Appendix C
Partial Proofs of *Boosting Accuracy of Differential Privacy Continuous Data Release for Federated Learning*

A. Proof of Theorem 2

Proof: Since AuBCR employs both BITM and identity matrices as strategy matrices, the domains of its private weights cover SimpleMech and the optimal ABCRG.

SimpleMech is equivalent to the case of AuBCR with all $\lambda_t = 0, \rho_t = 1$. Since both ABCRG and AuBCR adopt BITM as the strategy matrix, we can construct an AuBCR model equivalent to ABCRG by assigning the optimal private weights of ABCRG to λ_t and then setting ρ_t according to the residual sensitivity.

Once the above privacy budget allocation is not optimal, there is better privacy budget allocation for AuBCR boosting SimpleMech and ABCRG's release accuracy.

■

B. Proof of Property 5

Proof: According to Def. 3, the first $n^{[1]}$ rows of $\mathbf{B}^{(n)}$ can be directly proved as in Exp. (16).

For the rows $n^{[1]}+1 \sim n$ of $\mathbf{B}^{(n)}$, we consider $\mathbf{B}^{(\lfloor n \rfloor)}$ first. By Def. 4, b_{ij} of $\mathbf{B}^{(\lfloor n \rfloor)}$ satisfies

$$b_{ij} = 1 \Leftrightarrow i^{[1]} + 1 \le j \le i \wedge 1 \le i \le \lfloor n \rfloor . \tag{71}$$

Then, considering $b_{i'j'}$ satisfying $i' = i + n^{[1]}$ and $j' = j + n^{[1]}$, we have

$$\begin{aligned} & b_{i'j'} = 1 \\ \Leftrightarrow\ & {i'}^{[1]} + 1 \le j' \le i' \wedge n^{[1]} + 1 \le i' \le n^{[1]} + \lfloor n \rfloor \\ \Leftrightarrow\ & \left(i + n^{[1]}\right)^{[1]} + 1 \le j' \le i + n^{[1]} \wedge 1 \le i \le \lfloor n \rfloor \\ \Leftrightarrow\ & i^{[1]} + 1 \le j \le i \wedge 1 \le i \le \lfloor n \rfloor \text{ (by Prop.1)} \\ \Leftrightarrow\ & b_{ij} = 1. \end{aligned} \tag{72}$$

Thus, $\mathbf{B}^{(n)}_{n^{[1]}+1:n, n^{[1]}+1:n} = \mathbf{B}^{(\lfloor n \rfloor)}$.

Besides, we have ${i'}^{[1]} = \left(i + n^{[1]}\right)^{[1]} + 1 = i^{[1]} + n^{[1]} + 1 > n^{[1]}$ by Prop. 1, which means $\mathbf{B}^{(n)}_{0:n^{[1]}, n^{[1]}+1:n} = \mathbf{O}$. ■

C. Proof of Corollary 2

Proof: According to Cor. 1, there is

$$\begin{aligned} \mathbf{B}^{(n)} &= \bigoplus_{i=\mathfrak{l}(n)-1}^{0} \mathbf{B}^{\left(\lfloor n^{[i]} \rfloor\right)} \\ &= \mathbf{B}^{\left(\lfloor n^{[\mathfrak{l}(n)-1]} \rfloor\right)} \oplus \bigoplus_{i=\mathfrak{l}(n)-2}^{0} \mathbf{B}^{\left(\lfloor n^{[i]} \rfloor\right)} \\ &= \mathbf{B}^{\left(\lfloor n^{[\mathfrak{l}(n)-1]} \rfloor\right)} \oplus \bigoplus_{i=\mathfrak{l}(n)-2}^{0} \mathbf{B}^{\left(\sum_{i=1}^{\mathfrak{l}(n)-2} \lfloor n^{[i]} \rfloor\right)}. \end{aligned} \tag{73}$$

Since $2^k < n < 2^{k+1}$, there are $\lfloor n^{[\mathfrak{l}(n)-1]} \rfloor = 2^k$ and $n = \lfloor n^{[\mathfrak{l}(n)-1]} \rfloor + \sum_{i=1}^{\mathfrak{l}(n)-2} \lfloor n^{[i]} \rfloor = 2^k + \sum_{i=1}^{\mathfrak{l}(n)-2} \lfloor n^{[i]} \rfloor$ by Prop. 2 and 3, i.e. $\sum_{i=1}^{\mathfrak{l}(n)-2} \lfloor n^{[i]} \rfloor = n - 2^k$. Substituting them into Exp. (73), we get $\mathbf{B}^{(n)} = \mathbf{B}^{(2^k)} \oplus \mathbf{B}^{(n-2^k)}$. ■

D. Proof of Theorem 3

Proof: According to Prop. 5, there are

$$\mathbf{B}^{(t)} = \mathbf{B}^{(t^{[1]})} \oplus \mathbf{B}^{(\lfloor t \rfloor)} \text{ and } \boldsymbol{\rho}_{1:t} = \boldsymbol{\rho}_{1:t^{[1]}} \parallel \boldsymbol{\rho}_{t^{[1]}+1:t}. \tag{74}$$

Substituting them into Exp. (28) and (29), we have the sub-expressions in Exp. (28) satisfying

$$\begin{aligned} & \mathbf{B}^{(t)T} \mathrm{diag}^2 \left(\boldsymbol{\lambda}_{1:t}\right) \mathbf{B}^{(t)} + \mathrm{diag}^2 \left(\boldsymbol{\rho}_{1:t}\right) \\ =& \left(\mathbf{B}^{(t^{[1]})} \oplus \mathbf{B}^{(\lfloor t \rfloor)}\right)^T \mathrm{diag}^2 \left(\boldsymbol{\lambda}_{1:t^{[1]}} \parallel \boldsymbol{\lambda}_{t^{[1]}+1:t}\right) \\ & \left(\mathbf{B}^{(t^{[1]})} \oplus \mathbf{B}^{(\lfloor t \rfloor)}\right) + \mathrm{diag}^2 \left(\boldsymbol{\rho}_{1:t^{[1]}} \parallel \boldsymbol{\rho}_{t^{[1]}+1:t}\right) \\ =& \mathbf{B}^{(t^{[1]})} \mathrm{diag}^2 \left(\boldsymbol{\lambda}_{1:t^{[1]}}\right) \mathbf{B}^{(t^{[1]})} + \mathrm{diag}^2 \left(\boldsymbol{\rho}_{1:t^{[1]}}\right) \oplus \\ & \mathbf{B}^{(\lfloor t \rfloor)} \mathrm{diag}^2 \left(\boldsymbol{\lambda}_{t^{[1]}+1:t}\right) \mathbf{B}^{(\lfloor t \rfloor)} + \mathrm{diag}^2 \left(\boldsymbol{\rho}_{t^{[1]}+1:t}\right). \end{aligned} \tag{75}$$

