A Omitted Proofs

We will need the following helper Lemma in the proofs of consistency and robustness.

Lemma 2. *Recall that* $e_B \coloneqq \left(1+\frac{1}{B}\right)^B$ *and* $\alpha_B \coloneqq B\left(e_B^{\alpha/B}-1\right)$. *We have*

1. $e_B \leq e_{B_a}$ *and*
2. $\alpha_B \geq \alpha_{B_a}$.

Proof. It is well known that e_B converges to e from below for $B \to \infty$. Furthermore, we can show that α_B is decreasing in B for all $\alpha \geq 1$ by taking the derivative

$$\frac{\partial}{\partial B}\alpha_B = \left(1+\frac{1}{B}\right)^{\alpha} - 1 - \frac{\alpha}{B}\left(1+\frac{1}{B}\right)^{\alpha-1} = \left(1+\frac{1}{B}\right)^{\alpha-1}\left(1+\frac{1}{B}(1-\alpha)\right) - 1$$
$$\leq \left(1+\frac{1}{B}\right)^{\alpha-1}\left(1+\frac{1}{B}\right)^{1-\alpha} - 1 = 0$$

where the bound follows from Bernoulli's inequality, which states that $1+rx \leq (1+x)^r$ for $x \geq -1$ and $r \in \mathbb{R} \setminus (0,1)$. □

A.1 Proof of Theorem 1 (Robustness)

We write P and D to denote the objective value of the primal and dual solutions, i.e. $P = \sum_a \sum_{t \in \mathbf{S}_a} w_{at}$ and $D = \sum_a B_a\beta_a + \sum_t z_t$ where z_t is specified in the following proof to ensure feasibility. We can show that after the allocation of each impression t,

$$\Delta P \geq \frac{e_B^{\alpha}-1}{B e_B^{\alpha}\left(e_B^{\alpha/B}-1\right)}\Delta D$$

where ΔP and ΔD are the increase in the primal and dual solution values, respectively. Since we create feasible primal and dual solutions, this is sufficient to bound the robustness due to weak duality. There is one main difference to Feldman et al. (2009a): In their algorithm, setting the dual variable z_t to $w_{a_{(\mathrm{EXP})}t} - \beta_a$ ensures dual feasibility as $a_{(\mathrm{EXP})}$ is the advertiser with maximum discounted gain. However, in order not to violate dual feasibility when following the prediction, we need to increase the dual variables z_t by a factor of α_B. Note that for $\alpha = 1$, this recovers the competitiveness obtained by Feldman et al. (2009a).

Proof. Consider an iteration where we assign an impression t to advertiser a and let $w_1 \leq w_2 \leq \cdots \leq w_{B_a}$ be the values of impressions currently allocated to a in non-decreasing order. Let w_0 be the least valuable of the impressions allocated to a at the end of iteration $t-1$, i.e. the impression that is removed to make space for t. Assume that after allocating impression t to a, it becomes the k-th least valuable impression allocated to a with value $w_{at} = w_k$. Thus, using that $w_i \geq w_{i-1}$, we can bound

$$\begin{aligned}
\beta_a^{(t-1)} &= \frac{e_{B_a}^{\alpha/B_a}-1}{e_{B_a}^{\alpha}-1}\left(\sum_{i=0}^{k-1} w_i e_{B_a}^{\alpha(B_a-i-1)/B_a} + \sum_{i=k+1}^{B_a} w_i e_{B_a}^{\alpha(B_a-i)/B_a}\right)\\
&= \frac{e_{B_a}^{\alpha/B_a}-1}{e_{B_a}^{\alpha}-1}\left(\sum_{i=0}^{B_a-1} w_i e_{B_a}^{\alpha(B_a-i-1)/B_a} + \sum_{i=k+1}^{B_a} (w_i - w_{i-1})\, e_{B_a}^{\alpha(B_a-i)/B_a}\right)\\
&\geq \frac{e_{B_a}^{\alpha/B_a}-1}{e_{B_a}^{\alpha}-1}\left(\sum_{i=0}^{B_a-1} w_i e_{B_a}^{\alpha(B_a-i-1)/B_a}\right) \eqqcolon \hat{\beta}_a^{(t-1)}
\end{aligned}$$

which is tight when impression t becomes the most valuable impression assigned to a. We can now write $\beta_a^{(t)}$ as a function of the bound $\hat{\beta}_a^{(t-1)}$:

$$\beta_a^{(t)} = \frac{e_{B_a}^{\alpha/B_a}-1}{e_{B_a}^{\alpha}-1}\sum_{i=1}^{B_a} w_i e_{B_a}^{\alpha(B_a-i)/B_a}$$

$$= \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} \left(\sum_{i=0}^{B_a - 1} w_i e_{B_a}^{\alpha(B_a - i)/B_a} + w_{B_a} - w_0 e_{B_a}^{\alpha} \right)$$
$$= \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} e_{B_a}^{\alpha/B_a} \sum_{i=1}^{B_a} w_{i-1} e_{B_a}^{\alpha(B_a - i)/B_a} + \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} \left(w_{B_a} - w_0 e_{B_a}^{\alpha} \right)$$
$$= e_{B_a}^{\alpha/B_a} \hat{\beta}_a^{(t-1)} + \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} \left(w_{B_a} - w_0 e_{B_a}^{\alpha} \right).$$

We set $z_t := \alpha_B \left(w_{B_a} - \beta_a^{(t-1)} \right)$ which is feasible as the discounted value $w_{at} - \beta_a^{(t-1)}$ of the chosen advertiser a may only be α_B-times less the maximum discounted value $w_{a_{(\mathrm{EXP})}t} - \beta_{a_{(\mathrm{EXP})}}^{(t-1)}$ due to the advantage of the predicted advertiser. This yields a dual increase of

$$\Delta D = B_a \left(\beta_a^{(t)} - \beta_a^{(t-1)} \right) + z_t$$
$$= B_a \left(\beta_a^{(t)} - \beta_a^{(t-1)} \right) + \alpha_B \left(w_{B_a} - \beta_a^{(t-1)} \right)$$
$$\leq B_a \left(\beta_a^{(t)} - \hat{\beta}_a^{(t-1)} \right) + \alpha_B \left(w_{B_a} - \hat{\beta}_a^{(t-1)} \right)$$
$$= B_a \left(\left(e_{B_a}^{\alpha/B_a} - 1 \right) \hat{\beta}_a^{(t-1)} + \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} \left(w_{B_a} - w_0 e_{B_a}^{\alpha} \right) \right) + \alpha_B \left(w_{B_a} - \hat{\beta}_a^{(t-1)} \right)$$
$$= \alpha_{B_a} \hat{\beta}_a^{(t-1)} + \frac{\alpha_{B_a}}{e_{B_a}^{\alpha} - 1} \left(w_{B_a} - w_0 e_{B_a}^{\alpha} \right) + \alpha_B \left(w_{B_a} - \hat{\beta}_a^{(t-1)} \right)$$
$$= \alpha_{B_a} \frac{e_{B_a}^{\alpha}}{e_{B_a}^{\alpha} - 1} \underbrace{\left(\hat{\beta}_a^{(t-1)} - w_0 \right)}_{\geq 0} + \frac{\alpha_{B_a}}{e_{B_a}^{\alpha} - 1} \underbrace{\left(w_{B_a} - \hat{\beta}_a^{(t-1)} \right)}_{\geq 0} + \alpha_B \left(w_{B_a} - \hat{\beta}_a^{(t-1)} \right)$$
$$\leq \alpha_B \frac{e_B^{\alpha}}{e_B^{\alpha} - 1} \left(\hat{\beta}_a^{(t-1)} - w_0 \right) + \frac{\alpha_B}{e_B^{\alpha} - 1} \left(w_{B_a} - \hat{\beta}_a^{(t-1)} \right) + \alpha_B \left(w_{B_a} - \hat{\beta}_a^{(t-1)} \right)$$
$$= \alpha_B \frac{e_B^{\alpha}}{e_B^{\alpha} - 1} \left(\hat{\beta}_a^{(t-1)} - w_0 \right) + \alpha_B \frac{e_B^{\alpha}}{e_B^{\alpha} - 1} \left(w_{B_a} - \hat{\beta}_a^{(t-1)} \right)$$
$$= \alpha_B \frac{e_B^{\alpha}}{e_B^{\alpha} - 1} \left(w_{B_a} - w_0 \right) = B \frac{e_B^{\alpha/B} - 1}{e_B^{\alpha} - 1} e_B^{\alpha} \left(w_{B_a} - w_0 \right)$$

where the second inequality is due to $\alpha_B \geq \alpha_{B_a}$ and $e_B \leq e_{B_a}$, as shown in Lemma 2. $\square$

A.2 Proof of Theorem 1 (Consistency)

In the following, we upper bound PRD using the comparison in Line 7 of Algorithm 1.

Lemma 3. *We have*

$$\mathrm{PRD} \leq \sum_a \left((B_a - \ell_a) \beta_a^{(T)} + \frac{1}{\alpha_B} \sum_{t \in \mathbf{X}_a \setminus \mathbf{P}_a} \left(w_{at} - \beta_a^{(t-1)} \right) + \sum_{t \in \mathbf{P}_a \cap \mathbf{X}_a} w_{at} \right)$$

Proof. We first split impressions t into two categories: Either the algorithm followed the prediction and assigned t to $a^{(t)} = a_{(\mathrm{PRD})}^{(t)}$, or the algorithm ignored the prediction and assigned t to $a^{(t)} = a_{(\mathrm{EXP})}^{(t)} \neq a_{(\mathrm{PRD})}^{(t)}$. In the latter case, due to the selection rule in Line 7 of Algorithm 1,

$$\alpha_B \left(w_{a_{(\mathrm{PRD})}^{(t)} t} - \beta_{a_{(\mathrm{PRD})}^{(t)}}^{(t-1)} \right) \leq w_{a_{(\mathrm{EXP})}^{(t)} t} - \beta_{a_{(\mathrm{EXP})}^{(t)}}^{(t-1)}.$$

In symbols,

$$\mathrm{PRD} = \sum_a \left(\sum_{t \in \mathbf{P}_a \setminus \mathbf{X}_a} w_{at} + \sum_{t \in \mathbf{P}_a \cap \mathbf{X}_a} w_{at} \right)$$

$$
\begin{aligned}
&\le \sum_a \left(\sum_{t\in \mathbf{P}_a \setminus \mathbf{X}_a} \left(\beta_a^{(t-1)} + \frac{1}{\alpha_B} \left(w_{a^{(t)}_{(\mathrm{EXP})},t} - \beta^{(t-1)}_{a^{(t)}_{(\mathrm{EXP})}} \right) \right) + \sum_{t\in \mathbf{P}_a \cap \mathbf{X}_a} w_{at} \right) \\
&= \sum_a \left(\underbrace{\sum_{t\in \mathbf{P}_a \setminus \mathbf{X}_a} \beta_a^{(t-1)}}_{(\dagger)} + \frac{1}{\alpha_B} \sum_{t\in \mathbf{X}_a \setminus \mathbf{P}_a} \left(w_{at} - \beta_a^{(t-1)} \right) + \sum_{t\in \mathbf{P}_a \cap \mathbf{X}_a} w_{at} \right)
\end{aligned}
$$

where the last equality holds because $\{\mathbf{P}_a\}_a$ and $\{\mathbf{X}_a\}_a$ are both partitioning the set of all impressions due to the introduction of the dummy advertiser. For $(\dagger)$, we use that β_a can only increase in each round and bound

$$\sum_{t\in \mathbf{P}_a \setminus \mathbf{X}_a} \beta_a^{(t-1)} \le (B_a - \ell_a)\, \beta_a^{(T)}.$$

$\square$

For the remainder of this section, we consider a fixed advertiser a. Let us denote with t_i the i-th impression allocated to a. Let

$$(\star) = \frac{1}{\alpha_B} \sum_{t\in \mathbf{X}_a \setminus \mathbf{P}_a} \left(w_{at} - \beta_a^{(t-1)} \right) + \sum_{t\in \mathbf{P}_a \cap \mathbf{X}_a} w_{at}$$

as part of the the bound on PRD in Lemma 3.

