Journal of Machine Learning Research 18 (2017) 1-31 Submitted 2/17; Revised 4/17; Published 5/17

Perishability of Data: Dynamic Pricing under Varying-Coefficient Models

Adel Javanmard AJAVANMA@USC.EDU
Department of Data Sciences and Operations
Marshall School of Business
University of Southern California
Los Angeles, CA 90089 , USA

Editor: Qiang Liu

Abstract

We consider a firm that sells a large number of products to its customers in an online fashion. Each product is described by a high dimensional feature vector, and the market value of a product is assumed to be linear in the values of its features. Parameters of the valuation model are unknown and can change over time. The firm sequentially observes a product's features and can use the historical sales data (binary sale/no sale feedbacks) to set the price of current product, with the objective of maximizing the collected revenue. We measure the performance of a dynamic pricing policy via regret, which is the expected revenue loss compared to a clairvoyant that knows the sequence of model parameters in advance.

We propose a pricing policy based on projected stochastic gradient descent (PSGD) and characterize its regret in terms of time T, features dimension d, and the temporal variability in the model parameters, δ_t. We consider two settings. In the first one, feature vectors are chosen antagonistically by nature and we prove that the regret of PSGD pricing policy is of order $O(\sqrt{T} + \sum_{t=1}^{T} \sqrt{t}\delta_t)$. In the second setting (referred to as stochastic features model), the feature vectors are drawn independently from an unknown distribution. We show that in this case, the regret of PSGD pricing policy is of order $O(d^2 \log T + \sum_{t=1}^{T} t\delta_t/d)$.

Keywords: Dynamic Pricing, Revenue Management, Varying-Coefficient Models, Regret, Stochastic Gradient Descent, Hypothesis Testing

1. Introduction

Motivated by the prevalence of online marketplaces, we consider the problem of a firm selling a large number of products, that are significantly differentiated from each other, to customers that arrive over time. The firm needs to price the products in a dynamic manner, with the objective of maximizing the expected revenue.

The majority of work in dynamic pricing assume that a retailer sells *identical* items to its customers (Besbes and Zeevi, 2009; Farias and Van Roy, 2010; Broder and Rusmevichientong, 2012; den Boer and Zwart, 2013; Wang et al., 2014). Recently, feature-based models have been used to model the products differentiation by assuming that each product is described by vectors of high-dimensional features. These models are suitable for business settings where there are an enormous number of distinct products. One important example is online ad markets. In this context, products are the impressions (user view) that are sold by the web publisher to advertisers. Due to the ever-growing amount of data that is available on the Internet, for each impression there is large number of associated features, including demographic information, browsing history of the user, and context of the web-page. Many other online markets, such as Airbnb, eBay and Etsy also have a similar setting in which products to be sold are highly differentiated. For example, in the case of Aribnb, the products are "stays" and each is characterized by a large number of features including space properties, location, amenities, house rules, as well as arrival dates, events in the area, availability of near-by hotels, etc (Airbnb Documentation, 2015).

Here, we consider a feature-based model that postulates a linear relation between the market value of each product and its feature values. Further, from the firm's perspective, we treat distinct buyers independently, and hereafter focus on a single buyer. Put it formally, we start with the following model for the buyer's valuation:

$$v(x_t) = \langle x_t, \theta \rangle + z_t \,, \tag{1}$$

where $x_t \in \mathbb{R}^d$ denotes the product feature vector, θ represents the buyer's preferences and z_t, $t \ge 1$ are idiosyncratic shocks, referred to as noise, which are drawn independently and identically from a zero mean distribution. For two vectors a, b, we write $\langle a, b \rangle$ to refer to their inner product. Feature vectors x_t are observable, while model parameter θ is a-priori unknown to the firm (seller). Therefore, the buyer's valuation $v(x_t)$ is also hidden from the firm.

Parameters of the above model represents how different features are weighted by the buyer in assessing the product. Considering such model, a firm can use historical sales data to estimate parameters of the valuation model, while concurrently collecting revenue from new sales. In practice, though, the buyer's valuation of a product will change over time and this raises the concern of *perishability* of sales data.

In order to capture this point, we consider a richer model with varying coefficients:

$$v_t(x_t) = \langle x_t, \theta_t \rangle + z_t \,. \tag{2}$$

Model parameters θ_t may change over time and as a result, valuation of a product depends on both the product feature vector and the time index.

We study a dynamic pricing problem, where at each time period t, the firm has a product to sell and after observing the product feature vector x_t, posts a price p_t. If the buyer's valuation is above the posted price, $v_t(x_t) \ge p_t$, a sale occurs and the firm collects a revenue of p_t. If the posted price exceeds the buyer's valuation, $p_t > v_t(x_t)$, no sale occurs. Note that at each step, the firm has access to the previous feedbacks (sale/no sale) from the buyer and can use this information in setting the current price.

In this paper, we will analyze the varying-coefficient model (2) and answer two fundamental questions:

> First, what is the value of knowing the sequence of model parameters θ_t; in other words, what is the expected revenue loss (regret) compared to the clairvoyant policy that knows the parameters of the valuation model in advance? Second, what is a good pricing policy?

The answer to the first question intrinsically depends on the temporal variability in the sequence θ_t. If this variation is very large, then there is not much that can be learnt from previous feedback on the buyer's behavior and the problem turns into a random price experimentation. On the other hand, if all of the parameters θ_t are the same, then this feedback information can be used to learn the model parameters, which in turn helps in setting the future prices. In this case, an algorithm that performs a good balance between price exploration and best-guess pricing (exploitation) can lead to a small regret. In this work, we study this trade-off through a projected stochastic gradient descent algorithm and investigate the effect of variations of the sequence of θ_t on the regret bounds.

Feature-based models have recently attracted interest in dynamic pricing. (Amin et al., 2014) studied a similar model to (1) (without the noise terms z_t), where the features x_t are drawn from an unknown i.i.d distribution. A pricing strategy was proposed based on stochastic gradient descent, which results in a regret of the form $O(T^{2/3}\sqrt{\log T})$. This work also studied the problem of dynamic incentive compatibility in repeated posted-price auctions. Subsequently, (Cohen et al., 2016) studied model (1), wherein the feature vectors x_t are chosen antagonistically by nature and not sampled i.i.d. This work proposes a pricing policy based on the ellipsoid method from convex optimization (Boyd and Vandenberghe, 2004) with a regret bound of $O(d^2 \log(T/d))$, under a low-noise setting. More accurately, the regret scales as $O(d^2 \log(\min\{T/d, 1/\delta\}) + d\delta T)$, where δ measures the noise magnitude: in case of bounded noise, δ represents the uniform bound on noise and in case of gaussian noise with variance σ^2, it is defined as $\delta = 2\sigma\sqrt{\log(T)}$. In (Lobel et al., 2016), the regret bound of this policy was improved to $O(d \log T)$, under the noiseless setting. In (Javanmard and Nazerzadeh, 2016), authors study and highlight the role of the structure of demand curve in dynamic pricing. They introduce model (1), and assume that the feature vectors x_t are drawn i.i.d. from an unknown distribution. Further, motivated by real-world applications, it is assumed that the parameter vector θ is sparse in the sense that only a few of its entries are nonzero. A regularized log-likelihood approach is taken to get an improved regret bound of order $s_0(\log(d)+\log(T))$. We add to this body of work by considering feature-based models for valuation of products whose parameters vary over time.

Time-varying demand environments have also been studied recently by (Keskin and Zeevi, 2016). Explicitly, they consider a firm that sells one type of product to customers that arrive over a time horizon. After setting price p_t, the firm observes demand D_t given by $D_t = \alpha_t + \beta_t p_t + \epsilon_t$, where $\alpha_t, \beta_t \in \mathbb{R}$ are the unknown parameters of the demand model and ϵ_t are the unobserved demand shocks (noise). By contrast, in this work we consider different products, each characterized by a high-dimensional feature vector. Further, the seller only receives a binary feedback (sale/no sale) of the customer's behavior at each step, rather than observing the customer's valuation.

1.1 Organization of paper and our main contributions

The remainder of this paper is structured as follows. In Section 2, we formally define the model and formulate the problem. Technical assumptions and the notion of regret will be discussed in this section. We next propose a pricing policy based on projected stochastic gradient descent (PSGD) applied to the log-likelihood function. At each time period t, it returns an estimate $\widehat{\theta}_t$. The price p_t is then set to the optimal price as if $\widehat{\theta}_t$ was the actual parameter θ_t. We next analyze the regret of our PSGD algorithm. Let $\delta_t = \|\theta_{t+1} - \theta_t\|$ be the variation in model parameters at time period t. In Section 3.1, we consider the setting where the product feature vectors x_t are chosen antagonistically by nature and show that the regret of PSGD algorithm is of order $O(\sqrt{T} + \sum_{t=1}^{T} \sqrt{t}\delta_t)$. Interestingly, this bound is independent of the dimension d, which is a desirable property of our policy for high-dimensional applications. We next, in Section 4, consider a stochastic features model, where the feature vectors x_t are drawn independently from an unknown distribution (cf. Assumption 6). Under this setting, we show that the regret of PSGD is of order $O(d^2 \log T + \sum_{t=1}^{T} t\delta_t/d)$. Note that setting $\delta_t = 0$ corresponds to model (1) and our PSGD pricing obtains a logarithmic regret in T. Section 7 is devoted to the proof of main theorems and the main lemmas are proved in Section 8. Finally, proof of several technical steps are deferred to Appendices.

1.2 Related literature

Our works is at the intersection of dynamic pricing, online optimization and high-dimensional statistics. In the following, we briefly discuss the work most related to ours from these contexts.

Feature-based dynamic pricing. Recent papers on dynamic pricing consider models with features/covariates, motivated in part by new advances in big data technology that allow firms to collect large amount of fine-grained information. In the introduction, we discussed the work (Amin et al., 2014; Javanmard and Nazerzadeh, 2016; Cohen et al., 2016) which are closely related to our setting. Another recent work on feature-based dynamic pricing is (Qiang and Bayati, 2016). In this work, authors consider a model where the seller observes the demand entirely, rather than a binary feedback as in our setting. A greedy iterative least squares (GILS) algorithm is proposed that at each time period estimates the demand as a linear function of price by applying least squares to the set of prior prices and realized demands. The work underscores the role of feature-based approaches and show that they create enough price dispersion to achieve a regret of $O(\log(T))$. This is closely related to the work of (den Boer and Zwart, 2013) and (Keskin and Zeevi, 2014) in dynamic pricing (without demand covariates) that demonstrate the GILS is suboptimal and propose methods to integrate forced price-dispersion with GILS to achieve optimal regret.

Online optimization. This field offers a variety of tools for sequential prediction, where an agent measures its predictive performance according to a series of convex functions. Specifically, there is a sequence of a priori unknown reward functions $f_1, f_2, f_3, \dots$ and an agent must make a sequence of decisions: at each time period t, he selects a point z_t and a loss $f_t(z_t)$ is incurred. Note that the function f_t is not known to agent at step t, but he has access to all previous functions $f_1, \dots, f_{t-1}$. First order methods, like online gradient descent (OGD) or online mirror descent (OMD) only use the gradient of previous function at the selected points, i.e., $\partial f_t(z_t)$. The notion of regret here is defined by comparing the agent with the best fixed comparator (Shalev-Shwartz, 2011).

(Hall and Willett, 2015) proposed dynamic mirror descent that is capable of adapting adapts to a possibly non-stationary environment. In contrast to OMD (Beck and Teboulle, 2003; Shalev-Shwartz, 2011), the notion of regret is defined more generally with respect to the best comparator "sequence".

It is worth noting that the general framework of online learning does not directly apply to our problem. To see this, we define the the loss f_t to be the negative of the revenue obtained in time period t, i.e., $f_t = -p_t \mathbb{I}(p_t \geq v_t)$. Then (i) the loss functions are not convex; (ii) the (first order information) of previous loss functions depend on the corresponding valuations $v_1, \dots, v_{t-1}$ which are never revealed to the seller. That said, we borrow some of the techniques from online optimization in proving our results. (See proof of Lemma 3.)

High-dimensional statistics. Among the work in this area, perhaps the most related one to our setting is the problem of 1-bit compressed sensing (Plan and Vershynin, 2013a,b; Ai et al., 2014; Bhaskar and Javanmard, 2015). In this problem, a set of linear measurements are taken from an unknown vector and the goal is to recover this vector having access to the sign of these measurements (1-bit information). This is related to the dynamic pricing problem on model (1), as the seller observes 1-bit feedback (sale/no sale from previous time periods). However, there are a few important differences between these two problem that are worth noting: 1) In dynamic pricing, the crux of the matter is the decisions (prices) made by the firm. Of course this task entails learning the model parameters and therefore the firm gets into the realm of exploration (learning) and exploitation (earning revenue). By contrast, 1-bit compressed sensing is only a learning task; 2) In dynamic pricing, the prices are set based on the previous (sale/no sale) feedbacks. Therefore, the feedbacks are inherently correlated and this makes the learning task challenging. However, in 1-bit compressed sensing it is assumed that the measurements (and therefore the observed signs) are independent; 3) The majority of work on 1-bit compressed sensing consider an offline setting, while in the dynamic pricing, decision are made in an online manner.

