Optimal Non-parametric Learning in Repeated Contextual Auctions with Strategic Buyer: SUPPLEMENTARY MATERIALS*

Alexey Drutsa†

12 July 2020

Contents

A Missed proofs

A.1 Proof of Proposition 1

Here we present some details for the proof of Proposition 1. Some ideas used here are similar to the ones used in [7, 2, 4, 5].

I. Let

$$S_X(\sigma) = \mathbb{E}[\sum_{s=t}^{T} \gamma^{s-1} a_s (v(x_s) - p_s) \mid x_s \in X, \sigma] \tag{A.1}$$

be the future expected surplus of the buyer over those rounds $s \geq t$ whose feature vector x_s belongs to a set $X \subseteq \mathbb{X}$ when he follows a strategy σ[1]. It is easy to see that the full future surplus will be

*This is the supplementary materials for the work published in Proceedings of the 37-th International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020

†Yandex, Moscow, Russia (www.yandex.com); Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow, Russia; e-mail: adrutsa@yandex.ru.

[1]A buyer (pure) strategy is a map $\sigma : \sqcup_{t=1}^{T} \mathfrak{S}_t \to \{0, 1\}$ that maps any state $\mathfrak{s} \in \mathfrak{S}_t$ in a round t to a decision $\sigma(\mathfrak{s}) \in \{0, 1\}$. A mixed strategy is a probability distribution over pure strategies.

$S_{\mathbb{X}}(\sigma)$ and, for any $X \subseteq \mathbb{X}$ and any strategy σ, it can be decomposed as follows:

$$S_{\mathbb{X}}(\sigma) = S_X(\sigma) + S_{\mathbb{X}\setminus X}(\sigma). \tag{A.2}$$

The considered set $\tilde{X}$ is isolated (from $\mathbb{X} \setminus \tilde{X}$) after the current round t (as in Definition 3). Let two strategies σ_1 and σ_2 are s.t. buyer decisions coincide (between the strategies) in all rounds where x_s belongs to a set $\mathbb{X} \setminus X$ (outside of $\tilde{X}$), i.e., formally, in all the states $\mathfrak{L}(\tilde{\mathfrak{s}}, \mathbb{X} \setminus X) \cup \mathfrak{R}(\tilde{\mathfrak{s}}, \mathbb{X} \setminus X)$. Then $S_{\mathbb{X}\setminus\tilde{X}}(\sigma_1) = S_{\mathbb{X}\setminus\tilde{X}}(\sigma_2)$.

Therefore, when we compare strategies that differ in decisions within $\tilde{X}$ only, it is enough to investigate $S_{\tilde{X}}(\sigma)$ with strategies σ that are equal outside of $\tilde{X}$.

II. In our case (see the proof of Proposition 1 in the main text), we have two strategies: the optimal one σ^{Opt} and the strategy σ' that coincides with σ^{Opt} in all states outside $\tilde{X}$. So, applying the arguments above, we have (due to optimality):

$$S_{\mathbb{X}}(\sigma^{\mathrm{Opt}}) \geq S_{\mathbb{X}}(\sigma') \Rightarrow S_{\tilde{X}}(\sigma^{\mathrm{Opt}}) + S_{\mathbb{X}\setminus\tilde{X}}(\sigma^{\mathrm{Opt}}) \geq S_{\tilde{X}}(\sigma') + S_{\mathbb{X}\setminus\tilde{X}}(\sigma') \tag{A.3}$$

$$\Rightarrow S_{\tilde{X}}(\sigma^{\mathrm{Opt}}) + S_{\mathbb{X}\setminus\tilde{X}}(\sigma^{\mathrm{Opt}}) \geq S_{\tilde{X}}(\sigma') + S_{\mathbb{X}\setminus\tilde{X}}(\sigma^{\mathrm{Opt}}) \Rightarrow S_{\tilde{X}}(\sigma^{\mathrm{Opt}}) \geq S_{\tilde{X}}(\sigma'). \tag{A.4}$$

III. Note that, in the proof in the main text, we implicitly use that the strategy σ^{Opt} is pure (at least w.r.t. the decision in the round t). This is done for the sake of exposition, and the result holds for a mixed strategy as well. Indeed, let σ_0 be an optimal strategy among all the strategies that have decision $a_t = 0$ in the round t. Similarly, let σ_1 be an optimal strategy among all the strategies that have decision $a_t = 1$ in the round t. Then a mixed σ^{Opt} is a linear combination of these strategies: $\sigma^{\mathrm{Opt}} = \alpha\sigma_1 + (1-\alpha)\sigma_0, \alpha \in (0,1)$. Instead of σ' consider $\sigma'' = \alpha\sigma_1 + (1-\alpha)\sigma'_0$, where σ'_0 be the strategy, where the buyer accepts the good in the round t and rejects each future good from $\tilde{X}$ (σ'_0 coincides with σ_0 for goods in $\mathbb{X} \setminus \tilde{X}$).

