
Real-Time Optimisation for Online Learning in Auctions

Lorenzo Croissant¹ Marc Abeille¹ Clément Calauzènes¹

Abstract

In display advertising, a small group of sellers and bidders face each other in up to 10^{12} auctions a day. In this context, revenue maximisation via *monopoly price* learning is a high-value problem for sellers. By nature, these auctions are *online* and produce a very high frequency stream of data. This results in a computational strain that requires algorithms be *real-time*. Unfortunately, existing methods inherited from the batch setting suffer $\mathcal{O}(\sqrt{t})$ time/memory complexity at each update, prohibiting their use. In this paper, we provide the first algorithm for online learning of monopoly prices in online auctions whose update is constant in time and memory.

Introduction

Over the last two decades, online display advertising has become a key monetisation stream for many businesses. The market for the trading of these ads is controlled by a very small (< 10) number of large intermediaries who buy and sell at auction, which means that a seller-buyer pair might trade together in 10^{10} to 10^{12} auctions per day. Repeated auctions on this scale raise the stakes of revenue maximisation, while making computational efficiency a key consideration. In his 1981 seminal work on revenue maximisation, Myerson described *the* revenue-maximising auction when buyers' bid distributions are known. In the context of online display ads these distributions are private, but the large volume of data collected by sellers on buyers opens the way to *learning* revenue maximising auctions.

Optimal vs. tractable. The learning problem associated with the Myerson auction has infinite pseudo-dimension (Morgenstern & Roughgarden, 2015), making it unlearnable (Pollard, 1984). 2nd-price auctions with personalised reserve prices (*i.e.* different for each bidder) stand as the

commonly accepted compromise between optimality and tractability. They provide a 2-approximation (Roughgarden & Wang, 2016) to the revenue of the Myerson auction while securing finite pseudo-dimension.

Monopoly prices. 2nd-price auctions with personalised reserves can be either *eager* or *lazy*. In the *eager* format, the item goes to the highest bidder *amongst* those who cleared their reserve prices and goes unsold if none of them did. In the *lazy* format, the item goes to the highest bidder *if* he cleared his reserve price and goes unsold if he did not. While an optimised *eager* version leads to better revenue than an optimised *lazy* version, solving the *eager* auction's associated Empirical Risk Minimisation (ERM) problem is NP-hard (Paes Leme et al., 2016) and even APX-hard (Roughgarden & Wang, 2016). In contrast, solving the ERM for the *lazy* version can be done in polynomial time (Roughgarden & Wang, 2016): it amounts to computing a bidder-specific quantity called the *monopoly price*. Not only is the *monopoly price* the optimal reserve in the *lazy* 2nd-price auction, but it is also a provably good reserve in the *eager* one (Roughgarden & Wang, 2016), and the optimal reserve in posted-price (Paes Leme et al., 2016). This makes learning monopoly prices for revenue maximisation an important and popular research direction.

Online learning. Finding the monopoly price in a repeated 2nd-price auction is a natural sequential decision problem based on the incoming bids. All three aforementioned settings relating to the monopoly price have been studied: posted-price (Amin et al., 2014; Blum et al., 2004), *eager* (Cesa-Bianchi et al., 2014; Roughgarden & Wang, 2016; Kleinberg & Leighton, 2003), and *lazy* which we study (Blum & Hartline, 2005; Blum et al., 2004; Mohri & Medina, 2016; Rudolph et al., 2016; Bubeck et al., 2017; Shen et al., 2019). Each setting also corresponds to a different observability structure. The offline problems are well understood, but no online method offers the $\mathcal{O}(1)$ *efficiency* crucial for real-world settings. We focus, therefore, on the key problem of learning monopoly prices, online and efficiently, in the stationary and non-stationary cases.

Structure of the paper. We propose a real-time first-order algorithm which makes online learning of monopoly prices computationally feasible, when interacting with sta-

¹Criteo AI Lab, Paris, France. Correspondence to: Lorenzo Croissant <ld.croissant@criteo.com>.

tionary and non-stationary buyers. In Sec. 1, we detail the setting and problem we consider. We review, in Sec. 2, the existing approaches and stress the challenges of the problem including overcoming computational complexities. Our approach, based on convolution and the $\mathcal{O}(1)$ Online Gradient Ascent algorithm, is described in Sec. 3. We study performance for stationary bidders in Sec. 4 with $1/\sqrt{t}$ convergence rate to the monopoly price, and for non-stationary bidders in Sec. 5 with $\mathcal{O}(\sqrt{T})$ dynamic regret.

1. Setting

A key property of a personalised reserve price in a lazy second price auction is that it can be optimised separately for each bidder (Paes Leme et al., 2016). For a bidder with bid cdf F , the optimal reserve price is the monopoly price r^* , i.e. the maximiser of the *monopoly revenue* defined as

$$\Pi^F(r) = r(1 - F(r)). \quad (1)$$

Thus, without loss of generality, we study each bidder separately in the following repeated game: the seller sets a reserve price r and simultaneously the buyer submits a bid $b \in [0, \bar{b}]$ drawn from his private distribution F , whose pdf is f . The seller then observes b which determines the instantaneous revenue

$$p(r, b) = r\mathbb{1}_{r \leq b} \quad \text{with} \quad \mathbb{E}_F[p(r, b)] = \Pi^F(r). \quad (2)$$

In this work, we consider two settings, depending whether the bid distribution is stationary or not.

Stationary setting. Stationarity here means F is fixed for the whole game. We thus have a stream of i.i.d. bids from F , where the seller aims to maximise her long term revenue

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T p(r, b_t) = \Pi^F(r).$$

Or, equivalently, tries to construct a sequence of reserve prices $\{r_t\}_{t \geq 1}$ given the information available so far encoded in the filtration $\mathcal{F}_t = \sigma(r_1, b_1, \dots, b_{t-1})$ such that $\Pi^F(r_t) \rightarrow \Pi^F(r^*)$ as fast as possible.

Non-stationary setting. In real-world applications, bid distributions may change over time based on the current context. For example, near Christmas the overall value of advertising might go up since customers spend more readily, and thus bids might increase. The bidder could also refactor his bidding policy for reasons entirely independent of the seller. We relax the stationarity assumption by allowing bids to be drawn according to a sequence of distributions $\{F_t\}_{t \geq 1}$ that varies over time. As a result, the monopoly prices $\{r_t^*\}_{t \geq 1}$ and optimal monopoly revenues

$\{\Pi^{F_t}(r_t^*)\}_{t \geq 1}$ fluctuate and convergence is no longer defined. Instead, we evaluate the performance of an adaptive sequence of reserve prices by its expected dynamic regret

$$R(T) = \mathbb{E} \left(\sum_{t=1}^T \Pi^{F_t}(r_t^*) - \Pi^{F_t}(r_t) \right), \quad (3)$$

and our objective is to track the monopoly price as fast as possible to minimise the dynamic regret.

2. Related Work, Challenges and Contributions

2.1. Related Work

Lazy 2nd-price auctions have been studied both in batch (Mohri & Medina, 2016; Shen et al., 2019; Rudolph et al., 2016; Paes Leme et al., 2016) and online (Blum & Hartline, 2005; Blum et al., 2004; Bubeck et al., 2017) settings. All existing approaches aim to optimise, at least up to a precision of $1/\sqrt{t}$, the ERM objective

$$\Pi^{\hat{F}_t}(r) = r(1 - \hat{F}_t(r)) = \frac{1}{t} \sum_{i=1}^t r\mathbb{1}_{r \leq b_i}. \quad (4)$$

However, regardless of how well-behaved Π^F is, $\Pi^{\hat{F}_t}$ is very poorly behaved for optimisation: it is non-smooth, non-quasi-concave, discontinuous, and is increasing everywhere (see Fig.1, center, dashed). Thus direct optimisation with first order methods is not applicable. Some attempts have been made in the batch setting to optimise surrogate objectives, but ended up with an irreducible bias (Rudolph et al., 2016) or with hyper-parameters whose tuning is as hard as the initial problem (Shen et al., 2019). The classical approach relies on sorting the bids $\{b_i\}_{i=1}^t$ to be able to enumerate $\Pi^{\hat{F}_t}(\cdot)$ linearly over $\{b_i\}_{i=1}^t$ (Mohri & Medina, 2016; Paes Leme et al., 2016). A popular improvement in terms of complexity, especially used in online approaches (Blum et al., 2004; Blum & Hartline, 2005) consists in applying the same principle on a regular grid of resolution $1/\sqrt{t}$, which in the end provides an update with complexity $\mathcal{O}(\sqrt{t})$ and a memory requirement of $\mathcal{O}(\sqrt{t})$.

This idea of discretising the bid space was widely adopted in partially observable settings – e.g. online *eager* or online posted-price auctions – as it reduces the problem to a multi-armed bandit with its well-studied algorithms (Kleinberg & Leighton, 2003; Cesa-Bianchi et al., 2014; Roughgarden & Wang, 2016) at the price of still suffering the same update and memory complexities of $\mathcal{O}(\sqrt{t})$.

Numerous approaches with adversarial bandits also followed this discretisation approach to adapt Exp3/Exp4 (Kleinberg & Leighton, 2003; Cohen et al., 2016; Bubeck

et al., 2017) to all the settings. As a bonus, it also allows to handle the case of non-stationary bidders. However, the work of Amin et al. (2014) stresses that bidders cannot behave in an arbitrary way, as they optimise their own objective that is not incompatible with the seller’s¹. Hence, the non-stationarity mostly comes from the item’s value changing over time. This suggest adapting a regular stochastic algorithms (ERM, UCB...), e.g. using sliding windows (Garivier & Moulines, 2011; Lattimore & Szepesvári, 2018).

Non-smooth or non-differentiable objectives such as $\Pi^{\hat{F}}$ have been studied in both stochastic and 0-order optimisation. In both, convolution smoothing has been employed to circumvent these problems. In Duchi et al. (2012) stochastic gradient with decreasing convolution smoothing is studied for the convex case. Unfortunately, very few distributions yield a concave Π^F . In 0-order optimisation, the only feedback received for an input is the value of the objective at that input. In this setting, Flaxman et al. (2005) perturb their inputs to estimate a convolved gradient. In contrast, we obtain a closed form and do not need to perturb inputs.

2.2. Challenges

The directing challenge of our line of work is to devise an online learning algorithm for monopoly prices with minimal cost, to handle very large real-world data streams. With 10^{10} daily interactions in one seller-bidder pair, it is acceptable to forfeit some convergence speed in exchange for feasibility of the algorithm. It is not possible to accept update complexity or memory requirement scaling with t . Our objective is thus to find a method that **1**) converges to r^* in the stationary setting or has a low regret $R(T)$ in the non-stationary one, **2**) has $\mathcal{O}(1)$ memory footprint, and **3**) computes the next reserve r_{t+1} with $\mathcal{O}(1)$ computations.

Computational Complexity. Unfortunately, none of the previously proposed methods fit these requirements. On one hand, all methods based on solving ERM by sorting (Cesa-Bianchi et al., 2014; Roughgarden & Wang, 2016) need to keep all past bids in memory ($\mathcal{O}(t)$ dependence) and their update steps require at best $\mathcal{O}(\sqrt{t})$ computations. On the other hand, adversarial methods such as Exp3 or Exp4 (Cohen et al., 2016; Bubeck et al., 2017) are designed for finite action spaces and thus need to discretise $[0, \bar{b}]$ into \sqrt{t} intervals (to keep their regret guarantees), also leading to a complexity of $\mathcal{O}(\sqrt{t})$ from sampling to compute r_{t+1} .

Gradient bias. First order methods (e.g. Online Gradient Ascent a.k.a. OGA) are standard tools in online learning and enjoy $\mathcal{O}(1)$ update and memory. This makes them

¹An auction is not a zero-sum game: if the item goes unsold, neither player receives payoff.

great candidates for our problem. OGA requires 3 ingredients to converge: an objective whose gradients always point towards the optimum², and a gradient estimator which has bounded variance and is unbiased. Unfortunately, discontinuity of p makes ∇p a *biased* estimator of $\nabla \Pi^F$. A natural approach is to construct a surrogate for p which has unbiased gradients and preserves the other two conditions.

Surrogate consistency. Optimising a surrogate objective inherently creates a *bias*, which has to be reduced over-time. To do so without breaking the convergence of OGA, we must conduct a careful finite time analysis of the algorithm, which is an *analytical* challenge. We must re-analyse classical results (e.g. Bach & Moulines (2011); Duchi et al. (2012)) for varying objectives: the challenge is to design a bias reduction procedure, and then integrate it into these proofs to show we preserve consistency.