2. Model

We consider a pricing problem faced by a firm that sells products in a sequential manner. At each time period $t = 1, 2, \cdots, T$ the firm has a product to sell and the product is represented by an *observable* vector of features (covariates) $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$. The length of the time horizon, denoted by T, is *unknown* the to the firm and the set $\mathcal{X}$ is bounded.

The product at time t has a market value $v_t = v_t(x_t)$, depending on both t and x_t, which is *unobservable*. At each period t, the firm (seller) posts a price p_t. If $p_t \leq v_t$, a sale occurs, and the firm collects revenue p_t. If the price is set higher than the market value, $p_t > v_t$, no sale occurs and no revenue is generated. The goal of the firm is to design a pricing policy that maximizes the collected revenue.

We assume that the market value of a product is a linear function of its covariates, namely

$$v_t(x_t) = \langle \theta_t, x_t \rangle + z_t \,. \tag{3}$$

Here, θ_t and x_t are d-dimensional and $\{z_t\}_{t \geq 1}$ are idiosyncratic shocks, referred to as noise, which are drawn independently and identically from a zero-mean distribution over $\mathbb{R}$. We denote its cumulative distribution function by F, and the corresponding density by $f(x) = F'(x)$. Note that the noise can account for the features that are not measured. We refer to (Keskin and Zeevi, 2014; den Boer and Zwart, 2014; Qiang and Bayati, 2016) for a similar notion of demand shocks.

The sequence of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ is *unknown* to the firm and it may vary across time. This paper focuses on arbitrary sequences $\boldsymbol{\theta}$ and propose an efficient algorithm whose regret scale gracefully in time and the temporal variability in the sequences of θ_t.

The regret is measured with respect to the clairvoyant policy that knows the sequence $\boldsymbol{\theta}$ in advance. We will formally define the regret in Section 2.2.

Let y_t be the response variable that indicates whether a sale has occurred at period t:

$$y_t = \begin{cases} +1 & \text{if } v_t \ge p_t\,, \\ -1 & \text{if } v_t < p_t\,. \end{cases} \tag{4}$$

Note that the above model for y_t can be represented as the following probabilistic model:

$$y_t = \begin{cases} +1 & \text{with probability } \; 1 - F\left(p_t - \langle \theta_t, x_t\rangle\right) \\ -1 & \text{with probability } \; F\left(p_t - \langle \theta_t, x_t\rangle\right) \end{cases} \tag{5}$$

2.1 Technical assumptions and notations

For a vector v, we write $\|v\|_p$ for the standard ℓ_p norm of a vector v, i.e., $\|v\|_p = (\sum_i |v_i|^p)^{1/p}$. Whenever the subscript p is not mentioned it is deemed as the ℓ_2 norm. For a matrix A, $\|A\|$ denotes its ℓ_2 operator norm. For two vectors a, b, we use the notation $\langle a, b\rangle$ to refer to their inner product.

To simplify the presentation, we assume that $\|x_t\| \le 1$, for all $x_t \in \mathcal{X}$, and $\|\theta_t\| \le W$ for a known constant W. We denote by Θ the d-dimensional ℓ_2 ball of radius W (In fact, we can take Θ to be any convex set that contains parameters θ_t. The size of Θ effects our regret bounds up to a constant factor.)

We also make the following assumption on the distribution of noise F.

Assumption 1 *The function $F(v)$ is strictly increasing. Further, $F(v)$ and $1 - F(v)$ are log-concave in v.*

Log-concavity is a widely-used assumption in the economics literature (Bagnoli and Bergstrom, 2005). Note that if the density f is symmetric and the distribution F is log-concave, then $1-F$ is also log-concave. Assumption 1 is satisfied by several common probability distributions including normal, uniform, Laplace, exponential, and logistic. Note that the cumulative distribution function of all log-concave densities is also log-concave (Boyd and Vandenberghe, 2004).

We use the standard big-O notation. In particular $f(n) = O(g(n))$ if there exists a constant $C > 0$ such that $|f(n)| \le Cg(n)$ for all n large enough. We also use $\mathbb{R}_{\ge 0}$ to refer to the set of non-negative real-valued numbers.

2.2 Benchmark policy and regret minimization

For a pricing policy, we measures its performance via the notion of regret, which is the expected revenue loss compared to an oracle that knows the sequence of model parameters in advance (but not the realizations of $\{z_t\}_{t\ge 1}$).We first characterize this benchmark policy.

Using Eq. (3), the expected revenue from a posted price p is equal to $p \times \mathbb{P}(v_t \ge p) = p(1 - F(p - \theta_t \cdot x_t))$. First order condition for the optimal price $p^*(x_t, \theta_t)$ reads

$$p^*(x_t, \theta_t) = \frac{1 - F\left(p^*(x_t, \theta_t) - \langle \theta_t, x_t\rangle\right)}{f\left(p^*(x_t, \theta_t) - \langle \theta_t, x_t\rangle\right)}\,. \tag{6}$$

To lighten the notation, we drop the arguments x_t, θ_t and denote by p_t^* the optimal price at time t.

We next recall the *virtual valuation* function, commonly used in mechanism design (Myerson, 1981):

$$\varphi(v) \equiv v - \frac{1 - F(v)}{f(v)}\,.$$

Writing Eq. (6) in terms of function φ, we get

$$\langle \theta_t, x_t \rangle + \varphi\left(p_t^* - \langle \theta_t, x_t \rangle\right) = 0\,.$$

In order to solve for p_t^*, we define the pricing function g as follows:

$$g(v) \equiv v + \varphi^{-1}(-v)\,. \tag{7}$$

By Assumption 1, φ is injective and hence g is well-defined. Further, it is easy to verify that g is non-negative. Using the definition of g and rearranging the terms we obtain

$$p_t^* = g(\langle \theta_t, x_t \rangle)\,. \tag{8}$$

The performance metric we use in this paper is the worst-case regret with respect to a clairvoyant policy that knows the sequence $\boldsymbol{\theta}$ in advance. Formally, for a policy π to be the seller's policy that sets price p_t at period t, the worst-case regret is defined over T periods is defined as:

$$\mathsf{Regret}^{\pi}(T) \equiv \sup\left\{\Delta^{\pi}_{\boldsymbol{\theta},\boldsymbol{x}} :\, \theta_t \in \Theta,\, x_t \in \mathcal{X}\right\}\,, \tag{9}$$

where for $T \geq 1$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_T)$ and $\boldsymbol{x} = (x_1, x_2, \dots, x_T)$,

$$\Delta^{\pi}_{\boldsymbol{\theta},\boldsymbol{x}}(T) = \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}}\left[\sum_{t=1}^{T}\left(p_t^*\mathbb{I}(v_t \geq p_t^*) - p_t\mathbb{I}(v_t \geq p_t)\right)\right]\,. \tag{10}$$

Here the expectation $\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}}$ is with respect to the distributions of idiosyncratic noise, z_t. Note that v_t, p_t, and p_t^* depend on $\boldsymbol{\theta}$ and $\boldsymbol{x}$.

3. Pricing policy

Our dynamic pricing policy consists of a projected gradient descent algorithm to predict parameters $\widehat{\theta}_t$. With each new product, it computes the negative gradient of the loss and shirts its prediction in that direction. The result is projected onto set Θ to produce the next prediction. The policy then sets the prices as $p_t = g(\langle x_t, \widehat{\theta}_t \rangle)$. Note that by Eq. (7), p_t is the optimal price if $\widehat{\theta}_t$ was the true parameter θ_t. Also, by log-concavity assumption on F and $1 - F$, the function $\ell_t(\theta)$ is convex.

In projected gradient descent, the sequence of step sizes $\{\eta_t\}_{t\geq 1}$ is an arbitrary sequence of non-increasing values. In Sections 3.1 and 4, we analyze the regret of our pricing policy and provide guidelines for choosing step sizes.

PSGD (Projected stochastic gradient descent) pricing policy

Input: (at time 0**)** function g, set Θ,
Input: (arrives over time) covariate vectors $\{x_t\}_{t\in\mathbb{N}}$
Output: prices $\{p_t\}_{t\in\mathbb{N}}$

1: $p_1 \leftarrow 0$ and initialize $\widehat{\theta}_1 \in \Theta$
2: **for** $t = 1, 2, 3, \dots$ **do**
3: Set $\widehat{\theta}_{t+1}$ according to the following rule:

$$\widehat{\theta}_{t+1} = \Pi_\Theta(\widehat{\theta}_t - \eta_t \nabla \ell_t(\widehat{\theta}_t)) \tag{11}$$

with

$$\ell_t(\theta) = -\mathbb{I}(y_t = 1)\log(1 - F(p_t - \langle x_t, \theta\rangle)) - \mathbb{I}(y_t = -1)\log(F(p_t - \langle x_t, \theta\rangle)) \tag{12}$$

4: Set price p_{t+1} as

$$p_{t+1} \leftarrow g(\langle x_{t+1}, \widehat{\theta}_{t+1}\rangle) \tag{13}$$

3.1 Regret analysis

We first define a few useful quantities that appear in our regret bounds. Define

$$M \equiv W + \varphi^{-1}(0)\,, \tag{14}$$

$$u_M \equiv \sup_{|x|\le M}\left\{\max\left\{-\frac{\mathrm{d}}{\mathrm{d}x}\log F(x), -\frac{\mathrm{d}}{\mathrm{d}x}\log(1-F(x))\right\}\right\}, \tag{15}$$

$$\ell_M \equiv \inf_{|x|\le M}\left\{\min\left\{-\frac{\mathrm{d}^2}{\mathrm{d}x^2}\log F(x), -\frac{\mathrm{d}^2}{\mathrm{d}x^2}\log(1-F(x))\right\}\right\}, \tag{16}$$

where the derivatives are with respect to x. We note that M is an upper-bound on the maximum price offered and also, by the log-concavity property of F and $1-F$, we have

$$u_M = \max\left\{-\frac{\mathrm{d}}{\mathrm{d}x}\log F(-M), -\frac{\mathrm{d}}{\mathrm{d}x}\log(1-F(M))\right\}.$$

Further, by log-concavity property of F and $1-F$, we have $\ell_M > 0$.

We also let $B = \max_v f(v)$ and $B' = \max_v f'(v)$, respectively denote the maximum value of the density function f and the its derivative f'.

The following theorem bounds the regret of our PSGD policy.

Theorem 2 *Consider model* (3) *for the product market values and let Assumption 1 hold. Set* $M = 2W + \varphi^{-1}(0)$*, with* φ *being the virtual valuation function w.r.t distribution* F*. Then, the regret of PSGD pricing policy using a non-increasing sequence of step sizes* $\{\eta_t\}_{t\ge1}$ *is bounded as follows:*

$$\mathsf{Regret}(T) \le \frac{2(2B+MB')}{\ell_M}\max\left\{\frac{16}{\ell_M}\log T,\, \frac{2W^2}{\eta_{T+1}} + \frac{u_M^2}{2}\sum_{t=1}^T \eta_t + 2W\sum_{t=1}^T\frac{\delta_t}{\eta_t}\right\} + \frac{M}{T}\,, \tag{17}$$

where $\delta_t \equiv \|\theta_{t+1} - \theta_t\|$.

In particular, if $\eta_t \propto \frac{1}{\sqrt{t}}$, then there exists a constant $C = C(B, M, W, \ell_M, u_M) > 0$, independent of T, such that

$$\mathsf{Regret}(T) \le C\Big(\sqrt{T} + \sum_{t=1}^{T} \sqrt{t}\delta_t\Big)\,. \tag{18}$$

At the core of our regret analysis (proof of Theorem 2) is the following Lemma that provides a prediction error bound for the customer's valuations.

Lemma 3 *Consider model* (3) *for the product market values and let Assumption 1 hold. Set $M = 2W + \varphi^{-1}(0)$, with φ being the virtual valuation function w.r.t distribution F. Let $\{\widehat{\theta}_t\}_{t\ge1}$ be generated by PSGC pricing policy, using a non-increasing positive series $\eta_{t+1} \le \eta_t$. Then, with probability at least $1 - \frac{1}{T^2}$ the following holds true:*

$$\sum_{t=1}^{T} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \le \frac{4}{\ell_M} \max\Bigg\{\frac{16}{\ell_M}\log T,\; \frac{2W^2}{\eta_1} + \sum_{t=1}^{T}\Big(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\Big)\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2 + \frac{u_M^2}{2}\sum_{t=1}^{T}\eta_t + 2W\sum_{t=1}^{T}\frac{\delta_t}{\eta_t}\Bigg\}, \tag{19}$$

where u_M, ℓ_M are given by Equations (15), (16)*, respectively.*

Lemma 3 is presented in a form that can also be used in proving our next results under the stochastic features model. For proving Theorem 2, we simplify bound (19) as follows. Given that $\theta_{t+1}, \widehat{\theta}_{t+1} \in \Theta$, we have $\|\theta_{t+1} - \widehat{\theta}_{t+1}\| \le 2W$. Using the non-increasing property of sequence η_t, we write

$$\frac{2W^2}{\eta_1} + \sum_{t=1}^{T}\left(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\right)\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2 \le \frac{2W^2}{\eta_1} + \sum_{t=1}^{T}\left(\frac{2W^2}{\eta_{t+1}} - \frac{2W^2}{\eta_t}\right) \le \frac{2W^2}{\eta_{T+1}}\,.$$

Therefore, bound (19) simplifies to:

$$\sum_{t=1}^{T} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \le \frac{4}{\ell_M} \max\left\{\frac{16}{\ell_M}\log T, \frac{2W^2}{\eta_{T+1}} + \frac{u_M^2}{2}\sum_{t=1}^{T}\eta_t + 2W\sum_{t=1}^{T}\frac{\delta_t}{\eta_t}\right\}, \tag{20}$$

The regret bound (17) is derived by relating regret at each time period to the prediction error at that time. We refer to Section 7 for the proof of Theorem 2.

