Online Supplementary Document:

"Multi-block Parameter Calibration in Computer Models"

Cheoljoon Jeong, Ziang Xu, Albert S. Berahas, Eunshin Byon, and Kristen Cetin

1 Proof for Lemma 1

After the M_b iterations for the b^{th} block optimization (inner iteration) at the $(k+1)^{th}$ outer iteration, the iterate for the b^{th} block parameter is updated from $\boldsymbol{\theta}_b^k$ to $\boldsymbol{\theta}_b^{k+1}$. By L_b-Lipschitz continuous gradient and $\alpha_b \in (0, 1/L_b]$, we have

$$
\begin{aligned}
&F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1})\\
&\quad\le F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) + \nabla F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1})^T(\boldsymbol{\theta}_b^{k,M_b+1}-\boldsymbol{\theta}_b^{k,M_b}) + \frac{1}{2}L_b\|\boldsymbol{\theta}_b^{k,M_b+1}-\boldsymbol{\theta}_b^{k,M_b}\|_2^2\\
&\quad= F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) - \alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2 + \frac{1}{2}\alpha_b^2 L_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2\\
&\quad\le F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) - \frac{1}{2}\alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2\\
&\quad\le F_b(\boldsymbol{\theta}_b^{k,M_b-1};\boldsymbol{\theta}_{-b}^{k+1}) + \nabla F_b(\boldsymbol{\theta}_b^{k,M_b-1};\boldsymbol{\theta}_{-b}^{k+1})^T(\boldsymbol{\theta}_b^{k,M_b}-\boldsymbol{\theta}_b^{k,M_b-1})\\
&\qquad+\frac{1}{2}L_b\|\boldsymbol{\theta}_b^{k,M_b}-\boldsymbol{\theta}_b^{k,M_b-1}\|_2^2 - \frac{1}{2}\alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2\\
&\quad\le F_b(\boldsymbol{\theta}_b^{k,M_b-1};\boldsymbol{\theta}_{-b}^{k+1}) - \alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,M_b-1};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2 + \frac{1}{2}\alpha_b^2 L_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,M_b-1};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2\\
&\qquad-\frac{1}{2}\alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2\\
&\quad\le\cdots\\
&\quad\le F_b(\boldsymbol{\theta}_b^{k,1};\boldsymbol{\theta}_{-b}^{k+1}) - \frac{1}{2}\alpha_b\sum_{m=1}^{M_b}\|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2.
\end{aligned}
\tag{A.1}
$$

With $F_b(\boldsymbol{\theta}_b^{k,1};\boldsymbol{\theta}_{-b}^{k+1}) = F_b(\boldsymbol{\theta}_b^{k};\boldsymbol{\theta}_{-b}^{k+1})$ and $F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1}) = F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1})$, the following inequality holds.

$$F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1}) \leq F_b(\boldsymbol{\theta}_b^{k};\boldsymbol{\theta}_{-b}^{k+1}) - \frac{1}{2}\alpha_b \sum_{m=1}^{M_b} \|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2, \tag{A.2}$$

for $b = 1, 2, \ldots, B$.

□

2 Proof for Theorem 1

By Lemma 1, we have

$$F_b(\boldsymbol{\theta}_b^{k};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1}) \geq \frac{1}{2}\alpha_b \sum_{m=1}^{M_b} \|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2. \tag{A.3}$$

By the conditions in (16),

$$\begin{aligned} F_b(\boldsymbol{\theta}^{k}) - F_b(\boldsymbol{\theta}^{k+1}) &\geq F_b(\boldsymbol{\theta}_b^{k};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1}) \\ &\geq \frac{1}{2}\alpha_b \sum_{m=1}^{M_b} \|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2. \end{aligned} \tag{A.4}$$

Summing over $k \in \{0, 1, \ldots, K\}$, by the telescopic sum, it follows that

$$F_b(\boldsymbol{\theta}^{0}) - F_b(\boldsymbol{\theta}^{K+1}) \geq \frac{1}{2}\alpha_b \sum_{k=0}^{K}\sum_{m=1}^{M_b} \|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2. \tag{A.5}$$

Taking $K \to \infty$, and by the conditions in (16), we have

$$\infty > F_b(\boldsymbol{\theta}^{0}) - F_b^{\inf} \geq \frac{1}{2}\alpha_b \sum_{k=0}^{\infty}\sum_{m=1}^{M_b} \|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2. \tag{A.6}$$

The above inequality implies

$$\|\nabla F_b(\boldsymbol{\theta}_b^{k};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2 \to 0 \quad as \;\; k \to \infty. \tag{A.7}$$

