Online Appendix (Electronic Companion)

A. Proof of Proposition 3

The equivalence between Definition 3 and Condition (9) is established in both directions.

- For necessity, we need to show that Condition (9) implies Definition 3. Without loss of generality, we consider two paths $k, k' \in \hat{K}_w$ such that $f_k^w > 0$. It follows that $[\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})] \neq \emptyset$. Then, we need to prove that $C_k^w(\mathbf{f};\gamma) \leq C_{k'}^w(\mathbf{f};\gamma)$ for any $\gamma \in [\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})]$ and $k' \neq k$. According to Condition (9), it leads to

$$H_k^w(\mathbf{f}) = \mu_k^w(\mathbf{f}) + \sum_{i=k}^{|\hat{K}_w|} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] - h_w = 0, \tag{A.1}$$

$$H_{k'}^w(\mathbf{f}) = \mu_{k'}^w(\mathbf{f}) + \sum_{i=k'}^{|\hat{K}_w|} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] - h_w \geq 0. \tag{A.2}$$

Combing Eqs. (A.1) and (A.2) together yields

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) - \sum_{i=k'}^{|\hat{K}_w|-1} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] + \sum_{i=k}^{|\hat{K}_w|-1} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{A.3}$$

There are two cases in terms of the order between two paths: (i) $k' > k$ and (ii) $k' < k$. For case (i) such that $\sigma_k^w(\mathbf{f}) \geq \sigma_{k'}^w(\mathbf{f})$, Eq (A.3) can be simplified as

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) + \sum_{k \leq i < k'} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{A.4}$$

For any $\gamma \in [\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})]$, we have $\gamma \leq \hat{\gamma}_i^w(\mathbf{f})$ for each $i \geq k$. Given that $\sigma_i^w(\mathbf{f}) \geq \sigma_{i+1}^w(\mathbf{f})$ by path order, we have the following relationship as per Eq. (A.4):

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) + \sum_{k \leq i < k'} \gamma \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{A.5}$$

Simplifying Eq. (A.5) yields

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) + \gamma \cdot \sigma_k^w(\mathbf{f}) - \gamma \cdot \sigma_{k'}^w(\mathbf{f}) \leq 0 \Rightarrow C_k^w(\mathbf{f};\gamma) \leq C_{k'}^w(\mathbf{f};\gamma). \tag{A.6}$$

For case (ii) such that $\sigma_k^w(\mathbf{f}) \leq \sigma_{k'}^w(\mathbf{f})$, Eq (A.3) can be simplified as

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) - \sum_{k' \leq i < k} \hat{\gamma}_i^w(\mathbf{f}) \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{A.7}$$

For any $\gamma \in [\hat{\gamma}_{k-1}^w(\mathbf{f}), \hat{\gamma}_k^w(\mathbf{f})]$, we have $\gamma \geq \hat{\gamma}_i^w(\mathbf{f})$ for each $i \leq k$. Accordingly, Eq. (A.7) yields the following relationship:

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) - \sum_{k' \leq i < k} \gamma \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_{i+1}^w(\mathbf{f})\right] \leq 0. \tag{A.8}$$

Simplifying Eq. (A.8) also yields

$$\mu_k^w(\mathbf{f}) - \mu_{k'}^w(\mathbf{f}) + \gamma \cdot \sigma_k^w(\mathbf{f}) - \gamma \cdot \sigma_{k'}^w(\mathbf{f}) \leq 0 \Rightarrow C_k^w(\mathbf{f};\gamma) \leq C_{k'}^w(\mathbf{f};\gamma). \tag{A.9}$$

- For sufficiency, we need to show that Definition 3 implies Condition (9). Given an equilibrium path flow $\mathbf{f}$, we need to prove that all used paths have an equal cumulative cost, while paths with higher cumulative costs remain unused. Given an OD pair w, we split the set $\hat{K}_w$ into two sets: a used path set $\hat{K}_w^1 \subseteq \hat{K}_w$ and an unused path set $\hat{K}_w^2 \subseteq \hat{K}_w$. Therefore, we have $f_k^w > 0$ for all $k \in \hat{K}_w^1$, $f_k^w = 0$ for all $k \in \hat{K}_w^2$, and $\hat{K}_w = \hat{K}_w^1 \cup \hat{K}_w^2$. Note that all paths in $\hat{K}_w^1$ must lie on the Pareto frontier. Let $\Delta_k^w \geq 0$ denote the vertical quantity relative to the Pareto frontier for all paths in $\hat{K}_w^2$. There are two cases regarding the number of used paths, i.e., Case (i) for $|\hat{K}_w^1| = 1$ and Case (ii) for $|\hat{K}_w^1| > 1$.