Hence, due to linearity of surplus (since it is an expectation), we have:

$$S_{\mathbb{X}}(\sigma^{\mathrm{Opt}}) \geq S_{\mathbb{X}}(\sigma'') \Rightarrow \alpha S_{\mathbb{X}}(\sigma_1) + (1-\alpha)S_{\mathbb{X}}(\sigma_0) \geq \alpha S_{\mathbb{X}}(\sigma_1) + (1-\alpha)S_{\mathbb{X}}(\sigma'_0) \Rightarrow S_{\mathbb{X}}(\sigma_0) \geq S_{\mathbb{X}}(\sigma'_0) \tag{A.5}$$

In further steps (e.g., in Eq. (A.3)), use σ_0 instead of σ^{Opt} and σ'_0 instead of σ'.

A.2 Proof of Proposition 2

Proof of Proposition 2. Similarly to the proof of Prop. 1 (see also details in App. A.1), we analyze buyer surplus $S_{\tilde{X}}(\sigma)$ calculated only over rounds from $\tilde{X}$ (due to isolation of $\tilde{X}$ from $\mathbb{X} \setminus \tilde{X}$). Let σ^{Opt} be the optimal strategy of the buyer in the round t, what implies that $S_{\tilde{X}}(\sigma^{\mathrm{Opt}}) \geq 0$, because the buyer can reject all goods in $\tilde{X}$ and get at least 0. The left-hand side of the inequality can be upper bounded as follows

$$\begin{aligned} S_{\tilde{X}}(\sigma^{\mathrm{Opt}}) &= \gamma^{t-1}(v(x_t) - p_t) + \mathbb{E}[\sum_{s=t}^{T} \gamma^{s-1} a_s(v(x_s) - p_s) \mid x_s \in \tilde{X}, \sigma^{\mathrm{Opt}}] \\ &\leq \gamma^{t-1}(v(x_t) - p(\mathfrak{s})) + \sum_{s=t+1}^{T} \gamma^{s-1} \sup_{x\in\tilde{X}}[v(x) - u_{\mathfrak{s},1}(x)], \end{aligned} \tag{A.6}$$

where we upper bounded instant surpluses in all future rounds as maximal possible ones by $\sup_{x\in\tilde{X}}(v(x) - u_{\mathfrak{s},1}(x))$. The latter expression can be trivially bounded by $v(x_t) + L\mathrm{diam}(\tilde{X}) - \inf_{x\in\tilde{X}} u_{\mathfrak{s},1}(x)$ (see Def. 1 for $u_{\mathfrak{s},1}(\cdot)$). Combining all inequalities and dividing by γ^{t-1}, one gets:

$$v(x_t) - p(\mathfrak{s}) + \frac{\gamma}{1-\gamma}\Big(v(x_t) + L\mathrm{diam}(\tilde{X}) - \inf_{x\in\tilde{X}} u_{\mathfrak{s},1}(x)\Big) \geq 0, \tag{A.7}$$

what implies the proposition after using the condition on $p(\mathfrak{s}) - \inf_{x\in\tilde{X}} u_{\mathfrak{s},1}(x)$ and term rearrangement. ∎

A.3 Proof of Lemma 2

Proof of Lemma 2. First, we prove (a) under the assumption: let in an exploration round t, the current u^X and w^X are s.t. $(2\eta+3)L\text{diam}(X) \le w^X - u^X$. This implies w^X: $(w^X - u^X) - 2(\eta+1)L\text{diam}(X) \ge L\text{diam}(X)$. Since the gap $w^X - u^X$ reduces exactly by $L\text{diam}(X)$ after an acceptance and exactly by w^X: $(w^X - u^X) - 2(\eta+1)L\text{diam}(X)$ after a rejection, we get the statement (a) for this round.