In order to understand this bound, we make some useful observations in the following lemma to simplify the analysis. The key idea is that we may assume that impressions in $\mathbf{X}_a$ are ordered to be non-decreasing. In particular, we need to argue that the sum $\sum_{t\in \mathbf{X}_a \setminus \mathbf{P}_a} \beta_a^{(t-1)}$ can only decrease (as this term is negated in $(\star)$) when impressions in $\mathbf{X}_a$ are ordered to be non-decreasing: Intuitively, each $\beta_a^{(t-1)}$ depends only on the B_a most valuable impressions assigned before impression t, no matter the order in which $\mathbf{X}_a^{(t-1)}$ arrived. We can thus minimize each $\beta_a^{(t-1)}$ if the impressions allocated prior to t are the impressions of smallest value. To simultaneously minimize each $\beta_a^{(t)}$ in the sum, we order the impressions in $\mathbf{X}_a$ to have non-decreasing value. We prove this simplification formally in the following lemma.

Lemma 4. *Without loss of generality, we may assume that $\mathbf{P}_a \cap \mathbf{X}_a$ are the most valuable impressions in $\mathbf{X}_a$ and that impressions in $\mathbf{X}_a$ arrive such that their values are non-decreasing.*

Proof. We may assume that the impressions in $\mathbf{P}_a \cap \mathbf{X}_a$ are the most valuable impressions in $\mathbf{X}_a$: this can only increase the value of $\mathbf{P}_a$ but leaves $\mathbf{S}_a$ unaffected, as $\mathbf{S}_a$ are by design the B_a most valuable impressions in $\mathbf{X}_a$. All impressions in $(\star)$ are from $\mathbf{X}_a$, so reordering impressions only affects $(\star)$. Specifically, we can show that the sum $\sum_{t\in \mathbf{X}_a \setminus \mathbf{P}_a} \beta_a^{(t-1)}$ in $(\star)$ is minimized if the values in $\mathbf{X}_a$ are ordered to be non-decreasing. Assume to the contrary that the i-th impression added to a is the last that is in order. That is $w_{at_1} \le w_{at_2} \le \cdots \le w_{at_i}$ and there exists a $j \le i$ such that $w_{at_{j-1}} \le w_{at_{i+1}} < w_{at_j}$. Moving t_{i+1} ahead to its ranked position within the first i impressions allocated to a changes the ordering as follows (the first and second row show the impression values before and after changing the position of t_{i+1}, respectively):

$$
\begin{array}{ccccccccc}
w_{at_1} \le \cdots \le w_{at_{j-1}} & \le & w_{at_j} & \le & w_{at_{j+1}} & \le \cdots \le & w_{at_i} \\
w_{at_1} \le \cdots \le w_{at_{j-1}} & \le & w_{at_{i+1}} & < & w_{at_j} & \le \cdots \le & w_{at_{i-1}}
\end{array}
$$

Note that each position decreases in value, even strictly at the j-th position. As such, the exponential average $\beta_a^{(t-1)}$ decreases as well for $t < t_{i+1}$; it remains constant for $t \ge t_{i+1}$ as it only depends on the B_a most valuable impressions assigned up to t which remain the same. We can thus simultaneously minimize $\beta_a^{(t)}$ for each t by putting $\mathbf{X}_a$ in non-decreasing order. This reordering does not affect $\beta_a^{(T)}$ or the other terms in $(\star)$, so we may indeed assume that values are non-decreasing. $\square$

In light of Lemma 4, we can write $(\star)$ as follows.

Lemma 5. *We have*

$$(\star) = \frac{1}{\alpha_B} \sum_{i=1}^{I_a-\ell_a} \left(w_{at_i} - \beta_a^{(t_{i-1})} \right) + \sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i}.$$

Proof. Impression values are non-decreasing due to Lemma 4, so w_{at_i} is the i-th least valuable impression in $\mathbf{X}_a$. The impressions $\{I_a - \ell_a + 1, \ldots, I_a\} = \mathbf{X}_a \cap \mathbf{P}_a$ are thus the most valuable. We can now write $(\star)$ as

$$\frac{1}{\alpha_B} \sum_{t \in \mathbf{X}_a \setminus \mathbf{P}_a} \left(w_{at} - \beta_a^{(t-1)} \right) + \sum_{t \in \mathbf{P}_a \cap \mathbf{X}_a} w_{at} = \frac{1}{\alpha_B} \sum_{i=1}^{I_a-\ell_a} \left(w_{at_i} - \beta_a^{(t_{i-1})} \right) + \sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i}$$

where $\beta_a^{(t_i-1)} = \beta_a^{(t_{i-1})}$ as there was no change to the dual variable of advertiser a since no impression in $\{t_{i-1}+1, \ldots, t_i - 1\}$ was allocated to a. $\square$

Combining Lemmas 3 and 5, we obtain:

Lemma 6. $\mathrm{PRD} \leq \sum_a \mathrm{PRD}_a$ *where*

$$\mathrm{PRD}_a \coloneqq \frac{1}{\alpha_B} \sum_{i=1}^{I_a-\ell_a} \left(w_{at_i} - \beta_a^{(t_{i-1})} \right) + \sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i} + (I_a - \ell_a)\, \beta_a^{(T)}.$$

In the following, we use the non-decreasing ordering of impressions in $\mathbf{X}_a$ to compute $\beta_a^{(t_{i-1})}$ and bound PRD_a with a linear combination of values w_{at_i}. Consider the j-th impression t_j allocated to a. Since we assume that impression values are non-decreasing, we know that t_j becomes the most valuable impression right after it is allocated. After the allocation of the $(j+1)$-th impression to a, it becomes the second most valuable impression, and so forth, until it is disposed after the allocation of the $(j+B_a)$-th impression. The value w_{at_j} therefore appears alongside each coefficient in the convex combination that defines $\beta_a^{(t_{i-1})}$ for $i \in \{j+1, \ldots, j+B_a\}$. Expanding each $\beta_a^{(t_{i-1})}$ in the sum $\sum_{i=1}^{I_a-\ell_a} \beta_a^{(t_{i-1})}$ in PRD_a, we thus observe that the coefficients of values w_{at_j} for $j \leq I_a - \ell_a - B_a$ sum up to 1. We use this fact to cancel out most of the values in $\sum_{i=1}^{I_a-\ell_a} w_{at_i}$. What remains are only the values w_{at_i} for $i \in \{I_a - \ell_a - B_a + 1, \ldots, I_a - \ell_a\}$. For $i \in \{I_a - \ell_a - B_a + 1, \ldots, I_a - B_a\}$, we bound w_{at_i} by $w_{at_{I_a-B_a}}$ which is really the best we can hope for. Formally, we show:

Lemma 7. *We have*

$$\mathrm{PRD}_a \leq \sum_{i=I_a-B_a+1}^{I_a-\ell_a} \phi_i w_{at_i} + \sum_{i=I_a-\ell_a+1}^{I_a} \psi_i w_{at_i} + w_{at_{I_a-B_a}} \Omega_a$$

with coefficients

$$\begin{aligned}
\phi_i &\coloneqq (B_a - \ell_a)\, \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} e_{B_a}^{\alpha(I_a-i)/B_a} + \frac{1}{\alpha_B} \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a}}{e_{B_a}^{\alpha} - 1} \\
\psi_i &\coloneqq 1 + (B_a - \ell_a)\, \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} e_{B_a}^{\alpha(I_a-i)/B_a} \\
\Omega_a &\coloneqq \frac{1}{\alpha_B} \frac{1}{e_{B_a}^{\alpha} - 1} \left(\ell_a e_{B_a}^{\alpha} - \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(B_a-\ell_a)/B_a}}{e_{B_a}^{\alpha/B_a} - 1} \right).
\end{aligned}$$

Proof. We start by rewriting the terms in PRD_a individually. Since we assume that the values are non-decreasing, we can express $\beta_a^{(t_{i-1})}$ as the exponential average of values $w_{at_{i-B_a}}, w_{at_{i-B_a+1}}, \ldots, w_{at_{i-1}}$ of the last B_a impressions (for simplicity, we set $w_{at_j} = 0$ for $j \leq 0$). Summing over multiple iterations, we thus obtain for the sum over the dual variables that

$$\sum_{i=1}^{I_a-\ell_a} \beta_a^{(t_{i-1})} = \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} \sum_{i=1}^{I_a-\ell_a} \sum_{j=i-B_a}^{i-1} w_{at_j} e_{B_a}^{\alpha(i-j-1)/B_a}$$

$$
\begin{aligned}
&= \frac{e_{B_a}^{\alpha/B_a}-1}{e_{B_a}^{\alpha}-1} \sum_{j=1}^{I_a-\ell_a} w_{at_j} \sum_{i=j+1}^{\min\{j+B_a, I_a-\ell_a\}} e_{B_a}^{\alpha(i-j-1)/B_a} \\
&= \frac{e_{B_a}^{\alpha/B_a}-1}{e_{B_a}^{\alpha}-1} \sum_{j=1}^{I_a-\ell_a} w_{at_j} \sum_{i=1}^{\min\{B_a, I_a-\ell_a-j\}} e_{B_a}^{\alpha(i-1)/B_a} \\
&= \frac{e_{B_a}^{\alpha/B_a}-1}{e_{B_a}^{\alpha}-1} \sum_{j=1}^{I_a-B_a-\ell_a} w_{at_j} \sum_{i=1}^{I_a} e_{B_a}^{\alpha(i-1)/B_a} \\
&\quad + \frac{e_{B_a}^{\alpha/B_a}-1}{e_{B_a}^{\alpha}-1} \sum_{j=I_a-B_a-\ell_a+1}^{I_a-\ell_a} w_a \sum_{i=1}^{I_a-\ell_a-j} e_{B_a}^{\alpha(i-1)/B_a} \\
&= \sum_{i=1}^{I_a-B_a-\ell_a} w_{at_i} + \frac{1}{e_{B_a}^{\alpha}-1} \sum_{i=I_a-B_a-\ell_a+1}^{I_a-\ell_a} w_{at_i} \left(e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a} - 1 \right).
\end{aligned}
$$

where for the last equality, we use that the two inner sums are geometric. We can use this expression to cancel out most of the terms of the first sum in PRD_a:

$$
\begin{aligned}
&\sum_{i=1}^{I_a-\ell_a} \left(w_{at_i} - \beta_a^{(t_{i-1})} \right) \\
&= \sum_{i=1}^{I_a-\ell_a} w_{at_i} - \sum_{i=1}^{I_a-B_a-\ell_a} w_{at_i} - \frac{1}{\alpha_B \left(e_{B_a}^{\alpha}-1 \right)} \sum_{i=I_a-B_a-\ell_a+1}^{I_a-\ell_a} w_{at_i} \left(e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a} - 1 \right) \\
&= \sum_{i=I_a-B_a-\ell_a+1}^{I_a-\ell_a} w_{at_i} \left(1 - \frac{e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a}-1}{e_{B_a}^{\alpha}-1} \right) \\
&= \sum_{i=I_a-B_a-\ell_a+1}^{I_a-\ell_a} w_{at_i} \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a}}{e_{B_a}^{\alpha}-1} \\
&= \sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_i} \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a}}{e_{B_a}^{\alpha}-1} + \sum_{i=I_a-B_a-\ell_a+1}^{I_a-B_a} w_{at_i} \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a}}{e_{B_a}^{\alpha}-1}. \qquad (3)
\end{aligned}
$$

