Supplementary Material: Minimizing Adaptive Regret with One Gradient per Iteration

Paper ID: 1499

A Main Analysis

In this section, we provide the omitted proofs.

A.1 Proof of Theorem 1

We start our proof by investigating a new game between the learner and the adversary. In each round t, firstly the learner is asked to choose an action $\mathbf{x}_t$ in $\mathcal{D}$; next, an adversary *reveals a loss function* $L_t^e(\cdot)$, as defined in (8), instead of $f_t(\cdot)$ to the learner. Then the learner suffers a loss $L_t^e(\mathbf{x}_t)$. By definition, the SAR incurred under this new scenario is:

$$\begin{aligned}&\mathrm{SAR}_{L_1^e,\dots,L_T^e}(\tau)\\ =&\max_{I\subseteq[T],|I|=\tau}\left\{\sum_{t\in I}L_t^e(\mathbf{x}_t)-\sum_{t\in I}L_t^e(\mathbf{x}_*^{I,e})\right\}.\end{aligned}\tag{26}$$

where $\mathbf{x}_*^{I,e}=\arg\min_{\mathbf{x}\in\mathcal{D}}\sum_{t\in I}L_t^e(\mathbf{x})$. Our first observation is that for any algorithm, its SAR under the new problem, as defined in (26) , is an upper bound of that of the original problem, as defined in (1).

Lemma 3. *Given a parameter* τ, *let* $\mathrm{SAR}_f(\tau)$ *and* $\mathrm{SAR}_{L^e}(\tau)$ *be the SAR of any online learning algorithm which are defined in* (1) *and* (26) *respectively, then we have* $\forall\tau>0$,

$$\mathrm{SAR}_f(\tau)\le\mathrm{SAR}_{L^e}(\tau).\tag{27}$$

Proof of Lemma 3

Given τ, consider a time interval I_*, such that

$$I_*=\arg\max_{I\subseteq[T],|I|=\tau}\left\{\sum_{t\in I}f_t(\mathbf{x}_t)-\sum_{t\in I}f_t(\mathbf{x}_*^I)\right\}.\tag{28}$$

where $\mathbf{x}_*^I=\min_{\mathbf{x}\in\mathcal{D}}\sum_{t\in I}f_t(\mathbf{x})$. By Lemma 1, we have $\forall t\in I_*$,

$$\begin{aligned}f(\mathbf{x}_t)-f(\mathbf{x}_*^{I_*})\le&-\frac{\gamma}{2}(\mathbf{x}_t-\mathbf{x}_*^{I_*})^\top\nabla_t\nabla_t^\top(\mathbf{x}_t-\mathbf{x}_*^{I_*})\\&+\nabla_t^\top(\mathbf{x}_t-\mathbf{x}_*^{I_*})\end{aligned}$$

Combining with (8), we get

$$f(\mathbf{x}_t)-f(\mathbf{x}_*^{I_*})\le-L_t^e(\mathbf{x}_*^{I_*})$$

Since $L_t^e(\mathbf{x}_t)=0$, the equation above leads to

$$f(\mathbf{x}_t)-f(\mathbf{x}_*^{I_*})\le L_t^e(\mathbf{x}_t)-L_t^e(\mathbf{x}_*^{I_*})\tag{29}$$

By adding things together over I_*, we have

$$\begin{aligned}\mathrm{SAR}_f(\tau)=&\sum_{t\in I_*}\Big(f(\mathbf{x}_t)-f(\mathbf{x}_*^{I_*})\Big)\\\le&\sum_{t\in I_*}L_t^e(\mathbf{x}_t)-\sum_{t\in I_*}L_t^e(\mathbf{x}_*^{I_*})\\\le&\sum_{t\in I_*}L_t^e(\mathbf{x}_t)-\sum_{t\in I_*}L_t^e(\mathbf{x}_*^{I_*,e})\\\le&\max_{I\in[T],|I|=\tau}\sum_{t\in I}L_t^e(\mathbf{x}_t)-\sum_{t\in I}L_t^e(\mathbf{x}_*^{I,e})\\=&\,\mathrm{SAR}_{L^e}(\tau)\end{aligned}\tag{30}$$

The second inequality is derived by the definition of $\mathbf{x}_t^I$ and $\mathbf{x}_t^{I,e}$, and the third inequality is obtained by the definition of SAR_{L^e}. The proof can be finished by noticing that the (30) holds $\forall\tau>0$.

Next, we bound the SAR of MARSL-exp under the new problem. To start with, we prove the exp-concavity of $L_t^e(\cdot)$:

Lemma 4. *If Assumptions 1 and 2 hold, then* $L_t^e(\mathbf{u})$ *is* α' *exp-concave, where*

$$\alpha'=\gamma/\left(1+2\gamma DG+\gamma^2D^2G^2\right).$$

Proof of Lemma 4

Firstly we introduce the following lemma for exp-concave functions [Hazan, 2016]:

Lemma 5. *A twice differentiable function* f *is* α *exp-concave if and only if* $\nabla^2f_t(\mathbf{x})\succeq\alpha\nabla f(\mathbf{x})\nabla f(\mathbf{x})^\top$.