The proof of (b) is done by induction on the depth m^X of a current box X (s.t. $x_t \in X \in \mathcal{X}$). When $m^X = 0$: at the start of the game, the bounds hold by the construction of the initial partition $\mathcal{X}$ of $\mathbb{X}$: $1/\text{diam}(X) = \lceil(4\eta+6)L\rceil \in [(2\eta+3)L, \lceil(4\eta+6)L\rceil)$ and $w^X - u^X = 1$. Hence, in all exploration rounds within this box with $m^X = 0$, we will have $(2\eta+3)L\text{diam}(X) \le w^X - u^X$, because violation of this condition will result in a bisection of X. Assume (a) and (b) hold for boxes of depth $m-1 \in \mathbb{Z}_+$. Let t be the first exploration round in a box X of depth $m > 0$. This box is a result of the bisection of a box X_1 of depth $m-1$ in a exploration round t_1, what implies that $w^X - u^X < (2\eta+3)L\text{diam}(X_1) = 2(2\eta+3)L\text{diam}(X)$. The values of the function $u(\cdot)$ and $w(\cdot)$ in the round t_1 are denoted by u_1 and w_1. So, by induction for (a), we know that $(w_1 - u_1) - (w^X - u^X) \le (w_1 - u_1) - 2(\eta+1)L\text{diam}(X_1)$. Hence,

$$w^X - u^X \ge 2(\eta+1)L\text{diam}(X_1) = 4(\eta+1)L\text{diam}(X) \ge (2\eta+3)L\text{diam}(X).$$

Again, in all exploration rounds within this box with $m^X = m$, we will have $(2\eta+3)L\text{diam}(X) \le w^X - u^X$, because violation of this condition will result in a bisection of X. ∎

A.4 Proof of Theorem 1

We add the following remark to the proof of Theorem 1 (in the main text).

In the proof of Theorem 1, we used the following inequality

$$T \ge \sum_{m=1}^{M-1} g(m-1)N_m \tag{A.8}$$

to upper bound 2^M via a function of T. However, this inequality from Eq. (A.8) is useful only for $M > 1$. Indeed, note that, when $M = 1$, this inequality becomes $T \ge 0$ and cannot be used to upper bound M.

So, the case $M = 1$ should be considered separately. In fact, this case is trivial. Namely, we can show that, if $M = 1$, then the upper bound on the strategic regret holds even for $T = 1$. From the proof (in the main text) we know that $R \le \sum_{m=0}^{M} C(L,r,\eta)2^{md}N_0$; hence, for $M = 1$, we have:

$$\begin{aligned} R &\le \sum_{m=0}^{1} C(L,r,\eta)2^{md}N_0 = C(L,r,\eta)N_0(1+2^d) \le C(L,r,\eta)N_0 2^{d+1} \le \\ &\le C(L,r,\eta)N_0 2^{d+1}(\frac{1}{N_0}+1)^{d/(d+1)} = 2^{d+1}C(L,r,\eta)N_0^{1/(d+1)}(1+N_0)^{d/(d+1)}, \end{aligned} \tag{A.9}$$

where the right-hand side of the latter identity is exactly Eq.(6) with $T = 1$.

B The pseudo-code of the PELS algorithm

Algorithm B.1 Pseudo-code of Penalized Exploiting Lipschitz Search (PELS).

Input: $L > 0$, $\eta \in \mathbb{R}_+$, $r \in \mathbb{N}$, and $g : \mathbb{Z}_+ \to \mathbb{Z}_+$
Initialize: Split[] := partition of $[0,1]^d$ into $\lceil (6+4\eta)L \rceil^d$ cubes with $\operatorname{diam}(\cdot) = 1/\lceil (4\eta+6)L \rceil$; Boxes[] := $\varnothing$;
for all Cube $\in$ Split[ ] **do**
    Box := `NewStructure`($X :=$ Cube, $u := 0, w := 1, P := 0, E := 0, m := 0, l := 1/\lceil (4\eta+6)L \rceil$);
    Boxes[ ] := Boxes[ ] $\cup$ Box; // Add each cube with its associated data
end for
while the buyer plays **do**
    The seller receives a feature vector $x \in [0,1]^d$ from the nature, the buyer observes this vector x as well;
    j := find j such that $x \in$ Boxes[j].X;
    if Boxes[j].$P > 0$ **then**
        Offer the price $p := 1$ to the buyer; // Penalization
        Boxes[j].P := Boxes[j].$P - 1$;
        if this price is accepted **then** offer $p := 1$ for all remaining rounds;
    else
        if Boxes[j].$E > 0$ **then**
            Offer the price $p :=$ Boxes[j].u to the buyer; // Exploitation
            Boxes[j].E := Boxes[j].$E - 1$;
            if Boxes[j].$E == 0$ **then**
                Split[ ] := bisect each side of the cube Boxes[j].X to get 2^d cubes;
                for all Cube $\in$ Split[ ] **do**
                    Box := `CopyStructure`(Boxes[j]); // Copy the associated data of the parent cube
                    Box.X := Cube; // Replace some associated for the new cube
                    Box.m := Box.$m + 1$;
                    Box.l := Box.$l/2$;
                    Boxes[ ] := Boxes[ ] $\cup$ Box; // Add each cube with its associated data
                end for
                Remove Boxes[j] from Boxes[ ]; // Remove j-th cube with its associated data
            end if
        else
            Offer the price $p :=$ Boxes[j].$u + \eta L$Boxes[j].l to the buyer; // Exploration
            if the buyer accepts the price **then**
                Boxes[j].u := $p - (\eta - 1)L$Boxes[j].l;
            else
                Boxes[j].w := $p + (\eta + 2)L$Boxes[j].l;
                Boxes[j].P := r
            end if
            if Boxes[j].w $-$ Boxes[j].u $< (2\eta + 3)L$Boxes[j].l **then**
                Boxes[j].E := g(Boxes[j].m);
            end if
        end if
    end if
end while