We use that $w_{at_i} \leq w_{at_{I_a-B_a}}$ for all $i \leq I_a - B_a$ to upper bound the second sum, divided by α_B, in (3) to

$$
\begin{aligned}
&\frac{1}{\alpha_B} \sum_{i=I_a-B_a-\ell_a+1}^{I_a-B_a} w_{at_i} \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a}}{e_{B_a}^{\alpha}-1} \\
&\leq w_{at_{I_a-B_a}} \frac{1}{\alpha_B} \sum_{i=I_a-B_a-\ell_a+1}^{I_a-B_a} \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a}}{e_{B_a}^{\alpha}-1} \qquad (4) \\
&= w_{at_{I_a-B_a}} \frac{1}{\alpha_B} \frac{1}{e_{B_a}^{\alpha}-1} \left(\ell_a e_{B_a}^{\alpha} - \sum_{i=B_a-\ell_a}^{B_a-1} e_{B_a}^{\alpha i/B_a} \right) \\
&= w_{at_{I_a-B_a}} \underbrace{\frac{1}{\alpha_B} \frac{1}{e_{B_a}^{\alpha}-1} \left(\ell_a e_{B_a}^{\alpha} - \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(B_a-\ell_a)/B_a}}{e_{B_a}^{\alpha/B_a}-1} \right)}_{=\Omega_a}. \qquad (5)
\end{aligned}
$$

Furthermore, by definition of $\beta_a^{(T)} = \beta_a^{(t_{I_a})}$,

$$
(B_a - \ell_a)\, \beta_a^{(T)} = (B_a - \ell_a) \frac{e_{B_a}^{\alpha/B_a}-1}{e_{B_a}^{\alpha}-1} \sum_{i=I_a-B_a+1}^{I_a} w_{at_i} e_{B_a}^{\alpha(I_a-i)/B_a}. \qquad (6)
$$

We combine (3), (5), and (6) and group terms to obtain the desired bound

$$\begin{aligned}
\mathrm{PRD}_a &\leq \sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i} + \frac{1}{\alpha_B} \sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_i} \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a}}{e_{B_a}^{\alpha} - 1} + w_{at_{I_a-B_a}} \Omega_a \\
&\quad + (B_a - \ell_a) \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} \sum_{i=I_a-B_a+1}^{I_a} w_{at_i} e_{B_a}^{\alpha(I_a-i)/B_a} \\
&= \sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_i} \underbrace{\left((B_a - \ell_a) \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} e_{B_a}^{\alpha(I_a-i)/B_a} + \frac{1}{\alpha_B} \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(I_a-\ell_a-i)/B_a}}{e_{B_a}^{\alpha} - 1} \right)}_{=\phi_i} \\
&\quad + \sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i} \underbrace{\left(1 + (B_a - \ell_a) \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} e_{B_a}^{\alpha(I_a-i)/B_a} \right)}_{=\psi_i} + w_{I_a-B_a} \Omega_a
\end{aligned}$$

□

We can express ALG analogously:

Lemma 8. *We have* $\mathrm{ALG} = \sum_a \mathrm{ALG}_a$ *where*

$$\mathrm{ALG}_a \coloneqq \sum_{i=I_a-B_a+1}^{I_a} w_{at_i}.$$

Proof. We have $\mathrm{ALG} = \sum_a \sum_{t \in \mathbf{S}_a} w_{at}$. As we always dispose of the least valuable impression in Algorithm 1, $\mathbf{S}_a$ are the B_a most valuable impressions in $\mathbf{X}_a$. Due to Lemma 4, these are $\mathbf{S}_a = \{I_a - B_a + 1, \ldots, I_a\}$ and hence $\sum_{t \in \mathbf{S}_a} w_{at} = \sum_{i=I_a-B_a+1}^{I_a} w_{at_i} = \mathrm{ALG}_a$. □

We upper bound the ratio $\mathrm{PRD}/\mathrm{ALG}$ by $\max_a \mathrm{PRD}_a/\mathrm{ALG}_a$. To this end, we fix an advertiser a and upper bound the ratio $\mathrm{PRD}_a/\mathrm{ALG}_a$. Recall from Lemmas 7 and 8 that we can express PRD_a and ALG_a as linear combination over impression values. We could obtain a natural upper bound by comparing impression value coefficients. However, in the following lemma, we show how to use the non-decreasing ordering due to Lemma 4 to obtain a tighter bound.

We define

$$\Phi_a \coloneqq \sum_{i=I_a-B_a+1}^{I_a-\ell_a} \phi_i \qquad \text{and} \qquad \Psi_a \coloneqq \sum_{i=I_a-\ell_a+1}^{I_a} \psi_i$$

as the total factor mass on values w_{at_i} for ϕ_i and ψ_i, respectively. Let $\tau_a \coloneqq (\Phi_a + \Psi_a + \Omega_a)/B_a$ be the average factor. Recall that

$$\begin{aligned}
\mathrm{PRD}_a &\leq \sum_{i=I_a-B_a+1}^{I_a-\ell_a} \phi_i w_{at_i} + \sum_{i=I_a-\ell_a+1}^{I_a} \psi_i w_{at_i} + w_{at_{I_a-B_a}} \Omega_a \qquad (7) \\
\mathrm{ALG}_a &= \sum_{i=I_a-B_a+1}^{I_a} w_{at_i}.
\end{aligned}$$

In the following lemma, we use that $w_{at_i} \leq w_{at_j}$ for $i \leq j$ due to Lemma 4, to further upper bound the RHS of 7 by a linear combination of the values, where we move mass from coefficients on w_{at_i} to coefficients on w_{at_j}. Additionally, we move mass from Ω_a to coefficients ϕ_i for $i \in \{I_a - B_a + 1, \ldots, I_a - \ell_a\}$ and from ϕ_i to ψ_j for $j \in \{I_a - \ell_a + 1, \ldots, I_a\}$. In the best case, we are able to redistribute mass equally across all values, in which case the consistency is given as the average factor τ_a. Otherwise, the factors on the largest values dominate, giving us a consistency of Ψ_a/ℓ_a.

Lemma 9. *We have*

$$\frac{\mathrm{PRD}_a}{\mathrm{ALG}_a} \leq \begin{cases} \max\left\{\tau_a, \frac{\Psi_a}{\ell_a}\right\} & \text{if } \ell_a > 0 \\ \tau_a & \text{otherwise} \end{cases}$$

where

$$\tau_a = 1 + \frac{1}{e^{\alpha}_{B_a} - 1} \frac{1}{\alpha_B} \left(e^{\alpha}_{B_a} - \frac{e^{\alpha}_{B_a} - 1}{\alpha_{B_a}} \right)$$

and

$$\frac{\Psi_a}{\ell_a} = 1 + \left(\frac{B_a}{\ell_a} - 1 \right) \frac{e^{\alpha \ell_a / B_a}_{B_a} - 1}{e^{\alpha}_{B_a} - 1}.$$

Proof. We calculate τ_a and Ψ_a/ℓ_a separately in Lemma 10 below. Our main goal is to distribute mass from the factors ϕ_i, ψ_i, and from Ω_a equally to the values $w_{I_a - B_a + 1}, \ldots, w_{I_a}$. We begin by taking a closer look at the factors ϕ_i and ψ_i. First, note that ψ_i is always decreasing in i as

$$\psi_i = 1 + \underbrace{(B_a - \ell_a)}_{\geq 0} \underbrace{\frac{e^{\alpha/B_a}_{B_a} - 1}{e^{\alpha}_{B_a} - 1}}_{\geq 0} e^{\alpha(I_a - i)/B_a}_{B_a}.$$

We can therefore bound the linear combination over values in $\{I_a - \ell_a + 1, \ldots, I_a\}$ using the average value $\bar{w}_\Psi := \frac{1}{\ell_a} \sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i}$ as

$$\sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i} \psi_i \leq \sum_{i=I_a-\ell_a+1}^{I_a} \bar{w}_\Psi \psi_i = \bar{w}_\Psi \Psi_a. \tag{8}$$

However, ϕ_i is not always decreasing which can be seen by rearranging

$$\begin{aligned} \phi_i &= (B_a - \ell_a) \frac{e^{\alpha/B_a}_{B_a} - 1}{e^{\alpha}_{B_a} - 1} e^{\alpha(I_a - i)/B_a} + \frac{1}{\alpha_B} \frac{e^{\alpha}_{B_a} - e^{\alpha(I_a - \ell_a - i)/B_a}_{B_a}}{e^{\alpha}_{B_a} - 1} \\ &= \frac{1}{e^{\alpha}_{B_a} - 1} \left((B_a - \ell_a) \left(e^{\alpha/B_a}_{B_a} - 1 \right) - \frac{1}{\alpha_B} e^{-\alpha \ell_a / B_a}_{B_a} \right) e^{\alpha(I_a - i)/B_a}_{B_a} + \frac{1}{\alpha_B} \frac{e^{\alpha}_{B_a}}{e^{\alpha}_{B_a} - 1}. \end{aligned}$$

We observe that ϕ_i is decreasing if $(B_a - \ell_a) \left(e^{\alpha/B_a}_{B_a} - 1 \right)$ is at least $\frac{1}{\alpha_B} e^{-\alpha \ell_a / B_a}_{B_a}$, and we analyze two cases based on the relationship of both terms.

Let us first assume that $(B_a - \ell_a) \left(e^{\alpha/B_a}_{B_a} - 1 \right) \geq \frac{1}{\alpha_B} e^{-\alpha \ell_a / B_a}_{B_a}$ such that ϕ_i is decreasing in i which helps us to bound the linear combinations in Lemma 7 over $\{I_a - B_a + 1, \ldots, I_a - \ell_a\}$ and $\{I_a - \ell_a + 1, \ldots, T\}$ by the average values $\bar{w}_\Phi := \frac{1}{B_a - \ell_a} \sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_i}$ and $\bar{w}_\Psi$, respectively. We further use that $w_{at_{I_a - B_a}} \leq \bar{w}_\Phi$ to charge mass from Ω_a to Φ_a and obtain due to (8) that

$$\begin{aligned} &\sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_i} \phi_i + \sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i} \psi_i + w_{at_{I_a - B_a}} \Omega_a \\ &\leq \bar{w}_\Phi \Phi_a + \bar{w}_\Psi \Psi_a + w_{at_{I_a - B_a}} \Omega_a \qquad (9) \\ &\leq \bar{w}_\Phi \left(\Phi_a + \Omega_a \right) + \bar{w}_\Psi \Psi_a \\ &= \bar{w}_\Phi \left(B_a - \ell_a \right) \tau_a + \bar{w}_\Phi \left(\Phi_a + \Omega_a - \left(B_a - \ell_a \right) \tau_a \right) + \bar{w}_\Psi \Psi_a \\ &= \sum_{i=I_a-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} + \bar{w}_\Phi \left(\Phi_a + \Omega_a - \left(B_a - \ell_a \right) \tau_a \right) + \bar{w}_\Psi \Psi_a \qquad (10) \end{aligned}$$