Substituting Exp. (75) back into Exp. (28), we have

$$\begin{aligned} & \mathfrak{f}_t \left(\boldsymbol{\lambda}_{1:t}, \boldsymbol{\rho}_{1:t}\right) \\ =& \mathbf{1}^{(t^{[1]})T} \begin{pmatrix} \mathbf{B}^{(t^{[1]})} \mathrm{diag}^2 \left(\boldsymbol{\lambda}_{1:t^{[1]}}\right) \mathbf{B}^{(t^{[1]})} \\ + \mathrm{diag}^2 \left(\boldsymbol{\rho}_{1:t^{[1]}}\right) \end{pmatrix}^{-1} \mathbf{1}^{(t^{[1]})} + \\ & \mathbf{1}^{(\lfloor t \rfloor)T} \begin{pmatrix} \mathbf{B}^{(\lfloor t \rfloor)} \mathrm{diag}^2 \left(\boldsymbol{\lambda}_{t^{[1]}+1:t}\right) \mathbf{B}^{(\lfloor t \rfloor)} \\ + \mathrm{diag}^2 \left(\boldsymbol{\rho}_{t^{[1]}+1:t}\right) \end{pmatrix}^{-1} \mathbf{1}^{(\lfloor t \rfloor)} \\ =& \mathfrak{f}_{t^{[1]}} \left(\boldsymbol{\lambda}_{1:t^{[1]}}, \boldsymbol{\rho}_{1:t^{[1]}}\right) + \mathfrak{f}_{\lfloor t \rfloor} \left(\boldsymbol{\lambda}_{t^{[1]}+1:t}, \boldsymbol{\rho}_{t^{[1]}+1:t}\right). \end{aligned} \tag{76}$$

■

E. Proof of Theorem 4

Proof: According to the expressions of $\widetilde{c}$ and $\widetilde{q}$, there are x' and x'' satisfying

$$x' = \widetilde{c}/\lambda \ \text{ and } x'' = \widetilde{q}/\rho . \tag{77}$$

Thus, $D\left(x'\right) = \sigma^2/\lambda^2$ and $D\left(x''\right) = \sigma^2/\rho^2$.

Let $\overline{x} = \gamma x' + (1-\gamma) x''$. According to the principle of optimally consistent estimation, the minimized $D\left(\overline{x}\right)$ satisfies

$$\begin{aligned} D\left(\overline{x}\right) =& \frac{D\left(x'\right) D\left(x''\right)}{D\left(x'\right) + D\left(x''\right)} = \frac{\left(D\left(\xi\right)/\lambda^2\right)\left(D\left(\xi\right)/\rho^2\right)}{D\left(\xi\right)/\lambda^2 + D\left(\xi\right)/\rho^2} \\ =& D\left(\xi\right) \frac{\left(1/\lambda^2\right)\left(1/\rho^2\right)}{1/\lambda^2 + 1/\rho^2} = D\left(\xi\right) \frac{1/\lambda^2\rho^2}{(\lambda^2+\rho^2)/\lambda^2\rho^2} \\ =& D\left(\xi\right)/\left(\lambda^2+\rho^2\right) = D\left(\xi\right)/\Delta^2 . \end{aligned}$$

■

F. Proof of Property 7

Proof: According to Exp. (46), $\mathfrak{h}_t\left(\lambda \boldsymbol{\omega}_{1:t}\right)$ satisfies

$$\begin{aligned} \mathfrak{h}_t\left(\lambda \boldsymbol{\omega}_{1:t}\right) &= \mathfrak{f}_t\left(\sqrt{\lambda \boldsymbol{\omega}_{1:t}}, \mathbf{H}^{(t)} \sqrt{\lambda \boldsymbol{\omega}_{1:t}}\right) \\ &= \mathbf{1}^{(t)T} \begin{pmatrix} \mathbf{B}^{(t)T} \mathrm{diag}^2 \left(\sqrt{\lambda \boldsymbol{\omega}_{1:t}}\right) \mathbf{B}^{(t)} + \\ \mathrm{diag}^2 \left(\mathbf{H}^{(t)} \sqrt{\lambda \boldsymbol{\omega}_{1:t}}\right) \end{pmatrix}^{-1} \mathbf{1}^{(t)} \\ &= \frac{1}{\lambda} \mathbf{1}^{(t)T} \left(\begin{pmatrix} \mathbf{B}^{(t)T} \mathrm{diag}\left(\boldsymbol{\omega}_{1:t}\right) \mathbf{B}^{(t)} + \\ \mathbf{H}^{(t)} \mathrm{diag}\left(\boldsymbol{\omega}_{1:t}\right) \end{pmatrix} \right)^{-1} \mathbf{1}^{(t)} \\ &= \frac{1}{\lambda} \mathfrak{f}_t \left(\sqrt{\boldsymbol{\omega}_{1:t}}, \mathbf{H}^{(t)} \sqrt{\boldsymbol{\omega}_{1:t}}\right) \end{aligned} \tag{78}$$

Thus,

$$\begin{aligned}\mathcal{H}(\lambda\boldsymbol{\omega}) &= \sum\nolimits_i a_i \mathfrak{h}_{t_i}(\lambda\boldsymbol{\omega}_{1:t_i}) \\ &= \sum\nolimits_i a_i \mathfrak{h}_{t_i}(\boldsymbol{\omega}_{1:t_i})/\lambda \;= \mathcal{H}(\boldsymbol{\omega})/\lambda\,.\end{aligned} \tag{79}$$

■

G. Proof of Property 5

Proof: First, we check the constraints of $\mathbb{P}_{k,N}$. Since the constraints of $\mathbb{P}_{k,N}$ consist of nonnegative constraints and linear equation constraints, the feasible domain is convex. Thus, $\mathbb{P}_{k,N}$ can be proved to be convex as long as $\mathcal{H}^{(k)}(\boldsymbol{\omega}_{1:2^k}, N)$ is a convex function.