Non-stationarity. Resolving the above challenges is sufficient to achieve efficient convergence in the stationary setting. However, it is not sufficient in order to track non-stationary bid distributions. Taking a constant surrogate and learning rate, it is possible to adapt the stationary solution to the non-stationary case and keep its computational efficiency. The challenge is to devise this adaptation, and then to derive (sub-linear) regret for it.

2.3. Contributions

We propose a smoothing method for creating surrogates in pseudo-concave problems with biased gradients. We use it to create a first-order real-time optimisation algorithm which reduces the surrogate’s bias during optimisation. We prove convergence and give rates in the stationary setting and dynamic regret bounds for tracking. In more detail:

Smooth surrogates for first order methods. We first translate standard auction theory assumptions (e.g. increasing virtual value) into properties of generalised concavity of the monopoly revenue (Prop. 1). Next, we introduce our smoothing method and show (in Prop. 2) that it preserves the properties from Prop. 1 while offering arbitrary smoothness, which fixes the biased gradient problem. Finally, we provide controls, via the choice of the kernel, on the bias and variance of the gradient estimates of our surrogate, which is now ready for OGA (Prop. 3).

Consistent algorithm for stationary bidders. We construct an algorithm (V-CONV-OGA) which performs gradient ascent while simultaneously decreasing the strength of the smoothing over time, reducing the bias to zero. As a result our algorithm almost surely converges to the

²Pseudo-concave or variationally coherent.

monopoly price (Thm. 1) while enjoying computational efficiency. Further, under a minimum curvature assumption, we provide the rate of convergence and optimal tuning parameters (Thm. 2 and Cor. 1). At the cost of a slight degradation in convergence speed (from $\mathcal{O}(1/t)$ to $\mathcal{O}(1/\sqrt{t})$), our algorithm has update and memory complexity of $\mathcal{O}(1)$ which is vital for real-world applications. Results are summarised in Table 1.

	Update	Memory	Convergence
ERM	$\mathcal{O}(t)$	$\mathcal{O}(t)$	$\mathcal{O}(1/t)$
Discrete ERM	$\mathcal{O}(\sqrt{t})$	$\mathcal{O}(\sqrt{t})$	$\mathcal{O}(1/t)$
V-CONV-OGA	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1/\sqrt{t})$

Table 1. Comparison of our method (V-CONV-OGA) against solving ERM at each step and ERM discretised on a grid of resolution $1/\sqrt{t}$, in terms of complexity and convergence – i.e. $\|\Pi^F(r_t) - \Pi^F(r^*)\|$.

Tracking for non-stationary bidders. Contrary to the stationary setting, when tracking we do not decrease the strength of the smoothing over time. When the bias created is smaller than the noise, our algorithm can still achieve sub-linear dynamic regret when tracking changing bid distributions. For reasonably varying distribution, we show a regret bound of $\mathcal{O}(\sqrt{T})$ (see Thm. 3 and Cor. 2).

3. Smooth Surrogate for First Order Methods

Our objective, to reiterate, is to design an online optimisation procedure to learn or track the optimal reserve price whose updates require $\mathcal{O}(1)$ computational and memory cost. To this end, we focus on first order method and consider vanilla Online Gradient Ascent. Unfortunately, the specific problem of learning a monopoly price doesn’t provide a way to compute unbiased gradient estimates for Π^F from bid samples. We therefore want to design a surrogate that makes p sufficiently smooth so that differentiation and integration commute. This is a well known property of convolutional smoothing, suggesting its use. In addition, we must ensure our surrogate preserves the optimisation properties that Π^F already has. These must thus be studied first, before smoothing to obtain a surrogate.

3.1. Properties of the Monopoly Revenue

The standard assumptions of auction theory are made to guarantee that the monopoly price exists – and that the optimisation problem is well-posed. This is generally stated as “the monopoly revenue is *quasi-concave*”. We refine this characterisation by translating the assumptions we make into specific concavity properties of the monopoly revenue in Prop 1.

(A1). $F \in \mathcal{C}^2([0, \bar{b}])$ and $f = F' > 0$ on $(0, \bar{b})$.

(A2). F is strictly regular on its domain of definition i.e. the virtual value $\psi(b) = b - \frac{1-F(b)}{f(b)}$ is increasing.

(A3). F has strongly increasing hazard rate on its domain i.e. the hazard rate $\lambda(b) = \frac{f(b)}{1-F(b)}$ satisfies:

$$\forall 0 \leq b_1 \leq b_2 \leq \bar{b}, \lambda(b_2) - \lambda(b_1) \geq \mu(b_2 - b_1).$$

(A1) is made for ease of exposition. (A2) is a standard auction theory assumption (see Krishna (2009) for a review), and implies a pseudo-concave revenue, as shown by Prop 1. (A2) is satisfied by common distributions, exhaustively listed in (Ewerhart, 2013) and by real-world data – see e.g. (Ostrovsky & Schwarz, 2011). (A3) strengthens (A2) by requiring a minimum curvature around the maximum.

Proposition 1. Let F satisfy (A1) and Π^F be its associated monopoly revenue. Then, $\Pi^F \in \mathcal{C}^2([0, \bar{b}])$, $\Pi^F > 0$ on $(0, \bar{b})$ and:

- under (A2), Π^F is strictly pseudo-concave,
- under (A3), Π^F is μ -strongly log-concave, i.e. $\log \Pi^F$ is μ -strongly concave on $(0, \bar{b})$.

3.2. A Method Based on Smoothing

Prop. 1 ensures that the first condition for the convergence of OGA is met under standard assumptions (A2) or (A3). The main difficulty in the way of using OGA for revenue optimisation – it must be stressed – lies in the undesirable shape of the instantaneous revenue p . Indeed, p is non-smooth (discontinuous even) and cannot be used to construct an unbiased estimate of $\nabla \Pi^F(r)$, which is necessary for first order-methods.

Mohri & Medina (2016) suggests replacing $p(\cdot, b)$ by a continuous upper bound. This surrogate can be used for OGA, but it has potentially large areas of zero-gradient, which means it doesn’t learn from all samples. We give a general surrogate construction (based on convolutional smoothing) which **1**) can approximate the original monopoly revenue to arbitrary accuracy, **2**) preserves the concavity properties of Π^F , **3**) offers the desired level of smoothness, and **4**) exhibits no areas of zero gradient.

Formally, given a kernel k (a metaparameter), we use convolution smoothing to create surrogates for p and Π^F :

$$p_k(r, b) = (p(\cdot, b) \star k)(r), \quad \Pi_k^F(r) = (\Pi^F \star k)(r). \quad (5)$$

This smoothing guarantees that $\nabla p_k(\cdot, b)$ is an unbiased estimate of $\nabla \Pi_k^F$. On Fig. 1, we illustrate the effect of this smoothing on p , Π^F , and Π_k^F . We introduce a set \mathcal{K} of admissible kernels which contains all strictly positive, strictly log-concave, $\mathcal{C}^1(\mathbb{R})$, $\mathcal{L}^1(\mathbb{R})$, normalised (i.e. $\int_{\mathbb{R}} k(x)dx = 1$) functions. \mathcal{K} contains a large family of

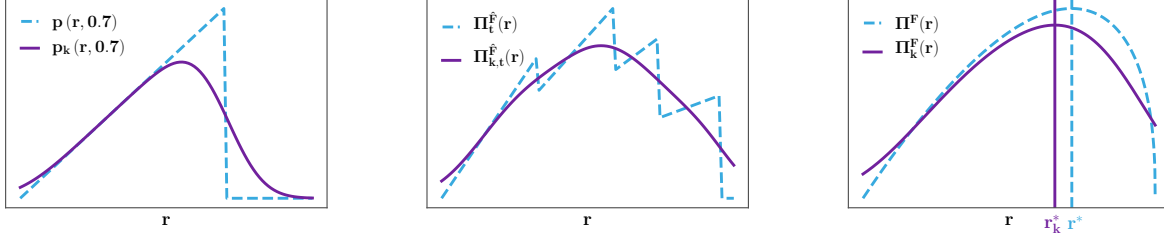


Figure 1. The effect of smoothing the monopoly revenue of a bidder with F a Kumaraswamy⁴(1,0.4) distribution with a Gaussian kernel. Left: smoothing of $p(r, b)$ for $b = 0.7$. Center: smoothing of the empirical revenue $\Pi_t^F(r)$ for some random values of b_t . Right: smoothing of the expected revenue (Π_k^F vs Π^F). Note the difference in the optima (i.e. the surrogate bias).

kernels, including standard smoothing ones such as Gaussians and mollifiers. Prop. 2 shows that convolution with elements of \mathcal{K} preserves pseudo- and log-concavity.

Proposition 2. *Let F satisfy (A1) and Π^F be its associated monopoly revenue. Let $k \in \mathcal{K}$, then:*

- Π_k^F and p_k are $\mathcal{C}^1(\mathbb{R})$,
- $\Pi_k^F(r) = \mathbb{E}_F(p_k(r, b))$ and $\nabla \Pi_k^F(r) = \mathbb{E}_F(\nabla p_k(r, b))$,
- under (A2), Π_k^F is strictly pseudo-concave on \mathbb{R} ,
- under (A3), Π_k^F is strictly log-concave on \mathbb{R} .

Proof. See App. B.2. \square

Algorithm 1 CONV-OGA

input: $r_0, \{\gamma_t\}_{t \in \mathbb{N}}, k \in \mathcal{K}, \mathcal{C} \subset [0, \bar{b}]$
for $t = 1$ **to** $+\infty$ **do**
 | observe b_t $r_t \leftarrow \text{proj}_{\mathcal{C}}(r_{t-1} + \gamma_t \nabla p_k(r_{t-1}, b_t))$

Prop. 2 guarantees that the surrogate satisfies the pseudo-concavity and unbiased gradient conditions of OGA. Applying OGA to the surrogate Π_k^F gives Alg. 1. Note that as a property of convolution, $\nabla p_k(\cdot, b) = p(\cdot, b) \star \nabla k$, which is a simple (generally closed-form) computation.

Prop. 3 will show that the bounded variance condition of OGA also holds. Since OGA’s three conditions are satisfied, we can guarantee convergence to the maximum r_k^* of Π_k^F (see e.g. (Bottou, 1998)). However, in general r_k^* is not the monopoly price r^* and the surrogate is biased. Prop. 3 also gives a control on this bias in terms of the L^1 distance between the cdf K of k and the cdf $\mathbb{1}_{\mathbb{R}^+}$ of the Dirac mass δ_0 , which is the only kernel to guarantee $r_{\delta_0}^* = r^*$.

⁴This distribution satisfies our concavity assumptions and can display highly eccentric behaviour for easy visualisation of the impact of the surrogate.

Proposition 3. *Let F satisfy (A1), $k \in \mathcal{K}$. Let r^* and r_k^* be the monopoly prices associated with Π^F and Π_k^F . Then, the bias $B_k = |\Pi^F(r^*) - \Pi^F(r_k^*)|$ and the instantaneous convolved gradient second moment $V_k = \max_{r \geq 0} \mathbb{E}_{b \sim F}(|\nabla p_k(r, b)|^2)$ are upper bounded by*

- $B_k \leq 2 \|\nabla \Pi^F\|_{\infty} \|K - \mathbb{1}_{\mathbb{R}^+}\|_1$,
- $V_k \leq 1 + \bar{b}(1 + \|\nabla \Pi^F\|_{\infty}) \|k\|_{\infty}$.

If one chooses a family of kernels, these bounds can be expressed in terms of its parameters. For instance, when k is zero-mean Gaussian with variance σ^2 , one easily recovers:

$$\|K - \mathbb{1}_{\mathbb{R}^+}\|_1 = \sigma \sqrt{2/\pi}, \quad \|k\|_{\infty} = (\sqrt{2\pi}\sigma)^{-1}. \quad (6)$$

CONV-OGA converges only to r_k^* . To remedy this, we would like to decrease B_k over time by letting $k \rightarrow \delta_0$. However, since $\|k\|_{\infty}$ diverges as $k \rightarrow \delta_0$, we will have to tread carefully in our analysis which occupies the next section.

4. Convergence with a Stationary Bidder

To decrease B_k over time, we introducing a decaying kernel sequence $\{k_t\}_{t \in \mathbb{N}}$ into CONV-OGA, giving V-CONV-OGA (Alg. 2). This section will demonstrate its consistency and convergence by controlling the trade-off between bias B_k and variance V_k , as B_k is reduced to zero over time. This trade-off decomposes the total error as:

$$\Pi^F(r^*) - \Pi^F(r_t) = \underbrace{\Pi^F(r^*) - \Pi^F(r_k^*)}_{(\text{surrogate bias})} + \underbrace{\Pi^F(r_k^*) - \Pi^F(r_t)}_{(\text{estimation})}.$$

This stresses that the kernel should converge to δ_0 *fast enough* to cancel the bias B_k , yet *slowly enough* to control V_k and preserve the convergence speed of OGA.