(i) For the case of $|\hat{K}_w^1| = 1$, i.e., only one path is used. Let i denote the index of that path. Therefore, $\hat{\gamma}_k^w = \gamma_{lb}^w$ for $k < i$, and $\hat{\gamma}_k^w = \gamma_{ub}^w$ for $k > i$. It follows that

$$H_i^w(\mathbf{f}) = \mu_i^w(\mathbf{f}) + \sum_{k=i}^{|\hat{K}_w|} \gamma_{ub}^w \cdot \left[\sigma_k^w(\mathbf{f}) - \sigma_{k+1}^w(\mathbf{f})\right] = \mu_i^w(\mathbf{f}) + \gamma_{ub}^w \cdot \sigma_i^w(\mathbf{f}). \tag{A.10}$$

Consider an unused path $k \in \hat{K}_w^2$ such that $k > i$, we have

$$H_k^w(\mathbf{f}) = \mu_k^w(\mathbf{f}) + \gamma_{ub}^w \cdot \sigma_k^w(\mathbf{f}) = \mu_i^w(\mathbf{f}) + \gamma_{ub}^w \cdot [\sigma_i^w(\mathbf{f}) - \sigma_k^w(\mathbf{f})] + \gamma_{ub}^w \cdot \sigma_k^w(\mathbf{f}) + \Delta_k^w \tag{A.11a}$$

$$= \mu_i^w(\mathbf{f}) + \gamma_{ub}^w \cdot \sigma_i^w(\mathbf{f}) + \Delta_k^w = H_i^w(\mathbf{f}) + \Delta_k^w \geq H_i^w(\mathbf{f}). \tag{A.11b}$$

Consider an unused path $k \in \hat{K}_w^2$ such that $k < i$, we have

$$H_k^w(\mathbf{f}) = \mu_k^w(\mathbf{f}) + \sum_{j=k}^{i-1} \gamma_{lb}^w \cdot \left[\sigma_j^w(\mathbf{f}) - \sigma_{j+1}^w(\mathbf{f})\right] + \sum_{j=i}^{|\hat{K}_w|} \gamma_{ub}^w \cdot \left[\sigma_j^w(\mathbf{f}) - \sigma_{j+1}^w(\mathbf{f})\right] \tag{A.12a}$$

$$= \mu_k^w(\mathbf{f}) + \gamma_{lb}^w \cdot [\sigma_k^w(\mathbf{f}) - \sigma_i^w(\mathbf{f})] + \gamma_{ub}^w \cdot \sigma_i^w(\mathbf{f}) \tag{A.12b}$$

$$= \mu_i^w(\mathbf{f}) + \gamma_{ub}^w \cdot \sigma_i^w(\mathbf{f}) + \Delta_k^w = H_i^w(\mathbf{f}) + \Delta_k^w \geq H_i^w(\mathbf{f}). \tag{A.12c}$$

Consequently, the above results show that $H_k^w(\mathbf{f}) \geq H_i^w(\mathbf{f})$ for all $k \in \hat{K}_w^2$. Paths in $\hat{K}_w^2$ with higher cumulative costs, i.e., $\Delta_k^w > 0$, strictly lie above the Pareto frontier and are not chosen by travelers.

(ii) For the case of $|\hat{K}_w^1| > 1$, without loss of generality, let i and j are two adjacent used paths in $\hat{K}_w^1$. Note that $\hat{\gamma}_k^w = \hat{\gamma}_i^w$ for all paths $k \in \{i+1, \cdots, j-1\} \subseteq \hat{K}_w^2$. We have the following relationship:

$$H_j^w(\mathbf{f}) = \mu_j^w(\mathbf{f}) + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] \tag{A.13a}$$

$$= \mu_i^w(\mathbf{f}) + \hat{\gamma}_i^w \cdot \left[\sigma_i^w(\mathbf{f}) - \sigma_j^w(\mathbf{f})\right] + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] \tag{A.13b}$$

$$= \mu_i^w(\mathbf{f}) + \sum_{k=i}^{j-1} \hat{\gamma}_i^w \cdot \left[\sigma_k^w(\mathbf{f}) - \sigma_{k+1}^w(\mathbf{f})\right] + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] \tag{A.13c}$$

$$= \mu_i^w(\mathbf{f}) + \sum_{l=i}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] = H_i^w(\mathbf{f}). \tag{A.13d}$$

