JMLR: Workshop and Conference Proceedings vol 35 (2014) 1–21

# A Second-order Bound with Excess Losses

**Pierre Gaillard** PIERRE-P.GAILLARD@EDF.FR
*EDF R&D, Clamart, France*
*GREGHEC: HEC Paris – CNRS, Jouy-en-Josas, France*

**Gilles Stoltz** STOLTZ@HEC.FR
*GREGHEC: HEC Paris – CNRS, Jouy-en-Josas, France*

**Tim van Erven** TIM@TIMVANERVEN.NL
*Université Paris-Sud, Orsay, France*

## Abstract

We study online aggregation of the predictions of experts, and first show new second-order regret bounds in the standard setting, which are obtained via a version of the Prod algorithm (and also a version of the polynomially weighted average algorithm) with multiple learning rates. These bounds are in terms of excess losses, the differences between the instantaneous losses suffered by the algorithm and the ones of a given expert. We then demonstrate the interest of these bounds in the context of experts that report their confidences as a number in the interval $[0, 1]$ using a generic reduction to the standard setting. We conclude by two other applications in the standard setting, which improve the known bounds in case of small excess losses and show a bounded regret against i.i.d. sequences of losses.

## 1. Introduction

In the (simplest) setting of prediction with expert advice, a learner has to make online sequential predictions over a series of rounds, with the help of $K$ experts (Freund and Schapire, 1997; Littlestone and Warmuth, 1994; Vovk, 1998; Cesa-Bianchi and Lugosi, 2006). In each round $t = 1, \dots, T$, the learner makes a prediction by choosing a vector $\boldsymbol{p}_t = (p_{1,t}, \dots, p_{K,t})$ of nonnegative weights that sum to one. Then every expert $k$ incurs a loss $\ell_{k,t} \in [a, b]$ and the learner's loss is $\widehat{\ell}_t = \boldsymbol{p}_t^\top \boldsymbol{\ell}_t = \sum_{k=1}^K p_{k,t}\ell_{k,t}$, where $\boldsymbol{\ell}_t = (\ell_{1,t}, \dots, \ell_{K,t})$. The goal of the learner is to control his cumulative loss, which he can do by controlling his regret $R_{k,T}$ against each expert $k$, where $R_{k,T} = \sum_{t \leqslant T} \bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)$. In the worst case, the best bound on the standard regret $R_{k,T}$ that can be guaranteed is of order $O\bigl(\sqrt{T \ln K}\bigr)$; see, e.g., Cesa-Bianchi and Lugosi (2006), but this can be improved. For example, when losses take values in $[0, 1]$, $R_{k,T} = O\bigl(\sqrt{L_{k,T} \ln K}\bigr)$, with $L_{k,T} = \sum_{t=1}^T \ell_{k,t}$, is also possible, which is better when the losses are small—hence the name *improvement for small losses* for this type of bounds (Cesa-Bianchi and Lugosi, 2006).

**Second-order bounds** Cesa-Bianchi et al. (2007) raised the question of whether it was possible to improve even further by proving second-order (variance-like) bounds on the regret. They could establish two types of bound, each with its own advantages. The first is of the form

$$R_{k,t} \leqslant \frac{\ln K}{\eta} + \eta \sum_{t=1}^T \ell_{k,t}^2 \qquad (1)$$

for all experts $k$, where $\eta \leqslant 1/2$ is a parameter of the algorithm. If one could optimize $\eta$ with hindsight knowledge of the losses, this would lead to the desired bound

$$\cancel{R_{k,T} = O\left(\sqrt{\ln K \sum_{t=1}^{T} \ell_{k,t}^2}\right)}, \tag{2}$$

but, unfortunately, no method is known that actually achieves (2) for all experts $k$ simultaneously without such hindsight knowledge. As explained by Cesa-Bianchi et al. (2007) and Hazan and Kale (2010), the technical difficulty is that the optimal $\eta$ would depend on $\sum_t \ell_{k_T^\star,t}^2$, where

$$k_T^\star \in \operatorname*{argmin}_{k=1,\dots,K} \left\{ \sum_{t=1}^{T} \ell_{k,t} + \sqrt{\ln K \sum_{t=1}^{T} \ell_{k,t}^2} \right\}.$$

But, because $k_T^\star$ can vary with $T$, the sequence of the $\sum \ell_{k_t^\star,t}^2$ is not monotonic and, as a consequence, standard tuning methods (like for example the doubling trick) cannot be applied directly on this sequence (only on the least non-decreasing sequence larger than it, which is then the key quantity in the regret bound though it is difficult to interpret).

This is why this issue — when hindsight bounds seem too good to be obtained in a sequential fashion — is sometimes referred to as the problem of *impossible tuning*. Improved bounds with respect to (1) have been obtained by Hazan and Kale (2010) and Chiang et al. (2012) but they suffer from the same impossible tuning issue.

The second type of bound distinguished by Cesa-Bianchi et al. (2007) is of the form

$$R_{k,T} = O\left(\sqrt{\ln K \sum_{t=1}^{T} v_t}\right), \tag{3}$$

uniformly over all experts $k$, where $v_t = \sum_{k \leqslant K} p_{k,t}\big(\widehat{\ell}_t - \ell_{k,t}\big)^2$ is the variance of the losses at instance $t$ under distribution $\boldsymbol{p}_t$. It can be achieved by a variant of the exponentially weighted average forecaster using the appropriate tuning of a time-varying learning rate $\eta_t$ (Cesa-Bianchi et al., 2007; de Rooij et al., 2013). The bound (3) was shown in the mentioned references to have several interesting consequences (see Section 5). Its main drawback comes from its uniformity: it does not reflect that it is harder to compete with some experts than with other ones.

**Excess losses** Instead of uniform regret bounds like (3), we aim to get expert-dependent regret bounds. We see this result as a steppingstone to solving the open problem of impossible tuning stated in (2).

The key quantities in our analysis turn out to be the instantaneous *excess losses* $\ell_{k,t} - \widehat{\ell}_t$, and we provide in Sections 2 and 3 a new second-order bound of the form

$$R_{k,T} = O\left(\sqrt{\ln K \sum_{t=1}^{T} (\widehat{\ell}_t - \ell_{k,t})^2}\right), \tag{4}$$

which holds for all experts $k$ simultaneously. To achieve this bound, we develop a variant of the Prod algorithm of Cesa-Bianchi et al. (2007) with two innovations: first we extend the analysis for Prod to multiple learning rates $\eta_k$ (one for each expert) in the spirit of a variant of the Hedge algorithm with multiple learning rates proposed by Blum and Mansour (2007). Standard tuning techniques for the learning rates would then still lead to an additional $O(\sqrt{K \ln T})$ multiplicative factor, so, secondly, we develop new techniques that bring this factor down to $O(\ln \ln T)$, which we consider to be essentially a constant. Duchi et al. (2011) also studied learning with multiple learning rates in a somewhat different context, namely, general online convex optimization; but the obtained regret bound is uniform over the experts.

The interest of the bound (4) is demonstrated in Sections 4 and 5, as well as in the recent paper by Wintenberger (2014). Section 4 considers the setting of prediction with experts that report their confidences as a number in the interval $[0, 1]$, which was first studied by Blum and Mansour (2007). Our general bound (4) leads to the first bound on the confidence regret that scales optimally with the confidences of each expert. Section 5 returns to the standard setting described at the beginning of this paper: we show an improvement for small excess losses, which supersedes the basic improvement for small losses described at the beginning of the introduction. Also, we prove that in the special case of independent, identically distributed losses, our bound leads to a constant regret. Finally, Wintenberger (2014) shows that bounds of the form (4) entail regret bounds on the cumulative predictive risks of the associated strategy without any assumption on the underlying stochastic process (in particular, without the usual dependency assumptions).

## 2. A new regret bound in the standard setting

We extend the Prod algorithm of Cesa-Bianchi et al. (2007) to work with multiple learning rates.