According to Exp. (46), the gradient of $\mathfrak{h}_t(\boldsymbol{\omega}_{1:t})$ satisfies

$$\begin{aligned}\frac{\partial \mathfrak{h}_t(\boldsymbol{\omega}_{1:t})}{\partial \boldsymbol{\omega}_{1:t}} = &-\left(\mathbf{B}^{(t)}\begin{pmatrix}\mathbf{B}^{(t)T}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{B}^{(t)}+\\ \mathbf{H}^{(t)}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{H}^{(t)T}\end{pmatrix}^{-1}\mathbf{1}^{(t)}\right)^2 \\ &-\left(\mathbf{H}^{(t)T}\begin{pmatrix}\mathbf{B}^{(t)T}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{B}^{(t)}+\\ \mathbf{H}^{(t)}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{H}^{(t)T}\end{pmatrix}^{-1}\mathbf{1}^{(t)}\right)^2\end{aligned} \tag{80}$$

and the Hessian matrix satisfies

$$\frac{\partial^2 \mathfrak{h}_t(\boldsymbol{\omega}_{1:t})}{\partial \boldsymbol{\omega}_{1:t}\partial \boldsymbol{\omega}_{1:t}^T} = 2\mathbf{A}^{(t)}\begin{pmatrix}\mathbf{B}^{(t)T}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{B}^{(t)}+\\ \mathbf{H}^{(t)}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{H}^{(t)T}\end{pmatrix}^{-1}\mathbf{A}^{(t)T}, \tag{81}$$

where

$$\begin{aligned}\mathbf{A}^{(t)} = &\operatorname{diag}\left(\mathbf{B}^{(t)}\begin{pmatrix}\mathbf{B}^{(t)T}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{B}^{(t)}+\\ \mathbf{H}^{(t)}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{H}^{(t)T}+\end{pmatrix}^{-1}\mathbf{1}^{(t)}\right)\mathbf{B}^{(t)}+ \\ &\operatorname{diag}\left(\mathbf{H}^{(t)T}\begin{pmatrix}\mathbf{B}^{(t)T}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{B}^{(t)}+\\ \mathbf{H}^{(t)}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{H}^{(t)T}\end{pmatrix}^{-1}\mathbf{1}^{(t)}\right)\mathbf{H}^{(t)T}.\end{aligned} \tag{82}$$

Since $\begin{pmatrix}\mathbf{B}^{(t)T}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{B}^{(t)}+\\ \mathbf{H}^{(t)}\operatorname{diag}(\boldsymbol{\omega}_{1:t})\,\mathbf{H}^{(t)T}\end{pmatrix}^{-1}$ is existing and positive-definite and $\mathbf{A}^{(t)}$ is also full-rank, we know $\frac{\partial^2 \mathfrak{h}_t(\boldsymbol{\omega}_{1:t})}{\partial \boldsymbol{\omega}_{1:t}\partial \boldsymbol{\omega}_{1:t}^T}$ is positive-definite, which means that $\mathfrak{h}_t(\boldsymbol{\omega}_{1:t})$ is a convex function. Since $\mathcal{H}^{(k)}(\boldsymbol{\omega}_{1:2^k}, N)$ is a linear polynomial composed of a series of $\mathfrak{h}_t(\boldsymbol{\omega}_{1:t})$ and the polynomial coefficients are all positive, $\mathcal{H}^{(k)}(\boldsymbol{\omega}_{1:2^k}, N)$ is also a convex function.

Meanwhile, the positive definite property of the Hesse matrix shows that there is a unique optimal solution of $\mathbb{P}_{k,N}.\boldsymbol{\omega}^*$ with respect to $\boldsymbol{\omega}_{1:2^k}$ for $N > 0$; and there is a unique optimal solution of $\mathbb{P}_{k,N}.\boldsymbol{\omega}^*$ with respect to $\boldsymbol{\omega}_{1:2^k-1}$ for $N = 0$. ■

H. Proof of Theorem 6

Proof: Cor. 2 and Prop. 6 show that in case $2^{k-1} < t < 2^k$, $\mathbf{B}^{(t)} = \mathbf{B}^{(2^{k-1})} \oplus \mathbf{B}^{(t-2^{k-1})}$ and $\mathbf{H}^{(t)} = \mathbf{H}^{(2^{k-1})} \oplus \mathbf{H}^{(t-2^{k-1})}$.

Substituting them into Exp. (46) and taking an analytical procedure like that of Exp. (76), we have

$$\mathfrak{h}_t(\boldsymbol{\omega}_{1:t}) = \mathfrak{h}_{2^{k-1}}(\boldsymbol{\omega}_{1:2^{k-1}}) + \mathfrak{h}_{t-2^{k-1}}(\boldsymbol{\omega}_{2^{k-1}+1:t})\,. \tag{83}$$

Applying the conclusion of Exp. (83) to (45), we get

$$\begin{aligned}&\mathcal{H}^{(k)}(\boldsymbol{\omega}_{1:2^k}, N) = \sum\nolimits_{t=1}^{2^k-1}\mathfrak{h}_t(\boldsymbol{\omega}_{1:t}) + N\mathfrak{h}_{2^k}(\boldsymbol{\omega}_{1:2^k}) \\ &= \sum\nolimits_{t=1}^{2^{k-1}-1}\mathfrak{h}_t(\boldsymbol{\omega}_{1:t}) + \mathfrak{h}_{2^{k-1}}(\boldsymbol{\omega}_{1:2^{k-1}}) + \\ &\quad \sum\nolimits_{t=2^{k-1}+1}^{2^k-1}\left(\mathfrak{h}_{2^{k-1}}(\boldsymbol{\omega}_{1:2^{k-1}}) + \mathfrak{h}_{t-2^{k-1}}(\boldsymbol{\omega}_{2^{k-1}+1:t})\right) \\ &\quad + N\mathfrak{h}_{2^k}(\boldsymbol{\omega}_{1:2^k}) \\ &= \sum\nolimits_{t=1}^{2^{k-1}-1}\mathfrak{h}_t(\boldsymbol{\omega}_{1:t}) + 2^{k-1}\mathfrak{h}_{2^{k-1}}(\boldsymbol{\omega}_{1:2^{k-1}}) + \\ &\quad \sum\nolimits_{t=1}^{2^{k-1}-1}\mathfrak{h}_{t-2^{k-1}}\left(\boldsymbol{\omega}_{2^{k-1}+(1:t)}\right) + N\mathfrak{h}_{2^k}(\boldsymbol{\omega}_{1:2^k}) \\ &= \mathcal{H}^{(k-1)}\left(\boldsymbol{\omega}_{1:2^{k-1}}, 2^{k-1}\right) + \mathcal{H}^{(k-1)}(\boldsymbol{\omega}_{2^{k-1}+1:2^k}, 0) \\ &\quad + N\mathfrak{h}_{2^k}(\boldsymbol{\omega}_{1:2^k})\end{aligned} \tag{84}$$