The above derivations can be applied to any paired adjacent paths in $\hat{K}_w^1$, indicating that all used paths have identical cumulative costs. We now proceed to consider an unused path $k \in \{i+1, \cdots, j-1\} \subseteq \hat{K}_w^2$, it leads to the following relationship:

$$H_k^w(\mathbf{f}) = \mu_k^w(\mathbf{f}) + \hat{\gamma}_i^w \cdot \left[\sigma_k^w(\mathbf{f}) - \sigma_j^w(\mathbf{f})\right] + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] \tag{A.14a}$$

$$= \mu_j^w(\mathbf{f}) + \sum_{l=j}^{|\hat{K}_w|} \hat{\gamma}_l^w(\mathbf{f}) \cdot \left[\sigma_l^w(\mathbf{f}) - \sigma_{l+1}^w(\mathbf{f})\right] + \Delta_k^w \tag{A.14b}$$

$$= H_j^w(\mathbf{f}) + \Delta_k^w \Rightarrow H_k^w(\mathbf{f}) \geq H_j^w(\mathbf{f}). \tag{A.14c}$$

The above derivation can be extended to every unused path between any paired adjacent used paths. For unused paths ordered before the first or after the last path in $\hat{K}_w^1$, the relationship holds by invoking the derivations in Case (1). Thus, the results indicate that the cumulative cost of any unused path is no less than that of used paths.

In summary, all used efficient paths must have identical cumulative costs, while paths with higher cumulative costs than those of used paths must carry no flows. This completes the proof. □

B. Formulation equivalence

B.1. Proof of Proposition 4

We need to show that the optimality of the convex Problem (11) satisfies Condition (9). Let us construct the Lagrangian function of Problem (11) as follows:

$$L(\mathbf{f},\mathbf{h}) = \sum_{a\in A} \int_0^{x_a(\mathbf{f})} e_a(z)dz + \sum_{w\in W} \sum_{k\in\hat{K}_w} \sigma_k^w \int_{Q_{k-1}^w(\mathbf{f})}^{Q_k^w(\mathbf{f})} \Gamma_w(z)dz + \sum_{w\in W} h_w \left(q_w - \sum_{k\in K_w} f_k^w\right), \tag{B.1}$$

where h_w denotes the multiplier associated with the flow conservation constraint.

To minimize L with respect to $\mathbf{f}$, he following KKT conditions have to hold:

$$f_k^w \frac{\partial L(\mathbf{f},\mathbf{u})}{\partial f_k^w} = 0, \quad \forall k \in \hat{K}_w, w \in W, \tag{B.2a}$$

$$\frac{\partial L(\mathbf{f},\mathbf{u})}{\partial f_k^w} \geq 0, \quad \forall k \in \hat{K}_w, w \in W, \tag{B.2b}$$

The first-order derivative of L with respect to f_k^w is derived as

$$\frac{\partial L(\mathbf{f},\mathbf{u})}{\partial f_k^w} = \sum_{a\in A} \frac{\partial \int_0^{x_a} e_a(z)dz}{\partial x_a} \cdot \frac{\partial x_a}{\partial f_k^w} + \sigma_k^w \cdot \Gamma_w\left(Q_k^w\right) + \sum_{i=k+1}^{|\hat{K}_w|} \sigma_i^w \cdot \left[\Gamma_w\left(Q_i^w\right) - \Gamma_w\left(Q_{i-1}^w\right)\right] \tag{B.3a}$$

$$= \sum_{a\in A} \delta_{ak} e_a(x_a) + \sum_{i=k}^{|\hat{K}_w|-1} \hat{\gamma}_i^w \cdot \left(\sigma_i^w - \sigma_{i+1}^w\right) + \sigma_{|\hat{K}_w|}^w \cdot \hat{\gamma}_{|\hat{K}_w|}^w \tag{B.3b}$$

$$= \mu_k^w + \sum_{i=k}^{|\hat{K}_w|} \hat{\gamma}_i^w \cdot \left(\sigma_i^w - \sigma_{i+1}^w\right). \tag{B.3c}$$

The KKT condition of Problem (11) is equivalent to Condition (9), which completes the proof. □

B.2. Proof of Proposition 5

The proof is completed by showing the equivalence between Condition (9) and VI (12) in both directions.

