Supplementary Material for
Parallel Asynchronous Stochastic Coordinate Descent with Auxiliary Variables

A Parallelize Second-Order Methods by AUX-PCD

Second-order methods are known to enjoy faster convergence than first-order methods such as CD or stochastic gradient descent in terms of iteration complexity. However, the cost per iteration of second-order methods is usually much higher as it requires n^2 entries in the Hessian to construct a quadratic approximation and obtain a Newton direction $\boldsymbol{s}^*$. Exploiting the structure of the Hessian matrix, some recent works successfully apply CD to solve the quadratic approximation with the explicit construction of Hessian matrix, such as GLMNET for ℓ_1-regularized logistic regression [27] and QUIC for ℓ_1-regularized Gaussian graphical model learning [8].

In this section we demonstrate that AUX-PCD can be embedded into a second-order method for the family of problems defined in (1). Let $\bar{f}(\boldsymbol{z}) := \sum_{i=1}^{n} f_i(z_i)$ and $\bar{d}(\boldsymbol{z}) := D(Q\boldsymbol{z})$. Thus, we have $F(\boldsymbol{z}) = \bar{f}(\boldsymbol{z}) + \bar{d}(\boldsymbol{z})$. We assume that $d_k(\cdot)$ is twice differentiable and f_i can be a non-smooth convex function. To apply a second-order method (e.g., Newton or proximal Newton) to minimize $F(\boldsymbol{z})$ requires obtaining the Newton direction $\boldsymbol{s}^*$ by minimizing the following quadratic approximation $\mathcal{Q}(\boldsymbol{s}) \approx F(\boldsymbol{z}+\boldsymbol{s}) - F(\boldsymbol{z})$:

$$\boldsymbol{s}^* = \arg\min_{\boldsymbol{s}} \mathcal{Q}(\boldsymbol{s}) := \bar{f}(\boldsymbol{z}+\boldsymbol{s}) + \nabla \bar{d}(\boldsymbol{z})^\top \boldsymbol{s} + \frac{1}{2}\boldsymbol{s}^T \nabla^2 \bar{d}(\boldsymbol{z})\boldsymbol{s}. \tag{14}$$

By the definition of $\bar{d}(\boldsymbol{z})$, we know that

$$\nabla^2 \bar{d}(\boldsymbol{z}) = Q^\top \mathcal{H} Q$$

where $\bar{\boldsymbol{q}}^\top$ is the k-th row of Q, and $\mathcal{H}$ is a diagonal matrix with $\mathcal{H}_{kk} = \nabla^2_{kk} D(\bar{\boldsymbol{q}}_k^\top \boldsymbol{z}),\ \forall k$. Note that $\nabla \bar{d}(\boldsymbol{z})$ and $\mathcal{H}$ can be pre-computed and stored as they do not change as $\boldsymbol{s}$ changes. We can maintain an auxiliary variable $\boldsymbol{r} = \mathcal{H}Q\boldsymbol{s}$ such that we can efficiently apply AUX-PCD to minimize (14) as each single variable subproblem becomes

$$\min_{u} \quad f_i(z_i + u) + \big(\nabla_i \bar{d}(\boldsymbol{z}) + \boldsymbol{q}_i^\top \boldsymbol{r}\big)u + \frac{\boldsymbol{q}_i^\top \mathcal{H} \boldsymbol{q}_i}{2} u^2,$$

which requires only $O(m)$ operations to construct the subproblem and obtain the optimal solution u^*. We can see that AUX-PCD can be easily applied to parallelize a second-order method. In Section 5, we will apply this approach to parallelize a second-order method for ℓ_1-regularized logistic regression.

B Proof for the Linear Convergence for AUX-PCD with Smooth f_i

Theorem 5. *Assume a function $F(\boldsymbol{z})$ in the family defined by (1) admits a global error bound from the beginning. We further assume that $F(\boldsymbol{z})$ is a L_{max}-smooth functions. If the upper bound of the staleness τ is small enough such that the following two conditions hold:*

$$(3(\tau+1)^2 e\bar{M})/\sqrt{n} \le 1, \tag{15}$$

$$\frac{2L_{max}c_0}{\sigma(1-2c_0)} \le 1, \tag{16}$$

where $\bar{M} = \sigma^{-1} L R_{max}^2 + 2LMR_{max}$, and $c_0 = \frac{\tau^2 M^2 \sigma^{-2} L^2 R_{max}^2 e^2}{n}$ then Algorithm 2 with atomic operations has a global linear convergence rate in expectation, that is,

$$\mathbb{E}[\![F(\boldsymbol{z}^{j+1})]\!] - F(\boldsymbol{z}^*) \le \eta\big(\mathbb{E}[\![F(\boldsymbol{z}^j)]\!] - F(\boldsymbol{z}^*)\big), \tag{17}$$

where $\boldsymbol{z}^*$ is the optimal solution and

$$\eta = 1 - \frac{\sigma}{2n\kappa^2 L_{max}}\left(1 - \frac{2L_{max}c_0}{\sigma(1-2c_0)}\right). \tag{18}$$

B.1 Notations and Propositions

B.1.1 Notations

- For all $i = 1, \ldots, n$, we have the following definitions:

$$T_i(\boldsymbol{r}, s) := \arg\min_u \quad f_i(u) + D(\boldsymbol{r} + (u - s)\boldsymbol{q}_i),$$

  where $\boldsymbol{r} \in R^d$ and $s \in R$. For convergence, we define $T(\boldsymbol{r}, \boldsymbol{s})$ as an n-dimension vector with $T_i(\boldsymbol{r}, \boldsymbol{s}) = T_i(\boldsymbol{r}, s_i)$ as the i-th element.
- Let $\{\boldsymbol{z}^j\}$ and $\{\hat{\boldsymbol{r}}^j\}$ be the sequence generated/maintained by Algorithm 2 using

$$z_t^{j+1} = \begin{cases} T_t(\hat{\boldsymbol{r}}^j, z_t^j) & \text{if } t = i(j), \\ z_t^j & \text{if } t \neq i(j), \end{cases}$$

  where $i(j)$ is the index selected at j-th iteration. For convenience, we define

$$\Delta z_j = z_{i(j)}^{j+1} - z_{i(j)}^j.$$

- Let $\{\tilde{\boldsymbol{z}}^j\}$ be the sequence defined by

$$\tilde{z}_t^{j+1} = T_t(\hat{\boldsymbol{r}}^j, z_t^j) \quad \forall t = 1, \ldots, n.$$

  Note that $\tilde{z}_{i(j)}^{j+1} = z_{i(t)}^{j+1}$.
- Let $\bar{\boldsymbol{r}}^j = \sum_{i=1}^n z_i^j \boldsymbol{q}_i$ be the "true" auxiliary variables corresponding to $\boldsymbol{z}^j$. Thus,

$$T_t(\bar{\boldsymbol{r}}^j, z_{i(j)}) = T(\boldsymbol{z}^j),$$

  where $T(\cdot)$ is the operator defined in Eq. (7).
- Let $\{\boldsymbol{y}^j\}$, $\{\tilde{\boldsymbol{y}}^j\}$ be the sequences defined by

$$y_t^{j+1} = \begin{cases} T_t(\bar{\boldsymbol{r}}^j, z_t^j) & \text{if } t = i(j), \\ z_t^j & \text{if } t \neq i(j), \end{cases}$$
$$\tilde{y}_t^{j+1} = T_t(\bar{\boldsymbol{r}}^j, z_t^j) \quad \forall t = 1, \ldots, n.$$

  Note that $\tilde{y}_{i(j)}^{j+1} = y_{i(j)}^{j+1}$ and $\tilde{\boldsymbol{y}}^{j+1} = T(\boldsymbol{z}^j)$, where $T(\cdot)$ is the operator defined in Eq. (7).
- Let $g^j(u)$ be the univariate function considered at the j-th iteration:

$$\begin{aligned} g^j(u) &:= F\Big(\boldsymbol{z}^j + (u - z_{i(j)}^j) \cdot \boldsymbol{e}_{i(j)}\Big) \\ &= f_{i(j)}(u) + D(\bar{\boldsymbol{r}} + \Big(u - z_{i(j)}^j \boldsymbol{q}_{i(j)}\Big)) + \text{constant} \end{aligned}$$

  Thus, $y_{i(j)}^{j+1} = T_{i(j)}(\bar{\boldsymbol{r}}^j, z_{i(j)}^j)$ is the minimizer for $g^j(u)$.

B.1.2 Propositions

Proposition 1.

$$E_{i(j)}\big(\|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|^2\big) = \frac{1}{n}\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2, \quad E_{i(j)}\big(\|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|\big) = \frac{1}{n}\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|, \tag{19}$$

$$E_{i(j)}\big(\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^j\|^2\big) = \frac{1}{n}\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2, \quad E_{i(j)}\big(\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^j\|\big) = \frac{1}{n}\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|. \tag{20}$$

Proof. Based on the assumption that $i(j)$ is uniformly random selected from $\{1, \ldots, n\}$, the above two identities follow from the definition of $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{y}}$. □

Proposition 2.

$$\|\bar{\boldsymbol{r}}^j - \hat{\boldsymbol{r}}^j\| \leq R_{max} \sum_{t=j-\tau}^{j-1} |\Delta z_t|. \tag{21}$$

Proof.

$$\|\bar{\boldsymbol{r}}^j - \hat{\boldsymbol{r}}^j\| = \| \sum_{(t,k) \in \mathcal{Z}^j \setminus \mathcal{U}^j} (\Delta z_t) q_{i(t)k} \boldsymbol{e}_k \|$$

$$\leq \sum_{t=j-\tau}^{j-1} |\Delta z_t| \|\boldsymbol{q}_{i(t)}\| \leq R_{max} \sum_{t=j-\tau}^{j-1} |\Delta z_t|$$

□

Proposition 3. *For any $\boldsymbol{r}_1, \boldsymbol{r}_2 \in R^m$ and $s_1, s_2 \in R$,*

$$|T_i(\boldsymbol{r}_1, s_1) - T_i(\boldsymbol{r}_2, s_2)| \leq \sigma^{-1} L \|\boldsymbol{q}_i\| \|\boldsymbol{r}_1 - \boldsymbol{r}_2\| + \sigma^{-1} L \|\boldsymbol{q}_i\|^2 |s_1 - s_2|, \tag{22}$$

Thus, for any $\boldsymbol{r}_1, \boldsymbol{r}_2 \in R^m$ and $\boldsymbol{s}_1, \boldsymbol{s}_2 \in R^n$,

$$\|T(\boldsymbol{r}_1, \boldsymbol{s}_1) - T(\boldsymbol{r}_2, \boldsymbol{s}_2)\| \leq \sigma^{-1} L M \|\boldsymbol{r}_1 - \boldsymbol{r}_2\| + \sigma^{-1} L R_{max}^2 \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|.$$

Proof. Consider the following two problems

$$G(u) = f_i(u) + D(\boldsymbol{r}_1 - s_1 \boldsymbol{q}_i + u \boldsymbol{q}_i) \tag{23}$$

$$\bar{G}(u) = f_i(u) + D(\boldsymbol{r}_2 - s_2 \boldsymbol{q}_i + u \boldsymbol{q}_i) \tag{24}$$

$$\tag{25}$$

Let u^*, $\bar{u}^*$ be the optimizers for $G(u)$ and $\bar{G}(u)$, respectively. Thus, we have $0 \in \partial G(u^*)$ and $0 \in \partial \bar{G}(\bar{u}^*)$. Based on the strong convex assumption and the non-expansiveness of sub-gradient, for all $\bar{g} \in \partial G(\bar{u}^*)$, we have

$$(u^* - \bar{u}^*)(0 - \bar{g}) \geq \sigma (u^* - \bar{u}^*)^2 \Rightarrow |u^* - \bar{u}^*| \leq \frac{1}{\sigma} |\bar{g}|. \tag{26}$$

Let's bound the size of the sub-gradient $\bar{g}$.