On the other hand, if $(B_a - \ell_a) \left(e^{\alpha/B_a}_{B_a} - 1 \right) \leq \frac{1}{\alpha_B} e^{-\alpha \ell_a / B_a}_{B_a}$ we can no longer bound the values over $\{I_a - B_a + 1, \ldots, I_a - \ell_a\}$ by the average value $\bar{w}_\Phi$. However, each factor ϕ_i is less than τ_a which can be seen by rearranging

$$\phi_i = \frac{1}{e^{\alpha}_{B_a} - 1}\left((B_a - \ell_a)\left(e^{\alpha/B_a}_{B_a} - 1\right) - \frac{1}{\alpha_B} e^{-\alpha\ell_a/B_a}_{B_a}\right) e^{\alpha(I_a - i)/B_a}_{B_a} + \frac{1}{\alpha_B}\frac{e^{\alpha}_{B_a}}{e^{\alpha}_{B_a} - 1}$$
$$\leq 1 + \frac{1}{e^{\alpha}_{B_a} - 1}\frac{1}{\alpha_B}\left(e^{\alpha}_{B_a} - \frac{e^{\alpha}_{B_a} - 1}{\alpha_{B_a}}\right) = \tau_a$$

to the equivalent expression

$$\underbrace{\left((B_a - \ell_a)\left(e^{\alpha/B_a}_{B_a} - 1\right) - \frac{1}{\alpha_B} e^{-\alpha\ell_a/B_a}_{B_a}\right)}_{\leq 0} \underbrace{e^{\alpha(I_a - i)/B_a}_{B_a}}_{\geq 0}$$
$$\leq e^{\alpha}_{B_a} - 1 - \frac{1}{\alpha_B}\frac{e^{\alpha}_{B_a} - 1}{\alpha_{B_a}} = \underbrace{\left(1 - \frac{1}{\alpha_B \cdot a_{B_a}}\right)}_{\geq 0} \underbrace{\left(e^{\alpha}_{B_a} - 1\right)}_{\geq 0}$$

which is true since the LHS is ≤ 0 and the RHS ≥ 0. We can thus charge $\tau_a - \phi_i$ of mass from Ω_a to the coefficients ϕ_i for each $i \in \{I_a - B_a + 1, \ldots, I_a - \ell_a\}$ which yields

$$\begin{aligned}
&\sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_i}\phi_i + \sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i}\psi_i + w_{at_{I_a-B_a}}\Omega_a \\
&\leq \sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_i}\phi_i + \bar{w}_\Psi\Psi_a + w_{at_{I_a-B_a}}\Omega_a \\
&= \sum_{i=I_a-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} - \sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_i}\underbrace{(\tau_a - \phi_i)}_{\geq 0} + \bar{w}_\Psi\Psi_a + w_{at_{I_a-B_a}}\Omega_a \\
&\leq \sum_{i=I_a-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} - \sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_{I_a-B_a}}(\tau_a - \phi_i) + \bar{w}_\Psi\Psi_a + w_{at_{I_a-B_a}}\Omega_a \\
&= \sum_{i=I_a-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} + w_{at_{I_a-B_a}}\left(\Phi_a + \Omega_a - (B_a - \ell_a)\tau_a\right) + \bar{w}_\Psi\Psi_a
\end{aligned} \tag{11}$$

In both cases (10) and (11), we have shown that

$$\sum_{i=I_a-B_a+1}^{I_a-\ell_a} w_{at_i}\phi_i + \sum_{i=I_a-\ell_a+1}^{I_a} w_{at_i}\psi_i + w_{at_{I_a-B_a}}\Omega_a$$
$$\leq \sum_{i=I_a-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} + v\left(\Phi_a + \Omega_a - (B_a - \ell_a)\tau_a\right) + \bar{w}_\Psi\Psi_a$$

for a $v \leq \bar{w}_\Psi$. If $\ell_a > 0$, we can use $v \leq \bar{w}_\Psi$ to charge the remaining mass to Ψ_a if the factors over $\{I_a - \ell_a + 1, \ldots, T\}$ leave enough space. In symbols, this means

$$\begin{aligned}
&\sum_{i=T-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} + v\left(\Phi_a + \Omega_a - (B_a - \ell_a)\tau_a\right) + \bar{w}_\Psi\Psi_a \\
&\leq \sum_{i=T-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} + \bar{w}_\Psi \max\left\{\Phi_a + \Omega_a - (B_a - \ell_a)\tau_a, 0\right\} + \bar{w}_\Psi\Psi_a \\
&= \sum_{i=T-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} + \bar{w}_\Psi \max\left\{\Phi_a + \Psi_a + \Omega_a - (B_a - \ell_a)\tau_a, \Psi_a\right\} \\
&= \sum_{i=T-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} + \bar{w}_\Psi \max\left\{\ell_a\tau_a, \Psi_a\right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \tau_a \sum_{i=T-B_a+1}^{I_a-\ell_a} w_{at_i} + \max\left\{\tau_a, \frac{\Psi_a}{\ell_a}\right\} \sum_{i=T-\ell_a+1}^{I_a} w_{at_i} \\
&\leq \max\left\{\tau_a, \frac{\Psi_a}{\ell_a}\right\} \sum_{t \in S_a} w_{at}.
\end{aligned}
$$

If $\ell_a = 0$, we have $\Psi_a = 0$ and immediately obtain by definition of τ_a that

$$
\sum_{i=I_a-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i} + v\left(\Phi_a + \Omega_a - (B_a - \ell_a)\,\tau_a\right) + \bar{w}_\Psi \Psi_a = \sum_{i=I_a-B_a+1}^{I_a-\ell_a} \tau_a w_{at_i}.
$$

□

Lemma 10. *We have*

$$
\frac{\Psi_a}{\ell_a} = 1 + \left(\frac{B_a}{\ell_a} - 1\right) \frac{e_{B_a}^{\alpha\ell_a/B_a} - 1}{e_{B_a}^{\alpha} - 1}
$$

and

$$
\tau_a = 1 + \frac{1}{e_{B_a}^{\alpha} - 1} \frac{1}{\alpha_B} \left(e_{B_a}^{\alpha} - \frac{e_{B_a}^{\alpha} - 1}{\alpha_{B_a}}\right).
$$

Proof. We compute

$$
\begin{aligned}
\Phi_a &= \sum_{i=T-B_a+1}^{T-\ell_a} \phi_i \\
&= \sum_{i=T-B_a+1}^{T-\ell_a} \left(\frac{1}{e_{B_a}^{\alpha} - 1} \left((B_a - \ell_a)\left(e_{B_a}^{\alpha/B_a} - 1\right) - \frac{1}{\alpha_B} e_{B_a}^{-\alpha\ell_a/B_a}\right) e_{B_a}^{\alpha(T-i)/B_a} + \frac{1}{\alpha_B} \frac{e_{B_a}^{\alpha}}{e_{B_a}^{\alpha} - 1}\right) \\
&= \frac{1}{e_{B_a}^{\alpha} - 1} \left((B_a - \ell_a)\left(e_{B_a}^{\alpha/B_a} - 1\right) - \frac{1}{\alpha_B} e_{B_a}^{-\alpha\ell_a/B_a}\right) \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha\ell_a/B_a}}{e_{B_a}^{\alpha/B_a} - 1} + (B_a - \ell_a) \frac{1}{\alpha_B} \frac{e_{B_a}^{\alpha}}{e_{B_a}^{\alpha} - 1} \\
&= \frac{1}{e_{B_a}^{\alpha} - 1} (B_a - \ell_a) \left(e_{B_a}^{\alpha} - e_{B_a}^{\alpha\ell_a/B_a} + \frac{1}{\alpha_B} e_{B_a}^{\alpha}\right) - \frac{1}{e_{B_a}^{\alpha} - 1} \frac{1}{\alpha_B} \frac{e_{B_a}^{\alpha - \alpha\ell_a/B_a} - 1}{e_{B_a}^{\alpha/B_a} - 1}
\end{aligned}
$$

and

$$
\begin{aligned}
\Psi_a &= \sum_{i=T-\ell_a+1}^{T} \psi_i \\
&= \sum_{i=T-\ell_a+1}^{T} \left(1 + (B_a - \ell_a) \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} e_{B_a}^{\alpha(T-i)/B_a}\right) \\
&= \ell_a + (B_a - \ell_a) \frac{e_{B_a}^{\alpha/B_a} - 1}{e_{B_a}^{\alpha} - 1} \frac{e_{B_a}^{\alpha\ell_a/B_a} - 1}{e_{B_a}^{\alpha/B_a} - 1} \\
&= \ell_a + (B_a - \ell_a) \frac{e_{B_a}^{\alpha\ell_a/B_a} - 1}{e_{B_a}^{\alpha} - 1}.
\end{aligned}
$$

Summing up,

$$
\begin{aligned}
&\Phi_a + \Psi_a + \Omega_a \\
&= \frac{1}{e^{\alpha} - 1} (B_a - \ell_a) \left(e_{B_a}^{\alpha} - e_{B_a}^{\alpha\ell_a/B_a} + \frac{1}{\alpha_B} e_{B_a}^{\alpha}\right) - \frac{1}{e_{B_a}^{\alpha} - 1} \frac{1}{\alpha_B} \frac{e_{B_a}^{\alpha - \alpha\ell_a/B_a} - 1}{e_{B_a}^{\alpha/B_a} - 1} \\
&\quad + \ell_a + (B_a - \ell_a) \frac{e_{B_a}^{\alpha\ell_a/B_a} - 1}{e_{B_a}^{\alpha} - 1} + \frac{1}{e_{B_a}^{\alpha} - 1} \frac{1}{\alpha_B} \left(\ell_a e_{B_a}^{\alpha} - \frac{e_{B_a}^{\alpha} - e_{B_a}^{\alpha(B_a - \ell_a)/B_a}}{e_{B_a}^{\alpha/B_a} - 1}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{e^{\alpha}_{B_a}-1}\left(B_a-\ell_a\right)\left(e^{\alpha}_{B_a}-1+\frac{1}{\alpha_B}e^{\alpha}_{B_a}\right)-\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}\frac{e^{\alpha}_{B_a}-1}{e^{\alpha/B_a}_{B_a}-1}\\
&\quad+\ell_a+\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}\ell_a e^{\alpha}_{B_a}\\
&= B_a+\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}\left(B_a-\ell_a\right)e^{\alpha}_{B_a}-\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}\frac{e^{\alpha}_{B_a}-1}{e^{\alpha/B_a}_{B_a}-1}+\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}\ell_a e^{\alpha}_{B_a}\\
&= B_a+\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}B_a e^{\alpha}_{B_a}-\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}\frac{e^{\alpha}_{B_a}-1}{e^{\alpha/B_a}_{B_a}-1}\\
&= B_a+\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}\left(B_a e^{\alpha}_{B_a}-\frac{e^{\alpha}_{B_a}-1}{e^{\alpha/B_a}_{B_a}-1}\right)
\end{aligned}
$$

which does no longer depend on ℓ_a. Dividing Ψ_a by ℓ_a and $\Phi_a+\Psi_a+\Omega_a$ by B_a yields the result. $\square$

Putting everything together, we have $\mathrm{PRD}/\mathrm{ALG} \le \max_a \max\left\{\tau_a, \max_{\ell_a\in\{1,\dots,B_a\}} \Psi_a/\ell_a\right\}$ as τ_a does not depend on ℓ_a. The reader can refer back to Figure 2 for an illustration of this upper bound. In the following lemma, we further analyze analytically $\max_{\ell_a\in\{1,\dots,B_a\}} \Psi_a/\ell_a$ and compare it with τ_a to obtain the upper bound:

Lemma 11. *The consistency of Algorithm 1 is given by*

$$
\mathrm{PRD}/\mathrm{ALG} \le \left(1+\frac{1}{e^{\alpha}_{B}-1}\max\left\{\frac{1}{\alpha_B}\left(e^{\alpha}_{B}-\frac{e^{\alpha}_{B}-1}{\alpha_B}\right), \ln\left(e^{\alpha}_{B}\right)\right\}\right).
$$

