Proceedings of Machine Learning Research vol vvv:1–3, 2024

Supplement to the paper: A Deep Learning Approach for Distributed Aggregative Optimization with Users' Feedback

Riccardo Brumali RICCARDO.BRUMALI@UNIBO.IT
Department of Electrical, Electronic and Information Engineering, Alma Mater Studiorum - Università di Bologna, Bologna, 40136, Italy.

Guido Carnevale GUIDO.CARNEVALE@UNIBO.IT
Department of Electrical, Electronic and Information Engineering, Alma Mater Studiorum - Università di Bologna, Bologna, 40136, Italy.

Giuseppe Notarstefano GIUSEPPE.NOTARSTEFANO@UNIBO.IT
Department of Electrical, Electronic and Information Engineering, Alma Mater Studiorum - Università di Bologna, Bologna, 40136, Italy.

1. Introduction

This external work contains the proofs of the paper lemmas. For the sake of reading, we report the assumptions and the preliminary considerations used in the proofs.

2. Assumptions and preliminary considerations

We have considered the following Assumptions.

Assumption 1 (Neural Network) *For each time instant $t > 0$, consider the estimate $\hat{U}_{i,t}(x_i, s_i)$ given from the network of agent i. Then, there exist $b_{1,t}, b_{2,t} > 0$ such that*

$$\left\|\nabla_1 \hat{U}_{i,t}(x_i, s_i) - \nabla_1 U_{i,t}(x_i, s_i)\right\| \leq b_{1,t} \ , \qquad \left\|\nabla_2 \hat{U}_{i,t}(x_i, s_i) - \nabla_2 U_{i,t}(x_i, s_i)\right\| \leq b_{2,t},$$

for all $x_i \in X_i$, $s_i \in \mathbb{R}^{n_\sigma}$, and $i \in \{1, \ldots, N\}$.

Assumption 2 (Global Problem Convexity) *The set X_i is closed, convex, and nonempty for all $i \in \{1, \ldots, N\}$, while the function $f_t(x, \sigma_t(x))$ is μ-strongly convex for all $t \geq 0$, with $\mu > 0$.*

Assumption 3 (Function Regularity) *For all $t \geq 0$, each function $f_{i,t}(x, \sigma_t(x))$ is differentiable and $f_t(x, \sigma_t(x))$ has L_1-Lipschitz continuous gradient, i.e.*

$$\left\|\nabla f_t(x, \sigma_t(x)) - \nabla f_t(x', \sigma_t(x'))\right\| \leq L_1 \left\|x - x'\right\| .$$

Further, for all $t \geq 0$, $G_t(x, s)$ and $\nabla_2 f_t(x, s)$ are Lipschitz continuous, respectively with constant $L_1, L_2 > 0$. Moreover, for all $i \in \{1, \ldots, N\}$ and $t \geq 0$, the function $\phi_{i,t}(x_i)$ is differentiable and L_3-Lipschitz continuous. Finally, assume both η_t and ω_t finite for all $t \geq 0$.

Assumption 2 implies the existence of a unique optimal solution $x_t^\star \in \mathbb{R}^n$ for all $t > 0$.

Assumption 4 (Communication Graph) *$\mathcal{G}$ is connected and $\mathcal{W}$ is doubly stochastic.*

© 2024 R. Brumali, G. Carnevale & G. Notarstefano.

Then we recall that ANN-PAT can be seen as a perturbed version of PAT for which the convergence properties have been already proved. Therefore, we have written the updates law of ANN-PAT as functions of the updates of PAT. In particular

$$x_{t+1} = n_x(x_t, s_t, r_t) + \delta(P_X[x_t - \alpha\hat{d}_t(x_t, s_t, r_t)] - P_X[x_t - \alpha d_t(x_t, s_t, r_t)]) \tag{1a}$$

$$r_{t+1} = n_r(x_t, s_t, r_t) + e_{y,t+1}(x_{t+1}, s_{t+1}) - e_{y,t}(x_t, s_t). \tag{1b}$$

With the same idea we have defined

$$\bar{r}_{t+1} = n_{\bar{r}}(x_{t+1}, s_{t+1}) + \bar{e}_{y,t+1}(x_{t+1}, s_{t+1}). \tag{2}$$

3. Proof of Lemma 1

For the sake of reading, we report the statement of the lemma.

Lemma 1 *Let Assumptions 1, 2, 3, and 4. If* $\alpha \leq \frac{1}{L_1}$*, then*

$$\left\|x_{t+1} - x^\star_{t+1}\right\| \leq (1 - \delta\mu\alpha)\left\|x_t - x^\star_t\right\| + \delta\alpha L_1\left\|s_t - \mathbf{1}\bar{s}_t\right\| + \delta\alpha L_3\left\|r_t - \mathbf{1}\bar{y}_t\right\| + \zeta_t\delta\alpha b_{1,t}.$$