$$0 \in \partial \bar{G}(\bar{u}^*) = \partial f_i(\bar{u}^*) + \boldsymbol{q}_i^\top \nabla D(\boldsymbol{r}_2 - s_2 \boldsymbol{q}_i + \bar{u}^* \boldsymbol{q}_i)$$

$$\bar{g} \in \partial G(\bar{u}^*) = \partial f_i(\bar{u}^*) + \boldsymbol{q}_i^\top \nabla D(\boldsymbol{r}_1 - s_1 \boldsymbol{q}_i + \bar{u}^* \boldsymbol{q}_i)$$

As a result, there exists a sub-gradient $\bar{g}$ such that

$$\begin{aligned} \bar{g} &= \boldsymbol{q}_i^\top (\nabla D(\boldsymbol{r}_1 - s_1 \boldsymbol{q}_i + \bar{u}^* \boldsymbol{q}_i) - \nabla D(\boldsymbol{r}_2 - s_2 \boldsymbol{q}_i + \bar{u}^* \boldsymbol{q}_i)) \\ &= \sum_{k=1}^m q_{ik} (d'_k (r_{1k} - s_1 q_{ik} + \bar{u}^* q_{ik}) - d'_k (r_{2k} - s_2 q_{ik} + \bar{u}^* q_{ik})) \end{aligned}$$

By the Lipschitz continuity, we can bound $|\bar{g}|$ as follows.

$$\begin{aligned} |\bar{g}| &\leq \sum_{k=1}^m L |q_{ik}| |(r_{1k} - r_{2k} + (s_1 - s_2) q_{ik})| \\ &\leq \sum_{k=1}^m L |q_{ik}| |r_{1k} - r_{2k}| + \sum_{k=1}^m L |s_1 - s_2| |q_{ik}|^2 \\ &\leq L \|\boldsymbol{q}_i\| \|\boldsymbol{r}_1 - \boldsymbol{r}_2\| + L |s_1 - s_2| \|\boldsymbol{q}_i\|^2 \end{aligned} \tag{27}$$

Combining (26), we obtain the following result

$$|T_i(\boldsymbol{r}_1, s_1) - T_i(\boldsymbol{r}_2, s_2)| \le \sigma^{-1} L \|\boldsymbol{q}_i\| \|\boldsymbol{r}_1 - \boldsymbol{r}_2\| + \sigma^{-1} L \|\boldsymbol{q}_i\|^2 |s_1 - s_2|$$

Consider the difference of the two operators $\Delta = T(\boldsymbol{r}_1, \boldsymbol{s}_1) - T(\boldsymbol{r}_2, \boldsymbol{s}_2)$, as $|\Delta_i| \le \sigma^{-1} L \|\boldsymbol{q}_i\| \|\boldsymbol{r}_1 - \boldsymbol{r}_2\| + \sigma^{-1} L \|\boldsymbol{q}_i\|^2 |s_{1i} - s_{2i}|$, we have

$$\|\Delta\| \le \sigma^{-1} L M \|\boldsymbol{r}_1 - \boldsymbol{r}_2\| + \sigma^{-1} L R_{max}^2 \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|.$$

□

Proposition 4. *Let $\bar{M} := \sigma^{-1}(LR_{max}^2 + 2LMR_{max}) \ge 1$, $q = \frac{3(\tau+1)e\bar{M}}{\sqrt{n}}$, $\rho = (1+q)^2$, and $\theta = \sum_{t=1}^{\tau} \rho^{t/2}$. If $q(\tau+1) \le 1$, then $\rho^{(\tau+1)/2} \le e$, and*

$$\rho^{-1} \le 1 - \frac{2 + 2\bar{M} + 2\bar{M}\theta}{\sqrt{n}}. \tag{28}$$

Proof. By the definition of ρ and the condition $q(\tau+1) \le 1$, we have

$$\begin{aligned} \rho^{(\tau+1)/2} &= \left(\left(\rho^{1/2}\right)^{1/q}\right)^{q(\tau+1)} \\ &= \left((1+q)^{1/q}\right)^{q(\tau+1)} \le e^{q(\tau+1)} \le e. \end{aligned}$$

By the definition of q, we know that

$$q = \rho^{1/2} - 1 = \frac{3(\tau+1)e\bar{M}}{\sqrt{n}} \Rightarrow \frac{3}{2} = \frac{\sqrt{n}(\rho^{1/2} - 1)}{2(\tau+1)e\bar{M}}.$$

We can derive

$$\begin{aligned} \frac{3}{2} &= \frac{\sqrt{n}(\rho^{1/2} - 1)}{2(\tau+1)eM} \\ &\le \frac{\sqrt{n}(\rho^{1/2} - 1)}{2(\tau+1)\rho^{(\tau+1)/2}\bar{M}} && \because \rho^{(\tau+1)/2} \le e \\ &\le \frac{\sqrt{n}(\rho^{1/2} - 1)}{2(1+\theta)\rho^{1/2}\bar{M}} && \because 1 + \theta = \sum_{t=0}^{\tau} \rho^{t/2} \le (\tau+1)\rho^{\tau/2} \\ &= \frac{\sqrt{n}(1 - \rho^{-1/2})}{2(1+\theta)\bar{M}} \\ &\le \frac{\sqrt{n}(1 - \rho^{-1})}{2(1+\theta)\bar{M}} && \because \rho^{-1/2} \le 1 \end{aligned}$$

Combining the condition that $\bar{M} \ge 1$ and $1 + \theta \ge 1$, we have

$$\frac{\sqrt{n}(1 - \rho^{-1}) - 2}{2(1+\theta)\bar{M}} \ge \frac{\sqrt{n}(1 - \rho^{-1})}{2(1+\theta)\bar{M}} - \frac{1}{2} \ge 1,$$

which leads to

$$\begin{aligned} 2(1+\theta)\bar{M} &\le \sqrt{n} - \sqrt{n}\rho^{-1} - 2 \\ \rho^{-1} &\le 1 - \frac{2 + 2\bar{M} + 2\bar{M}\theta}{\sqrt{n}}. \end{aligned}$$

□

Proposition 5. *For all $j > 0$, we have*

$$F(\boldsymbol{z}^j) \ge F(\boldsymbol{y}^{j+1}) + \frac{\sigma}{2} \|\boldsymbol{z}^j - \boldsymbol{y}^{j+1}\|^2 \tag{29}$$

$$F(\boldsymbol{z}^{j+1}) \le F(\boldsymbol{y}^{j+1}) + \frac{L_{max}}{2} \|\boldsymbol{z}^{j+1} - \boldsymbol{y}^{j+1}\|^2 \tag{30}$$

Proof. First, two properties of $g^j(u)$ are stated as follows.

- the strong convexity of $g^j(u)$: according to the assumption, $g^j(u)$ is σ-strongly convex.
- the Lipschitz assumption of $g^j(u)$: according to the global it follows from the smooth assumption on F

With the above two properties of $g^j(u)$ and the fact that $u^* := y_{i(j)}^{j+1}$ is the minimizer of $g^j(u)$ (which implies that $\nabla g^j(u^*) = 0$), we have the following inequalities:

$$
\begin{aligned}
g^j(z_{i(j)}^j) \geq & g^j(y_{i(j)}^{j+1}) + \frac{\sigma}{2}\|z_{i(j)}^j - y_{i(j)}^{j+1}\|^2 \\
& \text{(by strong convexity)}, \\
g^j(z_{i(j)}^{j+1}) \leq & g^j(y_{i(j)}^{j+1}) + \frac{L_{max}}{2}\|z_{i(j)}^{j+1} - y_{i(j)}^{j+1}\|^2 \\
& \text{(by Lipschitz continuity)}.
\end{aligned}
$$

By the definitions of g^j, $\boldsymbol{z}^j$, $\boldsymbol{z}^{j+1}$, and $\boldsymbol{y}^{j+1}$, we know that

$$
\begin{aligned}
g^j(z_{i(j)}^j) - g^j(y_{i(j)}^{j+1}) &= F(\boldsymbol{z}^j) - F(\boldsymbol{y}^{j+1}), \\
\|z_{i(j)}^j - y_{i(j)}^{j+1}\|^2 &= \|\boldsymbol{z}^j - \boldsymbol{y}^{j+1}\|^2, \\
g^j(z_{i(j)}^{j+1}) - g^j(y_{i(j)}^{j+1}) &= F(\boldsymbol{z}^{j+1}) - F(\boldsymbol{y}^{j+1}), \\
\|z_{i(j)}^{j+1} - y_{i(j)}^{j+1}\|^2 &= \|\boldsymbol{z}^{j+1} - \boldsymbol{y}^{j+1}\|^2,
\end{aligned}
$$

which imply (29) and (30). □

B.2 Lemma 1

To prove the convergence of asynchronous algorithms, we first show that the expected step size does not increase super-linearly by the following Lemma 1.

Lemma 1. *If τ is small enough such that (1) holds (i.e.,)*

$$
(3(\tau+1)^2 e\bar{M})/\sqrt{n} \leq 1,
$$

then $\{\boldsymbol{z}^j\}$ generated by Algorithm 2 with atomic operations satisfies the following inequality:

$$
E(\|\boldsymbol{z}^{j-1} - \boldsymbol{z}^j\|^2) \leq \rho E(\|\boldsymbol{z}^j - \boldsymbol{z}^{j+1}\|^2), \tag{31}
$$

where $\rho = (1 + \frac{3(\tau+1)e\bar{M}}{\sqrt{n}})^2$, and $\bar{M} = \sigma^{-1}\big(LR_{max}^2 + 2LMR_{max}\big)$.

Proof. Similar to [15], we prove Eq. (31) by induction. First, we know that for any two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, we have

$$
\|\boldsymbol{a}\|^2 - \|\boldsymbol{b}\|^2 \leq 2\|\boldsymbol{a}\|\|\boldsymbol{b} - \boldsymbol{a}\|.
$$

See [15] for a proof for the above inequality. Thus, for all j , we have

$$
\begin{aligned}
&\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|^2 - \|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\|^2 \\
&\leq 2\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^{j-1} + \tilde{\boldsymbol{z}}^j\|.
\end{aligned} \tag{32}
$$

The second factor in the r.h.s of (32) is bounded as follows:

$$
\begin{aligned}
&\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^{j-1} + \tilde{\boldsymbol{z}}^j\| \\
&\leq \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| + \|T(\hat{\boldsymbol{r}}^j, \boldsymbol{z}^j) - T(\hat{\boldsymbol{r}}^{j-1}, \boldsymbol{z}^{j-1})\| \\
&\leq \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| + \sigma^{-1} LM \|\hat{\boldsymbol{r}}^j - \hat{\boldsymbol{r}}^{j-1}\| + \sigma^{-1} L R_{max}^2 \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| \\
&\leq (1 + \sigma^{-1} L R_{max}^2) \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| + \sigma^{-1} LM \|\hat{\boldsymbol{r}}^j - \hat{\boldsymbol{r}}^{j-1}\| \\
&\leq (1 + \sigma^{-1} L R_{max}^2) \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| + \sigma^{-1} LM \big(\|\hat{\boldsymbol{r}}^j - \bar{\boldsymbol{r}}^j + \bar{\boldsymbol{r}}^j - \bar{\boldsymbol{r}}^{j-1} + \bar{\boldsymbol{r}}^{j-1} - \hat{\boldsymbol{r}}^{j-1}\|\big) \\
&\leq (1 + \sigma^{-1} L R_{max}^2) \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| + \sigma^{-1} LM \|\bar{\boldsymbol{r}}^j - \bar{\boldsymbol{r}}^{j-1}\| + \sigma^{-1} LM \|\hat{\boldsymbol{r}}^j - \bar{\boldsymbol{r}}^j\| + \sigma^{-1} LM \|\hat{\boldsymbol{r}}^{j-1} - \bar{\boldsymbol{r}}^{j-1}\| \\
&\leq (1 + \sigma^{-1} L R_{max}^2 + \sigma^{-1} LM R_{max}) \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| + \sigma^{-1} LM R_{max} \left(\sum_{t=j-\tau}^{j-1} \|\Delta z_t\| + \sum_{t=j-\tau-1}^{j-2} \|\Delta z_t\| \right) \\
&\leq (1 + \sigma^{-1} L R_{max}^2 + 2\sigma^{-1} LM R_{max}) \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| + 2\sigma^{-1} LM R_{max} \sum_{t=j-\tau-1}^{j-2} \|\Delta z_t\| \\
&= (1 + \bar{M}) \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| + 2\sigma^{-1} LM R_{max} \sum_{t=j-\tau-1}^{j-2} \|\Delta z_t\|
\end{aligned} \tag{33}
$$

Now we prove (31) by induction.