- For necessity, we need to show that Condition (9) implies VI (12). At equilibrium, all flows are assigned to paths with minimum and equal cumulative costs. Fixing cumulative path cost vector at $\mathbf{H}(\mathbf{f}^*)$, the total cumulative path costs can not be reduced by any changes in path flows. Equivalently, for any $\mathbf{f} \in \Omega$,it yields the following relationship:

$$\mathbf{H}(\mathbf{f}^*) \cdot \mathbf{f} \geq \mathbf{H}(\mathbf{f}^*) \cdot \mathbf{f}^* \Rightarrow \mathbf{H}(\mathbf{f}^*) \cdot (\mathbf{f} - \mathbf{f}^*) \geq 0. \tag{B.4}$$

- For sufficiency, we need to show that VI (12) implies Condition (9). Suppose that $\mathbf{f}^*$ is a solution to VI (12) but it is not an equilibrium flow. Then, there exists at least two paths $k, k' \in K_w$ such that $f_k^{w*} > 0$ and $H_k^w(\mathbf{f}^*) > H_{k'}^w(\mathbf{f}^*)$. Consider a flow δ shifted from path k to k', it leads to the reduction on total cumulative costs by $\delta \cdot [H_k^w(\mathbf{f}^*) - H_{k'}^w(\mathbf{f}^*)] > 0$. Let $\mathbf{f}$ denote the resulting path flow vector by flow adjustments, we have

$$\mathbf{H}(\mathbf{f}^*) \cdot \mathbf{f} < \mathbf{H}(\mathbf{f}^*) \cdot \mathbf{f}^* \Rightarrow \mathbf{H}(\mathbf{f}^*) \cdot (\mathbf{f} - \mathbf{f}^*) < 0, \tag{B.5}$$

which is a contradiction to the VI.

This completes the proof. □

C. Column generation algorithm

C.1. Labeling mean-variance bi-objective algorithm

The one-to-all labeling mean-variance bi-objective algorithm is presented below.

Input: Fixed $\{e_a(\mathbf{f}^n)\}$ and $\{s_a(\mathbf{f}^n)\}$ for all $a \in A$, origin node r

Output: A set of efficient labels L_i for each node $i \in N$.

1. Initialization. Let L_i be the label set for node i, where a label is defined to include four attributes (i.e., source node, mean costs from r, variance costs from r, predecessor label). Initialize a label $l = [r, 0, 0, \text{NULL}]$ and a scan list $SEL = \{l\}$. Set $L_r = \{l\}$ and $L_i = \emptyset, \forall i \in N \setminus \{r\}$. We use $l_{[n]}$ to denote the nth attribute of label l.
2. Select the first label l from SEL in the lexicographical order, and delete it from SEL.
3. For any outgoing link $a = (a^-, a^+)$ from the source node of l,
   - 3-1 Initialize a new label $\bar{l} = \{a^+, l_{[2]} + e_a, l_{[3]} + s_a^2, l\}$.
   - 3-2 If new label $\bar{l}$ is not dominated by any label in L_{a^+}, then update $L_{a^+} = L_{a^+} \cup \{\bar{l}\}$, and $SEL = SEL \cup \{\bar{l}\}$.
   - 3-3 If there exists label $l' \in L_{a^+}$ is dominated by new label $\bar{l}$, then update $L_{a^+} = L_{a^+} \setminus \{l'\}$, and $SEL = SEL \setminus \{l'\}$.

4. If $SEL = \emptyset$, stop; otherwise go to Step 2.

The algorithm initializes by assigning a label to the origin node r. Each iteration involves removing the foremost label from the sorted temporary label list SEL, based on mean as the primary criterion and variance as the secondary criterion. Subsequently, a new label $\bar{l}$ is generated for every outgoing link from the source node. If the label is dominated by existing ones, it is discarded; otherwise, it is retained while eliminating all dominated labels. Termination occurs when SEL becomes empty. To get a set of non-dominated paths between OD pair $w = (r, d)$, one simply traces each efficient label $l \in L_d$ backward to its predecessor label until reaching node r.

As per Sedeño-Noda and Colebrook (2019), the algorithm's complexity is roughly $O(R \times \log |N| + |A| \times R_{\max})$, where $R = \sum_{i=1}^{|N|} R_i$ denotes the total number of non-dominated labels, R_i represents the number of non-dominated labels at node i, and $R_{\max} = \max_{i=1,\ldots,|N|} R_i$.