By (8), the gradient of $L_t^e(\mathbf{u})$ is

$$\nabla L_t^e(\mathbf{u})=\gamma\nabla_t\nabla_t^\top(\mathbf{u}-\mathbf{x}_t)+\nabla_t,\tag{31}$$

Thus,

$$\begin{aligned}&\nabla L_t^e(\mathbf{u})(\nabla L_t^e(\mathbf{u}))^\top\\=&\gamma^2\nabla_t\nabla_t^\top(\mathbf{u}-\mathbf{x}_t)(\mathbf{u}-\mathbf{x}_t)^\top\nabla_t\nabla_t^\top\\&+\gamma\nabla_t\nabla_t^\top(\mathbf{u}-\mathbf{x}_t)\nabla_t^\top+\gamma\nabla_t(\mathbf{u}-\mathbf{x}_t)^\top\nabla_t\nabla_t^\top\\&+\nabla_t\nabla_t^\top\\=&\gamma^2\nabla_t\left[\nabla_t^\top(\mathbf{u}-\mathbf{x}_t)\right]\left[(\mathbf{u}-\mathbf{x}_t)^\top\nabla_t\right]\nabla_t^\top\\&+\gamma\nabla_t\left[\nabla_t^\top(\mathbf{u}-\mathbf{x}_t)\right]\nabla_t^\top+\gamma\nabla_t\left[(\mathbf{u}-\mathbf{x}_t)^\top\nabla_t\right]\nabla_t^\top\\&+\nabla_t\nabla_t^\top\\\preceq&\left(\gamma^2D^2G^2+2\gamma DG+1\right)\nabla_t\nabla_t^\top\end{aligned}$$

On the other hand,

$$\nabla^2 L_t^e(\mathbf{u}) = \gamma \nabla_t \nabla_t^\top. \tag{32}$$

The proof is finished by using Lemma 5.

Lemma 4 indicates that the series of loss function $L_1^e(\cdot), ..., L_T^e(\cdot)$ are all exp-concave. Besides, by the definition of $\nabla L_t^e(\cdot)$, we have $\forall \mathbf{u} \in \mathcal{D}$:

$$||\nabla L_t^e(\mathbf{u})|| \leq \frac{5}{4} GD =: G^e \tag{33}$$

Thus, by simply applying AFLH to the new problem, according to Theorem 1.2 in [Hazan and Seshadhri, 2007], we can immediately obtain the following lemma:

Lemma 6. *Suppose Assumptions 1 and 2 hold and all functions $f_1, \ldots, f_T$ are α-exp-concave. Then, MARSL-exp achieves:*

$$\begin{aligned} &\mathrm{SAR}_{L_1^e, \ldots, f_T^e}(\tau) \\ &\leq (\log T + 1) \left(\left(\frac{4 + 5d}{\alpha'} + dG^e D \right) \log T + 1 \right) \\ &= O(\frac{1}{\alpha'} \log^2 T) \end{aligned} \tag{34}$$

The proof of Theorem 1 can be finished by combining Lemma 3 and Lemma 6.

A.2 Proof of Theorem 2

For strongly convex functions, following the proof of Theorem 1, we define the SAR under the loss function series $L_t^{sc}(\cdot), ..., L_t^{sc}(\cdot)$ as

$$\begin{aligned} &\mathrm{SAR}^T_{L_1^{sc}, \ldots, L_T^{sc}}(\tau) \\ &= \max_{I \subseteq [T], |I| = \tau} \left\{ \sum_{t \in I} L_t^{sc}(\mathbf{x}_t) - \sum_{t \in I} L_t^c(\mathbf{x}_*^{I,sc}) \right\}. \end{aligned} \tag{35}$$

where $\mathbf{x}_*^{I,sc} = \arg\min_{\mathbf{x} \in \mathcal{D}} \sum_{t \in I} L_t^{sc}(\mathbf{x})$. It can be easily verified that $L_1^{sc}(\cdot), \ldots, L_T^{sc}(\cdot)$ are all λ strongly convex. Besides, we have $\forall \mathbf{u} \in \mathcal{D}$,

$$||\nabla L_t^{sc}(\mathbf{u})|| = ||\lambda(\mathbf{x}_t - \mathbf{u}) - \nabla_t|| \leq \lambda D + G =: G^{sc}.$$

By Lemma 2, it implies that all surrogate losses are λ / G^{sc}-exp-concave. Thus, according to Lemma 3, we derive the following lemma:

Lemma 7. *Given a parameter τ, let $\mathrm{SAR}_f(\tau)$ and $\mathrm{SAR}_{L^{sc}}(\tau)$ be the SAR of any online learning algorithm which are defined in* (1) *and* (35) *respectively, then we have $\forall \tau > 0$,*

$$\mathrm{SAR}_f(\tau) \leq \mathrm{SAR}_{L^e}(\tau). \tag{36}$$

Next, we focus on the SAR of our algorithm under the new problem. Following Lemma 4.5 in [Hazan and Seshadhri, 2007], we have the following lemma:

Lemma 8. *Suppose Assumptions 1 and 2 hold and all functions $L_1^{sc}(\cdot), \ldots, L_T^{sc}(\cdot)$ are λ strongly convex. Then, MARSL-sc attains*

$$\begin{aligned} &\mathrm{SAR}_{L_1^{sc}, \ldots, L_T^{sc}}(\tau) \\ &\leq (\log T + 1) \left(\frac{(G^{sc})^2}{2\lambda} (\log T + 1) + 1 \right) \\ &= O\left(\log^2 T\right) \end{aligned} \tag{37}$$

The proof of Theorem 2 can be finished by combining Lemma 7 and Lemma 8.