□

3 Proof for Corollary 1

By Lemma 1, we have

$$F_b(\boldsymbol{\theta}_b^k; \boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k+1}; \boldsymbol{\theta}_{-b}^{k+1}) \geq \frac{1}{2}\alpha_b \sum_{m=1}^{M_b} \|\nabla F_b(\boldsymbol{\theta}_b^{k,m}; \boldsymbol{\theta}_{-b}^{k+1})\|_2^2. \tag{A.8}$$

Summing over $k \in \{0, 1, \ldots, K\}$ on both sides, it follows that

$$\sum_{k=0}^{K} \left\{F_b(\boldsymbol{\theta}_b^k; \boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k+1}; \boldsymbol{\theta}_{-b}^{k+1})\right\} \geq \frac{1}{2}\alpha_b \sum_{k=0}^{K} \sum_{m=1}^{M_b} \|\nabla F_b(\boldsymbol{\theta}_b^{k,m}; \boldsymbol{\theta}_{-b}^{k+1})\|_2^2. \tag{A.9}$$

Taking $K \to \infty$, and by the condition in (18), we have

$$\infty > \lim_{K\to\infty} \sum_{k=0}^{K} \left\{F_b(\boldsymbol{\theta}_b^k; \boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k+1}; \boldsymbol{\theta}_{-b}^{k+1})\right\} \geq \frac{1}{2}\alpha_b \sum_{k=0}^{\infty} \sum_{m=1}^{M_b} \|\nabla F_b(\boldsymbol{\theta}_b^{k,m}; \boldsymbol{\theta}_{-b}^{k+1})\|_2^2. \tag{A.10}$$

The above inequality implies

$$\|\nabla F_b(\boldsymbol{\theta}_b^k; \boldsymbol{\theta}_{-b}^{k+1})\|_2^2 \to 0 \quad as \;\; k \to \infty. \tag{A.11}$$

□

4 Proof for Theorem 2

We use the fact that if a function $\phi(\mathbf{x})$ is μ-strongly convex, then $\phi(\mathbf{y}) \geq \phi(\mathbf{x}) + \nabla\phi(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{\mu}{2}\|\mathbf{y} - \mathbf{x}\|_2^2$ for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(\phi)$. This inequality implies

$$2\mu(\phi(\mathbf{x}) - \phi^*) \leq \|\nabla\phi(\mathbf{x})\|_2^2, \tag{A.12}$$

by minimizing both sides with respect to $\mathbf{y}$.

By L_b-Lipschitz continuous gradient, $\alpha_b \in (0, 1/L_b]$, and μ_b-strongly convexity for F_b, we have

$$\begin{aligned} F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1}) &\le F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) - \frac{1}{2}\alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2 \\ &\le F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) - \alpha_b\mu_b\big\{F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\}, \end{aligned} \tag{A.13}$$

where the first inequality is due to (A.1) in the proof of Lemma 1 and the last inequality holds with the property of μ_b-strongly convexity in (A.12).

Subtracting $F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})$ from both sides in (A.13), we obtain

$$\begin{aligned} F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1}) &\le F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1}) \\ &\quad - \alpha_b\mu_b\big\{F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\ &\le (1-\alpha_b\mu_b)\big\{F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\}. \end{aligned} \tag{A.14}$$

Applying (A.14) recursively with $F_b(\boldsymbol{\theta}_b^{k,1};\boldsymbol{\theta}_{-b}^{k+1}) = F_b(\boldsymbol{\theta}_b^{k};\boldsymbol{\theta}_{-b}^{k+1})$ and $F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1}) = F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1})$, we get

$$\begin{aligned} F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1}) &= F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1}) \\ &\le (1-\alpha_b\mu_b)\big\{F_b(\boldsymbol{\theta}_b^{k,M_b};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\ &\le (1-\alpha_b\mu_b)^2\big\{F_b(\boldsymbol{\theta}_b^{k,M_b-1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\ &\le \cdots \\ &\le (1-\alpha_b\mu_b)^{M_b}\big\{F_b(\boldsymbol{\theta}_b^{k,1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\ &= (1-\alpha_b\mu_b)^{M_b}\big\{F_b(\boldsymbol{\theta}_b^{k};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\}. \end{aligned} \tag{A.15}$$

By the conditions in (20), it follows that

$$\begin{aligned} |F_b(\boldsymbol{\theta}^{k+1}) - F_b(\boldsymbol{\theta}^{*})| &\le c_1\big\{F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\ &\le c_1(1-\alpha_b\mu_b)^{M_b}\big\{F_b(\boldsymbol{\theta}_b^{k};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\ &\le c_1c_2(1-\alpha_b\mu_b)^{M_b}|F_b(\boldsymbol{\theta}^{k}) - F_b(\boldsymbol{\theta}^{*})|. \end{aligned} \tag{A.16}$$

