# Self-Normalized Martingales and Uniform Regret Bounds for Linear Regression

**Fan Chen** FANCHEN@MIT.EDU
*Massachusetts Institute of Technology*

**Jian Qian** JIANQIAN@HKU.HK
*The University of Hong Kong*

**Alexander Rakhlin** RAKHLIN@MIT.EDU
*Massachusetts Institute of Technology*

**Nikita Zhivotovskiy** ZHIVOTOVSKIY@BERKELEY.EDU
*University of California, Berkeley*

**Editors:** Steve Hanneke and Tor Lattimore

## Abstract

Self-normalized martingale inequalities lie at the heart of confidence ellipsoids for online least squares and, more broadly, many bandit and reinforcement-learning results. Yet existing vector and scalar results typically rely on bounded covariates and an explicit regularization matrix, producing bounds that are *not scale-invariant*: although the self-normalized quantity is scale-invariant by definition, its standard upper bounds are not.

We characterize when scale-invariant upper bounds on self-normalized martingales are possible. Without further assumptions, we prove that nontrivial scale-invariant bounds exist only in dimension $d = 1$; moreover, in $d = 1$ we obtain $O(\log T)$ scale-invariant self-normalized bounds without any assumptions on the covariates. In contrast, for $d > 1$ we show that no nontrivial scale-invariant bound can hold in full generality. We then connect this dichotomy to *doubly-uniform* regret in online linear regression (i.e., regret bounds that are simultaneously independent of the covariate scale and the comparator norm) and use it to resolve the open question of Gaillard, Gerchinovitz, Huard, and Stoltz, *"Uniform regret bounds over $\mathbb{R}^d$ for the sequential linear regression problem with the square loss"* (ALT 2019): in $d = 1$ we give an explicit algorithm with $O(\log T)$ doubly-uniform regret, whereas for $d > 1$ sublinear doubly-uniform regret is impossible.

Finally, under a natural *smoothness* condition (bounded Radon–Nikodym derivatives of the conditional covariate laws with respect to a fixed base measure), we recover sublinear regret for $d > 1$ without bounded covariates and derive a self-normalized concentration inequality free of the usual regularization penalties, yielding arguably a first natural scale-invariant bound for adaptive, non-i.i.d. vector martingales.

## 1. Introduction

Our play opens with what appears to be a two-protagonist drama. They make their separate entrances across two acts, each insisting on a solo turn in the spotlight. In the final act they meet, only to realize that they are, in fact, a mirror image of each other.

**Act I: Self-normalized martingales.** In an introductory course on Probability, we learn that the sum of normal random variables is also normal, and, in particular, $\frac{\sum_{i=1}^T X_i}{\sqrt{\sum_{i=1}^T \sigma_i^2}} \sim \mathsf{N}(0,1)$ where $X_i \sim \mathsf{N}\big(0,\sigma_i^2\big)$ are independent normal random variables and $\sigma_i$ are constants. If one replaces the denominator by the random realizations of $X_i$, then the distribution of $\frac{\sum_{i=1}^T X_i}{\sqrt{\sum_{i=1}^T X_i^2}}$ is no longer normal, yet its tails are similar to those of the normal distribution. In particular, an application of Hoeffding's inequality (via symmetrization) shows that

$$\mathbb{P}\left(\frac{\sum_{i=1}^T X_i}{\sqrt{\sum_{i=1}^T X_i^2}} > u\right) \leqslant \exp\left(-\frac{u^2}{2}\right)$$

for any *symmetric* independent random variables $X_i$, remarkably under no further assumptions on their distributions (see e.g., van Handel (2014, Ex. 7.3)). To bring out the symmetry requirement, we may instead write the above inequality as

$$\mathbb{P}\left(\frac{\sum_{i=1}^T \varepsilon_i X_i}{\sqrt{\sum_{i=1}^T X_i^2}} > u\right) \leqslant \exp\left(-\frac{u^2}{2}\right), \tag{1}$$

where $\varepsilon_i$ are i.i.d. Rademacher random variables, and $X_i$ are independent with no further assumption on their distributions. Such inequalities for *self-normalized sums* are very attractive, both due to the lack of assumptions and due to their natural scale invariance.

It is then reasonable to ask whether inequalities similar to Eq. (1) hold for martingales. More precisely, suppose for simplicity that $X_i$ are deterministic functions of $\varepsilon_1, \dots, \varepsilon_{i-1}$, i.e. measurable with respect to the dyadic filtration. Clearly, the sum $\sum_{i=1}^T \varepsilon_i X_i$ is a martingale. Does the inequality Eq. (1) also hold for this martingale without assumptions on $X_i$'s?

Perhaps surprisingly, the answer is no. Even more interestingly, this answer is related to what is referred to as "doubly uniform regret" in online learning. But we are getting ahead of ourselves.

For simplicity of exposition, let us consider the first moment of the ratio, rather than the tail bound. The following inequality can be found in the classical book of de la Pena et al. (2009, p. 199):

$$\mathbb{E}\left[S_T/V_T^{1/2}\right] \leqslant C + c\,\mathbb{E}\Big[0 \vee \log\log(V_T^{1/2} \vee V_T^{-1/2})\Big], \tag{2}$$

where, henceforth, we abbreviate $S_T = \sum_{i=1}^T \varepsilon_i X_i$ and $V_T = \sum_{i=1}^T X_i^2$, and $C, c > 0$ are constants. In the multi-dimensional setting, described below, the upper bound also involves the logarithm of the condition number of the matrix $V_T$ (see additionally Whitehouse et al. (2023, Corollary 4.5) and references therein).

The inequality Eq. (2) lacks the scale-invariance property that we desire: the left-hand side does not change when multiplying all $X_i$'s by a constant, yet the right-hand side does. This lack of scale-invariance is present in many results in the literature on self-normalized martingales, and it stems from the pseudo-maximization (or the method of mixtures) technique pioneered by Robbins and Siegmund (1970) and used extensively by de la Peña et al. (2004); de la Pena et al. (2009). The

method aims to place a non-trivial mass on a parameter that can only be known after observing the scale of the realization.

One approach to remove scale-dependence in the upper bound on the self-normalized martingale is to change the denominator. This can be achieved by augmenting $V_T$ with a regularization term. For instance, de la Peña et al. (2004) establishes

$$\mathbb{P}\left(\frac{S_T}{(V_T + \mathbb{E}\, V_T)^{1/2}} > u\right) \leqslant \sqrt{2}\exp\left(-\frac{u^2}{4}\right). \tag{3}$$

The expected value $\mathbb{E}\, V_T$ can be viewed as fixing the scale of the problem, yet its presence in the denominator is not desirable. Another approach is to augment $V_T$ with a constant, making both the ratio and the ensuing upper bound scale-dependent.

In particular, the addition of a regularizing constant to $V_T$ has been employed in the multi-dimensional setting (that is, $X_i = X_i(\varepsilon_1, \ldots, \varepsilon_{i-1})$ are taking values in $\mathbb{R}^d$ and $V_T = \sum_{i=1}^T X_i X_i^\top$ is the sample covariance) by de la Pena et al. (2009, Theorem 14.7) and Abbasi-Yadkori et al. (2011) to establish tail bounds of the form

$$S_T^\top (V_T + \Gamma)^{-1} S_T \lesssim \log\left(\frac{\det(V_T + \Gamma)}{\det(\Gamma)}\right) + \log(1/\delta) \tag{4}$$

with probability at least $1 - \delta$, for some deterministic positive definite matrix $\Gamma$. Once again, both sides are not scale-invariant, limiting the applicability of the bound when the scale is unknown.

Before continuing our discussion, we mention that the analysis of self-normalized martingales plays a central role in bandits and reinforcement learning: it underlies confidence ellipsoids for online least-squares estimators (e.g., Dani et al. (2008); Abbasi-Yadkori et al. (2011)) and the resulting online-to-confidence set conversions (see Abbasi-Yadkori et al. (2012); Lee et al. (2024); Clerico et al. (2025) and the textbook (Lattimore and Szepesvári, 2020, Chapter 20) for bibliographic pointers), identification in Linear Time-Invariant systems (Simchowitz et al., 2018; Sarkar and Rakhlin, 2019), and beyond. More broadly, there is renewed interest in formulations and in weakening tail assumptions beyond the classical conditionally sub-Gaussian setting (e.g., Howard et al. (2020); Whitehouse et al. (2023); Zhao et al. (2023); Ziemann (2025); Akhavan et al. (2025)).

**Act II: Doubly uniform regret.** One of the first online prediction methods is the celebrated Vovk-Azoury-Warmuth (VAW) estimator, initially proposed by Vovk (1997) and later refined by Vovk (2001); Azoury and Warmuth (2001) (see also (Cesa-Bianchi and Lugosi, 2006, Theorem 11.9)). First, we recall that in online supervised learning with squared loss, on each round $t \in [T]$, the forecaster observes $x_t \in \mathbb{R}^d$, selects a prediction $\widehat{y}_t \in \mathbb{R}$ and observes $y_t \in [-1, 1]$. The VAW estimator, defined later in the text, is essentially a regularized least squares estimator with a regularization parameter $\lambda > 0$, and it achieves the following regret bound: for any $\theta \in \mathbb{R}^d$,

$$\sum_{t=1}^T (\widehat{y}_t - y_t)^2 - \sum_{t=1}^T (\langle \theta, x_t \rangle - y_t)^2 \leqslant \lambda \|\theta\|^2 + d \log\left(1 + \frac{T \max_t \|x_t\|^2}{\lambda}\right). \tag{5}$$

Notably, the bound is non-uniform in two ways: it depends both on the norm of the comparator vector $\theta$ and the scale of the covariates. Bartlett et al. (2015) raised the question of whether one can obtain a regret bound that is uniform over all $\theta$, as the existing lower bounds do not show this

necessity (see, e.g., (Gaillard et al., 2019, Theorem 4)). This led the authors of Gaillard et al. (2019) to further ask whether regret bounds that are doubly uniform—with respect to the norm of the target parameter and the covariates—are possible in online linear regression.

So far, this double uniformity is only known in the so-called transductive online setup, where all design vectors are arbitrary but known in advance so that the predictor can use them (Bartlett et al., 2015; Gaillard et al., 2019; Qian et al., 2026), and, roughly speaking, establish the scale of the prediction problem. This double uniformity also appears in the statistical (i.i.d.) setup, where variants of non-linear predictors allow one to bypass the dependence on both the distribution of the design and the norm of the parameter (Forster and Warmuth, 2002; Mourtada et al., 2022). More generally, in the context of GLMs there is recent interest in analyzing unbounded parameter spaces, for example in classification with logistic regression, where large parameter norms are very natural and relate to (almost) linearly separable samples. In the transductive setup and in the context of online logistic regression, see Drmota et al. (2026), while Qian et al. (2026) focuses on regression with square, hinge and logarithmic losses.

**Act III: On the equivalence of martingale bounds and regret inequalities.** Denote the regret of the learner in the online prediction problem with arbitrary $\theta \in \mathbb{R}^d$ as

$$\mathbf{Reg}(T) := \sum_{t=1}^{T} (\widehat{y}_t - y_t)^2 - \inf_{\theta \in \mathbb{R}^d} \sum_{t=1}^{T} (\langle \theta, x_t \rangle - y_t)^2. \tag{6}$$

Suppose the forecaster attempts to predict the following sequence. Covariates form a predictable process $x_t = X_t(\varepsilon_1, \dots, \varepsilon_{t-1})$, as earlier in the text, and the outcome variable $y_t = \varepsilon_t$ is an independent Rademacher random variable. It is clear that in this setting, the best strategy for the forecaster is to predict $\widehat{y}_t = 0$ for all $t$. Then, the expected regret of the forecaster is given by

$$\mathbb{E}_\varepsilon \left[ \sum_{t=1}^{T} y_t^2 - \inf_{\theta \in \mathbb{R}^d} \sum_{t=1}^{T} (\langle \theta, x_t \rangle - y_t)^2 \right] = \mathbb{E}_\varepsilon \sup_{\theta \in \mathbb{R}^d} 2\langle \theta, \sum_{t=1}^{T} \varepsilon_t X_t \rangle - \langle \theta, \sum_{t=1}^{T} X_t X_t^\top \theta \rangle \tag{7}$$

$$= \mathbb{E}_\varepsilon \big[ S_T^\top (V_T)^\dagger S_T \big]. \tag{8}$$

which is precisely the expected value of the self-normalized process that appeared in Act I. Furthermore, Rakhlin and Sridharan (2014) proved a converse statement (for a more general setting of regression with any class of functions): no matter what the sequence of $\{(x_t, y_t)\}_{t \in [T]}$ is, even if chosen adaptively by Nature, there *exists* a prediction strategy that achieves a regret bound that is, up to a multiplicative constant, in the above self-normalized form for the worst-case martingale (see below for more details). Thus, upper bounds on Eq. (6) for all sequences imply upper bounds for the expected self-normalized ratio, and vice versa.