A.3 Proof of Theorem 3

To begin with, following the proof of Theorem 1, we define the SAR under $L_t^c(\cdot), ..., L_t^c(\cdot)$ as

$$\begin{aligned} &\mathrm{SAR}_{L_1^c, \ldots, L_T^c}(\tau) \\ &= \max_{I \subseteq [T], |I| = \tau} \left\{ \sum_{t \in I} L_t^c(\mathbf{x}_t) - \sum_{t \in I} L_t^c(\mathbf{x}_*^{I,c}) \right\}. \end{aligned} \tag{38}$$

where $\mathbf{x}_*^{I,c} = \arg\min_{\mathbf{x} \in \mathcal{D}} \sum_{t \in I} L_t^c(\mathbf{x})$. The following Lemma illustrates the relations between (38) and (1):

Lemma 9. *Given a parameter τ, let $SAR_f(\tau)$ and $SAR_{L^c}(\tau)$ be the SAR which are defined in* (1) *and* (38)*, then we have $\forall \tau < T$,*

$$\mathrm{SAR}_f(\tau) \leq 2GD\left(SAR_{L^c}(\tau)\right). \tag{39}$$

Proof of Lemma 9

Given a parameter τ, consider a time interval I_*, which is defined in (28). By Definition 2, we have $\forall t \in I_*$,

$$f(\mathbf{x}_t) - f(\mathbf{x}_*^{I_*}) \leq \nabla f(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}_*^{I_*}). \tag{40}$$

By (20), we get

$$f(\mathbf{x}_t) - f(\mathbf{x}_*^{I_*}) \leq -2GD L_t^c(\mathbf{x}_*^{I_*}) + GD.$$

Since $L_t^c(\mathbf{x}_t) = 1/2$, the equation above leads to

$$f(\mathbf{x}_t) - f(\mathbf{x}_*^{I_*}) \leq 2GD\left(L_t^c(\mathbf{x}_t) - L_t^c(\mathbf{x}_*^{I_*})\right) \tag{41}$$

By summing the inequalities above in both side over I_*, we have

$$\begin{aligned} \mathrm{SAR}_f(\tau) &= \sum_{t \in I_*} \left(f(\mathbf{x}_t) - f(\mathbf{x}_*^{I_*}) \right) \\ &\leq 2GD \left(\sum_{t \in I_*} L_t^c(\mathbf{x}_t) - \sum_{t \in I_*} L_t^c(\mathbf{x}_*^{I_*}) \right) \\ &\leq 2GD \left(\sum_{t \in I_*} L_t^c(\mathbf{x}_t) - \sum_{t \in I_*} L_t^c(\mathbf{x}_*^{I_*,e}) \right) \\ &\leq 2GD \left(\max_{I \in [T], |I| = \tau} \sum_{t \in I} L_t^c(\mathbf{x}_t) - \sum_{t \in I} L_t^c(\mathbf{x}_*^{I,c}) \right) \\ &= \mathrm{SAR}_{L^c}(\tau) \end{aligned}$$

The proof can be finished by noticing that the inequalities above holds $\forall \tau > 0$.

Next, we bound the SAR of our algorithm under the new problem. To start with, it is easy to verify that $L_t^c(\mathbf{u})$ is convex, since it is a linear function with respect to $\mathbf{u}$. Besides, by the definition of $\nabla L_t^e(\cdot)$,we have $\forall \mathbf{u} \in \mathcal{D}$,

$$||\nabla L_t^e(\mathbf{u})||_2 \leq \frac{1}{2D} =: G^c. \tag{42}$$

Thus, by simply applying CBCE to the new problem, according to Theorem 2 in [Jun *et al.*, 2016], we can immediately obtain the following lemma:

Lemma 10. *Suppose Assumptions 1 and 2 hold, and all functions f_t are α exp-concave, then we have:*

$$
\begin{aligned}
&\mathrm{SAR}^{T}_{L_1^c,\ldots,L_T^c}(\tau) \\
&\leq \sqrt{\tau}\left(\frac{12D^2G^c}{\sqrt{2}-1} + 8\sqrt{7\log T + 5}\right) \\
&= O\left(\sqrt{\tau \log T}\right)
\end{aligned}
\tag{43}
$$

The Proof of Theorem 3 can be finished by combining Lemma 3 and Lemma 10.