Remark 4 *The regret bound* (17) *does not depend on the dimension d, which makes our pricing policy desirable for high-dimensional applications. Also, note that the temporal variation δ_t appears in our bound with coefficient $\sqrt{t}$. Therefore, variations at later times are more impactful on the regret of PSGD pricing policy. This is expected because at later times, the pricing policy is more relied on the accumulated information about the valuation model and an abrupt change in the model parameters can make this information worthless. On the other side, temporal changes at the beginning steps are not that effective since the policy is still experimenting different prices to learn the customer's behavior.*

Remark 5 *While the regret bound is dimension-free, the computational complexity of PSGD pricing policy scales with dimension d. Specifically, the complexity of each step is $O(d)$. To see this, we note that the gradient $\nabla \ell_t(\theta)$ can be computed in $O(d)$ by Equations* (70) *and* (71). *Projection onto set Θ (ℓ_2 projection) is also $O(d)$.*

4. Stochastic features model

In Theorem 2, we showed that our PSGD pricing policy achieves regret of order $O(\sqrt{T} + \sum_{t=1}^{T} \sqrt{t}\delta_t)$. Let us point out that in Theorem 2 the arrivals (feature vectors x_t) are modeled as adversarial. In this section, we assume that features x_t are independent and identically distributed according to a probability distribution on $\mathbb{R}^d$. Under such stochastic model, we show that the regret of PSGD pricing scales at most of order $O(d^2 \log T + \sum_{t=1}^{T} t\delta_t/d)$.

We proceed by formally defining the stochastic features model.

Assumption 6 *(*Stochastic features model*). Feature vectors x_t are generated independently according to a probability distribution $\mathbb{P}_{\boldsymbol{x}}$, with a bounded support in $\mathbb{R}^d$. We denote by Σ the covariance matrix of distribution $\mathbb{P}_{\boldsymbol{x}}$ and assume that Σ has bounded eigenvalues. Specifically, there exist constants $C_{\min}$ and $C_{\max}$ such that for every eigenvalue σ of Σ, we have $0 < \frac{1}{d} C_{\min} \le \sigma < \frac{1}{d} C_{\max}$.*

Without loss of generality and to simplify the presentation, we assume that $\mathbb{P}_{\boldsymbol{x}}$ is supported on the unit ℓ_2 ball in $\mathbb{R}^d$. The rationale behind the above assumption on the scaling of eigenvalues is that $\text{Trace}(\Sigma) = \mathbb{E}(\|x_t\|^2) \le 1$. Therefore, the assumption above on the eigenvalues of Σ states that all the eigenvalues are of the same order.

Under the stochastic features model, we define the notion of worst-case regret as follows. For a policy π be the seller's policy that sets price p_t at period t, the T-period regret is defined as:

$$\mathsf{Regret}^{\pi}(T) \equiv \sup \left\{ \Delta^{\pi}_{\boldsymbol{\theta}, \mathbb{P}_{\boldsymbol{x}}} : \, \theta_t \in \Theta, \, \mathbb{P}_{\boldsymbol{x}} \in Q \right\} , \tag{21}$$

where Q denotes the set of probability distribution supported on ℓ_2 unit ball satisfying Assumption 6 (bounded eigenvalues). Further, for $T \ge 1$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_T)$ and probability measure $\mathbb{P}_{\boldsymbol{x}}$, we define

$$\Delta^{\pi}_{\boldsymbol{\theta}, \mathbb{P}_{\boldsymbol{x}}}(T) = \mathbb{E}_{\boldsymbol{\theta}, \mathbb{P}_{\boldsymbol{x}}} \left[\sum_{t=1}^{T} \Big(p_t^* \mathbb{I}(v_t \ge p_t^*) - p_t \mathbb{I}(v_t \ge p_t) \Big) \right] . \tag{22}$$

where the expectation is with respect to the distributions of idiosyncratic noise, z_t, and $\mathbb{P}_{\boldsymbol{x}}$, the distribution of feature vectors. Note the subtle difference with definition (9), in that the worst case is computed over Q rather than $\mathcal{X}$.

We propose a similar PSGD pricing policy for this setting, with a specific choice of the step sizes. Ideally, we want to set $\eta_t = 6/(\ell_M C t)$, where C is an arbitrary fixed constant such that $0 < C < \sigma_{\min}$, with $\sigma_{\min}$ being the minimum eigenvalue of population covariance Σ. Of course, Σ is unknown and therefore we proceed as follows. We let $Q_t = (1/t) \sum_{\ell=1}^{t} x_\ell x_\ell^\mathsf{T}$ be the empirical covariance based on the first t features. Denote by σ_t the minimum eigenvalue of Q_t. We then use the sequence σ_t, and set the step size η_t as

$$\eta_t = \frac{1}{\lambda_t \cdot t} , \qquad \lambda_t = \frac{\ell_M}{6} \left\{ \frac{1}{t} \Big(1 + \sum_{\ell=1}^{t} \sigma_\ell \Big) \right\} .$$

PSGD pricing policy for stochastic features model

Input: (at time 0**)** function g, set Θ,
Input: (arrives over time) covariate vectors $\{x_t\}_{t\in\mathbb{N}}$
Output: prices $\{p_t\}_{t\in\mathbb{N}}$

$p_1 \leftarrow 0$ and initialize $\widehat{\theta}_1 \in \Theta$
$Q_1 \leftarrow x_1 x_1^\mathsf{T}$
for $t = 1, 2, 3, \dots$ **do**
  Define σ_t as the minimum eigenvalue of Q_t.
  Set

$$\lambda_t = \frac{\ell_M}{6t}(1 + \sum_{\ell=1}^{t} \sigma_\ell)\,. \tag{23}$$

  Set

$$\eta_t = \frac{1}{\lambda_t \cdot t} \tag{24}$$

  Set $\widehat{\theta}_{t+1}$ according to the following rule:

$$\widehat{\theta}_{t+1} = \Pi_\Theta(\widehat{\theta}_t - \eta_t \nabla \ell_t(\widehat{\theta}_t)) \tag{25}$$

  with

$$\ell_t(\theta) = -\mathbb{I}(y_t = 1)\log(1 - F(p_t - \langle x_t, \theta\rangle)) - \mathbb{I}(y_t = -1)\log(F(p_t - \langle x_t, \theta\rangle)) \tag{26}$$

  $Q_{t+1} \leftarrow (\frac{t}{t+1})Q_t + (\frac{1}{t+1})x_{t+1}x_{t+1}^\mathsf{T}$
  Set price p_{t+1} as

$$p_{t+1} \leftarrow g(\langle x_{t+1}, \widehat{\theta}_{t+1}\rangle) \tag{27}$$

Description of the PSGD pricing policy is given in Table above.

4.1 Logarithmic regret bound

The following theorem bounds the regret of our dynamics pricing policy.

Theorem 7 *Consider model* (3) *for the product market values and suppose Assumption 1 holds. Let* $M = 2W + \varphi^{-1}(0)$*, with* φ *being the virtual valuation function w.r.t distribution* F*. Under the stochastic features model (Assumption 6), the regret of PSGD pricing policy is bounded as follows:*

$$\mathsf{Regret}(T) \le C_1 d^2 \log T + C_2 \sum_{t=1}^{T} \frac{t}{d}\delta_t\,, \tag{28}$$

where $\delta_t \equiv \|\theta_{t+1} - \theta_t\|$ *and* C_1, C_2 *are constants that depend on* $C_{\max}, C_{\min}, u_M, \ell_M, M, B, W$ *but are independent of dimension* d.

Proof of Theorem 7 relies on the following lemma that is analogous to Lemma 3 and establishes a prediction error bound for the customer's valuations.

Lemma 8 *Consider model* (3) *for the product market values and the stochastic features model (Assumption 6). Suppose that Assumption 1 holds and set* $M = 2W + \varphi^{-1}(0)$*, with* φ *being the virtual valuation function w.r.t distribution* F*. Let* $\{\widehat{\theta}_t\}_{t\ge 1}$ *be generated by PSGD pricing policy. Then,*

$$C_{\min} \sum_{t=1}^{T} \mathbb{E}(\|\theta_t - \widehat{\theta}_t\|^2) \le \left[\frac{128}{\ell_M^2} + \frac{24 u_M^2}{\ell_M^2}\left(\tilde{c} + \frac{4}{C_{\min} d}\right)\right] \cdot d^3 \log T$$
$$+ 8W^2 d \left(\frac{1}{T} + \frac{12}{\ell_M^2} + \frac{1}{c_2 d}\right) + 4W \sum_{t=1}^{T} t\delta_t \,.$$

Here $\sigma_{\min}$ *denotes the minimum eigenvalue of covariance* Σ*. (See Assumption 6.)*

4.2 A lower bound on regret

In this section, we provide a theoretical lower bound on the minimum achievable regret of any pricing policy under the stochastic features model. Prior to that, we need to adopt a few notations.

For a given time horizon T and a sequence of valuations parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_T)$, let

$$V_{\boldsymbol{\theta}}(T) \equiv \sum_{t=1}^{T} t\|\theta_{t+1} - \theta_t\| \,. \tag{29}$$

We also define, for $\nu \in [1/2, 2]$,

$$\mathcal{V}(T, B, \nu) \equiv \{\boldsymbol{\theta} :\, \theta_t \in \Theta,\, V_{\boldsymbol{\theta}}(T) \le BdT^{\nu}\} \,. \tag{30}$$

By assuming $\boldsymbol{\theta} \in \mathcal{V}(T, B, \nu)$ for all T, we are assuming that nature has a finite temporal variation budget to use in changing the valuation parameters throughout the time horizon. Of course, different variation metrics can be considered such as total variation $\sum_{t=1}^{T} \delta_t$ or the maximum temporal variation $\sup_{1\le t\le T} \delta_t$ and the performance of a pricing policy can be studied under different variation budget constraints. The specific choice of (29) is putting higher weights at later variations in the sequence $\boldsymbol{\theta}$ and is reasonable for applications where one expects the buyer's preferences (valuation parameters) become stable over time. Note that designing favorable pricing policy for applications with gradual changes in buyer's preferences is more challenging than that for environments with bursty changes. This might look counterintuitive at first glance because at any time, the accumulated information about valuations can become useless by an abrupt change in the valuation model. However, as noticed and analyzed in (Keskin and Zeevi, 2016), this is not that case because, intuitively, gradual changes can be undetectable and lead to significant revenue loss, while for bursty changes, the policy can be designed in a way to detect the changes and reset its estimate of the valuation model after each change to avoid large estimation error and revenue loss. For a pricing policy π, consider the T-period regret, defined as

$$\mathsf{Regret}^{\pi}(T, B, \nu) \equiv \max\left\{\Delta^{\pi}_{\boldsymbol{\theta},\mathbb{P}_{\boldsymbol{x}}}(T) : \boldsymbol{\theta} \in \mathcal{V}(T, B, \nu), \, \mathbb{P}_{\boldsymbol{x}} \in Q\right\} \tag{31}$$

where we recall that

$$\Delta^{\pi}_{\boldsymbol{\theta},\mathbb{P}_{\boldsymbol{x}}}(T) \equiv \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\theta},\mathbb{P}_{\boldsymbol{x}}}\left(p^*_t \mathbb{I}(v_t \geq p^*_t) - p_t \mathbb{I}(v_t \geq p_t)\right). \tag{32}$$

Note that this is the same regret notion defined in (21), where we just make the variation budget constraint explicit in the notation.

Rephrasing the statement of Theorem 7, for PSGD pricing policy we have $\mathsf{Regret}^{\pi}(T, B, \nu) \leq C_1 d^2 \log T + C_2 B T^{\nu}$. We next provide a lower bound on the regret of any pricing policy. Indeed this lower bound applies to a powerful clairvoyant who fully observes the market values after the price is either accepted or rejected.

Theorem 9 *Consider linear model* (3) *where the market values $v_t(x_t)$, $1 \leq t \leq T$, are fully observed. We further assume that market value noises are generated as $z_t \sim \mathsf{N}(0, \sigma^2)$. There exists a constant c, depending on σ, $C_{\max}$, such that*

$$\mathsf{Regret}^{\pi}(T, B, \nu) \geq c \min\left(\left(B^2 d T^{2\nu-1}\right)^{1/3}, T/d\right),$$

for any pricing policy π and time horizon T.

