A Supplementary Material

A.1 The proposed DAve-QN method with exact time indices

Algorithm 2 Illustration of the DAve-QN Algorithm

Master:

Initialize $\mathbf{x}$, $\mathbf{B}_i$, $\mathbf{g} = \sum_{i=1}^n \nabla f_i(\mathbf{x})$, $\mathbf{B}^{-1} = (\sum_{i=1}^n \mathbf{B}_i)^{-1}, \mathbf{u} = \sum_{i=1}^n \mathbf{B}_i\mathbf{x}$,
for $t = 1$ **to** $T - 1$ **do**
  If a worker sends an update:
  Receive $\Delta\mathbf{u}$, $\mathbf{y}$, $\mathbf{q}$, α, β from it
  $\mathbf{u} = \mathbf{u} + \Delta\mathbf{u}$, $\mathbf{g} = \mathbf{g} + \mathbf{y}$, $\mathbf{v} = (\mathbf{B})^{-1}\mathbf{y}$
  $\mathbf{U} = (\mathbf{B})^{-1} - \frac{\mathbf{v}\mathbf{v}^\top}{\alpha + \mathbf{v}^\top\mathbf{y}}$
  $\mathbf{w} = \mathbf{U}\mathbf{q}$, $(\mathbf{B})^{-1} = \mathbf{U} + \frac{\mathbf{w}\mathbf{w}^\top}{\beta - \mathbf{q}^\top\mathbf{w}}$
  $\mathbf{x} = (\mathbf{B})^{-1}(\mathbf{u} - \mathbf{g})$
  Send $\mathbf{x}$ to the worker in return
end
Interrupt all workers
Output $\mathbf{x}^T$

Worker i:

Initialize $\mathbf{x}_i = \mathbf{x}$, $\mathbf{B}_i$
while *not interrupted by master* **do**
  Receive $\mathbf{x}$
  $\mathbf{s}_i = \mathbf{x} - \mathbf{z}_i$
  $\mathbf{y}_i = \nabla f_i(\mathbf{x}) - \nabla f_i(\mathbf{z}_i)$
  $\mathbf{q}_i = \mathbf{B}_i\mathbf{s}_i$
  $\alpha = \mathbf{y}_i^\top\mathbf{s}_i$
  $\beta = \mathbf{s}_i^\top\mathbf{B}_i^t\mathbf{s}_i$
  $\mathbf{u} = \mathbf{B}_i\mathbf{z}_i$
  $\mathbf{B}_i = \mathbf{B}_i + \frac{\mathbf{y}_i\mathbf{y}_i^\top}{\alpha} - \frac{\mathbf{q}_i\mathbf{q}_i^\top}{\beta}$
  $\Delta\mathbf{u} = \mathbf{B}_i\mathbf{x} - \mathbf{u}$
  $\mathbf{z}_i = \mathbf{x}$
  Send $\Delta\mathbf{u}, \mathbf{y}_i, \mathbf{q}_i, \alpha, \beta$ to the master
end

A.2 Proof of Lemma 1

Proof. To verify the claim, we need to show that $\mathbf{u}^t = \sum_{i=1}^n \mathbf{B}_i^t\mathbf{z}_i^t$ and $\mathbf{g}^t = \sum_{i=1}^n \nabla f_i(\mathbf{z}_i^t)$. They follow from our delayed vectors notation $\mathbf{z}_i^t = \mathbf{z}_i^{t-d_i^t}$ and how $\Delta\mathbf{u}^{t-d_i^t}$ and $\mathbf{y}_i^{t-d_i^t}$ are computed by the corresponding worker. □

A.3 Proof of Lemma 2

To prove the claim in Lemma 2 we first prove the following intermediate lemma using the result of Lemma 5.2 in Broyden et al. (1973b).