**Algorithm 1** Prod with multiple learning rates (ML-Prod)

*Parameters*: a vector $\boldsymbol{\eta} = (\eta_1, \dots, \eta_K)$ of positive learning rates
*Initialization*: a vector $\boldsymbol{w}_0 = (w_{1,0}, \dots, w_{K,0})$ of nonnegative weights that sum to 1

*For* each round $t = 1, 2, \dots$

1. form the mixture $\boldsymbol{p}_t$ defined component-wise by $p_{k,t} = \eta_k w_{k,t-1} / \boldsymbol{\eta}^\top \boldsymbol{w}_{t-1}$
2. observe the loss vector $\boldsymbol{\ell}_t$ and incur loss $\widehat{\ell}_t = \boldsymbol{p}_t^\top \boldsymbol{\ell}_t$
3. for each expert $k$ perform the update $w_{k,t} = w_{k,t-1}\bigl(1 + \eta_k\bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)\bigr)$

**Theorem 1** *For all sequences of loss vectors $\boldsymbol{\ell}_t \in [0,1]^K$, the cumulative loss of Algorithm 1 run with learning rates $\eta_k \leqslant 1/2$ is bounded by*

$$\sum_{t=1}^{T} \widehat{\ell}_t \leqslant \min_{1 \leqslant k \leqslant K} \left\{ \sum_{t=1}^{T} \ell_{k,t} + \frac{1}{\eta_k} \ln \frac{1}{w_{k,0}} + \eta_k \sum_{t=1}^{T} \bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)^2 \right\}.$$

If we could optimize the bound of the theorem with respect to $\eta_k$, we would obtain the desired result:

$$\sum_{t=1}^{T} \widehat{\ell}_t \leqslant \min_{1 \leqslant k \leqslant K} \left\{ \sum_{t=1}^{T} \ell_{k,t} + 2\sqrt{\sum_{t=1}^{T} V_{k,t} \ln \frac{1}{w_{k,0}}} \right\} \tag{5}$$

where $V_{k,t} = \bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)^2$. The question is therefore how to get the optimized bound (5) in a fully sequential way. Working in regimes (resorting to some doubling trick) seems suboptimal, since $K$ quantities $\sum_t V_{k,t}$ need to be controlled simultaneously and new regimes will start as soon as one of these quantities is larger than some dyadic threshold. This would lead to an additional $O(\sqrt{K \ln T})$ multiplicative factor in the bound. We propose in Section 3 a finer scheme, based on time-varying learning rates $\eta_{k,t}$, which only costs a multiplicative $O(\ln \ln T)$ factor in the regret bounds. Though the analysis of a single time-varying parameter is rather standard since the paper by Auer et al. (2002), the analysis of multiple such parameters is challenging and does not follow from a routine calculation. That the "impossible tuning" issue does not arise here was quite surprising to us.

**Empirical variance of the excess losses** A consequence of (5) is the following bound, which is in terms of the empirical variance of the excess losses $\ell_{k,t} - \widehat{\ell}_t$:

$$\sum_{t=1}^{T} \widehat{\ell}_t \leqslant \min_{1 \leqslant k \leqslant K} \left\{ \sum_{t=1}^{T} \ell_{k,t} + 4 \ln \frac{1}{w_{k,0}} + 2 \sqrt{ \sum_{t=1}^{T} \left( \widehat{\ell}_t - \ell_{k,t} - \frac{R_{k,T}}{T} \right)^2 \ln \frac{1}{w_{k,0}} } \right\}. \tag{6}$$

**Proposition 2** *Suppose losses take values in* $[0,1]$. *If* (5) *holds, then* (6) *holds.*

**Proof** A bias-variance decomposition indicates that, for each $k$,

$$\sum_{t=1}^{T} V_{k,t} = \sum_{t=1}^{T} \big(\widehat{\ell}_t - \ell_{k,t}\big)^2 = \sum_{t=1}^{T} \big(\widehat{\ell}_t - \ell_{k,t} - R_{k,T}/T\big)^2 + T \left(R_{k,T}/T\right)^2. \tag{7}$$

It is sufficient to prove the result when the minimum is restricted to $k$ such that $R_{k,T} \geqslant 0$. For such $k$, (5) implies that $R_{k,T}^2 \leqslant 4T \ln(1/w_{k,0})$. Substituting this into the rightmost term of (7), the result into (5), and using that $\sqrt{x+y} \leqslant \sqrt{x} + \sqrt{y}$ for $x, y \geqslant 0$ concludes the proof. ■

**Proof** [**of Theorem 1**] The proof follows from a simple adaptation of Lemma 2 in Cesa-Bianchi et al. (2007) and takes some inspiration from Section 6 of Blum and Mansour (2007).

For $t \geqslant 0$, we denote by $\boldsymbol{r}_t \in [-1,1]^K$ the instantaneous regret vector defined component-wise by $r_{k,t} = \widehat{\ell}_t - \ell_{k,t}$ and we define $W_t = \sum_{k=1}^{K} w_{k,t}$. We bound $\ln W_T$ from above and from below.

On the one hand, using the inequality $\ln(1+x) \geqslant x - x^2$ for all $x \geqslant -1/2$ (stated as Lemma 1 in Cesa-Bianchi et al., 2007), we have, for all experts $k$, that

$$\ln W_T \geqslant \ln w_{k,T} = \ln w_{k,0} + \sum_{t=1}^{T} \ln\big(1 + \eta_k r_{k,t}\big) \geqslant \ln w_{k,0} + \eta_k \sum_{t=1}^{T} r_{k,t} - \eta_k^2 \sum_{t=1}^{T} r_{k,t}^2 \,.$$

The last inequality holds because, by assumption, $\eta_k \leqslant 1/2$ and hence $\eta_k\big(\widehat{\ell}_t - \ell_{k,t}\big) \leqslant 1/2$ as well.

We now show by induction that, on the other hand, $W_T = W_0 = 1$ and thus that $\ln W_T = 0$. By definition of the weight update (step 3 of the algorithm), $W_t$ equals

$$\sum_{k=1}^{K} w_{k,t} = \sum_{k=1}^{K} w_{k,t-1}\Big(1 + \eta_k r_{k,t}\Big) = W_{t-1} + \underbrace{\left(\sum_{k=1}^{K} \eta_k w_{k,t-1}\right)}_{=\boldsymbol{\eta}^\top \boldsymbol{w}_{t-1}} \widehat{\ell}_t - \sum_{k=1}^{K} \underbrace{\eta_k w_{k,t-1}}_{=\boldsymbol{\eta}^\top \boldsymbol{w}_{t-1}\, p_{k,t}} \ell_{k,t} \,.$$

Substituting the definition of $\boldsymbol{p}_t$ (step 1 of the algorithm), as indicated in the line above, the last two sums are seen to cancel out, leading to $W_t = W_{t-1}$. Combining the lower bound on $\ln W_T$ with its value 0 and rearranging concludes the proof. ■

## 3. Algorithms and bound for parameters varying over time

To achieve the optimized bound (5), the learning parameters $\eta_k$ must be tuned using advance knowledge of the sums $\sum_{t=1}^{T}\bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)^2$. In this section we show how to remove this requirement, at the cost of a logarithmic factor $\ln\ln T$ only (unlike what would be obtained by working in regimes as mentioned above). We do so by having the learning rates $\eta_{k,t}$ for each expert vary with time.

### 3.1. Multiplicative updates (adaptive version of ML-Prod)

We generalize Algorithm 1 and Theorem 1 to Algorithm 2 and Theorem 3.

**Theorem 3** *For all sequences of loss vectors $\boldsymbol{\ell}_t \in [0,1]^K$, for all rules prescribing sequences of learning rates $\eta_{k,t} \leqslant 1/2$ that, for each $k$, are nonincreasing in $t$, the cumulative loss $\sum_{t\leqslant T}\widehat{\ell}_t$ of Algorithm 2 is bounded by*

$$\min_{1\leqslant k\leqslant K}\left\{\sum_{t=1}^{T}\ell_{k,t} + \frac{1}{\eta_{k,0}}\ln\frac{1}{w_{k,0}} + \sum_{t=1}^{T}\eta_{k,t-1}\bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)^2 \right.$$
$$\left. + \frac{1}{\eta_{k,T}}\ln\left(1 + \frac{1}{\mathrm{e}}\sum_{k'=1}^{K}\sum_{t=1}^{T}\left(\frac{\eta_{k',t-1}}{\eta_{k',t}} - 1\right)\right)\right\}.$$

---

**Algorithm 2** Prod with multiple adaptive learning rates (Adapt-ML-Prod)

---

*Parameter*: a rule to sequentially pick positive learning rates
*Initialization*: a vector $\boldsymbol{w}_0 = (w_{1,0},\dots,w_{K,0})$ of nonnegative weights that sum to 1
*For* each round $t = 1,\, 2,\, \dots$

0. pick the learning rates $\eta_{k,t-1} > 0$ according to the rule
1. form the mixture $\boldsymbol{p}_t$ defined component-wise by $p_{k,t} = \eta_{k,t-1}w_{k,t-1}\big/\boldsymbol{\eta}_{t-1}^{\top}\boldsymbol{w}_{t-1}$
2. observe the loss vector $\boldsymbol{\ell}_t$ and incur loss $\widehat{\ell}_t = \boldsymbol{p}_t^{\top}\boldsymbol{\ell}_t$
3. for each expert $k$ perform the update
$$w_{k,t} = \Bigl(w_{k,t-1}\Bigl(1 + \eta_{k,t-1}\bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)\Bigr)\Bigr)^{\frac{\eta_{k,t}}{\eta_{k,t-1}}}$$

---

**Corollary 4** *With uniform initial weights $\boldsymbol{w}_0 = (1/K,\dots,1/K)$ and learning rates, for $t \geqslant 1$,*

$$\eta_{k,t-1} = \min\left\{\frac{1}{2},\, \sqrt{\frac{\ln K}{1 + \sum_{s=1}^{t-1}\bigl(\widehat{\ell}_s - \ell_{k,s}\bigr)^2}}\right\},$$

*the cumulative loss of Algorithm 2 is bounded by*

$$\min_{1\leqslant k\leqslant K}\left\{\sum_{t=1}^{T}\ell_{k,t} + \frac{C_{K,T}}{\sqrt{\ln K}}\sqrt{1 + \sum_{t=1}^{T}\bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)^2} + 2C_{K,T}\right\},$$

*where* $C_{K,T} = 3\ln K + \ln\left(1 + \frac{K}{2\mathrm{e}}\bigl(1 + \ln(T+1)\bigr)\right) = O(\ln K + \ln\ln T)$.