Using the above inequality (A.16) recursively, we obtain

$$\begin{aligned} |F_b(\boldsymbol{\theta}^k) - F_b(\boldsymbol{\theta}^*)| &\le c_1 c_2 (1-\alpha_b\mu_b)^{M_b} |F_b(\boldsymbol{\theta}^{k-1}) - F_b(\boldsymbol{\theta}^*)| \\ &\le (c_1c_2)^2 (1-\alpha_b\mu_b)^{2M_b} |F_b(\boldsymbol{\theta}^{k-2}) - F_b(\boldsymbol{\theta}^*)| \\ &\le \cdots \\ &\le (c_1c_2)^k (1-\alpha_b\mu_b)^{kM_b} |F_b(\boldsymbol{\theta}^{0}) - F_b(\boldsymbol{\theta}^*)|. \end{aligned} \tag{A.17}$$

Letting $\rho = c_1c_2(1-\alpha_b\mu_b)^{M_b}$ with $c_1, c_2 > 0$ and $c_1c_2 < \frac{1}{(1-\alpha_b\mu_b)^{M_b}}$, we have

$$|F_b(\boldsymbol{\theta}^k) - F_b(\boldsymbol{\theta}^*)| \le \rho^k |F_b(\boldsymbol{\theta}^{0}) - F_b(\boldsymbol{\theta}^*)|. \tag{A.18}$$

The resulting inequality implies linear convergence rate, i.e., $\mathcal{O}(\rho^k)$.

□

5 Proof for Corollary 2

With the inequalities (A.12)-(A.15) in the proof of Theorem 2 and by the conditions in (23), we have

$$\begin{aligned} F_b(\boldsymbol{\theta}^{k+1}) - F_b(\boldsymbol{\theta}^*) &= |F_b(\boldsymbol{\theta}^{k+1}) - F_b(\boldsymbol{\theta}^*)| \\ &\le c_1 \big\{ F_b(\boldsymbol{\theta}_b^{k+1}; \boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*}; \boldsymbol{\theta}_{-b}^{k+1}) \big\} \\ &\le c_1 (1-\alpha_b\mu_b)^{M_b} \big\{ F_b(\boldsymbol{\theta}_b^{k}; \boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*}; \boldsymbol{\theta}_{-b}^{k+1}) \big\} \\ &\le c_1 c_2 (1-\alpha_b\mu_b)^{M_b} |F_b(\boldsymbol{\theta}^{k}) - F_b(\boldsymbol{\theta}^*)| \\ &= c_1 c_2 (1-\alpha_b\mu_b)^{M_b} \big\{ F_b(\boldsymbol{\theta}^{k}) - F_b(\boldsymbol{\theta}^*) \big\}. \end{aligned} \tag{A.19}$$

because Lemma 1 and the conditions in (23) imply $F_b(\boldsymbol{\theta}^k) - F_b(\boldsymbol{\theta}^{k+1}) \ge 0$ for all k, which tells $F_b(\boldsymbol{\theta}^k) - F_b(\boldsymbol{\theta}^*) \ge 0$. Thus, the first and last inequality hold in (A.19).

Using the above inequality (A.19) recursively, we obtain

$$\begin{aligned} F_b(\boldsymbol{\theta}^k) - F_b(\boldsymbol{\theta}^*) &\leq c_1c_2(1-\alpha_b\mu_b)^{M_b}\big\{F_b(\boldsymbol{\theta}^{k-1}) - F_b(\boldsymbol{\theta}^*)\big\} \\ &\leq (c_1c_2)^2(1-\alpha_b\mu_b)^{2M_b}\big\{F_b(\boldsymbol{\theta}^{k-2}) - F_b(\boldsymbol{\theta}^*)\big\} \\ &\leq \cdots \\ &\leq (c_1c_2)^k(1-\alpha_b\mu_b)^{kM_b}\big\{F_b(\boldsymbol{\theta}^{0}) - F_b(\boldsymbol{\theta}^*)\big\}. \end{aligned} \tag{A.20}$$

Letting $\rho = c_1c_2(1-\alpha_b\mu_b)^{M_b}$ with $c_1, c_2 > 0$ and $c_1c_2 < \frac{1}{(1-\alpha_b\mu_b)^{M_b}}$, we have

$$F_b(\boldsymbol{\theta}^k) - F_b(\boldsymbol{\theta}^*) \leq \rho^k\big\{F_b(\boldsymbol{\theta}^{0}) - F_b(\boldsymbol{\theta}^*)\big\}, \tag{A.21}$$