Lemma 4. *Consider the proposed method outlined in Algorithm 1. Let* $\mathbf{M}$ *be a nonsingular symmetric matrix such that*

$$\|\mathbf{M}\mathbf{y}_i^t - \mathbf{M}^{-1}\mathbf{s}_i^t\| \le \beta\|\mathbf{M}^{-1}\mathbf{s}_i^t\|, \tag{21}$$

for some $\beta \in [0, 1/3]$ *and vectors* $\mathbf{s}_i^t$ *and* $\mathbf{y}_i^t$ *in* $\mathbb{R}^p$ *with* $\mathbf{s}_i^t \neq \mathbf{0}$*. Let's denote* i *as the index that has been updated at time* t*. Then, there exist positive constants* α, α_1*, and* α_2 *such that, for any symmetric* $\mathbf{A} \in \mathbb{R}^{p\times p}$ *we have,*

$$\|\mathbf{B}_i^t - \mathbf{A}\|_\mathbf{M} \le \left[(1-\alpha\theta^2)^{1/2} + \alpha_1\frac{\|\mathbf{M}\mathbf{y}_i^{t-D_i^t} - \mathbf{M}^{-1}\mathbf{s}_i^{t-D_i^t}\|}{\|\mathbf{M}^{-1}\mathbf{s}_i^{t-D_i^t}\|}\right]\|\mathbf{B}_i^{t-D_i^t} - \mathbf{A}\|_\mathbf{M} + \alpha_2\frac{\|\mathbf{y}_i^{t-D_i^t} - \mathbf{A}\mathbf{s}_i^{t-D_i^t}\|}{\|\mathbf{M}^{-1}\mathbf{s}_i^{t-D_i^t}\|}, \tag{22}$$

where $\alpha = (1-2\beta)/(1-\beta^2) \in [3/8, 1]$, $\alpha_1 = 2.5(1-\beta)^{-1}$, $\alpha_2 = 2(1+2\sqrt{p})\|\mathbf{M}\|_\mathbf{F}$*, and*

$$\theta = \frac{\|\mathbf{M}(\mathbf{B}_i^{t-D_i^t} - \mathbf{A})\mathbf{s}_i^{t-D_i^t}\|}{\|\mathbf{B}_i^{t-D_i^t} - \mathbf{A}\|_\mathbf{M}\|\mathbf{M}^{-1}\mathbf{s}_i^{t-D_i^t}\|} \quad \text{for } \mathbf{B}_i^{t-D_i^t} \neq \mathbf{A}, \quad \theta = 0 \quad \text{for } \mathbf{B}_i^{t-D_i^t} = \mathbf{A}. \tag{23}$$

Proof. By definition of delays d_i^t, the function f_i was updated at step $t - d_i^t$ and $\mathbf{B}_i^{t-1}$ is equal to $\mathbf{B}_i^{t-D_i^t}$. Considering this observation and the result of Lemma 5.2 in Broyden et al. (1973b), the claim follows. □

Note that the result in Lemma 4 characterizes an upper bound on the difference between the Hessian approximation matrices $\mathbf{B}_i^t$ and $\mathbf{B}_i^{t-D_i^t}$ and any positive definite matrix $\mathbf{A}$. Let us show that matrices $\mathbf{M} = \nabla^2 f_i(\mathbf{x}^*)^{-1/2}$ and $\mathbf{A} = \nabla^2 f_i(\mathbf{x}^*)$ satisfy the conditions of Lemma 4. By strong convexity of f_i we have $\|\nabla^2 f_i(\mathbf{x}^*)^{1/2}\mathbf{s}_i^{t-D_i^t}\| \geq \sqrt{\mu}\|\mathbf{s}_i^{t-D_i^t}\|$. Combined with Assumption 2, it gives that

$$\frac{\|\mathbf{y}_i^{t-D_i^t} - \nabla^2 f_i(\mathbf{x}^*)\mathbf{s}_i^{t-D_i^t}\|}{\|\nabla^2 f_i(\mathbf{x}^*)^{1/2}\mathbf{s}_i^{t-D_i^t}\|} \leq \frac{\tilde{L}\|\mathbf{s}_i^{t-D_i^t}\|\max\{\|\mathbf{z}_i^{t-D_i^t} - \mathbf{x}^*\|, \|\mathbf{z}_i^t - \mathbf{x}^*\|\}}{\sqrt{\mu}\|\mathbf{s}_i^{t-D_i^t}\|} = \frac{\tilde{L}}{\sqrt{\mu}}\sigma_i^t \tag{24}$$