The connection between probabilistic martingale inequalities and regret bounds has been a focus of extensive research, including Abernethy et al. (2008); Rakhlin et al. (2010); Rakhlin and Sridharan (2017); Foster et al. (2018, 2017); Beiglböck and Siorpaes (2015); Orabona and Jun (2023), and, in particular, certain *equivalence* between these two seemingly unrelated fields was studied in Rakhlin and Sridharan (2017); Foster et al. (2018).

**Finale.** Due to the two-sided equivalence between minimax regret bounds for unbounded comparators $\theta \in \mathbb{R}^d$ and expected value of self-normalized martingales, we can establish lower/upper

bounds for one by studying the other, whichever is more convenient. In particular, the issues discussed in Act I regarding the knowledge of the scale of the problem are precisely the issues discussed in Act II regarding the knowledge of the norm of the comparator vector $\theta$ and the scale of the covariates. In particular, later in the paper, we describe the exact link between self-normalized bounds of the form Eq. (4) and the Vovk-Azoury-Warmuth forecaster.

In particular, our contributions are:

- We establish a sharp separation between the cases $d = 1$ and $d > 1$. When $d = 1$, we prove a *fully scale-invariant* bound of order $O(\log T)$ for self-normalized martingales *without any assumption on the covariates*. Via the regret–martingale connection developed in Rakhlin and Sridharan (2017), this implies a doubly-uniform $O(\log T)$ regret guarantee for online linear regression, thereby resolving the question of Gaillard et al. (2019) in dimension one. Moreover, we provide an explicit algorithm achieving this doubly-uniform $O(\log T)$ regret.

- In contrast, when $d > 1$ we show that no nontrivial scale-invariant control of self-normalized vector martingales is possible in full generality, and consequently sublinear doubly-uniform regret bounds for online linear regression cannot hold. This completes our answer to the question of Gaillard et al. (2019).

- On the positive side, still in the regime $d > 1$, we introduce a *smoothness* condition on the covariate process, requiring that each conditional law admits a bounded Radon–Nikodym derivative with respect to a fixed base measure. Under this assumption we obtain sublinear regret *without* assuming bounded covariates. Moreover, our bounds avoid the usual matrix regularization penalties (e.g., the log-determinant term in Eq. (4)), yielding what appears to be a first natural example of a scale-invariant self-normalized martingale bound in a genuinely non-i.i.d. setting.

## 2. Preliminaries

In the previous section, we motivated the study of dyadic self-normalized martingales of the form $\varepsilon_t X_t$, where $(\varepsilon_t)_{t\geqslant 1}$ are i.i.d. Rademacher signs and each $X_t$ is $\sigma(\varepsilon_1, \dots, \varepsilon_{t-1})$-measurable. In particular, if $(X_t)$ is deterministic (or more generally independent of $(\varepsilon_t)$), then conditioning on $(X_t)_{t\leqslant T}$ and applying Hoeffding's inequality yields the scale-invariant tail bound Eq. (1) for the ratio $\sum_{t=1}^T \varepsilon_t X_t / \sqrt{\sum_{t=1}^T X_t^2}$. Let us now present the more general filtered definition that is standard in the online regression and bandit literature (e.g., Abbasi-Yadkori et al., 2011; Ziemann, 2025), and that will serve as our main probabilistic object. Throughout this probabilistic discussion we use capitals $(X_t, Y_t)$; later, when we switch to the online learning protocol, we will revert to the conventional lowercase notation $(x_t, y_t)$ and use a separate notion of game history that also records predictions.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a filtration $(\mathcal{G}_t)_{t\geqslant 0}$. We assume that $X_t$ is *predictable* and that $(Y_t)_{t=1}^T$ is a real-valued martingale difference sequence with respect to $(\mathcal{G}_t)$:

$$X_t \in \mathbb{R}^d \text{ is } \mathcal{G}_{t-1}\text{-measurable and} \qquad \mathbb{E}\left[Y_t \,|\mathcal{G}_{t-1}\right] = 0 \qquad \text{for all } t \in [T]. \tag{9}$$

Depending on the application, one may further assume boundedness $|Y_t| \leqslant 1$ almost surely or a conditional sub-Gaussian condition. Define the cumulative vector and the Gram matrix

$$S_t := \sum_{i=1}^{t} Y_i X_i \in \mathbb{R}^d, \qquad V_t := \sum_{i=1}^{t} X_i X_i^\top \in \mathbb{R}^{d\times d}.$$

The canonical scale-free quantity is the *self-normalized* process

$$R_t := \|S_t\|^2_{V_t^\dagger} = S_t^\top V_t^\dagger S_t,$$

where $V_t^\dagger$ denotes the Moore–Penrose pseudoinverse. Controlling $R_t$ (in expectation or with high probability) is a central theme of the self-normalization literature (de la Pena, 1999; de la Pena et al., 2009; Bercu and Touati, 2019) and, in the online learning context, it is the quantity that underlies confidence ellipsoids and regret bounds in least squares and linear bandits (Abbasi-Yadkori et al., 2011; Ziemann, 2025).

The question we pursue is whether martingale analogues of Eq. (1) can hold for $R_t$ under minimal assumptions on the predictable covariates $(X_t)$. Since we aim for a uniform upper bound that holds for all martingales of the form (9), we define

$$\mathcal{R}_d(T) = \sup_{P_{X,Y}} \mathbb{E}_{P_{X,Y}}\left[\|S_T\|^2_{V_T^\dagger}\right], \tag{10}$$

where the supremum ranges over all laws $P_{X,Y}$ of dimension $d$ satisfying (9) and such that $|Y_t| \leqslant 1$ almost surely for all $t \in [T]$.

A *dyadic martingale* is a special case where $Y_t = \varepsilon_t$ are i.i.d. Rademacher and $X_t = X_t(\varepsilon_1, \dots, \varepsilon_{t-1})$ is a deterministic function of the past signs. Such a process can be viewed as an $\mathbb{R}^d$-valued *tree* $X$ of depth $T$, or a sequence of mappings $X_t \colon \{\pm 1\}^{t-1} \to \mathbb{R}^d$. Let

$$\mathcal{R}_d^{\mathrm{dyadic}}(T) := \sup_{X} \mathbb{E}_\varepsilon[R_T],$$

where the supremum is over all trees $X$ and expectation is over i.i.d. Rademacher $(\varepsilon_t)$.

**Lemma 1** *For any $d \geqslant 1$ and $T \geqslant 1$, if we only consider processes where $|Y_t| \leqslant 1$ a.s., then*

$$\mathcal{R}_d(T) = \mathcal{R}_d^{\mathrm{dyadic}}(T).$$

In words, worst-case martingales—from the point of view of self-normalized ratios—are the dyadic martingales, up to a factor 2. Since each dyadic martingale is defined by $2^T - 1$ values (the number of nodes in the binary tree), for each such dyadic martingale we can consider its rescaled (by the maximum norm) variant $X'$ with $\|X'_t\| \leqslant 1$. Since the value of the self-normalized ratio does not change when scaled by a constant, we also have the following conclusion. Let $\mathcal{R}_d^{\mathrm{bdd}}(T)$ denote the supremum over the processes of the form (9) with $\|X_t\| \leqslant 1$ almost surely, and let $\mathcal{R}_d^{\mathrm{bdd,dyadic}}(T)$ denote the corresponding supremum restricted to dyadic and bounded martingales.

**Corollary 2** *For any $d \geqslant 1$ and $T \geqslant 1$, if we only consider processes where $|Y_t| \leqslant 1$ a.s., then $\mathcal{R}_d^{\mathrm{dyadic}}(T) = \mathcal{R}_d^{\mathrm{bdd,dyadic}}(T)$, and thus*

$$\mathcal{R}_d^{\mathrm{dyadic}}(T) = \mathcal{R}_d^{\mathrm{bdd,dyadic}}(T) = \mathcal{R}_d^{\mathrm{bdd}}(T) = \mathcal{R}_d(T).$$

In the following section, this result will imply that the difficulty in doubly-uniform regret bounds is a consequence of unbounded $\theta$ rather than unbounded covariates. This is also reflected by our lower bounds, which hold for bounded covariates.

### 2.1. An online prediction game

We now use the standard online learning notation and write covariates, outcomes, and predictions as $(x_t, y_t, \widehat{y}_t)$. On each round $t \in [T]$, the environment Env reveals a covariate vector $x_t \in \mathbb{R}^d$, the learner Alg outputs a prediction $\widehat{y}_t \in \mathbb{R}$, and then Env reveals an outcome $y_t \in [-1, 1]$. Both Env and Alg may be adaptive. We denote the history prior to round $t$ by

$$\mathcal{H}^{t-1} := \sigma\Big(\{(x_s, \widehat{y}_s, y_s)\}_{s<t}\Big).$$

Given a comparator set $\Theta \subseteq \mathbb{R}^d$, the square-loss regret is

$$\mathbf{Reg}_\Theta(T) := \sum_{t=1}^{T} (\widehat{y}_t - y_t)^2 - \inf_{\theta\in\Theta} \sum_{t=1}^{T} \big(\langle \theta, x_t\rangle - y_t\big)^2. \tag{11}$$

Since our focus is on $\Theta = \mathbb{R}^d$, we abbreviate $\mathbf{Reg}(T) := \mathbf{Reg}_{\mathbb{R}^d}(T)$.

A convenient way to relate regret to self-normalization is to consider a stochastic environment with conditionally unbiased outcomes. Fix any algorithm Alg and a sequential law $P_{x,y}$ over $(x_1, y_1, \ldots, x_T, y_T)$ such that, when the environment is generated from $P_{x,y}$ independently of the predictions of Alg, the outcomes satisfy

$$\mathbb{E}\left[y_t \,|\, x_1, y_1, \ldots, x_{t-1}, y_{t-1}, x_t\right] = 0 \qquad \text{for all } t \in [T]. \tag{12}$$

Let Env be the environment that samples $(x_1, y_1, \ldots, x_T, y_T) \sim P_{x,y}$ and reveals it round by round. Then

$$\begin{aligned}\mathbb{E}^{\mathsf{Env},\mathsf{Alg}}\left[\sum_{t=1}^{T} (\widehat{y}_t - y_t)^2\right] &= \mathbb{E}^{\mathsf{Env},\mathsf{Alg}}\left[\sum_{t=1}^{T} \big(\widehat{y}_t^2 - 2\widehat{y}_t y_t + y_t^2\big)\right] \\ &= \mathbb{E}^{\mathsf{Env},\mathsf{Alg}}\left[\sum_{t=1}^{T} \big(\widehat{y}_t^2 + y_t^2\big)\right] \geqslant \mathbb{E}_{P_{x,y}}\left[\sum_{t=1}^{T} y_t^2\right],\end{aligned} \tag{13}$$

where the middle equality uses Eq. (12) (hence $\mathbb{E}[\widehat{y}_t y_t] = 0$). Next, define $S_T := \sum_{t=1}^{T} y_t x_t \in \mathbb{R}^d$ and $V_T := \sum_{t=1}^{T} x_t x_t^\top \in \mathbb{R}^{d\times d}$. A direct completion of squares gives the exact identity

$$\sum_{t=1}^{T} y_t^2 - \inf_{\theta\in\mathbb{R}^d} \sum_{t=1}^{T} \big(\langle \theta, x_t\rangle - y_t\big)^2 = \sup_{\theta\in\mathbb{R}^d} \Big\{2\langle \theta, S_T\rangle - \|\theta\|_{V_T}^2\Big\} = \|S_T\|_{V_T^\dagger}^2. \tag{14}$$

Combining Eq. (13) with Eq. (14) yields

$$\mathbb{E}^{\mathsf{Env},\mathsf{Alg}}[\mathbf{Reg}(T)] \geqslant \mathbb{E}_{P_{x,y}}\left[\|S_T\|_{V_T^\dagger}^2\right]. \tag{15}$$

This implies that for every algorithm Alg,

$$\max_{\mathsf{Env}} \mathbb{E}^{\mathsf{Env},\mathsf{Alg}}[\mathbf{Reg}(T)] \geqslant \mathcal{R}_d(T).$$

Conversely, the work of Rakhlin and Sridharan (2017) provides a minimax upper bound turning self-normalized control into regret guarantees (up to universal constants), yielding a two-sided link between self-normalization and optimal regret in online linear regression. We state here the upper bound of (Rakhlin and Sridharan, 2014, Lemma 4) for the linear function class:

**Lemma 3** *In the notation above, it holds that*

$$\min_{\mathsf{Alg}} \max_{\mathsf{Env}} \mathbb{E}^{\mathsf{Env},\mathsf{Alg}}[\mathbf{Reg}(T)] \leqslant 4\mathcal{R}_{d+1}^{\mathrm{dyadic}}(T) \leqslant 4\mathcal{R}_{d+1}(T).$$

We remark that the actual upper bound in the proof is smaller than $\mathcal{R}_{d+1}(T)$; for our purposes, this is only important for $d = 1$, which we treat separately.