The high-level intuition behind this result is that the nature can change the valuation parameters in a gradual manner such that the seller should pay a revenue loss in order to detect the changes and learn the new valuation parameter after a change. To be more specific, we divide the time horizon into cycles of length N periods, where N is of order $(T^{4-2\nu}/d)^{1/3}$ and consider a setting where the value of θ_t can change to one of two options θ^0, θ^1, only in the first period of a cycle. We choose the parameter change $\delta = \|\theta^1 - \theta^0\|$ of order $\sqrt{d/N}$ to ensure that (i) no policy can identify the change without incurring a revenue loss of order $N\delta^2/d$ (ii) The variation metric $V_{\boldsymbol{\theta}}(T)$ remains below the allowable limit of BdT^{ν}. Therefore, the total regret over T periods works out at $T\delta^2/d$. In particular, for proving point (i) we quantify the likelihood of valuations under the probability measures corresponding to θ^0 and θ^1, using Kullback-Leibler divergence. We use Pinsker inequality form probability theory and hypothesis testing results from information theory to show that there is a significant probability of not detecting the (potential) change, which consequently yields a revenue loss of order $N\delta^2/d$, over each cycle.

We refer to Section 7.3 for the proof of Theorem 9.

5. Numerical experiments

We numerically study the performance of our PSGD pricing policy on synthetic data. In our experiments, we set $W = 5$ and set $\theta_1 = (W/2)(Z/\|Z\|)$, with $Z \sim \mathsf{N}(0, \mathrm{I}_d)$ a multivariate normal variable. We then generate a sequence of parameters θ_t as follows:

$$\theta_{t+1} = \theta_t + r_t \,,$$

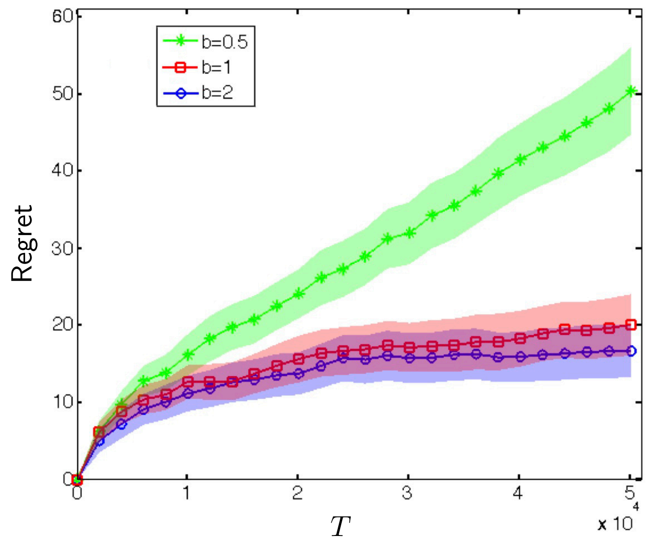

Figure 1: Cumulative regret of PSGD pricing policy for the synthetic data in Section 5. Temporal variations are $\delta_t = t^{-b}$ and the curves are obtained by averaging across 80 trials. Shaded region around each curve is the 95% confidence interval.

where $r_t = t^{-b}(\tilde{Z}/\|\tilde{Z}\|)$, with $\tilde{Z} \sim \mathsf{N}(0, \mathrm{I}_d)$. Note that $\delta_t = \|\theta_{t+1} - \theta_t\| = \|r_t\| = t^{-b}$.

Next, at each time t, product covariates x_t are independently sampled from a Gaussian distribution $\mathsf{N}(0, \mathrm{I}_d)$ and normalized so that $\|x_t\| = 1$. Further, the market shocks are generated as $z_t \sim \mathsf{N}(0, \sigma^2)$, with $\sigma = 1$. We run the PSGD pricing policy for stochastic features model.

Results. Figure 1 compares the cumulative regret (averaged over 80 trials) of the PSGD policy, for $b = 0.5, 1, 2$, on the aforementioned synthetic data for $T = 50,000$ steps. The shaded region around each curve correspond to the 95% confidence interval across the 80 trials. As expected, increase in b results in larger temporal variations and larger regret.

To better understand the behavior of regret for different values of b, we plotted the regret bounds in various scales in Figure 2. For $b = 0.5$, we have $\mathsf{Regret}(T) \sim T^{2/3}$, and for $b = 1, 2$, we have $\mathsf{Regret} \sim \log(T)$. Comparing with Theorem 7, we see that the empirical regret in case of $b = 0.5$, 1, is smaller than the upper bound given by Equation (28), order-wise. However, it is worth noting that bound given in Theorem 7 applies to any adversarial choice of temporal variations r_t, while in our experiments we generated these terms independently at random.

6. Extension to nonlinear model

Throughout the paper, we exclusively focused on linear models for buyer's valuation with varying coefficients. In order to generalize our results to nonlinear models, we consider a setting where the market value of a product with feature vector x_t is given by

$$v_t(x_t) = \psi(\langle x_t, \theta_t \rangle + z_t)\,. \tag{33}$$

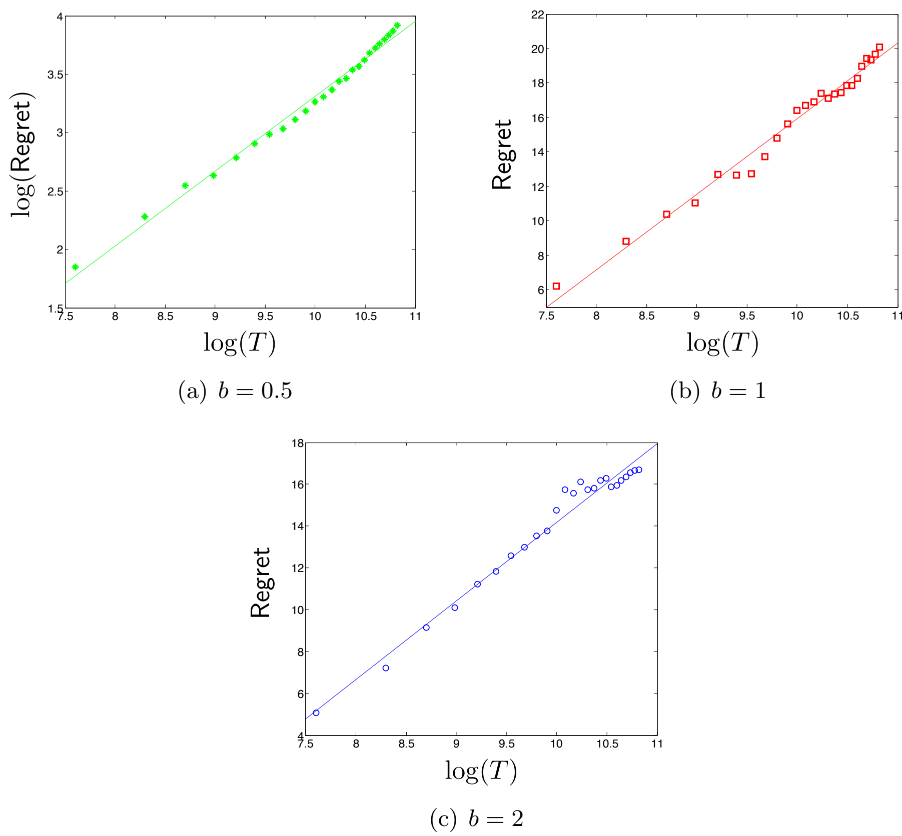

Figure 2: Cumulative regrets of PSGD for different values of b. For $b = 0.5$, $\mathsf{Regret}(T) \sim T^{2/3}$; for $b = 1, 2$, $\mathsf{Regret}(T) \sim \log(T)$.

This model is often referred to as generalized linear model and captures nonlinear dependencies on features to some extent. We assume that the link function $\psi : \mathbb{R} \mapsto \mathbb{R}$ is a general log-concave function and is strictly increasing.

We next compute the pricing function. Since ψ is strictly increasing, the expected revenue at a price p amounts to $p\left(1 - F\left(\psi^{-1}(p) - \langle x_t, \theta_t\rangle\right)\right)$. First order condition for the optimal price $p^*_t(x_t)$ reads as

$$\psi'(\psi^{-1}(p^*_t)) = \frac{pf\left(\psi^{-1}(p^*_t) - \langle x_t, \theta_t\rangle\right)}{1 - F\left(\psi^{-1}(p^*_t) - \langle x_t, \theta_t\rangle\right)}\,. \tag{34}$$

Define $\lambda(v) = f(v)/(1 - F(v))$ the hazard rate function for distribution F, and let $\tilde{p} = \psi^{-1}(p)$. Writing (34) in terms of λ function, we get

$$\langle x_t, \theta_t\rangle = \tilde{p}^*_t - \lambda^{-1}\left(\frac{\psi'(\tilde{p}^*_t)}{\psi(\tilde{p}^*_t)}\right)\,. \tag{35}$$

For real-valued v, define

$$g_\psi^{-1}(v) \equiv v - \lambda^{-1}\left(\frac{\psi'(v)}{\psi(v)}\right)\,. \tag{36}$$

Note that by log-concavity of $1 - F$, the hazard function λ is increasing. Also, by log-concavity of ψ, the term $\frac{\mathrm{d}}{\mathrm{d}v}\log\psi(v) = \psi'(v)/\psi(v)$ is decreasing. Putting these together, we obtain that $-\lambda^{-1}(\psi'(v)/\psi(v))$ is increasing. Therefore, the right-hand side of (36) is strictly increasing and the function g_ψ is well-defined. Invoking Equation (35), we derive the optimal price as

$$p^*_t = \psi\left(g_\psi(\langle x_t, \theta_t\rangle)\right)\,. \tag{37}$$

As noted before, since ψ is increasing, at each period t, a sale happens if $z_t \ge \psi^{-1}(p_t) - \langle x_t, \theta_t\rangle$. Hence, the log-likelihood function reads as

$$\ell_t(\theta) = -\mathbb{I}(y_t = 1)\log\left(1 - F\left(\psi^{-1}(p_t) - \langle x_t, \theta\rangle\right)\right) - \mathbb{I}(y_t = -1)\log\left(F\left(\psi^{-1}(p_t) - \langle x_t, \theta\rangle\right)\right)\,. \tag{38}$$

In PSGD pricing policy, we run gradient step with this log-likelihood function and then set price p_{t+1} at next step as $p_{t+1} = \psi\left(g_\psi(\langle x_{t+1}, \theta_{t+1}\rangle)\right)$.

The results on the regret of PSGD pricing policy carries over to the generalized linear model as well. The analysis goes along the same lines and is omitted.

7. Proof of main theorems

7.1 Proof of Theorem 2

Lemma 10 *Set $M = 2W + \varphi^{-1}(0)$, and for $\theta \in \Theta$ define $u_t(\theta) = p_t - \langle x_t, \theta\rangle$, where $p_t = g(\langle x_t, \widehat{\theta}_t\rangle)$ is the posted price at time t. Then $|u_t(\theta)| \le M$ for all $t \ge 1$.*

Define function $h(;u)$ from $\mathbb{R}_{\ge 0}$ to $\mathbb{R}_{\ge 0}$ as

$$h(p;u) = p(1 - F(p - u))$$

This is the expected revenue at price p when the noiseless valuation is u, i.e., $\langle x_t, \theta_t \rangle = u$. We let

$$R_t \equiv p_t^* \mathbb{I}(v_t \geq p_t^*) - p_t \mathbb{I}(v_t \geq p_t) \tag{39}$$

be the regret incurred at time t, and define $\mathcal{F}_t$ as the history up to time t (Formally, $\mathcal{F}_t$ is the σ-algebra generated by market noise $\{z_\ell\}_{\ell=1}^t$.) Then,

$$\mathbb{E}(R_t|\mathcal{F}_{t-1}) = p_t^* \mathbb{P}(v_t \geq p_t^*) - p_t \mathbb{P}(v_t \geq p_t) = h(p_t^*; \langle x_t, \theta_t \rangle) - h(p_t; \langle x_t, \widehat{\theta}_t \rangle) \,. \tag{40}$$

The optimal price p_t^* is the maximizer of $h(p; \langle x_t, \theta_t \rangle)$ and thus $h'(p_t^*; \langle x_t, \theta_t \rangle) = 0$. By Taylor expansion of function h, there exists a value p between p_t and p_t^*, such that,

$$h(p_t; \langle x_t, \theta_t \rangle) - h(p_t^*; \langle x_t, \theta_t \rangle) = \frac{1}{2} h''(p; \langle x_t, \theta_t \rangle)(p_t - p_t^*)^2 \,. \tag{41}$$

We next show that $|h''(p; \langle x_t, \theta_t \rangle)| \leq C$ with $C = 2B + MB'$. Recall that $B = \max_v f(v)$ and $B' = \max_v f'(v)$. To see this, we write

$$|h''(p; \langle x_t, \theta_t \rangle)| = \Big| 2f(p - \langle x_t, \theta_t \rangle) + pf'(p - \langle x_t, \theta_t \rangle) \Big| \leq 2B + MB' \,. \tag{42}$$

Putting Equations (39), (41), (42) and using the 1-Lipschitz property of price function g, we conclude:

$$\begin{aligned} \mathbb{E}[R_t|\mathcal{F}_{t-1}] &= h(p_t^*; \langle x_t, \theta_t \rangle) - h(p_t; \langle x_t, \widehat{\theta}_t \rangle) \leq \frac{2B + MB'}{2}(p_t - p_t^*)^2 \\ &= \frac{2B + MB'}{2} \Big(g(\langle x_t, \widehat{\theta}_t \rangle) - g(\langle x_t, \theta_t \rangle) \Big)^2 \leq \frac{2B + MB'}{2} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \end{aligned} \tag{43}$$

To ease the presentation, define the shorthand

$$A(T) \equiv \frac{4}{\ell_M} \max \left\{ \frac{16}{\ell_M} \log T, \, \frac{2W^2}{\eta_{T+1}} + \frac{u_M^2}{2} \sum_{t=1}^T \eta_t + 2W \sum_{t=1}^T \frac{\delta_t}{\eta_t} \right\} .$$

We further let $\mathcal{G}$ be the probabilistic event that $\sum_{t=1}^T \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \leq A(T)$. Employing Lemma 3 and using the fact that $\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2 \leq 4W^2$, we obtain that $\mathbb{P}(\mathcal{G}) \geq 1 - \frac{1}{T^2}$.