This observation implies that the left hand side of the condition in (21) for $\mathbf{M} = \nabla^2 f_i(\mathbf{x}^*)^{-1/2}$ is bounded above by

$$\frac{\|\mathbf{M}\mathbf{y}_i^{t-D_i^t} - \mathbf{M}^{-1}\mathbf{s}_i^{t-D_i^t}\|}{\|\mathbf{M}^{-1}\mathbf{s}_i^{t-D_i^t}\|} \leq \frac{\|\nabla^2 f_i(\mathbf{x}^*)^{-1/2}\|\|\mathbf{y}_i^{t-D_i^t} - \nabla^2 f_i(\mathbf{x}^*)\mathbf{s}_i^{t-D_i^t}\|}{\|\nabla^2 f_i(\mathbf{x}^*)^{1/2}\mathbf{s}_i^{t-D_i^t}\|} \leq \frac{\tilde{L}}{\mu}\sigma_i^t \tag{25}$$

Thus, the condition in (21) is satisfied since $\tilde{L}\sigma_i^t/\mu < 1/3$. Replacing the upper bounds in (24) and (25) into the expression in (22) implies the claim in (20) with

$$\beta = \frac{\tilde{L}}{\mu}\sigma_i^t,\ \alpha = \frac{1-2\beta}{1-\beta^2},\ \alpha_3 = \frac{5\tilde{L}}{2\mu(1-\beta)},\ \alpha_4 = \frac{2(1+2\sqrt{p})\tilde{L}}{\sqrt{\mu}}\|\nabla^2 f_i(\mathbf{x}^*)^{-\frac{1}{2}}\|_{\mathbf{F}}, \tag{26}$$

and the proof is complete.

A.4 Proof of Lemma 3

We first state the following result from Lemma 6 in Mokhtari et al. (2018a), which shows an upper bound for the error $\|\mathbf{x}^t - \mathbf{x}^*\|$ in terms of the gap between the delayed variables $\mathbf{z}_i^t$ and the optimal solution $\mathbf{x}^*$ and the difference between the Newton direction $\nabla^2 f_i(\mathbf{x}^*)\left(\mathbf{z}_i^t - \mathbf{x}^*\right)$ and the proposed quasi-Newton direction $\mathbf{B}_i^t\left(\mathbf{z}_i^t - \mathbf{x}^*\right)$.

Lemma 5. *If Assumptions 1 and 2 hold, then the sequence of iterates generated by Algorithm 1 satisfies*

$$\|\mathbf{x}^t - \mathbf{x}^*\| \leq \frac{\tilde{L}\Gamma^t}{n}\sum_{i=1}^n \left\|\mathbf{z}_i^t - \mathbf{x}^*\right\|^2 + \frac{\Gamma^t}{n}\sum_{i=1}^n \left\|\left(\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right)\left(\mathbf{z}_i^t - \mathbf{x}^*\right)\right\|, \tag{27}$$

where $\Gamma^t := \|((1/n)\sum_{i=1}^n \mathbf{B}_i^t)^{-1}\|$.

We use the result in Lemma 5 to prove the claim of Lemma 3. We will prove the claimed convergence rate in Lemma 3together with an additional claim

$$\left\|\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} \leq 2\delta$$

by inductions on m and on $t \in [T_m, T_{m+1})$. The base case of our induction is $m = 0$ and $t = 0$, which is the initialization step, so let us start with it.