4.1. General Convergence Result

Thm. 1 provides sufficient conditions on the schedules of k_t and γ_t that guarantees V-CONV-OGA converges *a.s.* to

Algorithm 2 V-CONV-OGA

input: $r_0, \{\gamma_t\}_{t \in \mathbb{N}}, \{k_t\}_{t \in \mathbb{N}} \in \mathcal{K}^{\mathbb{N}}, \mathcal{C} \subset [0, \bar{b}]$
for $t = 1$ **to** $+\infty$ **do**
 | observe b_t $r_t \leftarrow \text{proj}_{\mathcal{C}}(r_{t-1} + \gamma_t \nabla p_{k_t}(r_{t-1}, b_t))$

r^* . It is derived by adapting stochastic optimisation results (see e.g. Bottou (1998)) to the changing objective $\Pi_{k_t}^F$.

Theorem 1. Let F satisfy (A1) and (A2) and $\{k_t\}_{t \in \mathbb{N}} \in \mathcal{K}^{\mathbb{N}}$. Then, by running V-CONV-OGA with $\mathcal{C} = [0, \bar{b}]$, we have

$$r_t \xrightarrow{\text{a.s.}} r^*$$

as long as $\sum_{t=1}^{+\infty} \gamma_t = +\infty$, $\sum_{t=1}^{+\infty} \gamma_t \|K_t - \mathbb{1}_{\mathbb{R}^+}\|_1 < +\infty$ and $\sum_{t=1}^{+\infty} \gamma_t^2 \|k_t\|_{\infty} < +\infty$.

Proof Sketch. The proof relies on decomposing the error $\|r_t - r^*\|^2$ into three terms related respectively to the pseudo-concavity of Π^F , the bias B_{k_t} , and the instantaneous gradient second moment V_{k_t} . Then, following Bottou (1998), we use a quasi-martingale argument to ensure the convergence of the stochastic error process. The full proof is available in App. C.1. \square

If a constant kernel sequence were to be used in V-CONV-OGA, we would recover the usual stochastic approximation conditions on the step size γ_t , namely that $\sum_{t=1}^{\infty} \gamma_t = +\infty$ and $\sum_{t=1}^{\infty} \gamma_t^2 < +\infty$. This suggests setting $\gamma_t \propto 1/t$. For such a choice of step-size, Thm. 1 asserts convergence if $\sum_{t=1}^{\infty} \gamma_t \|K_t - \mathbb{1}_{\mathbb{R}^+}\|_1 < +\infty$, which is guaranteed by $k_t \rightarrow \delta_0$. This means $\|k\|_{\infty} \rightarrow \infty$ as $t \rightarrow \infty$, but $\sum_{t=1}^{+\infty} \gamma_t^2 \|k_t\|_{\infty} < +\infty$ tells us explicitly how slow our decay must be in terms of the family of kernels. For example in the case of a Gaussian kernel, (6) implies that a suitable choice of kernel variance is $\sigma_t \propto t^{-\alpha}$ for $\alpha \in (0, 1)$.

4.2. Finite-time Convergence Rates

While Thm. 1 provides sufficient conditions on the kernel sequence $\{k_t\}_{t \in \mathbb{N}}$ for V-CONV-OGA to be consistent, it does not characterise the rate of the convergence, and thus cannot be leveraged to *optimise* the step size γ_t and the decay rate of the kernel.

To obtain finite time guarantees on the rate of convergence, we must impose stronger conditions on the monopoly revenue Π^F . Recall that under (A2), Π^F is strictly pseudo-concave. It is well known that such functions can have large areas of arbitrarily small gradient. Since they can make first order methods arbitrarily slow, no meaningful rate can be obtained for them. Strengthening the assumption to (A3), i.e. excluding vanishing gradients by ensuring Π^F is μ -strongly log-concave (see Prop. 1), will give a rate in Thm. 2 under the further technical assumption (A4).

(A4). The seller is given a compact subset $\mathcal{C} \subseteq [0, \bar{b}]$ and a constant $c > 0$ such that $r^* \in \mathcal{C}$ and for all $r \in \mathcal{C}$, $\Pi^F(r) \geq c$.

(A4) ensures that the seller can lower bound revenue on a compact subset of $[0, \bar{b}]$. It should be understood as prior knowledge of the seller based on the format of the auction and the type of item sold. \mathcal{C} exists for any $c < \Pi^F(r^*)$, so this hypothesis is not restrictive relative to (A3).

Theorem 2. Let F satisfy (A1) and (A3), let \mathcal{C} and c as in (A4) be given, and let $\{k_t\}_{t \in \mathbb{N}} \in \mathcal{K}^{\mathbb{N}}$ such that there are $\nu_1, \nu_{\infty}, \alpha_1, \alpha_{\infty}$ with

$$\|K_t - \mathbb{1}_{\mathbb{R}^+}\|_1 \leq \nu_1 t^{-\alpha_1} \text{ and } \|k_t\|_{\infty} \leq \nu_{\infty} t^{\alpha_{\infty}}.$$

Then, by running V-CONV-OGA on \mathcal{C} for $\gamma_t = \nu t^{-\alpha}$ with $\nu \leq (2c\mu)^{-1}$, we have for all $t \geq 2$,

$$\begin{aligned} \alpha = 1 : \quad & \mathbb{E}(\|r_t - r^*\|^2) = \tilde{O}(t^{-\alpha_1} + t^{\alpha_{\infty}-1} + t^{-2\mu c\nu}) \\ \alpha \in (0, 1) : \quad & \mathbb{E}(\|r_t - r^*\|^2) = \tilde{O}(t^{-\alpha_1} + t^{\alpha_{\infty}-\alpha}) \\ & + \tilde{O}\left((t^{1-\alpha-\alpha_1} + t^{1+\alpha_{\infty}-2\alpha})e^{-\mu c\nu t^{1-\alpha}}\right) \end{aligned}$$

where \tilde{O} potentially hides a logarithmic term depending on the values of α, α_1 , and α_{∞} .

Proof Sketch. The extended statement of Thm. 2 with explicit constants and its proof are detailed in App. C.2. The proof builds on Bach & Moulines (2011, Thm.2), derived for log-concave functions, and is adapted to our varying kernel approach and its changing objective. In contrast with the proof of Thm. 1, where we only leveraged pseudo-concavity, we show here that (A3), together with (A4), guarantees more refined control on the curvature of Π^F around r^* :

$$\forall r \in \mathcal{C}, \quad (r - r^*) \nabla \Pi^F(r) \leq -\mu c \|r^* - r\|^2. \quad (7)$$

This way we can better control the stochastic process $\{\|r_t - r^*\|^2\}_{t \in \mathbb{N}}$: like for Thm. 1, we decompose the error into three terms related to concavity (Eq. 7), bias B_{k_t} , and instantaneous gradient smoothness V_{k_t} . The error is then bounded in expectation by manipulating finite series. \square

Thm. 2 show two distinct regimes for both choices of α : transient ($t^{-2\mu c\nu}$ and $e^{-\mu c\nu t^{1-\alpha}}$ resp.) and stationary ($t^{-\alpha_1} + t^{\alpha_{\infty}-\alpha}$ and $t^{-\alpha_1} + t^{\alpha_{\infty}-1}$ resp.). On Fig. 2 (top), the transient phase is visible up to 2×10^3 steps. Since the transient regime's rate depends only on $\nu = \gamma_0, c$ known from (A4), and μ known from (A3), we can set ν to make the *stationary* regime the driver of the rate.

To optimise the stationary regime we face a bias-variance trade-off. Like Thm. 1, Thm. 2 requires that $k \rightarrow \delta_0$ (via $\alpha_1 > 0$) while imposing a bound on the growth speed

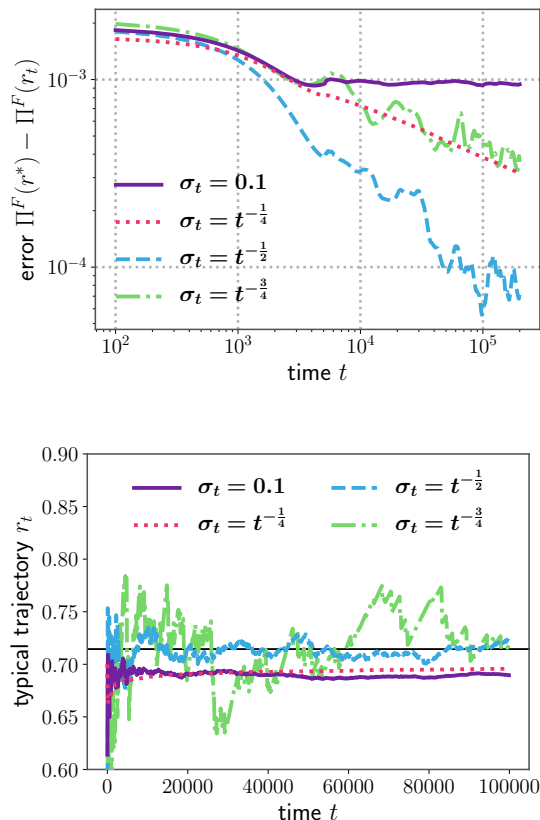


Figure 2. Stationary case. Numerical behaviour of V-CONV-OGA for different σ_t on i.i.d samples from a Kumaraswamy (1, 0.4). Top: averaged convergence speeds of instant regret (log-log scale). Bottom: representative reserve price trajectories.

of V_k (via $\alpha_\infty < \alpha$). This time, however, we have exact rates which we can use to determine optimal parameters for the trade-off, taking into account the antagonistic effects of α_1 and α_∞ . From Thm. 2, we recover that the optimal learning rate is $\gamma_t \propto 1/t$. To tune the kernels it is sensible to fix a parametric family and tune its parameter(s). For zero-mean Gaussian kernels, we have Cor. 1.

Corollary 1. *If we fix $\gamma_t \propto 1/t$, and let $\{k_t\}_{t \in \mathbb{N}}$ be Gaussian $(0, 1/t)$ kernels in Thm. 2 we have for all $t \geq 2$ that:*

$$\mathbb{E}(\|r_t - r^*\|^2) = \tilde{O}(t^{-1/2}).$$

This rate is optimal up to logarithmic factors.

Fig. 2 demonstrates this optimality: the $\sigma_t = 1/\sqrt{t}$ (blue) curve is the optimal rate on the top pane, and attains the rate of Cor. 1. The bottom pane illustrates the bias variance trade-off at hand in Thm. 2. If the kernel decays slower than $1/\sqrt{t}$ (red), the learning rate shrinks much faster and convergence is very slow but very smooth. If σ_t decreases too fast (green) the variance becomes overwhelming and

noise swallows the performance.

The novel analysis of V-CONV-OGA showed its a.s. convergence under (A2), and that with a bit of curvature (A3) and the technical (A4) we could fully characterise its convergence rates. We could thus derive optimal learning rates and place conditions on optimal kernel decay rates. We made the optimal decay rate explicit for Gaussian kernels. This concludes the primary discussion on V-CONV-OGA, and we now move to the non-stationary setting.

5. Tracking a Non-stationary Bidder

In practical applications of online auctions, such as display advertising, bidders might change their bid distribution over time. These changes often result from non-stationarity in the private information of bidders. It is therefore beneficial to be able to effectively adapt one’s reserve price over time to *track* changing bid distributions $\{F_t\}_{t \in \mathbb{N}}$. We use the dynamic regret $R(T)$ to measure the quality of an algorithm’s tracking.

The difficulty in the non-stationary setting is to trade-off adaptability (how fast a switch is detected) vs. accuracy (how accurate one is between switches). Convergent algorithms like ERM or V-CONV-OGA will have high accuracy in the first phase, but then suffer as they try to adapt to changes later on, when their learning rate is very small. Windowed methods are more adaptable but still carry with them a lag, directly dependent on their window size. First-order methods like CONV-OGA (with constant learning rate γ) are much more adaptable, but their convergence rate ($\mathcal{O}(1/\sqrt{t})$) hurts their accuracy. Nevertheless, we show that CONV-OGA is effective, with $\mathcal{O}(\sqrt{T})$ regret.

The dynamic regret $R(T)$ cannot be meaningfully controlled for arbitrary sequences $\{F_t\}_{t \in \mathbb{N}}$. As such it is customary to assume (A5) that $\{F_t\}_{t \in \mathbb{N}}$ contains at most $\tau - 1$ switches up to a horizon T (see e.g. (Garivier & Moulines, 2011; Lattimore & Szepesvári, 2018)). This corresponds to approximating a slowly changing sequence of F_t (e.g. Lipschitz) by a piece-wise constant sequence.