■

I. Proof of Theorem 7

Proof: First, we check the last two columns of $\mathbf{P}^{(2^k)}$. There are $p_{2^k-1,2^k-1} = p_{2^k-1,2^k} = p_{2^k,2^k} = 1$ from the definition of $\mathbf{B}^{(2^k)}$ and $p_{2^k,2^k-1} = 1$ from the definition of $\mathbf{H}^{(2^k)}$. We find that the last two columns of $\mathbf{P}^{(2^k)}$ are identical, which means that the last constraint of $\mathbf{P}^{(2^k)T}\boldsymbol{\omega}_{1:2^k} = a\mathbf{1}^{(2^k)}$ is redundant and can be removed. After removing the last constraint, we have

$$\begin{aligned}&\left[\mathbf{P}^{(2^k-1)T}, \mathbf{p}_{2^k}^T\right]\boldsymbol{\omega}_{1:2^k} = a\mathbf{1}^{(2^k-1)} \\ \Leftrightarrow\ &\mathbf{P}^{(2^k-1)T}\boldsymbol{\omega}_{1:2^k-1} + \mathbf{p}_{2^k}^T\omega_{2^k} = a\mathbf{1}^{(2^k-1)}\end{aligned} \tag{85}$$

where $\mathbf{p}_{2^k} = \mathbf{P}^{(2^k)}_{2^k,1:2^k-1}$. According to Def. 3, the 2^k-th row of $\mathbf{B}^{(2^k)}$ satisfies

$$b_{2^k,j} = 1, \left(2^k\right)^{[1]} + 1 \le j \le 2^k \Leftrightarrow 1 \le j \le 2^k, \tag{86}$$

which means that

$$\mathbf{p}_{2^k} = \mathbf{1}^{(2^k-1)T}. \tag{87}$$

Substituting Exp. 87 into Exp. 85, we have $\mathbf{P}^{(2^k-1)T}\boldsymbol{\omega}_{1:2^k-1} = (a - \omega_{2^k})\,\mathbf{1}^{(2^k-1)}$. ■

J. Proof of Theorem 8

Proof: Combined with Thm. 7, $\mathbb{P}_{k,0}$ can be expressed as

$$\begin{aligned}\min_{\boldsymbol{\omega}_{1:t}\ge\mathbf{0}}\quad &\sum\nolimits_{t=1}^{2^k-1}\mathfrak{h}_t(\boldsymbol{\omega}_{1:t}) \\ s.t.\quad &\mathbf{P}^{(2^k-1)T}\boldsymbol{\omega}_{1:2^k-1} = (1-\omega_{2^k})\,\mathbf{1}^{(2^k)}.\end{aligned} \tag{88}$$

Let $\boldsymbol{\omega}'_{1:t} = \boldsymbol{\omega}_{1:2^k-1}/(1-\omega_{2^k})$. By Prop. 7, there is

$$\begin{aligned}\min_{\boldsymbol{\omega}'_{1:t}\ge\mathbf{0},\omega_{2^k}\ge 0}\quad &\frac{1}{1-\omega_{2^k}}\sum\nolimits_{t=1}^{2^k-1}\mathfrak{h}_t(\boldsymbol{\omega}'_{1:t}) \\ s.t.\quad &\mathbf{P}^{(2^k-1)T}\boldsymbol{\omega}'_{1:2^k-1} = \mathbf{1}^{(2^k)}\end{aligned} \tag{89}$$

In the new optimization (89), the constraint is independent of ω_{2^k}. However, as the coefficient of the objective function, $\frac{1}{1-\omega_{2^k}}$ should be as small as possible. Thus, $\mathbb{P}_{k,0}.\omega^*_{2^k} \equiv 0$.

Notice that the last constraint in (89) is $\omega_{2^k-1} + \omega_{2^k} = 1$. We can conclude that $\mathbb{P}_{k,0}.\omega^*_{2^k-1} = 1$ and $\mathbb{P}_{1,0}.\mathcal{H}^* = 1$.

Since $\mathbf{B}^{(t)} = \mathbf{B}^{\left(2^{k-1}\right)} \oplus \mathbf{B}^{\left(t-2^{k-1}\right)}$ and $\mathbf{H}^{(t)} = \mathbf{H}^{\left(2^{k-1}\right)} \oplus \mathbf{H}^{\left(t-2^{k-1}\right)}$ by Cor. 2 and Prop. 6 for $2^{k-1} < t < 2^k$, the constraint on $\mathbb{P}_{k,0}$ can be expressed as

$$\begin{aligned} & \mathbf{P}^{\left(2^k\right)T}\boldsymbol{\omega}_{1:2^k} = \mathbf{1}^{\left(2^k\right)} \\ \Rightarrow\; & \mathbf{P}^{\left(2^k-1\right)T}\boldsymbol{\omega}_{1:2^k-1} = \mathbf{1}^{\left(2^k-1\right)} \\ \Rightarrow\; & \begin{cases} \mathbf{P}^{\left(2^{k-1}\right)T}\boldsymbol{\omega}_{1:2^{k-1}} = \mathbf{1}^{\left(2^{k-1}\right)} \\ \mathbf{P}^{\left(2^{k-1}-1\right)T}\boldsymbol{\omega}_{2^{k-1}+1:2^k-1} = \mathbf{1}^{\left(2^{k-1}-1\right)} \end{cases} \\ \Rightarrow\; & \begin{cases} \mathbf{P}^{\left(2^{k-1}\right)T}\boldsymbol{\omega}_{1:2^{k-1}} = \mathbf{1}^{\left(2^{k-1}\right)} \\ \mathbf{P}^{\left(2^{k-1}\right)T}\boldsymbol{\omega}_{2^{k-1}+1:2^k} = \mathbf{1}^{\left(2^k-1\right)} \end{cases}. \end{aligned} \tag{90}$$

Exp. (90) shows that the optimization of $\boldsymbol{\omega}_{1:2^{k-1}}$ and $\boldsymbol{\omega}_{2^{k-1}+1:2^k}$ are independent and corresponding to the subproblems $\mathbb{P}_{k-1,2^{k-1}}$ and $\mathbb{P}_{k-1,0}$, respectively. Thus, $\mathbb{P}_{k,0}.\boldsymbol{\omega}^* = \mathbb{P}_{k-1,2^{k-1}}.\boldsymbol{\omega}^* \parallel \mathbb{P}_{k-1,0}.\boldsymbol{\omega}^*$ and $\mathbb{P}_{k,0}.\mathcal{H}^* = \mathbb{P}_{k-1,2^{k-1}}.\mathcal{H}^* + \mathbb{P}_{k-1,0}.\mathcal{H}^*$ according to Exp. 48. ■

