required subroutines are the relative entropy projection step (3) and the decomposition step (5). For polytopes $\mathcal{U}$ that can be represented by few linear inequalities these can be deferred to general-purpose convex and linear optimizers. See Koolen et al. (2010) for more details, and for ideas regarding more efficient implementations for particular concept classes.

Input: Combinatorial class $\mathcal{C} \subseteq \{0,1\}^K$ with convex hull $\mathcal{U} = \text{conv}(\mathcal{C})$
Input: Prior distribution γ on a discrete grid $\mathcal{G} \subset [0, 1/2]$ and prior vector $\boldsymbol{\pi} \in [0,1]^K$

1: For each $\eta \in \mathcal{G}$, initialize $\tilde{\boldsymbol{u}}_1^\eta = \boldsymbol{\pi}$ and $L_1^\eta = -\ln(\gamma(\eta)\eta)$ ▷ (21)

2: **for** $t = 1, 2, \ldots$ **do**

3:      For each $\eta \in \mathcal{G}$, project $\boldsymbol{u}_t^\eta = \min_{\boldsymbol{u} \in \mathcal{U}} \triangle_2(\boldsymbol{u} \| \tilde{\boldsymbol{u}}_t)$ ▷ (21a)

4:      Compute usage $\boldsymbol{u}_t = \sum_\eta e^{-L_t^\eta} \boldsymbol{u}_t^\eta \Big/ \sum_\eta e^{-L_t^\eta}$ ▷ (23)

5:      Decompose $\boldsymbol{u}_t = \sum_i p_i \boldsymbol{c}_{t,i}$ into a convex combination of concepts $\boldsymbol{c}_{t,i} \in \mathcal{C}$

6:      Play $\boldsymbol{c}_{t,i}$ with probability p_i

7:      Receive loss vector $\boldsymbol{\ell}_t$, incur expected loss $\boldsymbol{u}_t^\mathsf{T} \boldsymbol{\ell}_t$

8:      For each $\eta \in \mathcal{G}$ and k, update $\tilde{u}_{t+1}^{\eta,k} = u_t^{\eta,k} \frac{1+\eta(u_t^k-1)\ell_t^k}{1+\eta(u_t^k-u_t^{\eta,k})\ell_t^k}$ ▷ (21b),(33),(24),(22)

9:      For each $\eta \in \mathcal{G}$, update $L_{t+1}^\eta = L_t^\eta - \sum_{k=1}^K \frac{1}{K} \ln(1 + \eta(u_t^k - u_t^{\eta,k})\ell_t^k)$ ▷ (20),(24),(22)

10: **end for**

Outlook Our approach will be based on a new potential function that aggregates over learning rates η explicitly and over the concept class $\mathcal{C}$ implicitly. Our inspiration for the latter comes from rewriting the factor featuring inside the $\mathbb{E}_{\gamma(\eta)}$ expectation in the iProd potential (9) as

$$\mathbb{E}_{\pi(k)} \left[\prod_{t=1}^T (1+\eta r_t^k)\right] = \prod_{t=1}^T \frac{\mathbb{E}_{\pi(k)}\left[\prod_{s=1}^t (1+\eta r_s^k)\right]}{\mathbb{E}_{\pi(k)}\left[\prod_{s=1}^{t-1}(1+\eta r_s^k)\right]} = e^{-\sum_{t=1}^T \ell_{\text{mix}}(p_t^\eta, \boldsymbol{x}_t^\eta)}, \qquad (19)$$

which we interpret as the *mix loss* (see De Rooij et al. 2014) of the *exponential weights distribution* p_t^η on *auxiliary losses*:

$$\ell_{\text{mix}}(p, \boldsymbol{x}) = -\ln \mathbb{E}_{p(k)}\left[e^{-x^k}\right], \quad p_t^\eta(k) = \frac{\pi(k)e^{-\sum_{s=1}^{t-1} x_s^{\eta,k}}}{\mathbb{E}_{\pi(k)}\left[e^{-\sum_{s=1}^{t-1} x_s^{\eta,k}}\right]}, \quad x_t^{\eta,k} = -\ln(1+\eta r_t^k).$$

Thus, for each fixed η, we have identified a sub-module in which the loss is the mix loss. It turns out that the Squint/iProd regret bounds can be reinterpreted as arising from (quantile) mix loss regret bounds for exponential weights in this sub-module. For combinatorial games, we hence need to upgrade exponential weights to a combinatorial algorithm for mix loss. No such algorithm was readily available, so we derive a new algorithm that we call Component Bayes (a variant of Component Hedge) in Section 4.1, and prove a quantile mix loss regret bound for it. Then in Section 4.2 we show that Component iProd, obtained by substituting Component Bayes for exponential weights in the sub-module above, inherits all of iProd's desirable features. That is, by aggregating the above sub-module over learning rates the Component iProd predictor delivers low second-order quantile regret. Component iProd is summarized as Algorithm 1. Proofs can be found in Appendix A.

4.1. Component Bayes

In this section we describe a combinatorial algorithm for mix loss, which will then be an essential subroutine in our Component iProd algorithm. We take as our action space some closed convex $\mathcal{U} \subseteq [0,1]^K$. The game then proceeds in rounds. Each round t the learner plays $\boldsymbol{u}_t \in \mathcal{U}$, which we interpret as making K independent plays in K parallel two-expert sub-games, putting weight u_t^k and $1 - u_t^k$ on experts 1 and 0 in sub-game k. The environment reveals a loss vector $\boldsymbol{x}_t \in \mathbb{R}^{K \times \{0,1\}}$ (we use $\boldsymbol{x}$ for the loss in this auxiliary game, and reserve $\boldsymbol{\ell}$ for the loss in the main game), and the loss of the learner is the sum of per-coordinate mix losses:

$$\ell_{\mathrm{mix}}(\boldsymbol{u}, \boldsymbol{x}) \;=\; \sum_{k=1}^{K} -\ln\left(u^k e^{-x^{k,1}} + (1 - u^k)e^{-x^{k,0}}\right). \tag{20}$$

