Proceedings of Machine Learning Research 129:1–26, 2020 ACML 2020

Exact Passive-Aggressive Algorithms for Multiclass Classification Using Bandit Feedbacks

Maanik Arora MAANIK.ARORA@RESEARCH.IIIT.AC.IN
Machine Learning Lab, Kohli Center on Intelligent Systems, IIIT Hyderabad, India

Naresh Manwani NARESH.MANWANI@IIIT.AC.IN
Machine Learning Lab, Kohli Center on Intelligent Systems, IIIT Hyderabad, India

1. EPABF Step Size λ_r^t Derivation

EPABF updates the parameters by solving the following optimization problem.

$$\begin{aligned}\mathbf{w}_1^{t+1}\dots\mathbf{w}_K^{t+1} &= \arg\min_{\mathbf{w}_1,\dots,\mathbf{w}_K}\frac{1}{2}\sum_{v=1}^{K}\|\mathbf{w}_v-\mathbf{w}_v^t\|^2 \;\; s.t. \;\; \sum_{r=1}^{K}\widetilde{l}_r = 0\\ &= \arg\min_{\mathbf{w}_1,\dots,\mathbf{w}_K}\frac{1}{2}\sum_{v=1}^{K}\|\mathbf{w}_v-\mathbf{w}_v^t\|^2 \;\; s.t. \;\; a_t\mathbf{w}_{\widetilde{y}^t}\cdot\mathbf{x}^t-\mathbf{w}_r\cdot\mathbf{x}^t\geq 1,\ \forall r\in[K]\end{aligned}$$

This is a quadratic optimization problem with K linear constraints. KKT conditions (Bertsekas (1999)) for optimal solution are as follows.

$$\begin{cases}\mathbf{w}_r=\mathbf{w}_r^t-\lambda_r^t\mathbf{x}^t+a_t\mathbb{I}_{\{\widetilde{y}^t=r\}}\sum_r\lambda_r^t\mathbf{x}^t, & \forall r\\ \lambda_r^t\left(1+\mathbf{w}_r\cdot\mathbf{x}^t-a_t\mathbf{w}_{\widetilde{y}^t}\cdot\mathbf{x}^t\right)=0, & \forall r\\ \left(1+\mathbf{w}_r\cdot\mathbf{x}^t-a_t\mathbf{w}_{\widetilde{y}^t}\cdot\mathbf{x}^t\right)\leq 0;\ \lambda_r^t\geq 0, & \forall r\end{cases}\tag{1}$$

where the Lagrange multipliers, λ_r^t, turn out to be the step sizes of the updates for each class. Using these equations, the weight vector could potentially be updated for every class. To complete the update rule, we need to determine the values of λ_r^t. Those with positive λ_r^t should satisfy

$$a_t\mathbf{w}_{\widetilde{y}^t}^{t+1}\cdot\mathbf{x}^t-\mathbf{w}_r^{t+1}\cdot\mathbf{x}^t=1\tag{2}$$

The classes for which $\lambda_r^t>0$ are called support classes. Let the support class set is denoted by S^t. We assume that S^t is known. Plugging values of $\mathbf{w}_{\widetilde{y}^t}^t$ and $\mathbf{w}_r^t$ in Eq. (2) we get the following.

$$a_t^2\sum_i\lambda_i^t-a_t\lambda_{\widetilde{y}^t}^t+\lambda_r^t=\frac{1+\mathbf{w}_r^t\cdot\mathbf{x}^t-a_t\mathbf{w}_{\widetilde{y}^t}^t\cdot\mathbf{x}^t}{\|\mathbf{x}^t\|^2}=\frac{\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2}\tag{3}$$

Summing the above equation over all $r\in S^t$, we get

$$(a_t^2|S^t|+1)\sum_{r\in S^t}\lambda_r^t-a_t|S^t|\lambda_{\widetilde{y}^t}^t=\frac{\sum_{r\in S^t}\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2}\tag{4}$$

We have two cases.

- **Case 1: $\widetilde{y}^t \in S^t$:** In this case we have, Taking $r = \widetilde{y}^t$ in Eq.(3), we get,

$$a_t^2 \sum_{r \in S^t} \lambda_i^t - a_t \lambda_{\widetilde{y}^t}^t + \lambda_{\widetilde{y}^t}^t = \frac{\widetilde{l}_{\widetilde{y}^t}^t}{\|\mathbf{x}^t\|^2} \tag{5}$$

Using Eq.(4) and (5), we find the values of $\lambda_{\widetilde{y}^t}^t$ and $\sum_{r \in S^t} \lambda_r^t$ as follows.

$$\lambda_{\widetilde{y}^t}^t = \frac{1 + |S^t| a_t^2}{(1 + |S^t| a_t^2 - a_t)\|\mathbf{x}^t\|^2} \left(\widetilde{l}_{\widetilde{y}^t}^t - \frac{a_t^2}{1 + |S^t| a_t^2} \sum_{r \in S^t} \widetilde{l}_r^t \right)$$
$$\sum_{r \in S^t} \lambda_r^t = \frac{(1 - a_t) \sum_{r \in S^t} \widetilde{l}_r^t + a_t |S^t| \widetilde{l}_{\widetilde{y}^t}^t}{(1 + |S^t| a_t^2 - a_t)\|\mathbf{x}^t\|^2}$$

Putting the values of $\lambda_{\widetilde{y}^t}^t$ and $\sum_{r \in S^t} \lambda_r^t$ in Eq.(3), we get,

$$\lambda_r^t = \frac{1}{\|\mathbf{x}^t\|^2} \left(\widetilde{l}_r^t + \frac{a_t \widetilde{l}_{\widetilde{y}^t}^t}{1 + |S^t| a_t^2 - a_t} - \frac{a_t^2 \sum_{i \in S^t} \widetilde{l}_i^t}{1 + |S^t| a_t^2 - a_t} \right) \tag{6}$$

- **Case 2:** $\widetilde{y}^t \notin S^t$: In this case $\lambda_{\widetilde{y}^t}^t = 0$. Using Eq.(3) and (4), we will get,

$$\sum_{r \in S^t} \lambda_r^t = \frac{1}{(\|\mathbf{x}^t\|^2)(a_t^2 |S^t| + 1)} \sum_{r \in S^t} \widetilde{l}_r^t$$

$$\lambda_r^t = \frac{1}{\|\mathbf{x}^t\|^2} \left(\widetilde{l}_r^t - \frac{a_t^2}{a_t^2 |S^t| + 1} \sum_{i \in S^t} \widetilde{l}_i^t \right) \tag{7}$$

2. Proof of Theorem 1

Proof We will have 2 cases:

- **Case 1:** If $\widetilde{y}^t \in S^t$
  Using the KKT conditions, we see that for any $r \notin S^t$, we have,

$$a_t(\mathbf{w}_{\widetilde{y}^t}^t - \lambda_{\widetilde{y}^t}^t \mathbf{x}^t + a_t \sum_i \lambda_i^t \mathbf{x}^t) \cdot \mathbf{x}^t - (\mathbf{w}_r^t - \lambda_r^t \mathbf{x}^t) \cdot \mathbf{x}^t \geq 1$$

Since $\lambda_r^t = 0$ for $r \notin S^t$, the above equation reduces to,

$$a_t^2 \sum_i \lambda_i^t - a_t \lambda_{\widetilde{y}^t}^t \geq \frac{1 + \mathbf{w}_r^t \cdot \mathbf{x}^t - a_t \mathbf{w}_{\widetilde{y}^t}^t \cdot \mathbf{x}^t}{\|\mathbf{x}^t\|^2} = \frac{\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2} \tag{8}$$

We know that,

$$\lambda_{\widetilde{y}^t}^t = \frac{1+|S^t|a_t^2}{(1+|S^t|a_t^2-a_t)\|\mathbf{x}^t\|^2}\left(\widetilde{l}_{\widetilde{y}^t}^t - \frac{a_t^2}{1+|S^t|a_t^2}\sum_{r\in S^t}\widetilde{l}_r^t\right) \tag{9}$$

$$\sum_{r\in S^t}\lambda_r^t = \frac{1}{(1+|S^t|a_t^2-a_t)\|\mathbf{x}^t\|^2}\left((1-a_t)\sum_{r\in S^t}\widetilde{l}_r^t + a_t|S^t|\widetilde{l}_{\widetilde{y}^t}^t\right) \tag{10}$$

Using Eq.(9) and Eq.(10), we can rewrite the equation Eq.(8) as,

$$\frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{1+|S^t|a_t^2-a_t} \geq \widetilde{l}_r^t + \frac{a_t\widetilde{l}_{\widetilde{y}^t}^t}{1+|S^t|a_t^2-a_t},\ \forall r\notin S^t \tag{11}$$

Also we have,

$$\lambda_r^t = \frac{1}{\|\mathbf{x}^t\|^2}\left(\widetilde{l}_r^t + \frac{a_t\widetilde{l}_{\widetilde{y}^t}^t}{1+|S^t|a_t^2-a_t} - \frac{a_t^2\sum_{i\in S^t}\widetilde{l}_i^t}{1+|S^t|a_t^2-a_t}\right) \tag{12}$$

To get support class, λ_r^t should be positive, so by Eq.(12), we get

$$\frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{1+|S^t|a_t^2-a_t} < \widetilde{l}_r^t + \frac{a_t\widetilde{l}_{\widetilde{y}^t}^t}{1+|S^t|a_t^2-a_t},\ \forall r\in S^t \tag{13}$$

Let $\sigma(k)$ be the k-th class when sorted in descending order of $\widetilde{l}_r^t$.
(Sufficiency) Assume that $\widetilde{l}_{\sigma(k)}^t$ satisfies the theorem, then we have,

$$\begin{aligned}
&\sum_{j=1}^{k-1}\widetilde{l}_{\sigma(j)}^t < \frac{1+(k-1)a_t^2-a_t}{a_t^2}\widetilde{l}_{\sigma(k)}^t + \frac{1}{a_t}\widetilde{l}_{\widetilde{y}^t}^t \\
\Rightarrow &\sum_{j=1}^{|S^t|}\widetilde{l}_{\sigma(j)}^t < \frac{1+|S^t|a_t^2-a_t}{a_t^2}\widetilde{l}_{\sigma(k)} + \frac{1}{a_t}\widetilde{l}_{\widetilde{y}^t}^t \\
\Rightarrow &\frac{a_t^2}{1+|S^t|a_t^2-a_t}\sum_{j=1}^{|S^t|}\widetilde{l}_{\sigma(j)}^t < \widetilde{l}_{\sigma(k)}^t + \frac{a_t}{1+|S^t|a_t^2-a_t}\widetilde{l}_{\widetilde{y}^t}^t
\end{aligned}$$

The second inequality is justified as the losses $\widetilde{l}_{\sigma(j)}^t$ are in decreasing order. This means $\sigma(k)$ corresponds to a label of some support classes (Eq.(13)).
(Necessity) Assume that $\widetilde{l}_{\sigma(k)}$ does not satisfy theorem, then

$$\sum_{j=1}^{k-1}\widetilde{l}_{\sigma(j)}^t \geq \frac{1+(k-1)a_t^2-a_t}{a_t^2}\widetilde{l}_{\sigma(k)}^t + \frac{1}{a_t}\widetilde{l}_{\widetilde{y}^t}^t$$

$$\sum_{j=1}^{k}\widetilde{l}_{\sigma(j)}^t \geq \frac{1+(k)a_t^2-a_t}{a_t^2}\widetilde{l}_{\sigma(k)}^t + \frac{1}{a_t}\widetilde{l}_{\widetilde{y}^t}^t$$

$$\frac{a_t^2}{1+ka_t^2-a_t}\sum_{j=1}^{k}\widetilde{l}^t_{\sigma(j)}-\frac{a_t}{1+ka_t^2-a_t}\widetilde{l}^t_{\widetilde{y}^t}\geq\widetilde{l}^t_{\sigma(k)}\geq\widetilde{l}^t_{\sigma(k+1)}$$

Therefore, any j larger than $\sigma(k)$ does not satisfy Eq.(13). It means $|S^t| < k$ and thus $\sigma(k)$ does not correspond to a label of a support class.