## 3. Self-normalized martingales and online prediction in the fully adversarial setting

In this section, we first discuss the one-dimensional case $d = 1$, and then the case $d \geqslant 2$. Remarkably, both regret and the self-normalized martingale exhibit very different behavior in these two regimes.

### 3.1. Dimension one: logarithmic self-normalized martingale and doubly-uniform regret

In this section, we focus on the one-dimensional case $d = 1$. We show that (i) the dyadic self-normalized martingale admits $O(\log T)$ control in expectation, and (ii) the minimax doubly-uniform regret in online linear regression is also $\Theta(\log T)$. Taken together, these results settle the behavior of both objects of interest in dimension one.

We start from the probabilistic side by establishing a linear bound on the exponential moment of the one-dimensional dyadic self-normalized martingale.

**Theorem 4** *For any dyadic martingale $X_1, \dots, X_T$ and any $c \in (0, 1/4]$,*

$$\mathbb{E}_\varepsilon\big[\exp(cR_T)\big] \leqslant T \exp\Big(\frac{c}{1-2c}\Big).$$

*Consequently,*

$$\mathbb{E}_\varepsilon[R_T] \leqslant \frac{1}{c}\log\left(T\exp\Big(\frac{c}{1-2c}\Big)\right) = \frac{\log T}{c} + \frac{1}{1-2c}.$$

Theorem 4 provides a homogeneous, scale-invariant control of the self-normalized martingale in dimension one. In particular, it improves upon the scale-sensitive behavior suggested by classical mixture-based bounds such as Eq. (2): the right-hand side grows only logarithmically with $T$ and requires no boundedness or moment assumptions on the predictable covariates $(X_t)$ beyond measurability with respect to the dyadic filtration.

In light of the regret–martingale connection discussed above, the logarithmic behavior in Theorem 4 suggests that $\log T$ is the correct scale for doubly-uniform regret in one dimension. We make this precise by giving a matching upper bound via an explicit procedure.

**Theorem 5** *Suppose that $d = 1$ and $|y_t| \leqslant m$ almost surely. Then there exists an algorithm (Algorithm 2) that achieves deterministically $\mathbf{Reg}(T) \lesssim m^2 \log T$.*

Complementarily, the minimax lower bound of order $\Omega(\log T)$ follows from Gaillard et al. (2019, Theorem 7) (adapted from Vovk (2001, Theorem 2)).

**Proposition 6** *In the setup of Theorem 5 with $T \geqslant 10$ and $m = 1$, there exists a dyadic martingale such that $\mathbb{E}[R_T] \gtrsim \log T$.*

Together, Theorem 5 and Proposition 6 yield the claimed $\Theta(\log T)$ characterization of doubly-uniform regret in dimension one, aligning with the logarithmic self-normalized control in Theorem 4.

### 3.2. Dimension two and higher: linear lower bounds

We now turn to $d \geqslant 2$. In sharp contrast to the one-dimensional case, the self-normalized *vector* martingale can grow linearly in the worst case. Through Eq. (15), this implies that doubly-uniform regret cannot be sublinear under a fully adversarial environment.

**Theorem 7** *Let $T \geqslant 1$ be any integer. When $d \geqslant 2$, for any $\varepsilon \in (0, 1)$, there exists a dyadic martingale $X_1, \dots, X_T$ such that*

$$\mathbb{E}\left[\|S_T\|^2_{V_T^\dagger}\right] \geqslant (1 - \varepsilon^2)T.$$

*In particular, by Eq. (15), there exists an environment such that for any algorithm,*

$$\mathbb{E}[\mathbf{Reg}(T)] \geqslant \mathbb{E}\left[\|S_T\|^2_{V_T^\dagger}\right] \geqslant (1 - \varepsilon^2)T.$$

**Proof sketch.** We construct an adaptive dyadic martingale that injects a constant amount of self-normalized energy at every step. Let $(\varepsilon_t)_{t\geqslant 1}$ be i.i.d. Rademacher variables, let $S_t = \sum_{i\leqslant t} \varepsilon_i X_i$, and define the regularized matrix

$$\widetilde{V}_t := \sum_{i\leqslant t} X_i X_i^\top + \lambda I,$$

for some fixed $\lambda > 0$. Using the Sherman–Morrison formula and conditioning on the past, one checks that

$$\mathbb{E}\left[\|S_{t+1}\|^2_{\widetilde{V}_{t+1}^{-1}} \mid \mathcal{F}_t\right] = \|S_t\|^2_{\widetilde{V}_t^{-1}} + \frac{\|X_{t+1}\|^2_{\widetilde{V}_t^{-1}} - \langle X_{t+1}, \widetilde{V}_t^{-1} S_t\rangle^2}{1 + \|X_{t+1}\|^2_{\widetilde{V}_t^{-1}}}.$$

We choose $X_{t+1} = r\,\widetilde{V}_t^{1/2} e_t$, where $e_t$ is any unit vector orthogonal to $\widetilde{V}_t^{-1/2} S_t$; this is always possible for $d \geqslant 2$. This choice kills the cross term and ensures $\|X_{t+1}\|^2_{\widetilde{V}_t^{-1}} = r^2$, so the conditional increment is the constant $r^2/(1 + r^2)$. Iterating yields $\mathbb{E}[\|S_T\|^2_{\widetilde{V}_T^{-1}}] = Tr^2/(1 + r^2)$. Moreover, $S_T \in \mathrm{range}(V_T)$, so $\|S_T\|^2_{V_T^\dagger} \geqslant \|S_T\|^2_{\widetilde{V}_T^{-1}}$, and hence $\mathbb{E}\left[\|S_T\|^2_{V_T^\dagger}\right] \geqslant \frac{Tr^2}{1+r^2}$. Taking $r^2 = (1-\varepsilon^2)/\varepsilon^2$ completes the lower bound (for any fixed $\lambda > 0$). ∎

Note that since the ratio is homogeneous, the construction in the proof can be rescaled so that $\|X_t\| \leqslant 1$ deterministically.

## 4. Online prediction with smooth environment

Even though the worst-case lower bound looks grim, the picture brightens if the environment is forced to "add randomness". We formalize this with a *smoothness* condition that caps how concentrated each $x_t$ can be relative to a fixed base measure $\mu$ (Block et al., 2022).

**Assumption 8 (Smoothness)** *There exists a probability measure $\mu$ on $\mathbb{R}^d$ and a parameter $C_{\mathsf{cov}} \geqslant 1$ such that for every round $t$, given the partial history $\mathcal{H}_{t-1}^x = \{x_1, \dots, x_{t-1}\}$, the conditional law of $x_t$ is $P_t(\cdot \mid \mathcal{H}_{t-1}^x)$ and satisfies*

$$\frac{\mathrm{d}P_t(\cdot \mid \mathcal{H}_{t-1}^x)}{\mathrm{d}\mu}(x) \leqslant C_{\mathsf{cov}} \qquad \text{for } \mu\text{-a.e. } x \in \mathbb{R}^d,$$

*for every history $\mathcal{H}_{t-1}^x$.* [1]

This assumption limits how concentrated the conditional law of $x_t$ can be; for instance, when $\mu$ is non-atomic it rules out choosing $x_t$ deterministically, as in the lower-bound construction of Theorem 7. The condition is trivially satisfied when $\mathcal{X}$ is finite (with $\mu = \mathrm{Unif}(\mathcal{X})$ and $C_{\mathsf{cov}} \leqslant |\mathcal{X}|$). When $\mathcal{X}$ is infinite, however, it can be a fairly strong assumption on the environment, and can make online prediction significantly easier.

To provide more intuition, Haghtalab et al. (2024) show that under Assumption 8, $x_1, \dots, x_T$ can be coupled with a subsequence of i.i.d. random vectors (Lemma 19), echoing the "smoothed analysis" philosophy in algorithms.

### 4.1. The Vovk–Azoury–Warmuth (VAW) algorithm under the smoothness assumption

For upper bounds under Assumption 8, we use the Vovk–Azoury–Warmuth (VAW) predictor, a classic forecaster for online linear regression. The regularized version is well known, but since both the comparator and the covariates may be unbounded here, standard analyses that rely on a fixed regularizer do not account for smoothness. We therefore study the *unregularized* VAW (Algorithm 1) and tailor the analysis to the smooth setting.

**Algorithm 1** Vovk–Azoury–Warmuth (VAW) predictor

**for** $t = 1, 2, \cdots, T$ **do**
  Set $\widehat{\theta}_t \in \arg\min_{\theta \in \mathbb{R}^d} \langle \theta, x_t \rangle^2 + \sum_{i<t} \big(\langle \theta, x_i \rangle - y_i\big)^2$.
  Predict $\widehat{y}_t = \langle \widehat{\theta}_t, x_t \rangle$.

We prove the following guarantee for VAW through an (almost) purely combinatorial argument, rather than the standard elliptical-potential analysis.

1. Here we assume smoothness with respect to the partial history $\mathcal{H}_{t-1}^x$. While it may be more natural to assume smoothness given the full history $\mathcal{H}_{t-1} = \{(x_s, y_s, \widehat{y}_s)\}_{s<t}$, requiring smoothness conditional on the partial history is weaker.

**Theorem 9** *Under Assumption 8, assuming $y_t \in [-1,1]$ for all $t \in [T]$, the VAW predictor in Algorithm 1 achieves the following regret for any $\delta \in (0,1)$, with probability at least $1-\delta$:*

$$\mathbf{Reg}(T) \lesssim \left(\sqrt{dC_{\mathsf{cov}} T \log(T/\delta)} + \log(1/\delta)\right).$$

We briefly indicate how the proof goes. It is standard to reduce the regret analysis of VAW-type algorithms to upper bounding the sum $\sum_{t=1}^{T} x_t^\top V_t^\dagger x_t$; in the usual analysis this leads to a dependence on the magnitudes of the $x_t$'s. In contrast, we use smoothed-analysis ideas. To control $\sum_{t=1}^{T} x_t^\top V_t^\dagger x_t$, we consider the longest subsequence $x_{t_1}, \ldots, x_{t_k}$ such that $\|x_{t_j}\|^2_{V_{t_j}^\dagger} \geqslant r$ for a fixed threshold $r > 0$, and then use the coupling result from Lemma 19 to reduce to an i.i.d. sequence. Compare this with the standard VAW bound Eq. (5), where a regularization parameter $\lambda$ is essential (one cannot take $\lambda \to 0$ without making the bound trivial), and the resulting regret bound depends explicitly on $\max_{t \leqslant T} \|x_t\|$.