Since all norms in finite dimensional spaces are equivalent, there exists a constant $\eta > 0$ such that $\|\mathbf{A}\| \leq \eta\|\mathbf{A}\|_{\mathbf{M}}$ for all $\mathbf{A}$. Define $\gamma := 1/\mu$ and $d := \max_m(T_{m+1} - T_m)$, and assume that $\epsilon(r) = \epsilon$ and $\delta(r) = \delta$ are chosen such that

$$(2\alpha_3\delta + \alpha_4)\frac{d\epsilon}{1-r} \leq \delta \quad \text{and} \quad \gamma(1+r)[\tilde{L}\epsilon + 2\eta\delta] \leq r, \tag{28}$$

where α_3 and α_4 are the constants from Lemma 2. As $\|\mathbf{B}_i^0 - \nabla^2 f_i(\mathbf{x}^*)\|_{\mathbf{M}} \leq \delta$, we also have

$$\|\mathbf{B}_i^0 - \nabla^2 f_i(\mathbf{x}^*)\| \leq \eta\delta.$$

Therefore, by triangle inequality from $\|\nabla^2 f_i(\mathbf{x}^*)\| \leq L$ we obtain $\|\mathbf{B}_i^0\| \leq \eta\delta + L$, so $\|(1/n)\sum_{i=1}^n \mathbf{B}_i^0\| \leq \eta\delta + L$. The second part of inequality (28) also implies $2\gamma(1+r)\eta\delta \leq r$. Moreover, it holds that $\|\mathbf{B}_i^0 - \nabla^2 f_i(\mathbf{x}^*)\| \leq \eta\delta < 2\eta\delta$ and by Assumption 1 $\gamma \geq \|\nabla^2 f_i(\mathbf{x}^*)^{-1}\|$, so we obtain by Banach Lemma that

$$\|(\mathbf{B}_i^0)^{-1}\| \leq (1+r)\gamma.$$

We formally prove this result in the following lemma.

Lemma 6. *If the Hessian approximation* $\mathbf{B}_i$ *satisfies the inequality* $\|\mathbf{B}_i - \nabla^2 f_i(\mathbf{x}^*)\| \leq 2\eta\delta$ *and* $\|\nabla^2 f_i(\mathbf{x}^*)^{-1}\| \leq \gamma$, *then we have* $\|\mathbf{B}_i^{-1}\| \leq (1+r)\gamma$.

Proof. Note that according to Banach Lemma, if a matrix $\mathbf{A}$ satisfies the inequality $\|\mathbf{A} - \mathbf{I}\| \leq 1$, then it holds $\|\mathbf{A}^{-1}\| \leq \frac{1}{1-\|\mathbf{A}-\mathbf{I}\|}$.

We first show that $\|\nabla^2 f_i(\mathbf{x}^*)^{-1/2}\mathbf{B}_i\nabla^2 f_i(\mathbf{x}^*)^{-1/2} - \mathbf{I}\| \leq 1$. To do so, note that

$$\begin{aligned}
\|\nabla^2 f_i(\mathbf{x}^*)^{-1/2}\mathbf{B}_i\nabla^2 f_i(\mathbf{x}^*)^{-1/2} - \mathbf{I}\| &\leq \|\nabla^2 f_i(\mathbf{x}^*)^{-1/2}\|\|\mathbf{B}_i - \nabla^2 f_i(\mathbf{x}^*)\|\|\nabla^2 f_i(\mathbf{x}^*)^{-1/2}\| \\
&\leq 2\eta\delta\gamma \\
&\leq \frac{r}{r+1} \\
&< 1. \qquad (29)
\end{aligned}$$

Now using this result and Banach Lemma we can show that

$$\begin{aligned}
\|\nabla^2 f_i(\mathbf{x}^*)^{1/2}\mathbf{B}_i^{-1}\nabla^2 f_i(\mathbf{x}^*)^{1/2}\| &\leq \frac{1}{1 - \|\nabla^2 f_i(\mathbf{x}^*)^{-1/2}\mathbf{B}_i\nabla^2 f_i(\mathbf{x}^*)^{-1/2} - \mathbf{I}\|} \\
&\leq \frac{1}{1 - \frac{r}{r+1}} \\
&= 1 + r \qquad (30)
\end{aligned}$$