- **Case 2:** If $\widetilde{y}^t \notin S^t$
  Using the KKT conditions, we see that for any $r \notin S^t$, we have,

$$a_t(\mathbf{w}^t_{\widetilde{y}^t}+a_t\sum_i\lambda^t_i\mathbf{x}^t)\cdot\mathbf{x}^t-(\mathbf{w}^t_r-\lambda^t_r\mathbf{x}^t)\cdot\mathbf{x}^t\geq 1$$

Since $\lambda^t_r = 0$ for $r \notin S^t$, the above equation reduces to,

$$a_t^2\sum_i\lambda^t_i\geq\frac{1+\mathbf{w}^t_r\cdot\mathbf{x}^t-a_t\mathbf{w}^t_{\widetilde{y}^t}\cdot\mathbf{x}^t}{\|\mathbf{x}^t\|^2}=\frac{\widetilde{l}^t_r}{\|\mathbf{x}^t\|^2}$$

We know that,

$$\sum_{r\in S^t}\lambda^t_r=\frac{1}{(1+|S^t|a_t^2)\|\mathbf{x}^t\|^2}\left(\sum_{r\in S^t}\widetilde{l}^t_r\right)\tag{14}$$

Using Eq.(14), we can rewrite this equation as,

$$\frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_{\sigma(j)}}{1+|S^t|a_t^2}\geq\widetilde{l}^t_{\sigma(r)},\ \forall r\notin S^t\tag{15}$$

Also we have,

$$\lambda^t_r=\frac{1}{\|\mathbf{x}^t\|^2}\left(\widetilde{l}^t_r-\frac{a_t^2\sum_{i\in S^t}\widetilde{l}^t_i}{1+|S^t|a_t^2}\right)\tag{16}$$

To get support class, λ^t_r should be positive, so by Eq.(16), we get

$$\frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_{\sigma(j)}}{1+|S^t|a_t^2}<\widetilde{l}^t_{\sigma(r)},\ \forall r\in S^t\tag{17}$$

Let $\sigma(k)$ be the k-th class when sorted in descending order of $\widetilde{l}^t_r$.
(Sufficiency) Assume that $\widetilde{l}^t_{\sigma(k)}$ satisfies the theorem, then we have,

$$\sum_{j=1}^{k-1}\widetilde{l}^t_{\sigma(j)}<\frac{1+(k-1)a_t^2}{a_t^2}\widetilde{l}^t_{\sigma(k)}$$

$$\Rightarrow\sum_{j=1}^{|S^t|}\widetilde{l}^t_{\sigma(j)}<\frac{1+|S^t|a_t^2}{a_t^2}\widetilde{l}^t_{\sigma(k)}$$

$$\Rightarrow\frac{a_t^2}{1+|S^t|a_t^2}\sum_{j=1}^{|S^t|}\widetilde{l}^t_{\sigma(j)}<\widetilde{l}^t_{\sigma(k)}$$

The second inequality is justified as the losses $\widetilde{l}^t_{\sigma(j)}$ are in decreasing order. This means $\sigma(k)$ corresponds to a label of some support classes Eq. (17).
(Necessity) Assume that $\widetilde{l}^t_{\sigma(k)}$ does not satisfy theorem, then

$$\sum_{j=1}^{k-1} \widetilde{l}^t_{\sigma(j)} \geq \frac{1+(k-1)a_t^2}{a_t^2}\widetilde{l}^t_{\sigma(k)}$$

$$\sum_{j=1}^{k} \widetilde{l}^t_{\sigma(j)} \geq \frac{1+(k)a_t^2}{a_t^2}\widetilde{l}^t_{\sigma(k)}$$

$$\frac{a_t^2}{1+ka_t^2}\sum_{j=1}^{k} \widetilde{l}^t_{\sigma(j)} \geq \widetilde{l}^t_{\sigma(k)} \geq \widetilde{l}^t_{\sigma(k+1)}$$

Therefore, any j larger than $\sigma(k)$ does not satisfy Eq. (17). It means $|S^t| < k$ and thus $\sigma(k)$ does not correspond to a label of a support class.

■

3. Proof of Theorem 2: EPABF bound

Proof

- **Case 1:** If $\widetilde{y}^t \in S^t$
  We define Δ_t as the following,

$$\Delta_t = \sum_{v=1}^{K} \|\mathbf{w}_v^t - \mathbf{u}_v\|^2 - \sum_{v=1}^{K} \|\mathbf{w}_v^{t+1} - \mathbf{u}_v\|^2$$

  We can write it as,

$$\Delta_t = \sum_{v=1}^{K} \|\mathbf{w}_v^t - \mathbf{u}_v\|^2 - \sum_{v=1}^{K} \|\mathbf{w}_v^t - \mathbf{u}_v - \lambda_v^t \mathbf{x}^t + \mathbb{I}_{\{\widetilde{y}^t = v\}} \frac{\mathbb{I}_{\{\widetilde{y}^t = y^t\}}}{P(\widetilde{y}_t)} \sum_i \lambda_i^t \mathbf{x}^t\|^2$$
$$= 2\sum_{v=1}^{K}\left(\lambda_v^t - \mathbb{I}_{\{\widetilde{y}^t = v\}} \frac{\mathbb{I}_{\{\widetilde{y}^t = y^t\}}}{P(\widetilde{y}^t)} \sum_i \lambda_i^t\right)(\mathbf{w}_v^t - \mathbf{u}_v)\cdot \mathbf{x}^t - \sum_{v=1}^{K}\left(\lambda_v^t - \mathbb{I}_{\{\widetilde{y}^t = v\}} \frac{\mathbb{I}_{\{\widetilde{y}^t = y^t\}}}{P(\widetilde{y}^t)} \sum_i \lambda_i^t\right)^2 \|\mathbf{x}^t\|^2$$
$$= 2\sum_{v\in S} \lambda_v^t \left(\mathbf{w}_v^t - \mathbf{u}_v\right).\mathbf{x}^t - 2a_t \sum_i \lambda_i \left(\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_{\widetilde{y}^t}\right)\cdot \mathbf{x}_t - \sum_{v\in S^t} \left(\lambda_v^t\right)^2 \|\mathbf{x}_t\|^2$$
$$- a_t^2 \left(\sum_v \lambda_v^t\right)^2 \|\mathbf{x}_t\|^2 + 2a_t \lambda_{\widetilde{y}^t}^t \sum_v \lambda_v^t \|\mathbf{x}_t\|^2$$

  We know that $\widetilde{l}_v^t = 1 - a_t \mathbf{w}_{\widetilde{y}^t}^t \cdot \mathbf{x}^t + \mathbf{w}_v^t \cdot \mathbf{x}^t,\ \forall v \in S^t$ and $\widetilde{l}_v^{*t} \geq 1 - a_t \mathbf{u}_{\widetilde{y}^t} \cdot \mathbf{x}^t + \mathbf{u}_v \cdot \mathbf{x}^t,\ \forall v \in [K]$. Thus, $(\mathbf{w}_v^t - \mathbf{u}_v)\cdot \mathbf{x}^t \geq \widetilde{l}_v^t - \widetilde{l}_v^{*t} + a(\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_{\widetilde{y}^t})\cdot \mathbf{x}^t$. So, we get the following.

$$\Delta_t \geq 2\sum_{v\in S^t} \lambda_v^t \left(\widetilde{l}_v^t - \widetilde{l}_v^{t*}\right) - \sum_{v\in S^t} \left(\lambda_v^t\right)^2 \|\mathbf{x}_t\|^2 - a_t^2 \left(\sum_v \lambda_v^t\right)^2 \|\mathbf{x}_t\|^2 + 2a_t \lambda_{\widetilde{y}^t}^t \sum_v \lambda_v^t \|\mathbf{x}_t\|^2$$

We use the following in the above equation.

$$\lambda^t_{\widetilde{y}^t} = \frac{1+|S^t|a_t^2}{(1+|S^t|a_t^2-a_t)\|\mathbf{x}^t\|^2}\left(\widetilde{l}^t_{\widetilde{y}^t} - \frac{a_t^2}{1+|S^t|a_t^2}\sum_{r\in S^t}\widetilde{l}^t_r\right)$$

$$\sum_{r\in S^t}\lambda^t_r = \frac{(1-a_t)\sum_{r\in S^t}\widetilde{l}^t_r + a_t|S^t|\widetilde{l}^t_{\widetilde{y}^t}}{(1+|S^t|a_t^2-a_t)\|\mathbf{x}^t\|^2}$$

$$\lambda^t_r = \frac{1}{\|\mathbf{x}^t\|^2}\left(\widetilde{l}^t_r + \frac{a_t\widetilde{l}^t_{\widetilde{y}^t}}{1+|S^t|a_t^2-a_t} - \frac{a_t^2\sum_{i\in S^t}\widetilde{l}^t_i}{1+|S^t|a_t^2-a_t}\right)$$

Using the above values to get,

$$\begin{aligned}\Delta_t \geq{}& \frac{1}{\|\mathbf{x}^t\|^2}\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2 - \frac{2}{\|\mathbf{x}^t\|^2}\sum_{v\in S^t}\widetilde{l}^t_v\widetilde{l}^{*t}_v - 2\left(\frac{a_t}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)}\right)\widetilde{l}^t_{\widetilde{y}^t}\sum_{v\in S^t}\widetilde{l}^{*t}_v\\
&+2\left(\frac{a_t^2}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)}\right)\sum_{v\in S^t}\widetilde{l}^t_v\sum_{v\in S^t}\widetilde{l}^{*t}_v + \left(\frac{a_t^2|S^t|(a_t^2|S^t|+1)}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)^2}\right)\left(\widetilde{l}^t_{\widetilde{y}^t}\right)^2\\
&+\left(\frac{a_t^2(a_t^2|S^t|+1-a_t^2)}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)^2}\right)\left(\sum_{v\in S^t}\widetilde{l}^t_v\right)^2 + \left(\frac{2a_t(a_t^4|S^t|^2+1-a_t+a_t^3|S^t|)}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)^2}\right)\widetilde{l}^t_{\widetilde{y}^t}\sum_{v\in S^t}\widetilde{l}^t_v\end{aligned}$$

We observe the following.

$$2\left(\frac{a_t^2}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)}\right)\sum_{v\in S^t}\widetilde{l}^t_v\sum_{v\in S^t}\widetilde{l}^{*t}_v \geq 0$$

$$\frac{a_t^2(a_t^2|S^t|^2-|S^t|+2)}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)^2}\left(\widetilde{l}^t_{\widetilde{y}^t}\right)^2 \geq 0$$

$$\frac{a_t^2(a_t^2|S^t|+1-a_t^2)}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)^2}\left(\sum_{v\in S^t}\widetilde{l}^t_v\right)^2 \geq 0$$

$$\left(\frac{2a_t(a_t^4|S^t|^2+1-a_t+a_t^3|S^t|)}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)^2}\right)\widetilde{l}^t_{\widetilde{y}^t}\sum_{v\in S^t}\widetilde{l}^t_v \geq 0$$

Using these, we get,

$$\Delta_t \geq \frac{1}{\|\mathbf{x}^t\|^2}\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2 - \frac{2}{\|\mathbf{x}^t\|^2}\sum_{v\in S^t}\widetilde{l}^t_v\widetilde{l}^{*t}_v - 2\left(\frac{a_t}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)}\right)\widetilde{l}^t_{\widetilde{y}^t}\sum_{v\in S^t}\widetilde{l}^{*t}_v.$$

Also,

$$2\left(\frac{a_t}{\|\mathbf{x}^t\|^2(a_t^2|S^t|+1-a_t)}\right)\widetilde{l}^t_{\widetilde{y}^t}\sum_{v\in S^t}\widetilde{l}^{*t}_v \leq \frac{2}{\|\mathbf{x}^t\|^2}\widetilde{l}^t_{\widetilde{y}^t}\sum_{v\in S^t}\widetilde{l}^{*t}_v.$$

