Proof sketch

Without loss of generality, assume $\boldsymbol{P}_k = \boldsymbol{I}_k$ for all $k = 1, \ldots, K$.

First show that: Suppose $\text{MCR}(\boldsymbol{M}_k,\ \hat{\boldsymbol{M}}_k) \geq \varepsilon$, then there exist $r_k \neq r'_k \in [R_k]$ such that at least one of the following two events holds: (1) $\boldsymbol{D}_{r_k r_k} \geq \boldsymbol{D}_{r_k r'_k} \geq \frac{\varepsilon}{R_k^2}$, or (2) $\boldsymbol{D}_{r_k r_k} \geq \boldsymbol{D}_{r'_k r_k} \geq \frac{\varepsilon}{R_k^2}$. We provide the proof when the case (2) holds. The proof under case (1) can be obtained similarly (please fill it in; we probably need the assumption that $d_{\min}$ is large enough).

Second show that: Suppose $\boldsymbol{D}_{r_1 r_1} \geq \boldsymbol{D}_{r'_1 r_1} \geq \frac{\varepsilon}{R_1^2}$ holds from some $r_1, r'_1 \in [R_1]$, where $r_1 \neq r'_1$. Then, for any $(r_2, \ldots, r_K) \in [d_2] \times \cdots \times [d_K]$ and any $(a_2, \ldots, a_K) \in [d_2] \times \cdots \times [d_K]$, the following inequality holds:

$$
\begin{aligned}
&\frac{\mathcal{N}(\boldsymbol{D}^{(1)}, \ldots, \boldsymbol{D}^{(K)})_{r_1 r_2 \ldots r_K}}{w_{r_1 r_2 \ldots r_K}} - f(z_{r_1 r_2 \ldots r_K}) \\
\geq &\frac{1}{2 w_{r_1 r_2 \ldots r_K}} \left(\boldsymbol{D}^{(1)}_{r_1 r_1} \boldsymbol{D}^{(2)}_{a_2 r_2} \cdots \boldsymbol{D}^{(K)}_{a_K r_K} (c_{r_1 a_2 \ldots a_K} - z_{r_1 r_2 \ldots r_K})^2 + \boldsymbol{D}^{(1)}_{r'_1 r_1} \boldsymbol{D}^{(2)}_{a_2 r_2} \cdots \boldsymbol{D}^{(K)}_{a_K r_K} (c_{r'_1 a_2 \ldots a_k} - z_{r_1 r_2 \ldots r_K})^2 \right) \\
\geq &\frac{1}{2 w_{r_1 r_2 \ldots r_K}} \min\left\{ \boldsymbol{D}^{(1)}_{r_1 r_1}, \boldsymbol{D}^{(1)}_{r'_1 r_1} \right\} \left((c_{r_1 a_2 \ldots a_k} - z_{r_1 r_2 \ldots r_K})^2 + (c_{r'_1 a_2 \ldots a_K} - z_{r_1 r_2 \ldots r_K})^2 \right) \boldsymbol{D}^{(2)}_{a_2 r_2} \cdots \boldsymbol{D}^{(K)}_{a_K r_K} \\
\geq &\frac{1}{4 w_{r_1 r_2 \ldots r_K}} \frac{\delta_{\min} \varepsilon}{R_1^2} \boldsymbol{D}^{(2)}_{a_2 r_2} \cdots \boldsymbol{D}^{(K)}_{a_K r_K}. \qquad (1)
\end{aligned}
$$

Here $\mathcal{N} = [\![f(c_{r_1 \ldots r_K})]\!] \in \mathbb{R}^{R_1 \times \cdots R_K}$ is the loss function evaluated at each block, $\mathcal{N}(\boldsymbol{D}^{(1)}, \ldots, \boldsymbol{D}^{(K)}) = \mathcal{N} \times_1 \boldsymbol{D}^{(1)^T} \times_2 \cdots \times_K \boldsymbol{D}^{(K)^T}$ is the weighted value of the loss function, $z_{r_1 \ldots, r_K} = \frac{1}{w_{r_1 \cdots r_K}} \mathcal{C}(\boldsymbol{D}^{(1)}, \ldots, \boldsymbol{D}^{(K)})_{r_1 \ldots r_K}$ is the $(r_1 \ldots r_k)$-th weighted entry of the block means.

Third: Taking sum of (1) over $(r_2, \ldots, r_K)$ gives

$$
\begin{aligned}
\sum_{r_2, \ldots, r_K} \left(w_{r_1 r_2 \ldots r_K} f(z_{r_1 r_2 \ldots r_K}) - \mathcal{N}(\boldsymbol{D}^{(1)}, \ldots, \boldsymbol{D}^{(K)})_{r_1 r_2 \ldots r_K} \right) &\leq -\frac{1}{4} \frac{\delta_{\min} \varepsilon}{R_1^2} \sum_{r_2, \ldots, r_K} \boldsymbol{D}^{(2)}_{a_2 r_2} \cdots \boldsymbol{D}^{(K)}_{a_K r_K} \\
&\leq -\frac{\delta_{\min} \tau^{K-1} \varepsilon}{4 R_1^2}. \qquad (2)
\end{aligned}
$$

Note that the inequality (2) holds for a certain $r_1 \in [R_1]$. For any other $a_1 = \{1, \ldots, R_1\}/\{r_1\}$, by Jensen's inequality we have

$$
\sum_{a_2, \ldots, a_K} \left(w_{a_1 a_2 \ldots a_K} f(z_{a_1 a_2 \ldots a_K}) - \mathcal{N}(\boldsymbol{D}^{(1)}, \ldots, \boldsymbol{D}^{(K)})_{a_1 a_2 \ldots a_K} \right) \leq 0, \quad \text{for all } a_1 \in [d_1]/\{r_1\}. \qquad (3)
$$

Combining (2) and (3) yields

$$
\sum_{a_1, \ldots, a_K} \ldots = \sum_{a_1 = r_1, (a_2, \ldots, a_K) \in [d_2] \times \cdots \times [d_K]} \ldots + \sum_{a_1 \in [R_1]/\{r_1\}, (a_2, \ldots, a_K) \in [d_2] \times \cdots \times [d_K]} \ldots \leq -\frac{\delta_{\min} \tau^{K-1} \varepsilon}{4 R_1^2}.
$$