Induction Hypothesis. Using Proposition 1, we prove the following equivalent statement. For all j,

$$
E(\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|^2) \leq \rho E(\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\|^2), \tag{34}
$$

Induction Basis. When $j = 1$,

$$
\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2 + \boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\| \leq (1 + \bar{M}) \|\boldsymbol{z}^1 - \boldsymbol{z}^0\|.
$$

Taking the expectation of (32), we have

$$
\begin{aligned}
&E[\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\|^2] - E[\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\|^2] \\
&\leq 2E[\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\| \|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2 - \boldsymbol{z}^0 + \tilde{\boldsymbol{z}}^1\|] \\
&\leq 2(1 + \bar{M}) E(\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\| \|\boldsymbol{z}^0 - \boldsymbol{z}^1\|).
\end{aligned}
$$

From (19) in Proposition 1, we have $E[\|\boldsymbol{z}^0 - \boldsymbol{z}^1\|^2] = \frac{1}{n} \|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\|^2$. Also, by AM-GM inequality, for any $\mu_1, \mu_2 > 0$ and any $c > 0$, we have

$$
\mu_1 \mu_2 \leq \frac{1}{2}(c\mu_1^2 + c^{-1}\mu_2^2). \tag{35}
$$

Therefore, we have

$$
\begin{aligned}
&E[\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\| \|\boldsymbol{z}^0 - \boldsymbol{z}^1\|] \\
&\leq \frac{1}{2} E\Big[n^{1/2} \|\boldsymbol{z}^0 - \boldsymbol{z}^1\|^2 + n^{-1/2} \|\tilde{\boldsymbol{z}}^1 - \boldsymbol{z}^0\|^2\Big] \\
&= \frac{1}{2} E\Big[n^{-1/2} \|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\|^2 + n^{-1/2} \|\tilde{\boldsymbol{z}}^1 - \boldsymbol{z}^0\|^2\Big] \quad \text{by (19)} \\
&= n^{-1/2} E[\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\|^2].
\end{aligned}
$$

Therefore,

$$
E[\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\|^2] - E[\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\|^2] \leq \frac{2(1 + \bar{M})}{\sqrt{n}} E[\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\|^2],
$$

which implies

$$
E[\|\boldsymbol{z}^0 - \boldsymbol{z}^1\|^2] \leq \frac{1}{1 - \frac{2(1+\bar{M})}{\sqrt{n}}} E[\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\|^2] \leq \rho E[\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\|^2], \tag{36}
$$

where the last inequality is based on Proposition 4 and the fact $\theta(\bar{M} + 1) \geq 1$.

Induction Step. By the induction hypothesis, we assume

$$E[\|\boldsymbol{z}^{t-1} - \tilde{\boldsymbol{z}}^{t}\|^2] \leq \rho E[\|\boldsymbol{z}^{t} - \tilde{\boldsymbol{z}}^{t+1}\|^2] \quad \forall t \leq j-1. \tag{37}$$

The goal is to show

$$E[\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2] \leq \rho E[\|\boldsymbol{z}^{j} - \tilde{\boldsymbol{z}}^{j+1}\|^2].$$

First, we show that for all $t < j$,

$$E\Big[\|\boldsymbol{z}^{t} - \boldsymbol{z}^{t+1}\|\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|\Big] \leq \frac{\rho^{(j-1-t)/2}}{\sqrt{n}} E\Big[\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2\Big] \tag{38}$$

By (35) with $c = n^{1/2}\gamma$, where $\gamma = \rho^{(t+1-j)/2}$,

$$\begin{aligned}
&E\Big[\|\boldsymbol{z}^{t} - \boldsymbol{z}^{t+1}\|\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|\Big] \\
&\leq \frac{1}{2}E\Big[n^{1/2}\gamma\|\boldsymbol{z}^{t} - \boldsymbol{z}^{t+1}\|^2 + n^{-1/2}\gamma^{-1}\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2\Big] \\
&= \frac{1}{2}E\Big[n^{1/2}\gamma E[\|\boldsymbol{z}^{t} - \boldsymbol{z}^{t+1}\|^2] + n^{-1/2}\gamma^{-1}\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2\Big] \\
&= \frac{1}{2}E\Big[n^{-1/2}\gamma\|\boldsymbol{z}^{t} - \tilde{\boldsymbol{z}}^{t+1}\|^2 + n^{-1/2}\gamma^{-1}\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2\Big] \\
&\qquad \text{(by Proposition 1)} \\
&\leq \frac{1}{2}E\Big[n^{-1/2}\gamma\rho^{j-1-t}\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2 + n^{-1/2}\gamma^{-1}\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2\Big] \\
&\qquad \text{(by Eq. (37))} \\
&\leq \frac{1}{2}E\Big[n^{-1/2}\gamma^{-1}\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2 + n^{-1/2}\gamma^{-1}\|\tilde{\boldsymbol{z}}^{j-1} - \boldsymbol{z}^{j}\|^2\Big] \\
&\qquad \text{(by the definition of } \gamma) \\
&\leq \frac{\rho^{(j-1-t)/2}}{\sqrt{n}} E\Big[\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2\Big]
\end{aligned}$$

Let $\theta = \sum_{t=1}^{\tau} \rho^{t/2}$ and recall that $\bar{M} = \sigma^{-1}\big(LR_{max}^2 + 2LMR_{max}\big)$. We have

$$\begin{aligned}
&E[\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2] - E[\|\boldsymbol{z}^{j} - \tilde{\boldsymbol{z}}^{j+1}\|^2] \\
&\leq E\Bigg[2\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|\Bigg((1+\bar{M})\|\boldsymbol{z}^{j} - \boldsymbol{z}^{j-1}\| \quad + 2\sigma^{-1}LMR_{max}\sum_{t=j-\tau-1}^{j-1}\|\boldsymbol{z}^{t} - \boldsymbol{z}^{t-1}\|\Bigg)\Bigg] \qquad \text{by (32), (33)} \\
&= (2+2\bar{M})E(\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|\|\boldsymbol{z}^{j} - \boldsymbol{z}^{j-1}\|) \\
&\quad + 4\sigma^{-1}LMR_{max}\sum_{t=j-\tau-1}^{j-1} E\Big[\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|\|\boldsymbol{z}^{t} - \boldsymbol{z}^{t-1}\|\Big] \\
&\leq (2+2\bar{M})n^{-1/2}E[\|\tilde{\boldsymbol{z}}^{j} - \boldsymbol{z}^{j-1}\|^2] \\
&\quad + 4\sigma^{-1}LMR_{max}n^{-1/2}E[\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2]\sum_{t=j-1-\tau}^{j-2}\rho^{(j-1-t)/2} \qquad \text{by (38)} \\
&\leq (2+2\bar{M})n^{-1/2}E[\|\tilde{\boldsymbol{z}}^{j} - \boldsymbol{z}^{j-1}\|^2] + 4\sigma^{-1}LMR_{max}n^{-1/2}\theta E[\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2] \\
&\leq \frac{2+2\bar{M}+2\bar{M}\theta}{\sqrt{n}}E[\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2],
\end{aligned}$$

which implies that

$$\begin{aligned}
E[\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^{j}\|^2] &\leq \frac{1}{1 - \frac{2+2\bar{M}+2\bar{M}\theta}{\sqrt{n}}}E[\|\boldsymbol{z}^{j} - \tilde{\boldsymbol{z}}^{j+1}\|^2] \\
&\leq \rho E[\|\boldsymbol{z}^{j} - \tilde{\boldsymbol{z}}^{j+1}\|^2],
\end{aligned}$$

where the last inequality is based on Proposition 4. □

Lemma 2. *Let $\{z^j\}$ be the sequence generated by Algorithm 2 with atomic operations. Assume the conditions in Lemma 1 hold and let*

$$c_0 = \frac{\left(\tau\sigma^{-1} L M R_{max} e\right)^2}{n},$$

we have

$$\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\|^2\right]\!\right] \le c_0\, \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right] \tag{39}$$

$$\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right] \le 2(1-2c_0)^{-1}\, \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right] \tag{40}$$

Proof. We prove (39) by the following derivation.

$$\begin{aligned}
&\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\|^2]\!] \\
&= \mathbb{E}\left[\!\left[\sum_{t=1}^{n} \left(T_t(\bar{\boldsymbol{r}}^j, z_t^j) - T_t(\hat{\boldsymbol{r}}^j, z_t^j)\right)^2\right]\!\right] \\
&\le \sigma^{-2} L^2 M^2 \mathbb{E}[\![\|\bar{\boldsymbol{r}}^j - \hat{\boldsymbol{r}}^j\|^2]\!] \quad \text{By Proposition 3} \\
&\le M^2\sigma^{-2}L^2R_{max}^2 \mathbb{E}\left[\!\left[\left(\sum_{t=j-\tau}^{j-1} \|\boldsymbol{z}^{t+1} - \boldsymbol{z}^t\|\right)^2\right]\!\right] \quad \text{By Proposition 2} \\
&\le M^2\sigma^{-2}L^2R_{max}^2 \mathbb{E}\left[\!\left[\tau\left(\sum_{t=j-\tau}^{j-1} \|\boldsymbol{z}^{t+1} - \boldsymbol{z}^t\|^2\right)\right]\!\right] \quad \text{By Cauchy Schwarz} \\
&\le \tau M^2\sigma^{-2}L^2R_{max}^2 \mathbb{E}\left[\!\left[\left(\sum_{t=1}^{\tau} \rho^t \|\boldsymbol{z}^j - \boldsymbol{z}^{j+1}\|^2\right)\right]\!\right] \quad \text{By Lemma 1} \\
&\le \frac{\tau M^2\sigma^{-2}L^2R_{max}^2}{n}\left(\sum_{t=1}^{\tau}\rho^t\right)\mathbb{E}[\![\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2]\!] \quad \text{By Proposition 1} \\
&\le \frac{\tau^2 M^2\sigma^{-2}L^2R_{max}^2}{n}\rho^\tau \mathbb{E}[\![\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2]\!]
\end{aligned}$$

Proposition 4 implies $\rho^{(\tau+1)/2} \le e$, which further implies $\rho^\tau \le e^2$ because $\rho \ge 1$. Thus,

$$\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\|^2]\!] \le \frac{\tau^2 M^2\sigma^{-2}L^2R_{max}^2 e^2}{n}\mathbb{E}[\![\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2]\!] = c_0\, \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right].$$

To prove (40), we applying triangle inequality and Cauchy-Schwarz Inequality as follows

$$\begin{aligned}
&\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right] \\
&= \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \tilde{\boldsymbol{y}}^{j+1} + \tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right] \\
&\le \mathbb{E}\left[\!\left[\left(\|\tilde{\boldsymbol{z}}^{j+1} - \tilde{\boldsymbol{y}}^{j+1}\| + \|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|\right)^2\right]\!\right] \quad \text{By triangle inequality} \\
&\le \mathbb{E}\left[\!\left[(1^2+1^2)\left(\|\tilde{\boldsymbol{z}}^{j+1} - \tilde{\boldsymbol{y}}^{j+1}\|^2 + \|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2\right)\right]\!\right] \quad \text{By Cauchy-Schwarz inequality} \\
&\le \frac{2\tau^2 M^2\sigma^{-2}L^2R_{max}^2 e^2}{n}\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right] + 2\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right] \quad \text{By (39)}
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right] &\le 2\left(1 - \frac{2\tau^2 M^2\sigma^{-2}L^2R_{max}^2 e^2}{n}\right)^{-1}\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right] \\
&= 2(1-2c_0)^{-1}\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2\right]\!\right].
\end{aligned}$$