C.2. Graphical illustration for identifying supported efficient paths

Algorithm 2 mainly consists of two sequential steps: (1) starting and ending path setup, and (2) Sequential slop verification. Figure C.1 provides graphical justifications of the procedure with nine efficient paths. In Figure C.1(a) for Step (1), P2 and P3 yield the shortest cost for γ_{lb}^w, while P7 and P8 do so for γ_{ub}^w. Furthermore, P2 and P8 are excluded because no potential flows will be assigned to them. Additionally, since the frontier should be constructed between P3 and P7, P1 and P9 are also excluded.

In Figure C.1(b) for Step (2), we now consider three candidate paths with orders between P3 and P7, i.e., P4, P5, and P6. The iterative process begins with P3, it includes P4 and P6 because the sequential slope is increasing, excluding P5 as the slope between P4 and P5 is greater than γ_{ub}^w. For the slope between P6 and P7, the non-increasing slope activates a comparison with that of previously included paths, leading to the exclusion of P6 and P4. The procedure concludes with the selection of P3 for RAF in $[\gamma_{lb}^w, \text{slope}_{37}]$, and P7 for RAF in $[\text{slope}_{37}, \gamma_{ub}^w]$.

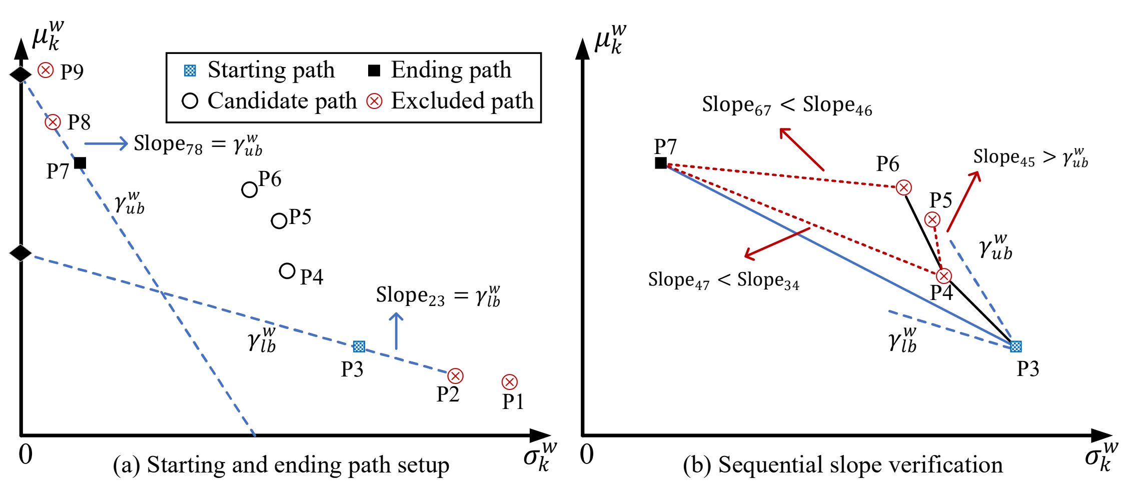

Figure C.1 Justifications of Pareto frontier construction.

D. Convergence Proof

D.1. Proof of Theorem 3

We first verify Conditions (a)-(c) in Zangwill's theorem concerning the map E_w for a single OD pair w. Consequently, we omit the index w for notional convenience.

