A. Proof of Theorem 2

Without loss of generality, let's assume that $\mathbf{r}_{\max}$ is 1. This can be obtained by dividing all $\mathbf{r}_{i,a}$ by $\mathbf{r}_{min}$. $\mathbf{r}_{min}$ in the processed instance is then in fact the ratio of $\mathbf{r}_{\min}$ and $\mathbf{r}_{\max}$ of the original instance. Let $\mathbf{x}^R_{i,a}$ and $\boldsymbol{\beta}^R_a$ 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 $\mathbf{x}^M_{i,a}$ denote the feasible assignments obtained after the processing in the last step of the algorithm. Let weight$(M) := \sum_{i,a\in\mathbb{E}} \mathbf{r}_{i,a}\mathbf{x}^M_{i,a}$ denote the weight of this feasible fractional matching M.

Now, initially, $\boldsymbol{\beta}_a = (1+\epsilon)^{-R}$. From Lemma 3, we have that for every a, either $\sum_{i\in\mathbf{N}_a}\mathbf{x}^R_{i,a} \in [(1+\epsilon)^{-2}\mathbf{C}_a, (1+\epsilon)^2\mathbf{C}_a]$, i.e., the advertiser budget constraint is approximately satisfied; or, we will have that $\boldsymbol{\beta}_a$ was continuously increased/decreased by $(1+\epsilon)$ factor for all R iterations, so that $\boldsymbol{\beta}^R_a$ 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, $|\mathbf{C}_a - \sum_{i\in\mathbf{N}_a}\mathbf{x}_{i,a}| \le 3\epsilon\mathbf{C}_a$ for any $\epsilon \le 1$. Also, $\boldsymbol{\beta}^R_a \ge (1+\epsilon)^{-2R}$. Among the second set, let $\mathcal{O}$ be the set of advertisers $a \in \mathbb{A}$ with $\boldsymbol{\beta}^R_a = (1+\epsilon)^{-2R}$. Here, $\boldsymbol{\beta}_a$ 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 $\boldsymbol{\beta}^R_a = 1$.

Using the upper bound from (10), and substituting the value of $\boldsymbol{\beta}^R_a$, we have that

$$
\begin{aligned}
\mathrm{OPT}_\lambda \;\le\; & \sum_{i,a}\mathbf{r}_{i,a}\mathbf{x}^R_{i,a} + \sum_{a\in\mathcal{O}} 2R\epsilon\lambda(\mathbf{C}_a - \sum_{i\in\mathbf{N}_a}\mathbf{x}^R_{i,a}) \\
& + \sum_{a\in\mathcal{E}} 2R\epsilon\lambda(3\epsilon\mathbf{C}_a) + \lambda\sum_{i,a}\mathbf{x}^R_{i,a}\log(1/\mathbf{x}^R_{i,a})
\end{aligned}
$$

The terms for rest of the advertisers $a \notin \mathcal{O}, a \notin \mathcal{E}$ do not appear in above because $\log(1/\boldsymbol{\beta}^R_a) = \log(1) = 0$ 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 $\sum_{i\in\mathbf{N}_a}\mathbf{x}^R_{i,a} - \mathbf{C}_a$ edges from $\{\mathbf{x}^R_{i,a}\}$ for every over-allocated advertiser a. Therefore, weight of matching M is at least

$$
\begin{aligned}
\mathrm{weight}(M) \;\ge\; & \sum_{i,a}\mathbf{r}_{i,a}\mathbf{x}^R_{i,a} - \sum_a(\sum_{i\in\mathbf{N}_a}\mathbf{x}^R_{i,a} - \mathbf{C}_a)^+ \\
\ge\; & \sum_{i,a}\mathbf{r}_{i,a}\mathbf{x}^R_{i,a} \\
& - \sum_{a\in\mathcal{E}} 3\epsilon\mathbf{C}_a - \sum_{a\in\mathcal{O}}(\sum_{i\in N_a}\mathbf{x}^R_{i,a} - \mathbf{C}_a)^+
\end{aligned}
$$