The goal is to compete with the best element $\boldsymbol{v} \in \mathcal{U}$. We define *Component Bayes*[2] inductively as follows. We set $\tilde{\boldsymbol{u}}_1 = \boldsymbol{\pi} \in [0,1]^K$ to some prior vector of our choice (which does not have to be a usage in $\mathcal{U}$), and then alternate

$$\boldsymbol{u}_t \;=\; \underset{\boldsymbol{u} \in \mathcal{U}}{\arg\min} \, \triangle_2(\boldsymbol{u} \| \tilde{\boldsymbol{u}}_t) \tag{21a}$$

$$\tilde{\boldsymbol{u}}_{t+1} \;=\; \underset{\boldsymbol{u} \in [0,1]^K}{\arg\min} \, \triangle_2(\boldsymbol{u} \| \boldsymbol{u}_t) + \sum_{k=1}^{K}(u^k x_t^{k,1} + (1 - u^k)x_t^{k,0}), \tag{21b}$$

where $\triangle_2$ denotes the binary relative entropy, defined from scalars x to y and vectors $\boldsymbol{v}$ to $\boldsymbol{u}$ by $\triangle_2(x \| y) = x \ln \frac{x}{y} + (1 - x) \ln \frac{1-x}{1-y}$ and $\triangle_2(\boldsymbol{v} \| \boldsymbol{u}) = \sum_{k=1}^{K} \triangle_2(v^k \| u^k)$. This simple scheme is all it takes to adapt to the combinatorial domain.

Lemma 5 *Fix any closed convex $\mathcal{U} \subseteq [0,1]^K$. For any loss sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$ in $\mathbb{R}^{K \times \{0,1\}}$, the mix loss regret of Component Bayes* (21) *with prior $\boldsymbol{\pi} \in [0,1]^K$ compared to any $\boldsymbol{v} \in \mathcal{U}$ is at most*

$$\sum_{t=1}^{T} \ell_{mix}(\boldsymbol{u}_t, \boldsymbol{x}_t) - \sum_{t=1}^{T}\sum_{k=1}^{K}\left(v^k x_t^{k,1} + (1 - v^k)x_t^{k,0}\right) \;\leq\; \triangle_2(\boldsymbol{v} \| \boldsymbol{\pi}).$$

The practicality of Component Bayes does depend on the computational cost of computing the binary relative entropy projection onto the convex set $\mathcal{U}$. Fortunately, in many applications $\mathcal{U}$ has a compact representation by means of a few linear inequalities; e.g. the Birkhoff polytope (permutations) and the Flow polytope (paths). See (Koolen et al., 2010) for examples. Component Bayes may then be implemented using off-the-shelf convex optimization subroutines like CVX.

4.2. Component iProd

We now return to our original problem of combinatorial prediction with linear loss. Using Component Bayes (which is for mix loss) as a sub-module, we construct an algorithm with second-order quantile bounds. We first have to extend our notion of regret vector . Suppose the learner predicts

---

2. Ignoring a small technically convenient switch from generalized to binary relative entropy we find that Component Bayes *equals* Component Hedge of Koolen et al. (2010). The new name stresses an important distinction in the game protocol: Component Hedge guarantees low *linear loss* regret, Component Bayes guarantees low *mix loss* regret.

usage $\boldsymbol{u}_t \in \mathcal{U} \subseteq [0,1]^K$ and encounters loss vector $\boldsymbol{\ell}_t \in [-1,+1]^K$. We then define the regret vector $\boldsymbol{r}_t \in [-1,+1]^{K \times \{0,1\}}$ by

$$r_t^{k,1} = u_t^k \ell_t^k - \ell_t^k \qquad \text{and} \qquad r_t^{k,0} = u_t^k \ell_t^k. \tag{22}$$

Fix a prior vector $\boldsymbol{\pi} \in [0,1]^K$ and prior distribution γ on $[0,1/2]$. We define the *Component iProd* potential function and predictor by

$$\Phi(\boldsymbol{r}_{1:T}) = \mathop{\mathbb{E}}_{\gamma(\eta)} \left[e^{-\frac{1}{K} \sum_{t=1}^{T} \ell_{\mathrm{mix}}(\boldsymbol{u}_t^\eta, \boldsymbol{x}_t^\eta)} - 1 \right], \quad \boldsymbol{u}_T = \frac{\mathbb{E}_{\gamma(\eta)} \left[e^{-\frac{1}{K} \sum_{t=1}^{T-1} \ell_{\mathrm{mix}}(\boldsymbol{u}_t^\eta, \boldsymbol{x}_t^\eta)} \eta \boldsymbol{u}_T^\eta \right]}{\mathbb{E}_{\gamma(\eta)} \left[e^{-\frac{1}{K} \sum_{t=1}^{T-1} \ell_{\mathrm{mix}}(\boldsymbol{u}_t^\eta, \boldsymbol{x}_t^\eta)} \eta \right]}, \tag{23}$$

where $\boldsymbol{u}_1^\eta, \boldsymbol{u}_2^\eta, \ldots$ denote the usages of Component Bayes with prior $\boldsymbol{\pi}$ on losses $\boldsymbol{x}_1^\eta, \boldsymbol{x}_2^\eta, \ldots$ set to

$$x_t^{\eta,k,b} = -\ln\left(1 + \eta r_t^{k,b}\right) \tag{24}$$

Note that $\boldsymbol{u}_T \in \mathcal{U}$ is a bona fide action, as it is a convex combination of $\boldsymbol{u}_T^\eta \in \mathcal{U}$. As can be seen from (19), this potential generalizes iProd (9): in the base case $K = 1$ and $\mathcal{C} = \{0,1\}$ Component iProd reduces to iProd (9) on $K = 2$ experts if we set the loss for Component iProd to the difference of the losses for iProd. We will now show that Component iProd has the desired regret guarantee.