□

B.3 Proof of Theorem 5

Next, we bound the decrease of objective function value by

$$
\begin{aligned}
&F(\boldsymbol{z}^j) - F(\boldsymbol{z}^{j+1}) \\
&= \big(F(\boldsymbol{z}^j) - F(\boldsymbol{y}^{j+1})\big) + \big(F(\boldsymbol{y}^{j+1}) - F(\boldsymbol{z}^{j+1})\big) \\
&\geq \left(\frac{\sigma}{2}\|\boldsymbol{z}^j - \boldsymbol{y}^{j+1}\|^2\right) - \left(\frac{L_{max}}{2}\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^{j+1}\|^2\right) \\
&\qquad \text{by Proposition 5} \\
&\geq \left(\frac{\sigma}{2}\|\boldsymbol{z}^j - \boldsymbol{y}^{j+1}\|^2\right) - \left(\frac{L_{max}}{2}\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^{j+1}\|^2\right)
\end{aligned}
$$

So

$$
\begin{aligned}
&\mathbb{E}[\![F(\boldsymbol{z}^j)]\!] - \mathbb{E}[\![F(\boldsymbol{z}^{j+1})]\!] \\
&\geq \frac{\sigma}{2n}\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2]\!] - \frac{L_{max}}{2n}\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\|^2]\!] \qquad \text{by Proposition 1} \\
&\geq \frac{\sigma}{2n}\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2]\!] - \frac{L_{max}}{2n}\frac{\tau^2M^2\sigma^{-2}L^2R_{max}^2e^2}{n}\mathbb{E}[\![\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|^2]\!] \qquad \text{by (39)} \\
&\geq \frac{\sigma}{2n}\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2]\!] - \frac{2L_{max}}{2n}\frac{\tau^2M^2\sigma^{-2}L^2R_{max}^2e^2}{n}\left(1 - \frac{2\tau^2M^2\sigma^{-2}L^2R_{max}^2e^2}{n}\right)^{-1}\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2]\!] \quad \text{by (40)} \\
&\geq \frac{\sigma}{2n}\left(1 - \frac{2L_{max}}{\sigma}\left(1 - \frac{2\tau^2M^2\sigma^{-2}L^2R_{max}^2e^2}{n}\right)^{-1}\left(\frac{\tau^2M^2\sigma^{-2}L^2R_{max}^2e^2}{n}\right)\right)\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2]\!] \qquad (41)
\end{aligned}
$$

Let $b = \frac{\sigma}{2n}\left(1 - \frac{2L_{max}}{\sigma}\left(1 - \frac{2\tau^2M^2\sigma^{-2}L^2R_{max}^2e^2}{n}\right)^{-1}\left(\frac{\tau^2M^2\sigma^{-2}L^2R_{max}^2e^2}{n}\right)\right)$ and combining the above inequality with Eq (7) we have

$$
\begin{aligned}
\mathbb{E}[\![F(\boldsymbol{z}^j)]\!] - \mathbb{E}[\![F(\boldsymbol{z}^{j+1})]\!] &\geq b\,\mathbb{E}\left[\!\!\left[\left\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\right\|^2\right]\!\!\right] \\
&= b\,\mathbb{E}\left[\!\!\left[\left\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\right\|^2\right]\!\!\right] \qquad \text{by the definition in Appendix B.1.1} \\
&\geq \frac{b}{\kappa^2}\mathbb{E}\left[\!\!\left[\left\|\boldsymbol{z}^j - P_S(\boldsymbol{z}^j)\right\|^2\right]\!\!\right] \qquad \text{by Eq. (7)} \\
&\geq \frac{b}{\kappa^2L_{max}}\mathbb{E}[\![F(\boldsymbol{z}^j) - F^*]\!], \qquad \text{by the smooth assumption on F}
\end{aligned}
$$

where F^* is the global minimum of $F(\boldsymbol{z})$. Therefore, we have

$$
\begin{aligned}
&\mathbb{E}[\![F(\boldsymbol{z}^{j+1})]\!] - F^* \\
&= \mathbb{E}[\![F(\boldsymbol{z}^j)]\!] - \big(\mathbb{E}[\![F(\boldsymbol{z}^j)]\!] - \mathbb{E}[\![F(\boldsymbol{z}^{j+1})]\!]\big) - F^* \\
&\leq (1 - \frac{b}{\kappa^2L_{max}})(\mathbb{E}[\![F(\boldsymbol{z}^j)]\!] - F^*) \\
&\leq \eta\big(\mathbb{E}[\![F(\boldsymbol{z}^j)]\!] - F^*\big),
\end{aligned}
$$

where $\eta = 1 - \frac{b}{\kappa^2L_{max}}$.

C Proof for the Linear Convergence for AUX-PCD with Non-smooth f_i

C.1 Notations and Propositions

C.1.1 Propositions

Proposition 6. *Let $\bar{M} := \sigma^{-1}(LR_{max}^2 + 2LMR_{max}) \geq 1$, $\bar{q} = \frac{2(\tau+1)e\bar{M}}{n}$, $\bar{\rho} = 1 + \bar{q}$, and $\bar{\theta} = \sum_{t=1}^{\tau} \bar{\rho}^t$. If $\bar{q}(\tau+1) \leq 1$, then $\bar{\rho}^{(\tau+1)} \leq e$, and*

$$\bar{\rho}^{-1} \leq 1 - \frac{1 + 1\bar{M} + 1\bar{M}\bar{\theta}}{n}. \tag{42}$$

Proof. The proof is similar to the one for Proposition 4. We include here for completeness. By the definition of $\bar{\rho}$ and the condition $\bar{q}(\tau+1) \leq 1$, we have

$$\begin{aligned}\bar{\rho}^{(\tau+1)} &= \left((\bar{\rho})^{1/\bar{q}}\right)^{\bar{q}(\tau+1)} \\ &= \left((1+\bar{q})^{1/\bar{q}}\right)^{\bar{q}(\tau+1)} \leq e^{\bar{q}(\tau+1)} \leq e.\end{aligned}$$

By the definition of $\bar{q}$, we know that

$$\bar{q} = \bar{\rho} - 1 = \frac{2(\tau+1)e\bar{M}}{n} \Rightarrow 2 = \frac{n(\bar{\rho}-1)}{(\tau+1)e\bar{M}}.$$

We can derive

$$\begin{aligned}2 &= \frac{n(\bar{\rho}-1)}{(\tau+1)e\bar{M}} \\ &\leq \frac{n(\bar{\rho}-1)}{(\tau+1)\bar{\rho}^{(\tau+1)}\bar{M}} && \because \bar{\rho}^{(\tau+1)} \leq e \\ &\leq \frac{n(\bar{\rho}-1)}{(1+\bar{\theta})\bar{\rho}\bar{M}} && \because 1 + \bar{\theta} = \sum_{t=0}^{\tau} \bar{\rho}^t \leq (\tau+1)\bar{\rho}^\tau \\ &= \frac{n(1-\bar{\rho}^{-1})}{(1+\bar{\theta})\bar{M}}\end{aligned}$$

Combining the condition that $\bar{M} \geq 1$ and $1 + \bar{\theta} \geq 1$, we have

$$\frac{n(1-\bar{\rho}^{-1}) - 1}{(1+\bar{\theta})M} \geq \frac{n(1-\bar{\rho}^{-1})}{(1+\bar{\theta})M} - 1 \geq 1,$$

which leads to

$$\begin{aligned}(1+\bar{\theta})\bar{M} &\leq n - n\bar{\rho}^{-1} - 1 \\ \bar{\rho}^{-1} &\leq 1 - \frac{1 + \bar{M} + \bar{M}\bar{\theta}}{n}.\end{aligned}$$

□

Proposition 7. *With the $\bar{L}$-Lipschitz continuous assumption on f_i, we have the following results. For all $j > 0$, we have*

$$F(\boldsymbol{z}^{j+1}) \leq F(\boldsymbol{y}^{j+1}) + 2\bar{L}\left\|\boldsymbol{z}^{j+1} - \boldsymbol{y}^{j+1}\right\|_1 + \frac{LR_{max}^2}{2}\left\|\boldsymbol{z}^{j+1} - \boldsymbol{y}^{j+1}\right\|^2. \tag{43}$$

Furthermore, for any given $\boldsymbol{y} \in \mathbb{R}^n$, we have:

$$F(\boldsymbol{y}) \leq F(\boldsymbol{z}^*) + 2\bar{L}\|\boldsymbol{y} - \boldsymbol{z}^*\|_1 + \frac{LM^2}{2}\|\boldsymbol{y} - \boldsymbol{z}^*\|^2. \tag{44}$$

where $\boldsymbol{z}^*$ is the optimal solution.

Proof. Let $\bar{d}(u) = D(\bar{\boldsymbol{r}} + (u - s)\boldsymbol{q}_i)$, where $i = i(j)$ is the index selected at the j-th iterate and $s = z_i^j$. Thus, we have $g^j(u) = f_i(u) + \bar{d}(u)$. We first establish the Lipschitz smoothness for $\bar{d}(u)$ as follows.

$$
\begin{aligned}
\left|\bar{d}'(u_1) - \bar{d}'(u_2)\right| &= \left|\boldsymbol{q}_i^\top(\nabla D(\bar{\boldsymbol{r}} + (u_1 - s)\boldsymbol{q}_i)) - \boldsymbol{q}_i^\top(\nabla D(\bar{\boldsymbol{r}} + (u_2 - s)\boldsymbol{q}_i))\right| \\
&\leq \|\boldsymbol{q}_i\| \left\| \begin{bmatrix} \vdots \\ d_k'(\bar{r}_k + (u_1 - s)q_{ik}) - d_k'(\bar{r}_k + (u_2 - s)q_{ik}) \\ \vdots \end{bmatrix} \right\| \\
&\leq \|\boldsymbol{q}_i\| \left\| \begin{bmatrix} \vdots \\ L|(u_1 - u_2)q_{ik}| \\ \vdots \end{bmatrix} \right\| \qquad \text{by } L\text{-smoothness assumption of } d_k(r) \text{ for all } k \\
&\leq \|\boldsymbol{q}_i\|\|\boldsymbol{q}_i\| L |u_1 - u_2| \\
&\leq R_{max}^2 L |u_1 - u_2|
\end{aligned}
$$

Let u^* be the optimal solution for the single variable problem, we have the following result based on the optimality condition.

$$
\begin{aligned}
& 0 \in \partial f_i(u^*) + \bar{d}'(u^*) \\
\Rightarrow\ & -\bar{d}'(u^*) \in \partial f_i(u^*) \\
\Rightarrow\ & \left|\bar{d}'(u^*)\right| \leq \bar{L} \qquad \text{by the } \bar{L}\text{-Lipschitz continuity assumption of } f_i.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
g^j(u) - g^j(u^*) &= f_i(u) - f_i(u^*) + \bar{d}(u) - \bar{d}(u^*) \\
&\leq \bar{L}|u - u^*| + \bar{d}'(u^*)(u - u^*) + \frac{1}{2}R_{max}^2 L (u - u^*)^2 \\
&\leq \bar{L}|u - u^*| + \left|\bar{d}'(u^*)\right||u - u^*| + \frac{1}{2}R_{max}^2 L (u - u^*)^2 \\
&\leq 2\bar{L}|u - u^*| + \frac{1}{2}R_{max}^2 L (u - u^*)^2
\end{aligned}
$$

By the definitions of g^j, $\boldsymbol{z}^{j+1}$, and $\boldsymbol{y}^{j+1}$, we know that $u^* = y_{i(j)}^{j+1}$ and

$$
\begin{aligned}
g^j(z_{i(j)}^{j+1}) - g^j(y_{i(j)}^{j+1}) &= F(\boldsymbol{z}^{j+1}) - F(\boldsymbol{y}^{j+1}), \\
\|z_{i(j)}^{j+1} - y_{i(j)}^{j+1}\|^2 &= \|\boldsymbol{z}^{j+1} - \boldsymbol{y}^{j+1}\|^2,
\end{aligned}
$$

which implies (43). To derive (44), we first define $\bar{d}(\boldsymbol{z}) = D(Q\boldsymbol{z})$ and derive the smoothness of $\bar{d}(\boldsymbol{z})$ as follows

$$
\begin{aligned}
\left\|\nabla\bar{d}(\boldsymbol{z}_1) - \nabla\bar{d}(\boldsymbol{z}_2)\right\| &= \left\|Q^\top(\nabla D(Q\boldsymbol{z}_1) - Q^\top(\nabla D(Q\boldsymbol{z}_2))\right\| \\
&\leq \|Q\| \left\| \begin{bmatrix} \vdots \\ d_k'(\bar{\boldsymbol{q}}_k^\top \boldsymbol{z}_1) - d_k'(\bar{\boldsymbol{q}}_k^\top \boldsymbol{z}_2) \\ \vdots \end{bmatrix} \right\| \qquad \text{where } \bar{\boldsymbol{q}}_k^\top \text{ is the } k\text{-th row of } Q \\
&\leq \|Q\| \left\| \begin{bmatrix} \vdots \\ L\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|\|\bar{\boldsymbol{q}}_k\| \\ \vdots \end{bmatrix} \right\| \qquad \text{by } L\text{-smoothness assumption of } d_k(r) \text{ for all } k \\
&\leq \|Q\|\|Q\|_F L \|\boldsymbol{z}_1 - \boldsymbol{z}_2\| \\
&\leq \|Q\|_F^2 L \|\boldsymbol{z}_1 - \boldsymbol{z}_2\| \qquad \because \|Q\|_2 \leq \|Q\|_F
\end{aligned}
$$