## 5. Bounding self-normalized martingales via regret analysis: high probability results

We are now back in the setup of self-normalized martingales, and we re-derive a self-normalized concentration inequality by combining (i) a deterministic regret bound for an online regression algorithm and (ii) a stochastic exponential supermartingale coming from the conditional sub-Gaussian noise assumption. Such a reduction is standard and is a key tool in (Rakhlin and Sridharan, 2017).

Let $(\mathcal{F}_t)_{t \geqslant 0}$ be a filtration such that, for each round $t$, the covariate $X_t$ and the prediction $\hat{y}_t$ are $\mathcal{F}_{t-1}$-measurable, and $Y_t$ is then revealed. Assume that $(Y_t)_{t \geqslant 1}$ is a martingale difference sequence with conditionally sub-Gaussian increments: for some $\sigma > 0$,

$$\mathbb{E}\left[Y_t \mid \mathcal{F}_{t-1}\right] = 0 \qquad \text{and} \qquad \mathbb{E}\left[\exp(\alpha Y_t) \mid \mathcal{F}_{t-1}\right] \leqslant \exp\left(\frac{\alpha^2\sigma^2}{2}\right), \tag{16}$$

for all $\alpha \in \mathbb{R}$ and $t \geqslant 1$. Fix a deterministic $0 \prec \Gamma \in \mathbb{R}^{d \times d}$ and define $S_t := \sum_{i=1}^{t} X_i Y_i$ and $V_t := \sum_{i=1}^{t} X_i X_i^\top$. As above, completion of squares yields

$$\sum_{t=1}^{T} Y_t^2 - \inf_{\theta \in \mathbb{R}^d} \left\{ \sum_{t=1}^{T} \big(\langle \theta, X_t \rangle - Y_t\big)^2 + \theta^\top \Gamma \theta \right\} \;=\; \|S_T\|^2_{(V_T+\Gamma)^{-1}}, \tag{17}$$

which equivalently, for *any* sequence of predictions $(\hat{y}_t)_{t=1}^T$, can be rewritten as

$$\|S_T\|^2_{(V_T+\Gamma)^{-1}} = \mathbf{Reg}_T(\hat{y}) + \sum_{t=1}^{T} \big(2\hat{y}_t Y_t - \hat{y}_t^2\big), \tag{18}$$

where the last term is the one we will control stochastically using Eq. (16). The next simple result shows that this term admits a sharp high-probability bound.

**Lemma 10** *Under the sub-Gaussian assumption Eq. (16), for any predictable sequence $(\hat{y}_t)_{t=1}^T$, the process $M_t := \exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^{t}\big(2\hat{y}_i Y_i - \hat{y}_i^2\big)\right)$ is a nonnegative supermartingale with $\mathbb{E}[M_t] \leqslant 1$ for all $t$. In particular, with probability at least $1-\delta$,*

$$\sum_{t=1}^{T} \big(2\hat{y}_t Y_t - \hat{y}_t^2\big) \leqslant 2\sigma^2 \log(1/\delta),$$

*and the same bound extends to stopping times.*

Finally, we combine the high-probability regret bound under smoothness with the generic reduction of (18) to obtain a self-normalized concentration inequality without any explicit bound on $\|X_t\|$ and without introducing a positive definite regularization matrix $\Gamma$ in the self-normalization.

**Theorem 11 (Self-normalized martingales under smoothness)** *Let $(\mathcal{F}_t)_{t\geqslant 0}$ be a filtration such that for each $t \leqslant T$, the covariate $X_t$ is $\mathcal{F}_{t-1}$-measurable and $Y_t$ is then revealed. Assume that $(Y_t)_{t=1}^T$ is a martingale difference sequence satisfying Eq. (16) with parameter $\sigma$. Assume also that $(X_t)_{t=1}^T$ satisfies the smoothness condition of Assumption 8 with parameter $C_{\mathsf{cov}} \geqslant 1$. Define $S_T := \sum_{t=1}^T X_t Y_t$ and $V_T := \sum_{t=1}^T X_t X_t^\top$. Then for any $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\|S_T\|^2_{V_T^\dagger} \lesssim \sigma^2 \Big( \sqrt{d\, C_{\mathsf{cov}}\, T\, \log(2T/\delta)} + \log(2/\delta) \Big). \tag{19}$$

Importantly, compared to the canonical bound (Abbasi-Yadkori et al., 2011), namely under the sub-Gaussian assumption Eq. (16) but without smoothness, for any positive definite $\Gamma$,

$$\|S_T\|^2_{(V_T+\Gamma)^\dagger} \leqslant \sigma^2 \left( \log\left( \frac{\det(V_T+\Gamma)}{\det(\Gamma)} \right) + 2\log(1/\delta) \right),$$

the right-hand side of Eq. (19) has no explicit dependence on $\max_t \|X_t\|$ and no regularization matrix $\Gamma$ is needed in the self-normalized quantity. Moreover, the base measure $\mu$ from Assumption 8 is only used for the analysis: even for the regret analysis of Theorem 9 the algorithm itself does not require knowing $\mu$, and the regret bound is used purely as a tool to prove concentration.

## Acknowledgments

We acknowledge support from AFOSR through award FA9550-25-1-0375, Simons Foundation and the NSF through awards DMS-2031883 and PHY-2019786, and DARPA AIQ award.

## References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Artificial Intelligence and Statistics*, pages 1–9. PMLR, 2012.

Jacob Abernethy, Peter L Bartlett, Alexander Rakhlin, and Ambuj Tewari. Optimal strategies and minimax lower bounds for online convex games. 2008.

Arya Akhavan, Amitis Shidani, Alex Ayoub, and David Janz. Bernstein-type dimension-free concentration for self-normalised martingales. *arXiv preprint arXiv:2507.20982*, 2025.

Katy S. Azoury and Manfred K. Warmuth. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning*, 43(3):211–246, June 2001.

Peter L. Bartlett, Wouter M. Koolen, Alan Malek, Eiji Takimoto, and Manfred K. Warmuth. Minimax fixed-design linear regression. In Peter Grünwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 226–239, Paris, France, 03–06 Jul 2015. PMLR.

Mathias Beiglböck and Pietro Siorpaes. Pathwise versions of the Burkholder–Davis–Gundy inequality. *Bernoulli*, 21(1):360–373, 2015.

Bernard Bercu and Taieb Touati. New insights on concentration inequalities for self-normalized martingales. 2019.

Adam Block, Yuval Dagan, Noah Golowich, and Alexander Rakhlin. Smoothed online learning is as easy as statistical learning. In *Conference on Learning Theory*, pages 1716–1786. PMLR, 2022.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Eugenio Clerico, Hamish Flynn, and Gergely Neu. Confidence sequences for generalized linear models via regret analysis. *arXiv preprint arXiv:2504.16555*, 2025.

Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *21st Annual Conference on Learning Theory*, number 101, pages 355–366, 2008.

Victor H de la Pena. A general class of exponential inequalities for martingales and ratios. *The Annals of Probability*, 27(1):537–564, 1999.

Victor H. de la Peña, Michael J. Klass, and Tze Leung Lai. Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *The Annals of Probability*, 32(3):1902 – 1933, 2004.

Victor H de la Pena, Tze Leung Lai, and Qi-Man Shao. *Self-normalized processes: Limit theory and Statistical Applications*. Springer, 2009.

Michael Drmota, Philippe Jacquet, Changlong Wu, and Wojciech Szpankowski. Phase transition of regret for logistic regression with large weights. *Proceedings of Machine Learning Research vol XXX*, 1:28, 2026.

Jürgen Forster and Manfred K Warmuth. Relative expected instantaneous loss bounds. *Journal of Computer and System Sciences*, 64(1):76–102, 2002.

Dylan J. Foster, Alexander Rakhlin, and Karthik Sridharan. Zigzag: A new approach to adaptive online learning. *30th Annual Conference on Learning Theory*, 2017.

Dylan J. Foster, Alexander Rakhlin, and Karthik Sridharan. Online learning: Sufficient statistics and the burkholder method. *Conference on Learning Theory*, 2018.

Pierre Gaillard, Sébastien Gerchinovitz, Malo Huard, and Gilles Stoltz. Uniform regret bounds over $\mathbb{R}^d$ for the sequential linear regression problem with the square loss. In *Algorithmic Learning Theory*, pages 404–432. PMLR, 2019.

Nika Haghtalab, Tim Roughgarden, and Abhishek Shetty. Smoothed analysis with adaptive adversaries. *Journal of the ACM*, 71(3):1–34, 2024.

Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317, 2020.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. Improved regret bounds of (multinomial) logistic bandits via regret-to-confidence-set conversion. In *International Conference on Artificial Intelligence and Statistics*, pages 4474–4482. PMLR, 2024.

Jaouad Mourtada, Tomas Vaškevičius, and Nikita Zhivotovskiy. Distribution-free robust linear regression. *Mathematical Statistics and Learning*, 4(3):253–292, 2022.

Francesco Orabona and Kwang-Sung Jun. Tight concentrations and confidence sequences from the regret of universal portfolio. *IEEE Transactions on Information Theory*, 70(1):436–455, 2023.

Jian Qian, Alexander Rakhlin, and Nikita Zhivotovskiy. Refined risk bounds for unbounded losses via transductive priors. *Journal of Machine Learning Research*, 27(26):1–64, 2026.

Alexander Rakhlin and Karthik Sridharan. Online nonparametric regression. In *Conference on Learning Theory*, 2014.

Alexander Rakhlin and Karthik Sridharan. On equivalence of martingale tail bounds and deterministic regret inequalities. In *Conference on Learning Theory*, pages 1704–1722, 2017.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Random averages, combinatorial parameters, and learnability. *Advances in Neural Information Processing Systems 23*, pages 1984–1992, 2010.

Herbert Robbins and David Siegmund. Boundary crossing probabilities for the Wiener process and sample sums. *The Annals of Mathematical Statistics*, pages 1410–1429, 1970.

Tuhin Sarkar and Alexander Rakhlin. Near optimal finite time identification of arbitrary linear dynamical systems. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5610–5618. PMLR, 2019.

Max Simchowitz, Horia Mania, Stephen Tu, Michael I. Jordan, and Benjamin Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 439–473. PMLR, 06–09 Jul 2018.

Ramon van Handel. Probability in High Dimension. 2014.

Vladimir Vovk. Competitive on-line statistics. *International Statistical Review*, 69:213–248, 2001.

Volodya Vovk. Competitive on-line linear regression. *Advances in Neural Information Processing Systems*, 10, 1997.

Justin Whitehouse, Zhiwei Steven Wu, and Aaditya Ramdas. Time-uniform self-normalized concentration for vector-valued processes. *arXiv preprint arXiv:2310.09100*, 2023.

Heyang Zhao, Jiafan He, Dongruo Zhou, Tong Zhang, and Quanquan Gu. Variance-dependent regret bounds for linear bandits and reinforcement learning: Adaptivity and computational efficiency. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 4977–5020. PMLR, 2023.

Ingvar Ziemann. A Vector Bernstein Inequality for Self-Normalized Martingales. *Transactions on Machine Learning Research*, 2025.