K. Proof of Theorem 9

Proof: Repeatedly applying Thm. 6, we have

$$\begin{aligned} \mathcal{H}^{(k)}\left(\boldsymbol{\omega}_{1:2^k}, N\right) = & \mathcal{G}\left(\left(1-\alpha_k-\beta_{k,k-1}\right)\boldsymbol{\theta}_{k,k-1}\right) + \\ & 2^{k-1}\mathfrak{h}_{2^{k-1}}\left(\left(1-\alpha_k-\beta_{k,k-1}\right)\boldsymbol{\theta}_{k,k-1} \parallel \beta_{k,k-1}\right) \\ & + \mathcal{H}^{(k-1)}\left(\boldsymbol{\omega}_{2^{k-1}+1:2^k}, 0\right) + N\mathfrak{h}_{2^k}\left(\boldsymbol{\omega}_{1:2^k}\right). \end{aligned} \tag{91}$$

Repeatedly applying Exp. (91), we further have

$$\begin{aligned} \mathcal{H}^{(k)}\left(\boldsymbol{\omega}_{1:2^k}, N\right) = & \sum\nolimits_{i=0}^{k-1}\mathcal{G}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i}\right) + \\ & \sum\nolimits_{i=0}^{k-1} 2^i\mathfrak{h}_{2^i}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i} \parallel \beta_{k,i}\right) \\ & + N\mathfrak{h}_{2^k}\left(\boldsymbol{\omega}_{1:2^k}\right). \end{aligned} \tag{92}$$

Notice $D^*\left(\widetilde{s}_{2^k}\right) = \mathfrak{h}_{2^i}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i} \parallel \beta_{k,i}\right)\sigma^2$ can be estimated by $D^*\left(\widetilde{s}_{2^k-1}\right)$, $D\left(\widetilde{c}_{2^k}\right)$ and $D\left(\widetilde{q}_{2^k}\right)$. Following the principle of optimal estimation, we conclude that

$$\begin{aligned} & \mathfrak{h}_{2^i}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i} \parallel \beta_{k,i}\right) \\ & = \Re\left(\mathfrak{g}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right), \end{aligned} \tag{93}$$

Correspondingly,

$$\mathfrak{h}_{2^k}\left(\boldsymbol{\omega}_{1:2^k}\right) = \Re\left(\mathfrak{h}_{2^k-1}\left(\boldsymbol{\omega}_{1:2^k-1}\right), \alpha_k, 0\right). \tag{94}$$

Further, by Cor. 1 and Prop. 6, there are

$$\mathbf{B}^{\left(2^k-1\right)} = \bigoplus_{i=k-1}^{0}\mathbf{B}^{\left(2^i\right)} \text{ and } \mathbf{H}^{\left(2^k-1\right)} = \bigoplus_{i=k-1}^{0}\mathbf{H}^{\left(2^i\right)}. \tag{95}$$

Substituting them into Exp. (46), we have

$$\mathfrak{h}_{2^k-1}\left(\boldsymbol{\omega}_{1:2^k-1}\right) = \sum\nolimits_{i=0}^{k-1}\mathfrak{h}_{2^i}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i} \parallel \beta_{k,i}\right). \tag{96}$$

Combining Exp. (94), (95) and (96), there is

$$\begin{aligned} & \mathfrak{h}_{2^k}\left(\boldsymbol{\omega}_{1:2^k}\right) \\ & = \Re\left(\sum\nolimits_{i=0}^{k-1}\Re\left(\mathfrak{g}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right), \alpha_k, 0\right). \end{aligned} \tag{97}$$

According to Prop. 7, we draw that

$$\begin{cases} \mathfrak{g}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i}\right) = \dfrac{\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)}{1-\alpha_k-\beta_{k,i}} \\ \mathcal{G}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i}\right) = \dfrac{\mathcal{G}\left(\boldsymbol{\theta}_{k,i}\right)}{1-\alpha_k-\beta_{k,i}} \end{cases} \tag{98}$$

Substituting Exp. (93), (97) and (98) into Exp. (92), we can obtain (57).

Next, we consider the constraints. According to the last constraint $\mathbb{P}_{k,0}$, there is

$$\omega_{2^k-1} + \omega_{2^k} = 1 \Leftrightarrow \omega_{2^k-1} + \alpha_k = 1 \Rightarrow\; 0 \le \alpha_k \le 1. \tag{99}$$

Similarly, the ς_i^k-th constraint satisfies

$$\begin{aligned} & \omega_{\varsigma_i^k-1} + \omega_{\varsigma_i^k} + \omega_{2^k} = 1 \Leftrightarrow \omega_{\varsigma_i^k-1} + \beta_{k,i} = 1-\alpha_k \\ & \Rightarrow 0 \le \beta_{k,i} \le 1-\alpha_k \end{aligned} \tag{100}$$

Thus,

$$\begin{gathered} \mathbf{P}^{\left(2^k\right)T}\boldsymbol{\omega}_{1:2^k} = \mathbf{1}^{\left(2^k\right)} \\ \Downarrow \\ \left(\bigoplus_{i=k-1}^{0}\mathbf{P}^{\left(2^i\right)T}\right)\left(\mathop{\Big\|}_{i=k-1}^{1}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i} \parallel \beta_{k,i}\right) \parallel \beta_{k,0}\right) \\ = \left(1-\alpha_k\right)\mathbf{1}^{\left(2^k-1\right)} \\ \Downarrow \\ \big\|_{i=k-1}^{1}\left(\mathbf{P}^{\left(2^i\right)T}\left(\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_{k,i} \parallel \beta_{k,i}\right)\right) \parallel \beta_{k,0} \\ = \left(1-\alpha_k\right)\big\|_{i=k-1}^{0}\mathbf{1}^{\left(2^i\right)} \\ \Downarrow \\ \big\|_{i=k-1}^{1}\left(1-\alpha_k-\beta_{k,i}\right)\mathbf{P}^{\left(2^i-1\right)T}\boldsymbol{\theta}_{k,i} \parallel \beta_{k,0} \\ = \big\|_{i=k-1}^{1}\left(1-\alpha_k-\beta_{k,i}\right)\mathbf{1}^{\left(2^i-1\right)} \parallel \left(1-\alpha_k\right) \\ \Downarrow \\ \begin{cases} \mathbf{P}^{\left(2^i-1\right)T}\boldsymbol{\theta}_{k,i} = \mathbf{1}^{\left(2^i-1\right)}, \forall i \in \{1,2,\ldots,k-1\} \\ \beta_{k,0} = 1-\alpha_k \end{cases} \end{gathered} \tag{101}$$

Integrating the results in Eq. (99), (100) and (101), we have that the constraint of the optimization (56) holds. ■