By the optimality condition of $\boldsymbol{z}^*$ on $F(\boldsymbol{z}) = \sum_{i=1}^n f_i(z_i) + \bar{d}(\boldsymbol{z})$, we know that

$$
-\nabla_i \bar{d}(\boldsymbol{z}^*) \in \partial f_i(z_i^*) \Rightarrow \left|\nabla_i \bar{d}(\boldsymbol{z}^*)\right| \leq \bar{L}.
$$

Thus, we have

$$\begin{aligned} F(\boldsymbol{z}) - F(\boldsymbol{z}^*) &= \sum_{i=1}^{n} (f_i(z_i) - f_i(z_i^*)) + \bar{d}(\boldsymbol{z}) - \bar{d}\boldsymbol{z}^* \\ &\leq \sum_{i=1}^{n} \bar{L}|z_i - z_i^*| + \sum_{i=1}^{n} \nabla_i \bar{d}(\boldsymbol{z}^*)(z_i - z_i^*) + \frac{\|Q\|_F^2 L}{2} \|\boldsymbol{z} - \boldsymbol{z}^*\|^2 \\ &\leq \sum_{i=1}^{n} \bar{L}|z_i - z_i^*| + \sum_{i=1}^{n} \left|\nabla_i \bar{d}(\boldsymbol{z}^*)\right| |z_i - z_i^*| + \frac{\|Q\|_F^2 L}{2} \|\boldsymbol{z} - \boldsymbol{z}^*\|^2 \\ &\leq 2\bar{L}\|\boldsymbol{z} - \boldsymbol{z}^*\|_1 + \frac{\|Q\|_F^2 L}{2} \|\boldsymbol{z} - \boldsymbol{z}^*\|^2, \end{aligned}$$

which implies (44). □

C.2 Lemma 3 and Lemma 4

To prove the convergence of asynchronous algorithms for non-smooth f_i, we first show that the expected step size does not increase super-linearly by the following Lemma 3.

Lemma 3. *If τ is small enough such that (45) holds (i.e.,)*

$$(\tau + 1)e\bar{M}/n \leq 1, \tag{45}$$

then $\{\boldsymbol{z}^j\}$ generated by Algorithm 2 with atomic operations satisfies the following inequality:

$$\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \boldsymbol{z}^j\|]\!] \leq \bar{\rho}\, \mathbb{E}[\![\|\boldsymbol{z}^j - \boldsymbol{z}^{j+1}\|]\!], \tag{46}$$

where $\bar{\rho} = (1 + \frac{(\tau+1)e\bar{M}}{n})$, and $\bar{M} = \sigma^{-1}\big(LR_{max}^2 + 2LMR_{max}\big)$.

Proof. The proof is similar to Lemma 1. We prove Eq. (46) by induction. First, by triangle inequality we know that for any two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, we have

$$\| - \boldsymbol{a}\| - \| - \boldsymbol{b}\| \leq \|\boldsymbol{b} - \boldsymbol{a}\|.$$

Let $\boldsymbol{b} = \boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}$ and $\boldsymbol{a} = \boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j$. By (33), we have

$$\begin{aligned} &\left\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\right\| - \left\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\right\| \leq \|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^{j-1} + \tilde{\boldsymbol{z}}^j\| \\ &\leq \underbrace{(1 + \bar{M})}_{A} \|\boldsymbol{z}^j - \boldsymbol{z}^{j-1}\| + \underbrace{2\sigma^{-1} LMR_{max}}_{B} \sum_{t=j-\tau-1}^{j-2} \|\Delta z_t\| \end{aligned} \tag{47}$$

Now we prove (46) by induction.

Induction Hypothesis. Using Proposition 1, we prove the following equivalent statement. For all j,

$$\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|]\!] \leq \bar{\rho}\, \mathbb{E}[\![\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\|]\!], \tag{48}$$

Induction Basis. When $j = 1$,

$$\left\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\right\| - \left\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\right\| \leq \|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2 + \boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\| \leq A\|\boldsymbol{z}^1 - \boldsymbol{z}^0\|.$$

Taking the expectation of the above equation we have

$$\mathbb{E}[\![\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\| - \|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\|]\!] \leq A\mathbb{E}[\![\|\boldsymbol{z}^1 - \boldsymbol{z}^0\|]\!] = \frac{A}{n}\, \mathbb{E}[\![\|\tilde{\boldsymbol{z}}^1 - \boldsymbol{z}^0\|]\!]$$

Thus, we have

$$
\begin{aligned}
(1 - A/n)\,\mathbb{E}[\![\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1)\|]\!] &\leq \mathbb{E}[\![\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\|]\!] \\
\mathbb{E}[\![\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\|]\!] &\leq (1 - A/n)^{-1}\,\mathbb{E}[\![\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\|]\!] \\
\mathbb{E}[\![\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\|]\!] &\leq \big(1 - (1 + \bar{M})/n\big)^{-1}\,\mathbb{E}[\![\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\|]\!] \\
\mathbb{E}[\![\|\boldsymbol{z}^0 - \tilde{\boldsymbol{z}}^1\|]\!] &\leq \bar{\rho}\,\mathbb{E}[\![\|\boldsymbol{z}^1 - \tilde{\boldsymbol{z}}^2\|]\!] \qquad \text{by Proposition 6}
\end{aligned}
$$

Induction Step. By the induction hypothesis, we assume

$$
\mathbb{E}[\![\|\boldsymbol{z}^{t-1} - \tilde{\boldsymbol{z}}^t\|]\!] \leq \bar{\rho}\,\mathbb{E}[\![\|\boldsymbol{z}^t - \tilde{\boldsymbol{z}}^{t+1}\|]\!] \quad \forall t \leq j-1,
$$

which implies

$$
\begin{aligned}
\mathbb{E}[\![\|\boldsymbol{z}^{t-1} - \boldsymbol{z}^t\|]\!] &= \frac{1}{n}\,\mathbb{E}[\![\|\boldsymbol{z}^{t-1} - \boldsymbol{z}^t\|]\!] \\
&\leq \frac{1}{n}\bar{\rho}^{j-1-t}\,\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|]\!] \quad \forall t \leq j-1. \qquad (49)
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
&\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|]\!] - \mathbb{E}[\![\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\|]\!] \leq \mathbb{E}[\![\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^{j-1} + \tilde{\boldsymbol{z}}^j\|]\!] \qquad (50) \\
&\leq A\,\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|]\!] + B \sum_{t=j-\tau-1}^{j-2} \mathbb{E}[\![\|\boldsymbol{z}^t - \boldsymbol{z}^{t-1}\|]\!] \\
&\leq \frac{A}{n}\,\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|]\!] + B \sum_{t=j-\tau-1}^{j-2} \bar{\rho}^{(j-1-t)} \frac{1}{n}\,\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|]\!] \qquad \text{by (49)} \\
&\leq \frac{A}{n}\,\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|]\!] + \frac{B\bar{\theta}}{n}\,\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|]\!] \qquad \text{See definition of } \bar{\theta} \text{ in Proposition 6} \\
&\leq n^{-1}\big(A + B\bar{\theta}\big)\,\mathbb{E}[\![\|\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j\|]\!] \qquad (51)
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\mathbb{E}[\![\boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j]\!] &\leq \big(1 - n^{-1}\big(A + B\bar{\theta}\big)\big)^{-1}\,\mathbb{E}[\![\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\|]\!] \\
&\leq \big(1 - n^{-1}\big(1 + \bar{M} + \big(2\sigma^{-1}LMR_{max}\big)\bar{\theta}\big)\big)^{-1}\,\mathbb{E}[\![\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\|]\!] \\
&\leq \big(1 - n^{-1}\big(1 + \bar{M} + \big(\sigma^{-1}LR_{max}^2 + 2\sigma^{-1}LMR_{max}\big)\bar{\theta}\big)\big)^{-1}\,\mathbb{E}[\![\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\|]\!] \\
&\leq \big(1 - n^{-1}\big(1 + \bar{M} + \bar{M}\bar{\theta}\big)\big)^{-1}\,\mathbb{E}[\![\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\|]\!] \\
&\leq \bar{\rho}\,\mathbb{E}[\![\|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}\|]\!] \qquad \text{by Proposition 6.}
\end{aligned}
$$

□

Lemma 4. *Let $\{\boldsymbol{z}^j\}$ be the sequence generated by Algorithm 2 with atomic operations. Assume the conditions in Lemma 3 hold and let*

$$
c_1 = \frac{\tau\sigma^{-1}LMeR_{max}}{n} \quad \textit{and} \quad \hat{\rho} = (1 + c_1)(1 - c_1)^{-1}\left(1 + \frac{1 + \bar{M} + \bar{M}\bar{\theta}}{n}\right),
$$

we have

$$
\begin{aligned}
\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\|]\!] &\leq c_1\,\mathbb{E}[\![\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|]\!] \qquad (52) \\
\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|]\!] &\leq (1 + c_1)\,\mathbb{E}[\![\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|]\!] \qquad (53) \\
\mathbb{E}[\![\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|]\!] &\leq (1 - c_1)^{-1}\,\mathbb{E}[\![\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|]\!] \qquad (54) \\
\mathbb{E}[\![\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\|]\!] &\leq \hat{\rho}\,\mathbb{E}[\![\|T(\boldsymbol{z}^{j-1}) - \boldsymbol{z}^{j-1}\|]\!] \qquad (55)
\end{aligned}
$$

Proof. Note that $T(\boldsymbol{z}^j) = \tilde{\boldsymbol{y}}^{j+1}$ and $T(\boldsymbol{z}^{j-1}) = \tilde{\boldsymbol{y}}^j$. We first prove (52) as follows.

$$\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\|\right]\!\right] \le \frac{\tau\sigma^{-1}LMR_{max}e}{n}\,\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right].$$

$$\begin{aligned}
&\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\|\right]\!\right] \\
&\le \mathbb{E}\left[\!\left[\sum_{t=1}\left|T_t(\bar{\boldsymbol{r}}^j, z_t^j) - T_t(\hat{\boldsymbol{r}}^j, z_t^j)\right|\right]\!\right] \qquad \text{by } \|\cdot\|_2 \le \|\cdot\|_1 \\
&\le \sigma^{-1}LM\,\mathbb{E}\left[\!\left[\|\bar{\boldsymbol{r}}^j - \hat{\boldsymbol{r}}^j\|\right]\!\right] \qquad \text{by Proposition 3} \\
&\le \sigma^{-1}LMR_{max}\sum_{t=j-\tau}^{j-1}\mathbb{E}\left[\!\left[\|\boldsymbol{z}^{t+1} - \boldsymbol{z}^t\|\right]\!\right] \qquad \text{by Proposition 2} \\
&\le \sigma^{-1}LMR_{max}\sum_{t=1}^{\tau}\mathbb{E}\left[\!\left[\bar{\rho}^t\|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \qquad \text{by Lemma 3} \\
&\le \sigma^{-1}LMR_{max}n^{-1}\left(\sum_{t=1}^{\tau}\bar{\rho}^t\right)\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \qquad \text{by Proposition 1} \\
&\le \sigma^{-1}LMR_{max}\tau n^{-1}\bar{\rho}^\tau\,\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \qquad \because \bar{\rho} \ge 1 \\
&\le \sigma^{-1}LMR_{max}\tau n^{-1}e\,\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \qquad \because \bar{\rho}^\tau \le \bar{\rho}^{\tau+1} \le e \text{ by Proposition 6}
\end{aligned}$$