Using the above approximation, the expression becomes,

$$\begin{aligned}\Delta_t &\geq \frac{1}{\|\mathbf{x}^t\|^2}\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2 - \frac{2}{\|\mathbf{x}^t\|^2}\sum_{v\in S^t}\widetilde{l}^t_v\widetilde{l}^{*t}_v - \frac{2}{\|\mathbf{x}^t\|^2}\widetilde{l}^t_{\widetilde{y}^t}\sum_{v\in S^t}\widetilde{l}^{*t}_v\\
&\geq \frac{1}{\|\mathbf{x}^t\|^2}\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2 - \frac{4}{\|\mathbf{x}^t\|^2}\sum_{v\in S^t}\widetilde{l}^t_v\widetilde{l}^{*t}_v.\end{aligned}$$

Taking expectation on both sides with respect to $\widetilde{y}^t$.

$$\mathbb{E}[\Delta_t] \geq \frac{1}{\|\mathbf{x}^t\|^2}\mathbb{E}[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2] - \frac{4}{\|\mathbf{x}^t\|^2}\mathbb{E}[\sum_{v\in S^t}\widetilde{l}_v^t\widetilde{l}_v^{*t}]$$

But, $\mathbb{E}[\sum_{v\in S^t}\widetilde{l}_v^t\widetilde{l}_v^{*t}] \leq \mathbb{E}[\sqrt{\sum_{v\in S^t}(\widetilde{l}_v^t)^2\sum_{v\in S^t}(\widetilde{l}_v^{*t})^2}$. Now using Cauchy Schwarz inequality $\mathbb{E}\left[xy\right] \leq \sqrt{\mathbb{E}\left[x^2\right]\mathbb{E}\left[y^2\right]}$, we get, $\mathbb{E}[\sum_{v\in S^t}\widetilde{l}_v^t\widetilde{l}_v^{*t}] \leq \sqrt{\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}_v^t)^2]}\sqrt{\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}_v^{*t})^2]}$. Using this and $\|\mathbf{x}^t\| \leq R$, we get the following.

$$\mathbb{E}[\Delta_t] \geq \frac{1}{R^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right] - \frac{4}{R^2}\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right]}\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]} \tag{18}$$

- **Case 2:** If $\widetilde{y}^t \notin S^t$
  We define Δ_t as the following,

$$\Delta_t = \sum_{v=1}^{K}\|\mathbf{w}_v^t - \mathbf{u}_v\|^2 - \sum_{v=1}^{K}\|\mathbf{w}_v^{t+1} - \mathbf{u}_v\|^2$$

We can write it as,

$$\begin{aligned}
\Delta_t &= \sum_{v\neq\widetilde{y}^t}\|\mathbf{w}_v^t - \mathbf{u}_v\|^2 + \|\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_{\widetilde{y}^t}\|^2 - \sum_{v\neq\widetilde{y}^t}\|\mathbf{w}_v^{t+1} - \mathbf{u}_v\|^2 - \|\mathbf{w}_{\widetilde{y}^t}^{t+1} - \mathbf{u}_{\widetilde{y}^t}\|^2 \\
&= \sum_{v\neq\widetilde{y}^t}\|\mathbf{w}_v^t - \mathbf{u}_v\|^2 + \|\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_{\widetilde{y}^t}\|^2 - \sum_{v\neq\widetilde{y}^t}\|\mathbf{w}_v^t - \lambda_v^t\mathbf{x}^t - \mathbf{u}_v\|^2 - \|\mathbf{w}_{\widetilde{y}^t}^t + a\sum_{i\neq\widetilde{y}^t}\lambda_i^t\mathbf{x}^t - \mathbf{u}_{\widetilde{y}^t}\|^2 \\
&= \sum_{v\neq\widetilde{y}^t}\|\mathbf{w}_v^t - \mathbf{u}_v\|^2 + \|\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_{\widetilde{y}^t}\|^2 - \sum_{v\neq\widetilde{y}^t}\|\mathbf{w}_v^t - \mathbf{u}_v\|^2 - \sum_{v\neq\widetilde{y}^t}\left(\lambda_v^t\right)^2\|\mathbf{x}^t\|^2 + 2\sum_{v\neq\widetilde{y}^t}\lambda_v^t\left(\mathbf{w}_v^t - \mathbf{u}_v\right)\cdot\mathbf{x}^t \\
&\quad - \|\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_{\widetilde{y}^t}\|^2 - a^2\left(\sum_{i\neq\widetilde{y}^t}\lambda_i^t\right)^2\|\mathbf{x}^t\|^2 - 2a\sum_{i\neq\widetilde{y}^t}\lambda_i^t\left(\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_{\widetilde{y}^t}\right)\cdot\mathbf{x}^t \\
&= 2\sum_{v\neq\widetilde{y}^t}\lambda_v^t\left(\mathbf{w}_v^t - \mathbf{u}_v\right)\cdot\mathbf{x}^t - 2a\sum_{i\neq\widetilde{y}^t}\left(\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_{\widetilde{y}^t}\right)\cdot\mathbf{x}^t - \sum_{v\neq\widetilde{y}^t}\left(\lambda_v^t\right)^2\|\mathbf{x}^t\|^2 - a^2\left(\sum_{i\neq\widetilde{y}^t}\lambda_i^t\right)^2\|\mathbf{x}^t\|^2 \\
&= 2\sum_{v\neq\widetilde{y}^t}\lambda_v^t\left(\mathbf{w}_v^t - a\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_v + a\mathbf{u}_{\widetilde{y}^t}\right)\cdot\mathbf{x}^t - \sum_{v\neq\widetilde{y}^t}(\lambda_v^t)^2\|\mathbf{x}^t\|^2 - a^2\left(\sum_{v\neq\widetilde{y}^t}\lambda_v^t\right)^2\|\mathbf{x}^t\|^2
\end{aligned}$$

We know that $\widetilde{l}_v^t = 1 - a\mathbb{I}_{\{\widetilde{y}^t=v\}} - a\mathbf{w}_{\widetilde{y}^t}^t\cdot\mathbf{x}^t + \mathbf{w}_v^t\cdot\mathbf{x}^t$, $\forall v \in S^t$ and $\widetilde{l}_v^{*t} \geq 1 - a\mathbb{I}_{\{\widetilde{y}^t=v\}} - a\mathbf{u}_{\widetilde{y}^t}\cdot\mathbf{x}^t + \mathbf{u}_v\cdot\mathbf{x}^t$, $\forall v \in [K]$. Thus, $(\mathbf{w}_v^t - \mathbf{u}_v)\cdot\mathbf{x}^t - a(\mathbf{w}_{\widetilde{y}^t}^t - \mathbf{u}_{\widetilde{y}^t})\cdot\mathbf{x}^t \geq \widetilde{l}_v^t - \widetilde{l}_v^{*t}$. Thus, we get the following.

$$\Delta_t \geq 2\sum_{v\neq\widetilde{y}^t}\lambda_v^t\left(\widetilde{l}_v^t - \widetilde{l}_v^{*t}\right) - \sum_{v\neq\widetilde{y}^t}(\lambda_v^t)^2\|\mathbf{x}^t\|^2 - a^2\left(\sum_{v\neq\widetilde{y}^t}\lambda_v^t\right)^2\|\mathbf{x}^t\|^2$$

Let S^t be the set of support classes in t^{th} trial. For all $v \in S^t$, we can see that

$$\|\mathbf{x}^t\|^2 \lambda_v^t = \widetilde{l}_v^t - \frac{a^2}{a^2|S^t|+1} \sum_{j \in S^t} \widetilde{l}_j^t$$

$$\|\mathbf{x}^t\|^2 \sum_{v \in S^t} \lambda_v^t = \frac{1}{(a^2|S^t|+1)} \sum_{v \in S^t} \widetilde{l}_v^t$$

We can ignore all $v \notin S^t$ in the sum for the last representation of Δ_t and substituting the above values in that to get,

$$\begin{aligned}
\Delta_t &\geq \sum_{v \in S} \lambda_v^t \left(2\widetilde{l}_v^t - 2\widetilde{l}_v^{*t} - \|\mathbf{x}^t\|^2 \lambda_v^t\right) - \|\mathbf{x}^t\|^2 a^2 \left(\sum_{v \in S} \lambda_v^t\right)^2 \\
&= \frac{1}{\|\mathbf{x}^t\|^2} \sum_{v \in S^t} \left(\widetilde{l}_v^t - \frac{a^2}{a^2|S^t|+1} \sum_{j \in S^t} \widetilde{l}_j^t\right) \left(\widetilde{l}_v^t - 2\widetilde{l}_v^{*t} + \frac{a^2}{a^2|S^t|+1} \sum_{j \in S^t} \widetilde{l}_j^t\right) - \frac{1}{\|\mathbf{x}^t\|^2} \frac{a^2}{(a^2|S^t|+1)^2} \left(\sum_{v \in S^t} \widetilde{l}_v^t\right)^2 \\
&= \frac{1}{\|\mathbf{x}^t\|^2} \sum_{v \in S^t} \left(\widetilde{l}_v^t - \frac{a^2}{a^2|S^t|+1} \sum_{j \in S^t} \widetilde{l}_j^t\right) \left(\widetilde{l}_v^t + \frac{a^2}{a^2|S^t|+1} \sum_{j \in S^t} \widetilde{l}_j^t\right) - \frac{1}{\|\mathbf{x}^t\|^2} \frac{a^2}{(a^2|S^t|+1)^2} \left(\sum_{v \in S^t} \widetilde{l}_v^t\right)^2 \\
&\quad - \frac{2}{\|\mathbf{x}^t\|^2} \sum_{v \in S^t} \widetilde{l}_v^{*t} \widetilde{l}_v^t + \frac{2a_t^2}{a_t^2|S^t|+1} \sum_{v \in S^t} \widetilde{l}_v^{*t} \sum_{j \in S^t} \widetilde{l}_v^t
\end{aligned}$$

We see that $\frac{2a_t^2}{a_t^2|S^t|+1} \sum_{v \in S^t} \widetilde{l}_v^{*t} \sum_{j \in S^t} \widetilde{l}_v^t \geq 0$. Thus, we get the following.

$$\begin{aligned}
\Delta_t &\geq \frac{1}{\|\mathbf{x}^t\|^2} \sum_{v \in S^t} \left(\widetilde{l}_v^t\right)^2 - \left(\frac{a^4|S|}{(a^2|S^t|+1)^2}\right) \left(\sum_{j \in S^t} \widetilde{l}_j^t\right)^2 - \frac{1}{\|\mathbf{x}^t\|^2} \frac{a^2}{(a^2|S^t|+1)^2} \left(\sum_{v \in S^t} \widetilde{l}_v^t\right)^2 - \frac{2}{\|\mathbf{x}^t\|^2} \sum_{v \in S^t} \widetilde{l}_v^{*t} \widetilde{l}_v^t \\
&\geq \frac{1}{\|\mathbf{x}^t\|^2} \sum_{v \in S^t} \left(\widetilde{l}_v^t\right)^2 - \left(\frac{a^2}{(a^2|S^t|+1)}\right) \left(\sum_{j \in S^t} \widetilde{l}_j^t\right)^2 - \frac{2}{\|\mathbf{x}^t\|^2} \sum_{v \in S^t} \widetilde{l}_v^{*t} \widetilde{l}_v^t \\
&= \frac{1}{\|\mathbf{x}^t\|^2 (a^2|S^t|+1)} \sum_{v \in S^t} \left(\widetilde{l}_v^t\right)^2 + \frac{a^2}{\|\mathbf{x}^t\|^2 (a^2|S^t|+1)} \left(|S^t| \sum_{v \in S^t} (\widetilde{l}_v^t)^2 - \left(\sum_{v \in S^t} \widetilde{l}_v^t\right)^2\right) \\
&\quad - \frac{2}{\|\mathbf{x}^t\|^2} \sum_{v \in S^t} \widetilde{l}_v^{*t} \widetilde{l}_v^t
\end{aligned}$$