We continue by bounding $E(R_t)$ as follows:

$$\begin{aligned} \mathbb{E}[R_t] &= \mathbb{E}[\mathbb{E}[R_t|\mathcal{F}_{t-1}]] = \mathbb{E}\Big[\mathbb{E}[R_t|\mathcal{F}_{t-1}] \cdot \Big(\mathbb{I}(\mathcal{G}) + \mathbb{I}(\mathcal{G}^c) \Big) \Big] \\ &= \frac{2B + MB'}{2} \mathbb{E}\Big[\langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \cdot \mathbb{I}(\mathcal{G}) \Big] + M\mathbb{P}(\mathcal{G}^c) \,. \end{aligned}$$

Consequently,

$$\mathsf{Regret}(T) \leq \sum_{t=1}^T \mathbb{E}[R_t] \leq \frac{2B + MB'}{2} \mathbb{E}\Big[\sum_{t=1}^T \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \cdot \mathbb{I}(\mathcal{G}) \Big] + MT \, \mathbb{P}(\mathcal{G}^c) \leq \frac{2B + MB'}{2} A(T) + \frac{M}{T} \,.$$

The proof is complete.

7.2 Proof of Theorem 7

Proof of Theorem 7 follows along the same lines as proof of Theorem 2. Let $\tilde{\mathcal{F}}_t$ be the σ-algebra generated by market noises $\{z_\ell\}_{\ell=1}^t$ and feature vectors $\{x_\ell\}_{\ell=1}^t$. Further, let $\mathcal{F}_t$ be the σ-algebra generated by $\tilde{\mathcal{F}}_t \cup \{x_{t+1}\}$. For term R_t defined by (39) and following the chain of inequalities as in (43),

$$\mathbb{E}[R_t|\mathcal{F}_{t-1}] \le \frac{2B+MB'}{2}\,\langle x_t, \theta_t - \widehat{\theta}_t\rangle^2\,. \tag{44}$$

For brevity in notation, let $\bar{B} = (2B+MB')/2$. Since, $\mathcal{F}_t \supseteq \tilde{\mathcal{F}}_t$, by iterated law of iteration,

$$\mathbb{E}(R_t|\tilde{\mathcal{F}}_{t-1}) = \mathbb{E}(\mathbb{E}(R_t|\mathcal{F}_{t-1})|\tilde{\mathcal{F}}_{t-1}) \le \bar{B}\langle \theta_t - \widehat{\theta}_t, \Sigma(\theta_t - \widehat{\theta}_t)\rangle \le \frac{1}{d}\bar{B}C_{\max}\|\theta_t - \widehat{\theta}_t\|^2 \tag{45}$$

Applying Lemma 8, we get

$$\begin{aligned}\mathsf{Regret}(T) &\le \sum_{t=1}^T \mathbb{E}[R_t] \le \frac{1}{d}\bar{B}C_{\max}\sum_{t=1}^T \mathbb{E}(\|\theta_t - \widehat{\theta}_t\|^2)\\ &\le \bar{B}\frac{C_{\max}}{C_{\min}}\left[\frac{128}{\ell_M^2} + \frac{24u_M^2}{\ell_M^2}\left(\tilde{c} + \frac{4}{C_{\min}d}\right)\right]\cdot d^2\log T\\ &+ 8W^2\bar{B}\frac{C_{\max}}{C_{\min}}\left(\frac{1}{T} + \frac{12}{\ell_M^2} + \frac{1}{c_2 d}\right) + \bar{B}\frac{C_{\max}}{C_{\min}}\Big(\frac{4W}{d}\Big)\sum_{t=1}^T t\delta_t\,.\end{aligned}$$

The result follows by taking

$$\begin{aligned}C_1 &= \bar{B}\frac{C_{\max}}{C_{\min}}\left[8W^2\left(\frac{1}{T} + \frac{12}{\ell_M^2} + \frac{1}{c_2 d}\right) + \frac{128}{\ell_M^2} + \frac{24u_M^2}{\ell_M^2}\left(\tilde{c} + \frac{4}{C_{\min}d}\right)\right],\\ C_2 &= 4W\bar{B}\frac{C_{\max}}{C_{\min}}\,.\end{aligned}$$

7.3 Proof of Theorem 9

The proof methodology is similar to the proof of (Keskin and Zeevi, 2016, Theorem 1).

We first propose a setting for constructing the sequence of valuation parameters $\boldsymbol{\theta} = (\theta_1,\dots,\theta_T)$. Divide the time horizon into cycles of length $N = \lceil m_0 T^{(4-2\nu)/3}\rceil$, where $m_0 = (\frac{\sigma^2}{C_{\max}B^2 d})^{1/3}$. Consider a setting wherein the noise markets are generated as $z_t \sim \mathsf{N}(0,\sigma^2)$ and the value of θ_t can change only in the first period of a cycle, taking one of the two values $\{\theta^0,\theta^1\}$. Here, $\theta^0,\theta^1 \in \mathbb{R}^d$ are two arbitrary vectors such that $\|\theta^0 - \theta^1\| = \delta$, with $\delta = \min(\sigma\sqrt{d/(C_{\max}N)}, \sqrt{c_2})$. Note that for this sequence of $\boldsymbol{\theta}$, we have

$$V_{\boldsymbol{\theta}}(T) \le \sum_{k=1}^{\lceil T/N\rceil}(kN)\delta \le \frac{T^2}{N}\delta \le BdT^\nu \tag{46}$$

We consider a clairvoyant who fully observes the market values $v_t(x_t)$. Focus on a single cycle and let $\mathbb{P}_0^\pi$ (resp. $\mathbb{P}_1^\pi$) denote the probability distribution of valuations $(v_1,v_2,\dots,v_N)$

when all the parameters θ_t are equal to θ^0 (resp. θ^1), for $1 \le t \le N$. The KL divergence between $\mathbb{P}_0^\pi$ and $\mathbb{P}_1^\pi$ amounts to

$$D_{\rm KL}(\mathbb{P}_0^\pi, \mathbb{P}_1^\pi) \equiv \mathbb{E}_0^\pi \log \left(\frac{\prod_{t=1}^N \phi\left(\frac{v_t - \langle x_t, \theta_0\rangle}{\sigma}\right)}{\prod_{t=1}^N \phi\left(\frac{v_t - \langle x_t, \theta_1\rangle}{\sigma}\right)} \right) . \tag{47}$$

where $\mathbb{E}_0^\pi$ denotes expectation w.r.t $\mathbb{P}_0^\pi$ and $\phi(s) = 1/(\sqrt{2\pi})e^{-s^2/2}$ is the standard Gaussian density. After simple algebraic manipulation, we obtain

$$\begin{aligned} D_{\rm KL}(\mathbb{P}_0^\pi, \mathbb{P}_1^\pi) &= -\frac{1}{2\sigma^2}\mathbb{E}_0^\pi\Big\{ \sum_{t=1}^N (2z_t - \langle x_t, \theta^1 - \theta^0\rangle)\langle x_t, \theta^1 - \theta^0\rangle \Big\} \\ &= \frac{1}{2\sigma^2}\sum_{t=1}^N \mathbb{E}_0^\pi(\langle x_t, \theta^1 - \theta^0\rangle^2) \le \frac{1}{2\sigma^2 d}\sum_{t=1}^N C_{\max}\|\theta^1 - \theta^0\|^2 \\ &= \frac{1}{2\sigma^2} C_{\max}\frac{\delta^2 N}{d} . \end{aligned}$$

We next relate the expected regret to the KL divergence between $\mathbb{P}_0^\pi$ and $\mathbb{P}_1^\pi$.

Lemma 11 *Let R_t be the regret incurred at time t, defined as $R_t \equiv p_t^* \mathbb{I}(v_t \ge p_t^*) - p_t\mathbb{I}(v_t \ge p_t)$. Then, there exist constants c_1, c_2 depending on σ, W, and $C_{\min}$, such that*

$$\mathbb{E}(R_t) \ge \frac{c_1}{d}\mathbb{E}\Big\{ \min\Big(\|\widehat{\theta}_t - \theta_t\|_2^2, c_2\Big) \Big\} . \tag{48}$$

Proof of Lemma 11 goes along the proof of (Javanmard and Nazerzadeh, 2016, Equation (55)) and is omitted.

By applying Lemma 11, we have

$$\Delta^\pi_{\boldsymbol{\theta}, \mathbb{P}_{\boldsymbol{x}}}(N) = \sum_{t=1}^N \mathbb{E}_{\boldsymbol{\theta}}(R_t) \ge \frac{c_1}{d}\sum_{t=1}^N \mathbb{E}\Big\{ \min\Big(\|\widehat{\theta}_t - \theta_t\|_2^2, c_2\Big) \Big\} . \tag{49}$$

For brevity in notations, for the sequence $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$, we define $\mathsf{d}_a(\boldsymbol{\theta}) = c_1 \sum_{t=1}^N \min(\|\theta_t - \theta^a\|_2^2, c_2)$, for $a = 1, 2$. Define two sets J_a, for $a = 1, 2$ as follows:

$$J_a = \left\{ \boldsymbol{\theta} = (\theta_1, \dots, \theta_N) :\, \theta_i \in \mathbb{R}^d \,,\, \mathsf{d}_a(\boldsymbol{\theta}) < \frac{1}{4}N\delta^2 \right\} . \tag{50}$$

Then,

$$\begin{aligned}
\max\left(\Delta^\pi_{0,\mathbb{P}_{\boldsymbol{x}}}(N), \Delta^\pi_{1,\mathbb{P}_{\boldsymbol{x}}}(N)\right) &\ge \frac{1}{d}\max\Big(\mathbb{E}^\pi_0(\mathsf{d}_0(\boldsymbol{\theta})), \mathbb{E}^\pi_1(\mathsf{d}_1(\boldsymbol{\theta}))\Big)\\
&\ge \frac{N}{4d}\delta^2 \max\Big(\mathbb{P}^\pi_0(\boldsymbol{\theta}\notin J_0), \mathbb{P}^\pi_1(\boldsymbol{\theta}\notin J_1)\Big)\\
&\overset{(a)}{\ge} \frac{N}{4d}\delta^2 \max\Big(\mathbb{P}^\pi_0(\boldsymbol{\theta}\notin J_0), \mathbb{P}^\pi_1(\boldsymbol{\theta}\in J_0)\Big)\\
&\ge \frac{N}{8d}\delta^2 \Big(\mathbb{P}^\pi_0(\boldsymbol{\theta}\notin J_0) + \mathbb{P}^\pi_1(\boldsymbol{\theta}\in J_0)\Big)\\
&\ge \frac{N}{8d}\delta^2 \Big(1-\mathbb{P}^\pi_0(\boldsymbol{\theta}\in J_0) + \mathbb{P}^\pi_1(\boldsymbol{\theta}\in J_0)\Big)\\
&\ge \frac{N}{8d}\delta^2 \Big(1-\sqrt{\frac{1}{2}D_{\rm KL}(\mathbb{P}^\pi_0,\mathbb{P}^\pi_1)}\Big) \quad \text{(By Pinsker inequality)}\\
&\ge \frac{N}{8d}\delta^2 \left(1-\frac{1}{2\sigma}\delta\sqrt{C_{\max}\frac{N}{d}}\right) \ge \frac{N\delta^2}{16d}\,.
\end{aligned}$$

Here (a) holds because $\boldsymbol{\theta}\in J_0$ implies $\boldsymbol{\theta}\notin J_1$. Otherwise, $\mathsf{d}_0(\boldsymbol{\theta}) < N\delta^2/4$ and $\mathsf{d}_1(\boldsymbol{\theta}) < N\delta^2/4$. Using the inequality $\min(a+b,c)\le \min(a,c)+\min(b,c)$ for $a,b,c\ge 0$, and applying triangle inequality, we get

$$N\min(\|\theta^0-\theta^1\|^2, c_2) \le 2\mathsf{d}_0(\boldsymbol{\theta}) + 2\mathsf{d}_1(\boldsymbol{\theta}) < N\delta^2\,, \tag{51}$$

which is a contradiction because $\delta^2 = \|\theta^0-\theta^1\|^2\le c_2$. Therefore, we conclude that

$$\begin{aligned}
\mathsf{Regret}^\pi(T,B,\nu) &\ge \Big\lfloor\frac{T}{N}\Big\rfloor \max\left(\Delta^\pi_{0,\mathbb{P}_{\boldsymbol{x}}}(N), \Delta^\pi_{1,\mathbb{P}_{\boldsymbol{x}}}(N)\right)\\
&\ge \frac{T\delta^2}{16d} = \frac{T}{16}\min\Big(\frac{\sigma^2}{C_{\max}N}, \frac{c_2}{d}\Big)\\
&= \frac{1}{16}\min\left\{\Big(\frac{\sigma^2}{C_{\max}}\Big)^{2/3}(B^2dT^{2\nu-1})^{1/3}, \frac{c_2T}{d}\right\}\,.
\end{aligned} \tag{52}$$

The result follows.