This optimized corollary is the adaptive version of (5). Its proof is postponed to Section A.3 of the additional material (and shows that meaningful bounds can be achieved as well with non-uniform initial weights). Here we only give the main ideas in the proof of Theorem 3. The complete argument is given in Section A.2 of the additional material. We point out that the proof technique is not a routine adaptation of well-known tuning tricks such as, for example, the ones of Auer et al. (2002).

**Proof** [**sketch for Theorem 3**] We follow the path of the proof of Theorem 1 and bound $\ln W_T$ from below and from above. The lower bound is easy to establish as it only relies on individual non-increasing sequences of rates, $(\eta_{k,t})_{t \geqslant 0}$ for a fixed $k$: the weight update (step 3 of the algorithm) was indeed tailored for it to go through. More precisely, by induction and still with the inequality $\ln(1+x) \geqslant x - x^2$ for $x \geqslant -1/2$, we get that

$$\ln W_T \geqslant \ln w_{k,T} \geqslant \frac{\eta_{k,T}}{\eta_{k,0}} \ln w_{k,0} + \eta_{k,T} \sum_{t=1}^{T} \bigl( r_{k,t} - \eta_{k,t-1} r_{k,t}^2 \bigr) \,.$$

The difficulties arise in proving an upper bound. We proceed by induction again and aim at upper bounding $W_t$ by $W_{t-1}$ plus some small term. The core difficulty is that the powers $\eta_{k,t}/\eta_{k,t-1}$ in the weight update are different for each $k$. In the literature, time-varying parameters could previously be handled using Jensen's inequality for the function $x \mapsto x^{\alpha_t}$ with a parameter $\alpha_t = \eta_t/\eta_{t-1} \geqslant 1$ that was the same for all experts: this is, for instance, the core of the argument in the main proof of Auer et al. (2002) as noticed by Györfi and Ottucsák (2007) in their re-worked version of the proof. This needs to be adapted here as we have $\alpha_{k,t} = \eta_{k,t-1}/\eta_{k,t}$, which depends on $k$. We quantify the cost for the $\alpha_{k,t}$ not to be all equal to a single power $\alpha_t$, say 1: we have $\alpha_{k,t} \geqslant 1$ but the gap to 1 should not be too large. This is why we may apply the inequality $x \leqslant x^{\alpha_{k,t}} + (\alpha_{k,t} - 1)/e$, valid for all $x > 0$ and $\alpha_{k,t} \geqslant 1$. We can then prove that

$$W_t \leqslant W_{t-1} + \frac{1}{e} \sum_{k=1}^{K} \left( \frac{\eta_{k,t-1}}{\eta_{k,t}} - 1 \right) ,$$

where the second term on the right-hand side is precisely the price to pay for having different time-varying learning rates — and this price is measured by how much they vary. ■

### 3.2. Polynomial potentials

As illustrated in Cesa-Bianchi and Lugosi (2003), polynomial potentials are also useful to minimize the regret. We present here an algorithm based on them (with order $p = 2$ in the terminology of the indicated reference). Its bound has the same poor dependency on the number of experts $K$ and on $T$ as achieved by working in regimes (see the discussion in Section 2), but its analysis is simpler and more elegant than that of Algorithm 2 (see Section A.4 in the appendix; the analysis resembles the proof of Blackwell's approachability theorem). The right dependencies might be achieved by considering polynomial functions of arbitrary orders $p$ as in Cesa-Bianchi and Lugosi (2003) but we were unable to provide an analysis for these values.

**Theorem 5** *For all sequences of loss vectors $\boldsymbol{\ell}_t \in [0,1]^K$, the cumulative loss of Algorithm 3 run with learning rates*

$$\eta_{k,t-1} = \frac{1}{1 + \sum_{s=1}^{t-1} \bigl( \widehat{\ell}_s - \ell_{k,s} \bigr)^2}$$

**Algorithm 3** Polynomially weighted averages with multiple learning rates (ML-Poly)

*Parameter*: a rule to sequentially pick positive learning rates $\boldsymbol{\eta}_t = \big(\eta_{1,t}, \ldots, \eta_{K,t}\big)$
*Initialization*: the vector of regrets with each expert $\boldsymbol{R}_0 = (0, \ldots, 0)$

*For* each round $t = 1,\, 2,\, \ldots$

0. pick the learning rates $\eta_{k,t-1}$ according to the rule
1. form the mixture $\boldsymbol{p}_t$ defined component-wise by $p_{k,t} = \eta_{k,t-1}\,(R_{k,t-1})_+ \;/\; \boldsymbol{\eta}_{t-1}^\top\,(\boldsymbol{R}_{t-1})_+$ where $\boldsymbol{x}_+$ denotes the vector of the nonnegative parts of the components of $\boldsymbol{x}$
2. observe the loss vector $\boldsymbol{\ell}_t$ and incur loss $\widehat{\ell}_t = \boldsymbol{p}_t^\top \boldsymbol{\ell}_t$
3. for each expert $k$ update the regret: $R_{k,t} = R_{k,t-1} + \widehat{\ell}_t - \ell_{k,t}$

*is bounded by*
$$\sum_{t=1}^T \widehat{\ell}_t \leqslant \min_{1 \leqslant k \leqslant K} \left\{ \sum_{t=1}^T \ell_{k,t} + \sqrt{K\big(1+\ln(1+T)\big)\left(1+\sum_{t=1}^T \big(\widehat{\ell}_t - \ell_{k,t}\big)^2\right)} \right\}.$$

## 4. First application: bounds with experts that report their confidences

We justify in this section why the second-order bounds exhibited in the previous sections are particularly adapted to the setting of prediction with experts that report their confidences, which was first considered by Blum and Mansour (2007). It differs from the standard setting in that, at the start of every round $t$, each expert $k$ expresses their confidence as a number $I_{k,t} \in [0,1]$.[1] In particular, confidence $I_{k,t} = 0$ expresses that expert $k$ is inactive (or sleeping) in round $t$. The learner now has to assign nonnegative weights $\boldsymbol{p}_t$, which sum up to 1, to the set $\mathcal{A}_t = \{k : I_{k,t} > 0\}$ of so-called *active experts* and suffers loss $\widehat{\ell}_t = \sum_{k \in \mathcal{A}_t} p_{k,t}\ell_{k,t}$. (It is assumed that, for any round $t$, there is at least one active expert $k$ with $I_{k,t} > 0$, so that $\mathcal{A}_t$ is never empty.)

The main difference in prediction with confidences comes from the definition of the regret. The *confidence regret* with respect to expert $k$ takes the numbers $I_{k,t}$ into account and is defined as $R^{\mathrm{c}}_{k,T} = \sum_{t=1}^T I_{k,t}\big(\widehat{\ell}_t - \ell_{k,t}\big)$.

When $I_{k,t}$ is always 1, prediction with confidences reduces to regular prediction with expert advice, and when the confidences $I_{k,t}$ only take on the values 0 and 1, it reduces to prediction with sleeping (or specialized) experts as introduced by Blum (1997) and Freund et al. (1997).

Because the confidence regret scales linearly with $I_{k,t}$, one would like to obtain bounds on the confidence regret that scale linearly as well. When confidences do not depend on $k$, this is achieved, e.g., by the bound (3). However, for confidences that do depend on $k$, the best available bound (Blum and Mansour, 2007, Theorem 16) is

$$R^{\mathrm{c}}_{k,T} = \sum_{t=1}^T I_{k,t}\big(\widehat{\ell}_t - \ell_{k,t}\big) = O\left(\sqrt{\sum_{t \leqslant T} I_{k,t}\ell_{k,t}}\right). \tag{8}$$

(We rederive this bound in Gaillard et al., 2014, Section 5.2.) If, in this bound, all confidences $I_{k,t}$ are scaled down by a factor $\lambda_k \in [0,1]$, then we would like the bound to also scale down by $\lambda_k$, but instead it scales only by $\sqrt{\lambda_k}$. In the remainder of this section we will show how our new

1. Technically, Blum and Mansour (2007) decouple the confidences $I_{k,t}$, which they refer to as “time selection functions”, from the experts, but as we explain in Gaillard et al. (2014, Section 5.2), the two settings are equivalent.

second-order bound (4) solves this issue via a generic reduction of the setting of prediction with confidences to the standard setting from Sections 1 and 2.