We observe that $n \sum_{i=1}^n m_i^2 - \left(\sum_{i=1}^n m_i\right)^2 = (n-1) \sum_{i=1}^n m_i^2 - 2 \sum_{i=1}^n \sum_{j=i+1}^n m_i m_j = \sum_{i=1}^n \sum_{j=i+1}^n (m_i - m_j)^2 \geq 0$. Thus,

$$\Delta_t \geq \frac{1}{\|\mathbf{x}^t\|^2 (a^2|S^t|+1)} \sum_{v \in S^t} \left(\widetilde{l}_v^t\right)^2 - \frac{2}{\|\mathbf{x}^t\|^2} \sum_{v \in S^t} \widetilde{l}_v^{*t} \widetilde{l}_v^t$$

Taking expectation on both side with respect to $\widetilde{y}^t$.

$$\begin{aligned}\mathbb{E}\left[\Delta_t\right] &\geq \frac{1}{\|\mathbf{x}^t\|^2}\mathbb{E}\left[\frac{1}{(a^2|S^t|+1)}\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2 - 2\sum_{v\in S^t}\widetilde{l}^{*t}_v\widetilde{l}^t_v\right]\\ &= \frac{1}{\|\mathbf{x}^t\|^2}\left(\mathbb{E}\left[\frac{1}{(a^2|S^t|+1)}\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2\right] - 2\mathbb{E}\left[\sum_{v\in S^t}\widetilde{l}^{*t}_v\widetilde{l}^t_v\right]\right)\end{aligned}$$

But, $\mathbb{E}[\sum_{v\in S^t}\widetilde{l}^t_v\widetilde{l}^{t*}_v] \leq \mathbb{E}[\sqrt{\sum_{v\in S^t}(\widetilde{l}^t_v)^2\sum_{v\in S^t}(\widetilde{l}^{*t}_v)^2}$. Now using Cauchy Schwartz inequality $\mathbb{E}\left[xy\right] \leq \sqrt{\mathbb{E}\left[x^2\right]\mathbb{E}\left[y^2\right]}$, we get, $\mathbb{E}[\sum_{v\in S^t}\widetilde{l}^t_v\widetilde{l}^{t*}_v] \leq \sqrt{\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}^t_v)^2]}\sqrt{\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}^{*t}_v)^2]}$ Using this we get the following.

$$\mathbb{E}[\Delta_t] \geq \frac{1}{\|\mathbf{x}^t\|^2}\mathbb{E}\left[\frac{1}{(a^2|S^t|+1)}\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2\right] - \frac{2}{\|\mathbf{x}^t\|^2}\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2\right]}\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^{*t}_v\right)^2\right]}$$

Since $\|\mathbf{x}^t\|^2 \leq R^2$, $a \leq \dfrac{K}{\gamma}$ and $|S^t| \leq K$, so $a^2|S^t|+1 \leq \dfrac{K^3}{\gamma^2}+1$, therefore $\dfrac{1}{\|\mathbf{x}^t\|^2}\dfrac{1}{a^2|S^t|+1} \geq \dfrac{1}{R^2}\dfrac{1}{\left(\frac{K^3}{\gamma^2}+1\right)}$ Using the above approximations to get,

$$\mathbb{E}\left[\Delta_t\right] \geq \frac{1}{R^2}\frac{1}{\left(\frac{K^3}{\gamma^2}+1\right)}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2\right] - \frac{2}{R^2}\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2\right]}\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^{*t}_v\right)^2\right]} \quad (19)$$

Combining the lower bounds in (18) and (19), we can get the following lower bound on $\mathbb{E}[\Delta_t]$.

$$\mathbb{E}\left[\Delta_t\right] \geq \frac{1}{R^2}\frac{1}{\left(\frac{K^3}{\gamma^2}+1\right)}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2\right] - \frac{4}{R^2}\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2\right]}\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^{*t}_v\right)^2\right]}$$

Summing $\Delta_t = \sum_{v=1}^K\|\mathbf{w}^t_v - \mathbf{u}_v\|^2 - \sum_{v=1}^K\|\mathbf{w}^{t+1}_v - \mathbf{u}_v\|^2$ from $t=1$ to T.

$$\sum_{t=1}^T\Delta_t = \sum_{v=1}^K\|\mathbf{w}^1_v - \mathbf{u}_v\|^2 - \sum_{v=1}^K\|\mathbf{w}^{T+1}_v - \mathbf{u}_v\|^2$$

Since $\mathbf{w}^1 = 0$ and $\|\mathbf{w}^{T+1}_v - \mathbf{u}_v\|^2 \geq 0$, we get $\sum_{t=1}^T\Delta_t \leq \sum_{v=1}^K\|\mathbf{u}_v\|^2$. Let $\alpha = \left(\dfrac{K^3}{\gamma^2}+1\right)$, then comparing the upper and lower bounds on $\sum_{t=1}^T\mathbb{E}[\Delta_t]$, we get

$$\sum_{t=1}^T\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2\right] \leq R^2\alpha\sum_{v=1}^K\|\mathbf{u}_v\|^2 + 4\alpha\sum_{t=1}^T\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^t_v\right)^2\right]}\sqrt{\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}^{*t}_v\right)^2\right]}.$$

Using Cauchy-Shwartz Inequality, we get

$$\sum_{t=1}^{T}\sqrt{\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}_v^t)^2]}\sqrt{\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}_v^{*t})^2]} \leq \sqrt{\sum_{t=1}^{T}\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}_v^t)^2]}\sqrt{\sum_{t=1}^{T}\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}_v^{*t})^2]}.$$

Let $L_T = \sqrt{\sum_{t=1}^{T}\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}_v^t)^2]}$ and $U_T = \sqrt{\sum_{t=1}^{T}\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}_v^{*t})^2]}$. So, we get

$$L_T^2 \leq R^2\alpha\sum_{v=1}^{K}\|\mathbf{u}_v\|^2 + 4\alpha L_T U_T.$$

The upper bound is bounded by largest root of the polynomial $L_T^2 - 4\alpha L_T U_T - R^2\alpha\sum_{v=1}^{K}\|\mathbf{u}_v\|^2$ which is $2\alpha U_T + \sqrt{4\alpha^2 U_T^2 + R^2\alpha\sum_{v=1}^{K}\|\mathbf{u}_v\|^2}$. Using the inequality that $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$, we get, $L_T \leq 4\alpha U_T + R\sqrt{\alpha\sum_{v=1}^{K}\|\mathbf{u}_v\|^2}$. As $\widetilde{l}_v^t = 0,\ \forall v \notin S_t$, we get, $\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2 = \sum_{v=1}^{K}\left(\widetilde{l}_v^t\right)^2$. Thus, we get,

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}[\sum_{v=1}^{K}(\widetilde{l}_v^t)^2] &\leq \left(R\sqrt{\alpha\sum_{v=1}^{K}\|\mathbf{u}_v\|^2} + 4\alpha\sqrt{\sum_{t=1}^{T}\mathbb{E}[\sum_{v\in S^t}(\widetilde{l}_v^{*t})^2]}\right)^2\\ &\leq \left(R\sqrt{\alpha\sum_{v=1}^{K}\|\mathbf{u}_v\|^2} + 4\alpha\sqrt{\sum_{t=1}^{T}\mathbb{E}[\sum_{v=1}^{K}(\widetilde{l}_v^{*t})^2]}\right)^2.\end{aligned}$$

■

4. Derivation Of EPABF-I Updates

EPABF-I updates the parameter by solving the following optimization problem.

$$\begin{aligned}\mathbf{w}_1^{t+1}\ldots\mathbf{w}_K^{t+1} &= \arg\min_{\mathbf{w}_1,\ldots,\mathbf{w}_K}\frac{1}{2}\sum_{v=1}^{K}\|\mathbf{w}_v-\mathbf{w}_v^t\|^2 + C\sum_{v=1}^{K}\widetilde{l}_v\\ &= \arg\min_{\mathbf{w}_1,\ldots,\mathbf{w}_K}\frac{1}{2}\sum_{v=1}^{K}\|\mathbf{w}_v-\mathbf{w}_v^t\|^2 + C\sum_{v=1}^{K}\xi_v\\ &s.t.\begin{cases}a_t\mathbf{w}_{\widetilde{y}^t}\cdot\mathbf{x}^t - \mathbf{w}_r\cdot\mathbf{x}^t \geq 1-\xi_r, r\in[K]\\ \xi_r\geq 0, r\in[K]\end{cases}\end{aligned} \tag{20}$$

KKT conditions for the optimal solution of the optimization problem (20) are as follows.

$$\begin{cases}\lambda_v^t\left(1-\xi_v+\mathbf{w}_v\cdot\mathbf{x}^t - a_t\mathbf{w}_{\widetilde{y}^t}\cdot\mathbf{x}^t\right)=0,\ \forall v\\ \lambda_v^t\geq 0;\ 1-\xi_v+\mathbf{w}_v\cdot\mathbf{x}^t - a_t\mathbf{w}_{\widetilde{y}^t}\cdot\mathbf{x}^t\leq 0,\ \forall v\\ \beta_v^t\geq 0;\ \ \xi_v\geq 0,\ \forall v\\ \mathbf{w}_v = \mathbf{w}_v^t - \lambda_v^t\mathbf{x}^t + \mathbb{I}_{\{\widetilde{y}^t=v\}}a_t\sum_{i=1}^{K}\lambda_i^t\mathbf{x}^t,\ \forall v\\ C=\lambda_v^t+\beta_v^t,\ \forall v\end{cases}$$

We now determine λ_r^t for the support classes. When $\lambda_r^t > 0$, we see that

$$\mathbf{w}_{\widetilde{y}^t}^{t+1} \cdot \mathbf{x}^t a_t - \mathbf{w}_r^{t+1} \cdot \mathbf{x}^t = 1 - \xi_r$$

Using the corresponding values of $\mathbf{w}_{\widetilde{y}^t}^{t+1}$ and $\mathbf{w}_r^{t+1}$, we get

$$\frac{\xi_r}{\|\mathbf{x}^t\|^2} + a_t^2 \sum_i \lambda_i^t - a_t \lambda_{\widetilde{y}^t}^t + \lambda_r^t = \frac{\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2} \tag{21}$$

Summing the above for $\forall r \in S^t$, we get,

$$\sum_{r \in S^t} \frac{\xi_r}{\|\mathbf{x}^t\|^2} + (a_t^2 |S^t| + 1) \sum_{r \in S^t} \lambda_r^t - a_t |S^t| \lambda_{\widetilde{y}^t}^t = \frac{\sum_{r \in S^t} \widetilde{l}_r^t}{\|\mathbf{x}^t\|^2} \tag{22}$$

Now, we have two cases:

- **Case 1:** $\widetilde{y}^t \in S^t$: Using $r = \widetilde{y}^t$ in Eq.(21) we get,

$$\frac{\xi_{\widetilde{y}^t}}{\|\mathbf{x}^t\|^2} + a_t^2 \sum_{r \in S^t} \lambda_r^t + (1 - a_t) \lambda_{\widetilde{y}^t}^t = \frac{\widetilde{l}_{\widetilde{y}^t}^t}{\|\mathbf{x}^t\|^2} \tag{23}$$

Using (22) and (23), we solve for $\lambda_{\widetilde{y}^t}^t$ and $\sum_{v \in S^t} \lambda_v^t$.

