A. Proof of Theorem 2

Without loss of generality, let’s assume that \( r_{\text{max}} \) is 1. This can be obtained by dividing all \( r_{\text{min}} \) by \( r_{\text{min}} \). \( r_{\text{min}} \) in the processed instance is then in fact the ratio of \( r_{\text{min}} \) and \( r_{\text{max}} \) of the original instance. Let \( \mathbf{x}_{i,a}^R \) and \( \beta_a^R \) denote the value of assignments and priority scores at the end of \( R \) iterations of Algorithm 2 (before the processing in the last step was done to handle over-allocated advertisers). And, let \( \mathbf{x}_{i,a}^M \) denote the feasible assignments obtained after the processing in the last step of the algorithm. Let \( \text{weight}(M) := \sum_{i,a \in \mathcal{E}} r_{i,a} \mathbf{x}_{i,a}^M \) denote the weight of this feasible fractional matching \( M \).

Now, initially, \( \beta_a = (1 + \epsilon)^{-R}. \) From Lemma 3, we have that for every \( a \), either \( \sum_{i \in N_a} \mathbf{x}_{i,a}^R \in [(1 + \epsilon)^{-2} C_a, (1 + \epsilon)^2 C_a] \), i.e., the advertiser budget constraint is approximately satisfied; or, we will have that \( \beta_a \) was continuously increased/decreased by \( (1 + \epsilon) \) factor for all \( R \) iterations, so that \( \beta_a^R \) is either 1 or \( (1 + \epsilon)^{-2R}. \) Let us call the first set of advertisers where the budget constraint is approximately satisfied as \( \mathcal{E}. \) For these advertisers, \( |C_a - \sum_{i \in N_a} \mathbf{x}_{i,a}| \leq 3\epsilon C_a \) for any \( \epsilon \leq 1. \) Also, \( \beta_a^R \geq (1 + \epsilon)^{-2R}. \) Among the second set, let \( \mathcal{O} \) be the set of advertisers \( a \in \mathcal{A} \) with \( \beta_a^R = (1 + \epsilon)^{-2R}. \) Here, \( \beta_a \) was continuously decreased in order to decrease the allocation, and these advertisers will be over-allocated in the end. For the remaining \( a \notin \mathcal{E}, a \notin \mathcal{O}, \) we have \( \beta_a^R = 1. \)

Using the upper bound from (10), and substituting the value of \( \beta_a^R \), we have that

\[
\text{OPT}_\lambda \leq \sum_{i,a} r_{i,a} \mathbf{x}_{i,a}^R + \sum_{a \in \mathcal{O}} 2R \epsilon \lambda (C_a - \sum_{i \in N_a} \mathbf{x}_{i,a}^R) + \sum_{a \in \mathcal{E}} 2R \epsilon \lambda (3\epsilon C_a) + \lambda \sum_{i,a} \mathbf{x}_{i,a}^R \log(1/\mathbf{x}_{i,a}^R)
\]

The terms for rest of the advertisers \( a \notin \mathcal{O}, a \notin \mathcal{E} \) do not appear in above because \( \log(1/\beta_a^R) = \log(1) = 0 \) for those \( a. \)

Next, we relate the above upper bound to the weight and entropy of the feasible fractional matching \( M. \) The matching \( M \) was created by removing \( \sum_{i \in N_a} \mathbf{x}_{i,a}^R - C_a \) edges from \( \mathbf{x}_{i,a}^R \) for every over-allocated advertiser \( a. \) Therefore, weight of matching \( M \) is at least

\[
\text{weight}(M) \geq \sum_{i,a} r_{i,a} \mathbf{x}_{i,a}^R - \sum_a \left( \sum_{i \in N_a} \mathbf{x}_{i,a}^R - C_a \right) + \sum_{a \in \mathcal{E}} 3\epsilon C_a - \sum_{a \in \mathcal{O}} \left( \sum_{i \in N_a} \mathbf{x}_{i,a}^R - C_a \right)
\]