## Contents of Appendix



## Appendix A. Proofs from Section 2

### A.1. Proof of Lemma 3

When the "centering" in (Rakhlin and Sridharan, 2014, Lemma 4) is incorporated into the $X$-process, the upper bound reads as

$$\sup_X \mathbb{E} \sup_{\theta \in \mathbb{R}^d} \sum_{t=1}^{T} 4\varepsilon_t \langle (\theta, 1), X_t \rangle - \langle (\theta, 1), X_t \rangle^2$$

where $X = (X_t)$ is an $\mathbb{R}^{d+1}$-valued tree with $X_t[d+1] \in [-1, 1]$. We over-bound by choosing a $(d+1)$-dimensional $\theta$ in the supremum. ■

## Appendix B. Proofs from Section 3

### B.1. Proof of Theorem 4

We note that

$$S_{T+1} = S_T + \varepsilon_{T+1} X_{T+1}, \qquad V_{T+1} = V_T + X_{T+1}^2.$$

Conditioning on $\mathcal{F}_T$ and expanding,

$$\begin{aligned}
\mathbb{E}[\exp(cR_{T+1}) \mid \mathcal{F}_T] &= \mathbb{E}\left[ \exp\left( \frac{c(S_T + \varepsilon_{T+1} X_{T+1})^2}{V_T + X_{T+1}^2} \right) \,\middle|\, \mathcal{F}_T \right] \\
&= \exp\left( \frac{cS_T^2 + cX_{T+1}^2}{V_T + X_{T+1}^2} \right) \mathbb{E}\left[ \exp\left( \frac{2c\,\varepsilon_{T+1} S_T X_{T+1}}{V_T + X_{T+1}^2} \right) \,\middle|\, \mathcal{F}_T \right].
\end{aligned}$$

Introduce

$$u := \frac{V_T}{V_T + X_{T+1}^2} \in [0,1], \qquad 1-u = \frac{X_{T+1}^2}{V_T + X_{T+1}^2}, \qquad \frac{S_T^2}{V_T + X_{T+1}^2} = u\frac{S_T^2}{V_T} = uR_T.$$

Also define

$$a := \frac{2c\, S_T X_{T+1}}{V_T + X_{T+1}^2}.$$

By Hoeffding's lemma for a Rademacher variable, $\mathbb{E}[\exp(\varepsilon_{T+1}a) \mid \mathcal{F}_T] \leqslant \exp(a^2/2)$, hence

$$\mathbb{E}[\exp(cR_{T+1}) \mid \mathcal{F}_T] \leqslant \exp\left(cuR_T + c(1-u) + \frac{a^2}{2}\right).$$

Next,

$$\frac{a^2}{2} = \frac{1}{2}\cdot\frac{4c^2 S_T^2 X_{T+1}^2}{(V_T + X_{T+1}^2)^2} = 2c^2\cdot\frac{S_T^2}{V_T}\cdot\frac{V_T X_{T+1}^2}{(V_T + X_{T+1}^2)^2} = 2c^2 R_T\, u(1-u).$$

Therefore,

$$\begin{aligned}\mathbb{E}[\exp(cR_{T+1}) \mid \mathcal{F}_T] &\leqslant \exp\big(cuR_T + c(1-u) + 2c^2 R_T\, u(1-u)\big)\\ &\leqslant \exp\big(cuR_T + c(1-u) + 2c^2 R_T\,(1-u)\big).\end{aligned}$$

where the second inequality is because $2c^2R_T(1-u)^2 \geqslant 0$. Now split into two cases.

*Case 1:* $R_T \leqslant \frac{1}{1-2c}$. Then we have

$$\begin{aligned}\exp\big(cuR_T + c(1-u) + 2c^2R_T\,(1-u)\big) &= \exp\big(c + 2c^2R_T + cuR_T - cu - 2c^2R_Tu\big)\\ &= \exp\big(c + 2c^2R_T + cu((1-2c)R_T - 1)\big)\\ &\leqslant \exp\big(c + 2c^2R_T\big)\\ &\leqslant \exp\Big(\frac{c}{1-2c}\Big) = C_0,\end{aligned}$$

where both inequalities are due to the condition of case 1.

*Case 2:* $R_T > \frac{1}{1-2c}$. Then $c(1-u)(1-(1-2c)R_T) \leqslant 0$. Reorgnizing the terms, we have $cuR_T + c(1-u) + 2c^2R_T(1-u) \leqslant cR_T$. This implies

$$\exp\big(cuR_T + c(1-u) + 2c^2R_T\,(1-u)\big) \leqslant \exp(cR_T).$$

Combining both cases yields the one-step inequality

$$\mathbb{E}[\exp(cR_{T+1}) \mid \mathcal{F}_T] \leqslant \exp(cR_T) + C_0.$$

Taking expectations and iterating gives

$$\mathbb{E}\exp(cR_T) \leqslant \mathbb{E}\exp(cR_1) + (T-1)C_0 \leqslant TC_0,$$

since $\exp(cR_1) = \exp(c) \leqslant \exp(c/(1-2c)) = C_0$. Finally, by Jensen's inequality, $c\,\mathbb{E}[R_T] \leqslant \log\mathbb{E}\exp(cR_T) \leqslant \log(TC_0)$, proving the stated bound. ∎

### B.2. Proof of Proposition 6

Let $n = \lfloor \frac{1}{2} \log T \rfloor$ and $K = \lfloor T/n \rfloor$. We set $M = \frac{2T}{n}$.

Consider the following sequence dyadic martingale difference sequence. Set $X_1 = 1$.

- For $t = jn + 1$, we set $X_t = M \cdot X_{(j-1)n+1}$ if there exists $\ell \in [(j-1)n + 1, jn]$ such that $\varepsilon_\ell = -1$. Otherwise we set $X_t = 0$.
- For $t \in (jn + 1, (j + 1)n]$, we set $X_t = X_{jn+1}$.

Let $i$ be the first index such that $\varepsilon_\ell = 1 \forall \ell \in [in + 1, (i + 1)n]$, and if no such index exists we set $i = K + 1$. Then, if $i \leqslant K$, we can bound $|S_T - M^i n| \leqslant M^{i-1} T$ and $V_T \leqslant \frac{nM^{2i}}{1-M^{-1}}$. Therefore $\frac{S_T^2}{V_T} \geqslant \frac{(n-T/M)^2}{2n} \geqslant \frac{n}{8}$. On the other hand, we have

$$\begin{aligned}\mathbb{P}(i = K + 1) &\leqslant \mathbb{P}(\forall 0 \leqslant j < K, \exists \ell \in [jn + 1, (j + 1)n], \varepsilon_\ell \neq 1) \\ &\leqslant (1 - 2^{-n})^K \leqslant e^{-2^{-n}K} \leqslant 1 - c_0,\end{aligned}$$

where $c_0 > 0$ is an absolute constant. This implies $\mathbb{E}[S_T^2/V_T] \geqslant \frac{c_0}{8} n = \Omega(\log T)$. ■

### B.3. Proof of Theorem 5

**Algorithm 2** Meta algorithm

**Input:** Parameter $M > 1$, round $T$, subroutine sequence $\{\mathsf{Alg}_k\}_{k\in\mathbb{Z}}$.

Initialize $k = -\infty$, $\mathcal{D} = \varnothing$.
**for** $t = 1, 2, \cdots, T$ **do**
  **if** $|x_t| \geqslant M^{k+1}$ **then**
    Update $k \leftarrow \lfloor \log_M |x_t| \rfloor$.
    Initialize $\mathsf{Alg}_k$ and set $\mathcal{D} = \varnothing$.
  Predict $\widehat{y}_t = \mathsf{Alg}_k(\mathcal{D}, x_t)$.
  Receive $y_t$ and update $\mathcal{D} \leftarrow \mathcal{D} \cup (x_t, y_t)$.

The above algorithm utilizes a sequence of subroutines $\{\mathsf{Alg}_k\}_{k\in\mathbb{Z}}$ and additionaly the trivial algorithm $\mathsf{Alg}_\infty$ that always predicts $\widehat{y}_t = 0$. We can then decompose the regret of Algorithm 2 to the regret of each subroutine $\mathsf{Alg}_k$.

**Assumption 12** *For any $k \in \mathbb{Z}$ and $n \leqslant T$, on any sequence $(x_1, y_1, \cdots, x_n, y_n)$ such that $\max_{i\in[n]} |x_i| \in [M^k, M^{k+1})$, the algorithm $\mathsf{Alg}_k$ achieves $\mathbf{Reg} \leqslant R_T$ and additionally*

$$\sum_{t=1}^{T} (\widehat{y}_t - y_t)^2 - \sum_{t=1}^{T} y_t^2 \leqslant \beta_T.$$

**Lemma 13** *Under Assumption 12, Algorithm 2 achieves*

$$\mathbf{Reg}(T) \leqslant T\beta_T + 2R_T + \frac{2T^2}{M}.$$

**Proof.** Let $\mathcal{K} \subset \{-\infty\} \cup \mathbb{Z}$ be the set of index $k$ such that $\mathsf{Alg}_k$ is executed, and suppose that $\mathsf{Alg}_k$ is executed on the time interval $T_k$. When $x_1 = \cdots = x_T = 0$ there is nothing to prove. Otherwise, we note that

$$\widehat{\theta} := \operatorname*{arg\,min}_{\theta \in \mathbb{R}} \sum_{t=1}^{T} (\langle \theta, x_t \rangle - y_t)^2 = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2}.$$

By Cauchy–Schwarz's inequality and $\max_t x_t \leqslant \sqrt{\sum_{t=1}^{T} x_t^2}$, we have

$$\max_t |x_t| \cdot \left( \sum_{t=1}^{T} y_t x_t \right) \leqslant \max_t |x_t| \cdot \sqrt{\sum_{t=1}^{T} y_t^2} \sqrt{\sum_{t=1}^{T} x_t^2} \leqslant \sqrt{T} \sum_{t=1}^{T} x_t^2.$$

In particular, $|\widehat{\theta}| \leqslant \frac{\sqrt{T}}{\max_t |x_t|}$. Therefore, we let $k^\star$ be the maximum of $\mathcal{K}$, and then $|\widehat{\theta}| \leqslant \frac{\sqrt{T}}{M^{k^\star}}$. We can then decompose

$$\begin{aligned} \mathbf{Reg} &= \sum_{t=1}^{T} (\widehat{y}_t - y_t)^2 - \sum_{t=1}^{T} (\widehat{\theta} x_t - y_t)^2 \\ &= \sum_{k \in \mathcal{K}} \sum_{t \in T_k} \left[ (\widehat{y}_t - y_t)^2 - (\widehat{\theta} x_t - y_t)^2 \right]. \end{aligned}$$

Note that when $k \leqslant k^\star - 2$, it holds that for any $t \in T_k$, $|x_t| \leqslant M^{k^\star - 1}$ and

$$(\widehat{\theta} x_t - y_t)^2 \geqslant y_t^2 - 2\widehat{\theta} x_t y_t \geqslant y_t^2 - \frac{2\sqrt{T}}{M}.$$

Therefore,

$$\begin{aligned} \mathbf{Reg} &= \sum_{k \in \mathcal{K}} \sum_{t \in T_k} \left[ (\widehat{y}_t - y_t)^2 - (\widehat{\theta} x_t - y_t)^2 \right] \\ &\leqslant \sum_{k \in \mathcal{K} : k \leqslant k^\star - 2} \sum_{t \in T_k} \left[ (\widehat{y}_t - y_t)^2 - y_t^2 + \frac{2\sqrt{T}}{M} \right] + \sum_{k \in \mathcal{K} : k \geqslant k^\star - 1} \sum_{t \in T_k} \left[ (\widehat{y}_t - y_t)^2 - (\widehat{\theta} x_t - y_t)^2 \right] \\ &\leqslant T \left( \beta_T + \frac{2\sqrt{T}}{M} \right) + 2R_T, \end{aligned}$$

where we use $\sum_{t \in T_k} \left[ (\widehat{y}_t - y_t)^2 - y_t^2 \right] \leqslant \beta_T$ and $\sum_{t \in T_k} \left[ (\widehat{y}_t - y_t)^2 - (\widehat{\theta} x_t - y_t)^2 \right] \leqslant R_T$ by our assumption (because we also know $\max_{t \in T_k} |x_t| \in [M^k, M^{k+1})$ unless $k = -\infty$). ∎

**Subroutines.** It remains to construct subroutines satisfying Assumption 12. We first recall that Vovk-Azoury-Warmuth forecaster has the following guarantee.