**Remark 6** *We consider the case of linear losses. The extension of our results to convex losses is immediate via the so-called gradient trick. The latter also applies in the setting of experts that report their confidences. The details were essentially provided by Devaine et al. (2013) (we recall them in Gaillard et al., 2014, Section B.1).*

**Generic reduction to the standard setting** There exists a generic reduction from the setting of sleeping experts to the standard setting of prediction with expert advice (Adamskiy et al., 2012; Koolen et al., 2013). This reduction generalizes easily to the setting of experts that report their confidences, as we will now explain.

Given any algorithm designed for the standard setting, we run it on modified losses $\widetilde{\ell}_{k,s}$, which will be defined shortly. At round $t \geqslant 1$, the algorithm takes as inputs the past modified losses $\widetilde{\ell}_{k,s}$, where $s \leqslant t-1$, and outputs a weight vector $\widetilde{\boldsymbol{p}}_t$ on $\{1,\ldots,K\}$. This vector is then used to form another weight vector $\boldsymbol{p}_t$, which has strictly positive weights only on $\mathcal{A}_t$:

$$p_{k,t} = \frac{I_{k,t}\,\widetilde{p}_{k,t}}{\sum_{k'=1}^{K} I_{k',t}\,\widetilde{p}_{k',t}} \qquad \text{for all } k. \tag{9}$$

This vector $\boldsymbol{p}_t$ is to be used with the experts that report their confidences. Then, the losses $\ell_{k,t}$ are observed and the modified losses are computed as follows: for all $k$,

$$\widetilde{\ell}_{k,t} = I_{k,t}\ell_{k,t} + (1-I_{k,t})\widehat{\ell}_t \qquad \text{where} \quad \widehat{\ell}_t = \sum_{k\in\mathcal{A}_t} p_{k,t}\ell_{k,t}\,.$$

**Proposition 7** *The induced confidence regret on the original losses $\ell_{k,t}$ equals the standard regret of the algorithm on the modified losses $\widetilde{\ell}_{k,t}$. In particular,*

$$I_{k,t}\big(\widehat{\ell}_t - \ell_{k,t}\big) = \sum_{i=1}^{K} \widetilde{p}_{i,t}\widetilde{\ell}_{i,t} - \widetilde{\ell}_{k,t} \qquad \textit{for all rounds } t \textit{ and experts } k.$$

**Proof** First we show that the loss in the standard setting (on the losses $\widetilde{\ell}_{k,t}$) is equal to the loss in the confidence regret setting (on the original losses $\ell_{k,t}$):

$$\begin{aligned}\sum_{k=1}^{K}\widetilde{p}_{k,t}\widetilde{\ell}_{k,t} &= \sum_{k=1}^{K}\widetilde{p}_{k,t}\Big(I_{k,t}\ell_{k,t} + (1-I_{k,t})\widehat{\ell}_t\Big) = \sum_{k=1}^{K}\widetilde{p}_{k,t}I_{k,t}\ell_{k,t} + \widehat{\ell}_t - \left(\sum_{k=1}^{K}\widetilde{p}_{k,t}I_{k,t}\right)\widehat{\ell}_t \\ &= \left(\sum_{k'=1}^{K}\widetilde{p}_{k',t}\,I_{k',t}\right)\sum_{k=1}^{K}p_{k,t}\ell_{k,t} + \widehat{\ell}_t - \left(\sum_{k=1}^{K}\widetilde{p}_{k,t}I_{k,t}\right)\widehat{\ell}_t = \widehat{\ell}_t.\end{aligned}$$

The proposition now follows by subtracting $\widetilde{\ell}_{k,t}$ on both sides of the equality. ■

**Corollary 8** *An algorithm with a standard regret bound of the form*

$$R_{k,T} \leqslant \Xi_1\sqrt{(\ln K)\sum_{t\leqslant T}\big(\widehat{\ell}_t - \ell_{k,t}\big)^2} + \Xi_2 \qquad \textit{for all } k, \tag{10}$$

*leads, via the generic reduction described above (and for losses $\ell_{k,t} \in [0,1]$), to an algorithm with a confidence regret bound of the form*

$$R^c_{k,T} \leqslant \Xi_1 \sqrt{(\ln K) \sum_{t \leqslant T} I^2_{k,t} \bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)^2} + \Xi_2 \leqslant \Xi_1 \sqrt{(\ln K) \sum_{t \leqslant T} I^2_{k,t}} + \Xi_2 \qquad \textit{for all } k. \tag{11}$$

We note that the second upper-bound, $\sqrt{\sum I^2_{k,t}}$, can be extracted from the proof of Theorem 11 in Chernov and Vovk (2010)—but not the first one, which, combined with the techniques of Section 5.1, yields a bound on the confidence regret for small (excess) losses.

**Comparison to the instantiation of other regret bounds** We now discuss why (11) improves on the literature. Consider first the improved bound for small losses from the introduction, which takes the form $\Xi_3 \sqrt{\sum_t \ell_{k,t}} + \Xi_4$. This improvement does not survive the generic reduction, as the resulting confidence regret bound is

$$\Xi_3 \sqrt{\sum_{t=1}^T \widetilde{\ell}_{k,t}} + \Xi_4 = \Xi_3 \sqrt{\sum_{t=1}^T I_{k,t} \ell_{k,t} + \underbrace{\sum_{t=1}^T (1 - I_{k,t}) \widehat{\ell}_t}_{\text{undesirable}}} + \Xi_4,$$

which is no better than plain $\Xi_3' \sqrt{T} + \Xi_4'$ bounds.

Alternatively, bounds (3) of Cesa-Bianchi et al. (2007) and de Rooij et al. (2013) are of the form

$$\Xi_5 \sqrt{\sum_{t=1}^T \sum_{k=1}^K p_{k,t} \bigl(\ell_{k,t} - \widehat{\ell}_t\bigr)^2} + \Xi_6,$$

uniformly over all experts $k$. These lead to a confidence regret bound against expert $k$ of the form

$$\Xi_5 \sqrt{\sum_{t=1}^T \sum_{k=1}^K p_{k,t} \, I^2_{k,t} \bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)^2} + \Xi_6 \leqslant \Xi_5 \sqrt{\sum_{t=1}^T \sum_{k=1}^K p_{k,t} \, I^2_{k,t}} + \Xi_6,$$

which depends not just on the confidences of this expert $k$, but also on the confidences of the other experts. It therefore does not scale proportionally to the confidences of the expert $k$ at hand.

We note that even bounds of the form (2), if they existed, would not be suitable either. They would indeed lead to

$$R^c_{k,T} = O\left( \sqrt{\sum_{t=1}^T \Bigl( I_{k,t} \ell_{k,t} + (1 - I_{k,t}) \widehat{\ell}_t \Bigr)^2} \right),$$

which also does not scale linearly with the confidences of expert $k$.

## 5. Other applications: bounds in the standard setting

We now leave the setting of prediction with confidences, and detail other applications of our new second-order bound (4). First, in Section 5.1, we show that, like (1) and (3), our new bound implies an improvement over the standard bound $O\big(\sqrt{\sum_t \ell_{k,t} \ln K}\big)$, which is itself already better than the worst-case bound if the losses of the reference expert are small. The key feature in our improvement is that excess losses $\ell_{k,t} - \widehat{\ell}_t$ can be considered instead of plain losses $\ell_{k,t}$. Then, in Section 5.2, we look at the non-adversarial setting in which losses are i.i.d., and show that our new bound implies constant regret of order $O\big(\ln K\big)$.

### 5.1. Improvement for small excess losses

It is known (Cesa-Bianchi et al., 2007; de Rooij et al., 2013) that (3) implies a bound of the form

$$R_{k^*,T} = O\left(\sqrt{\ln K \frac{L_{k^*,T}(T - L_{k^*,T})}{T}}\right), \tag{12}$$

where $k^* \in \operatorname{argmin}_k L_{k,T}$ is the expert with smallest cumulative loss. This bound symmetrizes the standard bound for small losses described in the introduction, because it is small also if $L_{k^*,T}$ is close to $T$, which is useful when losses are defined in terms of gains (Cesa-Bianchi et al., 2007).

However, if one is ready to lose symmetry, another way of improving the standard bound for small losses is to express it in terms of *excess* losses:

$$\sqrt{\ln K \sum_{t\colon \ell_{k,t} \geqslant \widehat{\ell}_t} \Big(\ell_{k,t} - \widehat{\ell}_t\Big)} \leqslant \sqrt{\ln K \sum_{t \leqslant T} \ell_{k,t}}\,,$$

where the inequality holds for nonnegative losses. As we show next, bounds of the form (4) indeed entail bounds of this form.