(A5). *Given some horizon T , there exists $\tau \leq T$ such that $\sum_{t=1}^{T-1} \mathbb{1}_{F_t \neq F_{t+1}} \leq \tau - 1$.*

Under (A5), the game (up to T) decomposes into τ phases. The first step towards controlling the regret is to bound the tracking performance in each phase. We do this in Thm. 3, which shows an incompressible asymptotic error (the bias of our surrogate plus the variance) and a transient phase with exponential decay.

Theorem 3. *Let F satisfy (A1) and (A3) with parameter μ , let \mathcal{C} and c be as in (A4) and $k \in \mathcal{K}$. Then, by running CONV-OGA on \mathcal{C} with a constant stepsize $\gamma > 0$, for any*

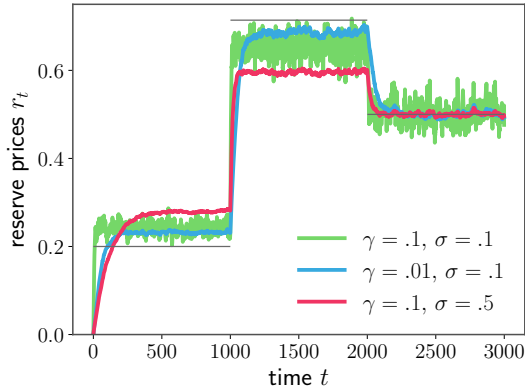


Figure 3. Non-stationary case. Tracking by CONV-OGA of three Kumaraswamy distributions (with parameters (1, 4), (1, 0.4), and (1, 1) resp.) with different Gaussian kernels and learning rates.

$t \geq 1$ we have

$$\mathbb{E}(\|r_t - r_k^*\|^2) \leq (\bar{b}^2 + C(\gamma, k)(t-1))e^{-\frac{\mu c \gamma}{2}t} + \frac{2C(\gamma, k)}{\mu c}$$

where $C(\gamma, k) = \mathcal{O}(\gamma \|K - \mathbb{1}_{\mathbb{R}^+}\|_1 + \gamma^2 \|k\|_\infty)$.

Thus, immediately after a switch there will be a transient regime of order $te^{-\frac{\mu c \gamma}{2}t}$ (high adaptability), but afterwards r_t will oscillate in a band of size $\frac{2C(\gamma, k)}{\mu c}$ around r_k^* (low accuracy). We can then use Thm. 3 to derive a sub-linear regret bound given T, τ (Cor. 2).

Corollary 2. Let $\{F_t\}_{t \geq 1}$ satisfy (A1), (A3), (A4), (A5) and $k \in \mathcal{K}$. Then, there exists $\Xi(k, \gamma)$ and $\Omega(k, \gamma)$ such that CONV-OGA has a non-stationary regret of

$$R(T) \leq \Xi(k, \gamma)T + \Omega(k, \gamma)\tau.$$

Further, if the horizon T is known in advance, running CONV-OGA with $\gamma = T^{-\frac{1}{2}}$ and k a kernel with $\|K - \mathbb{1}_{\mathbb{R}^+}\|_1 \leq T^{-\frac{1}{2}}$ and $\|k\|_\infty \leq T^{\frac{1}{2}}$, then $R(T) = \mathcal{O}(\sqrt{T})$.

Proof. See App. D. \square

Figure 3 illustrates the behaviour of CONV-OGA in a non-stationary environment. In agreement with Thm. 3 and Cor. 2, γ controls the length of the transient regime due to the $e^{-\frac{\mu c \gamma}{2}t}$ term. Increasing γ shortens it but increases the width of the band of the asymptotic regime as $C(\gamma, k)$ increases with γ (blue vs. green curves). For a fixed γ , the stationary regime in terms of k exhibits a bias-variance trade-off: $\|K - \mathbb{1}_{\mathbb{R}^+}\|_1$ corresponds to the bias and $\|k\|_\infty$ to the variance (see Prop. 3). In the case of a Gaussian kernel, increasing σ reduces variance but increases bias (green vs. red curve).

CONV-OGA with a constant learning rate is an efficient real-time algorithm for tracking monopoly prices of non-stationary bidders. It incurs $\mathcal{O}(\sqrt{T})$ regret given the horizon and τ , by tuning γ and k , while maintaining the computational efficiency of online methods.

6. Discussion

In this paper we introduced V-CONV-OGA, the first *real-time* ($\mathcal{O}(1)$ update-time and memory) method for monopoly price learning. We first gave some theoretical results bridging auction theory and optimisation. We then showed how to fix the biased gradient problem with smooth surrogates, giving CONV-OGA. Next, we let the smoothing decrease over time in V-CONV-OGA, for whom we showed convergence of $\mathcal{O}(1/\sqrt{t})$. Finally, we adapted CONV-OGA to perform tracking of non-stationary bid distributions with $\mathcal{O}(\sqrt{T})$ dynamic regret.

Towards optimal rates. In the context of high-frequency auctions, computational efficiency trumps numerical precision, so we traded $\mathcal{O}(\sqrt{t})$ complexity and $\mathcal{O}(1/t)$ speed for $\mathcal{O}(1)$ complexity and $\mathcal{O}(1/\sqrt{t})$ speed. Whether or not it is possible to reach the optimal rate with a real-time algorithm remains an open question. We conjecture this to be impossible in general, but we know it is possible in some instances. If F is a symmetric distribution, then CONV-OGA with a constant *symmetric* kernel has no bias and $\mathcal{O}(1/t)$ convergence. Adapting the choice kernel to some *a priori* knowledge on F is a possible direction to match the optimal rate.

Extension to stationary bandit. The second question concerns the extension to partially observable settings, such as online *eager* auctions, when the seller does not observe bids under the reserve. Obviously, extensions using a reduction to multi-armed bandits (UCB, Exp3, Exp4, etc.) via a discretisation of the bid space cannot be *real-time*: the discretisation creates a need for $\mathcal{O}(\sqrt{t})$ in memory and the same for the update. Yet, it is possible to obtain a straightforward extension of V-CONV-OGA in this setting, by plugging it into an Explore-The-Commit (ETC) (Perchet & Rigollet, 2013) algorithm: V-CONV-OGA learns an estimate of the monopoly price during the exploration period, which is then used during the exploitation period. As for other algorithms, by using a doubling trick to handle an unknown horizon, ETC+V-CONV-OGA exhibits a sub-linear regret. Unfortunately, like in the *lazy* auction setting, the regret is not optimal and the question of whether a *real-time* algorithm can match this optimal regret is still open.

Extension to non-stationary bandit. The question of the partially observable setting also applies to non-stationary bidders. In this case, extending CONV-OGA with ETC is no longer straight-forward, as the switching times are unknown. Thus, it is not obvious when to re-trigger an exploration phase of ETC to adapt to the change of the bidder’s distribution. A potential way to tackle this problem could be to use randomised resets for the algorithm (Allesiardo et al., 2017) or change-point detection algorithms to trigger exploration (Hartland et al., 2006).

References

- Allesiardo, R., Féraud, R., and Maillard, O.-A. The non-stationary stochastic multi-armed bandit problem. *International Journal of Data Science and Analytics*, 3(4): 267–283, Jun 2017.
- Amin, K., Rostamizadeh, A., and Syed, U. Repeated contextual auctions with strategic buyers. In *Advances in Neural Information Processing Systems*, pp. 622–630, 2014.
- Bach, F. and Moulines, E. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Advances in Neural Information Processing Systems*, 2011.
- Blum, A. and Hartline, J. D. Near-optimal online auctions. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1156–1163. Society for Industrial and Applied Mathematics, 2005.
- Blum, A., Kumar, V., Rudra, A., and Wu, F. Online learning in online auctions. *Theoretical Computer Science*, 324(2-3):137–146, 2004.
- Bottou, L. Online learning and stochastic approximations. *On-line learning in neural networks*, 17(9):142, 1998.
- Bubeck, S., Devanur, N. R., Huang, Z., and Niazadeh, R. Online auctions and multi-scale online learning. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 497–514. ACM, 2017.
- Cesa-Bianchi, N., Gentile, C., and Mansour, Y. Regret minimization for reserve prices in second-price auctions. *IEEE Transactions on Information Theory*, 61(1):549–564, 2014.
- Cohen, M., Lobel, I., and Leme, R. P. Feature-based dynamic pricing. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, 2016.
- Duchi, J. C., Bartlett, P. L., and Wainwright, M. J. Randomized smoothing for stochastic optimization. *SIAM Journal on Optimization*, 22(2):674–701, 2012.
- Ewerhart, C. Regular type distributions in mechanism design and ρ -concavity. *Economic Theory*, 53(3):591–603, 2013.
- Flaxman, A. D., Kalai, A. T., and McMahan, H. B. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA 05, pp. 385394, USA, 2005. Society for Industrial and Applied Mathematics. ISBN 0898715857.
- Garivier, A. and Moulines, E. On upper-confidence bound policies for switching bandit problems. In *International Conference on Algorithmic Learning Theory*, pp. 174–188. Springer, 2011.
- Hartland, C., Gelly, S., Baskiotis, N., Teytaud, O., and Sebag, M. Multi-armed Bandit, Dynamic Environments and Meta-Bandits. working paper or preprint, November 2006. URL <https://hal.archives-ouvertes.fr/hal-00113668>.
- Ibragimov, I. A. On the composition of unimodal distributions. *Theory of Probability & Its Applications*, 1(2): 255–260, 1956.
- Kleinberg, R. and Leighton, T. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pp. 594–605. IEEE, 2003.
- Krishna, V. *Auction theory*. Academic press, 2009.
- Lattimore, T. and Szepesvári, C. *Bandit algorithms*. preprint, 2018.
- Mohri, M. and Medina, A. M. Learning algorithms for second-price auctions with reserve. *The Journal of Machine Learning Research*, 17(1):2632–2656, 2016.
- Morgenstern, J. H. and Roughgarden, T. On the pseudo-dimension of nearly optimal auctions. In *Advances in Neural Information Processing Systems*, pp. 136–144, 2015.
- Myerson, R. B. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.
- Ostrovsky, M. and Schwarz, M. Reserve prices in internet advertising auctions: A field experiment. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, EC ’11, pp. 59–60, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0261-6.
- Paes Leme, R., Pal, M., and Vassilvitskii, S. A field guide to personalized reserve prices. In *Proceedings of the 25th International Conference on World Wide Web*, WWW

'16, pp. 1093–1102, Republic and Canton of Geneva, Switzerland, 2016. International World Wide Web Conferences Steering Committee.

Perchet, V. and Rigollet, P. The multi-armed bandit problem with covariates. *The Annals of Statistics*, 41(2):693–721, 04 2013.

Pollard, D. *Convergence of stochastic processes*. Springer Science & Business Media, 1984.

Roughgarden, T. and Wang, J. R. Minimizing Regret with Multiple Reserves. In *Proceedings of the 2016 ACM Conference on Economics and Computation - EC '16*, volume 9, pp. 601–616, 2016.

Rudolph, M. R., Ellis, J. G., and Blei, D. M. Objective variables for probabilistic revenue maximization in second-price auctions with reserve. In *Proceedings of the 25th International Conference on World Wide Web*, pp. 1113–1122. International World Wide Web Conferences Steering Committee, 2016.

Saumard, A. and Wellner, J. A. Log-concavity and strong log-concavity: A review. *Statistics Surveys*, 8:45–114, 2014.

Shen, W., Lahaie, S., and Paes Leme, R. Learning to clear the market. In *International Conference on Machine Learning*, pp. 5710–5718, 2019.

A. General Reminders on Pseudo- and Log-Concavity

This section is a stand-alone reminder and does not share notations with the rest of the paper.

A.1. Pseudo-Concavity

Definition 1. A function $f : \mathcal{X} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^1(\mathcal{X})$, is pseudo-concave on \mathcal{X} if $\forall (x, y) \in \mathcal{X}^2$,

$$\nabla f(x)^T(x - y) \geq 0 \Rightarrow f(x) \geq f(y).$$

Definition 2. A function $f : \mathcal{X} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^1(\mathcal{X})$, is strictly pseudo-concave on \mathcal{X} if it is pseudo-concave and has at most one critical point.

A.2. Log-Concavity

Definition 3. A function $f : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ is log-concave on \mathcal{X} if

$$\forall \alpha \in [0, 1], \forall (x, y) \in \mathcal{X}^2, f(\alpha x + (1 - \alpha)y) \geq f(x)^\alpha f(y)^{(1-\alpha)}$$

Note that if $f : \mathcal{X} \rightarrow \mathbb{R}_*^+$ this is equivalent to saying $f = e^{-\varphi}$ where φ is a convex function on \mathcal{X} .