Proof. Due to Lemma 9, it is sufficient to show

$$
1+\frac{1}{e^{\alpha}_{B}-1}\max\left\{\frac{1}{\alpha_B}\left(e^{\alpha}_{B}-\frac{e^{\alpha}_{B}-1}{\alpha_B}\right), \ln\left(e^{\alpha}_{B}\right)\right\} \ge \begin{cases}\max\left\{\tau_a, \Psi_a/\ell_a\right\} & \text{if } \ell_a>0\\ \tau_a & \text{otherwise.}\end{cases}
$$

By Lemma 10, we know for the first term in the maximum that

$$
\tau_a = 1+\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}\left(e^{\alpha}_{B_a}-\frac{e^{\alpha}_{B_a}-1}{\alpha_{B_a}}\right)
$$

This term is maximized for $B_a = B$ since

$$
\begin{aligned}
1+\frac{1}{e^{\alpha}_{B_a}-1}\frac{1}{\alpha_B}\left(e^{\alpha}_{B_a}-\frac{e^{\alpha}_{B_a}-1}{B_a\left(e^{\alpha/B_a}_{B_a}-1\right)}\right) &= 1+\frac{1}{\alpha_B}\left(\frac{e^{\alpha}_{B_a}}{e^{\alpha}_{B_a}-1}-\frac{1}{\alpha_{B_a}}\right)\\
&\le 1+\frac{1}{\alpha_B}\left(\frac{e^{\alpha}_{B}}{e^{\alpha}_{B}-1}-\frac{1}{\alpha_B}\right) = 1+\frac{1}{e^{\alpha}_{B}-1}\underbrace{\frac{1}{\alpha_B}\left(e^{\alpha}_{B}-\frac{e^{\alpha}_{B}-1}{\alpha_B}\right)}_{=:p(\alpha)}
\end{aligned}
$$

due to Lemma 2. The lemma statement therefore follows immediately if $\ell_a = 0$. We may thus assume that $\ell_a > 0$ and use Lemma 10 to determine the second term in the maximum as

$$
\frac{\Phi_a}{\ell_a} = 1+\frac{1}{e^{\alpha}_{B_a}-1}\left(\frac{1}{x}-1\right)\left(e^{\alpha x}_{B_a}-1\right)
$$

where $x =: \ell_a/B_a$. The second term behaves similarly to the first as

$$
\frac{\Phi_a}{\ell_a} = 1+\frac{1}{e^{\alpha}_{B_a}-1}\left(\frac{1}{x}-1\right)\left(e^{\alpha x}_{B_a}-1\right) \le 1+\frac{1}{e^{\alpha}_{B}-1}\left(\frac{1}{x}-1\right)\left(e^{\alpha x}_{B}-1\right)
$$

since $\left(e^{\alpha x}_{B_a}-1\right)/\left(e^{\alpha}_{B_a}-1\right) \le \left(e^{\alpha x}_{B}-1\right)/\left(e^{\alpha}_{B}-1\right)$. We define $g(\alpha,x) := \left(\frac{1}{x}-1\right)\left(e^{\alpha x}_{B}-1\right)$ such that we can write

$$
\max\left\{\frac{\Phi_a+\Psi_a+\Omega_a}{B_a}, \frac{\Psi_a}{\ell_a}\right\} \le 1+\frac{1}{e^{\alpha}_{B}-1}\max\left\{p(\alpha), g(\alpha,x)\right\}.
$$

We want to remove the dependency on x in g by maximizing g over $x \in [0,1]$ for a fixed α. As $g(\alpha, x)$ is continuous, it suffices to evaluate g in both endpoints and find the stationary points. We have

$$\begin{aligned} g(\alpha,0) = \lim_{x\to 0}\left(\frac{1}{x}-1\right)(e_B^{\alpha x}-1) &= \lim_{x\to 0}\frac{(1-x)\left(e_B^{\alpha x}-1\right)}{x} \\ &= \lim_{x\to 0} -\left(e_B^{\alpha x}-1\right) + (1-x)\ln(e_B^{\alpha})e_B^{\alpha x} = \ln(e_B^{\alpha}) \end{aligned}$$

by L'Hoptial. Further, $g(\alpha,1)=0$. Next, we find the stationary points $x^* \in [0,1]$ as solutions to the equation

$$\frac{\partial}{\partial x} g(\alpha, x^*) = \ln(e_B^{\alpha})\left(\frac{1}{x^*}-1\right)e_B^{\alpha x^*} - \frac{e_B^{\alpha x^*}-1}{(x^*)^2} = 0$$

which is equivalent to

$$e_B^{\alpha x^*} - 1 = \ln(e_B^{\alpha})(x^*)^2\left(\frac{1}{x^*}-1\right)e_B^{\alpha x^*}.$$

There is no closed form solution for x^*, but we can replace $e_B^{\alpha x}-1$ in g with the RHS of the above. This yields a new function

$$h(\alpha, y) = \left(\frac{1}{y}-1\right)\ln(e_B^{\alpha})y^2\left(\frac{1}{y}-1\right)e_B^{\alpha y} = \ln(e_B^{\alpha})\left(1-y\right)^2 e_B^{\alpha y}$$

with $h(\alpha, x^*) = g(\alpha, x^*)$. We can thus maximize h over $y \in [0,1]$ to obtain an upper bound on $g(x^*)$. Note that $h(\alpha, 0) = \ln(e_B^{\alpha}) = g(\alpha, 0)$ and $h(\alpha,1) = 0 = g(\alpha,1)$. To this end, let y^* be such that

$$\frac{\partial}{\partial y^*} h(\alpha, y^*) = \ln(e_B^{\alpha})^2\left(1-y^*\right)^2 e_B^{\alpha y^*} - 2\ln(e_B^{\alpha})\left(1-y^*\right)e_B^{\alpha y^*} = 0$$

which is equivalent to $\ln(e_B^{\alpha})\left(1-y^*\right) - 2 = 0$ or $y^* = 1 - \frac{2}{\ln(e_B^{\alpha})}$. We evaluate h in y^* and obtain

$$h(\alpha, y^*) = \alpha\ln(e_B)\left(\frac{2}{\alpha\ln(e_B)}\right)^2 e_B^{\alpha - \frac{2}{\ln(e_B)}} = \frac{4}{\alpha\ln(e_B)e^2}e_B^{\alpha}.$$

Note that $y^* \geq 0 \iff \alpha \geq 2/\ln(e_B)$. Furthermore, $h^*(\alpha)$ always exceeds the endpoint $g(\alpha, 0)$: We calculate

$$\begin{aligned} h^*(\alpha) := h(\alpha, y^*) &= \frac{4}{\ln(e_B^{\alpha})e^2}e_B^{\alpha} \\ &\geq \frac{4}{\ln(e_B^{\alpha})e^2}e^2\ln(e_B^{\alpha/2})^2 \\ &= \ln(e_B^{\alpha}) \end{aligned}$$

where the inequality is due to $e^z \geq ez$ for $z = \ln\left(e_B^{\alpha/2}\right) \geq 0$. Therefore, for all $x \in [0,1]$,

$$g(\alpha, x) \leq \begin{cases} \ln(e_B^{\alpha}) & \text{if } \alpha \leq \frac{2}{\ln(e_B)} \\ h^*(\alpha) & \text{otherwise.} \end{cases}$$

We consider both intervals separately. Let us first consider the the case when $\alpha \in \left[0, \frac{2}{\ln(e_B)}\right]$. If $B < \infty$, there could be multiple intersection points between $p(\alpha)$ and $\alpha\ln(e_B)$. However, the situation is easier if $B \to \infty$ as the intersection points given by

$$p(\alpha) = \frac{1}{\alpha}\left(e^{\alpha} - \frac{e^{\alpha}-1}{\alpha}\right) = \alpha \iff \alpha e^{\alpha} - e^{\alpha} + 1 = \alpha^3$$

are at $\alpha = 1$ and $\alpha^* \approx 1.79$, whereas $\alpha\ln(e_B)$ dominates $p(\alpha)$ between 1 and α^*.

It remains to consider the case $\alpha \geq \frac{2}{\ln(e_{B_a})}$. Again, there can be many intersection points of $p(\alpha)$ with $\alpha\ln(e_{B_a})$ and $h^*(\alpha)$. However, if $B \to \infty$, then $p(\alpha)$ already dominates $h^*(\alpha)$ for $\alpha > 2$ which we can see as follows. First,

$$h^*(\alpha) = \frac{4}{\alpha}e^{\alpha-2} \leq \frac{1}{\alpha}\left(e^{\alpha} - \frac{e^{\alpha}-1}{\alpha}\right) = p(\alpha)$$

$$\iff 4e^{-2} \leq 1 - \frac{1-e^{-\alpha}}{\alpha}.$$

We can see that $\frac{1-e^{-\alpha}}{\alpha}$ is decreasing in α as

$$\frac{\partial}{\partial \alpha}\frac{1-e^{-\alpha}}{\alpha} = \frac{e^{-\alpha}\left(\alpha - e^{\alpha} + 1\right)}{\alpha^2} \leq 0$$

which holds as $1+\alpha \leq e^{\alpha}$. Finally, we check that $h^*(2) = 2 \leq 2.10 \approx p(2)$. $\square$

B Generalized Assignment Problem

The generalized assignment problem (GAP) is a generalization of Display Ads where impressions t can take up any size u_{at} in the budget constraint of advertiser a. This formulation encompasses both Display Ads and Ad Words, and we empirically compare it to the Ad Words algorithm with predictions due to Mahdian et al. (2007) in Section C.3. For simplicity of presentation, we assume that budgets are all 1 and instead, $u_{at} \to 0$. However, as before it is possible to adapt the algorithm to work with large sizes u_{at}. We state the LP below.

GAP Primal	GAP Dual
$\max \sum_{a,t} w_{at} x_{at}$	$\min \sum_a \beta_a + \sum_t z_t$
$\forall a\colon \sum_t u_{at} x_{at} \leq 1$	$\forall a, t\colon z_t \geq w_{at} - u_{at}\beta_a$
$\forall t\colon \sum_a x_{at} \leq 1$	

Algorithm 2 is a generalization of Algorithm 1 to GAP. An immediate difference is that the discounted gain $w_{at} - u_{at}\beta_a$ respects the impression size u_{at} in accordance with the changed dual constraint. We still follow the predicted advertiser if its discounted gain still is a sufficiently high fraction of the maximum discounted gain. However, we might now have to remove multiple impressions with least value-size ratio to accommodate the new impression. The update for β_a also differs and is based on value-size ratios of impressions allocated to a: For a fixed advertiser a let $U_a = \sum_{t \in \mathbf{X}_a} u_{at}$, be the total size of all impressions ever allocated to a. For any $x \in (0, U_a]$ define $\frac{w_x}{u_x}$ as the minimal ratio such that

$$\sum_{t \in \mathbf{X}_a \colon \frac{w_{at}}{u_{at}} \leq \frac{w_x}{u_x}} u_{at} > x. \tag{12}$$

Then, we can naturally define β_a as the exponential average over ratios $\frac{w_x}{u_x}$. As before, we also assume that there exists a dummy advertiser that only receives impressions of zero value-size ratio and that all advertisers are initially filled up with impressions of zero value.