**Theorem 9** *If the regret of an algorithm satisfies* (10) *for all sequences of loss vectors* $\boldsymbol{\ell}_t \in [0,1]^K$*, then it also satisfies*

$$R_{k,T} \leqslant 2\,\Xi_1 \sqrt{\ln K \sum_{t\colon \ell_{k,t} \geqslant \widehat{\ell}_t} \Big(\ell_{k,t} - \widehat{\ell}_t\Big)} + \big(\Xi_2 + 2\,\Xi_1 \sqrt{\Xi_2 \ln K} + 4\,\Xi_1^2 \ln K\big)\,. \tag{13}$$

In general, losses take values in the range $[a, b]$. To apply our methods, they therefore need to be translated by $-a$ and scaled by $1/(b-a)$ to fit the canonical range $[0, 1]$. In the standard improvement for small losses, these operations remain visible in the regret bound, which becomes $R_{k,T} = O\big(\sqrt{(b-a)(L_{k,T} - Ta)\ln K}\big)$ in general. In particular, if $a < 0$, then no significant improvement over the worst-case bound $O\big(\sqrt{T \ln K}\big)$ is realized. By contrast, our original second-order bound (10) and its corollary (13) both have the nice feature that translations do not affect the bound because $(\ell_{k,t} - a) - (\widehat{\ell}_t - a) = \ell_{k,t} - \widehat{\ell}_t$, so that our new improvement for small losses remains meaningful even for $a < 0$.

**Proof** We define the positive and the negative part of the regret with respect to an expert $k$ by, respectively,

$$R^+_{k,T} = \sum_{t=1}^T \Big(\widehat{\ell}_t - \ell_{k,t}\Big)\,\mathbb{1}_{\ell_{k,t} \leqslant \widehat{\ell}_t} \qquad \text{and} \qquad R^-_{k,T} = \sum_{t=1}^T \Big(\ell_{k,t} - \widehat{\ell}_t\Big)\,\mathbb{1}_{\ell_{k,t} \geqslant \widehat{\ell}_t}\,.$$

The proof will rely on rephrasing the bound (10) in terms of $R^+_{k,T}$ and $R^-_{k,T}$ only. On the one hand, $R_{k,T} = R^+_{k,T} - R^-_{k,T}$, while, on the other hand,

$$\sqrt{\sum_{t\leqslant T}\bigl(\widehat{\ell}_t - \ell_{k,t}\bigr)^2} \leqslant \sqrt{\sum_{t\leqslant T}\Bigl|\widehat{\ell}_t - \ell_{k,t}\Bigr|} = \sqrt{R^+_{k,T} + R^-_{k,T}} \leqslant 2\sqrt{R^+_{k,T}}\,, \tag{14}$$

where we used $\ell_{k,t} \in [0,1]$ for the first inequality and where we assumed, with no loss of generality, that $R^+_{k,T} \geqslant R^-_{k,T}$. Indeed, if this was not the case, the regret would be negative and the bound would be true. Therefore for all experts $k$, substituting these (in)equalities in the initial inequality (10), we are left with the quadratic inequality

$$R^+_{k,T} - R^-_{k,T} \leqslant 2\Xi_1\sqrt{R^+_{k,T}\ln K} + \Xi_2\,. \tag{15}$$

Solving for $R^+_{k,T}$ using Lemma 10 below (whose proof can be found in Section A.1) yields

$$\sqrt{R^+_{k,T}} \leqslant \sqrt{R^-_{k,T} + \Xi_2} + 2\Xi_1\sqrt{\ln K} \leqslant \sqrt{R^-_{k,T}} + \sqrt{\Xi_2} + 2\Xi_1\sqrt{\ln K}\,,$$

which leads to the stated bound after re-substitution into (15). ■

**Lemma 10** *Let $a, c \geqslant 0$. If $x \geqslant 0$ satisfies $x^2 \leqslant a + cx$, then $x \leqslant \sqrt{a} + c$.*

### 5.2. Stochastic (i.i.d.) losses

Van Erven et al. (2011) provide a specific parameter-free algorithm that guarantees worst-case regret bounded by $O\left(\sqrt{L_{k^\star,T}\ln K}\right)$, but at the same time is able to adapt to the non-adversarial setting with independent, identically distributed (i.i.d.) loss vectors, for which its regret is bounded by $O(K)$. Theorem 9 already indicated that any algorithm satisfying a regret bound of the form (10) also achieves a worst-case bound that is at least as good as $O\left(\sqrt{L_{k^\star,T}\ln K}\right)$. Here we consider i.i.d. losses that satisfy the same assumption as the one imposed by Van Erven et al.:

**Assumption 1** *The loss vectors $\boldsymbol{\ell}_t \in [0,1]^K$ are independent random variables such that there exists an action $k^\star$ and some $\alpha \in (0,1]$ for which the expected differences in loss satisfy*

$$\forall t \geqslant 1, \qquad \min_{k\neq k^\star} \mathbb{E}\bigl[\ell_{k,t} - \ell_{k^\star,t}\bigr] \geqslant \alpha\,.$$

As shown by the following theorem, any algorithm that satisfies our new second-order bound (with a constant $\Xi_1$ factor and a $\Xi_2$ factor of order $\ln K$) is guaranteed to achieve constant regret of order $O(\ln K)$ under Assumption 1.

**Theorem 11** *If a strategy achieves a regret bound of the form (10) and the loss vectors satisfy Assumption 1, then the expected regret for that strategy is bounded by a constant: for all $T$,*

$$\mathbb{E}[R_{k^\star,T}] \leqslant C_K \stackrel{\text{def}}{=} (\Xi_1^2 \ln K)/\alpha + \Xi_1\sqrt{(\Xi_2\ln K)/\alpha} + \Xi_2\,;$$

*while for any $T$ and any $\delta \in (0,1)$, its regret is bounded with probability at least $1-\delta$ by*

$$R_{k^\star,T} \leqslant C_K + \frac{6\,\Xi_1}{\alpha}\sqrt{\left(\ln\frac{1}{\delta} + \ln\!\left(1 + \frac{1}{2\mathrm{e}}\ln\bigl(1 + C_K/4\bigr)\right)\right)\ln K}\,.$$

By the law of large numbers, the cumulative loss of any action $k \neq k^\star$ will exceed the cumulative loss of $k^\star$ by a linear term in the order of $\alpha T$, so that, for all sufficiently large $T$, the fact that $R_{k^\star,T}$ is bounded by a constant implies that the algorithm will have negative regret with respect to all other $k$.

Because we want to avoid using any special properties of the algorithm except for the fact that it satisfies (10), our proof of Theorem 11 requires a Bernstein-Freedman-type martingale concentration result (Freedman, 1975) rather than basic applications of Hoeffding's inequality, which are sufficient in the proof of Van Erven et al. (2011). However, this type of concentration inequalities is typically stated in terms of a deterministic bound $M$ on the cumulative conditional variance $\sum V_t$. To bound the deviations by the (random) quantity $\sqrt{\sum V_t}$ instead of the deterministic $\sqrt{M}$, peeling techniques can be applied as in Cesa-Bianchi et al. (2005, Corollary 16); this leads to an additional $\sqrt{\ln T}$ factor (in case of an additive peeling) or $\sqrt{\ln \ln T}$ (in case of a geometric peeling). Here, we replace these non-constant factors by a term of order $\ln \ln \mathbb{E}\big[\sum V_t\big]$, which will be seen to be less than a constant in our case.

**Theorem 12** *Let $(X_t)_{t \geqslant 1}$ be a martingale difference sequence with respect to some filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$ and let $V_t = \mathbb{E}\big[X_t^2 \,\big|\, \mathcal{F}_{t-1}\big]$ for $t \geqslant 1$. We assume that $X_t \leqslant 1$ a.s., for all $t \geqslant 1$. Then, for any $\delta \in (0,1)$ and any $T \geqslant 1$, with probability at least $1-\delta$,*

$$\sum_{t=1}^{T} X_t \leqslant 3 \sqrt{\left(1 + \sum_{t=1}^{T} V_t\right) \ln \frac{\gamma}{\delta}} + \ln \frac{\gamma}{\delta}\,, \quad \textit{where} \quad \gamma = 1 + \frac{1}{2\mathrm{e}}\left(1 + \ln\left(1 + \mathbb{E}\left[\sum_{t=1}^{T} V_t\right]\right)\right).$$

Theorem 12 and its proof (see Gaillard et al., 2014, Section A.5 for the latter) may be of independent interest, because our derivation uses new techniques that we originally developed for time-varying learning rates in the proof of Theorem 3. Instead of studying supermartingales of the form $\exp\big(\lambda \sum X_t - (\mathrm{e}-2)\lambda^2 \sum V_t\big)$ for some constant value of $\lambda$, as is typical, we are able to consider (predictable) random variables $\Lambda_t$, which in some sense play the role of the time-varying learning parameter $\eta_t$ of the (ML-)Prod algorithm.