Definition 4. A function $f : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ is strictly log-concave on \mathcal{X} if

$$\forall \alpha \in (0, 1), \forall (x, y) \in \mathcal{X}^2 \text{ s.t. } x \neq y, f(\alpha x + (1 - \alpha)y) > f(x)^\alpha f(y)^{(1-\alpha)}$$

Note that if $f : \mathcal{X} \rightarrow \mathbb{R}_*^+$ this is equivalent to saying $f = e^{-\varphi}$ where φ is a strictly convex function on \mathcal{X} .

Definition 5. A function $f : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ is μ -strongly log-concave on \mathcal{X} if $x \mapsto f(x)e^{-\mu x^2}$ is log-concave.

Note that if $f : \mathcal{X} \rightarrow \mathbb{R}_*^+$ this is equivalent to saying $f = e^{-\varphi}$ where φ is μ -strongly convex on \mathcal{X} .

We also recall a useful technical result for any log-concave function f , that is a straightforward consequence from the concavity characterization of $\log(f)$.

Proposition 4. Let f be a real strictly positive strictly log-concave function. Then, for all $u > v$, for all $\delta > 0$,

$$\frac{f(v + \delta)}{f(v)} > \frac{f(u + \delta)}{f(u)}.$$

Proof of Prop. 4. The proof is a straightforward application of properties of strictly concave functions applied to $\log(f)$.

Let $F(x, y) = \frac{\log f(x) - \log f(y)}{x - y}$, then F is strictly decreasing in x for every fixed y (and vice-versa). Thus,

$$\begin{aligned} F(v + \delta, v) > F(v + \delta, u) > F(u + \delta, u) &\Rightarrow \log f(v + \delta) - \log f(v) > \log f(u + \delta) - \log f(u) \\ &\Rightarrow \frac{f(v + \delta)}{f(v)} > \frac{f(u + \delta)}{f(u)}. \end{aligned}$$

□

A.3. Stability through Convolution

Theorem 4 (Ibragimov (1956)). Let $f : \mathcal{X} \subset \mathbb{R} \rightarrow \mathbb{R}^+$ be pseudo-concave on \mathcal{X} , $\mathcal{C}^1(\mathcal{X})$, $\mathcal{L}^1(\mathcal{X})$ and $g : \mathbb{R} \rightarrow \mathbb{R}^+$ be $\mathcal{L}^1(\mathbb{R})$ and log-concave. Then, $f \star g$ is pseudo-concave on \mathbb{R} .

We extend this theorem to *strict* pseudo-concavity.

Lemma 1 (Extension of Ibragimov (1956)). Let $f : \mathcal{X} \subset [x_1, x_2] \rightarrow \mathbb{R}^+$ be a strictly pseudo-concave on \mathcal{X} , $\mathcal{C}^1(\mathcal{X})$, $\mathcal{L}^1(\mathcal{X})$ such that $\lim_{x \rightarrow x_{1,2}} f(x) = 0$ and $g : \mathbb{R} \rightarrow \mathbb{R}^+$ be $\mathcal{L}^1(\mathbb{R})$ and strictly log-concave. Then, $f \star g$ is strictly pseudo-concave on \mathbb{R} .

Proof. The proof is conducted in two steps: **1)** we show $f \star g$ admits a maximum on the interior of its domain (which is a critical point) and we denote it x^* . **2)** we show that $f \star g$ is strictly increasing on $(-\infty, x^*)$ and strictly decreasing on $(x^*, +\infty)$ which immediately proves the strict pseudo-concavity (including unicity of x^*).

1. Since f and g are \mathcal{C}^1 and \mathcal{L}^1 , the convolution $f \star g$ is well defined, $\mathcal{C}^1(\mathbb{R})$, $\mathcal{L}^1(\mathbb{R})$ and positive (since f and g are positive). As a result, $f \star g(x) \rightarrow 0$ when $|x| \rightarrow \infty$ and Rolle's theorem guarantees that there exists at least one point $x^* \in \mathbb{R}$ such that $f \star g(x^*) \geq f \star g(x)$ for all $x \in \mathbb{R}$. Furthermore, Ibragimov's theorem (Thm. 4) ensures that $f \star g$ is pseudo-concave, hence $\nabla(f \star g)(x^*) = 0$.
2. Using the differentiation property of the convolution, one has that for all $x \in \mathbb{R}$,

$$\nabla(f \star g)(x) = \int_{-\infty}^{\infty} f(t) \nabla g(x-t) dt = \int_{x_1}^{x_2} \nabla f(t) g(x-t) dt, \quad (8)$$

where we used the fact that $\lim_{x \rightarrow x_{1,2}} f(x) = 0$. Let x^* be a critical point of $f \star g$. Taking Eq. 8 at $x = x^*$ leads to:

$$0 = \int_{x_1}^{x_2} \nabla f(t) g(x^* - t) dt.$$

Moreover, let $y^* \in (x_1, x_2)$ be the unique (by pseudo-concavity) critical point of f , whose existence is guaranteed by Rolle's theorem ($\lim_{x \rightarrow x_{1,2}} f(x) = 0$). We now split the integral in Eq. 8 to obtain:

$$\nabla(f \star g)(x) = \int_{x_1}^{y^*} \nabla f(t) g(x-t) dt + \int_{y^*}^{x_2} \nabla f(t) g(x-t) dt. \quad (9)$$

The core of the proof consists in proving that $\nabla(f \star g)(x^* + \delta) > 0$ for all $\delta > 0$ and $\nabla(f \star g)(x^* + \delta) < 0$ for all $\delta < 0$. Since the derivation is similar in both cases, we only display here the case $\delta > 0$. From Eq. 9, we have:

$$\begin{aligned} \nabla(f \star g)(x^* + \delta) &= \int_{x_1}^{y^*} \nabla f(t) g(x^* + \delta - t) dt + \int_{y^*}^{x_2} \nabla f(t) g(x^* + \delta - t) dt \\ &= \int_{x_1}^{y^*} \nabla f(t) g(x^* - t) \frac{g(x^* + \delta - t)}{g(x^* - t)} dt + \int_{y^*}^{x_2} \nabla f(t) g(x^* - t) \frac{g(x^* + \delta - t)}{g(x^* - t)} dt. \end{aligned}$$

We now provide upper and lower bounds for $\frac{g(x^* + \delta - t)}{g(x^* - t)}$ respectively on $[x_1, y^*]$ and $[y^*, x_2]$. Let

$$t^* = \operatorname{argmax}_{t \in [x_1, y^*]} \frac{g(x^* + \delta - t)}{g(x^* - t)}$$

which exists by our hypotheses on g . Then,

$$\forall t \in [x_1, y^*], \quad \frac{g(x^* + \delta - t)}{g(x^* - t)} \leq \frac{g(x^* + \delta - t^*)}{g(x^* - t^*)}.$$

Moreover, applying Prop. 4, we have for almost all $t \in [y^*, x_2]$,

$$\frac{g(x^* + \delta - t)}{g(x^* - t)} > \frac{g(x^* + \delta - t^*)}{g(x^* - t^*)}.$$

Since f is strictly pseudo-concave, $\nabla f(t) > 0$ on $[x_1, y^*]$ and $\nabla f(t) < 0$ on $(y^*, x_2]$, we obtain

$$\begin{aligned} \nabla(f \star g)(x^* + \delta) &= \int_{x_1}^{y^*} \nabla f(t) g(x^* - t) \frac{g(x^* + \delta - t)}{g(x^* - t)} dt + \int_{y^*}^{x_2} \nabla f(t) g(x^* - t) \frac{g(x^* + \delta - t)}{g(x^* - t)} dt \\ &< \int_{x_1}^{y^*} \nabla f(t) g(x^* - t) \frac{g(x^* + \delta - t^*)}{g(x^* - t^*)} dt + \int_{y^*}^{x_2} \nabla f(t) g(x^* - t) \frac{g(x^* + \delta - t^*)}{g(x^* - t^*)} dt \\ &< \frac{g(x^* + \delta - t^*)}{g(x^* - t^*)} \nabla(f \star g)(x^*) = 0, \end{aligned}$$

which proves the desired result.

□

Similar stability properties through convolution are asserted for strictly and strongly log-concave functions. The first result is standard and can be derived from the Prépoka-Leindler inequality, the second can be retrieved from (Saumard & Wellner, 2014).

Proposition 5. *Let $f : \mathcal{X} \subset \mathbb{R} \rightarrow \mathbb{R}^+$ and $g : \mathbb{R} \rightarrow \mathbb{R}^+$ be log-concave. Then, $f \star g$ is log-concave.*

Theorem 5 (Saumard & Wellner (2014), Thm. 6.6). *Let $f : \mathcal{X} \subset \mathbb{R} \rightarrow \mathbb{R}^+$ and $g : \mathbb{R} \rightarrow \mathbb{R}^+$ be μ and μ' strongly log-concave respectively. Then, $f \star g$ is $\mu\mu' / \sqrt{\mu^2 + \mu'^2}$ strongly log-concave. Further, the convolution of strictly log-concave functions is strictly log-concave.*

B. Proofs of Sec. 3

B.1. Pseudo- and Log-Concavity of the Monopoly Revenue

Proposition 1. *Let F satisfy (A1) and Π^F be its associated monopoly revenue. Then, $\Pi^F \in \mathcal{C}^2([0, \bar{b}])$, $\Pi^F > 0$ on $(0, \bar{b})$ and:*

- under (A2), Π^F is strictly pseudo-concave,
- under (A3), Π^F is μ -strongly log-concave, i.e. $\log \Pi^F$ is μ -strongly concave on $(0, \bar{b})$.

Proof.

- Under (A2), $\psi(r) = r - \frac{1-F(r)}{f(r)}$ is strictly increasing. Moreover, for all $r \in [0, \bar{b}]$,

$$\nabla \Pi^F(r) = 1 - F(r) - rf(r) = -\psi(r)f(r).$$

The objective is to show that for all $(r_1, r_2) \in [0, \bar{b}]^2$, $\nabla \Pi^F(r_1)(r_1 - r_2) \geq 0 \Rightarrow \Pi^F(r_1) \geq \Pi^F(r_2)$ and that Π^F has one critical point (by Rolle's theorem, since $\Pi^F(0) = \Pi^F(\bar{b}) = 0$). Without loss of generality, we only address the case where $r_1 \leq r_2$. Since ψ is strictly increasing, $\psi(r_1) \leq \psi(r_2)$, and as a result

$$\begin{aligned} \nabla \Pi^F(r_1) \leq 0 &\Leftrightarrow \psi(r_1) \geq 0 \\ \Rightarrow \forall r \in [r_1, r_2], \psi(r) \geq 0 &\Leftrightarrow \nabla \Pi^F(r) \leq 0 \\ \Rightarrow \Pi^F(r_2) - \Pi^F(r_1) &= \int_{r_1}^{r_2} \nabla \Pi^F(r) dr \leq 0. \end{aligned}$$

The case $\nabla \Pi^F(r_1) \geq 0$ is treated in a similar fashion. Finally, since ψ is strictly increasing it can only cross 0 once, which immediately ensures the uniqueness of the critical point since $f > 0$ on $(0, \bar{b})$.

- Under (A3), the hazard rate $\lambda(r) = \frac{f(r)}{1-F(r)}$ satisfies $\forall 0 \leq r_1 \leq r_2 \leq \bar{b}$, $\lambda(r_2) - \lambda(r_1) \geq \mu(r_2 - r_1)$. The objective is to show that $\log \Pi^F(r) = \log(r) + \log(1 - F(r))$ is μ -strongly concave. As $\log(r)$ is concave, we can simply show that $\log(1 - F(r))$ is μ -strongly concave. Since $F \in \mathcal{C}^1([0, \bar{b}])$, we can use a characterisation of strong concavity of $G(r) = \log(1 - F(r))$ based on its derivative:

$$G \text{ is } \mu\text{-strongly concave} \Leftrightarrow \forall (r_1, r_2) \in [0, \bar{b}]^2, (\nabla G(r_2) - \nabla G(r_1))^T (r_2 - r_1) \leq -\mu \|r_2 - r_1\|^2$$

Without loss of generality, we consider the case where $0 \leq r_1 \leq r_2 \leq \bar{b}$,

$$\nabla G(r_2) - \nabla G(r_1) = \frac{-f(r_2)}{1 - F(r_2)} - \frac{-f(r_1)}{1 - F(r_1)} = \lambda(r_1) - \lambda(r_2) \leq -\mu(r_2 - r_1).$$

Hence G is μ -strongly concave. □

B.2. Unbiased Gradient and Preservation of Concavity

Proposition 2. *Let F satisfy (A1) and Π^F be its associated monopoly revenue. Let $k \in \mathcal{K}$, then:*

- Π_k^F and p_k are $\mathcal{C}^1(\mathbb{R})$,
- $\Pi_k^F(r) = \mathbb{E}_F(p_k(r, b))$ and $\nabla \Pi_k^F(r) = \mathbb{E}_F(\nabla p_k(r, b))$,
- under (A2), Π_k^F is strictly pseudo-concave on \mathbb{R} ,
- under (A3), Π_k^F is strictly log-concave on \mathbb{R} .