**Lemma 14** *On any sequence $(x_1, y_1, \cdots, x_n, y_n)$, the following rule*

$$\begin{aligned} \widehat{\theta}_t &= \operatorname*{arg\,min}_{\theta \in \mathbb{R}^d} \lambda \|\theta\|^2 + \langle \theta, x_t \rangle^2 + \sum_{i < t} (\langle \theta, x_i \rangle - y_i)^2, \\ \widehat{y}_t &= \mathsf{clip}_{[-1,1]} \Big( \langle \widehat{\theta}_t, x_t \rangle \Big), \end{aligned} \tag{20}$$

*achieves the following for any $\theta \in \mathbb{R}^d$:*

$$\sum_{t=1}^{n}(\widehat{y}_t - y_t)^2 - \sum_{t=1}^{n}(\langle \theta, x_t\rangle - y_t)^2 \leqslant \lambda\|\theta\|^2 + d\log\left(1 + \frac{n\max_t\|x_t\|^2}{\lambda}\right).$$

*In particular, when $d = 1$, we have the following guarantee:*

$$\sum_{t=1}^{n}(\widehat{y}_t - y_t)^2 - \inf_{\theta\in\Theta}\sum_{t=1}^{n}(\langle \theta, x_t\rangle - y_t)^2 \leqslant \frac{n\lambda}{\max_t x_t^2} + \log\left(1 + \frac{n\max_t x_t^2}{\lambda}\right).$$

However, the forecaster Eq. (20) may not achieve good $\beta(n)$ bound. Therefore, we hedge it against the 0 forecaster.

**Lemma 15** *Consider the following two experts problem: For $t \geqslant 1$, expert 0 predicts $\widehat{y}_{t,0} = 0$ and expert 2 predicts $\widehat{y}_{t,1}$ following Eq. (20). The final prediction is given by*

$$\widehat{p}_t(j) \propto_{j\in\{0,1\}} p_t(j)\exp\left(-\eta\sum_{i<t}(\widehat{y}_{i,j} - y_i)^2\right), \qquad \widehat{y}_t = \mathbb{E}_{j\sim\widehat{p}_t}[\widehat{y}_{i,j}],$$

*where $p_0(1) = 1 - p_0(0) = \varepsilon$. Then as long as $\eta \leqslant \frac{1}{8}$, it holds that*

$$\sum_{t=1}^{n}(\widehat{y}_t - y_t)^2 - \sum_{t=1}^{n}(\widehat{y}_{t,1} - y_t)^2 \leqslant \frac{1}{\eta}\log\frac{1}{p_0(1)} = \frac{\log(1/\varepsilon)}{\eta},$$
$$\sum_{t=1}^{n}(\widehat{y}_t - y_t)^2 - \sum_{t=1}^{n} y_t^2 \leqslant \frac{1}{\eta}\log\frac{1}{p_0(0)} \leqslant \frac{\varepsilon}{(1-\varepsilon)\eta}.$$

*In particular, it holds that* $\mathbf{Reg} \leqslant \frac{n\lambda}{\max_t x_t^2} + \log\left(1 + \frac{n\max_t x_t^2}{\lambda}\right) + \frac{\log(1/\varepsilon)}{\eta}$.

To summarize, we have the following corollary.

**Corollary 16** *For any $k \in \mathbb{Z}$ and $\beta \in (0, 1]$, there exists a subroutine $\mathsf{Alg}_k$ (by choosing $\lambda = \frac{M^{2k}}{T}$, $\eta = \frac{1}{8}$, and $\varepsilon = \frac{1}{16}\beta$ in Lemma 15) such that Assumption 12 holds with $\beta_T = \beta$ and $R_T \leqslant O(\log(TM/\beta))$.*

*In other words, on any sequence $(x_1, y_1, \cdots, x_n, y_n)$ such that $\max_{i\in[n]}|x_i| \in [M^k, M^{k+1})$ and $n \leqslant T$, the subroutine achieves*

$$\sum_{t=1}^{n}(\widehat{y}_t - y_t)^2 - \sum_{t=1}^{n} y_t^2 \leqslant \beta, \qquad \mathbf{Reg} \leqslant O(\log(TM/\beta)).$$

In particular, we can choose $\beta = \frac{1}{T}$ and $M = T^2$, and by Lemma 13, Algorithm 2 can be suitably instantiated such that it achieves

$$\mathbf{Reg} \leqslant O(\log T).$$

As a remark, the dependence $\log(1/\beta)$ is crucial for this regret bound, and it can be regarded as the "price of super-efficiency" against 0.

### B.4. Proof of Theorem 7

The martingale is constructed as follows. Let $\varepsilon_1, \dots,$ be an i.i.d. sequence of Rademacher random variables, and we will define $x_t = X_t(\varepsilon_1, \dots, \varepsilon_{t-1})$ below. Let $S_t = \sum_{i=1}^t \varepsilon_i x_i$ and $V_t = \sum_{i=1}^t x_i x_i^\top$. Consider $\widetilde{V}_t = V_t + I > V_t$. Observe that under our construction,

$$\mathbb{E}\left[\|S_{t+1}\|^2_{\widetilde{V}_{t+1}^{-1}}\right] = \mathbb{E}\left[\|S_t\|^2_{\widetilde{V}_{t+1}^{-1}} + \|x_{t+1}\|^2_{\widetilde{V}_{t+1}^{-1}}\right].$$

Using

$$\widetilde{V}_{t+1}^{-1} = \widetilde{V}_t^{-1} - \frac{\widetilde{V}_t^{-1} x_{t+1} x_{t+1}^\top \widetilde{V}_t^{-1}}{1 + \|x_{t+1}\|^2_{\widetilde{V}_t^{-1}}},$$

we have

$$\mathbb{E}\left[\|S_{t+1}\|^2_{\widetilde{V}_{t+1}^{-1}}\right] = \mathbb{E}\left[\|S_t\|^2_{\widetilde{V}_t^{-1}} + \frac{\|x_{t+1}\|^2_{\widetilde{V}_t^{-1}} - \langle x_{t+1}, \widetilde{V}_t^{-1} S_t\rangle^2}{1 + \|x_{t+1}\|^2_{\widetilde{V}_t^{-1}}}\right].$$

We define $x_t$ recursively: $x_1$ is an arbitrarily fixed vector with norm $r > 0$, and for $t \geqslant 1$,

$$x_{t+1} = r\widetilde{V}_t^{1/2} e_t, \qquad \text{where } e_t \text{ is a unit vector such that } \langle e_t, \widetilde{V}_t^{-1/2} S_t\rangle = 0. \tag{21}$$

Then, it holds that

$$\mathbb{E}\left[\|S_{t+1}\|^2_{\widetilde{V}_{t+1}^{-1}}\right] = \mathbb{E}\left[\|S_t\|^2_{\widetilde{V}_t^{-1}} + \frac{\|x_{t+1}\|^2_{\widetilde{V}_t^{-1}}}{1 + \|x_{t+1}\|^2_{\widetilde{V}_t^{-1}}}\right] = \mathbb{E}\left[\|S_t\|^2_{\widetilde{V}_t^{-1}}\right] + \frac{r^2}{r^2+1}.$$

Hence, since $S_T \in \mathrm{Range}(V_T)$, it is clear that $\mathbb{E}[\|S_T\|^2_{V_T^\dagger}] \geqslant \mathbb{E}[\|S_T\|^2_{\widetilde{V}_T^{-1}}] = \frac{Tr^2}{r^2+1}$, and the proof is then completed by choosing $r = \frac{1}{\varepsilon}$. ∎

## Appendix C. Proofs from Section 4

### C.1. Proof of Lemma 10

Fix $t \geqslant 1$. Since $\hat{y}_t$ is $\mathcal{F}_{t-1}$-measurable, using Eq. (16) with $\alpha = \hat{y}_t/\sigma^2$ gives

$$\begin{aligned}\mathbb{E}\left[\exp\left(\frac{1}{2\sigma^2}(2\hat{y}_t y_t - \hat{y}_t^2)\right)\middle|\mathcal{F}_{t-1}\right] &= \exp\left(-\frac{\hat{y}_t^2}{2\sigma^2}\right)\mathbb{E}\left[\exp\left(\frac{\hat{y}_t}{\sigma^2}y_t\right)\middle|\mathcal{F}_{t-1}\right] \\ &\leqslant \exp\left(-\frac{\hat{y}_t^2}{2\sigma^2}\right)\exp\left(\frac{\hat{y}_t^2}{2\sigma^2}\right) = 1.\end{aligned}$$

Multiplying by $M_{t-1}$ and taking conditional expectations yields $\mathbb{E}[M_t \mid \mathcal{F}_{t-1}] \leqslant M_{t-1}$, so $(M_t)$ is a nonnegative supermartingale and hence $\mathbb{E}[M_t] \leqslant \mathbb{E}[M_0] = 1$. Finally, the proof follows from Ville's inequality for nonnegative supermartingales. ∎

We will make use of the following upper bound of VAW.

**Proposition 17** *Suppose that $y_t \in [-1,1]$ deterministically. Then Algorithm 1 achieves*

$$\sum_{t=1}^{T}(\widehat{y}_t - y_t)^2 - \inf_{\theta\in\mathbb{R}^d}\sum_{t=1}^{T}(\langle\theta, x_t\rangle - y_t)^2 \leqslant \sum_{t=1}^{T} x_t^\top V_t^\dagger x_t.$$

*More generally, suppose that $y_t$ satisfies $\mathbb{E}[e^{y_t^2/m^2} \mid \mathcal{F}_{t-1}] \leqslant e$. Then it holds that with probability at least $1-\delta$,*

$$\sum_{t=1}^{T}(\widehat{y}_t - y_t)^2 - \inf_{\theta\in\mathbb{R}^d}\sum_{t=1}^{T}(\langle\theta, x_t\rangle - y_t)^2 \leqslant m^2\sum_{t=1}^{T} x_t^\top V_t^\dagger x_t + m^2\log(1/\delta).$$

**Proof of Proposition 17.** Note that $\widehat{y}_t$ does not depend on the choice of $\widehat{\theta}_t$. Therefore, we only need to consider $\widehat{\theta}_t = V_t^\dagger S_{t-1}$, where $S_{t-1} = \sum_{i<t} x_i y_i$. Further, we know $\widehat{\theta} = V_T^\dagger S_T \in \arg\min_{\theta\in\mathbb{R}^d}\sum_{t=1}^{T}(\langle\theta, x_t\rangle - y_t)^2$. Then, we can calculate

$$\sum_{t=1}^{T}(\widehat{y}_t - y_t)^2 - \sum_{t=1}^{T}(\langle\widehat{\theta}, x_t\rangle - y_t)^2 = \sum_{t=1}^{T}\Big[(x_t^\top V_t^\dagger S_{t-1})^2 - 2y_t x_t^\top V_t^\dagger S_{t-1}\Big] + S_T^\top V_T^\dagger S_T.$$

We also have

$$S_t^\top V_t^\dagger S_t = y_t^2 x_t^\top V_t^\dagger x_t + 2y_t x_t V_t^\dagger S_{t-1} + S_{t-1}^\top V_t^\dagger S_{t-1}.$$

Further, we have the following basic inequality: For PSD matrix $V$, $v \in \mathrm{span}(V)$, and any $w$, it holds that

$$v(V + ww^\top)^\dagger v \leqslant vV^\dagger v - (w^\top(V + ww^\top)^\dagger v)^2.$$

When $w \in \mathrm{span}(V)$ this is straight-forward by restricting to the subspace $\mathrm{span}(V)$ where $V$ becomes invertible. Otherwise, we can consider $h = (V + ww^\top)^\dagger v$, and then $v = Vh + w\langle w, h\rangle$, and using $v \in \mathrm{span}(V)$ implies $\langle w, h\rangle = 0$. Plugging this in, and we can see the equality holds.

Now, combining the inequalities above, we know

$$S_t^\top V_t^\dagger S_t \leqslant y_t^2 x_t^\top V_t^\dagger x_t + 2y_t x_t V_t^\dagger S_{t-1} + S_{t-1}^\top V_{t-1}^\dagger S_{t-1} - (x_t^\top V_t^\dagger S_{t-1})^2.$$

Applying this inequality recursively, it holds

$$S_T^\top V_T^\dagger S_T \leqslant \sum_{t=1}^{T}\Big[-(x_t^\top V_t^\dagger S_{t-1})^2 + 2y_t x_t^\top V_t^\dagger S_{t-1} + y_t^2 x_t^\top V_t^\dagger x_t\Big].$$

Reorganizing yields

$$\sum_{t=1}^{T}(\widehat{y}_t - y_t)^2 - \inf_{\theta\in\mathbb{R}^d}\sum_{t=1}^{T}(\langle\theta, x_t\rangle - y_t)^2 \leqslant \sum_{t=1}^{T} y_t^2 x_t^\top V_t^\dagger x_t,$$

and the first inequality follows immediately. For the second upper bound, we use the fact that $\mathbb{E}[e^{\lambda y^2/m^2}|\mathcal{F}_{t-1}] \leqslant e^{\lambda}$ for $\lambda \in [0,1]$, and hence using this inequality recursively, we derive

$$\mathbb{E}\left[\exp\left(\frac{1}{m^2}\sum_{t=1}^{T} y_t^2 x_t^\top V_t^\dagger x_t - \sum_{t=1}^{T} x_t^\top V_t^\dagger x_t\right)\right] \leqslant 1.$$

By Markov's inequality, the desired upper bound follows. ■

To provide a upper bound for VAW, it remains to upper bound the quantity $\sum_{t=1}^{T} x_t^\top V_t^\dagger x_t$. This is nontrivial because the matrix $V_t$ can be ill-conditioned, and (as we have shown in Section 3.2) this sum can be $\Omega(T)$ without smoothness. In the following, we bound this quantity by a "combinatorial dimension" of the sequence $x$, and then apply the coupling trick and the backward analysis technique.