8. Proof of main lemmas

8.1 Proof of Lemma 3

We prove Lemma 3 by developing an upper bound and a lower bound for the quantity $\sum_{t=1}^T \ell_t(\widehat{\theta}_t) - \sum_{t=1}^T \ell_t(\theta_t)$. The result follows by combining these two bounds.

Lemma 12 (*Upper bound*) *Suppose $\{\theta_t\}_{t\ge1}$ is an arbitrary sequence in Θ, and $\|\theta\|\le W$ for all $\theta\in\Theta$. Set $M = 2W+\varphi^{-1}(0)$, with φ being the virtual valuation function w.r.t distribution F. Further, let $\{\widehat{\theta}_t\}_{t\ge1}$ be generated by PSGD policy using a non-increasing* *positive series* $\eta_{t+1} \le \eta_t$. *Then*

$$\sum_{t=1}^{T} \ell_t(\widehat{\theta}_t) - \sum_{t=1}^{T} \ell_t(\theta_t) \le \frac{2W^2}{\eta_1} + \sum_{t=1}^{T} \Big(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\Big) \|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2 + \frac{u_M^2}{2} \sum_{t=1}^{T} \eta_t + 2W \sum_{t=1}^{T} \frac{\delta_t}{\eta_t} - \frac{\ell_M}{2} \sum_{t=1}^{T} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \,, \tag{53}$$

where $\delta_t \equiv \|\theta_{t+1} - \theta_t\|$ *and we recall* u_M *from Equation* (15).

The proof of Lemma 12 uses similar ideas to the regret bounds established in (Hall and Willett, 2015), but uses the log-concavity of F and $1 - F$ and also definition of u_M and ℓ_M as per Equations (15) and (16) to get a more refined bound including quadratic terms $\langle x_t, \widehat{\theta}_t - \theta_t \rangle^2$. We refer to Appendix B for the proof of Lemma 12.

Our next Lemma provides a probabilistic lower bound on $\sum_{t=1}^{T} \ell_t(\widehat{\theta}_t) - \sum_{t=1}^{T} \ell_t(\theta_t)$.

Lemma 13 (*Lower bound*) *Consider model* (3) *for the product market values and suppose Assumption 1 holds. Let* $\{\widehat{\theta}_t\}_{t\ge 1}$ *be an arbitrary sequence in* Θ. *Then with probability at least* $1 - \frac{1}{T^2}$ *the following holds true*

$$\sum_{t=1}^{T} \ell_t(\widehat{\theta}_t) - \sum_{t=1}^{T} \ell_t(\theta_t) \ge -2\sqrt{\log T} \Big\{ \sum_{t=1}^{T} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \Big\}^{1/2} . \tag{54}$$

Proof of Lemma 13 is given in Appendix C. It uses convexity of $\ell_t(\widehat{\theta})$ and an application of a concentration bound on martingale difference sequences.

Combining Equations (53) and (54) we obtain that with probability at least $1 - \frac{1}{T^2}$ the following holds true

$$-2\sqrt{\log T} \Big\{ \sum_{t=1}^{T} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \Big\}^{1/2} \le \frac{2W^2}{\eta_1} + \sum_{t=1}^{T} \Big(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\Big) \|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2 + \frac{u_M^2}{2} \sum_{t=1}^{T} \eta_t + 2W \sum_{t=1}^{T} \frac{\delta_t}{\eta_t} - \frac{\ell_M}{2} \sum_{t=1}^{T} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \tag{55}$$

Rearranging the terms, we get

$$\frac{\ell_M}{2} \sum_{t=1}^{T} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 - 2\sqrt{\log T} \Big\{ \sum_{t=1}^{T} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2 \Big\}^{1/2} \le \frac{2W^2}{\eta_1} + \sum_{t=1}^{T} \Big(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\Big) \|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2 + \frac{u_M^2}{2} \sum_{t=1}^{T} \eta_t + 2W \sum_{t=1}^{T} \frac{\delta_t}{\eta_t} \tag{56}$$

Define $A \equiv \sum_{t=1}^{T} \langle x_t, \theta_t - \widehat{\theta}_t \rangle^2$ and denote by B the right-hand side of Equation (56).

Writing in terms of A and B, we have

$$A - \frac{4}{\ell_M} \sqrt{A \log T} \le \frac{2B}{\ell_M} \,. \tag{57}$$

We next upper bound A as follows. Consider two cases:

Case 1: Assume that

$$\sqrt{A \log T} \le \frac{\ell_M}{8} A\,.$$

Using this in Equation (57), we get $A \le 4B/\ell_M$.

Case 2: Assume that

$$\sqrt{A \log T} > \frac{\ell_M}{8} A\,.$$

Then, $A < (64/\ell_M^2)\log T$.

Combining the above two cases, we obtain

$$A \le \frac{4}{\ell_M} \max\Big(\frac{16}{\ell_M}\log T, B\Big)\,.$$

Substituting for A and B, we have

$$\sum_{t=1}^T \langle x_t, \theta_t - \widehat{\theta}_t\rangle^2 \le \frac{4}{\ell_M}\max\Bigg\{\frac{16}{\ell_M}\log T,$$
$$\frac{2W^2}{\eta_1} + \sum_{t=1}^T \Big(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\Big)\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2 + \frac{u_M^2}{2}\sum_{t=1}^T \eta_t + 2W\sum_{t=1}^T \frac{\delta_t}{\eta_t}\Bigg\}$$

The proof is complete.

8.2 Proof of Lemma 8

theorem 8.1 *Let σ_t denote the minimum eigenvalue of $Q_t \equiv (1/t)\sum_{\ell=1}^t x_\ell x_\ell^{\mathsf{T}}$. Further, let $\sigma_{\min}$ be the minimum eigenvalue of Σ, where Σ is the population covariance of feature vectors as in Assumption 6. Then, there exist constants $c_1, c_2 > 0$, such that*

$$\forall t \ge c_1 d:\ \mathbb{P}\Big(\frac{1}{2}\sigma_{\min} \le \sigma_t \le \frac{3}{2}\sigma_{\min}\Big) \ge 1 - 2e^{-c_2 t/d}\,. \tag{58}$$

Further, $\sigma_t \le 1$, for all $t \ge 1$.

Let $\mathcal{F}_t$ be the σ algebra generated by market shocks $\{z_\ell\}_{\ell=1}^t$ and features $\{x_\ell\}_{\ell=1}^t$. We further define $D_t = \langle x_t, \widehat{\theta}_t - \theta_t\rangle^2 - \|\Sigma^{1/2}(\widehat{\theta}_t - \theta_t)\|^2$. Note that $\widehat{\theta}_t$ is $\mathcal{F}_{t-1}$ measurable and x_t is independent of $\mathcal{F}_{t-1}$, which implies $\mathbb{E}(D_t|\mathcal{F}_{t-1}) = 0$. Hence, $\mathbb{E}(D_t) = 0$ by iterated law of expectation and therefore $\sum_{t=1}^T \mathbb{E}(D_t) = 0$. Equivalently,

$$\mathbb{E}\left[\sum_{t=1}^T \langle x_t, \widehat{\theta}_t - \theta_t\rangle^2\right] = \sum_{t=1}^T \mathbb{E}\left[\|\Sigma^{1/2}(\widehat{\theta}_t - \theta_t)\|^2\right] \ge \sigma_{\min}\,\mathbb{E}\left[\sum_{t=1}^T \|\widehat{\theta}_t - \theta_t\|^2\right] \tag{59}$$

Define $\mathcal{G}_T$ the event that bound (19) holds true. Then,

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\langle x_t, \widehat{\theta}_t - \theta_t\rangle^2\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\langle x_t, \widehat{\theta}_t - \theta_t\rangle^2 \cdot (\mathbb{I}_{\mathcal{G}} + \mathbb{I}_{\mathcal{G}^c})\right] \\
&\le \mathbb{E}\left[\sum_{t=1}^{T}\langle x_t, \widehat{\theta}_t - \theta_t\rangle^2 \cdot \mathbb{I}_{\mathcal{G}}\right] + 4W^2 T\, \mathbb{P}(\mathcal{G}^c) \\
&\le \mathbb{E}\left[\sum_{t=1}^{T}\langle x_t, \widehat{\theta}_t - \theta_t\rangle^2 \cdot \mathbb{I}_{\mathcal{G}}\right] + \frac{4W^2}{T}\,. \qquad (60)
\end{aligned}
$$

Further, using inequality $\max(a,b) \le |a| + |b|$, we get

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\langle x_t, \widehat{\theta}_t - \theta_t\rangle^2 \cdot \mathbb{I}_{\mathcal{G}}\right] \le \frac{4}{\ell_M}\Bigg\{ \frac{16}{\ell_M}\log T + \frac{12W^2}{\ell_M} + \frac{1}{2}\sum_{t=1}^{T}\mathbb{E}\left[\Big((t+1)\lambda_{t+1} - t\lambda_t\Big)\cdot \|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2\right] \\
+ \frac{u_M^2}{2}\sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{t\lambda_t}\right] + 2W\sum_{t=1}^{T}\mathbb{E}[t\lambda_t]\delta_t \Bigg\}. \qquad (61)
\end{aligned}
$$

We next bound the terms on the right-hand side individually.

$$
\begin{aligned}
&\sum_{t=1}^{T}\mathbb{E}\left[\Big((t+1)\lambda_{t+1} - t\lambda_t\Big)\cdot\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2\right] \le \frac{\ell_M}{6}\sum_{t=1}^{T}\mathbb{E}\left[\sigma_{t+1}\cdot\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2\right] \\
&\quad\le \frac{\ell_M}{6}\sum_{t=1}^{T}\mathbb{E}\left[\sigma_{t+1}\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2\,\mathbb{I}(\sigma_{t+1} < 3\sigma_{\min}/2)\right] + \frac{\ell_M}{6}\sum_{t=1}^{T}\mathbb{E}\left[\sigma_{t+1}\,\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2\,\mathbb{I}(\sigma_{t+1} > 3\sigma_{\min}/2)\right] \\
&\quad\le \frac{\ell_M}{4}\sigma_{\min}\sum_{t=1}^{T}\mathbb{E}\Big(\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2\Big) + \sum_{t=1}^{T}2\ell_M W^2 e^{-c_2 t/d} \\
&\quad\le \frac{\ell_M}{4}\sigma_{\min}\sum_{t=1}^{T}\mathbb{E}\Big(\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2\Big) + \frac{2\ell_M}{c_2 d}W^2\,, \qquad (62)
\end{aligned}
$$

where in the last inequality, we used $\mathbb{P}(\sigma_{t+1} > 3\sigma_{\min}/2) \le 2e^{-c_2 dt}$, $\sigma_t \le 1$ and $\|\widehat{\theta}_t - \theta_t\| \le 2W$, according to Proposition 8.1.