**Proof [of Theorem 11]** We recall the notation $r_{k,t} = \widehat{\ell}_t - \ell_{k,t}$ for the instantaneous regret. We start from $\mathcal{F}_0$, the trivial $\sigma$–algebra $\{\emptyset, \Omega\}$ (consisting of the empty set and the whole underlying probability space), and define by induction the following martingale difference sequence: for all $t \geqslant 1$,

$$Y_t = -r_{k^\star,t} + \mathbb{E}\big[r_{k^\star,t} \,\big|\, \mathcal{F}_{t-1}\big]$$

and $\mathcal{F}_t = \sigma(Y_1, \ldots, Y_t)$ is the $\sigma$–algebra generated by the random variables $Y_1, \ldots, Y_t$. We first bound the expectation of the regret. We note that

$$\mathbb{E}\big[r_{k^\star,t} \,\big|\, \mathcal{F}_{t-1}\big] = \sum_{k=1}^{K} p_{k,t}\, \mathbb{E}\big[\ell_{k,t} - \ell_{k^\star,t} \,\big|\, \mathcal{F}_{t-1}\big] = \sum_{k=1}^{K} p_{k,t}\, \mathbb{E}\big[\ell_{k,t} - \ell_{k^\star,t}\big] \geqslant \alpha(1 - p_{k^\star,t})\,, \quad (16)$$

while by convexity of $(\,\cdot\,)^2$,
$$r_{k^\star,t}^2 \leqslant \sum_{k=1}^{K} p_{k,t} \big(\ell_{k,t} - \ell_{k^\star,t}\big)^2 \leqslant 1 - p_{k^\star,t}\,, \quad (17)$$

so that
$$W_t = \mathbb{E}\big[Y_t^2 \,\big|\, \mathcal{F}_{t-1}\big] \leqslant \mathbb{E}\big[r_{k^\star,t}^2 \,\big|\, \mathcal{F}_{t-1}\big] \leqslant 1 - p_{k^\star,t}\,. \quad (18)$$

Therefore, using that expectations of conditional expectations are unconditional expectations,

$$\mathbb{E}[R_{k^\star,T}] \geqslant \alpha\, \mathbb{E}[S_T] \quad \text{and} \quad \mathbb{E}\left[\sum_{t=1}^{T} r_{k^\star,t}^2\right] \leqslant \mathbb{E}[S_T] \qquad \text{where} \quad S_T = \sum_{t=1}^{T} (1 - p_{k^\star,t})\,. \tag{19}$$

Substituting these inequalities in (10) using Jensen's inequality for $\sqrt{\cdot}$, we get

$$\mathbb{E}[S_T] \leqslant \frac{\Xi_1 \sqrt{\ln K}}{\alpha} \sqrt{\mathbb{E}[S_T]} + \frac{\Xi_2}{\alpha}\,.$$

Solving the quadratic inequality (see Lemma 10) yields $\mathbb{E}[S_T] \leqslant \bigl((\Xi_1 \sqrt{\ln K})/\alpha + \sqrt{\Xi_2/\alpha}\bigr)^2$. By (19) this bounds $\mathbb{E}\left[\sum_{t=1}^{T} r_{k^\star,t}^2\right]$, which we substitute into (10), together with Jensen's inequality, to prove the claimed bound on the expected regret.

Now, to get the high-probability bound, we apply Theorem 12 to $X_t = Y_t/2 \leqslant 1$ a.s. and $V_t = W_t/4$ and use the bounds (16) and (18). We find that, with probability at least $1 - \delta$,

$$\alpha S_T \leqslant R_{k^\star,T} + 3\sqrt{(4 + S_T)\ln(\gamma/\delta)} + 2\ln(\gamma/\delta) \leqslant R_{k^\star,T} + 3\sqrt{S_T \ln(\gamma/\delta)} + 8\ln(\gamma/\delta)$$

where $\gamma \leqslant 1 + (1/2\mathrm{e})\left[1 + \ln\bigl(1 + \mathbb{E}[S_T]/4\bigr)\right]$ and where we used $\sqrt{\ln(\gamma/\delta)} \geqslant 1$. Combining the bound (10) on the regret with (17) yields $R_{k^\star,T} \leqslant \Xi_1 \sqrt{S_T \ln K} + \Xi_2$, so that, still with probability at least $1 - \delta$,

$$\alpha S_T \leqslant \Bigl(\Xi_1 \sqrt{\ln K} + 3\sqrt{\ln(\gamma/\delta)}\Bigr)\sqrt{S_T} + \Bigl(8\ln(\gamma/\delta) + \Xi_2\Bigr)\,.$$

Solving for $\sqrt{S_T}$ with Lemma 10 and using that $\alpha \leqslant 1$, this implies

$$\sqrt{S_T} \leqslant \frac{\Xi_1\sqrt{\ln K} + 3\sqrt{\ln(\gamma/\delta)}}{\alpha} + \frac{1}{\sqrt{\alpha}}\sqrt{8\ln(\gamma/\delta) + \Xi_2} \leqslant \frac{\Xi_1 \sqrt{\ln K}}{\alpha} + \sqrt{\frac{\Xi_2}{\alpha}} + \frac{6}{\alpha}\sqrt{\ln\frac{\gamma}{\delta}}\,.$$

Substitution into the (deterministic) regret bound $R_{k^\star,T} \leqslant \Xi_1\sqrt{S_T \ln K} + \Xi_2$ concludes the proof. ■

## Acknowledgments

Van Erven was supported by NWO Rubicon grant 680-50-1112.

## References

Dmitry Adamskiy, Wouter M. Koolen, Alexey Chernov, and Vladimir Vovk. A closer look at adaptive regret. In *Algorithmic Learning Theory (ALT)*, pages 290–304, 2012.

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.

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

Avrim Blum. Empirical support for winnow and weighted-majority algorithms: Results on a calendar scheduling domain. *Machine Learning*, 26:5–23, 1997.

Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

N. Cesa-Bianchi and G. Lugosi. Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 3(51):239–261, 2003.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Nicolò Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. Minimizing regret with label efficient prediction. *IEEE Trans. Inform. Theory*, 51:77–92, 2005.

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.

Alexey V. Chernov and Vladimir Vovk. Prediction with advice of unknown number of experts. *CoRR*, abs/1006.0475, 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 (COLT)*, pages 6.1–6.20, 2012.

Steven de Rooij, Tim van Erven, Peter D. Grünwald, and Wouter M. Koolen. Follow the leader if you can, hedge if you must. Available at arXiv:1301.0534 [cs.LG], 2013.

Marie Devaine, Pierre Gaillard, Yannig Goude, and Gilles Stoltz. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning*, 90(2):231–260, 2013.

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. 12(Jul):2121–2159, 2011.

David Freedman. On tail probabilities for martingales. *Annals of Probability*, 3:100–118, 1975.

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.

Yoav Freund, Robert E. Schapire, Yoram Singer, and Manfred K. Warmuth. Using and combining predictors that specialize. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 334–343, 1997.

Pierre Gaillard, Gilles Stoltz, and Tim van Erven. A second-order bound with excess losses. Extended version available at arXiv:1301.0534 [cs.LG], 2014.

László Györfi and György Ottucsák. Sequential prediction of unbounded stationary time series. *IEEE Transactions on Information Theory*, 53(5):1866–1872, 2007.

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

Wouter M. Koolen, Dmitry Adamskiy, and Manfred K. Warmuth. Putting Bayes to sleep. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 135–143, 2013.

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

Tim van Erven, Peter Grünwald, Wouter M. Koolen, and Steven de Rooij. Adaptive Hedge. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2011.

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

Olivier Wintenberger. Optimal learning with bernstein online aggregation. Extended version available at arXiv:1404.1356 [stat.ML], 2014.

## Appendix A.

We gather in this appendix several facts and results whose proofs were omitted from the main body of the paper.

### A.1. Proof of Lemma 10

Solving $x^2 \leqslant a + cx$ for $x$, we find that

$$\frac{1}{2}c - \frac{1}{2}\sqrt{c^2+4a} \leqslant x \leqslant \frac{1}{2}c + \frac{1}{2}\sqrt{c^2+4a}\,.$$

In particular, focusing on the upper bound, we get $2x \leqslant c+\sqrt{c^2+4a} \leqslant c+\sqrt{c^2}+\sqrt{4a} = 2c+2\sqrt{a}$, which was to be shown.

### A.2. Proof of Theorem 3

The proof will rely on the following simple lemma.

**Lemma 13** *For all $x > 0$ and all $\alpha \geqslant 1$, we have $x \leqslant x^\alpha + (\alpha-1)/\mathrm{e}$.*

**Proof** The inequality is straightforward when $x \geqslant 1$, so we restrict our attention to the case where $x < 1$. The function $\alpha \mapsto x^\alpha = \mathrm{e}^{\alpha \ln x}$ is convex and thus is above any tangent line. In particular, considering the value $x \ln x$ of the derivative function $\alpha \mapsto (\ln x)\,\mathrm{e}^{\alpha \ln x}$ at $\alpha = 1$, we get

$$\forall\, \alpha > 0, \qquad x^\alpha - x \geqslant (x \ln x)\,(\alpha - 1)\,.$$

Now, since we only consider $\alpha \geqslant 1$, it suffices to lower bound $x \ln x$ for the values of interest for $x$, namely, the ones in $(0,1)$ as indicated at the beginning of the proof. On this interval, the stated quantity is at least $-1/\mathrm{e}$, which concludes the proof. ■

We now prove Theorem 3.