The resulting inequality implies linear convergence rate, i.e., $\mathcal{O}(\rho^k)$. Additionally, under the conditions in (23), the function of the b^{th} block is monotonically decreasing across outer iterations, and as such this yields the result of Corollary 2. □

6 Proof for Theorem 3

We know the fact that if a function $\phi(\mathbf{x})$ is convex, then for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(\phi)$ we have

$$\phi(\mathbf{y}) \geq \phi(\mathbf{x}) + \nabla\phi(\mathbf{x})^T(\mathbf{y} - \mathbf{x}). \tag{A.22}$$

If we rearrange the inequality (A.22), we get

$$\phi(\mathbf{x}) \leq \phi(\mathbf{y}) + \nabla\phi(\mathbf{x})^T(\mathbf{x} - \mathbf{y}). \tag{A.23}$$

By L_b-Lipschitz continous gradient and $\alpha_b \in (0, 1/L_b]$, we have

$$F_b(\boldsymbol{\theta}_b^{k,m+1}; \boldsymbol{\theta}_{-b}^{k+1}) \leq F_b(\boldsymbol{\theta}_b^{k,m}; \boldsymbol{\theta}_{-b}^{k+1}) - \frac{1}{2}\alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,m}; \boldsymbol{\theta}_{-b}^{k+1})\|_2^2. \tag{A.24}$$

By the block-wise convexity for F_b, and from the inequality (A.24), we have

$$\begin{aligned}
F_b(\boldsymbol{\theta}_b^{k,m+1};\boldsymbol{\theta}_{-b}^{k+1}) &\leq F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1}) - \frac{1}{2}\alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2 \\
&\leq F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1}) + \nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})^T(\boldsymbol{\theta}_b^{k,m} - \boldsymbol{\theta}_b^{k,*}) - \frac{1}{2}\alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2.
\end{aligned} \tag{A.25}$$

Subtracting $F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})$ on both sides in (A.25) and completing the square, we get

$$\begin{aligned}
F_b(\boldsymbol{\theta}_b^{k,m+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1}) &\leq \nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})^T(\boldsymbol{\theta}_b^{k,m} - \boldsymbol{\theta}_b^{k,*}) - \frac{1}{2}\alpha_b\|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2 \\
&\leq \frac{1}{2\alpha_b}\big\{\|\boldsymbol{\theta}_b^{k,m} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k,m} - \boldsymbol{\theta}_b^{k,*}\|_2^2 \\
&\qquad + 2\alpha_b\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})^T(\boldsymbol{\theta}_b^{k,m} - \boldsymbol{\theta}_b^{k,*}) - \alpha_b^2\|\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2\big\} \\
&= \frac{1}{2\alpha_b}\big\{\|\boldsymbol{\theta}_b^{k,m} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k,m} - \boldsymbol{\theta}_b^{k,*} - \alpha_b\nabla F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1})\|_2^2\big\} \\
&= \frac{1}{2\alpha_b}\big\{\|\boldsymbol{\theta}_b^{k,m} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k,m+1} - \boldsymbol{\theta}_b^{k,*}\|_2^2\big\}.
\end{aligned} \tag{A.26}$$

Summing over all inner iterations, $m = 1, 2, \ldots, M_b$, by a telescopic sum, we have

$$\begin{aligned}
&M_b\big\{F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\
&\quad \leq \big\{F_b(\boldsymbol{\theta}_b^{k,2};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} + \ldots + \big\{F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\
&\quad \leq \frac{1}{2\alpha_b}\big\{\|\boldsymbol{\theta}_b^{k,1} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k,2} - \boldsymbol{\theta}_b^{k,*}\|_2^2 + \ldots + \|\boldsymbol{\theta}_b^{k,M_b} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k,M_b+1} - \boldsymbol{\theta}_b^{k,*}\|_2^2\big\} \\
&\quad \leq \frac{1}{2\alpha_b}\big\{\|\boldsymbol{\theta}_b^{k,1} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k,M_b+1} - \boldsymbol{\theta}_b^{k,*}\|_2^2\big\} \\
&\quad = \frac{1}{2\alpha_b}\big\{\|\boldsymbol{\theta}_b^{k} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k+1} - \boldsymbol{\theta}_b^{k,*}\|_2^2\big\},
\end{aligned} \tag{A.27}$$

where the first inequality is satisfied because

$$F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1}) \leq F_b(\boldsymbol{\theta}_b^{k,m+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})$$

holds for all m by the fact $F_b(\boldsymbol{\theta}_b^{k,m};\boldsymbol{\theta}_{-b}^{k+1}) \geq F_b(\boldsymbol{\theta}_b^{k,m+1};\boldsymbol{\theta}_{-b}^{k+1})$ from the inequality (A.24).