The next term on the right-hand side of (61) is bounded in the following proposition.

theorem 8.2 *Using rule* (23) *for* λ_t*, we have*

$$
\mathbb{E}\left[\frac{1}{t\lambda_t}\right] \le \frac{6}{\ell_M}\left(\tilde{c}d^2\log T + \frac{4d}{C_{\min}}\log T\right), \qquad (63)
$$

where $\tilde{c} = \max(c_1, 1/c_2)$ *and constants* c_1 *and* c_2 *are defined in Proposition 8.1 .*

Finally, for the last term, we note that Q_t is rank deficient for $t \le d$ and hence $\sigma_t = 0$, for $1 \le t \le d$. Further, the minimum eigenvalue of a matrix is a concave function over PSD matrices. By Jensen inequality, we have

$$\mathbb{E}(\lambda_t) = \frac{\ell_M}{6t}(1 + \sum_{\ell=1}^{t} \mathbb{E}(\sigma_\ell)) = \frac{\ell_M}{6t}\Big(1 + \sum_{\ell=d+1}^{t} \mathbb{E}(\sigma_\ell)\Big)$$
$$\leq \frac{\ell_M}{6t}\Big(1 + \sum_{\ell=d+1}^{t} \sigma_{\min}\Big) \leq \frac{\ell_M}{6t}\Big(1 + \frac{t-d}{d}\Big) = \frac{\ell_M}{6d}\,. \tag{64}$$

In the last inequality, we used the fact that $\text{Trace}(\Sigma) = \mathbb{E}(\|x_t\|^2) = 1$, and thus $\sigma_{\min} \leq 1/d$. Hence,

$$\sum_{t=1}^{T} \mathbb{E}[t\lambda_t]\delta_t \leq \frac{\ell_M}{6d} \sum_{t=1}^{T} t\delta_t\,, \tag{65}$$

Using Equations (62), (63), (65) to bound the right-hand side of (61), we get

$$\mathbb{E}\left[\sum_{t=1}^{T} \langle x_t, \widehat{\theta}_t - \theta_t\rangle^2 \cdot \mathbb{I}_{\mathcal{G}}\right] \leq \left[\frac{64}{\ell_M^2} + \frac{12u_M^2}{\ell_M^2}\left(\tilde{c} + \frac{4}{C_{\min} d}\right)\right] \cdot d^2 \log T$$
$$+ \frac{48W^2}{\ell_M^2} + \frac{4W^2}{c_2 d} + \frac{2W}{d}\sum_{t=1}^{T} t\delta_t + \frac{\sigma_{\min}}{2}\sum_{t=1}^{T} \mathbb{E}(\|\theta_t - \widehat{\theta}_t\|^2)\,. \tag{66}$$

Combining bounds (59),(60) and (65), we obtain

$$\frac{\sigma_{\min}}{2}\sum_{t=1}^{T} \mathbb{E}(\|\theta_t - \widehat{\theta}_t\|^2) \leq \left[\frac{64}{\ell_M^2} + \frac{12u_M^2}{\ell_M^2}\left(\tilde{c} + \frac{4}{C_{\min} d}\right)\right] \cdot d^2 \log T$$
$$+ 4W^2\left(\frac{1}{T} + \frac{12}{\ell_M^2} + \frac{1}{c_2 d}\right) + \frac{2W}{d}\sum_{t=1}^{T} t\delta_t\,.$$

The result follows by recalling that $\sigma_{\min} \geq C_{\min}/d$ as stated by Assumption 6.

Acknowledgments

The author was partially supported by a grant from Google Faculty Research Award.

Appendix A. Proof of Lemma 10

We first state some properties of the the virtual valuation function φ and the price function g, given by Equation (7).

theorem A.1 *If $1 - F$ is log-concave, then the virtual valuation function φ is strictly monotone increasing and the price function g satisfies $0 < g'(v) < 1$, for all values of $v \in \mathbb{R}$.*

We refer to (Javanmard and Nazerzadeh, 2016) (Lemmas 1 and 2 in Appendix A therein) for a proof of Proposition A.1.

For $\theta \in \Theta$ we have $\|\theta\| \le W$ and hence $|\langle x_t, \theta\rangle| \le \|x_t\|\|\theta\| \le W$ for all t. Applying Proposition A.1 (1-Lipschitz property of g),

$$p_t = g(\langle x_t, \theta_t\rangle) \le g(0) + |\langle x_t, \theta_t\rangle| \le \varphi^{-1}(0) + W\,.$$

Therefore,

$$|u_t(\theta)| \le |p_t| + |\langle x_t, \theta\rangle| \le \varphi^{-1}(0) + 2W\,. \tag{67}$$

Appendix B. Proof of Lemma 12

We note that the update rule (11) can be recast as $\widehat{\theta}_{t+1} = \arg\min_{\theta\in\Theta} \mathcal{C}_t(\theta)$, where

$$\mathcal{C}_t(\theta) = \eta_t \langle \nabla \ell_t(\widehat{\theta}_t), \theta\rangle + \frac{1}{2}\|\theta - \widehat{\theta}_t\|^2\,.$$

By convexity of $\mathcal{C}_t$ and optimality of $\widehat{\theta}_{t+1}$, we have $\langle \theta - \widehat{\theta}_{t+1}, \nabla \mathcal{C}_t(\widehat{\theta}_{t+1})\rangle \ge 0$ for all $\theta \in \Theta$. Setting $\theta = \theta_t$,

$$\langle \theta_t - \widehat{\theta}_{t+1}, \eta_t \nabla \ell_t(\widehat{\theta}_t) + \widehat{\theta}_{t+1} - \widehat{\theta}_t\rangle \ge 0\,. \tag{68}$$

Expanding $\ell_t(\theta)$ around $\widehat{\theta}_t$, we have

$$\ell_t(\widehat{\theta}_t) - \ell(\theta_t) = \langle \nabla \ell_t(\widehat{\theta}_t), \widehat{\theta}_t - \theta_t\rangle - \frac{1}{2}\langle \theta_t - \widehat{\theta}_t, \nabla^2 \ell_t(\widetilde{\theta})(\theta_t - \widehat{\theta}_t)\rangle\,, \tag{69}$$

for some $\tilde{\theta}$ on the line segment between $\widetilde{\theta}_t$ and $\widehat{\theta}_t$. Recalling (12), the gradient and the hessian of ℓ_t read as

$$\nabla \ell_t(\theta) = \mu_t(\theta) x_t\,, \quad \nabla^2 \ell_t(\theta) = \eta_t(\theta) x_t x_t^{\mathsf{T}}\,, \tag{70}$$

with,

$$\begin{aligned}
\mu_t(\theta) \quad &= \quad -\frac{f(u_t(\theta))}{F(u_t(\theta))}\mathbb{I}(y_t = -1) + \frac{f(u_t(\theta))}{1 - F(u_t(\theta))}\mathbb{I}(y_t = +1) \\
&= \quad -\frac{\mathrm{d}}{\mathrm{d}x}\log F(u_t(\theta))\mathbb{I}(y_t = -1) - \frac{\mathrm{d}}{\mathrm{d}x}\log(1 - F(u_t(\theta)))\mathbb{I}(y_t = +1)
\end{aligned} \tag{71}$$

$$\eta_t(\theta) = \left(\frac{f(u_t(\theta))^2}{F(u_t(\theta))^2} - \frac{f'(u_t(\theta))}{F(u_t(\theta))}\right)\mathbb{I}(y_t=-1) + \left(\frac{f(u_t(\theta))^2}{(1-F(u_t(\theta)))^2} + \frac{f'(u_t(\theta))}{1-F(u_t(\theta))}\right)\mathbb{I}(y_t=+1)$$
$$= -\frac{\mathrm{d}^2}{\mathrm{d}x^2}\log F(u_t(\theta))\mathbb{I}(y_t=-1) - \frac{\mathrm{d}^2}{\mathrm{d}x^2}\log(1-F(u_t(\theta)))\mathbb{I}(y_t=+1)\,. \quad (72)$$

Here, $u_t(\theta) = p_t - \langle x_t, \theta\rangle$, and $\frac{\mathrm{d}}{\mathrm{d}x}\log F(x)$ and $\frac{\mathrm{d}^2}{\mathrm{d}x^2}\log F(x)$ represent first and second derivative w.r.t x, respectively. In addition, using Equation (73)

$$|u_t(\theta)| \le \varphi^{-1}(0) + 2W = M\,, \quad \forall \theta\in\Theta\,. \quad (73)$$

Hence, invoking the definition of ℓ_M, as per Equation (16), we get that $\eta_t(\theta) \ge \ell_M$ and hence $\nabla^2\ell_t(\tilde{\theta}) \succeq \ell_M x_t x_t^{\mathsf{T}}$.

Continuing from Equation (69), we get

$$\begin{aligned}
\ell_t(\widehat{\theta}_t) - \ell(\theta_t) &\le \langle \nabla\ell_t(\widehat{\theta}_t), \widehat{\theta}_t - \theta_t\rangle - \frac{\ell_M}{2}\langle x_t, \theta_t - \widehat{\theta}_t\rangle^2\\
&= \langle \nabla\ell_t(\widehat{\theta}_t), \widehat{\theta}_{t+1} - \theta_t\rangle + \langle \nabla\ell_t(\widehat{\theta}_t), \widehat{\theta}_t - \widehat{\theta}_{t+1}\rangle - \frac{\ell_M}{2}\langle x_t, \theta_t - \widehat{\theta}_t\rangle^2\\
&\le \frac{1}{\eta_t}\langle \theta_t - \widehat{\theta}_{t+1}, \widehat{\theta}_{t+1} - \widehat{\theta}_t\rangle + \langle \nabla\ell_t(\widehat{\theta}_t), \widehat{\theta}_t - \widehat{\theta}_{t+1}\rangle - \frac{\ell_M}{2}\langle x_t, \theta_t - \widehat{\theta}_t\rangle^2\\
&= \frac{1}{2\eta_t}\Big\{\|\theta_t - \widehat{\theta}_t\|^2 - \|\theta_t - \widehat{\theta}_{t+1}\|^2 - \|\widehat{\theta}_{t+1} - \widehat{\theta}_t\|^2\Big\}\\
&\quad + \langle \nabla\ell_t(\widehat{\theta}_t), \widehat{\theta}_t - \widehat{\theta}_{t+1}\rangle - \frac{\ell_M}{2}\langle x_t, \theta_t - \widehat{\theta}_t\rangle^2\\
&= \frac{1}{2\eta_t}\Big\{\|\theta_t - \widehat{\theta}_t\|^2 - \|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2\Big\} + \frac{1}{2\eta_t}\Big\{\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2 - \|\theta_t - \widehat{\theta}_{t+1}\|^2\Big\}\\
&\quad - \frac{1}{2\eta_t}\|\widehat{\theta}_{t+1} - \widehat{\theta}_t\|^2 + \langle \nabla\ell_t(\widehat{\theta}_t), \widehat{\theta}_t - \widehat{\theta}_{t+1}\rangle - \frac{\ell_M}{2}\langle x_t, \theta_t - \widehat{\theta}_t\rangle^2 \qquad (74)
\end{aligned}$$

We next note that the second term above can be bounded as

$$\frac{1}{2\eta_t}\Big\{\|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2 - \|\theta_t - \widehat{\theta}_{t+1}\|^2\Big\} = \frac{1}{\eta_t}\langle \theta_{t+1} - \widehat{\theta}_{t+1}, \theta_{t+1} - \theta_t\rangle \le \frac{2}{\eta_t}W\delta_t\,, \quad (75)$$

because $\theta_{t+1}, \widehat{\theta}_{t+1}\in\Theta$ and hence $\|\theta_{t+1} - \widehat{\theta}_{t+1}\| \le 2W$ by triangle inequality.

Further,

$$\begin{aligned}
\langle \nabla\ell_t(\widehat{\theta}_t), \widehat{\theta}_t - \widehat{\theta}_{t+1}\rangle &\le \frac{1}{2\eta_t}\|\widehat{\theta}_{t+1} - \widehat{\theta}_t\|^2 + \frac{\eta_t}{2}\|\nabla\ell_t(\widehat{\theta}_t)\|^2\\
&\le \frac{1}{2\eta_t}\|\widehat{\theta}_{t+1} - \widehat{\theta}_t\|^2 + \frac{\eta_t}{2}|\mu(\widehat{\theta}_t)|^2\|x_t\|^2 \le \frac{1}{2\eta_t}\|\widehat{\theta}_{t+1} - \widehat{\theta}_t\|^2 + \frac{\eta_t}{2}u_M^2\,, \quad (76)
\end{aligned}$$

where we used the inequality $2ab \le a^2 + b^2$ and the characterization of gradient (70). Note that by (73), $|u_t(\widehat{\theta})| \le M$ and by definition (15), $|\mu_t(\widehat{\theta}_t)| \le u_M$. Plugging in bounds from (75) and (76) in Equation (74), we arrive at

$$\ell_t(\widehat{\theta}_t) - \ell(\theta_t) \le \frac{1}{2\eta_t}\Big\{\|\theta_t - \widehat{\theta}_t\|^2 - \|\theta_{t+1} - \widehat{\theta}_{t+1}\|^2\Big\} + \frac{2}{\eta_t}W\delta_t + \frac{\eta_t}{2}u_M^2 - \frac{\ell_M}{2}\langle x_t, \theta_t - \widehat{\theta}_t\rangle^2 \quad (77)$$

We use the shorthand $D_t = \frac{1}{2}\|\theta_t - \widehat{\theta}_t\|^2$. The result follows by summing the above bound over time:

$$\sum_{t=1}^{T} \ell_t(\widehat{\theta}_t) - \sum_{t=1}^{T} \ell_t(\theta_t) = \sum_{t=1}^{T} \Big(\frac{D_t}{\eta_t} - \frac{D_{t+1}}{\eta_{t+1}}\Big) + \sum_{t=1}^{T} D_{t+1}\Big(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\Big) + \frac{u_M^2}{2}\sum_{t=1}^{T}\eta_t + 2W\sum_{t=1}^{T}\frac{\delta_t}{\eta_t} - \frac{\ell_M}{2}\sum_{t=1}^{T}\langle x_t, \theta_t - \widehat{\theta}_t\rangle^2 .$$

The proof is concluded because $D_1 \le 2W^2$ as $\widehat{\theta}_1, \theta_1 \in \Theta$; therefore

$$\sum_{t=1}^{T}\Big(\frac{D_t}{\eta_t} - \frac{D_{t+1}}{\eta_{t+1}}\Big) = \frac{D_1}{\eta_1} - \frac{D_{T+1}}{\eta_{T+1}} \le \frac{D_1}{\eta_1} \le \frac{2W^2}{\eta_1} .$$

Appendix C. Proof of Lemma 13

By convexity of $\ell_t(\theta)$, we have

$$\ell_t(\theta_t) - \ell_t(\widehat{\theta}_t) \le \langle \nabla \ell_t(\theta_t), \widehat{\theta}_t - \theta_t\rangle = \mu_t(\theta_t)\langle x_t, \theta_t - \widehat{\theta}_t\rangle . \tag{78}$$

We denote $D_t = \mu_t(\theta_t)\langle x_t, \theta_t - \widehat{\theta}_t\rangle$ and let $\mathcal{F}_t$ be the σ-algebra generated by $\{z_t\}_{t=1}^T$. Since $\widehat{\theta}_t$ is $\mathcal{F}_{t-1}$ measurable, we have

$$\mathbb{E}(D_t|\mathcal{F}_{t-1}) = \mathbb{E}(\mu_t(\theta_t)|\mathcal{F}_{t-1})\langle x_t, \theta_t - \widehat{\theta}_t\rangle = 0 , \tag{79}$$

where $\mathbb{E}(\mu_t(\theta_t)|\mathcal{F}_{t-1}) = 0$ follows readily from Equation (71). Therefore, $D(T) \equiv \sum_{t=1}^T D_t$ is a martingale adapted to the filtration $\mathcal{F}_t$.