Proof. The proof is a straightforward application of the convolution's properties, the Fubini-Tonelli theorem and of the stability of concavity w.r.t. convolution detailed in App. A.3.

1. Since $k \in \mathcal{C}^1(\mathbb{R}) \cap \mathcal{L}^1(\mathbb{R})$, and since Π^F and p are \mathcal{L}^1 , Π_k^F and p_k are in $\mathcal{C}^1 \cap \mathcal{L}^1$.
2. Since p , Π^F , and k are positive, so are Π_k^F and p_k . Thus, the Fubini-Tonelli theorem ensures that $\Pi_k^F(r) = \mathbb{E}_{b \sim F}(p_k(r, b))$. Further, $\nabla \Pi_k^F(r) = \mathbb{E}_{b \sim F}(\nabla p_k(r, b))$.
3. Under (A2), Π^F is strictly pseudo-concave (Prop. 1) and $k \in \mathcal{K}$ is strictly log-concave. Further, $\Pi^F(0) = \Pi^F(\bar{b}) = 0$ and we can apply Lem. 1 to guarantee the strict pseudo-concavity of Π_k^F .
4. Under (A3), Π^F is strictly log-concave and $k \in \mathcal{K}$ is strictly log-concave. Since the convolution preserves strict-concavity (see Thm. 5), Π_k^F is strictly log-concave.

□

B.3. Bias and Bounded Gradient

Proposition 3. *Let F satisfy (A1), $k \in \mathcal{K}$. Let r^* and r_k^* be the monopoly prices associated with Π^F and Π_k^F . Then, the bias $B_k = |\Pi^F(r^*) - \Pi^F(r_k^*)|$ and the instantaneous convolved gradient second moment $V_k = \max_{r \geq 0} \mathbb{E}_{b \sim F}(|\nabla p_k(r, b)|^2)$ are upper bounded by*

- $B_k \leq 2 \|\nabla \Pi^F\|_\infty \|K - \mathbf{1}_{\mathbb{R}^+}\|_1$,
- $V_k \leq 1 + \bar{b}(1 + \|\nabla \Pi^F\|_\infty) \|k\|_\infty$.

Proof.

1. The bound on B_k relies on Lem. 2, which guarantees that for all $r \geq 0$,

$$|\Pi^F(r) - \Pi_k^F(r)| \leq \|\nabla \Pi^F\|_\infty \int_{-\infty}^{\infty} |r|k(r)dr.$$

Decomposing B_k as

$$B_k \leq \Pi^F(r^*) - \Pi_k^F(r^*) + \Pi_k^F(r^*) - \Pi_k^F(r_k^*) + \Pi_k^F(r_k^*) - \Pi^F(r_k^*) \leq \Pi^F(r^*) - \Pi_k^F(r^*) + \Pi_k^F(r_k^*) - \Pi^F(r_k^*)$$

and applying two times Lem. 2 proves the desired result.

2. For all $(r, b) \in \mathbb{R}_+ \times [0, \bar{b}]$, using properties of the convolution, one has:

$$\begin{aligned} \nabla p_k(r, b) &= \int_{-\infty}^{\infty} p(\tau, b) \nabla k(r - \tau) d\tau = \int_0^b \tau \nabla k(r - \tau) d\tau \\ &= [-\tau k(r - \tau)]_0^b + \int_0^b k(r - \tau) d\tau = \int_0^b k(r - \tau) d\tau - bk(r - b). \end{aligned}$$

Since $k > 0$, and $\|k\|_1 = 1$ it is clear that

$$-bk(r-b) \leq \nabla p_k(r, b) \leq 1 \quad \Rightarrow \quad |\nabla p_k(r, b)|^2 \leq \max(1, b^2 k(r-b)^2) \leq 1 + b^2 k(r-b)^2.$$

Taking the expectation w.r.t. F (with pdf f), one obtains:

$$\mathbb{E}_B[(\nabla p_k(r, B))^2] \leq 1 + \int_0^{\bar{b}} b^2 k(r-b)^2 f(b) db \leq 1 + \|k\|_\infty \int_0^{\bar{b}} b^2 f(b) k(r-b) db.$$

Finally, under (A1), for all $b \in [0, \bar{b}]$,

$$bf(b) = 1 - F(b) - \nabla \Pi^F(b) \quad \Rightarrow \quad bf(b) \leq 1 + \|\nabla \Pi^F\|_\infty < \infty.$$

As a result, we have that

$$\mathbb{E}_B[(\nabla p_k(r, B))^2] \leq 1 + (1 + \|\nabla \Pi^F\|_\infty) \|k\|_\infty \int_0^{\bar{b}} bk(r-b) db \leq 1 + \bar{b}(1 + \|\nabla \Pi^F\|_\infty) \|k\|_\infty.$$

□

The following intermediate results provide uniform bounds on the distance between the monopoly revenue (resp. gradient) and the convoluted monopoly revenue (resp. gradient).

Lemma 2. *Let F satisfies (A1) and $k \in \mathcal{K}$ be a convolution kernel. For any $r \in [0, \bar{b}]$, we have*

$$|\Pi^F(r) - \Pi_k^F(r)| \leq \|\nabla \Pi^F\|_\infty \|K - \mathbf{1}_{\mathbb{R}^+}\|_1 = \|\nabla \Pi^F\|_\infty \int_{-\infty}^{\infty} |r|k(r) dr$$

Proof. First, notice that since $\nabla \Pi^F$ is continuous on the closed interval $[0, \bar{b}]$, it is bounded – i.e. $\|\nabla \Pi^F\|_\infty < +\infty$. Thus, integrating by parts leads to

$$\begin{aligned} |\Pi^F(r) - \Pi_k^F(r)| &= |(\Pi^F \star (\delta_0 - k))(r)| \\ &\leq \left| \underbrace{\left[\Pi^F(t)(\mathbf{1}_{\mathbb{R}^+}(r-t) - K(r-t)) \right]_{-\infty}^{+\infty}}_{=0} - \nabla \Pi^F \star (\mathbf{1}_{\mathbb{R}^+} - K)(r) \right| \\ &\leq |\nabla \Pi^F \star (\mathbf{1}_{\mathbb{R}^+} - K)(r)| \\ &\leq \|\nabla \Pi^F\|_\infty \|K - \mathbf{1}_{\mathbb{R}^+}\|_1 \quad (\text{Young's convolution inequality}) \end{aligned}$$

Finally, a last integration by parts leads to

$$\begin{aligned} \|K - \mathbf{1}_{\mathbb{R}^+}\|_1 &= \int_{-\infty}^0 K(r) dr + \int_0^{\infty} (1 - K(r)) dr \\ &= [rK(r)]_{-\infty}^0 - \int_{-\infty}^0 rk(r) dr + [r(1 - K(r))]_0^{\infty} + \int_0^{\infty} rk(r) dr \\ &= \int_{-\infty}^{\infty} |r|k(r) dr. \end{aligned}$$

□

Then, another important bias to control is the one of the gradient.

Lemma 3. *Assuming F satisfies (A1) and $k \in \mathcal{K}$ be a convolution kernel. For any $r \in [0, \bar{b}]$, we have*

$$|\nabla \Pi_k^F(r) - \nabla \Pi^F(r)| \leq \|\nabla^2 \Pi^F\|_\infty \|K - \mathbf{1}_{\mathbb{R}^+}\|_1 = \|\nabla^2 \Pi^F\|_\infty \int_{-\infty}^{\infty} |r|k(r) dr$$

Proof. The proof is essentially the same as the one of Lemma 2. First, note that since $\nabla^2 \Pi^F$ is continuous on the closed interval $[0, \bar{b}]$, it is bounded – i.e. $\|\nabla^2 \Pi^F\|_\infty < +\infty$.

$$\begin{aligned}
 |\nabla \Pi_k^F(r) - \nabla \Pi^F(r)| &= |(\nabla \Pi^F \star (k - \delta_0))(r)| \\
 &= \left| \underbrace{\left[\nabla \Pi^F(x)(K(r-x) - \mathbf{1}_{\mathbb{R}^+}(r-x)) \right]_{-\infty}^{+\infty}}_{=0} - \nabla^2 \Pi^F \star (K - \mathbf{1}_{\mathbb{R}^+})(r) \right| \\
 &= |\nabla^2 \Pi^F \star (K - \mathbf{1}_{\mathbb{R}^+})(r)| \\
 &\leq \|\nabla^2 \Pi^F\|_\infty \|K - \mathbf{1}_{\mathbb{R}^+}\|_1 \quad (\text{Young's convolution inequality}) \\
 &= \|\nabla^2 \Pi^F\|_\infty \int_{-\infty}^{\infty} |r| k(r) dr.
 \end{aligned}$$

□

C. Proofs of Sec. 4

In this section we consider only a stationary F , and therefore, for simplicity, we will denote Π^F simply by Π .

C.1. Almost Sure Convergence

Theorem 1. *Let F satisfy (A1) and (A2) and $\{k_t\}_{t \in \mathbb{N}} \in \mathcal{K}^{\mathbb{N}}$. Then, by running V-CONV-OGA with $\mathcal{C} = [0, \bar{b}]$, we have*

$$r_t \xrightarrow{\text{a.s.}} r^*$$

as long as $\sum_{t=1}^{+\infty} \gamma_t = +\infty$, $\sum_{t=1}^{+\infty} \gamma_t \|K_t - \mathbf{1}_{\mathbb{R}^+}\|_1 < +\infty$ and $\sum_{t=1}^{+\infty} \gamma_t^2 \|k_t\|_\infty < +\infty$.

Proof. The proof inherits a lot from classical methods, see e.g. Bottou (1998). The main difference lies in the type of “concavity” required. The proof in Bottou (1998) is derived for *variationally coherent* functions i.e. those such that

$$\forall t \in \mathbb{N}, \forall \epsilon > 0, \sup_{(r-r_{k_t}^*)^2 > \epsilon} (r - r_{k_t}^*)^T \nabla \Pi_{k_t}(r) < 0.$$

However, such assumption is clearly violated here since $\nabla \Pi_{k_t}(r) \rightarrow 0$ as $r \rightarrow \infty$. Nevertheless since Π_{k_t} is strictly pseudo-concave and strictly positive, one can obtain a similar result.

Following Bottou (1998), we introduce the Lyapunov process $h_t = \|r_t - r^*\|^2$. Using the fact that the projection operator is a contraction, one obtains:

$$\begin{aligned}
 h_{t+1} &= (\text{proj}_{\mathcal{C}}(r_t + \gamma_t \nabla p_{k_t}(r_t, b_t)) - r^*)^2 \\
 &\leq (r_t + \gamma_t \nabla p_{k_t}(r_t, b_t) - r^*)^2 \\
 &\leq h_t + 2\gamma_t (r_t - r^*)^T \nabla p_{k_t}(r_t, b_t) + \gamma_t^2 (\nabla p_{k_t}(r_t, b_t))^2.
 \end{aligned}$$

Hence, h_t satisfies the recursion:

$$h_{t+1} - h_t \leq 2\gamma_t (r_t - r^*)^T \nabla p_{k_t}(r_t, b_t) + \gamma_t^2 (\nabla p_{k_t}(r_t, b_t))^2$$

Taking the conditional expectation w.r.t. $\mathcal{F}_t = \sigma(b_0, \dots, b_{t-1}, r_0, \dots, r_t, \gamma_0, \dots, \gamma_t)$, one obtains:

$$\begin{aligned}
 \mathbb{E}[h_{t+1} - h_t | \mathcal{F}_t] &\leq 2\gamma_t (r_t - r^*)^T \mathbb{E}[\nabla p_{k_t}(r_t, b_t) | \mathcal{F}_t] + \gamma_t^2 \mathbb{E}[(\nabla p_{k_t}(r_t, b_t))^2 | \mathcal{F}_t] \\
 &\leq 2\gamma_t (r_t - r^*)^T \nabla \Pi_{k_t}(r_t) + \gamma_t^2 \mathbb{E}[(\nabla p_{k_t}(r_t, b_t))^2 | \mathcal{F}_t]
 \end{aligned} \tag{10}$$

as Prop. 2 provides that $\mathbb{E}[\nabla p_{k_t}(r_t, b_t) | \mathcal{F}_t] = \nabla \Pi_{k_t}(r_t)$. Then, we decompose the gradient term to isolate the gradient bias:

$$\mathbb{E} [h_{t+1} - h_t | \mathcal{F}_t] \leq \underbrace{2\gamma_t(r_t - r^*)^T \nabla \Pi(r_t)}_{\leq 0 \text{ by pseudo-concavity}} + \underbrace{2\gamma_t(r_t - r^*)^T (\nabla \Pi_{k_t}(r_t) - \nabla \Pi(r_t))}_{\text{bias}} + \underbrace{\gamma_t^2 \mathbb{E} \left[(\nabla p_{k_t}(r_t, b_t))^2 | \mathcal{F}_t \right]}_{\text{Bounded gradient}} \quad (11)$$