L. Proof of Theorem 10

Proof: To simplify the subsequent description, we define

$$\Re'\left(s,\beta,\alpha\right) = \Re\left(\frac{s}{1-\alpha-\beta}, \beta, \alpha\right) = \frac{s+1}{s\beta+\left(1-\alpha\right)}. \tag{102}$$

The gradient of $\mathcal{L}^{(k)}\left(\alpha_k, \{\beta_{k,i}\}_{i=0}^{k-1}, \{\boldsymbol{\theta}_{k,i}\}_{i=0}^{k-1}, N\right)$ is

$$\begin{aligned}
&\frac{\partial \mathcal{L}^{(k)}}{\partial \boldsymbol{\theta}_{k,i}} \\
=& \frac{1}{1-\alpha_k-\beta_{k,i}} \times \frac{\partial \mathcal{G}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}} + \\
&\frac{2^i}{1-\alpha_k-\beta_{k,i}} \times \frac{\partial \Re'\left(\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right)}{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)} \times \frac{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}} \\
&+ \frac{N}{1-\alpha_k-\beta_{k,i}} \times \left.\frac{\partial \Re\left(s, \alpha_k, 0\right)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}_{1:2^k-1}\right)} \times \\
&\frac{\partial \Re'\left(\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right)}{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)} \times \frac{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}} \\
=& \frac{1}{1-\alpha_k-\beta_{k,i}} \times \\
&\left(\begin{array}{l}
\frac{\partial \mathcal{G}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}} + \frac{\partial \Re'\left(\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right)}{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)} \times \\
\left(2^i + N \left.\frac{\partial \Re\left(s, \alpha_k, 0\right)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}_{1:2^k-1}\right)}\right) \times \frac{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}}
\end{array}\right) \\
=& \frac{1}{1-\alpha_k-\beta_{k,i}} \left[\begin{array}{ll} \frac{\partial \mathcal{G}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}} & \frac{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}} \end{array}\right] \times \\
&\left[\begin{array}{c} 1 \\ \frac{\partial \Re'\left(\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right)}{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)} \times \left(2^i + N \left.\frac{\partial \Re\left(s, \alpha_k, 0\right)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}_{1:2^k-1}\right)}\right) \end{array}\right],
\end{aligned} \tag{103}$$

and

$$\begin{aligned}
&\frac{\partial \mathcal{L}^{(k)}}{\partial \beta_{k,i}} \\
=& \frac{\mathcal{G}\left(\boldsymbol{\theta}_{k,i}\right)}{\left(1-\alpha_k-\beta_{k,i}\right)^2} + 2^i \frac{\partial \Re'\left(\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right)}{\partial \beta_{k,i}} + \\
&N \left.\frac{\partial \Re\left(s, \alpha_k, 0\right)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}_{1:2^k-1}\right)} \times \frac{\partial \Re'\left(\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right)}{\partial \beta_{k,i}} \\
=& \left(2^i + N \left.\frac{\partial \Re\left(s, \alpha_k, 0\right)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}_{1:2^k-1}\right)}\right) \times \\
&\frac{\partial \Re'\left(\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right)}{\partial \beta_{k,i}} + \frac{\mathcal{G}\left(\boldsymbol{\theta}_{k,i}\right)}{\left(1-\alpha_k-\beta_{k,i}\right)^2} \\
=& -\left(2^i + N \left.\frac{\partial \Re\left(s, \alpha_k, 0\right)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}_{1:2^k-1}\right)}\right) \times \\
&\frac{\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)\left(\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)+1\right)}{\left(\beta_{k,i}\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)+\left(1-\alpha_k\right)\right)^2} + \frac{\mathcal{G}^{(k-i)}\left(\boldsymbol{\theta}_{k,i}\right)}{\left(1-\alpha_k-\beta_{k,i}\right)^2}
\end{aligned} \tag{104}$$

Analyzing the gradients of $\Re(s, \beta, \alpha)$ and $\Re'(s, \beta, \alpha)$, we have

$$\frac{\partial \Re(s, \beta, \alpha)}{\partial s} = \frac{1}{\left(s\beta + (1-\alpha)/(1-\alpha-\beta)\ \right)^2} \geq 0, \tag{105}$$

and

$$0 \leq \frac{\partial \Re'\left(s, \beta_{k,i}, \alpha_k\right)}{\partial s} = \frac{\left(1-\alpha_k-\beta_{k,i}\right)^2}{\left(1-\alpha_k+\beta_{k,i}s\right)^2} \leq \frac{\left(1-\alpha_k\right)^2}{\left(1-\alpha_k\right)^2} = 1. \tag{106}$$

When β is equal to the boundary 0 or $1-\alpha_k$, there are

$$\begin{cases} \frac{\partial \Re'\left(s, 0, \alpha_k\right)}{\partial s} = 1 \\ \frac{\partial \Re'\left(s, 1-\alpha_k, \alpha_k\right)}{\partial s} = 0 \end{cases}. \tag{107}$$

It is known that $\mathbb{P}_{k,0}.\boldsymbol{\omega}^* = \left(\|\|_{i=k-1}^{1}\left(1-\alpha_i\right)\boldsymbol{\theta}_i^* \parallel \alpha_i^*\right) \parallel 1 \parallel 0$, which satisfies the KKT condition that

$$\begin{cases} \exists \boldsymbol{\psi},\ \left.\frac{\partial \mathcal{L}^{(k)}}{\partial \boldsymbol{\theta}_{k,i}}\right|_{\boldsymbol{\omega}_{1:2^k}=\mathbb{P}_{k,0}.\boldsymbol{\omega}^*, N=0} + \mathbf{P}^{\left(2^{k-i}-1\right)}\boldsymbol{\psi} = \mathbf{0} \\ \left.\frac{\partial \mathcal{L}^{(k)}}{\partial \beta_{k,i}}\right|_{\boldsymbol{\omega}_{1:2^k}=\mathbb{P}_{k,0}.\boldsymbol{\omega}^*, N=0} = 0 \end{cases}. \tag{108}$$

For $\mathbb{P}_{k,N}$, we assume $\boldsymbol{\omega}'$ has the form like Exp. (61), i.e.

$$\boldsymbol{\omega}' = \left(\|\|_{i=k-1}^{1}\left(1-\alpha_k-\beta_{k,i}\right)\boldsymbol{\theta}_i^* \parallel \beta_{k,i}\right) \parallel \beta_{k,0} \parallel \alpha_k. \tag{109}$$

Next, we will demonstrate that regardless of the value of $\alpha_k \in [0,1]$, there is always a group of $\beta_{k,i}$ such that $0 \leq \beta_{k,i} \leq 1-\alpha_k, 1 \leq i \leq k-1$ making $\boldsymbol{\omega}'_{1:2^k-1}$ also satisfies the KKT condition. At this point, according to Thm. 5, $\boldsymbol{\omega}'$ has reached the optimal solution for a given α_k. Once $\alpha_k = \mathbb{P}_{k,0}.\boldsymbol{\omega}^*_{2^k}$, $\boldsymbol{\omega}' = \mathbb{P}_{k,0}.\boldsymbol{\omega}^*$.