To prove (53), we apply the triangle inequality as follows:

$$\begin{aligned}
\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] &= \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1} + \tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \\
&\le \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\|\right]\!\right] + \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \\
&\le \left(1 + \frac{\tau\sigma^{-1}LMR_{max}e}{n}\right)\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \qquad \text{by (52)}
\end{aligned}$$

To prove (54), we applying the triangle inequality again as follows:

$$\begin{aligned}
\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] &= \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \tilde{\boldsymbol{y}}^{j+1} + \tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \\
&\le \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \tilde{\boldsymbol{y}}^{j+1}\|\right]\!\right] + \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \\
&\le \left(\frac{\tau\sigma^{-1}LMR_{max}e}{n}\right)\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] + \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \qquad \text{by (52)}
\end{aligned}$$

Thus, we have

$$\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] \le \left(1 - \frac{\tau\sigma^{-1}LMR_{max}e}{n}\right)^{-1}\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right].$$

To prove (55), it suffices to bound the R.H.S. of (53). Based on the triangle inequality $\|\boldsymbol{a}\| - \|\boldsymbol{b}\| \le \|\boldsymbol{a} - \boldsymbol{b}\|$ with $\boldsymbol{a} = \boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1}$ and $\boldsymbol{b} = \boldsymbol{z}^{j-1} - \tilde{\boldsymbol{z}}^j$, we have

$$\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\| - \|\tilde{\boldsymbol{z}}^j - \boldsymbol{z}^{j-1}\| \le \|\boldsymbol{z}^j - \tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^{j-1} + \tilde{\boldsymbol{z}}^j\|,$$

where the expectation of the R.H.S of the above inequality is exactly the same as the R.H.S. of (50). Thus the (51) is also an upper bound of the L.H.S of the above inequality as follows

$$\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] - \mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^j - \boldsymbol{z}^{j-1}\|\right]\!\right] \le \left(n^{-1}(A + B\bar{\theta})\right)\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^j - \boldsymbol{z}^{j-1}\|\right]\!\right]$$

Thus, we have

$$\begin{aligned}
\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\|\right]\!\right] &\le \left(1 + \frac{1 + \bar{M} + \bar{M}\bar{\theta}}{n}\right)\mathbb{E}\left[\!\left[\|\tilde{\boldsymbol{z}}^j - \boldsymbol{z}^{j-1}\|\right]\!\right] \\
&\le \left(1 + \frac{1 + \bar{M} + \bar{M}\bar{\theta}}{n}\right)\left(1 - \frac{\tau\sigma^{-1}LMR_{max}e}{n}\right)^{-1}\mathbb{E}\,\|\tilde{\boldsymbol{y}}^j - \boldsymbol{z}^{j-1}\| \qquad \text{by (54)}
\end{aligned}$$

□

C.3 Proof of Theorem 1

Theorem 1. *Assume a function $F(\boldsymbol{z})$ in the family defined by (1) admits a global error bound from the beginning. Assume there is a constant $\bar{L}$ such that all f_i are $\bar{L}$-Lipschitz continuous. If the upper bound of the staleness τ is small enough such that the following two conditions hold:*

$$(3(\tau+1)^2 e\bar{M})/\sqrt{n} \le 1, \quad \textit{and} \quad \frac{2LM^2c_0}{\sigma(1-2c_0)} \le 1, \tag{56}$$

where $\bar{M} = \sigma^{-1}\big(LR_{max}^2 + 2LMR_{max}\big)$, and $c_0 = \frac{\tau^2 M^2 \sigma^{-2} L^2 R_{max}^2 e^2}{n}$ then Algorithm 2 with atomic operations has a global linear convergence rate in expectation, that is, there is a constant c_2 and $\eta < 1$ such that

$$\mathbb{E}\big[\!\big[F(\boldsymbol{z}^{j+1}) - F^* + c_2\big\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\big\|\big]\!\big] \le \eta\big(\mathbb{E}\big[\!\big[F(\boldsymbol{z}^j) - F^* + c_2\big\|T(\boldsymbol{z}^{j-1}) - \boldsymbol{z}^{j-1}\big\|\big]\!\big]\big), \tag{57}$$

where F^* is the minimum of $F(\boldsymbol{z})$.

First, based on Proposition 7 and the generalized error bound (7), for all $\boldsymbol{z}^j$, we have

$$\begin{aligned}
\big\|\boldsymbol{z}^j - \boldsymbol{z}^*\big\|^2 &\ge \frac{2}{LM^2}\big\{F(\boldsymbol{z}^j) - F^* - 2\bar{L}\sqrt{n}\big\|\boldsymbol{z}^j - P_S(\boldsymbol{z}^j)\big\|\big\} \\
&\ge \frac{2}{LM^2}\big\{F(\boldsymbol{z}^j) - F^* - 2\kappa\bar{L}\sqrt{n}\big\|\boldsymbol{z}^j - T(\boldsymbol{z}^j)\big\|\big\}.
\end{aligned} \tag{58}$$

Next, we bound the decrease of objective function value by

$$\begin{aligned}
&F(\boldsymbol{z}^j) - F(\boldsymbol{z}^{j+1}) \\
&= \big(F(\boldsymbol{z}^j) - F(\boldsymbol{y}^{j+1})\big) + \big(F(\boldsymbol{y}^{j+1}) - F(\boldsymbol{z}^{j+1})\big) \\
&\ge \Big(\frac{\sigma}{2}\big\|\boldsymbol{z}^j - \boldsymbol{y}^{j+1}\big\|^2\Big) - \Big(2\bar{L}\big\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^j\big\|_1 + \frac{LM^2}{2}\big\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^{j+1}\big\|^2\Big) \qquad \text{by Proposition 5 and Proposition 7} \\
&\ge \Big(\frac{\sigma}{2}\big\|\boldsymbol{z}^j - \boldsymbol{y}^{j+1}\big\|^2\Big) - \Big(2\bar{L}n^{1/2}\big\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^j\big\| + \frac{LM^2}{2}\big\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^{j+1}\big\|^2\Big) \qquad \because \|\cdot\|_1 \le \sqrt{n}\|\cdot\|_2
\end{aligned}$$

By taking the expectation, we have

$$\begin{aligned}
&\mathbb{E}\big[\!\big[F(\boldsymbol{z}^j) - F(\boldsymbol{z}^{j+1})\big]\!\big] \\
&\ge \frac{\sigma}{2n}\,\mathbb{E}\Big[\!\Big[\big\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\big\|^2\Big]\!\Big] - \frac{2\bar{L}}{\sqrt{n}}\,\mathbb{E}\big[\!\big[\big\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\big\|\big]\!\big] - \frac{LM^2}{2n}\,\mathbb{E}\Big[\!\Big[\big\|\tilde{\boldsymbol{y}}^{j+1} - \tilde{\boldsymbol{z}}^{j+1}\big\|^2\Big]\!\Big] \qquad \text{by Proposition 1} \\
&\ge \frac{\sigma}{2n}\,\mathbb{E}\Big[\!\Big[\big\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\big\|^2\Big]\!\Big] - \frac{2\bar{L}c_1}{\sqrt{n}}\,\mathbb{E}\big[\!\big[\big\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\big\|\big]\!\big] - \frac{LM^2c_0}{2n}\,\mathbb{E}\Big[\!\Big[\big\|\tilde{\boldsymbol{z}}^{j+1} - \boldsymbol{z}^j\big\|^2\Big]\!\Big] \qquad \text{by Lemma 2 and Lemma 4} \\
&\ge \frac{\sigma}{2n}\,\mathbb{E}\Big[\!\Big[\big\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\big\|^2\Big]\!\Big] - \frac{2\bar{L}c_1(1-c_1)^{-1}}{\sqrt{n}}\,\mathbb{E}\big[\!\big[\big\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\big\|\big]\!\big] - \frac{LM^2 2c_0(1-2c_0)^{-1}}{2n}\,\mathbb{E}\Big[\!\Big[\big\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\big\|^2\Big]\!\Big] \\
&\qquad \text{by Lemma 2 and Lemma 4} \\
&\ge \Big(\frac{\sigma}{2n} - \frac{LM^2c_0(1-2c_0)^{-1}}{n}\Big)\,\mathbb{E}\Big[\!\Big[\big\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\big\|^2\Big]\!\Big] - \frac{2\bar{L}c_1(1-c_1)^{-1}}{\sqrt{n}}\,\mathbb{E}\big[\!\big[\big\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\big\|\big]\!\big] \\
&\ge \frac{1}{\kappa^2}\Big(\frac{\sigma}{2n} - \frac{LM^2c_0(1-2c_0)^{-1}}{n}\Big)\,\mathbb{E}\Big[\!\Big[\big\|P_S(\boldsymbol{z}^j) - \boldsymbol{z}^j\big\|^2\Big]\!\Big] - \frac{2\bar{L}c_1(1-c_1)^{-1}}{\sqrt{n}}\,\mathbb{E}\big[\!\big[\big\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\big\|\big]\!\big] \qquad \text{by (7)} \\
&\ge a_0\,\mathbb{E}\big[\!\big[F(\boldsymbol{z}^j) - F^*\big]\!\big] - b_0\,\mathbb{E}\,\big\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\big\| \qquad \text{by (58)},
\end{aligned} \tag{59}$$

where

$$\begin{aligned}
a_0 &= \frac{2}{LM^2\kappa^2}\Big(\frac{\sigma}{2n} - \frac{LM^2c_0(1-2c_0)^{-1}}{n}\Big), \\
b_0 &= 2a_0\kappa\bar{L}\sqrt{n} + \frac{2\bar{L}c_1(1-c_1)^{-1}}{\sqrt{n}}.
\end{aligned}$$

Now let

$$c_2 = \frac{b_0 \hat{\rho}}{1 - a_0 - \hat{\rho}},$$

where $\hat{\rho}$ is the constant defined in Lemma 4. The definition of c_2 implies

$$(b_0 + c_2)\hat{\rho} = (1 - a_0)c_2. \tag{60}$$

Thus, we have

$$\begin{aligned}
&\mathbb{E}\llbracket F(\boldsymbol{z}^{j+1}) - F^* + c_2\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\|\rrbracket \\
=&\mathbb{E}\llbracket F(\boldsymbol{z}^j)\rrbracket - \mathbb{E}\llbracket F(\boldsymbol{z}^j) - F(\boldsymbol{z}^{j+1})\rrbracket - F^* + \mathbb{E}\llbracket c_2\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\|\rrbracket \\
\leq&\mathbb{E}\llbracket F(\boldsymbol{z}^j) - F^*\rrbracket - \left\{a_0\left(\mathbb{E}\llbracket F(\boldsymbol{z}^j) - F^*\rrbracket - b_0\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\|\right)\right\} + \mathbb{E}\llbracket c_2\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\|\rrbracket && \text{by (59)} \\
\leq&(1 - a_0)\mathbb{E}\llbracket F(\boldsymbol{z}^j) - F^*\rrbracket + (b_0 + c_2)\mathbb{E}\llbracket\|T(\boldsymbol{z}^j) - \boldsymbol{z}^j\|\rrbracket \\
\leq&(1 - a_0)\mathbb{E}\llbracket F(\boldsymbol{z}^j) - F^*\rrbracket + (b_0 + c_2)\hat{\rho}\mathbb{E}\llbracket\|T(\boldsymbol{z}^{j-1}) - \boldsymbol{z}^{j-1}\|\rrbracket && \text{by (55)} \\
=&(1 - a_0)\mathbb{E}\llbracket F(\boldsymbol{z}^j) - F^* + c_2\|T(\boldsymbol{z}^{j-1}) - \boldsymbol{z}^{j-1}\|\rrbracket && \text{by (60)}
\end{aligned}$$

D Convergence for AUX-PCD with Wild Updates

D.1 Modeling the Wild updates

In the wild process, at each iteration j, some of the writes to $\boldsymbol{r}$ is missing. Therefore, we define the $\boldsymbol{r}$ vector to be

$$r_t^{j+1} = r_t^j + Q_{t,i(j)}\Delta z_j \delta_t^j \quad \forall t,$$

where δ_t^j is a random variable with

$$\delta_t^j = \begin{cases} 1 & \text{with probability } \ 1 - \theta \\ 0 & \text{with probability } \ \theta \end{cases}$$

where $0 \leq \theta \ll 1$ is the missing rate for the writes to $\boldsymbol{r}$ vector.