B.1 Robustness

Theorem 12. *Algorithm 1 has a robustness of*

$$\frac{\text{ALG}}{\text{OPT}} \geq \frac{e^{\alpha}-1}{\alpha e^{\alpha}}$$

Proof. Assume we assign impression t to advertiser a while disposing of some impressions to make space. We will bound the dual increase as a multiple of the primal increase. We now assume that after allocating t to a, it becomes the impression with highest value-size ratio (a general proof follows analogously to the proof of robustness for Display Ads in Section A.1). The primal increase is simply

$$\Delta P = \int_{U_a - u_{at}}^{U_a} \frac{w_x}{u_x} dx - \int_{U_a - 1 - u_{at}}^{U_a - 1} \frac{w_x}{u_x} dx = w_{at} - \int_{U_a - 1 - u_{at}}^{U_a - 1} \frac{w_x}{u_x} dx.$$

Algorithm 2 Exponential Averaging with Predictions for GAP

```
Input: Robustness-consistency trade-off parameter α ∈ [1, ∞)
For each advertiser a, initialize β_a ← 0 and fill up a with zero-value impressions
for all arriving impressions t do
    a_(PRD) ← PRD(t)
    a_(EXP) ← arg max_a {w_at − u_at β_a}
    if α (w_{a(PRD),t} − u_{a(PRD),t} β_{a(PRD)}) ≥ w_{a(EXP),t} − u_{a(EXP),t} β_{a(EXP)} then
        a ← a_(PRD)
    else
        a ← a_(EXP)
    end if
    Dispose of impressions with least value-size ratio currently allocated to a until there is u_at of
    free space and allocate t to a
    Let w_x/u_x as in (12) and update β_a ← α/(e^α − 1) ∫_{U_a−1}^{U_a} (w_x/u_x) e^{α(U_a−x)} dx
end for
```

At the same time,

$$
\begin{aligned}
\beta_a^{(t)} &= \frac{\alpha}{e^\alpha - 1} \int_{U_a-1}^{U_a} \frac{w_x}{u_x} e^{\alpha(U_a-x)} dx \\
&= \frac{\alpha}{e^\alpha - 1} \left(\int_{U_a-1-u_{at}}^{U_a-u_{at}} \frac{w_x}{u_x} e^{\alpha(U_a-x)} dx + \int_{U_a-u_{at}}^{U_a} \frac{w_x}{u_x} e^{\alpha(U_a-x)} dx - \int_{U_a-1-u_{at}}^{U_a-1} \frac{w_x}{u_x} e^{\alpha(U_a-x)} dx \right) \\
&= \frac{\alpha}{e^\alpha - 1} \left(e^{\alpha u_{at}} \int_{U_a-1-u_{at}}^{U_a-u_{at}} \frac{w_x}{u_x} e^{\alpha(U_a-x-u_{at})} dx + w_{at} - \int_{U_a-1-u_{at}}^{U_a-1} \frac{w_x}{u_x} e^{\alpha(U_a-x)} dx \right) \\
&= e^{\alpha u_{at}} \beta_a^{(t-1)} + \frac{\alpha}{e^\alpha - 1} \left(w_{at} - \int_{U_a-1-u_{at}}^{U_a-1} \frac{w_x}{u_x} e^{\alpha(U_a-x)} dx \right)
\end{aligned}
$$

We set $z_t = \alpha \left(w_{at} - u_{at} \beta_a^{(t-1)} \right)$ and obtain, since $e^{\alpha u_{at}} - 1 = \alpha u_{at}$ due to $u_{at} \to 0$,

$$
\begin{aligned}
\Delta D &= \beta_a^{(t)} - \beta_a^{(t-1)} + z_t \\
&= \left(e^{\alpha u_{at}} - 1\right) \beta_a^{(t-1)} + \frac{\alpha}{e^\alpha - 1} \left(w_{at} - \int_{U_a-1-u_{at}}^{U_a-1} \frac{w_x}{u_x} e^{\alpha(U_a-x)} dx \right) + \alpha \left(w_{at} - u_{at} \beta_a^{(t-1)} \right) \\
&= \alpha u_{at} \beta_a^{(t-1)} + \frac{\alpha}{e^\alpha - 1} \left(w_{at} - \int_{U_a-1-u_{at}}^{U_a-1} \frac{w_x}{u_x} e^{\alpha(U_a-x)} dx \right) + \alpha \left(w_{at} - u_{at} \beta_a^{(t-1)} \right) \\
&= \frac{\alpha e^\alpha}{e^\alpha - 1} w_{at} - \frac{\alpha e^\alpha}{e^\alpha - 1} \int_{U_a-1-u_{at}}^{U_a-1} \frac{w_x}{u_x} dx \\
&= \frac{\alpha e^\alpha}{e^\alpha - 1} \Delta P.
\end{aligned}
$$

□

B.2 Consistency

Theorem 13. *Algorithm 1 has a consistency of*

$$
\frac{\mathrm{ALG}}{\mathrm{PRD}} \geq \left(1 + \frac{1}{e^\alpha - 1} \max \left\{ \frac{1}{\alpha} \left(e^\alpha - \frac{e^\alpha - 1}{\alpha} \right), \alpha \right\} \right)^{-1}.
$$

As before, we split the impressions t based on whether the algorithm followed the prediction or not. If the algorithm ignores the prediction, we can use that $\alpha \left(w_{a_{(\mathrm{PRD})},t} - u_{a_{(\mathrm{PRD})},t} \beta_{a_{(\mathrm{PRD})}} \right) \leq$

$w_{a_{(\mathrm{EXP})},t} - u_{a_{(\mathrm{EXP})},t}\beta_{a_{(\mathrm{EXP})}}$ due to Line 6 in Algorithm 2. With a similar calculation, we obtain

$$\begin{aligned}\mathrm{PRD} &= \sum_a \mathrm{PRD}_a \\ &= \sum_a \left(\sum_{t \in \mathbf{P}_a \cap \mathbf{X}_a} w_{at} + \frac{1}{\alpha} \sum_{t \in \mathbf{X}_a \setminus \mathbf{P}_a} w_{at} - \frac{1}{\alpha} \sum_{t \in \mathbf{X}_a \setminus \mathbf{P}_a} u_{at}\beta_a^{(t-1)} + \sum_{t \in \mathbf{P}_a \setminus \mathbf{X}_a} u_{at}\beta_a^{(t-1)} \right).\end{aligned}$$

Once again, we fix an advertiser a. Let $\rho_a \coloneqq \sum_{t \in \mathbf{P}_a \cap \mathbf{X}_a} u_{at}$ so that we can bound

$$\sum_{t \in \mathbf{P}_a \setminus \mathbf{X}_a} u_{at}\beta_a^{(t-1)} \leq (1 - \rho_a)\, \beta_a^{(T)}.$$

We still have to argue that the worst-case is when all impressions are ordered such that their value-size ratios are non-decreasing and the impressions in $\mathbf{P}_a \cap \mathbf{X}_a$ are the ones with maximum value-size ratio among $\mathbf{X}_a$. The latter is obvious as it can only increase the value of PRD, so it remains to show that the non-decreasing value-size ordering minimizes the third sum in PRD_a (the first two sums are invariant under reordering). To this end, note that the value of β_a for GAP is the limit of β_a for Display Ads in the following sense: For positive $\epsilon \to 0$, we can split each GAP-impression $t \in \mathbf{X}_a$ into $\frac{u_t}{\epsilon}$ identical Display Ads-impressions with value $\frac{w_t}{u_t}$, while assuming a budget of $1/\epsilon$. Then, the GAP β_a and Display Ads β_a are identical. As we know from Display Ads, the worst case is achieved when the Display Ads-impressions with value $\frac{w_t}{u_t}$ are in non-decreasing order. In this ordering, consecutive Display-Ads impressions with identical value $\frac{w_t}{u_t}$ still correspond to the same GAP-impression t, so we also know that this ordering is the worst-case for GAP. We may therefore assume that the impressions are ordered such that their value-size ratios are non-decreasing. As such, we obtain

$$\beta_a^{(t)} = \frac{\alpha}{e^\alpha - 1} \int_{U_a^{(t)}-1}^{U_a^{(t)}} \frac{w_x}{u_x} e^{\alpha(U_a^{(t)} - x)} dx = \int_{y-1}^{y} \frac{w_x}{u_x} e^{\alpha(y-x)} dx =: \beta_a^{(y)}$$

where $y = U_a^{(t)}$. Combined with the fact that $\mathbf{P}_a \cap \mathbf{X}_a$ are last impressions in $\mathbf{X}_a$, we can now write

$$\begin{aligned}&\sum_{t \in \mathbf{P}_a \cap \mathbf{X}_a} w_{at} + \frac{1}{\alpha} \sum_{t \in \mathbf{X}_a \setminus \mathbf{P}_a} w_{at} - \frac{1}{\alpha} \sum_{t \in \mathbf{X}_a \setminus \mathbf{P}_a} u_{at}\beta_a^{(t-1)} \\ &\qquad = \int_{U_a - \rho}^{U_a} \frac{w_x}{u_x} dx + \frac{1}{\alpha} \int_0^{U_a - \rho_a} \frac{w_x}{u_x} dx - \frac{1}{\alpha} \int_0^{U_a - \rho_a} \beta_a^{(x)} dx\end{aligned}$$

This helps us to compute $\beta_a^{(x)}$ in PRD_a and rewrite the whole term as a linear combination of value-size ratios.

Lemma 14. *We have*

$$\mathrm{PRD}_a \leq \int_{U_a - 1}^{U_a - \rho_a} \frac{w_x}{u_x} \phi_x dx + \int_{U_a - \rho_a}^{U_a} \frac{w_x}{u_x} \psi_x dx + \frac{w_{U_a - 1}}{u_{U_a - 1}} \Omega_a$$

where

$$\begin{aligned}\phi_x &\coloneqq (1 - \rho_a)\, \frac{\alpha}{e^\alpha - 1} e^{\alpha(U_a - x)} + \frac{1}{\alpha} \frac{e^\alpha - e^{\alpha(U_a - \rho_a - x)}}{e^\alpha - 1} \\ \psi_x &\coloneqq 1 + (1 - \rho_a)\, \frac{\alpha}{e^\alpha - 1} e^{\alpha(U_a - x)} \\ \Omega_a &\coloneqq \frac{1}{\alpha} \frac{1}{e^\alpha - 1} \left(\rho_a e^\alpha - \frac{1}{\alpha} \left(e^\alpha - e^{\alpha(1 - \rho_a)} \right) \right)\end{aligned}$$