Further, we know that

$$\|\nabla^2 f_i(\mathbf{x}^*)^{1/2}\mathbf{B}_i^{-1}\nabla^2 f_i(\mathbf{x}^*)^{1/2}\| \geq \frac{\|\mathbf{B}_i^{-1}\|}{\gamma} \qquad (31)$$

By combining these results we obtain that

$$\|\mathbf{B}_i^{-1}\| \leq (1+r)\gamma. \qquad (32)$$

□

Similarly, for matrix $((1/n)\sum_{i=1}^n \mathbf{B}_i^0)^{-1}$ we get from $\|(1/n)\sum_{i=1}^n \mathbf{B}_i^0 - (1/n)\sum_{i=1}^n \nabla^2 f_i(\mathbf{x}^*)\| \leq (1/n)\sum_{i=1}^n \|\mathbf{B}_i^0 - \nabla^2 f_i(\mathbf{x}^*)\| \leq \eta\delta$ and $\|\nabla^2 f(\mathbf{x}^*)^{-1}\| \leq \gamma$ that

$$\left\|\left(\frac{1}{n}\sum_{i=1}^n \mathbf{B}_i^0\right)^{-1}\right\| \leq (1+r)\gamma.$$

We have by Lemma 2 and induction hypothesis

$$\begin{aligned}
\left\|\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} - \left\|\mathbf{B}_i^{t-D_i^t} - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} &\leq \alpha_3\sigma_i^{t-D_i^t}\left\|\mathbf{B}_i^{t-D_i^t} - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} + \alpha_4\sigma_i^{t-D_i^t} \\
&\leq \left(\alpha_3\left\|\mathbf{B}_i^{t-D_i^t} - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} + \alpha_4\right) r^{m-1}\epsilon \\
&\leq (2\alpha_3\delta + \alpha_4)\, r^{m-1}\epsilon,
\end{aligned}$$

By summing this inequality over all moments in the current epoch when worker i performed its update, we obtain that

$$\left\|\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} - \left\|\mathbf{B}_i^{T_m - d_i^{T_m}} - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} \leq (2\alpha_3\delta + \alpha_4)\, d r^{m-1}\epsilon,$$

Summing the new bound again, but this time over all passed epoch, we obtain

$$\left\|\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} - \left\|\mathbf{B}_i^0 - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} \leq (2\alpha_3\delta + \alpha_4)\, d\epsilon \sum_{k=0}^{m-1} r^k \leq \frac{(2\alpha_3\delta + \alpha_4)\, d\epsilon}{1-r} \leq \delta.$$

Therefore, $\left\|\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right\|_{\mathbf{M}} \leq 2\delta$. By using the Banach argument again, we can show that $\|(\frac{1}{n}\sum_{i=1}^n \mathbf{B}_i^t)^{-1}\| \leq (1+r)\gamma$. Using this result, for any $t \in [T_m, T_{m+1})$ we have $z_i^t = x^{t-D_i^t} \in [T_{m-1}, t)$ and we can write

$$\begin{aligned}
\|\mathbf{x}^t - \mathbf{x}^*\| &\leq (1+r)\gamma\left[\frac{\tilde{L}}{n}\sum_{i=1}^n \left\|\mathbf{z}_i^t - \mathbf{x}^*\right\|^2 + \frac{1}{n}\sum_{i=1}^n \left\|\left[\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right]\left(\mathbf{z}_i^t - \mathbf{x}^*\right)\right\|\right] \\
&\leq (1+r)\gamma\left[\tilde{L}\epsilon + 2\eta\delta\right]\max_i \|\mathbf{z}_i^t - \mathbf{x}^*\| \\
&\leq r \max_i \|\mathbf{z}_i^t - \mathbf{x}^*\| \\
&\leq r^m \|\mathbf{x}^0 - \mathbf{x}^*\|. \qquad (33)
\end{aligned}$$