The first term in Eq. 11 is negative by the pseudo-concavity of Π (Prop. 1). The second term is bounded by $2\gamma_t \|\nabla^2 \Pi\|_\infty \|K - \mathbb{1}_{\mathbb{R}^+}\|_1$ by Lem. 3. The third term is bounded by $\gamma_t^2 (1 + \bar{b}(1 + \|\nabla \Pi\|_\infty) \|k\|_\infty)$ by Prop. 3. Using the same quasi-martingale argument as in Bottou (1998), we have that Eq. 11, together with $\sum \gamma_t \|K_t - \mathbb{1}_{\mathbb{R}^+}\|_1 < \infty$ and $\sum \gamma_t^2 \|k_t\|_\infty < \infty$, implies that

$$h_t \xrightarrow{a.s.} h_\infty < \infty, \quad \sum \mathbb{E} [h_{t+1} - h_t | \mathcal{F}_t] < \infty.$$

Thus, using Eq. 10, we have that

$$0 \leq \sum \gamma_t (r^* - r_t)^T \nabla \Pi(r_t) < \infty. \quad (12)$$

Suppose now that $(r_t - r^*) \xrightarrow{a.s.} h_\infty \neq 0$, then since $\sum \gamma_t = +\infty$, it would lead to $\sum \gamma_t (r^* - r_t)^T \nabla \Pi(r_t) = +\infty$ which is in contradiction with Eq. 12. As a result,

$$(r_t - r^*)^T \nabla \Pi(r_t) \xrightarrow{a.s.} 0.$$

Finally, since $(r_t - r^*)^2 \xrightarrow{a.s.} h_\infty < \infty$, necessarily, $\nabla \Pi(r_t) \xrightarrow{a.s.} 0$ and $r_t \xrightarrow{a.s.} r^*$. \square

C.2. Finite-time Convergence Speed

We provide here the full statement of Thm. 2, along with explicit rates and constants. The rates are expressed in terms of an auxiliary function φ , which allows us to handle all configurations of the step-size and kernel decay schedules. Depending on the value of $\alpha, \alpha_1, \alpha_\infty$, it may introduce some logarithmic term $\log(t)$. This explains the \tilde{O} notation used in the abridged version of Thm. 2 (Sec. 4), which can easily be recovered from this extended version.

Theorem 2. *Let F satisfy (A1) and (A3), \mathcal{C} be given in (A4) and $\{k_t\}_{t \in \mathbb{N}} \in \mathcal{K}^{\mathbb{N}}$ such that $\|K_t - \mathbb{1}_{\mathbb{R}^+}\|_1 \leq \nu_1 t^{-\alpha_1}$ and $\|k_t\|_\infty \leq \nu_\infty t^{\alpha_\infty}$. Then, by running V-CONV-OGA on \mathcal{C} with $\gamma_t = \nu t^{-\alpha}$ with $\nu \leq (2c\mu)^{-1}$, we have for all $t \geq 2$,*

$$\begin{aligned} \text{if } \alpha = 1 \quad & \mathbb{E}(\|r_t - r^*\|^2) \leq \left(\bar{b}^2 + 2C\nu\nu_1\varphi_{2\mu c\nu - \alpha_1}(t) + 2C_\infty\nu^2\nu_\infty\varphi_{2\mu c\nu + \alpha_\infty - 1}(t) \right) t^{-2\mu c\nu} \\ \text{if } \alpha \in (0, 1) \quad & \mathbb{E}(\|r_t - r^*\|^2) \leq \left(\bar{b}^2 + C_1\nu\nu_1\varphi_{1-\alpha-\alpha_1}(t) + C_\infty\nu^2\nu_\infty\varphi_{1+\alpha_\infty-2\alpha}(t) \right) \exp(-\mu c\nu t^{1-\alpha}) \\ & + C_1 \frac{\nu_1}{\mu c} t^{-\alpha_1} + C_\infty \frac{\nu\nu_\infty}{\mu c} t^{\alpha_\infty - \alpha} \end{aligned}$$

as long as $\alpha, \alpha_1, \alpha_\infty$ satisfy the condition of Thm. 1 i.e., $\alpha \leq 1, \alpha + \alpha_1 > 1$ and $2\alpha - \alpha_\infty > 1$. The function φ and the constant C_1, C_∞ are given by

$$\varphi_\beta(t) = \log(t)\mathbb{1}_{\beta=0} + \frac{t^\beta - 1}{\beta}\mathbb{1}_{\beta \neq 0}; \quad C_1 = 2\bar{b}\|\nabla^2 \Pi\|_\infty; \quad C_\infty = 1 + \bar{b}(1 + \|\nabla \Pi\|_\infty).$$

Proof. The proof builds on Bach & Moulines (2011, Thm. 2). The main differences are that **1**) we don't require the local function p_{k_t} to be concave **2**) we don't rely on the strong concavity of Π_{k_t} but on its strong log-concavity and one of its lower bounds **3**) our objective function varies over time because of the sequence of convolution kernels $\{k_t\}_{t \geq 1}$.

We first stress that (A3) together with the lower bounded revenue Π_{k_t} leads to some sort of local strong concavity of Π . From Prop. 1, the strongly increasing hazard rate ensures that Π is strongly log-concave with parameter μ . Further, Π admits a unique maximum r^* (since $\Pi(0) = \Pi(\beta) = 0$) such that $r^* \in \mathcal{C}$ by assumption. As a result, for any $r \in \mathcal{C}$,

$$\begin{aligned} (r - r^*)^T (\nabla(\log \Pi)(r) - \nabla(\log \Pi)(r^*)) & \leq -\mu \|r^* - r\|^2 \\ \Rightarrow (r - r^*)^T (\nabla \Pi(r)/\Pi(r) - \nabla \Pi(r^*)/\Pi(r^*)) & \leq -\mu \|r^* - r\|^2 \\ \Rightarrow (r - r^*)^T \nabla \Pi(r) & \leq -\Pi(r)\mu \|r^* - r\|^2 \leq -\mu c \|r^* - r\|^2. \end{aligned}$$

We denote as $\tilde{\mu} = \mu c$ the quantity which plays the role of the strong-concavity parameter in Bach & Moulines (2011, Thm. 2).

As for the proof of Thm. 1, we introduce the Lyapunov process $h_t = \|r_t - r^*\|_2^2$ and its expectation $\bar{h}_t = \mathbb{E}(h_t)$. From Eq. 11, we have

$$\mathbb{E}(h_{t+1} - h_t) | \mathcal{F}_t \leq 2\gamma_t (r_t - r^*)^T \nabla \Pi(r_t) + 2\gamma_t \bar{b} \|\nabla^2 \Pi\|_\infty \|K_t - \mathbb{1}_{\mathbb{R}^+}\|_1 + \gamma_t^2 (1 + \bar{b}(1 + \|\nabla \Pi\|_\infty)) \|k\|_\infty.$$

Using the local strong-concavity of Π and that $\|K_t - \mathbb{1}_{\mathbb{R}^+}\|_1 \leq \gamma_t^1$, $\|k\|_\infty \leq \gamma_t^\infty$, we obtain:

$$\mathbb{E}(h_{t+1} - h_t) | \mathcal{F}_t \leq 2\tilde{\mu}\gamma_t h_t + C_1 \gamma_t \gamma_t^1 + C_\infty \gamma_t^2 \gamma_t^\infty,$$

where $C_1 = 2\bar{b}\|\nabla^2 \Pi\|_\infty$ and $C_\infty = 1 + \bar{b}(1 + \|\nabla \Pi\|_\infty)$.⁵ Taking the expectation leads to:

$$\bar{h}_{t+1} \leq (1 - 2\tilde{\mu}\gamma_t) \bar{h}_t + C_1 \gamma_t \gamma_t^1 + C_\infty \gamma_t^2 \gamma_t^\infty. \quad (13)$$

In line with Bach & Moulines (2011), we split the proof depending whether $\alpha = 1$ or $\alpha \in (0, 1)$.

1. The case $\alpha = 1$: using that $1 - x \leq \exp(-x)$ for all $x \in \mathbb{R}$ and applying the recursion t times in Eq. 13, we have

$$\begin{aligned} \bar{h}_t &\leq \bar{h}_1 \exp\left(-2\tilde{\mu} \sum_{s=1}^{t-1} \gamma_s\right) + C_1 \sum_{s=1}^{t-1} \gamma_s \gamma_s^1 \exp\left(-2\tilde{\mu} \sum_{\tau=s+1}^{t-1} \gamma_\tau\right) + C_\infty \sum_{s=1}^{t-1} \gamma_s^2 \gamma_s^\infty \exp\left(-2\tilde{\mu} \sum_{\tau=s+1}^{t-1} \gamma_\tau\right) \\ &\leq \bar{h}_1 \exp\left(-2\tilde{\mu}\nu \sum_{s=1}^{t-1} s^{-1}\right) + \left(C_1 \nu \nu_1 \sum_{s=1}^{t-1} s^{-\alpha-\alpha_1} + C_\infty \nu^2 \nu_\infty \sum_{s=1}^{t-1} s^{-2\alpha+\alpha_0}\right) \exp\left(-2\tilde{\mu}\nu \sum_{\tau=s+1}^{t-1} \tau^{-1}\right). \end{aligned}$$

Further, for all $t \geq 2$,

$$\begin{aligned} \sum_{s=1}^{t-1} s^{-1} &\geq \log(t) \\ \sum_{\tau=s+1}^{t-1} \tau^{-1} &\geq \log(t/s + 1) \end{aligned}$$

we obtain that (under $\tilde{\mu}\nu \leq 1/2$):

$$\begin{aligned} \bar{h}_t &\leq \bar{h}_1 t^{-2\tilde{\mu}\nu} + C_1 \nu \nu_1 t^{-2\tilde{\mu}\nu} \sum_{s=1}^{t-1} s^{-1-\alpha_1} (s+1)^{2\tilde{\mu}\nu} + C_\infty \nu^2 \nu_\infty t^{-2\tilde{\mu}\nu} \sum_{s=1}^{t-1} s^{-2+\alpha_\infty} (s+1)^{2\tilde{\mu}\nu} \\ &\leq \bar{h}_1 t^{-2\tilde{\mu}\nu} + 2C_1 \nu \nu_1 t^{-2\tilde{\mu}\nu} \sum_{s=1}^{t-1} s^{2\tilde{\mu}\nu-1-\alpha_1} + 2C_\infty \nu^2 \nu_\infty t^{-2\tilde{\mu}\nu} \sum_{s=1}^{t-1} s^{-2+\alpha_\infty+2\tilde{\mu}\nu}, \\ &\leq \left(\bar{b}^2 + 2C_1 \nu \nu_1 \varphi_{2\tilde{\mu}\nu-\alpha_1}(t) + 2C_\infty \nu^2 \nu_\infty \varphi_{2\tilde{\mu}\nu+\alpha_\infty-1}(t)\right) t^{-2\tilde{\mu}\nu}. \end{aligned}$$

2. The case $\alpha \in (0, 1)$: applying the recursion t times in Eq. 13, we have

$$\bar{h}_t \leq \bar{h}_1 \underbrace{\prod_{s=1}^{t-1} (1 - 2\tilde{\mu}\gamma_s)}_{A_t^1} + C_1 \sum_{s=1}^{t-1} \gamma_s \gamma_s^1 \underbrace{\prod_{\tau=s+1}^{t-1} (1 - 2\tilde{\mu}\gamma_\tau)}_{A_t^2} + C_\infty \sum_{s=1}^{t-1} \gamma_s^2 \gamma_s^\infty \underbrace{\prod_{\tau=s+1}^{t-1} (1 - 2\tilde{\mu}\gamma_\tau)}_{A_t^3} \quad (14)$$

⁵Without loss of generality, we assume that $\nu_\infty \geq 1$.