C Auxiliary statements

C.1 Statement on linear programming

Statement C.1. *Consider the linear program:*

$$\begin{aligned} \textit{maximize} \quad R(z_1,..,z_{\overline{M}}) &= \sum_{m=1}^{\overline{M}} z_m \quad \textit{s.t.} \\ \sum_{m=1}^{\overline{M}} \beta_m z_m &\leq C \quad \textit{and} \\ 0 \leq z_m &\leq c_m \,\forall m, \end{aligned} \tag{C.1}$$

where $c_m > 0\, \forall m$*,* $0 < \beta_1 \leq ... \leq \beta_{\overline{M}}$*, and* $C \leq \sum_{m=1}^{\overline{M}} \beta_m c_m$*. Then the maximum of* R *is achieved at*

$$(z_1, \ldots, z_{\overline{M}}) = \left(c_1, \ldots, c_{M-1}, \beta_M^{-1}\left(C - \sum_{m=1}^{M-1} \beta_m c_m\right), 0, \ldots, 0\right), \tag{C.2}$$

where M *is such that the following inequality hold:* $\sum_{m=1}^{M-1} \beta_m c_m < C \leq \sum_{m=1}^{M} \beta_m c_m$*.*

Proof. The proof trivially follows from the theory of linear programs. In particular, note that the structure of the solution in Eq. (C.2) is as follows:

- all z_i have the maximal available value c_i for all consecutive $i = 1, ..., M-1$;
- the next z_M has value s.t. it is maximal possible to fit the first condition in Eq. (C.1);
- all remaining z_i are 0 the maximal available value c_i for all consecutive $i = M+1, ..., \overline{M}$.

This is because of the following: if there is two consecutive z_i and z_{i+1} s.t. $z_i < c_i$ and $z_{i+1} > 0$, then we can without change of R consider new vector $\{z_j^{\text{new}}\}_j$: $z_i^{\text{new}} := z_i + \epsilon$ and $z_{i+1}^{\text{new}} := z_{i+1} - \epsilon$, where $\epsilon > 0$ is s.t. $z_i + \epsilon \leq c_i$ and $z_{i+1} - \epsilon \geq 0$. Then $R^{\text{new}} = R$, while the conditions are satisfied due to the properties ϵ and the monotonicity of $\{\beta_j\}_j$. □

C.2 Remarks

In the definition of PELS, we used $\{[0, l], (l, 2l], \ldots, (1-l, 1]\}^d$ to denote all boxes that are obtained from the split of each side of the cube $[0, 1]^d$ into $\lceil(4\eta + 6)L\rceil$ equal parts. This is done for sake of short notations. Formally, the set of all $\lceil(4\eta + 6)L\rceil^d$ boxes should be written as follows:

$$\mathcal{X} = \left\{ I_1 \times I_2 \times \ldots \times I_d \,\middle|\, (I_1, I_2, \ldots, I_d) \in \Big\{[0, l], (l, 2l], \ldots, (1-l, 1]\Big\}^d \right\}.$$

D The game's workflow and structure of the knowledge

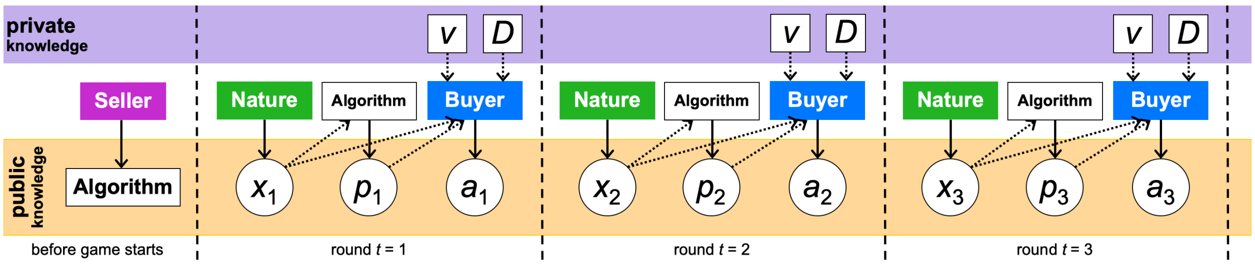

Figure D.1: The game's workflow (an algorithm and the buyer exploit public knowledge available in the previous rounds).