A.5 Proof of Theorem 1

Dividing both sides of (27) by $(1/n)\sum_{i=1}^n \|\mathbf{z}_i^t - \mathbf{x}^*\|$, we get

$$\frac{\|\mathbf{x}^t - \mathbf{x}^*\|}{\frac{1}{n}\sum_{i=1}^n \|\mathbf{z}_i^t - \mathbf{x}^*\|} \leq \tilde{L}\Gamma^t \sum_{i=1}^n \frac{\|\mathbf{z}_i^t - \mathbf{x}^*\|^2}{\sum_{i=1}^n \|\mathbf{z}_i^t - \mathbf{x}^*\|} + \Gamma^t \sum_{i=1}^n \frac{\left\|\left(\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right)\left(\mathbf{z}_i^t - \mathbf{x}^*\right)\right\|}{\sum_{i=1}^n \|\mathbf{z}_i^t - \mathbf{x}^*\|} \qquad (34)$$

As every term in $\sum_{i=1}^n \|\mathbf{z}_i^t - \mathbf{x}^*\|$ is non-negative, the upper bound in (34) will remain valid if we keep only one summand out of the whole sum in the denominators of the right-hand side, so

$$\begin{aligned}
\frac{\|\mathbf{x}^t - \mathbf{x}^*\|}{\frac{1}{n}\sum_{i=1}^n \|\mathbf{z}_i^t - \mathbf{x}^*\|} &\leq \tilde{L}\Gamma^t \sum_{i=1}^n \frac{\|\mathbf{z}_i^t - \mathbf{x}^*\|^2}{\|\mathbf{z}_i^t - \mathbf{x}^*\|} + \Gamma^t \sum_{i=1}^n \frac{\left\|\left(\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right)\left(\mathbf{z}_i^t - \mathbf{x}^*\right)\right\|}{\|\mathbf{z}_i^t - \mathbf{x}^*\|} \\
&= \tilde{L}\Gamma^t \sum_{i=1}^n \left\|\mathbf{z}_i^t - \mathbf{x}^*\right\| + \Gamma^t \sum_{i=1}^n \frac{\left\|\left(\mathbf{B}_i^t - \nabla^2 f_i(\mathbf{x}^*)\right)\left(\mathbf{z}_i^t - \mathbf{x}^*\right)\right\|}{\|\mathbf{z}_i^t - \mathbf{x}^*\|}. \qquad (35)
\end{aligned}$$

Now using the result in Lemma 5 of Mokhtari et al. (2018a), the second sum in (35) converges to zero. Further, Γ^t is bounded above by a positive constant. Hence, by computing the limit of both sides in (35) we obtain

$$\lim_{t\to\infty} \frac{\|\mathbf{x}^t - \mathbf{x}^*\|}{\frac{1}{n}\sum_{i=1}^n \|\mathbf{z}_i^t - \mathbf{x}^*\|} = 0.$$

Therefore, if T is big enough, for $t > T$ we have

$$\|\mathbf{x}^t - \mathbf{x}^*\| \leq \frac{1}{n}\sum_{i=1}^n \left\|\mathbf{z}_i^t - \mathbf{x}^*\right\| = \frac{1}{n}\sum_{i=1}^n \left\|\mathbf{x}_i^{t-D_i^t} - \mathbf{x}^*\right\| \leq \max_i \left\|\mathbf{x}_i^{t-D_i^t} - \mathbf{x}^*\right\|. \qquad (36)$$