- For Condition (a), the feasible domain of f_k is $\Omega_{\bar{k}k} = [0, \phi_{\bar{k}k}]$, which is a compact set. The update rule by Eq. (17) ensures that the new point $f_k^{(r+1)}$ is always contained within the compact set $\Omega_{\bar{k}k}$. Hence, Condition (a) is satisfied because all points $f_k^{(r)}$ are in a compact set.
- For condition (b), we consider two feasible points $f_{k[1]}$ and $f_{k[2]}$ such that $f_{k[2]} \in E_w(f_{k[1]})$. Suppose that $f_{k[1]}$ is not a solution to the paired-path problem, then the derivative $\tilde{H}_k(f_{k[1]}) \neq 0$ and $f_{k[1]} > 0$. The BB stepsize scheme with the Amoij rule ensures the strict reduction on the objective function value, i.e., $Z(f_{k[2]}) < Z(f_{k[1]})$. If $f_{k[1]}$ is a solution, then Algorithm 3 must induce a zero-flow shift because either the two paths achieve zero cost difference ($\tilde{H}_k(f_{k[1]}) = 0$), or path k is unused with a higher cost ($\tilde{H}_k(f_{k[1]}) > 0$ and $f_{k[1]} = 0$). The algorithm is terminated (as per Step 3 in Algorithm 3) and yields $Z(f_{k[2]}) = Z(f_{k[1]})$. It follows that $Z(f_{k[2]}) \leq Z(f_{k[1]})$, with equality holds only when $f_{k[1]}$ is a solution. This verifies Condition (b) to show the monotonically decreasing of the sequence until the algorithm is terminated.
- For Condition (c) to show E_w is closed, we need to prove that at any non-optimal point $f_{k[1]}^* \in \Omega_{\bar{k}k}$ if for any two sequences $\{f_{k[1]}^{(r)}\}$ and $\{f_{k[2]}^{(r)}\}$ satisfying $\lim_{r\to\infty}\{f_{k[1]}^{(r)}\} = f_{k[1]}^*$, $f_{k[2]}^{(r)} \in E\left(f_{k[1]}^{(r)}\right)$ for any r and $\lim_{r\to\infty}\{f_{k[2]}^{(r)}\} = f_{k[2]}^*$, then it follows that $f_{k[2]}^* \in E\left(f_{k[1]}^*\right)$ (Bazaraa et al. 2013). According to the update rule by Eq. (17), E_w is a continuous function of f_k because (i) $\tilde{H}_{\bar{k}k}(f_k)$ is continuous, (ii) the stepsize $\alpha^{(r)}$ is well-defined for each iteration r, and (iii) the min and max operators still preserve continuity regarding a continuous function and the constant (i.e., 0 and $\phi_{k\bar{k}}$). Given that $f_{k[2]}^{(r)} \in E_w(f_{k[1]}^{(r)})$ for all r, and E_w is a point-to-set map, we have $\lim_{r\to\infty}\{f_{k[2]}^{(r)}\} \subseteq \lim_{r\to\infty}\{E_w(f_{k[1]}^{(r)})\} = E_w(f_{k[1]}^*)$, where the set limit is taken in the sense of Painlevé-Kuratowski convergence. Since $\lim_{r\to\infty}\{f_{k[2]}^{(r)}\} = f_{k[2]}^*$ is a singleton set by assumption, it leads to $f_{k[2]}^* \in E_w(f_{k[1]}^*)$, which satisfies the condition for closeness.

Given that the algorithm map E is the composition of E_w for all $w \in W$, it follows from Zangwill's theorem that any sequence $\{\mathbf{f}^l\}$ generated by E converges to $\mathbf{f}_n^*$, which corresponds to the equilibrium solution for the restricted master problem at outer iteration n. □

D.2. Proof of Theorem 4

Let $\mathbf{f}^*$ denote the optimality that satisfies the UE condition (9) regarding Problem (11), and $\mathcal{K}_w^n$ the working path set at iteration n. Suppose that $Z(\mathbf{f}_n^*) > Z(\mathbf{f}^*)$ at iteration n, the algorithm moves to $n+1$ to yield $\mathcal{K}_w^{n+1} = \mathcal{K}_{w+}^n \cup \bar{\mathcal{K}}_w^{n+1}$, where $\mathcal{K}_{w+}^n = \{k \mid f_k^{n*} > 0, k \in \mathcal{K}_w^n\}$ is the set of used working paths after accurately solving the restricted master problem, and $\bar{\mathcal{K}}_w^{n+1}$ is the set of generated new paths by Algorithm 2. It's worth emphasizing that $\cup_{w\in W}\bar{\mathcal{K}}_w^{n+1} \neq \emptyset$ since currently there must exists promising new columns outside

$\mathcal{K}_{w+}^{n}$. Given that solving restricted master problem decreases the objective function value by invoking Theorem 3, it follows that $Z(\mathbf{f}_{n+1}^{*} \mid \cup_{w\in W}\mathcal{K}_{w}^{n+1}) < Z(\mathbf{f}_{n}^{*} \mid \cup_{w\in W}\mathcal{K}_{w}^{n}) < \cdots < Z(\mathbf{f}_{1}^{*} \mid \cup_{w\in W}\mathcal{K}_{w}^{1})$. Recall that the objective $Z(\mathbf{f})$ is convex to induce a unique minimum for a given working path set, it yields $\cup_{w\in W}\mathcal{K}_{w}^{n+1} \nsubseteq \cup_{w\in W}\mathcal{K}_{w}^{j}$ for any $j = 1, \cdots, n$, ensuring the algorithm explores a distinct set of working paths at each iteration. We further introduce B_w as the set of all possible OD-based column combinations (a subset of K_w), then the set of all possible column combinations across all OD pairs can be defined by $B = \prod_{w\in W} B_w$, where $|B| = \prod_{w\in W}(2^{|K_w|} - 1)$ is finite. Because (i) the objective of Problem (11) is bounded below by its optimal value $Z(\mathbf{f}^{*})$ and (ii) the maximum number of iterations is bounded by $|B|$, it leads to $\{Z(\mathbf{f}_{n}^{*})\} \to Z(\mathbf{f}^{*})$ after a finite number of iterations. □