$$\sum_{r \in S^t} \lambda_r^t = \frac{(1 - a_t) \sum_{r \in S^t} \widetilde{l}_r^t + a_t |S^t| \widetilde{l}_{\widetilde{y}^t}^t}{\|\mathbf{x}^t\|^2 (a_t^2 |S^t| + 1 - a_t)} - \frac{\left(a_t |S^t| \xi_{\widetilde{y}^t} + (1 - a_t) \sum_{r \in S^t} \xi_r\right)}{\|\mathbf{x}^t\|^2 (a_t^2 |S^t| + 1 - a_t)}$$

$$\lambda_{\widetilde{y}^t}^t = \frac{(a_t^2 |S^t| + 1) \widetilde{l}_{\widetilde{y}^t}^t - a_t^2 \sum_{r \in S^t} \widetilde{l}_r^t + a_t^2 \sum_{r \in S^t} \xi_r}{\|\mathbf{x}^t\|^2 (a_t^2 |S^t| + 1 - a_t)} - \frac{(a_t^2 |S^t| + 1) \xi_{\widetilde{y}^t}}{\|\mathbf{x}^t\|^2 (a_t^2 |S^t| + 1 - a_t)}$$

Plugging the values of $\lambda_{\widetilde{y}^t}^t$ and $\sum_{r \in S^t} \lambda_r^t$ in Eq.(21) , to get λ_r^t as follows.

$$\begin{aligned}\lambda_r^t = \frac{1}{\|\mathbf{x}^t\|^2} &\left(\widetilde{l}_r^t + \frac{a_t \widetilde{l}_{\widetilde{y}^t}^t}{a_t^2 |S^t| + 1 - a_t} - \frac{a_t^2 \sum_{v \in S^t} \widetilde{l}_v^t}{a_t^2 |S^t| + 1 - a_t} \right) - \\ &\frac{1}{\|\mathbf{x}^t\|^2} \left(\xi_r - \frac{a_t \xi_{\widetilde{y}^t}}{a_t^2 |S^t| + 1 - a_t} + \frac{a_t^2 \sum_{v \in S^t} \xi_v}{a_t^2 |S^t| + 1 - a_t} \right)\end{aligned}$$

Since $\left(\xi_r - \frac{a_t \xi_{\widetilde{y}^t}}{a_t^2 |S^t| + 1 - a_t} + \frac{a_t^2 \sum_{v \in S^t} \xi_v}{a_t^2 |S^t| + 1 - a_t} \right) \geq 0$, so we have,

$$\lambda_r^t \leq \frac{1}{\|\mathbf{x}^t\|^2} \left(\widetilde{l}_r^t + \frac{a_t \widetilde{l}_{\widetilde{y}^t}^t}{a_t^2 |S^t| + 1 - a_t} - \frac{a_t^2 \sum_{v \in S^t} \widetilde{l}_v^t}{a_t^2 |S^t| + 1 - a_t} \right)$$

From the KKT conditions we know that $\lambda_r \leq C$. Thus,

$$\lambda_r^t = \min \left(C, \frac{1}{\|\mathbf{x}^t\|^2} \left(\widetilde{l}_r^t + \frac{a_t \widetilde{l}_{\widetilde{y}^t}^t}{a_t^2 |S^t| + 1 - a_t} - \frac{a_t^2 \sum_{v \in S^t} \widetilde{l}_v^t}{a_t^2 |S^t| + 1 - a_t} \right) \right)$$

- **Case 2:** $\widetilde{y}^t \notin S^t$**:** In this case $\lambda^t_{\widetilde{y}^t} = 0$. Using Eq.(21) and (22) , we will get,

$$\lambda_r^t = \min\left(C, \frac{1}{\|\mathbf{x}^t\|^2}\left(\widetilde{l}_r^t - \frac{a_t^2 \sum_{v\in S^t} \widetilde{l}_v^t}{a_t^2|S^t| + 1}\right)\right) \tag{24}$$

5. Proof of Theorem 3: EPABF-I bound

Proof In the previous proof, we had

$$\Delta_t \geq 2\sum_{v\in S^t} \lambda_v^t \left(\widetilde{l}_v^t - \widetilde{l}_v^{*t}\right) - \sum_{v\in S^t} \left(\lambda_v^t\right)^2 \|\mathbf{x}_t\|^2 - a_t^2 \left(\sum_v \lambda_v^t\right)^2 \|\mathbf{x}_t\|^2 + 2a_t\lambda^t_{\widetilde{y}^t} \sum_v \lambda_v^t \|\mathbf{x}_t\|^2$$

- **Case 1:** If $\widetilde{y}^t \in S^t$ The step size for EPA-I is

$$\lambda_v^t = \min\left(C, \frac{1}{\|\mathbf{x}^t\|^2}\left(\widetilde{l}_v^t + \frac{a_t}{1 + |S^t|a_t^2 - a_t}\widetilde{l}^t_{\widetilde{y}^t} - \frac{a_t^2}{1 + |S^t|a_t^2 - a_t}\sum_{i\in S}\widetilde{l}_i^t\right)\right)$$

- **Case 2:** If $\widetilde{y}^t \notin S^t$ The step size for EPA-I is

$$\lambda_v^t = \min\left(C, \frac{1}{\|\mathbf{x}^t\|^2}\left(\widetilde{l}_v^t - \frac{a_t^2}{1 + |S^t|a_t^2}\sum_{i\in S}\widetilde{l}_i^t\right)\right)$$

So in both the cases we have $\lambda_v^t \leq C, \lambda_v^t \widetilde{l}_v^{*t} \leq C\widetilde{l}_v^{*t}$, $\sum_{v\in S^t} \lambda_v^t \widetilde{l}_v^{*t} \leq C \sum_{v\in S^t} \widetilde{l}_v^{*t}$, $\left(\lambda_v^t\right)^2 \leq C^2$, $\sum_{v\in S^t}\left(\lambda_v^t\right)^2 \leq C^2|S^t|$
So,

$$\Delta_t \geq 2\sum_{v\in S^t} \lambda_v^t \widetilde{l}_v^t - 2\sum_{v\in S^t} \lambda_v^t \widetilde{l}_v^{*t} - \sum_{v\in S^t} \left(\lambda_v^t\right)^2 \|\mathbf{x}_t\|^2 - a_t^2 \left(\sum_v \lambda_v^t\right)^2 \|\mathbf{x}_t\|^2 + 2a_t\lambda^t_{\widetilde{y}^t} \sum_v \lambda_v^t \|\mathbf{x}_t\|^2$$

By using the above mentioned approximations, we get,

$$\Delta_t \geq 2\sum_{v\in S^t} \lambda_v^t \widetilde{l}_v^t - 2C\sum_{v\in S^t} \widetilde{l}_v^{*t} - C^2|S^t|R^2 - \frac{C^2K^2|S^t|^2R^2}{\gamma^2}$$

Since $|S^t| \leq K$, we get,

$$\Delta_t \geq 2\sum_{v\in S^t} \lambda_v^t \widetilde{l}_v^t - 2C\sum_{v\in S^t} \widetilde{l}_v^{*t} - C^2KR^2 - \frac{C^2K^4R^2}{\gamma^2}$$

$$\mathbb{E}\left[\Delta_t\right] \geq 2\mathbb{E}\left[\sum_{v\in S^t} \lambda_v^t \widetilde{l}_v^t\right] - 2C\mathbb{E}\left[\sum_{v\in S^t} \widetilde{l}_v^{*t}\right] - C^2KR^2 - \frac{C^2K^4R^2}{\gamma^2}$$

If we consider Case 1:

Adding and subtracting $2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t \frac{a_t\widetilde{l}^t_{\widetilde{y}^t}}{a_t^2|S^t|+1-a_t}\right]$ $-2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1-a_t}\right]$, and $\mathbb{E}\left[\sum_{v\in S^t}(\lambda_v^t)^2\right]$ and using the fact that $\sum_{v\in S^t}\lambda_v^t \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1-a_t} - \sum_{v\in S^t}\lambda_v^t \frac{a_t\widetilde{l}^t_{\widetilde{y}^t}}{a_t^2|S^t|+1-a_t} \geq 0$ and $\sum_{v\in S^t}(\lambda_v^t)^2 \geq 0$ and simplifying to get,

$$\mathbb{E}\left[\Delta_t\right] \geq 2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t\left(\widetilde{l}^t_v + \frac{a_t\widetilde{l}^t_{\widetilde{y}^t}}{a_t^2|S^t|+1-a_t} - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1-a_t} - \frac{\lambda_v^t}{2}\right)\right] - 2C\mathbb{E}\left[\sum_{v\in S^t}\widetilde{l}^{*t}_v\right] - C^2KR^2 - \frac{C^2K^4R^2}{\gamma^2}$$

If we consider Case 2:

Adding and subtracting $-2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1}\right]$, and $\mathbb{E}\left[\sum_{v\in S^t}(\lambda_v^t)^2\right]$ and using the fact that $\sum_{v\in S^t}\lambda_v^t \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1}$ and $\sum_{v\in S^t}(\lambda_v^t)^2 \geq 0$ and simplifying to get,

$$\mathbb{E}\left[\Delta_t\right] \geq 2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t\left(\widetilde{l}^t_v - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1} - \frac{\lambda_v^t}{2}\right)\right] - 2C\mathbb{E}\left[\sum_{v\in S^t}\widetilde{l}^{*t}_v\right] - C^2KR^2 - \frac{C^2K^4R^2}{\gamma^2}$$

Combining the two cases we get,

$$\begin{aligned}\mathbb{E}\left[\Delta_t\right] &\geq 2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t\left(\widetilde{l}^t_v - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1} - \frac{\lambda_v^t}{2}\right)\right] - 2C\mathbb{E}\left[\sum_{v\in S^t}\widetilde{l}^{*t}_v\right] - C^2KR^2 - \frac{C^2K^4R^2}{\gamma^2}\\ &\geq 2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t\left(\frac{1}{\|\mathbf{x}^t\|^2}\left(\widetilde{l}^t_v - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1}\right) - \frac{\lambda_v^t}{2}\right)\right] - 2C\mathbb{E}\left[\sum_{v\in S^t}\widetilde{l}^{*t}_v\right] - C^2KR^2 - \frac{C^2K^4R^2}{\gamma^2}\end{aligned}$$

Now the above expression becomes,

$$\mathbb{E}\left[\Delta_t\right] \geq 2C\mathbb{E}\left[\sum_{v=1}^{K}\phi\left(\frac{1}{\|\mathbf{x}^t\|^2}\left(\widetilde{l}^t_v - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1}\right)\right)\right] - 2C\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}^{*t}_v\right] - C^2KR^2 - \frac{C^2K^4R^2}{\gamma^2}$$

where $\phi(z) = \frac{1}{C}\left(\min(z,C)\left(z - \frac{1}{2}\min(z,C)\right)\right)$,Shalev-Shwartz and Singer (2007). Summing the above from t=1 to T to get,

$$\sum_{t=1}^{T}\mathbb{E}\left[\Delta_t\right] \geq 2C\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\phi\left(\frac{1}{\|\mathbf{x}^t\|^2}\left(\widetilde{l}^t_v - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}^t_j}{a_t^2|S^t|+1}\right)\right)\right]\right) - 2C\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}^{*t}_v\right]\right) - TC^2KR^2 - \frac{TC^2K^4R^2}{\gamma^2}$$

Also $\phi(.)$ is a convex function, so we get,

$$\sum_{t=1}^{T}\mathbb{E}\left[\Delta_t\right] \geq \frac{2CT}{R^2}\phi\left(\frac{1}{T}\left(\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^t - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{a_t^2|S^t|+1}\right]\right)\right)\right) - 2C\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^{*t}\right]\right) - TC^2KR^2 - \frac{TC^2K^4R^2}{\gamma^2}$$

We had,

$$\sum_{t=1}^{T}\Delta_t \leq \sum_{v=1}^{K}\|\mathbf{u}_v\|^2$$

$$\sum_{t=1}^{T}\mathbb{E}\left[\Delta_t\right] \leq \sum_{v=1}^{K}\|\mathbf{u}_v\|^2$$