Proof By using (1a), we can write

$$\begin{aligned}
\left\|x_{t+1} - x^\star_{t+1}\right\| &= \left\|n_x(x_t, s_t, r_t) - x^\star_{t+1} + \delta\Big(P_X[x_t - \alpha\hat{d}_t(x_t, s_t, r_t)] - P_X[x_t - \alpha d_t(x_t, s_t, r_t)]\Big)\right\| \\
&\overset{(a)}{\leq} \left\|n_x(x_t, s_t, r_t) - x^\star_{t+1}\right\| + \left\|\delta\Big(P_X[x_t - \alpha\hat{d}_t(x_t, s_t, r_t)] - P_X[x_t - \alpha d_t(x_t, s_t, r_t)]\Big)\right\| \\
&\overset{(b)}{\leq} \left\|n_x(x_t, s_t, r_t) - x^\star_{t+1}\right\| + \left\|\delta\Big(x_t - \alpha\hat{d}_t(x_t, s_t, r_t) - x_t + \alpha d_t(x_t, s_t, r_t)\Big)\right\| \\
&\overset{(c)}{\leq} \left\|n_x(x_t, s_t, r_t) - x^\star_{t+1}\right\| + \delta\alpha b_{1,t},
\end{aligned}$$

where in (a) we use the triangular inequality, in (b) the non-expansiveness of the projection (see Bertsekas (2015)), and in (c) Assumption 1 with the definition of $d_t(x_t, s_t, r_t)$ and $\hat{d}_t(x_t, s_t, r_t)$. The proof follows by applying (Carnevale et al., 2022, Lemma 1) in which they define a bound for $\left\|n_x(x_t, s_t, r_t) - x^\star_{t+1}\right\|$. ■

4. Proof of Lemma 3

For the sake of reading, we report the statement of the lemma.

Lemma 3 *Let Assumptions 1,2, 3, and 4. Then*

$$\begin{aligned}
\left\|r_{t+1} - \mathbf{1}\bar{r}_{t+1}\right\| \leq\ & (\Lambda + \delta\alpha L_3(L_2 + L_2 L_3))\left\|r_t - \bar{r}_t\right\| \\
& + \delta(2 + \alpha L_1 + \alpha L_1 L_3)(L_2 + L_2 L_3)\left\|x_t - x^\star_t\right\| \\
& + (\delta\alpha L_1(L_2 + L_2 L_3) + 2L_2)\left\|s_t - \mathbf{1}\bar{s}_t\right\| + L_2\omega_t + \eta_t + (1 + \sqrt{N})b_{2,t+1} + b_{2,t},
\end{aligned}$$

where Λ *is the maximum eigenvalue of the matrix* $W - \frac{\mathbf{1}\mathbf{1}^\top}{N}$.

Proof By using (1b) and (2) we can write

$$\begin{aligned}
\|r_{t+1} - \mathbf{1}\bar{r}_{t+1}\| &= \|n_r(x_t, s_t, r_t) - \mathbf{1}n_{\bar{r}}(x_{t+1}, s_{t+1}) + e_{y,t+1}(x_{t+1}, s_{t+1}) - \mathbf{1}\bar{e}_{y,t+1}(x_{t+1}, s_{t+1}) \\
&\quad - e_{y,t}(x_t, s_t)\| \\
&\overset{(a)}{\leq} \|n_r(x_t, s_t, r_t) - \mathbf{1}n_{\bar{r}}(x_{t+1}, s_{t+1})\| + \|e_{y,t+1}(x_{t+1}, s_{t+1})\| \\
&\quad + \|\mathbf{1}\bar{e}_{y,t+1}(x_{t+1}, s_{t+1})\| + \|e_{y,t}(x_t, s_t)\| \\
&\overset{(b)}{\leq} \|n_r(x_t, s_t, r_t) - \mathbf{1}n_{\bar{r}}(x_{t+1}, s_{t+1})\| + \|e_{y,t+1}(x_{t+1}, s_{t+1})\| \\
&\quad + \|\mathbf{1}\| \, \|\bar{e}_{y,t+1}(x_{t+1}, s_{t+1})\| + \|e_{y,t}(x_t, s_t)\| \\
&\overset{(c)}{\leq} \|n_r(x_t, s_t, r_t) - \mathbf{1}n_{\bar{r}}(x_{t+1}, s_{t+1})\| + \|e_{y,t+1}(x_{t+1}, s_{t+1})\| \\
&\quad + \frac{\sqrt{N}}{N} \sum_{i=1}^{N} \left\|e^i_{y,t+1}(x_{i,t+1}, s_{i,t+1})\right\| + \|e_{y,t}(x_t, s_t)\| \\
&\overset{(d)}{\leq} \|n_r(x_t, s_t, r_t) - \mathbf{1}n_{\bar{r}}(x_{t+1}, s_{t+1})\| + (1 + \sqrt{N})b_{2,+}b_{2,t},
\end{aligned}$$

where in (a) we have used the triangular inequality, in (b) the Cauchy-Schwarz inequality, in (c) once again the triangular inequality and in (d) Assumption 1. Finally, the proof concludes considering (Carnevale et al., 2022, Lemma 1) which bounds $\|n_r(x_t, s_t, r_t) - \mathbf{1}n_{\bar{r}}(x_{t+1}, s_{t+1})\|$. ■

References

Dimitri Bertsekas. *Convex optimization algorithms*. Athena Scientific, 2015.

Guido Carnevale, Andrea Camisa, and Giuseppe Notarstefano. Distributed online aggregative optimization for dynamic multi-robot coordination. *IEEE Transactions on Automatic Control*, 2022.