E. Path-based algorithm for DMRTA

Algorithm 4 *The CG-GP algorithm for solving DMRTA problems.*

Input: Network $G(N, A)$, a set of user classes M_w for each $w \in W$, class-specific OD demand vector $\check{\mathbf{q}} = (q_{wm})_{w\in W, m\in M_w}$, and link performance functions $(e_a(\cdot))_{a\in A}$ and $(s_a(\cdot))_{a\in A}$.

Output: Equilibrium class-specific path flow $\check{\mathbf{f}}^{*}$.

1. **Initialization.** Initialize counter $n = 0$, assign class-specific demand to minimum-mean path k for each $m \in M_w, w \in W$, i.e., $f_k^{wm} = q_{wm}$. Set the working path set $\mathcal{K}_{wm} = \{k\}$ for class m of OD pair w.
2. **Solving pricing problem.** (i) Set $n = n + 1$; (ii) Obtain the set of efficient paths $\tilde{K}_w$ for OD pair w by one-to-all labeling bi-objective algorithm in Appendix C.1; (iii) For each $m \in M_w, w \in W$, identify path $k = \arg\min_{k'}\{k' | C_{k'}^{w}(\check{\mathbf{f}}^{n}; \gamma_m), k' \in \tilde{K}_w\}$, and update $\mathcal{K}_{wm} = \mathcal{K}_{wm} \cup \{k\}$.
3. **Termination criterion.** The algorithm is terminated when achieving a satisfactory outer convergence, which is measured by the following relative gap:

$$RG_{outer} = 1 - \frac{\sum_{w\in W}\sum_{m\in M_w} q_{wm}\cdot u_{wm}(\check{\mathbf{f}}^{n})}{\sum_{w\in W}\sum_{m\in M_w}\sum_{k\in\mathcal{K}_{wm}} f_k^{wm}\cdot C_k^{w}(\check{\mathbf{f}}^{n};\gamma_m)}, \tag{E.1}$$

   where $u_{wm}(\check{\mathbf{f}}^{n})$ is the minimum path cost obtained by column generation at iteration n.
4. **Solving restricted master problem.** Initialize the inner counter $l = 0$. (i) Set $l = l + 1$; (ii) For each $m \in M_w, w \in W$, update f_k^{wm} with gradient projection as follows:

   4-1. Initialize k and $\bar{k}$ as the shortest-cost path and longest-cost path respectively, set iteration counter $r = 0$ and stepsize $\alpha^{(0)} = 0.1$.

   4-2. Update path cost difference $\tilde{C}_{\bar{k}k}^{wm(r)} = C_{\bar{k}}^{w}(f_k^{wm(r)};\gamma_m) - C_k^{w}(f_k^{wm(r)};\gamma_m)$.

   4-3. If $\tilde{C}_{\bar{k}k}^{wm}(f_k^{wm(r)}) = 0$ or ($\tilde{C}_{\bar{k}k}^{wm(r)} > 0$ and $f_k^{wm(r)} = 0$), then stop updating f_k^{wm}; otherwise move to Step 4-4.

   4-4. Update $f_k^{wm(r+1)} = \min\left\{\max\left\{f_k^{wm(r)} - \alpha^{(r)}\cdot\tilde{C}_{\bar{k}k}^{wm}(f_k^{wm(r)}), 0\right\}, \phi_{k\bar{k}}^{wm}\right\}$, where $\alpha^{(r)}$ is the Barzilai-Borwein stepsize, and $\phi_{k\bar{k}}^{wm} = f_k^{wm(0)} + f_{\bar{k}}^{wm(0)}$ denotes the total flows on paired paths. Set $f_{\bar{k}}^{wm(r+1)} = \phi_{k\bar{k}}^{wm} - f_k^{wm(r+1)}$.

4-5. Update $e_a(x_a)$ and $s_a(x_a)$ for each link a on path k and $\bar{k}$.

4-6. Set $r = r + 1$, and return to Step 4-2.