Now, let $t_0 = t_0(m) := \min\{\tilde{t} \in [T_{m+1}, T_{m+2}) : \|\mathbf{x}^{\tilde{t}} - \mathbf{x}^*\| = \max_{t\in[T_{m+1},T_{m+2})} \|\mathbf{x}^t - \mathbf{x}^*\|\}$. In other words, t_0 is the first moment in epoch $m+1$ attaining the maximal distance from x^*. Then, for all $t \in [T_{m+1}, t_0)$ we have $\|\mathbf{x}^t - \mathbf{x}^*\| < \|\mathbf{x}^{t_0} - \mathbf{x}^*\|$. Furthermore, from equation (36) and the fact that, according to Proposition 1, $t_0 - D_i^{t_0} \in [T_m, t_0)$ we get

$$\max_{t\in[T_{m+1},T_{m+2})} \|\mathbf{x}^t - \mathbf{x}^*\| = \|\mathbf{x}^{t_0} - \mathbf{x}^*\| \leq \max_i \left\|\mathbf{x}_i^{t_0-D_i^{t_0}} - \mathbf{x}^*\right\| \leq \max_{t\in[T_m,t_0)} \|\mathbf{x}^t - \mathbf{x}^*\|.$$

Note that it can not happen that $\max_{t\in[T_m,t_0)} \|\mathbf{x}^t - \mathbf{x}^*\| = \max_{t\in[T_{m+1},t_0)} \|\mathbf{x}^t - \mathbf{x}^*\|$ as that would mean that there exists a $\hat{t} \in [T_{m+1}, t_0)$ such that $\|\mathbf{x}^{\hat{t}} - \mathbf{x}^*\| \geq \|\mathbf{x}^{t_0} - \mathbf{x}^*\|$, which we made impossible when defining t_0. Then, the only option is that in fact

$$\max_{t\in[T_m,t_0)} \|\mathbf{x}^t - \mathbf{x}^*\| = \max_{t\in[T_m,T_{m+1})} \|\mathbf{x}^t - \mathbf{x}^*\|.$$

Finally,

$$\begin{aligned}
\lim_{t\to\infty} \frac{\max_{t\in[T_{m+1},T_{m+2})}\|\mathbf{x}^t-\mathbf{x}^*\|}{\max_{t\in[T_m,T_{m+1})}\|\mathbf{x}^t-\mathbf{x}^*\|} &= \lim_{t\to\infty} \frac{\|\mathbf{x}^{t_0(m)}-\mathbf{x}^*\|}{\max_{t\in[T_m,T_{m+1})}\|\mathbf{x}^t-\mathbf{x}^*\|} \\
&= \lim_{t\to\infty} \frac{\|\mathbf{x}^{t_0(m)}-\mathbf{x}^*\|}{\max_{t\in[T_m,t_0(m))}\|\mathbf{x}^t-\mathbf{x}^*\|} \\
&\le \lim_{t\to\infty} \frac{\|\mathbf{x}^{t_0(m)}-\mathbf{x}^*\|}{\max_i \|\mathbf{x}^{t_0(m)-D_i^{t_0(m)}}-\mathbf{x}^*\|} \\
&\le \lim_{t\to\infty} \frac{\|\mathbf{x}^{t_0(m)}-\mathbf{x}^*\|}{\frac{1}{n}\sum_{i=1}^n \|\mathbf{z}_i^{t_0(m)}-\mathbf{x}^*\|} = 0,
\end{aligned}$$

where at the last step we used again the fact that $\mathbf{z}_i^t = \mathbf{x}^{t-D_i^t}$.

B Implementation of Dave-QN

In Algorithm 2, we provide a simplified version of the Dave-QN in terms of notation and indices of the variables, which illustrates how Dave-QN can be implemented from master's and worker nodes' side further. We observe that steps at worker i is devoted to performing the update in (11). Using the computed matrix $\mathbf{B}_i$, node i evaluates the vector $\Delta\mathbf{u}$. Then, it sends the vectors $\Delta\mathbf{u}$, $\mathbf{y}_i$, and $\mathbf{q}_i$ as well as the scalars α and β to the master node. The master node uses the variation vectors $\Delta\mathbf{u}$ and $\mathbf{y}$ to update $\mathbf{u}$ and $\mathbf{g}$. Then, it performs the update $\mathbf{x}^{t+1} = (\mathbf{B}^{t+1})^{-1}\left(\mathbf{u}^{t+1} - \mathbf{g}^{t+1}\right)$ by following the efficient procedure presented in (16)–(17).