On comparing the lower and upper bounds we get,

$$\frac{2CT}{R^2}\phi\left(\frac{1}{T}\left(\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^t - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{a_t^2|S^t|+1}\right]\right)\right)\right) \leq \sum_{v=1}^{K}\|\mathbf{u}_v\|^2 + 2C\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^{*t}\right]\right) + TC^2KR^2 + \frac{TC^2K^4R^2}{\gamma^2}$$

Simplifying it to get,

$$\frac{1}{TR^2}\left(\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^t - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{a_t^2|S^t|+1}\right]\right)\right) \leq \phi^{-1}\left(\frac{1}{2CT}\sum_{v=1}^{K}\|\mathbf{u}_v\|^2 + \frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^{*t}\right]\right) + \frac{CKR^2}{2} + \frac{CK^4R^2}{2\gamma^2}\right)$$

Notice that,

$$\begin{aligned}
\frac{1}{TR^2}\left(\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^t - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{a_t^2|S^t|+1}\right]\right)\right) &\geq \frac{1}{TR^2}\left(\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^t - \frac{1}{a_t|S^t|}\sum_{j\in S^t}\widetilde{l}_j^t\right]\right)\right) \\
&= \frac{1}{TR^2}\left(\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^t - \frac{1}{a_t}\widetilde{l}_v^t\right]\right)\right) \\
&\geq \frac{1}{TR^2}\left(1-\frac{\gamma}{K}\right)\left(\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^t\right]\right)\right)
\end{aligned}$$

we know that $\phi^{-1}(z) \leq z + \frac{C}{2}$,Shalev-Shwartz and Singer (2007). Hence we get,

$$\phi^{-1}\left(\frac{1}{2CT}\sum_{v=1}^{K}\|\mathbf{u}_v\|^2 + \frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^{*t}\right]\right) + \frac{CKR^2}{2} + \frac{CK^4R^2}{2\gamma^2}\right) \leq \frac{1}{2CT}\sum_{v=1}^{K}\|\mathbf{u}_v\|^2 + \frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^{*t}\right]\right) + \frac{CKR^2}{2} + \frac{CK^4R^2}{2\gamma^2} + \frac{C}{2}$$

Combining the last two inequalities we get,

$$\frac{1}{R^2}\left(1-\frac{\gamma}{K}\right)\left(\sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^t\right]\right)\right) \leq \frac{1}{2C}\sum_{v=1}^{K}\|\mathbf{u}_v\|^2 + \sum_{t=1}^{T}\left(\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^{*t}\right]\right) + \frac{TCKR^2}{2} + \frac{TCK^4R^2}{2\gamma^2} + \frac{CT}{2}$$

We use $C = \frac{\sqrt{\sum_{v=1}^{K}\|\mathbf{u}_v\|^2}}{\sqrt{TKR^2 + \frac{TK^4R^2}{\gamma^2} + T}}$ as it minimizes the upper bound. Using that we get,

$$\frac{1}{R^2}\left(1-\frac{\gamma}{K}\right)\sum_{t=1}^{T}\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^t\right] \leq \frac{1}{2}\sqrt{\sum_{v=1}^{K}\|\mathbf{u}_v\|^2}\sqrt{TKR^2 + \frac{TK^4R^2}{\gamma^2} + T} + \sum_{t=1}^{T}\mathbb{E}\left[\sum_{v=1}^{K}\widetilde{l}_v^{*t}\right]$$

■

6. Derivation of EPABF-II Updates

The optimization problem associated to the EPABF-II is as follows.

$$\begin{aligned}
\mathbf{w}_1^{t+1}\ldots\mathbf{w}_K^{t+1} &= \underset{\mathbf{w}_1\ldots\mathbf{w}_K}{\arg\min}\frac{1}{2}\sum_{v=1}^{K}\|\mathbf{w}_v - \mathbf{w}_v^t\|^2 + C\sum_{v=1}^{K}(\widetilde{l}_v)^2 \\
&= \underset{\mathbf{w}_1\ldots\mathbf{w}_K}{\arg\min}\ \frac{1}{2}\sum_{v=1}^{K}\|\mathbf{w}_v - \mathbf{w}_v^t\|^2 + C\sum_{v=1}^{K}\xi_v^2 \\
&s.t.\ \ a_t\mathbf{w}_{\widetilde{y}^t}\cdot\mathbf{x}^t - \mathbf{w}_r\cdot\mathbf{x}^t \geq 1-\xi_v,\ v\in[K]
\end{aligned}$$

The optimal solution satisfies the following KKT conditions.

$$\begin{cases}
\lambda_r^t\left(1-\xi_r + \mathbf{w}_r\cdot\mathbf{x}^t - a_t\mathbf{w}_{\widetilde{y}^t}\cdot\mathbf{x}^t\right) = 0,\ \forall r \\
\lambda_r^t \geq 0;\ \left(1-\xi_r + \mathbf{w}_r\cdot\mathbf{x}^t - a_t\mathbf{w}_{\widetilde{y}^t}\cdot\mathbf{x}^t\right) \leq 0,\ \forall r \\
\mathbf{w}_r = \mathbf{w}_r^t - \lambda_r^t\mathbf{x}_t + \mathbb{I}_{\{\widetilde{y}^t = r\}}a_t\sum_i\lambda_i^t\mathbf{x}^t,\ \ \forall r \\
\xi_r = \dfrac{\lambda_r^t}{2C},\ \ \forall r
\end{cases}$$

Now we determine λ_r^t for the support classes.

$$\begin{aligned}
& a_t\left(\mathbf{w}_{\widetilde{y}_t^t}^t - \lambda_{\widetilde{y}^t}^t\mathbf{x}_t + a_t\sum_i\lambda_i^t\mathbf{x}^t\right)\cdot\mathbf{x}^t - \left(\mathbf{w}_r^t - \lambda_r^t\mathbf{x}^t\right)\cdot\mathbf{x}^t = 1-\xi_r \\
\Rightarrow\ & a_t\mathbf{w}_{\widetilde{y}_t^t}^t\cdot\mathbf{x}^t - a_t\lambda_{\widetilde{y}^t}^t\|\mathbf{x}^t\|^2 + a_t^2\|\mathbf{x}^t\|^2\sum_i\lambda_i^t - \mathbf{w}_r^t\cdot\mathbf{x}^t + \lambda_r^t\|\mathbf{x}^t\|^2 = 1-\xi_r \\
\Rightarrow\ & \xi_r + (a_t^2\sum_i\lambda_i^t - a_t\lambda_{\widetilde{y}^t}^t + \lambda_r^t)\|\mathbf{x}^t\|^2 = 1 - a_t\mathbf{w}_{\widetilde{y}_t^t}^t\cdot\mathbf{x}^t + \mathbf{w}_r^t\cdot\mathbf{x}^t
\end{aligned}$$

Using value of ξ_r and rearrange to get,

$$\frac{1}{2C\|\mathbf{x}^t\|^2}\lambda_r^t + a_t^2\sum_i \lambda_i^t - a_t\lambda_{\widetilde{y}^t}^t + \lambda_r^t = \frac{\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2} \tag{25}$$

Summing the above for $\forall r \in S^t$, we get,

$$\left(a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)\sum_{r\in S^t}\lambda_r^t - a_t|S^t|\lambda_{\widetilde{y}^t}^t = \frac{\sum_{r\in S^t}\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2} \tag{26}$$

Now, we have two cases:

- **Case 1 : $\widetilde{y}^t \in S^t$:** Taking $r = \widetilde{y}^t$, the Eq.(25) becomes,

$$a_t^2\sum_{r\in S^t}\lambda_r^t + \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right)\lambda_{\widetilde{y}^t}^t = \frac{\widetilde{l}_{\widetilde{y}^t}^t}{\|\mathbf{x}^t\|^2} \tag{27}$$

  Using Eq.(26) and (27), we get the following.

$$\sum_{r\in S^t}\lambda_r^t = \frac{a_t|S^t|\widetilde{l}_{\widetilde{y}^t}^t + \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right)\sum_{r\in S^t}\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2\left(a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right)\left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)} \tag{28}$$

$$\lambda_{\widetilde{y}^t}^t = \frac{\left(a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)\widetilde{l}_{\widetilde{y}^t}^t - a_t^2\sum_{r\in S^t}\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2\left(a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right)\left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)} \tag{29}$$

  Using Eq.(28) and Eq.(29) in Eq.(25), we can find λ_r^t as follows.

$$\lambda_r^t = \frac{1}{\|\mathbf{x}^t\|^2\left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\left(\widetilde{l}_r^t + \frac{a_t\widetilde{l}_{\widetilde{y}^t}^t - a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t}\right) \tag{30}$$

- $\widetilde{y}^t \notin S^t$: In this case $\lambda_{\widetilde{y}^t}^t = 0$. Using Eq.(25) and (26), we will get,

$$\lambda_r^t = \frac{1}{\|\mathbf{x}^t\|^2\left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\left(\widetilde{l}_r^t - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2}}\right)$$

7. Proof of Theorem 4

Proof We will have 2 cases:

- **Case 1:** If $\widetilde{y}^t \in S^t$
  Using the KKT conditions, we see that for any $r \notin S^t$, we have,

$$a_t(\mathbf{w}^t_{\widetilde{y}^t} - \lambda^t_{\widetilde{y}^t}\mathbf{x}^t + a_t \sum_i \lambda^t_i \mathbf{x}^t) \cdot \mathbf{x}^t - (\mathbf{w}^t_r - \lambda^t_r \mathbf{x}^t) \cdot \mathbf{x}^t \geq 1 - \xi^t_r$$

Since $\xi^t_r = \frac{\lambda^t_r}{2C}, \lambda^t_r = 0$ for $r \notin S^t$, the above equation reduces to,

$$a_t^2 \sum_i \lambda^t_i - a_t \lambda^t_{\widetilde{y}^t} \geq \frac{1 + \mathbf{w}^t_r \cdot \mathbf{x}^t - a_t \mathbf{w}^t_{\widetilde{y}^t} \cdot \mathbf{x}^t}{\|\mathbf{x}^t\|^2} = \frac{\widetilde{l}^t_r}{\|\mathbf{x}^t\|^2} \tag{31}$$

We know that,

$$\sum_{r \in S^t} \lambda^t_r = \frac{a_t |S^t| \widetilde{l}^t_{\widetilde{y}^t} + \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right) \sum_{r \in S^t} \widetilde{l}^t_r}{\|\mathbf{x}^t\|^2 \left(a_t^2 |S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right)\left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)} \tag{32}$$

$$\lambda^t_{\widetilde{y}^t} = \frac{\left(a_t^2 |S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right) \widetilde{l}^t_{\widetilde{y}^t} - a_t^2 \sum_{r \in S^t} \widetilde{l}^t_r}{\|\mathbf{x}^t\|^2 \left((a_t^2 |S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right)\left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)} \tag{33}$$

Using Eq.(33) and (32) in Eq.(31), we get

$$\frac{a_t^2 \sum_{j \in S^t} \widetilde{l}^t_j}{|S^t| a_t^2 + 1 - a_t + \frac{1}{2C\|\mathbf{x}^t\|^2}} \geq \widetilde{l}^t_r + \frac{a_t \widetilde{l}^t_{\widetilde{y}^t}}{|S^t| a_t^2 + 1 - a_t + \frac{1}{2C\|\mathbf{x}^t\|^2}}, \quad \forall r \notin S^t$$

On the other hand, if r lies in the support set S^t, we have $\lambda^t_r > 0$. Using Eq.(30), we get the following condition for $r \in S^t$.

$$\frac{a_t^2 \sum_{j \in S^t} \widetilde{l}^t_{\sigma(j)}}{|S^t| a_t^2 + 1 - a_t + \frac{1}{2C\|\mathbf{x}^t\|^2}} < \widetilde{l}^t_{\sigma(r)} + \frac{a_t \widetilde{l}^t_{\widetilde{y}^t}}{|S^t| a_t^2 + 1 - a_t + \frac{1}{2C\|\mathbf{x}^t\|^2}}, \; \forall r \in S^t \tag{34}$$

Let $\sigma(k)$ be the k-th class when sorted in descending order of $\widetilde{l}^t_r$.