**Proof [of Theorem 3]** As in the proof of Theorem 1, we bound $\ln W_T$ from below and from above. For the lower bound, we start with $\ln W_T \geqslant \ln w_{k,T}$. We then show by induction that for all $t \geqslant 0$,

$$\ln w_{k,t} \geqslant \eta_{k,t} \sum_{s=1}^{t} \Bigl( r_{k,s} - \eta_{k,s-1} r_{k,s}^2 \Bigr) + \frac{\eta_{k,t}}{\eta_{k,0}} \ln w_{k,0}\,,$$

where $r_{k,s} = \widehat{\ell}_s - \ell_{k,s}$ denotes the instantaneous regret with respect to expert $k$. The inequality is trivial for $t = 0$. If it holds at a given round $t$, then by the weight update (step 3 of the algorithm),

$$\begin{aligned}
\ln w_{k,t+1} &= \frac{\eta_{k,t+1}}{\eta_{k,t}} \biggl( \ln w_{k,t} + \ln\Bigl(1 + \eta_{k,t} r_{k,t+1}\Bigr) \biggr) \\
&\geqslant \frac{\eta_{k,t+1}}{\eta_{k,t}} \left( \frac{\eta_{k,t}}{\eta_{k,0}} \ln w_{k,0} + \eta_{k,t} \sum_{s=1}^{t} \Bigl( r_{k,s} - \eta_{k,s-1} r_{k,s}^2 \Bigr) \right) + \frac{\eta_{k,t+1}}{\eta_{k,t}} \bigl( \eta_{k,t} r_{k,t+1} - \eta_{k,t}^2 r_{k,t+1}^2 \bigr)
\end{aligned}$$

$$= \quad \eta_{k,t+1} \sum_{s=1}^{t+1} \Bigl( r_{k,s} - \eta_{k,s-1} r_{k,s}^2 \Bigr) + \frac{\eta_{k,t+1}}{\eta_{k,0}} \ln w_{k,0} \,,$$

where the inequality comes from the induction hypothesis and from the inequality $\ln(1+x) \geqslant x - x^2$ for all $x \geqslant -1/2$ already used in the proof of Theorem 1.

We now bound from above $\ln W_T$, or equivalently, $W_T$ itself. We show by induction that for all $t \geqslant 0$,

$$W_t \leqslant 1 + \frac{1}{\mathrm{e}} \sum_{k=1}^{K} \sum_{s=1}^{t} \left( \frac{\eta_{k,s-1}}{\eta_{k,s}} - 1 \right).$$

The inequality is trivial for $t = 0$. To show that if the property holds for some $t \geqslant 0$ it also holds for $t+1$, we prove that

$$W_{t+1} \leqslant W_t + \frac{1}{\mathrm{e}} \sum_{k=1}^{K} \left( \frac{\eta_{k,t}}{\eta_{k,t+1}} - 1 \right). \tag{20}$$

Indeed, since $x \leqslant x^\alpha + (\alpha - 1)/\mathrm{e}$ for all $x > 0$ and $\alpha \geqslant 1$ (see Lemma 13), we have, for each expert $k$,

$$w_{k,t+1} \leqslant \bigl( w_{k,t+1} \bigr)^{\frac{\eta_{k,t}}{\eta_{k,t+1}}} + \frac{1}{\mathrm{e}} \left( \frac{\eta_{k,t}}{\eta_{k,t+1}} - 1 \right); \tag{21}$$

we used here $x = w_{k,t+1}$ and $\alpha = \eta_{k,t}/\eta_{k,t+1}$, which is larger than 1 because of the assumption that the learning rates are nonincreasing in $t$ for each $k$. Now, by definition of the weight update (step 3 of the algorithm),

$$\sum_{k=1}^{K} \bigl( w_{k,t+1} \bigr)^{\frac{\eta_{k,t}}{\eta_{k,t+1}}} = \sum_{k=1}^{K} w_{k,t} \bigl( 1 + \eta_{k,t} r_{k,t+1} \bigr) = W_t \,,$$

where the second inequality follows from the same argument as in the last display of the proof of Theorem 1, by using that $\eta_{k,t} w_{k,t}$ is proportional to $p_{k,t+1}$. Summing (21) over $k$ thus yields (20) as desired.

Finally, combining the upper and lower bounds on $\ln W_T$ and rearranging leads to the inequality of Theorem 3. ■

### A.3. Proof of Corollary 4

The following lemma will be useful.

**Lemma 14** *Let $a_0 > 0$ and $a_1, \ldots, a_m \in [0,1]$ be real numbers and let $f : (0, +\infty) \to [0, +\infty)$ be a nonincreasing function. Then*

$$\sum_{i=1}^{m} a_i \, f\bigl( a_0 + \ldots + a_{i-1} \bigr) \leqslant f(a_0) + \int_{a_0}^{a_0 + a_1 + \ldots + a_m} f(u) \, \mathrm{d}u \,.$$

**Proof** Abbreviating $s_i = a_0 + \ldots + a_i$ for $i = 0, \ldots, m$, we find that

$$\begin{aligned}\sum_{i=1}^{m} a_i\, f(s_{i-1}) &= \sum_{i=1}^{m} a_i\, f(s_i) + \sum_{i=1}^{m} a_i\big(f(s_{i-1}) - f(s_i)\big) \\ &\leqslant \sum_{i=1}^{m} a_i\, f(s_i) + \sum_{i=1}^{m} \big(f(s_{i-1}) - f(s_i)\big) \leqslant \sum_{i=1}^{m} a_i\, f(s_i) + f(s_0),\end{aligned}$$

where the first inequality follows because $f(s_{i-1}) \geqslant f(s_i)$ and $a_i \leqslant 1$ for $i \geqslant 1$, while the second inequality stems from a telescoping argument together with the fact that $f(s_m) \geqslant 0$. Using that $f$ is nonincreasing together with $s_i - s_{i-1} = a_i$ for $i \geqslant 1$, we further have

$$a_i\, f(s_i) = \int_{s_{i-1}}^{s_i} f(s_i)\, \mathrm{d}y \leqslant \int_{s_{i-1}}^{s_i} f(y)\, \mathrm{d}y\,.$$

Substituting this bound in the above inequality completes the proof. ∎

We will be slightly more general and take

$$\eta_{k,t} = \min\left\{\frac{1}{2},\ \sqrt{\frac{\gamma_k}{1+\sum_{s=1}^{t} r_{k,s}^2}}\right\}$$

for some constant $\gamma_k > 0$ to be defined by the analysis.

Because of the choice of nonincreasing learning rates, the first inequality of Theorem 3 holds true, and the regret $R_{k,t}$ is upper-bounded by

$$\frac{1}{\eta_{k,0}} \ln\frac{1}{w_{k,0}} + \frac{1}{\eta_{k,T}} \ln\left(1 + \frac{1}{\mathrm{e}} \sum_{k'=1}^{K} \underbrace{\sum_{t=1}^{T} \left(\frac{\eta_{k',t-1}}{\eta_{k',t}} - 1\right)}_{\text{first term}}\right) + \underbrace{\sum_{t=1}^{T} \eta_{k,t-1} r_{k,t}^2}_{\text{second term}}\,. \tag{22}$$

For the first term in (22), we note that for each $k'$ and $t \geqslant 1$ one of three possibilities must hold, all depending on which of the inequalities in $\eta_{k',t} \leqslant \eta_{k',t-1} \leqslant 1/2$ are equalities or strict inequalities. More precisely, either $\eta_{k',t} = \eta_{k',t-1} = 1/2$; or

$$\sqrt{\frac{\gamma_{k'}}{1+\sum_{s=1}^{t} r_{k',s}^2}} = \eta_{k',t} < \eta_{k',t-1} = \frac{1}{2} \leqslant \sqrt{\frac{\gamma_{k'}}{1+\sum_{s=1}^{t-1} r_{k',s}^2}}\,;$$

or $\eta_{k',t} \leqslant \eta_{k',t-1} < 1/2$. In all cases, the ratios $\eta_{k',t-1}/\eta_{k',t} - 1$ can be bounded as follows:

$$\begin{aligned}\sum_{t=1}^{T}\left(\frac{\eta_{k',t-1}}{\eta_{k',t}} - 1\right) &\leqslant \sum_{t=1}^{T}\left(\sqrt{\frac{1+\sum_{s=1}^{t} r_{k',t}^2}{1+\sum_{s=1}^{t-1} r_{k',t}^2}} - 1\right) \\ &= \sum_{t=1}^{T}\left(\sqrt{1 + \frac{r_{k',t}^2}{1+\sum_{s=1}^{t-1} r_{k',s}^2}} - 1\right) \leqslant \frac{1}{2}\sum_{t=1}^{T} \frac{r_{k',t}^2}{1+\sum_{s=1}^{t-1} r_{k',s}^2}\,,\end{aligned} \tag{23}$$

where we used, for the second inequality, that $g(1+z) \leqslant g(1) + z\, g'(1)$ for $z \geqslant 0$ for any concave function $g$, in particular the square root. We apply Lemma 14 with $f(x) = 1/x$ to further bound the sum in (23), which gives

$$\sum_{t=1}^{T} \frac{r_{k',t}^2}{1+\sum_{s=1}^{t-1} r_{k',s}^2} \leqslant 1 + \ln\left(1 + \sum_{t=1}^{T} r_{k',t}^2\right) - \cancel{\ln(1)} \leqslant 1 + \ln(T+1)\,. \tag{24}$$