**Step 1: Connect to combinatorial quantities of the sequence.** To proceed, we introduce the notion of "bad" subsequence. For any sequence $z = (z_1, \cdots, z_k)$, we define $V(z) := \sum_{i \leqslant k} z_i z_i^\top$. We call a sequence $z = (z_1, \cdots, z_k)$ *$r$-bad* if for any $t \in [k]$, $\|z_t\|^2_{V(z_{1:t})^\dagger} \geqslant r$. We define $N(r; z)$ to be the length of the longest $r$-bad subsequence of $z$.

In the following, we bound the sum $\sum_{t=1}^{T} x_t^\top V_t^\dagger x_t$ by $N(r; x)$ for $r \in \{2^{-1}, 2^{-2}, \cdots\}$.

**Lemma 18** *For any sequence $x = (x_1, \cdots, x_T)$, it holds that*

$$\sum_{t=1}^{T} x_t^\top V_t(x)^\dagger x_t \leqslant \int_0^1 N(r; x) dr \leqslant 1 + \sum_{1 \leqslant i \leqslant \log n} 2^{-i} N(2^{-i}; x).$$

**Proof of Lemma 18.** By definition, $x_t^\top V_t(x)^\dagger x_t \in [0,1]$, and hence

$$x_t^\top V_t(x)^\dagger x_t = \int_0^1 \mathbf{1}\left\{x_t^\top V_t(x)^\dagger x_t \geqslant r\right\} dr.$$

Therefore,

$$\sum_{t=1}^{T} x_t^\top V_t(x)^\dagger x_t = \int_0^1 \sum_{t=1}^{T} \mathbf{1}\left\{x_t^\top V_t(x)^\dagger x_t \geqslant r\right\} dr.$$

Note that $\sum_{t=1}^{T} \mathbf{1}\left\{x_t^\top V_t(x)^\dagger x_t \geqslant r\right\} \leqslant N(r; x)$, because if we consider all the indices $t_1 < \cdots < t_k$ such that $x_t^\top V_t(x)^\dagger x_t \geqslant r$, then $(x_{t_1}, \cdots, x_{t_k})$ is a $r$-bad sequence. Hence, we have

$$\begin{aligned}\sum_{t=1}^{T} x_t^\top V_t(x)^\dagger x_t &= \int_0^1 \sum_{t=1}^{T} \mathbf{1}\left\{x_t^\top V_t(x)^\dagger x_t \geqslant r\right\} dr \\ &\leqslant \int_0^1 N(r; x) dr \leqslant 1 + \sum_{1 \leqslant i \leqslant \log n} 2^{-i} N(2^{-i}; x).\end{aligned}$$

The claim follows.

■

**Step 2: Coupling smooth sequence to a i.i.d sequence.** We invoke the following coupling lemma for smooth sequence (Haghtalab et al., 2024). The idea of this lemma is quite simple: Any smooth sequence can be generated by performing rejection sampling. We sketch the proof below.

**Lemma 19** *Suppose that Assumption 8 holds. Then for any $\delta \in (0,1)$, $K \geqslant C_{\mathsf{cov}} \log(T/\delta)$, there exists a coupling between $x = (x_1, \cdots, x_T)$ with a sequence $z = (z_{1,1}, \cdots, z_{1,K}; \cdots; z_{T,1}, \cdots, z_{T,K})$ such that:*

*(1) Marginally, $(z_{t,j})_{1 \leqslant t \leqslant T, 1 \leqslant j \leqslant K} \sim \mu$ are i.i.d random vectors from $\mu$.*

*(2) With probability at least $1-\delta$, it holds that $x_t \in \{z_{t,i} : i = 1, 2, \cdots, K\}$ for all $t \in [T]$.*

Note that this lemma implies (with probability at least $1-\delta$) $N(r; x) \leqslant N(r; z)$, and it remains to control $N(r; z)$ under the i.i.d sequence $z$.

**Proof of Lemma 19.** Consider the randomness $z^\infty = (z_{t,j})_{1 \leqslant t \leqslant T, j \geqslant 1} \sim \mu$ are i.i.d random vectors from $\mu$. Consider an environment Env that adopts the following protocol for each $t = 1, 2, \cdots, T$:

- Given the history $\mathcal{H}_{t-1}^x$, the environment fix the distribution $p_t(\cdot) = P_t(\cdot \mid \mathcal{H}_{t-1}^x)$ and perform *rejection sampling:*
- For $j = 1, 2, \cdots$, with probability $\frac{p_t(z_{t,j})}{C_{\mathsf{cov}} \mu(z_{t,j})}$, the environment set $x_t = z_{t,j}$ and break. Otherwise, the environment goes to the next step $j+1$.

By the guarantee of rejection sampling, we know that conditional on $\mathcal{H}_{t-1}$, the vector $x_t$ is generated from the distribution $p_t = P_t(\cdot \mid \mathcal{H}_{t-1}^x)$. Further, for any fixed $t \in [T]$, with probability at least $1-\delta$ it holds that $x_t = z_{t,j}$ for some $j \leqslant C_{\mathsf{cov}} \log(1/\delta)$. Therefore, by union bound, the above construction gives a coupling between the sequence $x$ and the i.i.d sequence $z^\infty$, such that $x_t \in \{z_{t,i} : i = 1, 2, \cdots, K\}$ for all $t \in [T]$. ■

**Step 3: Bad i.i.d sequence is rare.** Finally, the problem is now reduced to bounding the probability that a i.i.d sequence $(z_1, \cdots, z_k)$ being $r$-bad, as we handle in the following proposition.

**Proposition 20** *Fix any $r \in (0, 1]$. Suppose that $z = (z_1, \cdots, z_n)$ is a sequence of $n$ i.i.d vectors drawn from $\mu$. Then with probability at least $1-\delta$ (over the randomness of $z$)*

$$N(r; z) \leqslant 3\sqrt{nd/r} + 6\log(1/\delta).$$

**Proof of Proposition 20.** Fix any $k \geqslant 1$, we bound the probability $p := \mathbb{P}_{z \sim \mu}(N(r; z) \geqslant k)$. Note that $N(r; z) \geqslant k$ if and only if there exists a subset $I = \{i_1 < \cdots < i_k\} \subseteq [n]$ such that the sequence $z_I = (z_{i_1}, \cdots, z_{i_k})$ is $r$-bad. Also, note that there are $\binom{n}{k}$ many such subsets. Therefore, we can bound

$$p := \mathbb{P}_{z \sim \mu}(N(r; z) \geqslant k) \leqslant \sum_{I \subseteq [n], |I| = k} \mathbb{P}_{z \sim \mu}(z_I \text{ is } r\text{-bad}) = \binom{n}{k} p_0, \tag{22}$$

where we denote $p_0 = \mathbb{P}_{z \sim \mu}((z_1, \cdots, z_k) \text{ is } r\text{-bad})$ and use the exchangeability of i.i.d random variables.

In the following, we proceed to upper bound $p_0$. To this end, we consider the unordered multiset $\mathcal{S}_i = \{z_1, \cdots, z_i\}$ and the following random process:

$$\mathcal{S}_n \to \mathcal{S}_{n-1} \to \cdots \to \mathcal{S}_1. \tag{23}$$

Note that this is a Markov chain, such that given $\mathcal{S}_i$, the multiset $\mathcal{S}_{i-1}$ is generated as first randomly select $z_i \sim \mathrm{Unif}(\mathcal{S}_i)$, and the set $\mathcal{S}_{i-1} = \mathcal{S}_i \backslash \{z_i\}$. Further, we note that $(z_1, \cdots, z_k)$ is $r$-bad if and only if

$$w_i := z_i \left( \sum_{j \leqslant i} z_j z_j^\top \right)^\dagger z_i \geqslant r, \qquad \forall i \in [k]. \tag{24}$$

Now, we can consider the backward expectation, where we define $\mathcal{H}_i = (S_n, \cdots, S_i)$ to be the history up to step $i$:

$$\mathbb{E}[w_i \mid \mathcal{H}_i] = \mathbb{E}[w_i \mid \mathcal{S}_i] = \mathbb{E}\left[ z_i \left( \sum_{z \in \mathcal{S}_i} z z^\top \right)^\dagger z_i \mid \mathcal{S}_i \right] = \mathbb{E}_{z_i \sim \mathrm{Unif}(\mathcal{S}_i)} \left[ z_i \left( \sum_{z \in \mathcal{S}_i} z z^\top \right)^\dagger z_i \right] \tag{25}$$

$$= \mathrm{tr}\left( \frac{1}{i} \left( \sum_{z \in \mathcal{S}_i} z z^\top \right) \left( \sum_{z \in \mathcal{S}_i} z z^\top \right)^\dagger \right) \leqslant \frac{d}{i}, \tag{26}$$

where we use $\mathrm{tr}(AA^\dagger) \leqslant d$ for any $d \times d$ positive semi-definite matrix $A$. In particular, this implies $\mathbb{P}(w_i \geqslant r \mid \mathcal{H}_i) \leqslant \min\{1, \frac{d}{ri}\} =: a_i$ for any $i \geqslant 1$. Now, we can bound

$$p_0 = \mathbb{P}_{z \sim \mu}((z_1, \cdots, z_k) \text{ is } r\text{-bad}) = \mathbb{P}_{z \sim \mu}(w_1 \geqslant r, \cdots, w_k \geqslant r) \tag{27}$$

$$= \mathbb{E}[\mathbf{1}\{w_1 \geqslant r, \cdots, w_k \geqslant r\}] = \mathbb{E}[\mathbb{P}(w_1 \geqslant r \mid \mathcal{H}_1) \cdot \mathbf{1}\{w_2 \geqslant r, \cdots, w_k \geqslant r\}] \tag{28}$$

$$\leqslant a_1 \, \mathbb{E}[\mathbf{1}\{w_2 \geqslant r, \cdots, w_k \geqslant r\}] = a_1 \, \mathbb{E}[\mathbb{P}(w_2 \geqslant r \mid \mathcal{H}_2) \cdot \mathbf{1}\{w_3 \geqslant r, \cdots, w_k \geqslant r\}] \tag{29}$$

$$\leqslant a_1 a_2 \, \mathbb{E}[\mathbf{1}\{w_3 \geqslant r, \cdots, w_k \geqslant r\}] \leqslant \cdots \tag{30}$$

$$\leqslant a_1 a_2 \cdots a_k = \prod_{i=1}^{k} \min\left\{1, \frac{d}{ri}\right\} \leqslant \frac{d^k}{k!} \leqslant \left(\frac{ed}{rk}\right)^k, \tag{31}$$

where we use $k! \geqslant (k/e)^k$. Therefore, we can conclude that

$$p \leqslant \binom{n}{k} \left(\frac{ed}{rk}\right)^k \leqslant \left(\frac{e^2 nd}{rk^2}\right)^k. \tag{32}$$

Then, as long as $k \geqslant 3\sqrt{nd/r} + 6\log(1/\delta)$, it holds that $p = \mathbb{P}_{z \sim \mu}(N(r; z) \geqslant k) \leqslant \delta$. This is the desired result. ■