Rearranging the resulting inequality and using the facts that $F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1}) = F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1})$, we have

$$
\begin{aligned}
F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1}) &= F_b(\boldsymbol{\theta}_b^{k,M_b+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1}) \\
&\leq \frac{1}{2\alpha_b M_b}\big\{\|\boldsymbol{\theta}_b^{k} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k+1} - \boldsymbol{\theta}_b^{k,*}\|_2^2\big\}.
\end{aligned} \tag{A.28}
$$

By the conditions in (25), it follows that

$$
\begin{aligned}
|F_b(\boldsymbol{\theta}^{k+1}) - F_b(\boldsymbol{\theta}^{*})| &\leq c_1\big\{F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\
&\leq \frac{c_1}{2\alpha_b M_b}\big\{\|\boldsymbol{\theta}_b^{k} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k+1} - \boldsymbol{\theta}_b^{k,*}\|_2^2\big\} \\
&\leq \frac{c_1}{2\alpha_b M_b}\big\{\|\boldsymbol{\theta}_b^{k} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k+1} - \boldsymbol{\theta}_b^{k+1,*}\|_2^2\big\}.
\end{aligned} \tag{A.29}
$$

Summing over $\tilde{l} \in \{0, 1, \ldots, k-1\}$ and taking the average, by the telescopic sum, it follows that

$$
\begin{aligned}
\frac{1}{k}\sum_{l=1}^{k}|F_b(\boldsymbol{\theta}^{l}) - F_b(\boldsymbol{\theta}^{*})| &\leq \frac{c_1}{2\alpha_b k M_b}\sum_{\tilde{l}=0}^{k-1}\big\{\|\boldsymbol{\theta}_b^{\tilde{l}} - \boldsymbol{\theta}_b^{\tilde{l},*}\|_2^2 - \|\boldsymbol{\theta}_b^{\tilde{l}+1} - \boldsymbol{\theta}_b^{\tilde{l}+1,*}\|_2^2\big\} \\
&\leq \frac{c_1}{2\alpha_b k M_b}\big\{\|\boldsymbol{\theta}_b^{0} - \boldsymbol{\theta}_b^{0,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k} - \boldsymbol{\theta}_b^{k,*}\|_2^2\big\} \\
&\leq \frac{c_1}{2\alpha_b k M_b}\|\boldsymbol{\theta}_b^{0} - \boldsymbol{\theta}_b^{0,*}\|_2^2.
\end{aligned} \tag{A.30}
$$

The resulting inequality implies sublinear convergence rate, i.e., $\mathcal{O}(1/(kM_b))$.

□

7 Proof for Corollary 3

With the inequalities (A.22)-(A.28) in the proof of Theorem 3 and by the conditions in (27) and (28), we have

$$
\begin{aligned}
F_b(\boldsymbol{\theta}^{k+1}) - F_b(\boldsymbol{\theta}^{*}) &= |F_b(\boldsymbol{\theta}^{k+1}) - F_b(\boldsymbol{\theta}^{*})| \\
&\leq c_1\big\{F_b(\boldsymbol{\theta}_b^{k+1};\boldsymbol{\theta}_{-b}^{k+1}) - F_b(\boldsymbol{\theta}_b^{k,*};\boldsymbol{\theta}_{-b}^{k+1})\big\} \\
&\leq \frac{c_1}{2\alpha_b M_b}\big\{\|\boldsymbol{\theta}_b^{k} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k+1} - \boldsymbol{\theta}_b^{k,*}\|_2^2\big\} \\
&\leq \frac{c_1}{2\alpha_b M_b}\big\{\|\boldsymbol{\theta}_b^{k} - \boldsymbol{\theta}_b^{k,*}\|_2^2 - \|\boldsymbol{\theta}_b^{k+1} - \boldsymbol{\theta}_b^{k+1,*}\|_2^2\big\}.
\end{aligned} \tag{A.31}
$$

because Lemma 1 and the conditions in (28) imply $F_b(\boldsymbol{\theta}^k) - F_b(\boldsymbol{\theta}^{k+1}) \geq 0$ for all k, which tells $F_b(\boldsymbol{\theta}^k) - F_b(\boldsymbol{\theta}^*) \geq 0$. Thus, the first inequality holds in (A.31).