For the second term in (22), we write

$$\sum_{t=1}^{T} \eta_{k,t-1} r_{k,t}^2 \leqslant \sqrt{\gamma_k} \sum_{t=1}^{T} \frac{r_{k,t}^2}{\sqrt{1+\sum_{s=1}^{t-1} r_{k,s}^2}}\,.$$

We apply Lemma 14 again, with $f(x) = 1/\sqrt{x}$, and get

$$\sum_{t=1}^{T} \frac{r_{k,t}^2}{\sqrt{1+\sum_{s=1}^{t-1} r_{k,s}^2}} \leqslant \underbrace{1 - 2\sqrt{1}}_{\leqslant 0} + 2\sqrt{\left(1 + \sum_{t=1}^{T} r_{k,t}^2\right)}\,. \tag{25}$$

We may now get back to (22). Substituting the obtained bounds on its first and second terms, and using $\eta_{k,0} \geqslant \eta_{k,T}$, we find it is no greater than

$$\frac{1}{\eta_{k,T}}\left(\ln \frac{1}{w_{k,0}} + B_{K,T}\right) + 2\sqrt{\gamma_k \left(1 + \sum_{t=1}^{T} r_{k,t}^2\right)}\,, \tag{26}$$

where $B_{K,T} = \ln\Bigl(1 + \frac{K}{2\mathrm{e}}\bigl(1 + \ln(T+1)\bigr)\Bigr)$.

Now if $\sqrt{1+\sum_{t=1}^{T} r_{k,t}^2} > 2\sqrt{\gamma_k}$ then $\eta_{k,T} < 1/2$ and (26) is bounded by

$$\sqrt{1 + \sum_{t=1}^{T} r_{k,t}^2}\left(2\sqrt{\gamma_k} + \frac{\ln \frac{1}{w_{k,0}} + B_{K,T}}{\sqrt{\gamma_k}}\right).$$

Alternatively, if $\sqrt{1+\sum_{t=1}^{T} r_{k,t}^2} \leqslant 2\sqrt{\gamma_k}$, then $\eta_{k,T} = 1/2$ and (26) does not exceed

$$2 \ln \frac{1}{w_{k,0}} + 2B_{K,T} + 4\gamma_k.$$

In either case, (26) is smaller than the sum of the latter two bounds, from which the corollary follows upon taking $\gamma_k = \ln(1/w_{k,0}) = \ln K$. (Although the derivation also works for non-uniform initial weights, they cannot provide any significant gain, because $B_{K,T} = O(\ln K + \ln \ln T)$ already contains an $O(\ln K)$ term.)

### A.4. Proof of Theorem 5

The proof has a geometric flavor—the same as in the proof of the approachability theorem (Blackwell, 1956). With a diagonal matrix $D = \mathrm{diag}(d_1, \ldots, d_K)$, with positive on-diagonal elements $d_i$, we associate an inner product and a norm as follows:

$$\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^K, \qquad \langle \boldsymbol{x},\, \boldsymbol{y} \rangle_D = \boldsymbol{x}^\top D \boldsymbol{y} \qquad \text{and} \qquad \|\boldsymbol{x}\|_D = \sqrt{\boldsymbol{x}^\top D \boldsymbol{x}}\,.$$

We denote by $\pi_D$ the projection on $\mathbb{R}^K_-$ under the norm $\|\cdot\|_D$. It turns out that this projection is independent of the considered matrix $D$ satisfying the constraints described above: it equals

$$\forall \boldsymbol{x} \in \mathbb{R}^K, \qquad \pi_D(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{x}_+\,,$$

where we recall that $\boldsymbol{x}_+$ denotes the vector whose components are the nonnegative parts of the components of $\boldsymbol{x}$. This entails that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^K$

$$\|(\boldsymbol{x}+\boldsymbol{y})_+\|_D^2 = \|\boldsymbol{x}+\boldsymbol{y} - \pi_D(\boldsymbol{x}+\boldsymbol{y})\|_D^2 \leqslant \|\boldsymbol{x}+\boldsymbol{y} - \pi_D(\boldsymbol{x})\|_D^2 = \|\boldsymbol{x}_+ + \boldsymbol{y}\|_D^2\,. \tag{27}$$

Now, we consider, for each instance $t \geqslant 1$, the diagonal matrix $D_t = \mathrm{diag}(\eta_{1,t}, \ldots, \eta_{K,t})$, with positive elements on the diagonal. As all sequences $(\eta_{k,t})_{t\geqslant 0}$ are non-increasing for a fixed $k$, we have, for all $t \geqslant 1$, that

$$\forall \boldsymbol{x} \in \mathbb{R}^K, \qquad \|\boldsymbol{x}\|_{D_t} \leqslant \|\boldsymbol{x}\|_{D_{t-1}}\,. \tag{28}$$

This entails that

$$\big\|(\boldsymbol{R}_t)_+\big\|_{D_t} \leqslant \big\|(\boldsymbol{R}_t)_+\big\|_{D_{t-1}} = \big\|(\boldsymbol{R}_{t-1} + \boldsymbol{r}_t)_+\big\|_{D_{t-1}} \leqslant \big\|(\boldsymbol{R}_{t-1})_+ + \boldsymbol{r}_t\big\|_{D_{t-1}}\,, \tag{29}$$

where we denoted by $\boldsymbol{r}_t$ the vector $(r_{k,t})_{1\leqslant k\leqslant K}$ of the instantaneous regrets and where we applied (27). Taking squares and developing the squared norm, we get

$$\big\|(\boldsymbol{R}_t)_+\big\|_{D_t}^2 \leqslant \big\|(\boldsymbol{R}_{t-1})_+\big\|_{D_{t-1}}^2 + \|\boldsymbol{r}_t\|_{D_{t-1}}^2 + 2\,\boldsymbol{r}_t^\top D_{t-1}\,(\boldsymbol{R}_{t-1})_+\,. \tag{30}$$

But the inner product equals

$$2\,\boldsymbol{r}_t^\top D_{t-1}\,(\boldsymbol{R}_{t-1})_+ = 2\sum_{k=1}^K \eta_{k,t-1}\,(R_{k,t-1})_+\,r_{k,t} = 2\,\boldsymbol{\eta}_{t-1}^\top\,(\boldsymbol{R}_{t-1})_+ \underbrace{\sum_{k=1}^K p_{k,t} r_{k,t}}_{=0} = 0\,,$$

where the last but one equality follows from step 1 of the algorithm.

Hence (30) entails $\big\|(\boldsymbol{R}_t)_+\big\|_{D_t}^2 - \big\|(\boldsymbol{R}_{t-1})_+\big\|_{D_{t-1}}^2 \leqslant \|\boldsymbol{r}_t\|_{D_{t-1}}^2$, which, summing over all rounds $t \geqslant 1$, leads to

$$\begin{aligned}\big\|(\boldsymbol{R}_T)_+\big\|_{D_T}^2 - \cancel{\big\|(\boldsymbol{R}_0)_+\big\|_{D_0}^2} &\leqslant \sum_{t=1}^T \|\boldsymbol{r}_t\|_{D_{t-1}}^2 = \sum_{t=1}^T\sum_{k=1}^K \eta_{k,t-1} r_{k,t}^2 \\ &= \sum_{k=1}^K\sum_{t=1}^T \frac{r_{k,t}^2}{1+\sum_{s=1}^{t-1} r_{k,s}^2} \leqslant K\big(1+\ln(1+T)\big)\,,\end{aligned} \tag{31}$$

where the last equality follows from substituting the value of $\eta_{k,t-1}$ and the last inequality was proved in (24). Finally, (31) implies that, for any expert $k = 1, \ldots, K$,

$$\eta_{k,T} \left(R_{k,T}\right)_+^2 \leqslant \big\|(\boldsymbol{R}_T)_+\big\|_{D_T}^2 \leqslant K\big(1 + \ln(1+T)\big) \,,$$

so that

$$R_{k,T} \leqslant \sqrt{K\big(1 + \ln(1+T)\big)\, \eta_{k,T}^{-1}} \,.$$

The proof is concluded by substituting the value of $\eta_{k,T}$.