**Finalizing the proof.** Combining the results above, we can conclude that VAW achieves a sublinear regret.

**Proposition 21** *Under Assumption 8, with probability at least* $1 - \delta$*:*

$$\sum_{t=1}^{T} x_t^\top V_t^\dagger x_t \lesssim \sqrt{d C_{\mathsf{cov}} T \log(T/\delta)} + \log(1/\delta).$$

**Proof of Proposition 21.** By Lemma 18, we can further upper bound

$$\sum_{t=1}^{T} x_t^\top V_t^\dagger x_t =: \bar{R}_T \leqslant 1 + \sum_{1 \leqslant i \leqslant \log n} 2^{-i} N(2^{-i}; x). \tag{33}$$

Now, we take the coupling $\gamma$ constructed in Lemma 19 (with $K = \lceil C_{\mathsf{cov}} \log(2T/\delta) \rceil$). We know that under $\gamma$ and a success event $\mathcal{E}$ such that $\mathbb{P}(\mathcal{E}) \geqslant 1 - \frac{\delta}{2}$, the sequence $z = (z_{t,j})_{1 \leqslant t \leqslant T, 1 \leqslant j \leqslant K}$ are i.i.d random vectors from $\mu$, and $x$ is a subsequence of $z$. This immediately implies that under $\mathcal{E}$, we have $N(r; x) \leqslant N(r; z)$ for any $r \in [0, 1]$. Finally, applying Proposition 20 to $r_i = 2^{-i}$ for $1 \leqslant i \leqslant \log_2 n$ and take a union bound, we know that with probability at least $1 - \delta/2$,

$$N(r_i; z) \leqslant 3\sqrt{nd/r_i} + 6 \log(4 \log_2(n)/\delta), \qquad \forall 1 \leqslant i \leqslant \log_2 n, \tag{34}$$

where $n = TK$. Combining the inequalities above and taking union bound gives the desired result. ■

**Proof of Theorem 9.** By Proposition 17 and Proposition 21, we have

$$\begin{aligned}
\mathbf{Reg}(T) &:= \sum_{t=1}^{T} (\widehat{y}_t - y_t)^2 - \inf_{\theta \in \mathbb{R}^d} \sum_{t=1}^{T} (\langle \theta, x_t \rangle - y_t)^2 \\
&\leqslant \sum_{t=1}^{T} x_t^\top V_t^\dagger x_t \\
&\lesssim \sqrt{d C_{\mathsf{cov}} T \log(T/\delta)} + \log(1/\delta).
\end{aligned}$$

■

### C.2. Tightness of our analysis

We argue that in the worst-case, the sum $\sum_{t=1}^{T} x_t^\top V_t^\dagger x_t = \Omega(\sqrt{C_{\mathsf{cov}} T})$ even when $d = 1$, demonstrating that our analysis is nearly tight.

**Lemma 22** *For any $C > 1$, there exists a smooth environment with $C_{\mathsf{cov}} \leqslant C$ such that*

$$\mathbb{E}\left[ \sum_{t=1}^{T} \frac{x_t^2}{1 + \sum_{i \leqslant t} x_i^2} \right] \geqslant \Omega(\sqrt{(C-1)T} \wedge T).$$

**Proof.** Fix $1 \leqslant n \leqslant \frac{(C-1)T}{4} + 1$ and set $p = \min\{\frac{C-1}{n}, 1\}$. We consider the following environment:

- Initialize $k = 1$.
- For $t = 1, 2, \cdots$: With probability $p$, set $x_t = 2^k$ and set $k \leftarrow k + 1$. If $k > n$, terminates (i.e., outputs $x_t = 0$ afterwards). Otherwise, set $x_t = 0$.

Then it is clear that the environment is $C$-smooth with measure $\mu$ given by $\mu(2^k) = \frac{p}{C}$ for $k \in [n]$ and $\mu(0) = 1 - \frac{np}{C} \geqslant \frac{1}{C}$. Further, when $x_t > 0$, it must hold that $\frac{x_t^2}{1+\sum_{i \leqslant t} x_i^2} \geqslant \frac{1}{2}$. Therefore, we can lower bound

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{x_t^2}{1 + \sum_{i \leqslant t} x_i^2}\right] \geqslant \frac{1}{2}\,\mathbb{E}[\min\{n, L\}],$$

where $L \sim \mathrm{Binomial}(T, p)$. As long as $p \geqslant \frac{1}{4T}$, it is clear that there is a constant $c > 0$ such that $\mathbb{P}(L \leqslant cTp) \leqslant \frac{1}{2}$, and hence $\mathbb{E}[\min\{n, L\}] \geqslant \frac{1}{4}\min\{2n, cTp\}$. Suitably choosing $n$ gives the desired lower bound. ∎

### C.3. Dyadic filtration is enough

**Proof of Lemma 1.** The inequality $\mathcal{R}_d^{\mathrm{dyadic}}(T) \leqslant \mathcal{R}_d(T)$ is immediate, since every dyadic martingale is admissible in the definition of $\mathcal{R}_d(T)$.

For the reverse inequality, let

$$\Phi(s, V) := \|s\|^2_{V^\dagger}, \qquad s \in \mathbb{R}^d, \quad V \succeq 0.$$

Thus, for any admissible process $(X_t, Y_t)_{t=1}^T$,

$$R_T = \Phi(S_T, V_T), \qquad S_t = \sum_{i \leqslant t} Y_i X_i, \quad V_t = \sum_{i \leqslant t} X_i X_i^\top.$$

Let $\tilde{\Delta}$ denote the set of all probability laws on $[-1, 1]$ with mean zero. Define recursively, for $t = T, T-1, \ldots, 1$,

$$F_{T+1}(s, V) := \Phi(s, V), \qquad F_t(s, V) := \sup_{x \in \mathbb{R}^d} \sup_{p \in \tilde{\Delta}} \mathbb{E}_{Y \sim p}\big[F_{t+1}(s + Yx,\ V + xx^\top)\big].$$

We first claim that

$$\mathcal{R}_d(T) \leqslant F_1(0, 0).$$

Indeed, fix any admissible process $(X_t, Y_t)_{t=1}^T$, and we prove by backward induction on $t$ that

$$\mathbb{E}[\Phi(S_T, V_T) \mid \mathcal{G}_{t-1}] \leqslant F_t(S_{t-1}, V_{t-1}) \qquad \text{a.s.}$$

The case $t = T + 1$ is tautological. Assuming the claim at time $t + 1$, we have

$$\begin{aligned}
\mathbb{E}[\Phi(S_T, V_T) \mid \mathcal{G}_{t-1}] &= \mathbb{E}[\mathbb{E}[\Phi(S_T, V_T) \mid \mathcal{G}_t] | \mathcal{G}_{t-1}] \\
&\leqslant \mathbb{E}[F_{t+1}(S_t, V_t) | \mathcal{G}_{t-1}] \\
&= \mathbb{E}\big[F_{t+1}(S_{t-1} + Y_t X_t,\ V_{t-1} + X_t X_t^\top) \big| \mathcal{G}_{t-1}\big].
\end{aligned}$$

Conditionally on $\mathcal{G}_{t-1}$, the vector $X_t$ is fixed, while the conditional law of $Y_t$ belongs to $\tilde{\Delta}$ (because $|Y_t| \leqslant 1$ a.s. and $\mathbb{E}[Y_t \mid \mathcal{G}_{t-1}] = 0$). Hence the last display is at most $F_t(S_{t-1}, V_{t-1})$ by the

definition of $F_t$. Taking expectations at $t = 1$ and then the supremum over all admissible processes yields $\mathcal{R}_d(T) \leqslant F_1(0,0)$.

Next we claim that, for every $t$ and every $V \succeq 0$, the map $s \mapsto F_t(s,V)$ is convex. This is proved by backward induction on $t$. At time $T+1$, the claim is immediate since

$$\Phi(s,V) = s^\top V^\dagger s$$

and $V^\dagger \succeq 0$. If $F_{t+1}(\cdot, V')$ is convex for every $V' \succeq 0$, then for fixed $x \in \mathbb{R}^d$ and fixed $p \in \tilde{\Delta}$, the map

$$s \mapsto \mathbb{E}_{Y\sim p}\big[F_{t+1}(s + Yx,\ V + xx^\top)\big]$$

is convex, because translation and expectation preserve convexity. Taking the supremum over $x$ and $p$ shows that $F_t(\cdot, V)$ is convex as well.

Now define the dyadic Bellman recursion by

$$G_{T+1}(s,V) := \Phi(s,V), \qquad G_t(s,V) := \sup_{x\in\mathbb{R}^d} \mathbb{E}_\varepsilon\big[G_{t+1}(s + \varepsilon x,\ V + xx^\top)\big],$$

where $\varepsilon$ is a Rademacher random variable.

We show by backward induction that $F_t = G_t$ for all $t$. The terminal condition is clear. Assume $F_{t+1} = G_{t+1}$. Fix $s \in \mathbb{R}^d$, $V \succeq 0$, and $x \in \mathbb{R}^d$, and define

$$g_x(y) := F_{t+1}(s + yx,\ V + xx^\top), \qquad y \in [-1,1].$$

Since $F_{t+1}(\cdot, V + xx^\top)$ is convex, the function $g_x$ is convex on $[-1,1]$. Therefore, for every $y \in [-1,1]$,

$$g_x(y) \leqslant \frac{1+y}{2} g_x(1) + \frac{1-y}{2} g_x(-1).$$

Taking expectation with respect to any $p \in \tilde{\Delta}$ and using $\mathbb{E}_{Y\sim p}[Y] = 0$, we obtain

$$\mathbb{E}_{Y\sim p}[g_x(Y)] \leqslant \frac{1}{2} g_x(1) + \frac{1}{2} g_x(-1).$$

Taking the supremum over $p \in \tilde{\Delta}$ and then over $x \in \mathbb{R}^d$ gives

$$F_t(s,V) \leqslant \sup_{x\in\mathbb{R}^d} \frac{F_{t+1}(s+x, V+xx^\top) + F_{t+1}(s-x, V+xx^\top)}{2} = G_t(s,V).$$

The reverse inequality is immediate, since the symmetric Rademacher law $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ belongs to $\tilde{\Delta}$. Hence $F_t = G_t$ for all $t$.

Finally, let $H_t(s,V)$ denote the supremum over all dyadic trees from rounds $t, \dots, T$ of the expected terminal payoff starting from state $(s,V)$. Then $H_{T+1} = \Phi$, and $H_t$ satisfies the same recursion as $G_t$:

$$H_t(s,V) = \sup_{x\in\mathbb{R}^d} \mathbb{E}_\varepsilon\big[H_{t+1}(s + \varepsilon x,\ V + xx^\top)\big].$$

Therefore $H_t = G_t$ for all $t$, and in particular

$$G_1(0,0) = H_1(0,0) = \mathcal{R}_d^{\mathrm{dyadic}}(T).$$

Combining the previous steps, we conclude that

$$\mathcal{R}_d(T) \leqslant F_1(0,0) = G_1(0,0) = \mathcal{R}_d^{\mathrm{dyadic}}(T).$$

Together with the trivial inequality $\mathcal{R}_d^{\mathrm{dyadic}}(T) \leqslant \mathcal{R}_d(T)$, this yields $\mathcal{R}_d^{\mathrm{dyadic}}(T) = \mathcal{R}_d(T)$. ■

## Appendix D. Proofs from Section 5

### D.1. Proof of Theorem 11

By Lemma 10 and Eq. (18), taking $\Gamma$ in Eq. (18) to 0, we have with probability at least $1 - \delta/2$,

$$\|S_T\|^2_{V_T^\dagger} \leqslant \mathbf{Reg}_T(\hat{y}) + 2\sigma^2 \log(2/\delta). \tag{35}$$

On the other hand, applying Propositions 17 and 21 to the VAW predictor with confidence level $\delta/2$ gives, with probability at least $1 - \delta/2$,

$$\mathbf{Reg}_T(\hat{y}) \lesssim \sigma^2 \Big( \sqrt{d\, C_{\mathsf{cov}}\, T\, \log(2T/\delta)} + \log(2/\delta) \Big). \tag{36}$$

By the union bound, with probability at least $1 - \delta$ both Eq. (35) and Eq. (36) hold. ■