(Sufficiency) Assume that $\widetilde{l}^t_{\sigma(k)}$ satisfies the theorem, then we have,

$$\sum_{j=1}^{k-1} \widetilde{l}^t_{\sigma(j)} < \frac{\left(1+(k-1)a_t^2 - a_t + \dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)}{a_t^2} \widetilde{l}^t_{\sigma(k)} + \frac{1}{a_t}\widetilde{l}^t_{\widetilde{y}^t}$$

$$\Rightarrow \sum_{j=1}^{|S^t|} \widetilde{l}^t_{\sigma(j)} < \frac{\left(1+|S^t|a_t^2 - a_t + \dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)}{a_t^2} \widetilde{l}_{\sigma(k)} + \frac{1}{a_t}\widetilde{l}^t_{\widetilde{y}^t}$$

$$\Rightarrow \frac{a_t^2}{\left(1+|S^t|a_t^2 - a_t + \dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)} \sum_{j=1}^{|S^t|} \widetilde{l}^t_{\sigma(j)} < \widetilde{l}^t_{\sigma(k)} + \frac{a_t}{\left(1+|S^t|a_t^2 - a_t + \dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)} \widetilde{l}^t_{\widetilde{y}^t}$$

The second inequality is justified as the losses $\widetilde{l}^t_{\sigma(j)}$ are in decreasing order. This means $\sigma(k)$ belongs to the support set S^t.

(Necessity) Assume that $\widetilde{l}_{\sigma(k)}$ does not satisfy theorem, then

$$\sum_{j=1}^{k-1} \widetilde{l}^t_{\sigma(j)} \geq \frac{\left(1+(k-1)a_t^2 - a_t + \dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)}{a_t^2} \widetilde{l}^t_{\sigma(k)} + \frac{1}{a_t}\widetilde{l}^t_{\widetilde{y}^t}$$

$$\sum_{j=1}^{k} \widetilde{l}^t_{\sigma(j)} \geq \frac{\left(1+(k)a_t^2 - a_t + \dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)}{a_t^2} \widetilde{l}^t_{\sigma(k)} + \frac{1}{a_t}\widetilde{l}^t_{\widetilde{y}^t}$$

$$\frac{a_t^2}{\left(1+ka_t^2 - a_t + \dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)} \sum_{j=1}^{k} \widetilde{l}^t_{\sigma(j)} - \frac{a_t}{\left(1+ka_t^2 - a_t + \dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)} \widetilde{l}^t_{\widetilde{y}^t} \geq \widetilde{l}^t_{\sigma(k)} \geq \widetilde{l}^t_{\sigma(k+1)}$$

Therefore, any j larger than $\sigma(k)$ does not satisfy Eq.(34). It means $|S^t| < k$ and thus k does not correspond to a label of a support class.

- **Case 2:** If $\widetilde{y}^t \notin S^t$
  Using the KKT conditions, we see that for any $r \notin S^t$, we have,

$$a_t(\mathbf{w}^t_{\widetilde{y}^t} + a_t \sum_i \lambda^t_i \mathbf{x}^t) \cdot \mathbf{x}^t - (\mathbf{w}^t_r - \lambda^t_r \mathbf{x}^t) \cdot \mathbf{x}^t \geq 1 - \xi^t_r$$

Since $\xi^t_r = \dfrac{\lambda^t_r}{2C}, \lambda^t_r = 0$ for $r \notin S^t$, the above equation reduces to,

$$a_t^2 \sum_i \lambda^t_i \geq \frac{1 + \mathbf{w}^t_r \cdot \mathbf{x}^t - a_t \mathbf{w}^t_{\widetilde{y}^t} \cdot \mathbf{x}^t}{\|\mathbf{x}^t\|^2} = \frac{\widetilde{l}^t_r}{\|\mathbf{x}^t\|^2}$$

We know that,

$$\sum_{r\in S^t} \lambda_r^t = \frac{1}{(1+\frac{1}{2C\|\mathbf{x}^t\|^2}+|S^t|a_t^2)\|\mathbf{x}^t\|^2}\left(\sum_{r\in S^t}\widetilde{l}_r^t\right) \tag{35}$$

Using Eq.(35), we can rewrite this equation as,

$$\frac{a_t^2\sum_{j\in S^t}\widetilde{l}_{\sigma(j)}^t}{1+\frac{1}{2C\|\mathbf{x}^t\|^2}+|S^t|a_t^2} \geq \widetilde{l}_{\sigma(r)}^t,\ \forall r\notin S^t \tag{36}$$

Also we have,

$$\lambda_r^t = \frac{1}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\left(\widetilde{l}_r^t - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{\left(a_t^2|S^t|+1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right) \tag{37}$$

To get support class, λ_r^t should be positive, so by Eq.(37), we get

$$\frac{a_t^2\sum_{j\in S^t}\widetilde{l}_{\sigma(j)}^t}{1+\frac{1}{2C\|\mathbf{x}^t\|^2}+|S^t|a_t^2} < \widetilde{l}_{\sigma(r)}^t,\ \forall r\in S^t \tag{38}$$

Let $\sigma(k)$ be the k-th class when sorted in descending order of $\widetilde{l}_r^t$.
(Sufficiency) Assume that $\widetilde{l}_{\sigma(k)}^t$ satisfies the theorem, then we have,

$$\sum_{j=1}^{k-1}\widetilde{l}_{\sigma(j)}^t < \frac{1+\frac{1}{2C\|\mathbf{x}^t\|^2}+(k-1)a_t^2}{a_t^2}\widetilde{l}_{\sigma(k)}^t$$

$$\Rightarrow \sum_{j=1}^{|S^t|}\widetilde{l}_{\sigma(j)}^t < \frac{1+\frac{1}{2C\|\mathbf{x}^t\|^2}+|S^t|a_t^2}{a_t^2}\widetilde{l}_{\sigma(k)}^t$$

$$\Rightarrow \frac{a_t^2}{1+\frac{1}{2C\|\mathbf{x}^t\|^2}+|S^t|a_t^2}\sum_{j=1}^{|S^t|}\widetilde{l}_{\sigma(j)}^t < \widetilde{l}_{\sigma(k)}^t$$

The second inequality is justified as the losses $\widetilde{l}_{\sigma(j)}^t$ are in decreasing order. This means $\sigma(k)$ corresponds to a label of some support classes Eq. (38).
(Necessity) Assume that $\widetilde{l}_r^t$ does not satisfy theorem, then

$$\sum_{j=1}^{k-1}\widetilde{l}_{\sigma(j)}^t \geq \frac{1+\frac{1}{2C\|\mathbf{x}^t\|^2}+(k-1)a_t^2}{a_t^2}\widetilde{l}_{\sigma(k)}^t$$

$$\sum_{j=1}^{k} \widetilde{l}^t_{\sigma(j)} \geq \frac{1 + \frac{1}{2C\|\mathbf{x}^t\|^2} + (k)a_t^2}{a_t^2} \widetilde{l}^t_{\sigma(k)}$$

$$\frac{a_t^2}{1 + \frac{1}{2C\|\mathbf{x}^t\|^2} + ka_t^2} \sum_{j=1}^{k} \widetilde{l}^t_{\sigma(j)} \geq \widetilde{l}^t_{\sigma(k)} \geq \widetilde{l}^t_{\sigma(k+1)}$$

Therefore, any j larger than $\sigma(k)$ does not satisfy Eq. (38). It means $|S^t| < k$ and thus $\sigma(k)$ does not correspond to a label of a support class.

■

8. Proof of Theorem 5: EPABF-II bound

Proof

- **Case 1:** If $\widetilde{y}^t \in S^t$ We had

$$\mathbb{E}\left[\Delta_t\right] \geq 2\mathbb{E}\left[\sum_{v \in S^t} \lambda_v^t \left(\widetilde{l}^t_v - \widetilde{l}^{*t}_v\right)\right] - \mathbb{E}\left[\sum_{v \in S^t} \left(\lambda_v^t\right)^2 \|\mathbf{x}_t\|^2\right] - \mathbb{E}\left[a_t^2 \left(\sum_v \lambda_v^t\right)^2 \|\mathbf{x}_t\|^2\right] + \mathbb{E}\left[2a_t \lambda^t_{\widetilde{y}^t} \sum_v \lambda_v^t \|\mathbf{x}_t\|^2\right]$$

And the step sizes for EPABF-II are as follows,

$$\sum_{r \in S^t} \lambda_r^t = \frac{a_t|S^t|\widetilde{l}^t_{\widetilde{y}^t} + \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right) \sum_{r \in S^t} \widetilde{l}^t_r}{\|\mathbf{x}^t\|^2 \left(a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right)\left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)}$$

$$\lambda^t_{\widetilde{y}^t} = \frac{\left(a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)\widetilde{l}^t_{\widetilde{y}^t} - a_t^2 \sum_{r \in S^t} \widetilde{l}^t_r}{\|\mathbf{x}^t\|^2 \left(a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2} - a_t\right)\left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)}$$

$$\lambda_r^t = \frac{1}{\|\mathbf{x}^t\|^2 \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)} \left(\widetilde{l}^t_r + \frac{a_t \widetilde{l}^t_{\widetilde{y}^t} - a_t^2 \sum_{j \in S^t} \widetilde{l}^t_j}{\left(a_t^2|S^t| + 1 - a_t + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)$$

The Inequality is true even if we subtract $\mathbb{E}\left[\sum_{v \in S^t}\left(\alpha\lambda_v^t - \frac{\widetilde{l}^{*t}_v}{\alpha}\right)^2\right]$ from it. Here,

$\alpha = \frac{1}{\sqrt{2C\|\mathbf{x}^t\|^2}}$

$$
\begin{aligned}
\mathbb{E}\left[\Delta_t\right] \geq\ & 2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t\left(\widetilde{l}_v^t-\widetilde{l}_v^{*t}\right)\right]-\mathbb{E}\left[\sum_{v\in S^t}\left(\lambda_v^t\right)^2\|\mathbf{x}_t\|^2\right]-\mathbb{E}\left[a_t^2\left(\sum_{v}\lambda_v^t\right)^2\|\mathbf{x}_t\|^2\right]\\
&+\mathbb{E}\left[2a_t\lambda_{\widetilde{y}^t}^t\sum_{v}\lambda_v^t\|\mathbf{x}_t\|^2\right]-\mathbb{E}\left[\sum_{v\in S^t}\left(\alpha\lambda_v^t-\frac{\widetilde{l}_v^{*t}}{\alpha}\right)^2\right]\\
=\ & 2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t\widetilde{l}_v^t\right]-(\|\mathbf{x}_t\|^2+\alpha^2)\mathbb{E}\left[\sum_{v\in S^t}\left(\lambda_v^t\right)^2\right]-\mathbb{E}\left[a_t^2\left(\sum_{v}\lambda_v^t\right)^2\|\mathbf{x}_t\|^2\right]\\
&+\mathbb{E}\left[2a_t\lambda_{\widetilde{y}^t}^t\sum_{v}\lambda_v^t\|\mathbf{x}_t\|^2\right]-\frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]\\
=\ & \frac{2}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t+\frac{a_t\widetilde{l}_{\widetilde{y}^t}^t-a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{\left(a_t^2|S^t|+1-a_t+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)\widetilde{l}_v^t\right]\\
&-\frac{(\|\mathbf{x}_t\|^2+\alpha^2)}{\|\mathbf{x}^t\|^4\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t+\frac{a_t\widetilde{l}_{\widetilde{y}^t}^t-a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{\left(a_t^2|S^t|+1-a_t+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)^2\right]\\
&-\mathbb{E}\left[a_t^2\left(\frac{a_t|S^t|\widetilde{l}_{\widetilde{y}^t}^t}{\|\mathbf{x}^t\|^2\left(a_t^2|S^t|+1+\frac{1}{2C\|\mathbf{x}^t\|^2}-a_t\right)\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right.\right.\\
&\left.\left.+\frac{\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}-a_t\right)\sum_{r\in S^t}\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2\left(a_t^2|S^t|+1+\frac{1}{2C\|\mathbf{x}^t\|^2}-a_t\right)\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)^2\|\mathbf{x}_t\|^2\right]\\
&+\mathbb{E}\left[2a_t\left(\frac{\left(a_t^2|S^t|+1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)\widetilde{l}_{\widetilde{y}^t}^t-a_t^2\sum_{r\in S^t}\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2\left(a_t^2|S^t|+1+\frac{1}{2C\|\mathbf{x}^t\|^2}-a_t\right)\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)\right.\\
&\left.\left(\frac{a_t|S^t|\widetilde{l}_{\widetilde{y}^t}^t+\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}-a_t\right)\sum_{r\in S^t}\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2\left(a_t^2|S^t|+1+\frac{1}{2C\|\mathbf{x}^t\|^2}-a_t\right)\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)\|\mathbf{x}_t\|^2\right]-\frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]
\end{aligned}
$$