Also, \( M \) retains all the edges allocated to \( a \in \mathcal{E} \) within a

\[
(1 + \epsilon)^2 \text{ factor, so that}
\]

\[
\text{weight}(M) \geq \frac{r_{\text{min}}}{(1 + \epsilon)^2} \sum_{a \in \mathcal{E}} C_a
\]

Substituting these observations in above upper bound for \( \text{OPT}_\lambda, \) along with

\[
R = \frac{1}{2\lambda} (1 + \lambda \log(N))
\]

(17)

Now, let,

\[
\epsilon = \frac{r_{\text{min}}}{8(2 + \lambda \log(N))} \delta
\]

so that the first term in the upper bound of (18) is at most \((1 + \frac{\delta}{2})\text{weight}(M). \) Now, we show that the next two terms approximate Entropy \((M) := \sum_{i,a} \mathbf{x}_{i,a}^R \log(1/\mathbf{x}_{i,a}^R). \) Recall that \( \mathbf{x}_{i,a}^M \) is the assignment of \( i,a \) in the fractional matching \( M, \) i.e., the assignment obtained after adjusting \( \mathbf{x}_{i,a}^R \) in the last step of Algorithm 2. This adjustment step ensures that \( \mathbf{x}_{i,a}^R \geq \mathbf{x}_{i,a}^M, \) and for any \( a \) with \( \mathbf{x}_{i,a}^R - \mathbf{x}_{i,a}^M > 0, \) we have \( \mathbf{x}_{i,a}^M \geq \frac{C_a}{|N_a|} \geq \frac{1}{N}. \) Therefore, it is easy to see that

\[
\sum_{i,a} \mathbf{x}_{i,a}^R \log(\frac{1}{\mathbf{x}_{i,a}^R}) - \sum_{i,a} \mathbf{x}_{i,a}^M \log(\frac{1}{\mathbf{x}_{i,a}^M}) \leq \sum_{i,a} (\mathbf{x}_{i,a}^R - \mathbf{x}_{i,a}^M) \log(N) = \sum_{a} (\sum_{i} \mathbf{x}_{i,a}^R - C_a)^+ \log(N).
\]

Then, using \( \sum_{i,a} \mathbf{x}_{i,a}^R - C_a \leq 3\epsilon C_a \) for \( a \in \mathcal{E}, \) relating \( \sum_{a \in \mathcal{E}} C_a \) to \( \text{weight}(M) \) as in (16), and substituting the choice of \( \epsilon, \) we obtain,

\[
\sum_{i,a} \mathbf{x}_{i,a}^R \log(\frac{1}{\mathbf{x}_{i,a}^R}) - \sum_{i,a} \mathbf{x}_{i,a}^M \log(\frac{1}{\mathbf{x}_{i,a}^M}) \leq \\sum_{a} (\sum_{i \in N_a} \mathbf{x}_{i,a}^R - C_a)^+ \log(N) \leq \sum_{a} (\sum_{i \in N_a} (\mathbf{x}_{i,a}^R - C_a)^+) \log(N) + \sum_{a \in \mathcal{O}} 3\epsilon C_a \log(N) \leq \sum_{a \in \mathcal{O}} (\sum_{i \in N_a} \mathbf{x}_{i,a}^R - C_a)^+ \log(N) + \frac{\delta}{2\lambda} \text{weight}(M)
\]
Substituting back in (18),

\[ \text{OPT} \leq (1 + \delta) \text{weight}(M) + \lambda \text{Entropy}(M) \]

Finally, from (17), substituting value of \( \epsilon \) from (19), we have the number of iterations

\[ R = \frac{1}{2e\lambda} (1 + \lambda \log(N)) \leq \frac{8}{r_{\text{min}}} \frac{(1 + \lambda \log(N))^2}{\lambda \delta} \]

Then, the theorem statement is obtained on substituting back \( r_{\text{min}}/r_{\text{max}} \) for \( r_{\text{min}} \).