Lemma 6 *Fix any closed convex $\mathcal{U} \subseteq [0,1]^K$. Component iProd (23) ensures that for any loss sequence $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_T$ in $[-1,+1]^K$ we have $\Phi(\boldsymbol{r}_{1:T}) \leq \ldots \leq \Phi(\emptyset) = 0$.*

We now establish that non-positive potential implies our desired regret bound. We express our quantile bound in terms of the $\boldsymbol{v}$-weighted cumulative coordinate-wise regret and uncentered variance

$$\begin{aligned}
R_T^{\boldsymbol{v}} &= \sum_{t=1}^{T} \sum_{k=1}^{K} \left(v^k r_t^{k,1} + (1 - v^k) r_t^{k,0} \right) \overset{(22)}{=} \sum_{t=1}^{T} (\boldsymbol{u}_t - \boldsymbol{v})^\mathsf{T} \boldsymbol{\ell}_t, \\
V_T^{\boldsymbol{v}} &= \sum_{t=1}^{T} \sum_{k=1}^{K} \left(v^k (r_t^{k,1})^2 + (1 - v^k)(r_t^{k,0})^2 \right).
\end{aligned} \tag{25}$$

Lemma 7 *Suppose γ is supported on a discrete grid $\mathcal{G} \subset [0,1/2]$. Component iProd (23) guarantees that for every $\eta \in \mathcal{G}$ and for every comparator $\boldsymbol{v} \in \mathcal{U}$ the regret is at most*

$$\eta R_T^{\boldsymbol{v}} - \eta^2 V_T^{\boldsymbol{v}} \leq \triangle_2(\boldsymbol{v} \| \boldsymbol{\pi}) - K \ln \gamma(\eta).$$

We now discuss the choice of the discrete prior γ on η. Here we face a trade-off between regret and computation. More discretization points reduce the regret overhead for mis-tuning, but since we need to run one instance of Component Bayes per grid point the computation time also grows linearly in the number of grid points. Fortunately, Lemma 7 implies that exponential spacing suffices, as missing the optimal tuning $\hat{\eta} = \frac{R_T^{\boldsymbol{v}}}{2V_T^{\boldsymbol{v}}}$ by a constant factor affects the regret bound by another constant factor. To see this, apply Lemma 7 to $\eta = \alpha \hat{\eta}$. We find

$$R_T^{\boldsymbol{v}} \leq \frac{2}{\sqrt{\alpha(2-\alpha)}} \sqrt{V_T^{\boldsymbol{v}} \left(\triangle_2(\boldsymbol{v} \| \boldsymbol{\pi}) - K \ln \gamma(\alpha \hat{\eta}) \right)}. \tag{26}$$

It is therefore sufficient to choose η from an exponentially spaced grid:

Theorem 8 *Let $\mathcal{U} \subseteq [0,1]^K$ be closed and convex. Component iProd (23), with γ the uniform prior on grid $\mathcal{G} = \{2^{-i} \mid i = 1, \ldots, \lceil 1 + \log_2 T \rceil\}$ and arbitrary $\boldsymbol{\pi} \in [0,1]^K$, ensures that, for any sequence $\boldsymbol{\ell}_1 \ldots, \boldsymbol{\ell}_T$ of $[-1, +1]$-valued loss vectors, the regret compared to any $\boldsymbol{v} \in \mathcal{U}$ is at most*

$$R_T^{\boldsymbol{v}} \leq \frac{4}{\sqrt{3}} \sqrt{V_T^{\boldsymbol{v}} \big(\triangle_2(\boldsymbol{v} \| \boldsymbol{\pi}) + K \ln \lceil 1 + \log_2 T \rceil \big)} + 4 \triangle_2(\boldsymbol{v} \| \boldsymbol{\pi}) + K \max\{4 \ln \lceil 1 + \log_2 T \rceil, 1\}. \tag{27}$$

Discussion of Component iProd We showed that if we have an algorithm for keeping the *mix loss* regret small compared to some concept class, we can run multiple instances, each with a different learning rate factored into the losses, and as a result also keep the linear loss small with second order quantile bounds. Another setting where this could be applied is to switching experts. The Fixed Share algorithm by Herbster and Warmuth (1998) applies to all Vovk mixable losses, so in particular to the mix loss, and delivers adaptive regret bounds (Adamskiy et al., 2012). Aggregating over $\log_2 T$ exponentially spaced η to learn the *learning rate* would indeed be very cheap. Yet another setting is matrix-valued prediction under linear loss (Tsuda et al., 2005), where our method would transport the mix loss bounds of Warmuth and Kuzmin (2010) to second-order quantile bounds.

In Lemma 7 we see that the cost $-\ln \gamma(\eta)$ for learning the learning rate η occurs multiplied by the ambient dimension K. Intuitively this seems wasteful, as we are not trying to learn a separate rate for each component. But we could not reduce K to 1. For example, defining the potential (23) *without* the division by K escalates its dependency on the loss $\boldsymbol{\ell}$ from linear to polynomial of order K. Unfortunately this potential cannot be kept below zero even for $K = 2$.

5. Conclusion and Future Work

We have constructed second-order quantile methods for both the expert setting (Squint) and for general combinatorial games (Component iProd). The key in both cases is the ability to learn the appropriate learning rate, which is reflected by the integrals over η in our potential functions (5) and (23). As discussed under related work, there is a whole variety of different ways to adapt to the optimal η. This raises the question of whether there is a unifying perspective that explains when and how it is possible to learn the learning rate in general.

Another issue for future work is to find matching lower bounds. Although lower bounds in terms of $\sqrt{T \ln K}$ are available for the worst possible sequence (Cesa-Bianchi and Lugosi, 2006), the issue is substantially more complex when considering either variances or quantiles. We are not aware of any lower bounds in terms of the variance V_T^k. Gofer and Mansour (2012) provide lower bounds that hold for *any* sequence, in terms of the squared loss ranges in each round, but these do not apply to methods that adaptively tune their learning rate. For quantile bounds, Koolen (2013) takes a first step by characterizing the Pareto optimal quantile bounds for 2 experts in the $\sqrt{T}$ regime.

Finally, we have assumed throughout that all losses are normalized to the range $[0,1]$. But there exist second-order methods that do not require this normalization and can adapt automatically to the loss range (Cesa-Bianchi et al., 2007; De Rooij et al., 2014; Wintenberger, 2014). It is an open question how such adaptive techniques can be incorporated elegantly into our methods.