$$\geq \left(\frac{2}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)} - \frac{\|\mathbf{x}^t\|^2+\alpha^2}{\|\mathbf{x}^t\|^4\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\right)\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right]$$
$$-\frac{1}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\widetilde{l}_{\widetilde{y}^t}^t\sum_{v\in S^t}\widetilde{l}_v^t\right] - \frac{1}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\left(\sum_{v\in S^t}\widetilde{l}_v^t\right)^2\right]$$
$$+\frac{1}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\left(\widetilde{l}_{\widetilde{y}^t}^t\right)^2\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]$$

$$= \left(\frac{2}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)} - \frac{\|\mathbf{x}^t\|^2+\alpha^2}{\|\mathbf{x}^t\|^4\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\right)\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right]$$
$$+\frac{1}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\left(\sum_{v\in S^t}\widetilde{l}_v^t - \widetilde{l}_{\widetilde{y}^t}^t\right)^2\right] - \frac{2}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\left(\sum_{v\in S^t}\widetilde{l}_v^t\right)^2\right]$$
$$+\frac{1}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\widetilde{l}_{\widetilde{y}^t}^t\sum_{v\in S^t}\widetilde{l}_v^t\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]$$
$$= \left(\frac{2}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C}\right)} - \frac{\|\mathbf{x}^t\|^2+\alpha^2}{\|\mathbf{x}^t\|^4\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\right)\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right]$$
$$+\frac{1}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\left(\sum_{v\in S^t}\widetilde{l}_v^t - \widetilde{l}_{\widetilde{y}^t}^t\right)^2 - 2\left(\sum_{v\in S^t}\widetilde{l}_v^t\right)^2 + \widetilde{l}_{\widetilde{y}^t}^t\sum_{v\in S^t}\widetilde{l}_v^t\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]$$

Since $(x-y)^2 + xy = \frac{x^2}{2} + \frac{y^2}{2} + \frac{(x-y)^2}{2} \geq \frac{x^2}{2} + \frac{y^2}{2}$

$$\geq \left(\frac{2}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)} - \frac{\|\mathbf{x}^t\|^2+\alpha^2}{\|\mathbf{x}^t\|^4\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\right)\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right]$$
$$-\left(\frac{3}{2}\right)\frac{1}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\left(\sum_{v\in S^t}\widetilde{l}_v^t\right)^2\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]$$

$$= \left(\frac{1 + \frac{1}{C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^2 \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \frac{\alpha^2}{\|\mathbf{x}^t\|^4 \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} \right) \mathbb{E}\left[\sum_{v\in S^t} \left(\widetilde{l}_v^t\right)^2\right]$$

$$- \left(\frac{3}{2}\right) \frac{1}{\|\mathbf{x}^t\|^2 \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} \mathbb{E}\left[\left(\sum_{v\in S^t} \widetilde{l}_v^t\right)^2\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t} \left(\widetilde{l}_v^{*t}\right)^2\right]$$

$$\geq \left(\frac{1 + \frac{1}{C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^2 \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \frac{\frac{1}{2C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^4 \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \left(\frac{3}{2}\right) \frac{K}{\|\mathbf{x}^t\|^2 \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} \right) \mathbb{E}\left[\sum_{v\in S^t} \left(\widetilde{l}_v^t\right)^2\right]$$

$$- \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t} \left(\widetilde{l}_v^{*t}\right)^2\right]$$

We have used $\mathbb{E}\left[\left(\sum_{v\in S^t} \widetilde{l}_v^t\right)^2\right] \leq K\mathbb{E}\left[\sum_{v\in S^t} \left(\widetilde{l}_v^t\right)^2\right]$ in the last inequality

- **Case 2:** If $\widetilde{y}^t \notin S^t, \lambda_{\widetilde{y}^t} = 0$, then we had

$$\mathbb{E}\left[\Delta_t\right] \geq 2\mathbb{E}\left[\sum_{v\in S^t} \lambda_v^t \left(\widetilde{l}_v^t - \widetilde{l}_v^{*t}\right)\right] - \mathbb{E}\left[\sum_{v\in S^t} \left(\lambda_v^t\right)^2 \|\mathbf{x}_t\|^2\right] - \mathbb{E}\left[a_t^2 \left(\sum_{v} \lambda_v^t\right)^2 \|\mathbf{x}_t\|^2\right]$$

And the step sizes for EPABF-II are as follows,

$$\sum_{r\in S^t} \lambda_r^t = \frac{\sum_{r\in S^t} \widetilde{l}_r^t}{\|\mathbf{x}^t\|^2 \left(a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)}$$

$$\lambda_r^t = \frac{1}{\|\mathbf{x}^t\|^2 \left(1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)} \left(\widetilde{l}_r^t - \frac{a_t^2 \sum_{j\in S^t} \widetilde{l}_j^t}{\left(a_t^2|S^t| + 1 + \frac{1}{2C\|\mathbf{x}^t\|^2}\right)} \right)$$

The Inequality is true even if we subtract $\mathbb{E}\left[\sum_{v\in S^t} \left(\alpha\lambda_v^t - \frac{\widetilde{l}_v^{*t}}{\alpha}\right)^2\right]$ from it. Here,

$\alpha = \frac{1}{\sqrt{2C\|\mathbf{x}^t\|^2}}$

$$
\begin{aligned}
\mathbb{E}\left[\Delta_t\right] \geq{}& 2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t\left(\widetilde{l}_v^t - \widetilde{l}_v^{*t}\right)\right] - \mathbb{E}\left[\sum_{v\in S^t}\left(\lambda_v^t\right)^2\|\mathbf{x}_t\|^2\right] - \mathbb{E}\left[a_t^2\left(\sum_v \lambda_v^t\right)^2\|\mathbf{x}_t\|^2\right] \\
&- \mathbb{E}\left[\sum_{v\in S^t}\left(\alpha\lambda_v^t - \frac{\widetilde{l}_v^{*t}}{\alpha}\right)^2\right] \\
={}& 2\mathbb{E}\left[\sum_{v\in S^t}\lambda_v^t\widetilde{l}_v^t\right] - (\|\mathbf{x}_t\|^2+\alpha^2)\mathbb{E}\left[\sum_{v\in S^t}\left(\lambda_v^t\right)^2\right] - \mathbb{E}\left[a_t^2\left(\sum_v\lambda_v^t\right)^2\|\mathbf{x}_t\|^2\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right] \\
={}& \frac{2}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{\left(a_t^2|S^t|+1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)\widetilde{l}_v^t\right] \\
&- \frac{(\|\mathbf{x}_t\|^2+\alpha^2)}{\|\mathbf{x}^t\|^4\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t - \frac{a_t^2\sum_{j\in S^t}\widetilde{l}_j^t}{\left(a_t^2|S^t|+1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)^2\right] \\
&- \mathbb{E}\left[a_t^2\left(\frac{\sum_{r\in S^t}\widetilde{l}_r^t}{\|\mathbf{x}^t\|^2\left(a_t^2|S^t|+1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)^2\|\mathbf{x}_t\|^2\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right] \\
\geq{}& \left(\frac{1+\frac{1}{C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \frac{\frac{1}{2C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^4\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \left(\frac{\left(3+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)K}{\|\mathbf{x}^t\|^2\left(1+\frac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)\right)\times \\
&\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]
\end{aligned}
$$

We have used $\mathbb{E}\left[\left(\sum_{v\in S^t}\widetilde{l}_v^t\right)^2\right] \leq K\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right]$ in the last inequality

Combining the two cases we get,

$$\mathbb{E}\left[\Delta_t\right] \geq \min\left(\left(\frac{1+\dfrac{1}{C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^2\left(1+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \frac{\dfrac{1}{2C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^4\left(1+\dfrac{1}{2CR^2}\right)^2} - \left(\frac{3}{2}\right)\frac{K}{\|\mathbf{x}^t\|^2\left(1+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)^2}\right)\times \right.$$
$$\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right],$$
$$\left(\frac{1+\dfrac{1}{C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^2\left(1+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \frac{\dfrac{1}{2C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^4\left(1+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \left(\frac{\left(3+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)K}{\|\mathbf{x}^t\|^2\left(1+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)\right)\times$$
$$\left.\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]\right)$$

So we can write,

$$\mathbb{E}\left[\Delta_t\right] \geq \left(\frac{1+\dfrac{1}{C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^2\left(1+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \frac{\dfrac{1}{2C\|\mathbf{x}^t\|^2}}{\|\mathbf{x}^t\|^4\left(1+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)^2} - \left(\frac{\left(3+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)K}{\|\mathbf{x}^t\|^2\left(1+\dfrac{1}{2C\|\mathbf{x}^t\|^2}\right)}\right)\right)\times$$
$$\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^t\right)^2\right] - \frac{1}{\alpha^2}\mathbb{E}\left[\sum_{v\in S^t}\left(\widetilde{l}_v^{*t}\right)^2\right]$$

Since, we have

$$\Delta_t = \mathbb{E}\left[\sum_{v=1}^{K}\|\mathbf{w}_v^t - \mathbf{u}_v\|^2\right] - \mathbb{E}\left[\sum_{v=1}^{K}\|\mathbf{w}_v^{t+1} - \mathbf{u}_v\|^2\right]$$

Summing it over t to get,

$$\sum_{t=1}^{T}\Delta_t = \mathbb{E}\left[\sum_{v=1}^{K}\|\mathbf{w}_v^1 - \mathbf{u}_v\|^2\right] - \mathbb{E}\left[\sum_{v=1}^{K}\|\mathbf{w}_v^{T+1} - \mathbf{u}_v\|^2\right]$$

Since $\mathbf{w}^1 = 0$ and $\|\mathbf{w}_v^{T+1} - \mathbf{u}_v\|^2$ is a positive quantity, we will get

$$\sum_{t=1}^{T}\Delta_t \leq \mathbb{E}\left[\sum_{v=1}^{K}\|\mathbf{u}_v\|^2\right]$$

which is same as

$$\sum_{t=1}^{T}\Delta_t \leq \sum_{v=1}^{K}\|\mathbf{u}_v\|^2$$

Comparing the upper and lower bounds on $\sum_{t=1}^{T} \Delta_t$, we get

$$\sum_{t=1}^{T} \mathbb{E}\left[\sum_{v=1}^{K} \left(\widetilde{l}_v^t\right)^2\right] \leq \frac{\left(R^2 + \frac{1}{2C}\right)^2}{\left(2K + \frac{1}{C}\right)} \left(\sum_{v=1}^{K} \|\mathbf{u}_v\|^2 + 2CR^2 \sum_{t=1}^{T} \mathbb{E}\left[\sum_{v=1}^{K} \left(\widetilde{l}_v^{*t}\right)^2\right]\right)$$

■

References

Dimitri P. Bertsekas. *Nonlinear Programming.* Athena Scientific, September 1999.

Shai Shalev-Shwartz and Yoram Singer. A primal-dual perspective of online learning algorithms. *Machine Learning*, 69:115–142, 10 2007.