(iii) Compute the relative gap for inner convergence by

$$RG_{inner} = 1 - \frac{\sum_{w \in W} \sum_{m \in M_w} \sum_{k \in \mathcal{K}_{wm}} q_{wm} \cdot C^w_{min}(\check{\mathbf{f}}^l; \gamma_m)}{\sum_{w \in W} \sum_{m \in M_w} \sum_{k \in \mathcal{K}_{wm}} f^{wm}_k \cdot C^w_k(\mathbf{f}^l; \gamma_m)}, \tag{E.2}$$

where $C^w_{min}(\check{\mathbf{f}}^l; \gamma_m) = \min_{k \in \mathcal{K}_{wm}} C^w_k(\check{\mathbf{f}}^l; \gamma_m)$ is the minimum cumulative cost at inner iteration l. Stop and go to Step 5 if *inner convergence* is achieved, otherwise return to (i).

5. **Dropping.** Remove unused paths for each OD pair, and return to Step 2.

The CG-GP algorithm follows the same fundamental logic when solving either CMRTA or DMRTA problems. However, the primary distinctions between Algorithm 1 (CMRTA) and Algorithm 4 (DMRTA) lie in two key steps: Step 2 (Solving pricing problem) and Step 4 (Solving restricted master problem). The following two remarks provide a detailed analysis of the algorithmic complexity.

REMARK 1 (SOLVING PRICING PROBLEM). In Step 2, the procedure for solving both DMRTA and CMRTA problems consists of two stages. The first stage, which is common to both problems, runs the one-to-all bi-objective labeling algorithm to find all possible efficient paths for OD pairs with the same origin. The time complexity of this stage is discussed in Appendix C.1. The second stage generates promising paths for the considered user classes given the set of efficient paths $\tilde{K}_w$ for each OD pair. For CMRTA, the time complexity is $O(|\tilde{K}_w| \cdot \log(|\tilde{K}_w|))$ (cf. Section 5.1). For DMRTA, the procedure directly searches for the minimum-cost path for each user class, resulting in the time complexity of $O(|\tilde{K}_w| \cdot |M_w|)$. Notably, generating new paths for the DMRTA problem would incur higher computational time for a sufficiently large number of $|M_w|$.

REMARK 2 (SOLVING RESTRICTED MASTER PROBLEM). In Step 4, we apply the gradient projection algorithm to sequentially solve the paired path problem defined upon the working path set. The key distinction between CMRTA and DMRTA lies in the selection of paired paths: CMRTA uses paths with the largest cumulative cost difference, while DMRTA uses those with the largest path cost difference. The one-time projection involves traversing stored links on paired paths four times: twice for computing first-order derivatives (Step 4-2) and twice for updating affected link flows (Step 4-5). Supposing there are two working paths for all OD pairs, with each path containing approximately $\sqrt{|N|}$ links (Jayakrishnan et al. 1994). Under these conditions, the time complexity for a one-time projection across all OD pairs is roughly $O(4|W|\sqrt{|N|})$ for CMRTA and $O(4\sum_w |M_w|\sqrt{|N|})$ for DMRTA. This analysis reveals that DMRTA's time complexity scales linearly with the number of user classes, potentially resulting in higher computational costs as the number of classes increases.

F. Code and data

We provide the source code and network data in the supplementary material, including:

- **Solution** folder: Contains the top-level file *AlgSolution.sln*, which stores information about the projects, dependencies, build configurations, and other settings.
- **DllProj** folder: The DLL project contains functionality resources related to the CG-GP algorithm.
- **ExeDriver** folder: The executable project contains the entry point for executing the algorithm.
- **Networks** folder: Contains all network data applied in this paper.

Please refer to the "README.md" file within the material for a detailed description of each component.

References

1. Jayakrishnan R, Tsai W, Prashker J, Rajadhyaksha S (1994) A faster path-based algorithm for traffic assignment. *Transportation Research Record* 1443:75-83.
2. Prakash AA, Seshadri R, Srinivasan KK (2018) A consistent reliability-based user-equilibrium problem with risk-averse users and endogenous travel time correlations: Formulation and solution algorithm. Transportation Research Part B: Methodological 114:171-198.
3. Sedeño-Noda A, Colebrook M (2019) A biobjective dijkstra algorithm. *European Journal of Operational Research* 276(1):106-118.
4. Seshadri R, Srinivasan KK (2017) Robust traffic assignment model: Formulation, solution algorithms and empirical application. Journal of Intelligent Transportation Systems 21(6):507-524.