Supplementary material for the paper: A coordinator-driven communication reduction scheme for distributed optimization using the projected gradient method

Giorgos Stathopoulos and Colin N. Jones *

Theorem 1 *Let N agents update with probabilities p_i and $p_{\min} = \min_i p_i$. Let $\nu = 1 - \sqrt{(1 - 2\gamma\mu + \mu\gamma^2 L)}$, where L is the Lipschitz constant of ∇h and μ the strong convexity constant of h, while $\gamma < 2/L$. If z^* is the unique optimizer of Problem 7 in [1], for any time instant $k > K$, the sequence $\{z^k\}$ generated by Algorithm 1 satisfies*

$$\begin{aligned} &\mathbb{E}[\|z^{k+K} - z^*\|^2] \leq \left(1 - \tfrac{\rho(\mu-\epsilon)}{N}\right)^k \mathbb{E}[\|z^K - z^*\|^2] \\ &+ \tfrac{\rho}{N}\left(\tfrac{1}{\epsilon} + \tfrac{\rho(1+\delta)}{N p_{\min}\delta}\right) \textstyle\sum_{j=1}^k \left(1 - \tfrac{\rho(\mu-\epsilon)}{N}\right)^{k-j} \mathbb{E}[\|e^{K-1+j}\|^2] \ , \end{aligned} \tag{1}$$

for $\rho \in (0, Np_{\min}/(2(1+\delta)))$, $\delta > 0$, $\nu > \epsilon > 0$, and $e^k = (e_1^k, \dots, e_N^k) \in \mathbb{R}^{Nn}$ the vector that is constituted of the components $e_i^k = \gamma(\nabla\phi_i^\gamma(v^k) - g_i^k)$, $i = 1, \dots, N$, while $e^K = 0$.

Proof 1 *The key point is to observe that the approximate iteration $z_i^{k+1} = v_i^k - \gamma g_i^k$ can be expressed as an inexact projected gradient iteration. To this end, we introduce the error sequence $\{e^k\}$ so as to write*

$$e_i^k + \mathcal{P}_{\mathcal{Z}_i}\left(v_i^k\right) = v_i^k - \gamma g_i^k \ , \tag{2}$$

while

$$\mathcal{P}_{\mathcal{Z}_i}\left(v_i^k\right) = v_i^k - \gamma\nabla\phi_i^\gamma(v_i^k) \ . \tag{3}$$

Substituting (3) *in* (2) *we have that*

$$e_i^k = \gamma(\nabla\phi_i^\gamma(v_{i,k}) - g_i^k) \ . \tag{4}$$

Using the error (4)*, the randomized coordinate descent iteration can be expressed as*

$$\begin{cases} z_{i_k}^{k+1} &= z_{i_k}^k + \eta^k\left(e_{i_k}^k + \mathcal{P}_{\mathcal{Z}_{i_k}}\left(z_{i_k}^k - \gamma\nabla_{i_k} h(z^k)\right) - z_{i_k}^k\right) \\ z_{i\neq i_k}^{k+1} &= z_{i\neq i_k}^k \ , \end{cases}$$

*The authors are with Laboratoire d'Automatique, EPFL, CH-1015 Lausanne, Switzerland, {georgios.stathopoulos,colin.jones}@epfl.ch

or, more compactly, as

$$z^{k+1} = z^k + \eta^k U_{i_k}\Big(\mathcal{P}_{\mathcal{Z}}\left(z^k - \gamma\nabla h(z^k)\right) - z^k + e^k\Big) \ . \tag{5}$$

The matrix $U_{i_k} : \mathbb{R}^{Nn} \mapsto \mathbb{R}^{Nn}$ *is drawn from a set of orthogonal projection matrices* $\{U_i\}_{i=1}^N$ *such that* $U_i : z \mapsto (0, \ldots, 0, z_i, 0, \ldots, 0)$, $i = 1, \ldots, N$ *and* $\sum_{i=1}^N U_i = I$. *Consequently,* U_{i_k} *isolates the* i_k^{th} *component of its argument, thus it updates the corresponding component of* z*, while the other components (agents) are set to their previous values. The projection operator* $\mathcal{P}_{\mathcal{Z}}$ *is defined as* $\mathcal{P}_{\mathcal{Z}} = \mathcal{P}_{\mathcal{Z}_1} \times \mathcal{P}_{\mathcal{Z}_2} \times \cdots \times \mathcal{P}_{\mathcal{Z}_N}$.