Note that $\mathbb{P}_{k,0}$ and $\mathbb{P}_{k,N}$ have the same linear constraint. As long as the following two conditions (110) and (111) are satisfied simultaneously, $\boldsymbol{\omega}'_{1:2^k}$ satisfies the KKT condition.

$$\left.\frac{\partial \mathcal{L}^{(k)}}{\partial \boldsymbol{\theta}_{k,i}}\right|_{\boldsymbol{\omega}_{1:2^k}=\mathbb{P}_{k,0}.\boldsymbol{\omega}^*, N=0} = \gamma \left.\frac{\partial \mathcal{L}^{(k)}}{\partial \boldsymbol{\theta}_{k,i}}\right|_{\boldsymbol{\omega}_{1:2^k}=\boldsymbol{\omega}'} \tag{110}$$

$$\left.\frac{\partial \mathcal{L}^{(k)}}{\partial \beta_{k,i}}\right|_{\boldsymbol{\omega}_{1:2^k}=\boldsymbol{\omega}'} = 0 \tag{111}$$

First, we find $\beta_{k,i}$ satisfying the condition (110). As shown in the gradient (103), there are two components of $\frac{\partial \mathcal{G}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}}$ and $\frac{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}}$. The coefficient of $\frac{\partial \mathcal{G}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}}$ is known to be constant 1. When the coefficient of $\frac{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}}$ is also constant, the condition (110) is satisfied. Let

$$\begin{aligned} \mathcal{J}\left(\alpha_k, \beta_{k,i}, \boldsymbol{\theta}_{k,i}, z, N\right) =& \frac{\partial \Re'\left(\mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right), \beta_{k,i}, \alpha_k\right)}{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)} \times \\ &\left(2^i + N \left.\frac{\partial \Re\left(s, \alpha_k, 0\right)}{\partial s}\right|_{s=z}\right) \end{aligned} \tag{112}$$

denote the coefficient of the gradient component $\frac{\partial \mathfrak{g}\left(\boldsymbol{\theta}_{k,i}\right)}{\partial \boldsymbol{\theta}_{k,i}}$. We have

$$\begin{aligned} \mathfrak{F}_i\left(\beta_{k,i}\right) =& \mathcal{J}\left(\alpha_k, \beta_{k,i}, \boldsymbol{\theta}_i^*, \mathfrak{g}\left(\boldsymbol{\omega}'_{1:2^k-1}\right), N\right) - \\ &\mathcal{J}\left(0, \alpha_i^*, \boldsymbol{\theta}_i^*, \mathfrak{g}\left(\mathbb{P}_{k,0}.\boldsymbol{\omega}^*_{1:2^k-1}\right), 0\right), \end{aligned} \tag{113}$$

and $\mathfrak{F}_i(\beta_{k,i}) = 0 \Rightarrow$ the condition (110) is satisfied. If $\beta_{k,i} = 0$, there is

$$\begin{aligned}\mathfrak{F}_i(0) &= \left(2^i + N\left.\frac{\partial\mathfrak{R}(s,\alpha_k,0)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}'_{1:2^k-1}\right)}\right) \\ &\quad - 2^i\left.\frac{\partial\mathfrak{R}'(s,\alpha_i^*,0)}{\partial s}\right|_{s=\mathfrak{g}(\boldsymbol{\theta}_i^*)} \\ &= 2^i\left(1-\left.\frac{\partial\mathfrak{R}'(s,\alpha_i^*,0)}{\partial s}\right|_{s=\mathfrak{g}(\boldsymbol{\theta}_i^*)}\right) \\ &\quad + N\left.\frac{\partial\mathfrak{R}(s,\alpha_k,0)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}'_{1:2^k-1}\right)} \\ &\geq N\left.\frac{\partial\mathfrak{R}(s,\alpha_k,0)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}'_{1:2^k-1}\right)} \geq 0.\end{aligned} \tag{114}$$

For $\beta_{k,i} = 1-\alpha_k$, there is

$$\mathfrak{F}_i\left(1-\alpha_k^{(2)}\right) = -2^i\left.\frac{\partial\mathfrak{R}'(s,\alpha_i^*,0)}{\partial s}\right|_{s=\mathfrak{g}(\boldsymbol{\theta}_i^*)} \leq 0. \tag{115}$$

Since $\mathfrak{F}_i(\beta_{k,i})$ is continuous about $\beta_{k,i}$, according to Existence Theorem of Zero Points, there must exist a $\beta_{k,i} \in [0, 1-\alpha_k]$ such that $\mathfrak{F}_i(\beta_{k,i}) = 0$, which makes KKT condition (110) satisfied.

Next, we verify whether condition (111) is satisfied while $\mathfrak{F}_i(\beta_{k,i}) = 0$. $\mathfrak{F}_i(\beta_{k,i}) = 0$ means that

$$\begin{aligned}&\frac{\partial\mathfrak{R}'(\mathfrak{g}(\boldsymbol{\theta}_i^*),\beta_{k,i},\alpha_k)}{\partial\mathfrak{g}(\boldsymbol{\theta}_i^*)} \times \left(2^i + N\left.\frac{\partial\mathfrak{R}(s,\alpha_k,0)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}'_{1:2^k-1}\right)}\right) \\ &\qquad -2^i\frac{\partial\mathfrak{R}'(\mathfrak{g}(\boldsymbol{\theta}_i^*),\alpha_i^*,0)}{\partial\mathfrak{g}(\boldsymbol{\theta}_i^*)} = 0.\end{aligned} \tag{116}$$

Meanwhile, by (108), we have

$$\begin{aligned}&\left.\frac{\partial\mathcal{L}^{(k)}}{\partial\beta_{k,i}}\right|_{\boldsymbol{\omega}_{1:2^k}=\mathbb{P}_{k,0}.\boldsymbol{\omega}^*,N=0} \\ &= \frac{\mathcal{G}(\boldsymbol{\theta}_i^*)}{(1-\alpha_i^*)^2} - 2^i\frac{\mathfrak{g}(\boldsymbol{\theta}_i^*)(\mathfrak{g}(\boldsymbol{\theta}_i^*)+1)}{(\alpha_i^*\mathfrak{g}(\boldsymbol{\theta}_i^*)+1)^2} = 0.\end{aligned} \tag{117}$$