We define the set $\mathcal{Z}^j$ to be all the updates that are not missing until step j:

$$\mathcal{Z}^j = \{(t,k) \mid t < j, \ k \in N(i(t)), \ \delta_k^t = 1\},$$

where $N(i)$ is the nonzero elements in $\boldsymbol{q}_i$. The set U^j is the updates in the observed auxiliary vector $\hat{\boldsymbol{r}}$ at iteration j, so

$$\begin{aligned}
\hat{\boldsymbol{r}}^j &= \boldsymbol{r}^0 + \sum_{(t,k)\in\mathcal{U}^j} Q_{k,i(t)}\Delta z_t \boldsymbol{e}_k \\
\boldsymbol{r}^j &= \boldsymbol{r}^0 + \sum_{(t,k)\in\mathcal{Z}^j} Q_{k,i(t)}\Delta z_t \boldsymbol{e}_k.
\end{aligned}$$

We assume that the delay of the writes is smaller than τ, so

$$\mathcal{Z}^{j-\tau} \subseteq \mathcal{U}^j \subseteq \mathcal{Z}^j.$$

Since there are missing updates,

$$\boldsymbol{r}^j = Q\boldsymbol{z}^j - \boldsymbol{\epsilon}^j,$$

where

$$\boldsymbol{\epsilon}^j = \sum_{t<j} \sum_{k\in N(i(t))} Q_{k,i(t)}\Delta z_{i(t)} \delta_k^t \boldsymbol{e}_k.$$

Note that most of the δ_k^t are 0 because the missing rate $\theta \ll 1$.

To model the behavior of the algorithm, we define the "shifted" objective function at the j-th iteration:

$$\begin{aligned} F^j(\boldsymbol{z}) &= \sum_{i=1}^{n} f_i(z_i) + \sum_{k=1}^{m} d_k((Q\boldsymbol{z} - \boldsymbol{\epsilon}^j)_k) \\ &= \sum_{i=1}^{n} f_i(z_i) + \sum_{k=1}^{m} d_k(r_k^j) \text{ where } \boldsymbol{r}^j = Q\boldsymbol{z}^j - \boldsymbol{\epsilon}^j. \end{aligned}$$

Clearly, $F^j(\boldsymbol{z})$ satisfies all the condition for the F function listed in the paper, so all the previous results hold for all $F^j(\boldsymbol{z})$.

At the j-th iteration, the algorithm updates the coordinate $i(j)$ by

$$z_t^{j+1} = \begin{cases} T_t(\hat{\boldsymbol{r}}^j, z_t^j) & \text{if } t = i(j) \\ z_t^j & \text{if } t \neq i(j). \end{cases}$$

And the update using the real $\boldsymbol{r}^j$ is

$$y_t^{j+1} = \begin{cases} T_t(\boldsymbol{r}^j, z_t^j) & \text{if } t = i(j) \\ z_t^j & \text{if } t \neq i(j). \end{cases}$$

It is easy to verify the following equality:

$$\begin{aligned} T_t(\boldsymbol{r}^j, z_t^j) &= \arg\min_u f_t(u) + D(\boldsymbol{r} + \delta \boldsymbol{q}_t) \\ &= \arg\min_u f_t(u) + D(Q\boldsymbol{z}^j - \boldsymbol{\epsilon}^j + \delta \boldsymbol{q}_t) \\ &= \arg\min_u F^j(\boldsymbol{z} + (u - z_i)\boldsymbol{e}_i) \end{aligned}$$

So at the j-th iteration, the update $T_t(\boldsymbol{r}^j, z_t^j)$ is actually trying to minimize $F^j(\boldsymbol{z})$ by one step of coordinate descent. Therefore, Theorem 5 holds if the assumptions in Theorem 5 hold. Instead of using the final result for Theorem 5, we are going to use the following inequalities from (41):

$$E[F^j(\boldsymbol{z}^j)] - E[F^j(\boldsymbol{z}^{j+1})] \geq bE[\|\tilde{\boldsymbol{y}}^{j+1} - \boldsymbol{z}^j\|^2],$$

$$\text{where } b = \tfrac{\sigma}{2n}\left(1 - \tfrac{2L_{max}}{\sigma}\left(1 - \tfrac{2\tau^2 M^2 \sigma^{-2} L^2 R_{max}^2 e^2}{n}\right)^{-1}\left(\tfrac{\tau^2 M^2 \sigma^{-2} L^2 R_{max}^2 e^2}{n}\right)\right).$$

Our goal is to show

$$E[F^{j+1}(\boldsymbol{z}^{j+1}) - F^{*(j+1)}] \leq \eta E[F^j(\boldsymbol{z}^j) - F^{*(j)}]$$

with some $\eta < 1$ under certain condition, where $F^{*(j)}$ is the optimal solution for F^j. This will imply:

1. $E[F^j(\boldsymbol{x}^j) - F^{*(j)}] \leq \eta^j(F^0(\boldsymbol{z}^0) - F^{*(0)})$, so the error converges linearly.
2. The program stop if we set the stopping condition such as $\|T^j(\boldsymbol{z}^j)\| < 10^-3$.
3. The total missing updates $\boldsymbol{\epsilon}$ converges linearly to a bounded value.

D.2 Lemmas and Proof of Theorem

Lemma 5. *Assume $\boldsymbol{z}^{*(j)}$ is the optimal solution of F^j, then*

$$\mathbb{E}\left[\left\|\boldsymbol{z}^{j+1} - \boldsymbol{y}^{j+1}\right\|^2\right] \leq \frac{2c_0(1-2c_0)^{-1}}{n}\mathbb{E}\left[\left\|T^j(\boldsymbol{z}^j) - \boldsymbol{z}^j\right\|^2\right] \tag{61}$$

$$\mathbb{E}\left[\left\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^{*(j)}\right\|^2\right] \leq \frac{2}{n}\left(\kappa\sigma^{-1}L\right)^2\left(R_{max}^4 + M^4\right)\mathbb{E}\left[\left\|T^j(\boldsymbol{z}^j) - \boldsymbol{z}^j\right\|^2\right] \tag{62}$$

$$\mathbb{E}\left[\left\|\boldsymbol{z}^{*(j)} - \boldsymbol{z}^{*(j+1)}\right\|^2\right] \leq \frac{2(1-2c_0)^{-1}}{n}\left(\kappa\sigma^{-1}LMR_{max}\right)^2\theta R_{max}^2\mathbb{E}\left[\left\|T^j(\boldsymbol{z}^j) - \boldsymbol{z}^j\right\|^2\right] \tag{63}$$

Proof. The derivation for (61):

$$
\begin{aligned}
&\mathbb{E}\left[\!\left[\left\|\boldsymbol{z}^{j+1}-\boldsymbol{y}^{j+1}\right\|^2\right]\!\right] \\
\leq&\frac{1}{n}\mathbb{E}\left[\!\left[\left\|\tilde{\boldsymbol{z}}^{j+1}-\tilde{\boldsymbol{y}}^{j+1}\right\|^2\right]\!\right] \qquad \text{by Proposition 1} \\
\leq&\frac{2c_0(1-2c_0)^{-1}}{n}\mathbb{E}\left[\!\left[\left\|\tilde{\boldsymbol{y}}^{j+1}-\boldsymbol{z}^{j}\right\|^2\right]\!\right] \qquad \text{by (39) and (40) in Lemma 2} \\
\leq&\frac{2c_0(1-2c_0)^{-1}}{n}\mathbb{E}\left[\!\left[\left\|T^j(\boldsymbol{z}^j)-\boldsymbol{z}^{j}\right\|^2\right]\!\right]
\end{aligned}
$$

The derivation for (62):

$$
\begin{aligned}
&\mathbb{E}\left[\!\left[\left\|\boldsymbol{y}^{j+1}-\boldsymbol{z}^{*(j)}\right\|^2\right]\!\right] \\
\leq&\frac{1}{n}\mathbb{E}\left[\!\left[\left\|\tilde{\boldsymbol{y}}^{j+1}-\boldsymbol{z}^{*(j)}\right\|^2\right]\!\right] \qquad \text{by Proposition 1} \\
\leq&\frac{1}{n}\mathbb{E}\left[\!\left[\left\|T^j(\boldsymbol{z}^j)-\boldsymbol{z}^{*(j)}\right\|^2\right]\!\right] \qquad \text{by definition} \\
\leq&\frac{1}{n}\mathbb{E}\left[\!\left[\left\|T^j(\boldsymbol{z}^j)-T^j(\boldsymbol{z}^{*(j)})\right\|^2\right]\!\right] \qquad \text{by } \boldsymbol{z}^{*(j)}=T^j(\boldsymbol{z}^{*(j)}) \\
=&\frac{1}{n}\mathbb{E}\left[\!\left[\left\|T(Q\boldsymbol{z}^j-\boldsymbol{\epsilon}^j,\boldsymbol{z}^j)-T(Q\boldsymbol{z}^{*(j)}-\boldsymbol{\epsilon}^j,\boldsymbol{z}^{*(j)})\right\|^2\right]\!\right] \\
\leq&\frac{1}{n}\mathbb{E}\left[\!\left[\left\{\left(\sigma^{-1}LM\right)\left\|Q\boldsymbol{z}^j-Q\boldsymbol{z}^{*(j)}\right\|+\left(\sigma^{-1}LR_{max}^2\right)\left\|\boldsymbol{z}^j-\boldsymbol{z}^{*(j)}\right\|\right\}^2\right]\!\right] \qquad \text{by Proposition 3} \\
\leq&\frac{1}{n}\mathbb{E}\left[\!\left[\left\{\left(\sigma^{-1}LM^2\right)\left\|\boldsymbol{z}^j-\boldsymbol{z}^{*(j)}\right\|+\left(\sigma^{-1}LR_{max}^2\right)\left\|\boldsymbol{z}^j-\boldsymbol{z}^{*(j)}\right\|\right\}^2\right]\!\right] \qquad \text{by } M=\|Q\|_F \\
\leq&\frac{1}{n}\mathbb{E}\left[\!\left[2\left\{\left(\sigma^{-1}LM^2\right)\left\|\boldsymbol{z}^j-\boldsymbol{z}^{*(j)}\right\|\right\}^2+2\left\{\left(\sigma^{-1}LR_{max}^2\right)\left\|\boldsymbol{z}^j-\boldsymbol{z}^{*(j)}\right\|\right\}^2\right]\!\right] \qquad \text{by } (a+b)^2\leq 2a^2+2b^2 \\
\leq&\frac{2}{n}\left(\sigma^{-1}L\right)^2\left(R_{max}^4+M^4\right)\mathbb{E}\left[\!\left[\left\|\boldsymbol{z}^j-\boldsymbol{z}^{*(j)}\right\|^2\right]\!\right] \\
\leq&\frac{2}{n}\left(\kappa\sigma^{-1}L\right)^2\left(R_{max}^4+M^4\right)\mathbb{E}\left[\!\left[\left\|T^j(\boldsymbol{z}^j)-\boldsymbol{z}^{j}\right\|^2\right]\!\right] \qquad \text{by (7)}
\end{aligned}
$$

The derivation for (63):