Summing over $\tilde{l} \in \{0, 1, \ldots, k-1\}$ and by the telescopic sum, it follows that

$$\begin{aligned} F_b(\boldsymbol{\theta}^k) - F_b(\boldsymbol{\theta}^*) &\leq \frac{c_1}{2\alpha_b k M_b} \sum_{\tilde{l}=0}^{k-1} \left\{ \|\boldsymbol{\theta}_b^{\tilde{l}} - \boldsymbol{\theta}_b^{\tilde{l},*}\|_2^2 - \|\boldsymbol{\theta}_b^{\tilde{l}+1} - \boldsymbol{\theta}_b^{\tilde{l}+1,*}\|_2^2 \right\} \\ &\leq \frac{c_1}{2\alpha_b k M_b} \left\{ \|\boldsymbol{\theta}_b^0 - \boldsymbol{\theta}_b^{0,*}\|_2^2 - \|\boldsymbol{\theta}_b^k - \boldsymbol{\theta}_b^{k,*}\|_2^2 \right\} \\ &\leq \frac{c_1}{2\alpha_b k M_b} \|\boldsymbol{\theta}_b^0 - \boldsymbol{\theta}_b^{0,*}\|_2^2. \end{aligned} \tag{A.32}$$

Under the conditions in (28), the function of the b^{th} block is monotonically decreasing across outer iterations and as such this yields the result of this corollary. It shows the sublinear rate of convergence, i.e., $\mathcal{O}(1/(kM_b))$. □

8 Scalability Analysis

We design a new problem structure, Example IV, to show the scalability of the method in various settings such as the number of samples, the extent of block overlaps, and the number of blocks/parameters.

Example IV

- Physical process

$$y(x) = \zeta(x) + \varepsilon,$$

with

$$\begin{aligned} \zeta(x) = &\sum_{b=1}^{\left\lfloor \frac{B+1}{2} \right\rfloor} \mathbb{I}\left(\frac{T(4b-4+\omega)}{2B+1} \leq x < \frac{T(4b-1-\omega)}{2B+1} \right) \exp\left(\frac{x}{10}\right) \sin(x) \\ &+ \sum_{b=1}^{\left\lfloor \frac{B}{2} \right\rfloor} \mathbb{I}\left(\frac{T(4b-6+\omega)}{2B+1} \leq x < \frac{T(4b-3-\omega)}{2B+1} \right) \exp\left(\frac{x}{10}\right) \cos(x). \end{aligned}$$

- Computer model

$$
\begin{aligned}
\eta(x,\boldsymbol{\theta}) = \zeta(x) &+ \sum_{b=1}^{B} \mathbb{I}\left(\frac{T(2b-1)}{2B+1} \le x < \frac{T(2b)}{2B+1}\right) \frac{(\theta_b - b)\sqrt{x^2-x+1}}{5} \\
&+ \sum_{b=1}^{B-1} \mathbb{I}\left(\frac{T(2b+\omega)}{2B+1} \le x < \frac{T(2b+1-\omega)}{2B+1}\right) (\theta_b - b)\big(\theta_{b+1} - (b+1)\big)\sqrt{x^2-x+1} \\
&+ \frac{(\theta_1 - 1)\sqrt{x^2-x+1}}{5} \mathbb{I}\left(\frac{T\omega}{2B+1} \le x < \frac{T}{2B+1}\right) \\
&+ \frac{(\theta_B - B)\sqrt{x^2-x+1}}{5} \mathbb{I}\left(x \ge T\Big(1 - \frac{\omega}{2B+1}\Big)\right),
\end{aligned}
$$

where $x \sim \mathcal{U}(0,T)$ and $\varepsilon \sim \mathcal{N}(0, 0.01)$.

By adjusting the problem parameter B, a problem can be generalized to the problem that has a different number of blocks, each of which is associated with the parameter θ_b and the corresponding input domain $(2b-2)T/(2B+1) \le x < (2b+1)T/(2B+1)$ for $b = 1, 2, \ldots, B$ within the overall domain $[0, T]$. Here, we set $T = 2\pi$ for all scenarios. Next, ω quantifies the block overlapping degree for $-0.5 \le \omega \le 0.5$. We consider three cases; with $\omega = -0.5$, the blocks highly overlap with adjacent blocks, $\omega = 0$ represents a weak overlapping setting, and $\omega = 0.5$ implies no overlap. In each setting, we conduct 100 experiments, each with 8 starting points.

Tables 1-3 below show that M-BC yields consistently good results with a different sample size, degree of block overlap, and number of parameters in terms of calibration accuracy and uncertainty quantification. In particular, it demonstrates that the proposed method is suitable even for the problems with significant data overlap. Without data overlap (i.e., $\omega = 0.5$), calibrating each block of parameters does not affect other blocks. Therefore, M-BC solves each block separately and converges to a stationary point after the first outer iteration.