Proof. We rewrite the third sum in PRD_a to

$$\int_0^{U_a - \rho_a} \beta_a^{(x)} dx$$

$$
\begin{aligned}
&= \frac{\alpha}{e^{\alpha}-1}\int_0^{U_a-\rho_a}\int_{x-1}^{x}\frac{w_y}{u_y}e^{\alpha(x-y)}dydx \\
&= \frac{\alpha}{e^{\alpha}-1}\int_0^{U_a-\rho_a}\frac{w_y}{u_y}\int_0^{\min\{1,U_a-\rho_a-y\}}e^{\alpha x}dxdy \\
&= \frac{\alpha}{e^{\alpha}-1}\int_0^{U_a-1-\rho_a}\frac{w_y}{u_y}\int_0^{1}e^{\alpha x}dxdy + \frac{\alpha}{e^{\alpha}-1}\int_{U_a-1-\rho_a}^{U_a-\rho_a}\frac{w_y}{u_y}\int_0^{U_a-\rho_a-y}e^{\alpha x}dxdy \\
&= \int_0^{U_a-1-\rho_a}\frac{w_y}{u_y}dy + \frac{1}{e^{\alpha}-1}\int_{U_a-1-\rho_a}^{U_a-\rho_a}\frac{w_y}{u_y}\left(e^{\alpha(U_a-\rho_a-y)}-1\right)dy
\end{aligned}
$$

where for the last equality, we simply evaluated the integral. Using this in place of the second sum in PRD_a cancels out most of the terms of the second sum:

$$
\begin{aligned}
&\int_0^{U_a-\rho_a}\frac{w_x}{u_x}dx - \int_0^{U_a-\rho_a}\beta_a^{(x)}dx \\
&= \int_0^{U_a-\rho_a}\frac{w_x}{u_x}dx - \int_0^{U_a-1-\rho_a}\frac{w_y}{u_y}dy - \frac{1}{e^{\alpha}-1}\int_{U_a-1-\rho_a}^{U_a-\rho_a}\frac{w_y}{u_y}\left(e^{\alpha(U_a-\rho_a-y)}-1\right)dy \\
&= \int_{U_a-1-\rho_a}^{U_a-\rho_a}\frac{w_y}{u_y}\left(1-\frac{e^{\alpha(U_a-\rho_a-y)}-1}{e^{\alpha}-1}\right)dy \\
&= \int_{U_a-1-\rho_a}^{U_a-\rho_a}\frac{w_y}{u_y}\frac{e^{\alpha}-e^{\alpha(U_a-\rho_a-y)}}{e^{\alpha}-1}dy \\
&= \int_{U_a-1}^{U_a-\rho_a}\frac{w_y}{u_y}\frac{e^{\alpha}-e^{\alpha(U_a-\rho_a-y)}}{e^{\alpha}-1}dy + \int_{U_a-1-\rho_a}^{U_a-1}\frac{w_y}{u_y}\frac{e^{\alpha}-e^{\alpha(U_a-\rho_a-y)}}{e^{\alpha}-1}dy.
\end{aligned}
\tag{13}
$$

We upper bound the third sum

$$
\begin{aligned}
\frac{1}{\alpha}\int_{U_a-1-\rho_a}^{U_a-1}\frac{w_y}{u_y}\frac{e^{\alpha}-e^{\alpha(U_a-\rho_a-y)}}{e^{\alpha}-1}dy &\le \frac{w_{U_a-1}}{u_{U_a-1}}\frac{1}{\alpha}\int_{U_a-1-\rho_a}^{U_a-1}\frac{e^{\alpha}-e^{\alpha(U_a-\rho_a-y)}}{e^{\alpha}-1}dy \\
&= \frac{w_{U_a-1}}{u_{U_a-1}}\frac{1}{\alpha}\frac{1}{e^{\alpha}-1}\left(\rho_a e^{\alpha}-\int_{1-\rho_a}^{1}e^{\alpha y}dy\right) \\
&= \frac{w_{U_a-1}}{u_{U_a-1}}\underbrace{\frac{1}{\alpha}\frac{1}{e^{\alpha}-1}\left(\rho_a e^{\alpha}-\frac{1}{\alpha}\left(e^{\alpha}-e^{\alpha(1-\rho_a)}\right)\right)}_{=\Omega_a}
\end{aligned}
\tag{14}
$$

Furthermore,

$$
(1-\rho_a)\,\beta_a^{(U_a)} = (1-\rho_a)\,\frac{\alpha}{e^{\alpha}-1}\int_{U_a-1}^{U_a}\frac{w_x}{u_x}e^{\alpha(U_a-x)}dx
\tag{15}
$$

Combining (13), (14), and (15) and grouping terms yields

$$
\begin{aligned}
&\int_{U_a-\rho_a}^{U_a}\frac{w_x}{u_x}dx + \frac{1}{\alpha}\int_{U_a-1}^{U_a-\rho_a}\frac{w_y}{u_y}\frac{e^{\alpha}-e^{\alpha(U_a-\rho_a-y)}}{e^{\alpha}-1}dy + \frac{w_{U_a-1}}{u_{U_a-1}}\Omega_a \\
&\quad + (1-\rho_a)\,\frac{\alpha}{e^{\alpha}-1}\int_{U_a-1}^{U_a}\frac{w_x}{u_x}e^{\alpha(U_a-x)}dx \\
&= \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\underbrace{\left((1-\rho_a)\,\frac{\alpha}{e^{\alpha}-1}e^{\alpha(U_a-x)}+\frac{1}{\alpha}\frac{e^{\alpha}-e^{\alpha(U_a-\rho_a-x)}}{e^{\alpha}-1}\right)}_{=\phi_x}dx \\
&\quad + \int_{U_a-\rho_a}^{U_a}\frac{w_x}{u_x}\underbrace{\left(1+(1-\rho_a)\,\frac{\alpha}{e^{\alpha}-1}e^{\alpha(U_a-x)}\right)}_{=\psi_x}dx + \frac{w_{U_a-1}}{u_{U_a-1}}\Omega_a.
\end{aligned}
$$

$\square$

Analogously to Display Ads, we define

$$\Phi_a \coloneqq \int_{U_a-1}^{U_a-\rho_a} \phi_x dx \qquad \text{and} \qquad \Psi_a \coloneqq \int_{U_a-\rho}^{U_a} \psi_x dx$$

and the total coefficient $\tau_a \coloneqq \Phi_a + \Psi_a + \Omega_a$ which by a calculation similar to Lemma 10 can be shown to be

$$\tau_a = 1 + \frac{1}{e^\alpha - 1}\frac{1}{\alpha}\left(e^\alpha - \frac{e^\alpha - 1}{\alpha}\right).$$

Lemma 15. *We have*

$$\text{PRD} \le \max\left\{\tau_a, \frac{\Psi_a}{\rho_a}\right\}\text{ALG}$$

if $\rho_a > 0$ *and otherwise,*

$$\text{PRD} \le \tau_a \text{ALG}.$$

Proof. Again, let

$$\bar{w}_\Phi \coloneqq \frac{1}{1-\rho_a}\int_{U_a-1}^{U_a-\rho_a} \frac{w_x}{u_x} dx$$
$$\bar{w}_\Psi \coloneqq \frac{1}{\rho_a}\int_{U_a-\rho_a}^{U_a} \frac{w_x}{u_x} dx$$

be the average coefficients on the intervals $[U_a - 1, U_a - \rho_a]$ and $[U_a - \rho_a, U_a]$, respectively. The latter coefficients are still decreasing as

$$\psi_x = 1 + \underbrace{(1-\rho_a)\frac{\alpha}{e^\alpha - 1}}_{\ge 0} e^{\alpha(U_a - x)}$$

so we can bound the linear combination

$$\int_{U_a-\rho_a}^{U_a} \frac{w_x}{u_x}\psi_x dx \le \bar{w}_\Psi \int_{U_a-\rho}^{U_a} \psi_x dx = \bar{w}_\Psi \Psi_a.$$

However, ϕ_x is not always decreasing which can be seen by rearranging

$$\begin{aligned}\phi_x &= (1-\rho_a)\frac{\alpha}{e^\alpha - 1}e^{\alpha(U_a-x)} + \frac{1}{\alpha}\frac{e^\alpha - e^{\alpha(U_a-\rho_a-x)}}{e^\alpha - 1} \\ &= \frac{1}{e^\alpha - 1}\left((1-\rho_a)\alpha - \frac{1}{\alpha}e^{-\alpha\rho_a}\right)e^{\alpha(U_a-x)} + \frac{1}{\alpha}\frac{1}{e^\alpha - 1}e^\alpha\end{aligned}$$

We observe that ϕ_x is decreasing if $(1-\rho_a)\alpha$ is at least $\frac{1}{\alpha}e^{-\alpha\rho_a}$, and we analyze two cases based on the relationship of both terms:

- $(1-\rho_a)\alpha \ge \frac{1}{\alpha}e^{-\alpha\rho_a}$: We have $\int_{U_a-1}^{U_a-\rho_a} \frac{w_x}{u_x}\phi_x dx \le \bar{w}_\Phi \Phi_a$ and thus

$$\begin{aligned}&\int_{U_a-1}^{U_a-\rho_a} \frac{w_x}{u_x}\phi_x dx + \int_{U_a-\rho}^{U_a} \frac{w_x}{u_x}\psi_x dx + \frac{w_{U_a-1}}{u_{U_a-1}}\Omega_a \\ &\le \bar{w}_\Phi \Phi_a + \bar{w}_\Psi \Psi_a + w_{a,I_a-B_a}\Omega_a \\ &\le \bar{w}_\Phi(\Phi_a + \Omega_a) + \bar{w}_\Psi \Psi_a \\ &= \bar{w}_\Phi(1-\rho_a)\tau_a + \bar{w}_\Phi(\Phi_a + \Omega_a - (1-\rho_a)\tau_a) + \bar{w}_\Psi\Psi_a \\ &= \int_{U_a-1}^{U_a-\rho_a} \frac{w_x}{u_s}\tau_a dx + \bar{w}_\Phi(\Phi_a + \Omega_a - (1-\rho_a)\tau_a) + \bar{w}_\Psi\Psi_a \end{aligned} \tag{16}$$

- $(1-\rho_a)\alpha \le \frac{1}{\alpha}e^{-\alpha\rho_a}$: We can still show that $\phi_x \le \tau_a$ as

$$\phi_x = (1-\rho_a)\frac{\alpha}{e^\alpha - 1} e^{\alpha(U_a - x)} + \frac{1}{\alpha}\frac{e^\alpha - e^{\alpha(U_a-\rho_a-x)}}{e^\alpha - 1}$$

$$\leq 1 + \frac{1}{e^\alpha - 1}\frac{1}{\alpha}\left(e^\alpha - \frac{e^\alpha - 1}{\alpha}\right) = \tau_a$$

$$\iff \underbrace{\left((1-\rho_a)\,\alpha - \frac{1}{\alpha}e^{-\alpha\rho_a}\right)}_{\leq 0}\underbrace{e^{\alpha(U_a-x)}}_{\geq 0} \leq e^\alpha - 1 - \frac{1}{\alpha}\frac{e^\alpha - 1}{\alpha} = \underbrace{\left(1 - \frac{1}{\alpha^2}\right)}_{\geq 0}\underbrace{(e^\alpha - 1)}_{\geq 0}.$$

Therefore,

$$\begin{aligned}
&\int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\phi_x dx + \int_{U_a-\rho}^{U_a}\frac{w_x}{u_x}\psi_x dx + \frac{w_{U_a-1}}{u_{U_a-1}}\Omega_a \\
&\leq \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\phi_x dx + \bar{w}_\Psi \Psi_a + \frac{w_{U_a-1}}{u_{U_a-1}}\Omega_a \\
&= \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\tau_a dx - \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}(\tau_a - \phi_x)\,dx + \bar{w}_\Psi \Psi_a + \frac{w_{U_a-1}}{u_{U_a-1}}\Omega_a \\
&\leq \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\tau_a dx - \int_{U_a-1}^{U_a-\rho_a}\frac{w_{U_a-1}}{u_{U_a-1}}(\tau_a - \phi_x)\,dx + \bar{w}_\Psi \Psi_a + \frac{w_{U_a-1}}{u_{U_a-1}}\Omega_a \\
&= \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\tau_a dx + \frac{w_{U_a-1}}{u_{U_a-1}}(\Phi_a + \Omega_a - (1-\rho_a)\,\tau_a) + \bar{w}_\Psi \Psi_a \qquad (17)
\end{aligned}$$

In both cases (16) and (17), we have shown that

$$\int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\phi_x dx + \int_{U_a-\rho}^{U_a}\frac{w_x}{u_x}\psi_x dx + \frac{w_{U_a-1}}{u_{U_a-1}}\Omega_a \leq \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\tau_a dx + v\,(\Phi_a + \Omega_a - (1-\rho_a)\,\tau_a) + \bar{w}_\Psi \Psi_a$$

for a $v \leq \bar{w}_\Phi$.

$$\begin{aligned}
&\int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\tau_a dx + v\,(\Phi_a + \Omega_a - (1-\rho_a)\,\tau_a) + \bar{w}_\Psi \Psi_a \\
&\leq \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\tau_a dx + \bar{w}_\Psi \max\{\Phi_a + \Omega_a - (1-\rho_a)\,\tau_a, 0\} + \bar{w}_\Psi \Psi_a \\
&= \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\tau_a dx + \bar{w}_\Psi \max\{\Phi_a + \Psi_a + \Omega_a - (1-\rho_a)\,\tau_a, \Psi_a\} \\
&= \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}\tau_a dx + \bar{w}_\Psi \max\{\rho_a\tau_a, \Psi_a\} \\
&\leq \tau_a \int_{U_a-1}^{U_a-\rho_a}\frac{w_x}{u_x}dx + \max\left\{\tau_a, \frac{\Psi_a}{\rho_a}\right\}\int_{U_a-\rho_a}^{U_a}\frac{w_x}{u_x}dx \\
&\leq \max\left\{\tau_a, \frac{\Psi_a}{\rho_a}\right\}\int_{U_a-1}^{U_a}\frac{w_x}{u_x}dx
\end{aligned}$$

or $\leq \tau_a \int_{U_a-1}^{U_a}\frac{w_x}{u_x}dx$ if $\rho_a = 0$ □

Note that for the bound of Lemma 11, we did not require that ℓ_a is integral. We can thus apply Lemma 11 to bound $\max\left\{\tau_a, \frac{\Psi_a}{\rho_a}\right\}$ and obtain the same result, which proves Theorem 13.