Equation (5) *is an instance of a more general inexact fixed-point iteration. By introducing the operator*

$$T : \mathbb{R}^n \mapsto \mathbb{R}^n, \quad T := \mathcal{P}_{\mathcal{Z}}\left(I - \gamma\nabla h\right)$$

and

$$S : \mathbb{R}^n \mapsto \mathbb{R}^n, \quad S = I - T \ .$$

equation (5) *can be written as*

$$z^{k+1} = z^k + \eta^k U_{i_k}(Tz^k - z^k + e^k) = z^k - \eta^k U_{i_k} s^k \ , \tag{6}$$

where $s^k = Sz^k - e^k$*, and* e^k *is given by* (4)*. We set the relaxation parameter to* $\eta^k = \frac{\rho}{Np_{i_k}}$*, where* $\rho > 0$ *will be bounded from above later on.*

Our purpose is to bound the distance of z^{k+1} *to the fixed point* z^* *as a function of* $\|z^k - z^*\|$ *and* $\|e^k\|$*, always in expectation. We thus introduce* $Z^k = \left\{z^0, z^1, \ldots, z^k\right\}$*, and by taking the conditional expectation and squaring* (6)*, we get*

$$\begin{aligned}
\mathbb{E}[\|z^{k+1} - z^*\|^2 \mid Z^k] &= \|z^k - z^*\|^2 - 2\frac{\rho}{N}\mathbb{E}[\langle z^k - z^*, \frac{1}{p_{i_k}} U_{i_k} s^k\rangle \mid Z^k] + \frac{\rho^2}{N^2}\mathbb{E}[\|\frac{1}{p_{i_k}} U_{i_k} s^k\|^2 \mid Z^k] \\
&= \|z^k - z^*\|^2 - 2\frac{\rho}{N}\sum_{i=1}^N p_i \langle z^k - z^*, \frac{1}{p_i} U_i s^k\rangle + \frac{\rho^2}{N^2}\sum_{i=1}^N p_i \langle \frac{1}{p_i} U_i s^k, \frac{1}{p_i} U_i s^k\rangle \\
&\le \|z^k - z^*\|^2 - 2\frac{\rho}{N}\langle z^k - z^*, s^k\rangle + \frac{\rho^2}{N^2 p_{\min}}\|s^k\|^2 \ ,
\end{aligned} \tag{7}$$

where the second equality follows from the definition of the expectation and the third one from the fact that U_i *is an orthogonal projection operator.*

Let us now analyze the second and third term in (7).

- *Bound* $-2\frac{\rho}{N}\langle z^k - z^*, s^k\rangle$*: From the definition of* $s^k = Sz^k - e^k$*, it holds that*

$$\langle z^k - z^*, s^k\rangle = \langle z^k - z^*, Sz^k\rangle - \langle z^k - z^*, e^k\rangle \ . \tag{8}$$

We will now upper-bound the resulting inner product terms. In order to do so, we must use both the Lipschitz continuity of ∇h *and the strong convexity of* h.

Lemma 1 *Let $S = I - \mathcal{P}_{\mathcal{Z}}(I - \gamma\nabla h)$ as defined above. Then*

$$\langle z^k - z^*, Sz^k\rangle \geq \frac{1}{2}\|Sz^k\|^2 \ .$$

Proof 2 *If $T = \mathcal{P}_{\mathcal{Z}}(I - \gamma\nabla h)$ is a nonexpansive operator, then the property holds for $S = I - T$ from [3, Proposition 4.33]. Nonexpansivity of T can be easily shown (see, e.g., [2, Proposition 2.2]), from where the result follows.*

Lemma 2 *Denoting as L be the Lipschitz continuous gradient constant of h and μ its strong convexity modulus, it holds that*

$$\langle z^k - z^*, Sz^k\rangle \geq \nu\|z^k - z^*\|^2 \ ,$$

where $\nu = 1 - \sqrt{(1 - 2\gamma\mu + \mu\gamma^2 L)}$ for $\gamma < 2/L$.

Proof 3 *From [3, Example 22.5] we have that if T is β-Lipschitz continuous for some $\beta \in [0, 1)$ then $I - T$ is $(1-\beta)$-strongly monotone. It is proven in [2, Proposition 2.2] that $\|Tz - Tz^*\| \leq \sqrt{(1 - 2\gamma\mu + \mu\gamma^2 L)}\|z - z^*\|$ for $\gamma < 2/L$, so T is β-Lipschitz continuous with $\beta = \sqrt{(1 - 2\gamma\mu + \mu\gamma^2 L)}$, which concludes the proof.*

Using Lemmata 1 and 2 we get

$$-2\frac{\rho}{N}\langle z^k - z^*, Sz^k\rangle \leq -\frac{\rho\nu}{N}\|z^k - z^*\|^2 - \frac{\rho}{2N}\|Sz^k\|^2 \ . \tag{9}$$

For the second inner product term in (8) *we can easily derive the bound*

$$2\frac{\rho}{N}\langle z^k - z^*, e^k\rangle \leq 2\frac{\rho}{N}\|z^k - z^*\|\|e^k\| \ . \tag{10}$$