9 Example of Block Dominance Case

To demonstrate the performance when block domination occurs, we modify Example I and design a new example. Specifically, we set an input domain as $x \in \mathcal{U}(0, 4\pi)$ with $x < 8\pi/13$ for the first block, $4\pi/13 \le x < 48\pi/13$ for the second block, and $x \ge 4\pi/13$ for the third block, each of which is associated with calibration parameters θ_1, θ_2, and θ_3, respectively. In this setting, the second block dominates the first and third blocks, and the problem structure is as follows.

Table 1: M-BC results with different sample sizes and MSE in Example IV (Note: the values inside parentheses are standard deviations)

Ex	No. of samples (n)	Block (b)	True θ_b	Calibrated parameter (θ_b)	MSE (F_b)	Half CI width	Coverage (%)
IV	500	1	1	1.00 (0.01)	0.01 (0.00)	0.015 (0.001)	96.0
		2	2	2.00 (0.00)	0.01 (0.00)	0.008 (0.001)	93.0
		3	3	3.00 (0.00)	0.01 (0.00)	0.003 (0.000)	97.0
	1000	1	1	1.00 (0.01)	0.01 (0.00)	0.011 (0.000)	96.0
		2	2	2.00 (0.00)	0.01 (0.00)	0.006 (0.000)	96.0
		3	3	3.00 (0.00)	0.01 (0.00)	0.002 (0.000)	97.0
	1500	1	1	1.00 (0.00)	0.01 (0.00)	0.009 (0.000)	97.0
		2	2	2.00 (0.00)	0.01 (0.00)	0.005 (0.000)	93.0
		3	3	3.00 (0.00)	0.01 (0.00)	0.002 (0.000)	95.0

Table 2: M-BC results with different block overlapping degrees in Example IV (Note: the values inside parentheses are standard deviations)

Ex	Overlap degree (ω)	Block (b)	True θ_b	Calibrated parameter (θ_b)	MSE (F_b)	Half CI width	Coverage (%)
IV	High ($\omega = -0.5$)	1	1	1.00 (0.01)	0.01 (0.00)	0.011 (0.000)	96.0
		2	2	2.00 (0.00)	0.01 (0.00)	0.006 (0.000)	96.0
		3	3	3.00 (0.00)	0.01 (0.00)	0.002 (0.000)	97.0
	Low ($\omega = 0$)	1	1	1.00 (0.01)	0.01 (0.00)	0.011 (0.000)	96.0
		2	2	2.00 (0.00)	0.01 (0.00)	0.006 (0.000)	96.0
		3	3	3.00 (0.00)	0.01 (0.00)	0.002 (0.000)	97.0
	No ($\omega = 0.5$)	1	1	1.00 (0.01)	0.01 (0.00)	0.012 (0.001)	95.0
		2	2	2.00 (0.00)	0.01 (0.00)	0.006 (0.000)	96.0
		3	3	3.00 (0.00)	0.01 (0.00)	0.003 (0.000)	98.0

- Physical process

$$y(x) = \zeta(x) + \varepsilon,$$

with

$$\zeta(x) = \mathbb{I}\left(x < \frac{8}{13}\pi\right)\exp\left(\frac{x}{10}\right)\sin(x) + \mathbb{I}\left(\frac{4}{13}\pi \le x < \frac{48}{13}\pi\right)\exp\left(\frac{x}{10}\right)\cos(x) + \mathbb{I}\left(x \ge \frac{44}{13}\pi\right)\exp\left(\frac{x}{10}\right)\sin(x).$$

- Computer model

$$\begin{aligned}\eta(x, \boldsymbol{\theta}) = \zeta(x) &- \mathbb{I}\left(x < \frac{4}{13}\pi\right)\frac{(\theta_1 - 1)\sqrt{x^2 - x + 1}}{5} \\ &- \mathbb{I}\left(\frac{4}{13}\pi \le x < \frac{8}{13}\pi\right)(\theta_1 - 1)(\theta_2 - 2)\sqrt{x^2 - x + 1} - \mathbb{I}\left(\frac{8}{13}\pi \le x < \frac{44}{13}\pi\right)\frac{(\theta_2 - 2)\sqrt{x^2 - x + 1}}{5} \\ &- \mathbb{I}\left(\frac{44}{13}\pi \le x < \frac{48}{13}\pi\right)(\theta_2 - 2)(\theta_3 - 3)\sqrt{x^2 - x + 1} - \mathbb{I}\left(x \ge \frac{48}{13}\pi\right)\frac{(\theta_3 - 3)\sqrt{x^2 - x + 1}}{5}.\end{aligned}$$