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\boldsymbol{z}^{*(j)}-\boldsymbol{z}^{*(j+1)}\right\|^{2}\right]\\
\leq&\kappa^{2}\mathbb{E}\left[\left\|T^{j+1}(\boldsymbol{z}^{*(j)})-\boldsymbol{z}^{*(j)}\right\|^{2}\right] \qquad \text{by (7)}\\
\leq&\kappa^{2}\mathbb{E}\left[\left\|T^{j+1}(\boldsymbol{z}^{*(j)})-T^{j}(\boldsymbol{z}^{*(j)})\right\|^{2}\right] \qquad \text{by } \boldsymbol{z}^{*(j)}=T^{j}(\boldsymbol{z}^{*(j)})\\
=&\kappa^{2}\mathbb{E}\left[\left\|T(Q\boldsymbol{z}^{*(j)}-\boldsymbol{\epsilon}^{j+1},\boldsymbol{z}^{*(j)})-T(Q\boldsymbol{z}^{*(j)}-\boldsymbol{\epsilon}^{j},\boldsymbol{z}^{*(j)})\right\|^{2}\right]\\
\leq&\left(\kappa\sigma^{-1}LMR_{max}\right)^{2}\mathbb{E}\left[\left\|\boldsymbol{\epsilon}^{j+1}-\boldsymbol{\epsilon}^{j}\right\|^{2}\right] \qquad \text{by Proposition 3}\\
\leq&\left(\kappa\sigma^{-1}LMR_{max}\right)^{2}\mathbb{E}\left[\sum_{k=1}^{m}(1-\delta_{k}^{j})Q_{k,i(j)^{2}}\Delta z_{j}^{2}\right]\\
\leq&\left(\kappa\sigma^{-1}LMR_{max}\right)^{2}\theta R_{max}^{2}\mathbb{E}\left[\left\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\right\|^{2}\right]\\
\leq&\left(\kappa\sigma^{-1}LMR_{max}\right)^{2}\mathbb{E}\left[\sum_{k=1}^{m}(1-\delta_{k}^{j})Q_{k,i(j)^{2}}\Delta z_{j}^{2}\right]\\
\leq&\left(\kappa\sigma^{-1}LMR_{max}\right)^{2}\theta R_{max}^{2}\mathbb{E}\left[\left\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\right\|^{2}\right]\\
\leq&\frac{1}{n}\left(\kappa\sigma^{-1}LMR_{max}\right)^{2}\theta R_{max}^{2}\mathbb{E}\left[\left\|\tilde{\boldsymbol{z}}^{j+1}-\boldsymbol{z}^{j}\right\|^{2}\right] \qquad \text{by Proposition 1}\\
\leq&\frac{2(1-2c_{0})^{-1}}{n}\left(\kappa\sigma^{-1}LMR_{max}\right)^{2}\theta R_{max}^{2}\mathbb{E}\left[\left\|\tilde{\boldsymbol{y}}^{j+1}-\boldsymbol{z}^{j}\right\|^{2}\right] \qquad \text{by (39) in Lemma 2}\\
=&\frac{2(1-2c_{0})^{-1}}{n}\left(\kappa\sigma^{-1}LMR_{max}\right)^{2}\theta R_{max}^{2}\mathbb{E}\left[\left\|T^{j}(\boldsymbol{z}^{j})-\boldsymbol{z}^{j}\right\|^{2}\right]
\end{aligned}
$$

□

D.3 Main Theorems for AUX-PCD with Wild Updates

We first prove that if $\{\boldsymbol{\epsilon}^{j}\}$ is bounded and convergent, then a limit point of $\{\boldsymbol{z}^{j}\}$ is the exact solution of a perturbed problem as follows:

Theorem 2. *If $\boldsymbol{\epsilon}^{j}$ converges to $\boldsymbol{\epsilon}^{\infty}$, and $\bar{\boldsymbol{z}}$ is a limit point of $\{\boldsymbol{z}^{j}\}$, then we have the following properties:*

- *$\bar{\boldsymbol{z}}$ is a minimizer of the following "perturbed" problem:*

$$\operatorname*{argmin}_{\boldsymbol{z}}\Big\{\sum_{i=1}^{n}f_i(z_i)+D(Q\boldsymbol{z}-\boldsymbol{\epsilon}^{\infty})\Big\}:=F^{\infty}(\boldsymbol{z})$$

- *Furthermore, the distance between real and computed solution can be bounded by*

$$\|\bar{\boldsymbol{z}}-P_S(\bar{\boldsymbol{z}})\|\leq\kappa\sigma^{-1}LM\|\boldsymbol{\epsilon}^{\infty}\|,$$

where $P_S(\cdot)$ is the projection on to the set of optimal solution of the original objective function F.

Proof. Taking $j\to\infty$, we can easily see that $F^{j}\to F^{\infty}$. Combining this with Theorem 3, we can conclude

$$\lim_{k\to\infty}F^{\infty}(\boldsymbol{z}^{k})=F^{\infty}(\boldsymbol{z}^{*(\infty)}),$$

where $\boldsymbol{z}^{*(\infty)}$ is the optimal solution of F^{∞}, which concludes the first result.

For 2., we can easily derive as follows:

$$\begin{aligned}\|\boldsymbol{z}^* - \hat{\boldsymbol{z}}\| &\le \kappa\|T(\bar{\boldsymbol{z}}) - \bar{\boldsymbol{z}}\| \\ &= \kappa\|T(\bar{\boldsymbol{z}}) - \hat{T}(\bar{\boldsymbol{z}})\| \\ &= \kappa\|T(Q\bar{\boldsymbol{z}}, \bar{\boldsymbol{z}}) - T(Q\bar{\boldsymbol{z}} - \boldsymbol{\epsilon}^\infty, \bar{\boldsymbol{z}})\| \\ &\le \kappa\sigma^{-1}LM\|\boldsymbol{\epsilon}^\infty\| \ \text{(Proposition 3)}\end{aligned}$$

□

Theorem 3. *Let $\boldsymbol{z}^{*(j+1)}, \boldsymbol{z}^{*(j)}$ be the optimal solution for F^{j+1}, F^j respectively. Assume F is L_{max}-smooth and the conditions in Theorem 5 hold. The sequence $\{\boldsymbol{z}^j\}$ generated by the wild algorithm satisfies*

$$\mathbb{E}\Big[\!\Big[F^{j+1}(\boldsymbol{z}^{j+1}) - F^{j+1}(\boldsymbol{z}^{*(j+1)})\Big]\!\Big] \le \bar{\eta}\mathbb{E}\Big[\!\Big[F^j(\boldsymbol{z}^j) - F^j(\boldsymbol{z}^{*(j)})\Big]\!\Big],$$

where

$$\begin{aligned}\bar{\eta} &= \bar{c}_0 + \theta\bar{c}_2, \\ \bar{c}_0 &= \frac{3L_{max}b^{-1}}{n}\left(c_0(1-2c_0)^{-1} + (\kappa\sigma^{-1}L)^2(R_{max}^4 + M^4)\right) \\ \bar{c}_2 &= \frac{3L_{max}b^{-1}}{n}(1-2c_0)^{-1}\left(\kappa\sigma^{-1}LMR_{max}^2\right).\end{aligned}$$

c_0 and b are the constants defined in Theorem 5, which depends on the bounded delay parameter τ. This implies the algorithm converges linearly when

$$\bar{c}_0 + \theta\bar{c}_2 < 1, \tag{64}$$

which can be satisfied when θ is small enough.

Proof. Based on Lemma 5 and the smoothness of F, we can derive the results as follows:

$$\begin{aligned}&\mathbb{E}\Big[\!\Big[F^{j+1}(\boldsymbol{z}^{j+1}) - F^{j+1}(\boldsymbol{z}^{*(j+1)})\Big]\!\Big] \le \frac{L_{max}}{2}\mathbb{E}\Big[\!\Big[\Big\|\boldsymbol{z}^{j+1} - \boldsymbol{z}^{*(j+1)}\Big\|^2\Big]\!\Big] \\ =&\frac{3L_{max}}{2}\mathbb{E}\Big[\!\Big[\|\boldsymbol{z}^{j+1} - \boldsymbol{y}^{j+1}\|^2\Big]\!\Big] + \frac{3L_{max}}{2}\mathbb{E}\Big[\!\Big[\Big\|\boldsymbol{y}^{j+1} - \boldsymbol{z}^{*(j)}\Big\|^2\Big]\!\Big] + \frac{3L_{max}}{2}\mathbb{E}\Big[\!\Big[\Big\|\boldsymbol{z}^{*(j)} - \boldsymbol{z}^{*(j+1)}\Big\|^2\Big]\!\Big] \\ \le&\underbrace{\frac{3L_{max}}{n}\left(c_0(1-2c_0)^{-1} + (\kappa\sigma^{-1}L)^2(R_{max}^4 + M^4) + \theta(1-2c_0)^{-1}\left(\kappa\sigma^{-1}LMR_{max}\right)^2 R_{max}^2\right)}_{:=C}\mathbb{E}\Big[\!\Big[\|T^j(\boldsymbol{z}^j) - \boldsymbol{z}^j\|^2\Big]\!\Big] \\ \le&b^{-1}C\mathbb{E}[\![F^j(\boldsymbol{z}^j) - F^j(\boldsymbol{z}^{j+1})]\!] \qquad \text{by (41)} \\ =&b^{-1}C\mathbb{E}\Big[\!\Big[F^j(\boldsymbol{z}^j) - F^j(\boldsymbol{z}^{*(j)}) + F^j(\boldsymbol{z}^{*(j)}) - F^j(\boldsymbol{z}^{j+1})\Big]\!\Big] \\ \le&b^{-1}C\mathbb{E}\Big[\!\Big[F^j(\boldsymbol{z}^j) - F^j(\boldsymbol{z}^{*(j)})\Big]\!\Big] \qquad \text{by } F^j(\boldsymbol{z}^{*(j)}) \le F^j(\boldsymbol{z}^{j+1}).\end{aligned}$$

By plugging the definition of b into the above equation, it is not hard to see that there are two constants $\bar{c}_0, \bar{c}_2$ such that $\bar{\eta} = b^{-1}C = \bar{c}_0 + \theta\bar{c}_2$. As long as $\bar{c}_0 + \theta\bar{c}_2 \le 1$, we have linear convergence. □

Theorem 4. *With Theorem 3, we can show that $\{\boldsymbol{\epsilon}^j\}$ converges, and*

$$\mathbb{E}[\![\|\boldsymbol{\epsilon}^\infty\|]\!] \le \theta R_{max}\sqrt{F(\boldsymbol{z}^0) - F^*}\left(\frac{1}{1-\sqrt{\bar{\eta}}}\right)\sqrt{\frac{1}{2\sigma}},$$

where $\bar{\eta} \in (0,1)$ is the linear convergence rate in Theorem 3.

Proof. Since each $f_i(z)$ is σ-strongly convex, we let g be the single variable function, and z be the current z_t^j, $u^* = \arg\min_u g(u)$, then

$$
\begin{aligned}
& g(z) \geq g(u^*) + \langle \nabla g(u^*), z - u^* \rangle + \frac{\sigma}{2}(u^* - z)^2 \\
\Rightarrow & |u^* - z| \leq \sqrt{\frac{2}{\sigma}} \sqrt{(g(z) - g(u^*))}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\mathbb{E}[\![\|\boldsymbol{\epsilon}^\infty\|]\!] &\leq \mathbb{E}\left[\!\!\left[\sum_{j=1}^{\infty} \|\boldsymbol{\epsilon}^{j+1} - \boldsymbol{\epsilon}^j\|\right]\!\!\right] \\
&\leq \mathbb{E}\left[\!\!\left[\sum_{j=1}^{\infty} \theta \|\boldsymbol{y}^{j+1} - \boldsymbol{z}^j\| \|\boldsymbol{q}_{i(j)}\|\right]\!\!\right] \\
&\leq \theta \mathbb{E}\left[\!\!\left[\sum_{j=1}^{\infty} \sqrt{F^j(\boldsymbol{z}^j) - F^j(\boldsymbol{y}^{j+1})}\right]\!\!\right] R_{max} \sqrt{\frac{2}{\sigma}} \\
&\leq \theta \mathbb{E}\left[\!\!\left[\sum_{j=1}^{\infty} \sqrt{F^j(\boldsymbol{z}^j) - F^{*(j)}}\right]\!\!\right] R_{max} \sqrt{\frac{2}{\sigma}} \\
&\leq \theta R_{max} \sqrt{\frac{2}{\sigma}} \frac{\sqrt{F(\boldsymbol{z}^0) - F^*}}{1 - \sqrt{\bar{\eta}}}.
\end{aligned}
$$

□