Equations (9) *and* (10) *result in the bound*

$$-2\frac{\rho}{N}\langle z^k - z^*, s^k\rangle \leq -\frac{\rho\nu}{N}\|z^k - z^*\|^2 - \frac{\rho}{2N}\|Sz^k\|^2 + 2\frac{\rho}{N}\|z^k - z^*\|\|e^k\| \ . \tag{11}$$

- *Bound $\frac{\rho^2}{N^2 p_{\min}}\|s^k\|^2$: Using again the definition of s^k, we have that*

$$\begin{aligned}\|s^k\|^2 &= \|Sz^k\|^2 + \|e^k\|^2 - 2\langle Sz^k, e^k\rangle \\ &\leq \|Sz^k\|^2 + \|e^k\|^2 + \frac{\delta}{p_{\min}}\|Sz^k\|^2 + \frac{1}{\delta p_{\min}}\|e^k\|^2 \ ,\end{aligned} \tag{12}$$

where the inner product term was bounded by employing Young's inequality.[1] *We finally get the bound:*

$$\frac{\rho^2}{N^2 p_{\min}}\|s^k\|^2 \leq \frac{\rho^2}{N^2 p_{\min}}(1+\delta)\|Sz^k\|^2 + \frac{\rho^2}{N^2 p_{\min}\delta}(1+\delta)\|e^k\|^2 \ . \tag{13}$$

[1] For two nonnegative real numbers x and y, it holds that $xy \leq \frac{\delta x^2}{2} + \frac{y^2}{2\delta}$ for every $\delta > 0$.

Using (11) *and* (13), *inequality* (7) *can be written as*

$$\begin{aligned}\mathbb{E}[\|z^{k+1}-z^*\|^2 \mid Z^k] \le \|z^k-z^*\|^2 &- \frac{\rho\nu}{N}\|z^k-z^*\|^2 \\ &+ \frac{\rho}{N}\left(\frac{\rho(1+\delta)}{Np_{\min}}-\frac{1}{2}\right)\|Sz^k\|^2 \\ &+ 2\frac{\rho}{N}\|z^k-z^*\|\|e^k\| + \frac{\rho^2}{N^2p_{\min}\delta}(1+\delta)\|e^k\|^2 \ . \end{aligned} \tag{14}$$

The third term in the sum can be eliminated by asumming that

$$\frac{\rho(1+\delta)}{Np_{\min}}-\frac{1}{2} < 0 \Rightarrow \rho < \frac{Np_{\min}}{2(1+\delta)} \ , \tag{15}$$

which gives rise to the inequality

$$\mathbb{E}[\|z^{k+1}-z^*\|^2 \mid Z^k] \le \|z^k-z^*\|^2 - \frac{\rho\nu}{N}\|z^k-z^*\|^2 + 2\frac{\rho}{N}\|z^k-z^*\|\|e^k\| + \frac{\rho^2}{N^2p_{\min}\delta}(1+\delta)\|e^k\|^2 \ . \tag{16}$$

The complicating term on the right hand side can be eliminated by using once more Young's inequality, i.e.,

$$\begin{aligned} 2\frac{\rho}{N}\|z^k-z^*\|\|e^k\| &\le 2\frac{\rho}{N}\left(\frac{\epsilon}{2}\|z^k-z^*\|^2 + \frac{1}{2\epsilon}\|e^k\|^2\right) \\ &= \frac{\rho\epsilon}{N}\|z^k-z^*\|^2 + \frac{\rho}{N\epsilon}\|e^k\|^2 \ . \end{aligned}$$

Using the above in (16) *and taking the expectation in both sides, we recover the inequality*

$$\mathbb{E}[\|z^{k+1}-z^*\|^2] \le \left(1-\frac{\rho(\nu-\epsilon)}{N}\right)\mathbb{E}[\|z^k-z^*\|^2] + \frac{\rho}{N}\left(\frac{1}{\epsilon}+\frac{\rho(1+\delta)}{Np_{\min}\delta}\right)\mathbb{E}[\|e^k\|^2] \ ,$$

for $\rho \in (0, Np_{\min}/(2(1+\delta)))$ *and any* $\delta > 0, \epsilon > 0$, *which concludes the proof.*

References

[1] G. Stathopoulos and C. N. Jones *A coordinator-driven communication reduction scheme for distributed optimization using the projected gradient method*. Proceedings of the 17th IEEE European Control Conference, 2018.

[2] Peng, Z. and Xu, Y. and Yan, M. and Yin, W. *ARock: an Algorithmic Framework for Asynchronous Parallel Coordinate Updates*. SIAM Journal on Scientific Computing, 2016.

[3] H.H. Bauschke and P.L. Combettes *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.