Acknowledgments

References

Jacob Abernethy, Chansoo Lee, Abhinav Sinha, and Ambuj Tewari. Online linear optimization via smoothing. In *Proceedings of The 27th Conference on Learning Theory, COLT*, volume 35 of *JMLR Proceedings*, pages 807–823, 2014.

Dmitry Adamskiy, Wouter M. Koolen, Alexey Chernov, and Vladimir Vovk. A closer look at adaptive regret. In *Proceedings of the 23rd International Conference on Algorithmic Learning Theory (ALT)*, LNAI 7568, pages 290–304. Springer, 2012.

Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. *Math. Oper. Res.*, 39(1):31–45, 2014.

Peter Auer, Nicolò Cesa-Bianchi, and Claudio Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64:48–75, 2002.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

Nicolò Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *J. Comput. Syst. Sci.*, 78(5):1404–1422, 2012.

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.

Kamalika Chaudhuri, Yoav Freund, and Daniel Hsu. A parameter-free hedging algorithm. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, pages 297–305, 2009.

Alexey V. Chernov and Vladimir Vovk. Prediction with advice of unknown number of experts. In *Uncertainty in Artificial Intelligence*, pages 117–125, 2010.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Proceedings of the 25th Annual Conference on Learning Theory*, number 23 in JMLR W&CP, pages 6.1–6.20, 2012.

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*, 55:119–139, 1997.

Pierre Gaillard, Gilles Stoltz, and Tim van Erven. A second-order bound with excess losses. In *JMLR Workshop and Conference Proceedings*, volume 35: Proceedings of the 27th Conference on Learning Theory (COLT), pages 176–196, 2014.

Eyal Gofer and Yishay Mansour. Lower bounds on individual sequence regret. In *Algorithmic Learning Theory (ALT)*, volume 7568, pages 275–289. Springer Berlin Heidelberg, 2012.

Peter D. Grünwald. Safe learning: bridging the gap between Bayes, MDL and statistical learning theory via empirical convexity. In *JMLR W&CP 19: Proceedings of the 24th Annual Conference on Learning Theory, COLT*, pages 397–420, 2011.

Peter D. Grünwald. The safe Bayesian. In *Algorithmic Learning Theory (ALT)*, volume 7568, pages 169–183. Springer Berlin Heidelberg, 2012.

Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. *Machine learning*, 80(2-3):165–188, 2010.

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

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

Marcus Hutter and Jan Poland. Adaptive online prediction by following the perturbed leader. *Journal of Machine Learning Research*, 6:639–660, 2005.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Wouter M. Koolen. The Pareto regret frontier. In *Advances in Neural Information Processing Systems (NIPS) 26*, pages 863–871, 2013.

Wouter M. Koolen, Manfred K. Warmuth, and Jyrki Kivinen. Hedging structured concepts. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*, pages 93–105, 2010.

Wouter M. Koolen, Tim van Erven, and Peter D. Grünwald. Learning the learning rate for prediction with expert advice. In *Advances in Neural Information Processing Systems (NIPS) 27*, pages 2294–2302, 2014.

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

Haipeng Luo and Robert E Schapire. A drifting-games analysis for online learning and applications to boosting. In *Advances in Neural Information Processing Systems 27*, pages 1368–1376. Curran Associates, Inc., 2014.

Haipeng Luo and Robert E. Schapire. Achieving all with no parameters: Adaptive NormalHedge. *CoRR*, abs/1502.05934, 2015. To appear at COLT.

Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. Localization and adaptation in online learning. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, AISTATS*, volume 31 of *JMLR Proceedings*, pages 516–526. JMLR.org, 2013.

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.

Amir Sani, Gergely Neu, and Alessandro Lazaric. Exploiting easy data in online optimization. In *Advances in Neural Information Processing Systems 27*, pages 810–818. Curran Associates, Inc., 2014.

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

Jacob Steinhardt and Percy Liang. Adaptivity and optimism: An improved exponentiated gradient algorithm. In *International Conference on Machine Learning (ICML)*, 2014.

Eiji Takimoto and Manfred K. Warmuth. Path kernels and multiplicative updates. *Journal of Machine Learning Research*, 4:773–818, 2003.

Koji Tsuda, Gunnar Rätsch, and Manfred K. Warmuth. Matrix Exponentiated Gradient updates for on-line learning and Bregman projections. *Journal of Machine Learning Research*, 6:995–1018, 2005.

Vladimir Vovk. Aggregating strategies. In *COLT Proceedings*, pages 371–383, 1990.

Vladimir Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153–173, 1998.

Manfred K. Warmuth and Dima Kuzmin. Randomized online PCA algorithms with regret bounds that are logarithmic in the dimension. *Journal of Machine Learning Research*, 9:2287–2320, 2008.

Manfred K. Warmuth and Dima Kuzmin. Bayesian generalized probability calculus for density matrices. *Machine Learning*, 78(1-2):63–101, 2010.

Manfred K. Warmuth, Wouter M. Koolen, and David P. Helmbold. Combining initial segments of lists. *Theoretical Computer Science*, 519:29–45, 2014. The special issue on Algorithmic Learning Theory for ALT 2011.

Olivier Wintenberger. Optimal learning with Bernstein online aggregation. Preprint, available from http://arxiv.org/abs/1404.1356v2, 2014.

Appendix A. Proofs

This section collects the proofs omitted from Sections 3 and 4.

A.1. Theorem 2

Proof Abbreviate $R = R_T^{\mathcal{K}} + a$ and $V = V_T^{\mathcal{K}} + b$. Then from (7) and Jensen's inequality we obtain

$$1 \geq \mathbb{E}_{\pi(k)\gamma(\eta)}\left[e^{\eta R_T^k - \eta^2 V_T^k}\right] \geq \frac{\pi(\mathcal{K})\,\mathbb{E}_{\pi(k|\mathcal{K})} \int_0^{1/2} e^{\eta(R_T^k + a) - \eta^2(V_T^k + b)}\,\mathrm{d}\eta}{Z(a,b)} \geq \frac{\pi(\mathcal{K}) \int_0^{1/2} e^{\eta R - \eta^2 V}\,\mathrm{d}\eta}{Z(a,b)}.$$