Also, M retains all the edges allocated to $a \in \mathcal{E}$ within a $(1+\epsilon)^2$ factor, so that

$$
\mathrm{weight}(M) \ge \frac{\mathbf{r}_{\min}}{(1+\epsilon)^2}\sum_{a\in\mathcal{E}}\mathbf{C}_a \tag{16}
$$

Substituting these observations in above upper bound for OPT_λ, along with

$$
R = \frac{1}{2\epsilon\lambda}\left(1 + \lambda\log(\bar{N})\right), \tag{17}
$$

(where $\bar{N} = \max_a \frac{\mathbf{C}_a}{|\mathbf{N}_a|}$) we get

$$
\begin{aligned}
\mathrm{OPT}_\lambda \;\le\; & \mathrm{weight}(M)(1 + 3\epsilon(2+\lambda\log(\bar{N}))\frac{(1+\epsilon)^2}{\mathbf{r}_{\min}}) \\
& - \sum_{a\in\mathcal{O}}\lambda\log(\bar{N})(\sum_{i\in N_a}\mathbf{x}^R_{i,a} - C_a)^+ \\
& + \sum_{i,a}\lambda\mathbf{x}^R_{i,a}\log(\frac{1}{\mathbf{x}^R_{i,a}})
\end{aligned} \tag{18}
$$

Now, let

$$
\epsilon = \frac{\mathbf{r}_{\min}}{8(2+\lambda\log(\bar{N}))}\delta, \tag{19}
$$

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

$$
\begin{aligned}
& \sum_{i,a}\mathbf{x}^R_{i,a}\log(\frac{1}{\mathbf{x}^R_{i,a}}) - \sum_{i,a}\mathbf{x}^M_{i,a}\log(\frac{1}{\mathbf{x}^M_{i,a}}) \\
& \le (\sum_{i,a}\mathbf{x}^R_{i,a} - \mathbf{x}^M_{i,a})\log(\bar{N}) \\
& = \sum_a(\sum_i\mathbf{x}^R_{i,a} - \mathbf{C}_a)^+\log(\bar{N}).
\end{aligned}
$$

Then, using $|\sum_{i,a}\mathbf{x}^R_{i,a} - \mathbf{C}_a| \le 3\epsilon\mathbf{C}_a$ for $a \in \mathcal{E}$, relating $\sum_{a\in\mathcal{E}}\mathbf{C}_a$ to weight(M) as in (16), and substituting the choice of ϵ, we obtain,

$$
\begin{aligned}
& \sum_{i,a}\mathbf{x}^R_{i,a}\log(\frac{1}{\mathbf{x}^R_{i,a}}) - \sum_{i,a}\mathbf{x}^M_{i,a}\log(\frac{1}{\mathbf{x}^M_{i,a}}) \\
\le\; & \sum_a(\sum_{i\in\mathbf{N}_a}(\mathbf{x}^R_{i,a} - \mathbf{C}_a))^+\log(\bar{N}) \\
\le\; & \sum_{a\in\mathcal{O}}(\sum_{i\in\mathbf{N}_a}(\mathbf{x}^R_{i,a} - \mathbf{C}_a))^+\log(\bar{N}) + \sum_{a\in\mathcal{E}}3\epsilon\mathbf{C}_a\log(\bar{N}) \\
\le\; & \sum_{a\in\mathcal{O}}(\sum_{i\in\mathbf{N}_a}\mathbf{x}^R_{i,a} - \mathbf{C}_a)^+\log(\bar{N}) + \frac{\delta}{2\lambda}\mathrm{weight}(M)
\end{aligned}
$$

Substituting back in (18),

$$\text{OPT}_\lambda \quad \leq \quad (1+\delta)\text{weight}(M) + \lambda\text{Entropy}(M)$$

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

$$R = \frac{1}{2\epsilon\lambda}\left(1 + \lambda\log(\bar{N})\right) \leq \frac{8}{\mathbf{r}_{min}}\frac{(1+\lambda\log(\bar{N}))^2}{\lambda\delta}$$

Then, the theorem statement is obtained on substituting back $\mathbf{r}_{\min}/\mathbf{r}_{\max}$ for $\mathbf{r}_{\min}$.