The derivation slightly differs from the case $\alpha = 1$. Following [Bach & Moulines \(2011\)](#), one has:

$$\begin{aligned} A_t^1 &\leq \exp\left(-2\tilde{\mu}\sum_{s=1}^{t-1}\gamma_s\right) \\ A_t^2 &\leq \frac{\gamma_{\lfloor t/2 \rfloor}^1}{2\tilde{\mu}} + \exp\left(-2\tilde{\mu}\sum_{\tau=\lfloor t/2 \rfloor}^{t-1}\gamma_\tau\right)\sum_{s=1}^{t-1}\gamma_s\gamma_s^1 \\ A_t^3 &\leq \frac{\gamma_{\lfloor t/2 \rfloor}\gamma_{\lfloor t/2 \rfloor}^\infty}{2\tilde{\mu}} + \exp\left(-2\tilde{\mu}\sum_{\tau=\lfloor t/2 \rfloor}^{t-1}\gamma_\tau\right)\sum_{s=1}^{t-1}\gamma_s^2\gamma_s^\infty. \end{aligned}$$

Using the expression of γ , γ_1 and γ_∞ , where $\alpha_1 \leq 1$, together with $\varphi_{1-\alpha}(x) - \varphi_{1-\alpha}(x/2) \geq x^{1-\alpha}/2$ and $\varphi_{1-\alpha}(t) \geq t^{1-\alpha}/2$ we have:

$$\begin{aligned} A_t^1 &\leq \exp(-2\tilde{\mu}\nu\varphi_{1-\alpha}(t)) \leq \exp(-\tilde{\mu}\nu t^{1-\alpha}), \\ A_t^2 &\leq \frac{\nu_1}{\tilde{\mu}}t^{-\alpha_1} + \nu\nu_1 \exp(-\tilde{\mu}\nu t^{1-\alpha})\sum_{s=1}^{t-1}s^{-\alpha-\alpha_1} \leq \frac{\nu_1}{\tilde{\mu}}t^{-\alpha_1} + \nu\nu_1 \exp(-\tilde{\mu}\nu t^{1-\alpha})\varphi_{1-\alpha-\alpha_1}(t), \\ A_t^3 &\leq \frac{\nu\nu_\infty}{\tilde{\mu}}t^{\alpha_\infty-\alpha} + \nu^2\nu_\infty \exp(-\tilde{\mu}\nu t^{1-\alpha})\sum_{s=1}^{t-1}s^{\alpha_\infty-2\alpha} \leq \frac{\nu\nu_\infty}{\tilde{\mu}}t^{\alpha_\infty-\alpha} + \nu^2\nu_\infty \exp(-\tilde{\mu}\nu t^{1-\alpha})\varphi_{1+\alpha_\infty-2\alpha}(t). \end{aligned}$$

Putting everything together, we obtain the final bound for $\alpha \in (0, 1)$:

$$\bar{h}_t \leq \left(\bar{b}^2 + C_1\nu\nu_1\varphi_{1-\alpha-\alpha_1}(t) + C_\infty\nu^2\nu_\infty\varphi_{1+\alpha_\infty-2\alpha}(t)\right) \exp(-\tilde{\mu}\nu t^{1-\alpha}) + C_1\frac{\nu_1}{\tilde{\mu}}t^{-\alpha_1} + C_\infty\frac{\nu\nu_\infty}{\tilde{\mu}}t^{\alpha_\infty-\alpha}.$$

□

D. Proof of Sec. 5

Theorem 3. *Let F satisfy (A1) and (A3) with parameter μ , let \mathcal{C} and c be as in (A4) and $k \in \mathcal{K}$. Then, by running CONV-OGA on \mathcal{C} with a constant stepsize $\gamma > 0$, for any $t \geq 1$ we have*

$$\mathbb{E}(\|r_t - r_k^*\|^2) \leq (\bar{b}^2 + C(\gamma, k)(t-1))e^{-\frac{\mu c \gamma}{2}t} + \frac{2C(\gamma, k)}{\mu c}$$

where $C(\gamma, k) = \mathcal{O}(\gamma\|K - \mathbb{1}_{\mathbb{R}^+}\|_1 + \gamma^2\|k\|_\infty)$.

Proof. Similarly to the one of Th.2, this proof builds on the one of [Bach & Moulines \(2011\)](#).

Since Π^F is μ -strongly log-concave, one has for all $r \in \mathcal{C}$,

$$(r_t - r^*)^T \nabla \Pi^F(r) \leq -\Pi^F(r)\mu\|r - r^*\|^2 \leq -\tilde{\mu}\|r - r^*\|^2 \quad (15)$$

where $\tilde{\mu} = \mu c$. As a result, although we do not assume the function to be strongly concave, it still enjoys a similar property in r^* on the bounded subset \mathcal{C} . Then, let $h_t = \|r_t - r^*\|^2$ be the Lyapunov process (similarly to the proof of Thm. 1). Since the projection operator over \mathcal{C} is 1-Lipschitz, from Eq. 11, one has:

$$\mathbb{E}[h_{t+1} - h_t | \mathcal{F}_t] \leq 2\gamma(r_t - r^*)^T \nabla \Pi^F(r_t) + 2\gamma\bar{b}\|\nabla^2 \Pi^F\|_\infty\|K - \mathbb{1}_{\mathbb{R}^+}\|_1 + \gamma_t^2(1 + \bar{b}(1 + \|\nabla \Pi^F\|_\infty))\|k\|_\infty. \quad (16)$$

Denoting $C(\gamma, k) = 2\gamma\bar{b}\|\nabla^2 \Pi^F\|_\infty\|K - \mathbb{1}_{\mathbb{R}^+}\|_1 + \gamma^2(1 + \bar{b}(1 + \|\nabla \Pi^F\|_\infty))\|k\|_\infty$ and $\bar{h}_t = \mathbb{E}(h_t)$, and then taking the expectation in Eq. 16, one obtains:

$$\bar{h}_{t+1} \leq (1 - 2\gamma_t\tilde{\mu})\bar{h}_t + C(\gamma, k). \quad (17)$$

Further, Eq. 17 is exactly the same as Eq. 25 in (Bach & Moulines, 2011) with different definitions for the constants, and the rest of the proof is identical. As a result, one has:

$$\bar{h}_t \leq (\bar{h}_0 + C(\gamma, k)(t-1)) \exp\left(-\frac{\tilde{\mu}\gamma}{2}t\right) + \frac{2C(\gamma, k)}{\tilde{\mu}}.$$

□

Corollary 2. Let $\{F_t\}_{t \geq 1}$ satisfy (A1), (A3), (A4), (A5) and $k \in \mathcal{K}$. Then, there exists $\Xi(k, \gamma)$ and $\Omega(k, \gamma)$ such that CONV-OGA has a non-stationary regret of

$$R(T) \leq \Xi(k, \gamma)T + \Omega(k, \gamma)\tau.$$

Further, if the horizon T is known in advance, running CONV-OGA with $\gamma = T^{-\frac{1}{2}}$ and k a kernel with $\|K - \mathbb{1}_{\mathbb{R}^+}\|_1 \leq T^{-\frac{1}{2}}$ and $\|k\|_\infty \leq T^{\frac{1}{2}}$, then $R(T) = \mathcal{O}(\sqrt{T})$.

Proof. Denoting $\{[s_i, t_i]\}_{i=1}^\tau$ the intervals on which the distribution is constant, we have

$$\begin{aligned} R(T) &= \mathbb{E} \left(\sum_{t=1}^T \Pi^{F_t}(r_t^*) - \Pi^{F_t}(r_t) \right) \\ &\leq \mathbb{E} \left(\sum_{t=1}^T \frac{\|\nabla^2 \Pi^{F_t}\|_\infty}{2} \|r_t^* - r_t\|_2^2 \right) \\ &\leq \sum_{i=1}^\tau \frac{\|\nabla^2 \Pi^{F_{s_i}}\|_\infty}{2} \sum_{t=s_i}^{t_i} \mathbb{E} (\|r_{s_i}^* - r_t\|_2^2) \\ &\leq \sum_{i=1}^\tau \frac{\|\nabla^2 \Pi^{F_{s_i}}\|_\infty}{2} \sum_{t=1}^{t_i - s_i + 1} \mathbb{E} (\|r_{s_i}^* - r_{t+s_i-1}\|_2^2) \end{aligned}$$

Applying Thm. 3 on $\mathbb{E} (\|r_{s_i}^* - r_{t+s_i-1}\|_2^2)$ and denoting $\bar{C}(\gamma, k) = \max_{i \in [\tau]} 2\gamma\bar{b}\|\nabla^2 \Pi^{F_{s_i}}\|_\infty \|K - \mathbb{1}_{\mathbb{R}^+}\|_1 + \gamma^2(1 + \bar{b}(1 + \|\nabla \Pi^{F_{s_i}}\|_\infty)\|k\|_\infty)$ we obtain

$$\begin{aligned} R(T) &\leq \sum_{i=1}^\tau \frac{\|\nabla^2 \Pi^{F_{s_i}}\|_\infty}{2} \sum_{t=1}^{t_i - s_i + 1} \left((\bar{b}^2 + \bar{C}(\gamma, k)(t-1))e^{-\frac{\mu c \gamma}{2}t} + \frac{2\bar{C}(\gamma, k)}{\mu c} \right) \\ &\leq \frac{1}{2} \max_{i \in [\tau]} \|\nabla^2 \Pi^{F_{s_i}}\|_\infty \left(\frac{2\bar{C}(\gamma, k)}{\mu c} T + \sum_{i=1}^\tau \sum_{t=1}^{t_i - s_i + 1} \bar{b}^2 e^{-\frac{\mu c \gamma}{2}t} + \bar{C}(\gamma, k)(t-1)e^{-\frac{\mu c \gamma}{2}t} \right) \\ &\leq \frac{1}{2} \max_{i \in [\tau]} \|\nabla^2 \Pi^{F_{s_i}}\|_\infty \left(\frac{2\bar{C}(\gamma, k)}{\mu c} T + \sum_{i=1}^\tau \sum_{t=0}^{t_i - s_i} \bar{b}^2 e^{-\frac{\mu c \gamma}{2}(t+1)} + \bar{C}(\gamma, k)te^{-\frac{\mu c \gamma}{2}(t+1)} \right) \\ &\leq \frac{1}{2} \max_{i \in [\tau]} \|\nabla^2 \Pi^{F_{s_i}}\|_\infty \left(\frac{2\bar{C}(\gamma, k)}{\mu c} T + \sum_{i=1}^\tau \left(\frac{\bar{b}^2 e^{-\frac{\mu c \gamma}{2}}}{1 - e^{-\frac{\mu c \gamma}{2}}} + \sum_{t=0}^{t_i - s_i} \bar{C}(\gamma, k)te^{-\frac{\mu c \gamma}{2}(t+1)} \right) \right) \\ &\leq \frac{1}{2} \max_{i \in [\tau]} \|\nabla^2 \Pi^{F_{s_i}}\|_\infty \left(\frac{2\bar{C}(\gamma, k)}{\mu c} T + \left(\frac{\bar{b}^2}{e^{\frac{\mu c \gamma}{2}} - 1} \tau + \bar{C}(\gamma, k)e^{-\mu c \gamma} \sum_{i=1}^\tau \frac{1 - e^{-\frac{\mu c \gamma}{2}(t_i - s_i)}}{(1 - e^{-\frac{\mu c \gamma}{2}})^2} (1 + (t_i - s_i)(1 - e^{-\frac{\mu c \gamma}{2}})) \right) \right) \\ &\leq \frac{1}{2} \max_{i \in [\tau]} \|\nabla^2 \Pi^{F_{s_i}}\|_\infty \left(\frac{2\bar{C}(\gamma, k)}{\mu c} T + \left(\frac{\bar{b}^2}{e^{\frac{\mu c \gamma}{2}} - 1} \tau + \frac{\bar{C}(\gamma, k)e^{-\mu c \gamma}}{(1 - e^{-\frac{\mu c \gamma}{2}})^2} \tau \right) \right) \\ &\leq \frac{1}{2} \max_{i \in [\tau]} \|\nabla^2 \Pi^{F_{s_i}}\|_\infty \left(\frac{2\bar{C}(\gamma, k)}{\mu c} T + \left(\frac{\bar{b}^2}{e^{\frac{\mu c \gamma}{2}} - 1} + \frac{\bar{C}(\gamma, k)}{(e^{\frac{\mu c \gamma}{2}} - 1)^2} \right) \tau \right) \end{aligned}$$

Getting $R(T) = \mathcal{O}(\sqrt{T})$ when T is known in advance just amounts to plugging $\gamma = \frac{1}{\sqrt{T}}$, $\|K - \mathbb{1}_{\mathbb{R}^+}\|_1 \propto \frac{1}{\sqrt{T}}$ and $\|k\|_\infty \propto \sqrt{T}$ in the last equation. □