The η that maximizes $\eta R - \eta^2 V$ is $\hat{\eta} = \frac{R}{2V}$. Without loss of generality, we can assume that $\hat{\eta} \geq \frac{1}{\sqrt{2V}} \geq 0$, because otherwise $R \leq \sqrt{2V}$, from which (12) follows directly. Now let $[u,v] \subseteq [0,\frac{1}{2}]$ be any interval such that $v \leq \hat{\eta}$. Then, because $\eta R - \eta^2 V$ is non-decreasing in η for $\eta \leq \hat{\eta}$, we have

$$\int_0^{1/2} e^{\eta R - \eta^2 V}\,\mathrm{d}\eta \geq \int_u^v e^{\eta R - \eta^2 V}\,\mathrm{d}\eta \geq (v-u)e^{uR - u^2 V},$$

so that the above two equations imply

$$uR - u^2 V \leq \ln\left(\frac{Z(a,b)}{\pi(\mathcal{K})(v-u)}\right). \tag{28}$$

Suppose first that $\hat{\eta} \leq 1/2$. Then we take $v = \hat{\eta}$ and $u = \hat{\eta} - \frac{1}{\sqrt{2V}}$. Plugging these into (28) we obtain

$$R \leq 2\sqrt{V\left(\frac{1}{2} + \ln\left(\frac{Z(a,b)\sqrt{2V}}{\pi(\mathcal{K})}\right)\right)},$$

which implies (12). Alternatively, we may have $\hat{\eta} > 1/2$, which is equivalent to $R > V$. Then the left-hand side of (28) is at least $u(1-u)R$ and hence we obtain

$$R \leq \frac{1}{(1-u)u}\ln\left(\frac{Z(a,b)}{\pi(\mathcal{K})(v-u)}\right).$$

Taking $u = \frac{5-\sqrt{5}}{10}$ and $v = 1/2$ then leads to the bound

$$R \leq 5\ln\left(\frac{2\sqrt{5}Z(a,b)}{\pi(\mathcal{K})}\right),$$

which again implies (12). ∎

A.2. Theorem 3

Proof Abbreviate $R = R_T^{\mathcal{K}}$ and $V = V_T^{\mathcal{K}}$, and let $\hat{\eta} = \frac{R}{2V}$ be the η that maximizes $\eta R - \eta^2 V$. Then $\eta R - \eta^2 V$ is non-decreasing in η for $\eta \leq \hat{\eta}$ and hence, for any interval $[u,v] \subseteq [0,1/2]$ such that $v \leq \hat{\eta}$, we obtain from (7) and Jensen's inequality that

$$1 \geq \pi(\mathcal{K}) \mathbb{E}_{\pi(k|\mathcal{K})\gamma(\eta)}\left[e^{\eta R_T^k - \eta^2 V_T^k}\right] \geq \pi(\mathcal{K}) \mathbb{E}_{\gamma(\eta)}\left[e^{\eta R - \eta^2 V}\right] \geq \pi(\mathcal{K})\gamma([u,v])e^{uR - u^2 V}, \tag{29}$$

where

$$\gamma([u, v]) = \int_u^v \frac{\ln 2}{\eta (\ln \eta)^2} \, d\eta = \frac{\ln(2)}{\ln(\frac{1}{v})} - \frac{\ln(2)}{\ln(\frac{1}{u})} \geq \frac{\ln(2) \ln(\frac{v}{u})}{\ln^2(\frac{1}{u})}. \tag{30}$$

If $R \leq 2\sqrt{V}$, then (14) follows by considering the cases $V \leq 4$ and $V \geq 4$, so suppose that $R \geq 2\sqrt{V}$, which implies that $\hat{\eta} \geq \frac{1}{\sqrt{2V}}$.

Now suppose first that $\hat{\eta} \leq 1/2$. Then we take $v = \hat{\eta}$ and $u = \hat{\eta} - \frac{1}{\sqrt{2V}} \geq 0$, for which

$$e^{uR - u^2 V} \frac{\ln(\frac{v}{u})}{\ln^2(\frac{1}{u})} = \frac{e^{\frac{R^2}{4V} - \frac{1}{2}} \ln \left(\frac{1}{1 - \frac{\sqrt{2V}}{R}} \right)}{\ln^2 (\frac{2V}{R - \sqrt{2V}})} \geq \frac{e^{\frac{R^2}{4V} - \frac{1}{2}} \ln \left(\frac{1}{1 - \frac{\sqrt{2V}}{R}} \right)}{\ln^2 (\frac{2\sqrt{V}}{2 - \sqrt{2}})},$$

where the inequality follows from $R \geq 2\sqrt{V}$. By $e^{\frac{1}{2}(x^2 - 1)} = e^{\frac{1}{2}(x-1)^2} e^{x-1} \geq e^{\frac{1}{2}(x-1)^2} x$ and $\ln \frac{1}{1-x} \geq x$, we can lower bound the numerator with

$$e^{\frac{R^2}{4V} - \frac{1}{2}} \ln \left(\frac{1}{1 - \frac{\sqrt{2V}}{R}} \right) \geq e^{\frac{1}{2}(\frac{R}{\sqrt{2V}} - 1)^2} \frac{R}{\sqrt{2V}} \frac{\sqrt{2V}}{R} = e^{\frac{1}{2}(\frac{R}{\sqrt{2V}} - 1)^2}.$$

Putting everything together, we obtain

$$1 \geq \frac{\pi(\mathcal{K}) \ln(2) e^{\frac{1}{2}(\frac{R}{\sqrt{2V}} - 1)^2}}{\ln^2 \left(\frac{2\sqrt{V}}{2 - \sqrt{2}} \right)},$$

which implies

$$R \leq \sqrt{2V} \left(1 + \sqrt{2 \ln \left(\frac{\ln^2 \left(\frac{2\sqrt{V}}{2 - \sqrt{2}} \right)}{\pi(\mathcal{K}) \ln(2)} \right)} \right),$$

and (14) is satisfied.