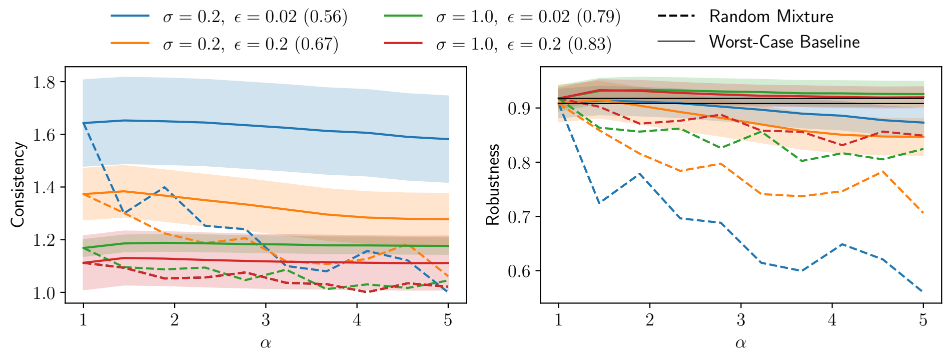

Figure 6: Experimental results for varying values of α on synthetic data with 12 advertisers and 2000 impressions of 10 types, where we report the same quantities as in Figure 3. We use Dual Base predictions for different σ and ϵ. Note that there are two black lines indicating the performance of the worst-case algorithm without predictions, corresponding to the datasets with differing σ.

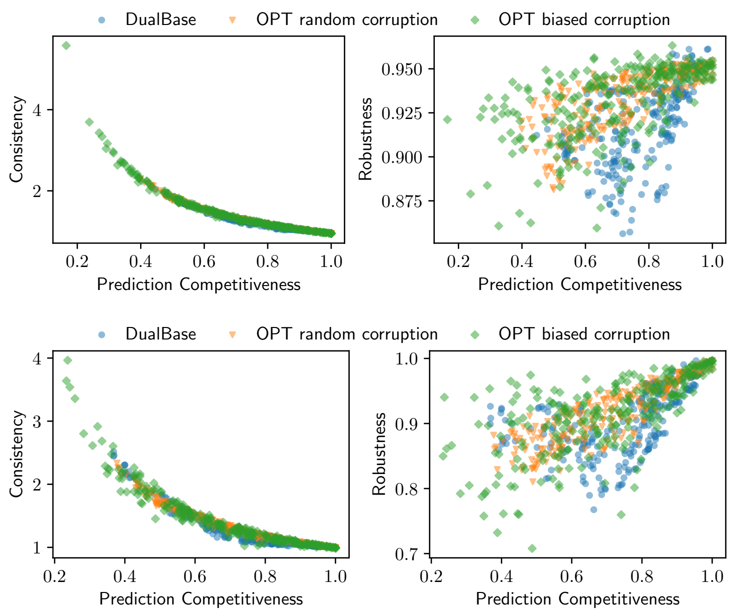

Figure 7: Performance for varying prediction quality with the data from Figure 6 (top) for $\alpha = 2$ (top) and $\alpha = 5$ (bottom).

C Further Experimental Results

C.1 Real-World Data

Description of the Yahoo Dataset: The original dataset contains impression allocations to 16268 advertisers throughout 123 days, each tagged with the advertiser that bought the impression and a set of keyphrases that categorize the user for whom the impression is displayed. Lavastida et al. (2021) then consider the 20 most common keyphrases and create an impression type for each non-empty subset thereof. Whenever an advertiser buys an impression with a certain set of keyphrases, we assume that all impression types that correspond to a superset of these keyphrases are relevant for this advertiser, and that it derives some constant value (say, 1) from this allocation. At the same time, the number of impressions we create from each impression type (i.e. the supply) is the number of impression allocations in the original dataset that show that the impression type is relevant for an advertiser. As such, we obtain around 2 million impressions. Lavastida et al. (2021) try multiple

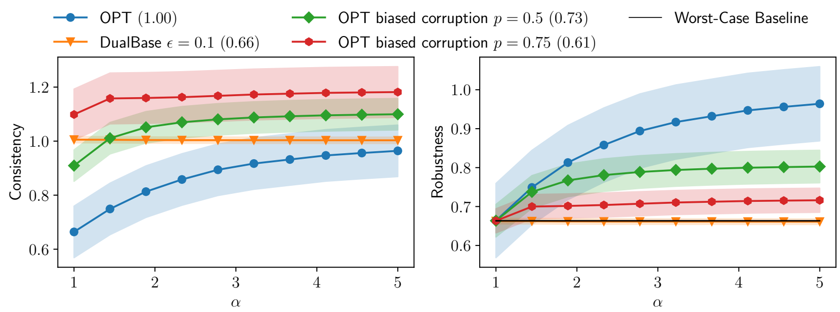

Figure 8: Performance on a worst-case instance with different predictors.

impression orders and budgets for the advertisers, but due to space constraints we restrict ourselves to display all impressions of a type at once, in supply-ascending order. We determine advertisers' budgets by allocating each impression to one of the advertisers with non-zero valuation uniformly at random and taking the number of allocated impressions at the end to be the advertiser's budget.

C.2 Synthetic Data

Results: Figure 5 shows consistency and robustness of our algorithm on synthetic data on $T = 2000$ impressions of 10 types and $k = 12$ advertisers, for a variation of predictions. The plot shows the performance for predictions from the optimum solutions (with varying corruption) and the dual base prediction. Our algorithm converges to almost perfect consistency and robustness for $\alpha = 10$, given the optimum solution. At the same time, we observe that the algorithm is robust against both random and biased corruption, as the robustness does not drop to the prediction's low competitiveness of around 0.7. Furthermore, the algorithm performs well in combination with the dual base prediction for $\epsilon = 0.1$ even though the first 200 impressions are clearly not representative of all synthetically generated impressions.

To investigate the our algorithm in conjunction with an easily available prediction, we also analyze the behavior of the dual base algorithm for different values of σ and ϵ in Figure 6. The performance of our algorithm under dual base predictions clearly improves for increasing values of σ as impressions become more evenly distributed across the day. Generally, sampling more impressions helps but dual base predictions may also lead to a drop in robustness, and more samples can even lead to a more adversarial prediction, as we explore further below. Yet, the robustness does still stays above the prediction's competitiveness in these cases.

Figure 7 shows consistency and robustness for different predictions with varying competitiveness on $\alpha \in \{2, 5\}$. We achieve this by varying the fraction $\epsilon \in [0, 1]$ of samples for the dual base algorithm and the corruption rate $p \in [0, 1]$ for random and biased corruptions. For $\alpha = 2$, the consistency exceeds 1 if the prediction is not very good (competitiveness below 0.9). The algorithm is not heavily influenced by a bad prediction since $\alpha = 2$ is low, so the total obtained value remains relatively constant. For $\alpha = 5$, the algorithm might however follow the bad choices of the prediction, so the competitiveness varies more. As expected, the average robustness decreases for increasing α, but the dual base prediction starts out with a much lower robustness than the corrupted predictions. The reason for that is that both the dual base algorithm and exponential averaging make their decisions based on the discounted gain. Our algorithm might therefore easily disregard a corrupted prediction as its discounted gain is low (or even negative), but the dual base prediction looks like a sensible choice. The dual base algorithm therefore manages to fool the algorithm for low α, while a biased corruption leads to the worst corruption for larger values of α.

Hard Instances: We consider the worst-case instance for the Display Ads problem described in Mehta et al. (2007). For k advertisers, we create impressions of types $r \in \{1, \dots, k\}$. An impression t of type r has zero value for the first $r - 1$ advertisers $w_{1,t} = \cdots = w_{r-1,t} = 0$ and value 1 for the following advertisers $w_{r,t} = \cdots = w_{k,t} = 1$. We first show all impressions of type 1, then all impressions of type 2, and so forth. The instance is difficult as the algorithm—not knowing about future impressions—has to allocate impressions of a type equally among advertisers that can

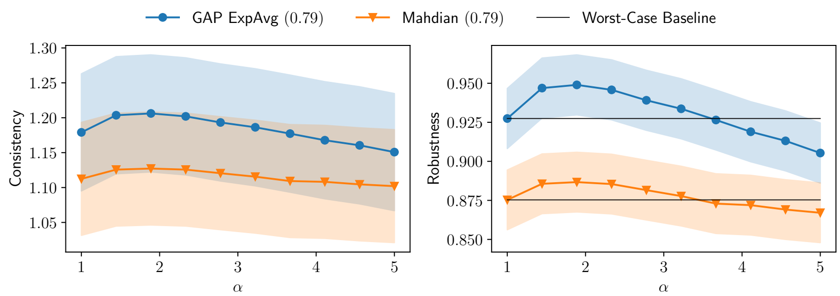

Figure 9: Performance on synthetic Ad Words instances, compared to the algorithm of Mehta et al. (2007). The black lines show the robustness of two worst-case algorithms without predictions: The algorithm due to Feldman et al. (2009a) which is the basis for our algorithm, and the algorithm of Mehta et al. (2007), which serves as a basis for the algorithm of Mahdian et al. (2007).

derive value from this impression type. As shown by Mehta et al. (2007), the competitiveness of the exponential averaging algorithm reaches $1 - \frac{1}{e}$ for $k \to \infty$ on this instance.

We evaluate the performance of our algorithm on this worst-case instance in Figure 8. Providing the optimum solution as prediction allows the algorithm to quickly ascend to a perfect robustness of 1. We also consider two (biased) corrupted versions of this prediction with $p \in \{50\%, 75\%\}$. In both cases, the algorithm still achieves a robustness above the competitiveness of the prediction. The dual base algorithm cannot deliver meaningful predictions as it only sees impressions of the first type, which are clearly not representative of the following impressions by construction.

C.3 Evaluation of GAP on an Ad Words Instance

With an algorithm for GAP, we can also solve AdWords instances. This allows us to compare our generalized algorithm to the algorithm of Mahdian et al. (2007) under the same predictions. In Figure 9, we run both algorithms on synthetic instances from Section C.2 with an optimum prediction and random corruption ($p = 0.5$). Both algorithms seem to have similar consistency, but our algorithm achieves a better robustness, due to a different choice of constants in the underlying algorithms.