Table 3: M-BC results with a different number of blocks and parameters in Example IV (Note: the values inside parentheses are standard deviations)

Ex	No. of blocks (B)	Block (b)	True θ_b	Calibrated parameter (θ_b)	MSE (F_b)	Half CI width	Coverage (%)
IV	3	1	1	1.00 (0.01)	0.01 (0.00)	0.011 (0.000)	96.0
		2	2	2.00 (0.00)	0.01 (0.00)	0.006 (0.000)	96.0
		3	3	3.00 (0.00)	0.01 (0.00)	0.002 (0.000)	97.0
	5	1	1	1.00 (0.01)	0.01 (0.00)	0.016 (0.001)	96.0
		2	2	2.00 (0.01)	0.01 (0.00)	0.012 (0.001)	97.0
		3	3	3.00 (0.00)	0.01 (0.00)	0.007 (0.001)	96.0
		4	4	4.00 (0.00)	0.01 (0.00)	0.005 (0.000)	96.0
		5	5	5.00 (0.00)	0.01 (0.00)	0.003 (0.000)	94.0
	7	1	1	1.00 (0.01)	0.01 (0.00)	0.019 (0.001)	94.0
		2	2	2.00 (0.01)	0.01 (0.00)	0.019 (0.002)	95.0
		3	3	3.00 (0.01)	0.01 (0.00)	0.012 (0.001)	95.0
		4	4	4.00 (0.00)	0.01 (0.00)	0.009 (0.001)	93.0
		5	5	5.00 (0.00)	0.01 (0.00)	0.007 (0.000)	91.0
		6	6	6.00 (0.00)	0.01 (0.00)	0.005 (0.000)	93.0
		7	7	7.00 (0.00)	0.01 (0.00)	0.003 (0.000)	90.0
	9	1	1	1.00 (0.01)	0.01 (0.00)	0.021 (0.002)	96.0
		2	2	2.00 (0.01)	0.01 (0.00)	0.025 (0.002)	93.0
		3	3	3.00 (0.01)	0.01 (0.00)	0.017 (0.002)	98.0
		4	4	4.00 (0.01)	0.01 (0.00)	0.013 (0.001)	90.0
		5	5	5.00 (.000)	0.01 (0.00)	0.010 (0.001)	93.0
		6	6	6.00 (0.00)	0.01 (0.00)	0.008 (0.001)	92.0
		7	7	7.00 (0.00)	0.01 (0.00)	0.007 (0.001)	96.0
		8	8	8.00 (0.00)	0.01 (0.00)	0.006 (0.000)	95.0
		9	9	9.00 (0.00)	0.01 (0.00)	0.003 (0.000)	94.0

Table 4: Comparison results when the second block dominates the first and third blocks (Note: the values inside parentheses are standard deviations)

Method	Block (b)	True θ_b	Calibrated parameter (θ_b)	MSE (F_b)	Half CI width	Coverage (%)
M-BC	1	1	1.00 (0.04)	0.01 (0.00)	0.067 (0.007)	90.0
	2	2	2.00 (0.00)	0.01 (0.00)	0.005 (0.001)	95.0
	3	3	3.00 (0.00)	0.01 (0.00)	0.007 (0.000)	95.0
H-BC	1	1	0.99 (0.07)	0.01 (0.00)	0.122 (0.043)	81.1
	2	2	2.00 (0.00)	0.01 (0.00)	0.006 (0.000)	100.0
	3	3	3.00 (0.00)	0.01 (0.00)	0.009 (0.001)	97.3
S-BC	1	1	0.22 (0.12)	0.01 (0.00)	0.034 (0.014)	0.0
	2	2	1.81 (0.00)	101.65 (7.82)	0.000 (0.000)	0.0
	3	3	11.43 (0.12)	0.24 (0.01)	0.036 (0.022)	0.0

Table 4 summarizes the results. Notably, even in the block dominance setting, M-BC performs very well with respect to all performance metrics. The small standard deviations indicate that it also provides similar results in multiple experiments. It also produces narrower CIs with the coverage rate close to the nominal rate 95%. While H-BC estimates all parameters well, the half

CIs are wider than those from M-BC. Even with the wider CI for θ_1, H-BC's coverage rate remains at 81.1%. Lastly, S-BC calibrates θ_2 relatively well, but poorly calibrate θ_1 and θ_3. Overall, M-BC consistently provides superior performance when some blocks dominate others.