It remains to consider the case $\hat{\eta} > 1/2$, which implies $R > V$. Then we take $v = 1/2$, and (29) leads to

$$uR - u^2 V \leq -\ln \pi(\mathcal{K}) - \ln \left(1 - \frac{\ln(2)}{\ln(\frac{1}{u})} \right).$$

Using $R > V$, the left-hand side is at most $u(1 - u)R$. The choice $u = \frac{5 - \sqrt{5}}{10}$ then again implies (14), which completes the proof. ∎

A.3. Theorem 4

Proof The proof of Lemma 1 goes through unchanged for the improper prior, but we have to be careful, because we cannot pull out the constant 1 from the integral over η in the potential function any more. So abbreviate $R = R_T^{\mathcal{K}}$ and $V = V_T^{\mathcal{K}}$. Then, by (7), $R \geq -T$, $V \leq T$, and Jensen's

inequality,

$$
\begin{aligned}
0 \;\geq\; \Phi(\boldsymbol{r}_{1:T}) \;&=\; \underset{\pi(k)}{\mathbb{E}}\left[\int_0^{1/2} \frac{e^{\eta R_T^k - \eta^2 V_T^k} - 1}{\eta}\, \mathrm{d}\eta\right] \\
&\geq\; \pi(\mathcal{K}) \underset{\pi(k|\mathcal{K})}{\mathbb{E}}\left[\int_0^{1/2} \frac{e^{\eta R_T^k - \eta^2 V_T^k} - 1}{\eta}\, \mathrm{d}\eta\right] + (1 - \pi(\mathcal{K})) \int_0^{1/2} \frac{e^{-\eta T - \eta^2 T} - 1}{\eta}\, \mathrm{d}\eta \\
&\geq\; \pi(\mathcal{K}) \int_0^{1/2} \frac{e^{\eta R - \eta^2 V} - 1}{\eta}\, \mathrm{d}\eta + (1 - \pi(\mathcal{K})) \int_0^{1/2} \frac{e^{-\eta T - \eta^2 T} - 1}{\eta}\, \mathrm{d}\eta.
\end{aligned}
$$

Now first for the bad experts that are not in $\mathcal{K}$, we will show that

$$
\int_0^{1/2} \frac{e^{-\eta T - \eta^2 T} - 1}{\eta}\, \mathrm{d}\eta \;\geq\; -\frac{1}{2} - \ln(T+1). \tag{31}
$$

Let $\epsilon \in [0, 1/2]$ be arbitrary. Then, using $e^x \geq 1 + x$ and $e^x \geq 0$, we obtain

$$
\begin{aligned}
\int_0^{1/2} \frac{e^{-\eta T - \eta^2 T} - 1}{\eta}\, \mathrm{d}\eta &= \int_0^{\epsilon} \frac{e^{-\eta T - \eta^2 T} - 1}{\eta}\, \mathrm{d}\eta + \int_\epsilon^{1/2} \frac{e^{-\eta T - \eta^2 T} - 1}{\eta}\, \mathrm{d}\eta \\
&\geq \int_0^{\epsilon} \frac{-\eta T - \eta^2 T}{\eta}\, \mathrm{d}\eta + \int_\epsilon^{1/2} \frac{-1}{\eta}\, \mathrm{d}\eta = -\epsilon T - \frac{\epsilon^2}{2} T + \ln(2\epsilon).
\end{aligned}
$$

The choice $\epsilon = \frac{1}{2(T+1)}$ implies (31) for all $T \geq 0$.

Second, for the good experts that are in $\mathcal{K}$, we proceed as follows. Let $\hat{\eta} = \frac{R}{2V}$ be the η that maximizes $\eta R - \eta^2 V$. Then $\eta R - \eta^2 V$ is non-decreasing in η for $\eta \leq \hat{\eta}$ and hence, for any interval $[u, v] \subseteq [0, 1/2]$ such that $v \leq \hat{\eta}$,

$$
\begin{aligned}
\int_0^{1/2} \frac{e^{\eta R - \eta^2 V} - 1}{\eta}\, \mathrm{d}\eta &\geq \int_0^{u} \frac{e^{0R - 0V} - 1}{\eta}\, \mathrm{d}\eta + \left(e^{uR - u^2 V} - 1\right) \int_u^{v} \frac{1}{\eta}\, \mathrm{d}\eta - \int_v^{1/2} \frac{1}{\eta}\, \mathrm{d}\eta \\
&= \left(e^{uR - u^2 V} - 1\right) \ln\frac{v}{u} + \ln(2v). \tag{32}
\end{aligned}
$$

We may assume without loss of generality that $R \geq 2\sqrt{V}$ (otherwise (16) follows directly), which implies that $\hat{\eta} \geq \frac{1}{\sqrt{2V}}$.

We now have two cases. Suppose first that $\hat{\eta} \leq 1/2$. Then we plug in $v = \hat{\eta}$ and $u = \hat{\eta} - \frac{1}{\sqrt{2V}}$ and use $R \geq 2\sqrt{V}$ to find that

$$
\begin{aligned}
\int_0^{1/2} \frac{e^{\eta R - \eta^2 V} - 1}{\eta}\, \mathrm{d}\eta &\geq \left(e^{\frac{R^2}{4V} - \frac{1}{2}} - 1\right) \ln\left(\frac{1}{1 - \frac{\sqrt{2V}}{R}}\right) + \ln\left(\frac{R}{V}\right) \\
&\geq \left(e^{\frac{R^2}{4V} - \frac{1}{2}} - 1\right) \ln\left(\frac{1}{1 - \frac{\sqrt{2V}}{R}}\right) - \frac{1}{2}\ln\left(\frac{V}{4}\right).
\end{aligned}
$$

Using $e^{\frac{1}{2}(x^2 - 1)} = e^{\frac{1}{2}(x-1)^2} e^{x-1} \geq e^{\frac{1}{2}(x-1)^2} x$, $-1 \geq -\frac{R}{\sqrt{2V}}$ and $\ln\frac{1}{1-x} \geq x$, we find

$$
\left(e^{\frac{R^2}{4V} - \frac{1}{2}} - 1\right) \ln\left(\frac{1}{1 - \frac{\sqrt{2V}}{R}}\right) \;\geq\; \left(e^{\frac{1}{2}\left(\frac{R}{\sqrt{2V}} - 1\right)^2} \frac{R}{\sqrt{2V}} - \frac{R}{\sqrt{2V}}\right) \frac{\sqrt{2V}}{R} = e^{\frac{1}{2}\left(\frac{R}{\sqrt{2V}} - 1\right)^2} - 1.
$$