Thus,

$$\begin{aligned}&\left.\frac{\partial\mathcal{L}^{(k)}}{\partial\beta_{k,i}}\right|_{\boldsymbol{\omega}_{1:2^k}=\boldsymbol{\omega}'} \\ &= \frac{\mathcal{G}(\boldsymbol{\theta}_i^*)}{(1-\alpha_k-\beta_{k,i})^2} - \frac{\mathfrak{g}(\boldsymbol{\theta}_i^*)(\mathfrak{g}(\boldsymbol{\theta}_i^*)+1)}{(\beta_{k,i}\mathfrak{g}(\boldsymbol{\theta}_i^*)+(1-\alpha_k))^2} \\ &\quad\times\left(2^i + N\left.\frac{\partial\mathfrak{R}(s,\alpha_k,0)}{\partial s}\right|_{s=\mathfrak{g}\left(\boldsymbol{\omega}'_{1:2^k-1}\right)}\right) \\ &= \frac{\mathcal{G}(\boldsymbol{\theta}_i^*)}{(1-\alpha_k-\beta_{k,i})^2} - 2^i\frac{\frac{\partial\mathfrak{R}'(\mathfrak{g}(\boldsymbol{\theta}_i^*),\alpha_i^*,0)}{\partial\mathfrak{g}(\boldsymbol{\theta}_i^*)}}{\frac{\partial\mathfrak{R}'(\mathfrak{g}(\boldsymbol{\theta}_i^*),\beta_{k,i},\alpha_k)}{\partial\mathfrak{g}(\boldsymbol{\theta}_i^*)}} \\ &\quad\times\frac{\mathfrak{g}(\boldsymbol{\theta}_i^*)(\mathfrak{g}(\boldsymbol{\theta}_i^*)+1)}{(\beta_{k,i}\mathfrak{g}(\boldsymbol{\theta}_i^*)+(1-\alpha_k))^2} \\ &= \frac{\mathcal{G}(\boldsymbol{\theta}_i^*)}{(1-\alpha_k-\beta_{k,i})^2} - \frac{\mathfrak{g}(\boldsymbol{\theta}_i^*)(\mathfrak{g}(\boldsymbol{\theta}_i^*)+1)}{(\beta_{k,i}\mathfrak{g}(\boldsymbol{\theta}_i^*)+(1-\alpha_k))^2} \\ &\quad\times 2^i\frac{(1-\alpha_i^*)^2(1-\alpha_k+\beta_{k,i}\mathfrak{g}(\boldsymbol{\theta}_i^*))^2}{(1+\alpha_i^*\mathfrak{g}(\boldsymbol{\theta}_i^*))^2(1-\alpha_k-\beta_{k,i})^2} \\ &= \frac{(1-\alpha_i^*)^2}{(1-\alpha_k-\beta_{k,i})^2}\times\left(\frac{\mathcal{G}(\boldsymbol{\theta}_i^*)}{(1-\alpha_i^*)^2} - 2^i\frac{\mathfrak{g}(\boldsymbol{\theta}_i^*)(\mathfrak{g}(\boldsymbol{\theta}_i^*)+1)}{(1+\alpha_i^*\mathfrak{g}(\boldsymbol{\theta}_i^*))^2}\right) \\ &= \frac{(1-\alpha_i^*)^2}{(1-\alpha_k-\beta_{k,i})^2}\left.\frac{\partial\mathcal{L}^{(k)}}{\partial\beta_{k,i}}\right|_{\boldsymbol{\omega}_{1:2^k}=\mathbb{P}_{k,0}.\boldsymbol{\omega}^*_{1:2^k},N=0} = 0.\end{aligned} \tag{118, 119}$$

Therefore, conditions (110) and (111) are satisfied simultaneously, meaning that $\boldsymbol{\omega}'_{1:2^k}$ satisfies the KKT condition. Since $\alpha_k \in [0,1]$, there must exist an α_k such that $\boldsymbol{\omega}' = \mathbb{P}_{k,0}.\boldsymbol{\omega}^*$. ■

M. Proof of Corollary 3

Proof: First, we consider the constraints of $\mathbb{Q}_{k,N}$. Since the constraints contain only equation and bounds constraints, the feasible domains of α_k and $\{\beta_{k,i}\}_{i=0}^{k-1}$ are convex. Take any two groups of inputs in the domain of definition, namely α'_k, $\left\{\beta'_{k,i}\right\}_{i=0}^{k-1}$ and α''_k, $\left\{\beta''_{k,i}\right\}_{i=0}^{k-1}$ respectively. Then, we represent Exp. (55) as the mapping $\mathfrak{T}: \alpha_k, \{\beta_{k,i}\}_{i=0}^{k-1} \mapsto \boldsymbol{\omega}_{1:2^k}$, and obtain

$$\begin{cases}\boldsymbol{\omega}'_{1:2^k} = \mathfrak{T}\left(\alpha'_k, \left\{\beta'_{k,i}\right\}_{i=0}^{k-1}\right) \\ \boldsymbol{\omega}''_{1:2^k} = \mathfrak{T}\left(\alpha''_k, \left\{\beta''_{k,i}\right\}_{i=0}^{k-1}\right)\end{cases} \tag{120}$$

Since $\mathfrak{T}$ is a linear mapping, the midpoint $\overline{\alpha}_k$, $\left\{\overline{\beta}_{k,i}\right\}_{i=0}^{k-1}$ of the inputs satisfies $\overline{\alpha}_k = \frac{\alpha'_k+\alpha''_k}{2}$, $\overline{\beta}_{k,i} = \frac{\beta'_{k,i}+\beta''_{k,i}}{2}$. Thus,

$$\begin{aligned}\overline{\boldsymbol{\omega}}_{1:2^k} &= \mathfrak{T}\left(\overline{\alpha}_k, \left\{\overline{\beta}_{k,i}\right\}_{i=0}^{k-1}\right) \\ &= \frac{\mathfrak{T}\left(\alpha'_k, \left\{\beta'_{k,i}\right\}_{i=0}^{k-1}\right) + \mathfrak{T}\left(\alpha''_k, \left\{\beta''_{k,i}\right\}_{i=0}^{k-1}\right)}{2} \\ &= \frac{\boldsymbol{\omega}'_{1:2^k} + \boldsymbol{\omega}''_{1:2^k}}{2}\end{aligned} \tag{121}$$

Since Thm. 5 has demonstrated that $\mathbb{P}_{k,N}$ is convex and Thm. 9 has analyzed the equivalence of $\mathbb{Q}_{k,N}$ and $\mathbb{P}_{k,N}$. Therefore, $\mathbb{Q}_{k,N}$ is also a convex optimization problem and there exists a unique optimal solution. ■