We next bound $\mathbb{E}[e^{\lambda D_t}|\mathcal{F}_{t-1}]$ for any $\lambda \in \mathbb{R}$. Conditional on $\mathcal{F}_{t-1}$, we have $|D_t| \le \beta_t$, with $\beta_t \equiv u_M|\langle x_t, \theta_t - \widehat{\theta}_t\rangle|$. Since $e^{\lambda z}$ is convex,

$$\begin{aligned}\mathbb{E}[e^{\lambda D_t}|\mathcal{F}_{t-1}] &\le \mathbb{E}\Big[\frac{\beta_t - D_t}{2\beta_t}e^{-\lambda\beta_t} + \frac{\beta_t + D_t}{2\beta_t}e^{\lambda\beta_t}\Big|\mathcal{F}_{t-1}\Big] \\ &= \mathbb{E}\Big[\frac{e^{-\lambda\beta_t} + e^{\lambda\beta_t}}{2}\Big] + \mathbb{E}[D_t|\mathcal{F}_{t-1}]\Big(\frac{e^{-\lambda\beta_t} + e^{\lambda\beta_t}}{2\beta_t}\Big) = \cosh(\lambda\beta_t) \le e^{\lambda^2\beta_t^2/2} .\end{aligned} \tag{80}$$

We are now ready to apply the following Bernstein-type concentration bound for martingale difference sequences, whose proof is given in Appendix D for the reader's convenience.

theorem C.1 *Consider a martingale difference sequence D_t adapted to a filtration $\mathcal{F}_t$, such that for any $\lambda \ge 0$, $\mathbb{E}[e^{\lambda D_t}|\mathcal{F}_{t-1}] \le e^{\lambda^2\sigma_t^2/2}$. Then, for $D(T) = \sum_{t=1}^T D_t$, the following holds true:*

$$\mathbb{P}(D(T) \ge \xi) \le e^{-\xi^2/(2\sum_{t=1}^T \sigma_t^2)} . \tag{81}$$

Combining Equation (78) and the result of Proposition C.1 we obtain

$$\mathbb{P}\Big(\sum_{t=1}^{T}\ell_t(\widehat{\theta}_t) - \sum_{t=1}^{T}\ell_t(\theta_t) \le -2\sqrt{\log T}\Big\{\sum_{t=1}^{T}\langle x_t, \theta_t - \widehat{\theta}_t\rangle^2\Big\}^{1/2}\Big) \le \frac{1}{T^2} . \tag{82}$$

The result follows.

Appendix D. Proof of Proposition C.1

We follow the standard approach of controlling the moment generating function of $D(T)$.Conditioning on $\mathcal{F}_{t-1}$ and applying iterated expectation yields

$$\mathbb{E}[e^{\lambda D(T)}] = \mathbb{E}\Big[e^{\lambda \sum_{t=1}^{T-1} D_t} \cdot \mathbb{E}[e^{\lambda D_T} | \mathcal{F}_{T-1}]\Big] \le \mathbb{E}\Big[e^{\lambda \sum_{t=1}^{T-1} D_t}\Big] e^{\lambda^2 \sigma_T^2/2} . \tag{83}$$

Iterating this procedure gives the bound $\mathbb{E}[e^{\lambda \sum_{t=1}^{T} D_t}] \le e^{\lambda^2 \sum_{t=1}^{T} \sigma_t^2/2}$, for all $\lambda \ge 0$.

Now by applying the exponential Markov inequality we get

$$\mathbb{P}(D(T) \ge \xi) = \mathbb{P}(e^{\lambda D(T)} \ge e^{\lambda \xi}) \le e^{-\lambda\xi} \mathbb{E}[e^{\lambda \sum_{t=1}^{T} D_t}] \le e^{-\lambda\xi} e^{\lambda^2 (\sum_{t=1}^{T} \sigma_t^2)/2} . \tag{84}$$

Choosing $\lambda = \xi/(\sum_{t=1}^{T} \sigma_t^2)$ gives the desired result.

Appendix E. Proof of Proposition 8.1

We prove the result in a more general case, namely when the features are independent random vectors with bounded subgaussian norms.

Definition 14 *For a random variable z, its subgaussian norm, denoted by $\|z\|_{\psi_2}$ is defined as*

$$\|z\|_{\psi_2} = \sup_{p \ge 1} \; p^{-1/2} (\mathbb{E}|z|^p)^{1/p} . \tag{85}$$

Further, for a random vector z its subgaussian norm is defined as

$$\|z\|_{\psi_2} = \sup_{\|u\| \ge 1} \; \|\langle z, u\rangle\|_{\psi_2} . \tag{86}$$

We next recall the following result from (Vershynin, 2012) about random matrices with independent rows.

theorem E.1 *Suppose $x_\ell \in \mathbb{R}^d$ are independent random vectors generated from a distribution with covariance Σ and their subgaussian norms are bounded by K. Further, let $Q_t = (1/t) \sum_{\ell=1}^{t} x_\ell x_\ell^\mathsf{T}$. Then for every $s \ge 0$, the following inequality holds with probability at least $1 - 2\exp(-cs^2)$:*

$$\Big\| Q_t - \Sigma \Big\| \le \max(\delta, \delta^2) \qquad\qquad \text{where } \delta = C\sqrt{\frac{d}{t}} + \frac{s}{\sqrt{t}} . \tag{87}$$

Here C and $c > 0$ are constants that depend solely on K.

We next show that the feature vectors in our problem have bounded subgaussian norm. Given that $\|x_\ell\| \le 1$, for $\|u\| \le 1$, we have

$$\|\langle x_\ell, u\rangle\|_{\psi_2} = \sup_{p\ge 1} \; p^{-1/2} (\mathbb{E}|\langle x_\ell, u\rangle|^p)^{1/p} \le \sup_{p \ge 1} \; p^{-1/2} (\mathbb{E}[\|x_\ell\| \|u\|]^p)^{1/p} \le 1 .$$

Applying Proposition (E.1) with $K = 1$, there exist constants c_1, c_2 (depending on $C_{\min}$), such that for $t \ge c_1 d^2$, we have

$$\|Q_t - \Sigma\| \le \frac{1}{2d} C_{\min} \le \frac{1}{2}\sigma_{\min}\,, \tag{88}$$

with probability at least $1 - 2e^{-c_2 t/d}$. Weyl's inequality then implies that $|\sigma_t - \sigma_{\min}| \le \sigma_{\min}/2$.

Also note that for $t \ge 1$,

$$\sigma_t \le \|Q_t\| \le \frac{1}{t}\sum_{\ell=1}^{t} \|x_\ell x_\ell^\mathsf{T}\| = \frac{1}{t}\sum_{\ell=1}^{t} \|x_\ell\|^2 = 1\,.$$

The proof is complete.

Appendix F. Proof of Lemma 8.2

The way we set λ_t (see Equation (23)), we have

$$\frac{1}{t\lambda_t} = \left(\frac{6}{\ell_M}\right)\frac{1}{1+\sigma_1+\sigma_2+\ldots+\sigma_t}$$

Clearly, for $t \ge 1$, $1/(t\lambda_t) \le 6/\ell_M$. Let $t_0 = \tilde{c} d^2 \log T$, with $\tilde{c} = \max(c_1, 1/c_2)$. For $T \ge t_0$, define the event $\mathcal{E}_T$ as follows

$$\mathcal{E}_T = \{\sigma_t \ge \sigma_{\min}/2, \text{ for } t_0 \le t \le T\}\,. \tag{89}$$

By applying Proposition 8.1 and union bounding over t, we get

$$\mathbb{P}(\mathcal{E}_T) \ge 1 - \sum_{t=t_0}^{T} 2e^{-c_2 t/d} \ge 1 - \frac{2d}{c_2} e^{-c_2 t_0/d} \tag{90}$$

Therefore,

$$\begin{aligned}
\sum_{t=t_0}^{T} \mathbb{E}\left[\frac{1}{t\lambda}\right] &\le \mathbb{E}\left[\left(\sum_{t=t_0}^{T}\frac{1}{t\lambda}\right)\mathbb{I}(\mathcal{E}_T)\right] + \frac{6T}{\ell_M}\mathbb{P}(\mathcal{E}_T^c)\\
&= \frac{6}{\ell_M}\mathbb{E}\left[\left(\sum_{t=t_0}^{T}\frac{1}{1+\sigma_1+\ldots+\sigma_t}\right)\cdot\mathbb{I}(\mathcal{E}_T)\right] + \frac{6T}{\ell_M}\mathbb{P}(\mathcal{E}_T^c)\\
&\le \frac{6}{\ell_M}\left(\sum_{t=1}^{T}\frac{1}{1+\frac{t}{2}\sigma_{\min}} + \frac{2d}{c_2}T^{1-c_2\tilde{c}d}\right)\\
&\le \frac{12}{\ell_M}\left(\frac{1}{\sigma_{\min}}\log T + \frac{d}{c_2}T^{1-d}\right) \le \frac{24d}{\ell_M C_{\min}}\log T\,.
\end{aligned} \tag{91}$$

For $t \ge 1$, we use the bound $1/(t\lambda_t) \le 6/\ell_M$. Hence,

$$\sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{t\lambda}\right] \le \frac{6}{\ell_M}\left(t_0 + \frac{4d}{C_{\min}}\log T\right) \le \frac{6}{\ell_M}\left(\tilde{c}d^2\log T + \frac{4d}{C_{\min}}\log T\right) \tag{92}$$

The proof is complete.

References

Albert Ai, Alex Lapanowski, Yaniv Plan, and Roman Vershynin. One-bit compressed sensing with non-gaussian measurements. *Linear Algebra and its Applications*, 441:222–239, 2014.

Airbnb Documentation. Smart pricing: Set prices based on demand. `https://www.airbnb.com/help/article/1168/smart-pricing--set-prices-based-on-demand`, 2015.

Kareem Amin, Afshin Rostamizadeh, and Umar Syed. Repeated contextual auctions with strategic buyers. In *Advances in Neural Information Processing Systems*, pages 622–630, 2014.

Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. *Economic theory*, 26(2):445–469, 2005.

Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: risk bounds and near-optimal algorithms. *Operations Research*, 57:1407–1420, 2009.

Sonia A Bhaskar and Adel Javanmard. 1-bit matrix completion under exact low-rank constraint. In *Information Sciences and Systems (CISS), 2015 49th Annual Conference on*, pages 1–6. IEEE, 2015.

Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Josef Broder and Paat Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012.

Maxime C Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. *ACM Conference on Economics and Computation*, 2016.

Arnoud V. den Boer and Bert Zwart. Simultaneously learning and optimizing using controlled variance pricing. *Management Science*, 60(3):770–783, 2013.

Arnoud V. den Boer and Bert Zwart. Mean square convergence rates for maximum quasi-likelihood estimators. *Stochastic Systems*, 4(2):375–403, 2014.

Vivek F Farias and Benjamin Van Roy. Dynamic pricing with a prior on market response. *Operations Research*, 58(1):16–29, 2010.

Eric C. Hall and Rebecca M. Willett. Online convex optimization in dynamic environments. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):647–662, June 2015.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. arXiv:1609.07574, 2016.

Bora Keskin and Assaf Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167, 2014.

Bora Keskin and Assaf Zeevi. Chasing demand: Learning and earning in a changing environment. *To appear in Mathematics of Operations Research*, 2016. doi: 10.1287/moor.2016.0807.

Ilan Lobel, Renato Paes Leme, and Adrian Vladu. Multidimensional binary search for contextual decision-making. *arXiv preprint arXiv:1611.00829*, 2016.

Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1): 58–73, 1981.

Yaniv Plan and Roman Vershynin. One-bit compressed sensing by linear programming. *Communications on Pure and Applied Mathematics*, 66(8):1275–1297, 2013a.

Yaniv Plan and Roman Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory*, 59(1):482–494, 2013b.

Sheng Qiang and Mohsen Bayati. Dynamic pricing with demand covariates. Working Paper, 2016.

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

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press, 2012.

Zizhuo Wang, Shiming Deng, and Yinyu Ye. Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Operations Research*, 62(2): 318–331, 2014.