Putting everything together and using $V \leq T$ together with $1 + \frac{1}{2}\ln\frac{T}{4} \leq \frac{1}{2} + \ln(T+1)$ for $T \geq 1$, we get

$$0 \geq \pi(\mathcal{K}) \left(e^{\frac{1}{2}\left(\frac{R}{\sqrt{2V}}-1\right)^2} - 1 - \frac{1}{2}\ln\frac{V}{4} \right) - (1 - \pi(\mathcal{K})) \left(\frac{1}{2} + \ln(T+1) \right)$$

$$\geq \pi(\mathcal{K}) \left(e^{\frac{1}{2}\left(\frac{R}{\sqrt{2V}}-1\right)^2} \right) - \left(\frac{1}{2} + \ln(T+1) \right),$$

which implies

$$R \leq \sqrt{2V} \left(1 + \sqrt{2\ln\left(\frac{\frac{1}{2} + \ln(T+1)}{\pi(\mathcal{K})} \right)} \right),$$

and (16) follows.

It remains to consider the case that $\hat{\eta} > 1/2$, which implies $R > V$. We then use $v = 1/2$, for which (32) leads to

$$\int_0^{1/2} \frac{e^{\eta R - \eta^2 V} - 1}{\eta} \, d\eta \geq (e^{uR - u^2 V} - 1)\ln\frac{1}{2u} \geq (e^{u(1-u)R} - 1)\ln\frac{1}{2u}.$$

Putting everything together then gives

$$R \leq \frac{1}{u(1-u)} \ln\left(1 + \frac{(1 - \pi(\mathcal{K}))\left(\frac{1}{2} + \ln(T+1)\right)}{-\ln(2u)\pi(\mathcal{K})} \right) \leq \frac{1}{u(1-u)} \ln\left(1 + \frac{\frac{1}{2} + \ln(T+1)}{-\ln(2u)\pi(\mathcal{K})} \right).$$

And (16) follows by plugging in $u = \frac{5-\sqrt{5}}{10}$. ∎

A.4. Lemma 5

Proof Note that (21b) is minimized at the independent component-wise posteriors

$$\tilde{u}_{t+1}^k = \frac{u_t^k e^{-x_t^{k,1}}}{u_t^k e^{-x_t^{k,1}} + (1 - u_t^k)e^{-x_t^{k,0}}}. \tag{33}$$

The instantaneous mix loss regret in coordinate k in round t hence equals

$$-\ln\left(u_t^k e^{-x_t^{k,1}} + (1 - u_t^k)e^{-x_t^{k,0}} \right) - v^k x_t^{k,1} - (1 - v^k)x_t^{k,0}$$

$$= v^k \ln\frac{\tilde{u}_{t+1}^k}{u_t^k} + (1 - v^k)\ln\frac{1 - \tilde{u}_{t+1}^k}{1 - u_t^k}$$

$$= \triangle_2(v^k\|u_t^k) - \triangle_2(v^k\|\tilde{u}_{t+1}^k),$$

and we can write the cumulative regret as

$$\sum_{t=1}^T \sum_{k=1}^K \left(\triangle_2(v^k\|u_t^k) - \triangle_2(v^k\|\tilde{u}_{t+1}^k) \right) = \sum_{t=1}^T \left(\triangle_2(\boldsymbol{v}\|\boldsymbol{u}_t) - \triangle_2(\boldsymbol{v}\|\tilde{\boldsymbol{u}}_{t+1}) \right). \tag{34}$$

As $\triangle_2$ is a Bregman divergence (for convex generator $F(\boldsymbol{x}) = \sum_k x_k \ln x_k + (1-x_k)\ln(1-x_k)$), it is non-negative and satisfies the generalized Pythagorean inequality for Bregman divergences (Cesa-Bianchi and Lugosi, 2006, Lemma 11.3). Since $\boldsymbol{v} \in \mathcal{U}$ and $\boldsymbol{u}_{t+1}$ is the projection of $\tilde{\boldsymbol{u}}_{t+1}$ onto $\mathcal{U}$, these properties together imply that

$$\triangle_2(\boldsymbol{v}\|\boldsymbol{u}_{t+1}) \leq \triangle_2(\boldsymbol{v}\|\boldsymbol{u}_{t+1}) + \triangle_2(\boldsymbol{u}_{t+1}\|\tilde{\boldsymbol{u}}_{t+1}) \leq \triangle_2(\boldsymbol{v}\|\tilde{\boldsymbol{u}}_{t+1}).$$

Hence the cumulative mix loss regret satisfies

$$(34) \leq \sum_{t=1}^{T}\Big(\triangle_2(\boldsymbol{v}\|\boldsymbol{u}_t) - \triangle_2(\boldsymbol{v}\|\boldsymbol{u}_{t+1})\Big) = \triangle_2(\boldsymbol{v}\|\boldsymbol{u}_1) - \triangle_2(\boldsymbol{v}\|\boldsymbol{u}_{T+1}) \leq \triangle_2(\boldsymbol{v}\|\boldsymbol{\pi}),$$

as required. ∎

A.5. Lemma 6

Proof First, observe that, for any η,

$$e^{-\frac{1}{K}\ell_{\mathrm{mix}}(\boldsymbol{u}_t^\eta,\boldsymbol{x}_t^\eta)} \overset{(20),\,\mathrm{Jensen}}{\leq} \frac{1}{K}\sum_{k=1}^{K}\Big(u_t^{\eta,k}e^{-x_t^{\eta,k,1}} + (1-u_t^{\eta,k})e^{-x_t^{\eta,k,0}}\Big)$$

$$\overset{(24)}{=} \frac{1}{K}\sum_{k=1}^{K}\Big(u_t^{\eta,k}(1+\eta r_t^{k,1}) + (1-u_t^{\eta,k})(1+\eta r_t^{k,0})\Big)$$

$$\overset{(22)}{=} 1 + \frac{\eta}{K}\sum_{k=1}^{K}\Big(u_t^k - u_t^{\eta,k}\Big)\ell_t^k = 1 + \frac{\eta}{K}(\boldsymbol{u}_t - \boldsymbol{u}_t^\eta)^{\mathsf{T}}\boldsymbol{\ell}_t.$$