E Examples for the notion of a start of r-length penalization

Note that the definition of a start of r-length penalization *does not mean* that all the next $r-1$ future rounds are penalizations: instead, penalization rounds will be those whose feature vectors belong to the penalization domain $\tilde{X}$ (i.e., a round s is *not* a penalization if $x_s \notin \tilde{X}$).

Also note that a start of r-length penalization in a round t *does not mean* that rounds after $t+r$ will not be penalizations (connected to this start): instead, the counter of penalization rounds is increased only when context of a future round belongs to the penalization domain $\tilde{X}$. For instance, if the contexts $x_{t+1}, \ldots, x_{t+r-1} \notin \tilde{X}$ in future rounds $t+1, \ldots, t+r-1$, but we have $x_{t+r} \in \tilde{X}$ in the round $t+r$, then this round $t+r$ will be the 1-st penalization round associated with the considered start (i.e., $r-2$ penalizations will remain after the round $t+r$). See also the work flow of penalization in the pseudo-code of PELS in Alg. B.1.

Let consider an example:

$$d = 1, \quad \tilde{X} = [0.5, 1], \quad r = 3$$

Let $x_1 = 0.75$ be a start of 3-length penalization with domain $\tilde{X} = [0.5, 1]$ and the buyer rejects the price in $t = 1$.

Then, the penalization rounds (for this penalization start) will be those whose sequence of context $x_{1:s}$ is , e.g.,

- $(0.75, 0.6)$, because $0.6 \in \tilde{X}$;
- $(0.75, 0.2, 0.9)$, because $0.9 \in \tilde{X}$;
- $(0.75, 0.2, 0.4, 0.7)$, because $0.7 \in \tilde{X}$ and contexts of rounds $2, 3, 4$ do not belong to $\tilde{X}$;
- $(0.75, 0.2, 0.4, 0.7, 0.8)$, because $0.8 \in \tilde{X}$ and only one context ($x_4 = 0.7$) is in $\tilde{X}$ among contexts of rounds $2, 3, 4$;

but the following sequences of context $x_{1:s}$ terminate with **NON**-penalization rounds (for the considered penalization start):

- $(0.75, 0.2)$, because $0.2 \notin \tilde{X}$;
- $(0.75, 0.2, 0.4)$, because $0.4 \notin \tilde{X}$;

- $(0.75, 0.2, 0.9, 0.3)$, because $0.3 \notin \tilde{X}$;
- $(0.75, 0.2, 0.4, 0.7, 0.8, 0.9)$, because, in previous rounds $(2, 3, 4, 5)$, already 2 contexts (0.7 and 0.8) belonged to $\tilde{X}$ and $r - 1 = 2$.

References

[1] K. Amin, A. Rostamizadeh, and U. Syed. Learning prices for repeated auctions with strategic buyers. In *NIPS'2013*, pages 1169–1177, 2013.

[2] A. Drutsa. Horizon-independent optimal pricing in repeated auctions with truthful and strategic buyers. In *WWW'2017*, pages 33–42, 2017.

[3] A. Drutsa. On consistency of optimal pricing algorithms in repeated posted-price auctions with strategic buyer. *CoRR*, abs/1707.05101, 2017.

[4] A. Drutsa. Weakly consistent optimal pricing algorithms in repeated posted-price auctions with strategic buyer. In *ICML'2018*, pages 1318–1327, 2018.

[5] A. Drutsa. Reserve pricing in repeated second-price auctions with strategic bidders. In *ICML'2020*, 2020.

[6] J. Mao, R. Leme, and J. Schneider. Contextual pricing for lipschitz buyers. In *Advances in Neural Information Processing Systems*, pages 5648–5656, 2018.

[7] M. Mohri and A. Munoz. Optimal regret minimization in posted-price auctions with strategic buyers. In *NIPS'2014*, pages 1871–1879, 2014.

[8] A. Vanunts and A. Drutsa. Optimal pricing in repeated posted-price auctions with different patience of the seller and the buyer. In *Advances in Neural Information Processing Systems*, pages 939–951, 2019.

[9] A. Zhiyanov and A. Drutsa. Bisection-based pricing for repeated contextual auctions against strategic buyer. In *ICML'2020*, 2020.