We hence have

$$\Phi(\boldsymbol{r}_{1:T+1}) - \Phi(\boldsymbol{r}_{1:T}) = \mathbb{E}_{\gamma(\eta)}\left[e^{-\frac{1}{K}\sum_{t=1}^{T}\ell_{\mathrm{mix}}(\boldsymbol{u}_t^\eta,\boldsymbol{x}^\eta)}\Big(e^{-\frac{1}{K}\ell_{\mathrm{mix}}(\boldsymbol{u}_{T+1}^\eta,\boldsymbol{x}_{T+1}^\eta)} - 1\Big)\right]$$

$$\leq \mathbb{E}_{\gamma(\eta)}\left[e^{-\frac{1}{K}\sum_{t=1}^{T}\ell_{\mathrm{mix}}(\boldsymbol{u}_t^\eta,\boldsymbol{x}^\eta)}\frac{\eta}{K}(\boldsymbol{u}_{T+1} - \boldsymbol{u}_{T+1}^\eta)^{\mathsf{T}}\boldsymbol{\ell}_{T+1}\right] = 0,$$

where the last equality is by design of the weights (23). ∎

A.6. Lemma 7

Proof Lemma 6 tells us that Component iProd ensures $\Phi(\boldsymbol{r}_{1:T}) \leq 0$. For any η, this implies

$$-K\ln\gamma(\eta) \overset{\mathrm{Lemma\,6}}{\geq} -\sum_{t=1}^{T}\ell_{\mathrm{mix}}(\boldsymbol{u}_t^\eta,\boldsymbol{x}_t^\eta) \overset{\mathrm{Lemma\,5}}{\geq} -\sum_{t=1}^{T}\sum_{k=1}^{K}\Big(v^k x_t^{\eta,k,1} + (1-v^k)x_t^{\eta,k,0}\Big) - \triangle_2(\boldsymbol{v}\|\boldsymbol{\pi})$$

$$\overset{(24),(8)}{\geq} -\sum_{t=1}^{T}\sum_{k=1}^{K}\Big(v^k\big(\eta^2(r_t^{k,1})^2 - \eta r_t^{k,1}\big) + (1-v^k)\big(\eta^2(r_t^{k,0})^2 - \eta r_t^{k,0}\big)\Big) - \triangle_2(\boldsymbol{v}\|\boldsymbol{\pi})$$

$$\overset{(25)}{=} \eta R_T^{\boldsymbol{v}} - \eta^2 V_T^{\boldsymbol{v}} - \triangle_2(\boldsymbol{v}\|\boldsymbol{\pi}),$$

from which the result follows. ∎

A.7. Theorem 8

Proof The exponentially spaced grid of learning rates ensures that, for any $\eta \in [\frac{1}{2T}, \frac{1}{2}]$, there always exists an $\alpha \in [\frac{1}{2}, 1]$ for which $\alpha\eta$ is a grid point. Hence, whenever $\hat{\eta} = \frac{R_T^v}{2V_T^v} \in [\frac{1}{2T}, 1/2]$, (26) implies that

$$R_T^v \leq \frac{4}{\sqrt{3}} \sqrt{V_T^v \big(\triangle_2(\boldsymbol{v}\|\boldsymbol{\pi}) + K \ln\lceil 1 + \log_2 T \rceil \big)},$$

and (27) is satisfied. Alternatively, if $\hat{\eta} < \frac{1}{2T}$, then $R_T^v < V_T^v/T \leq K$, and (27) again holds. Finally, supppose that $\hat{\eta} > 1/2$. Then $R_T^v > V_T^v$, and plugging $\eta = 1/2$ into Lemma 7 results in

$$\frac{1}{2}R_T^v - \frac{1}{4}V_T^v \leq \triangle_2(\boldsymbol{v}\|\boldsymbol{\pi}) + K \ln\lceil 1 + \log_2 T \rceil.$$

Using that $R_T^v > V_T^v$, the left-hand side is at most $\frac{1}{4}R_T^v$, from which (27) follows. ∎

Appendix B. Numerical stability

Some care should be taken when evaluating the weight expressions for the conjugate prior (11) and improper prior (15). Initially, and as long as $V = 0$, we have $R = 0$ and hence by (5) Squint sets the weights $\boldsymbol{w}$ equal to the prior π. We now assume $V > 0$, and look at (11) and (15). Both involve a contribution of the form

$$\frac{\sqrt{\pi} e^{\frac{R^2}{4V}} \left(\operatorname{erf}\left(\frac{R}{2\sqrt{V}}\right) - \operatorname{erf}\left(\frac{R-V}{2\sqrt{V}}\right) \right)}{2\sqrt{V}}. \tag{35}$$

This expression is empirically numerically stable unless both erf arguments fall outside $[-5, 5]$ to the same side. In other words, it can be used when

$$-5 \leq \frac{R}{2\sqrt{V}} \quad \text{and} \quad \frac{R-V}{2\sqrt{V}} \leq 5, \qquad \text{that is} \qquad R \in [-10\sqrt{V}, V + 10\sqrt{V}]. \tag{36}$$

If we are not in this range, then we are feeding extreme arguments into both erfs. Hence we may Taylor expand (35) around $R = \pm\infty$ (both of which give the same result)[3] to get

$$\frac{e^{\frac{R}{2} - \frac{V}{4}} - 1}{R} \quad \text{(0th and 1st order)} \qquad \text{or} \qquad \frac{e^{\frac{R}{2} - \frac{V}{4}}(R + V) - R}{R^2} \quad \text{(2nd order)}.$$

Note that this 0th order expansion is negative for $R \in [0, V/2]$, but that falls well within the stable range (36) where we should use (35) directly.

The range $[-5, 5]$ can be extended to $[-22, 22]$ by a careful use of the erfc function instead of erf, but the case distinction cannot be avoided entirely.

---

3. Taylor expansion of $f(x)$ around ∞ can be defined as expanding $f(1/x)$ around 0.