Appendices

A. The upper bound of r_2

When testing the conditions for Theorem 1, if the calculation of r_2 is hard or not necessary, one can turn to finding the upper bound of r_2 instead. According to the Hoffman' error bound theorem (Facchinei and Pang 2007), there exists a constant $H(\hat{Q})$ that depends only on the matrix $\hat{Q}$, such that the L_2 distance from (d^0, v^0) to S is bounded above by $H(\hat{Q}) \cdot (\|\hat{Q}d^0 - v^0\| + \|min(0,(d^0,v^0))\|)$, where $H(\hat{Q})$ is called the Hoffman's constant. If we use r_3 to denote such an L_2 distance, it turns out that $r_3 \leq H(\hat{Q}) \cdot (\|\hat{Q}d^0 - v^0\| + \|min(0,(d^0,v^0))\|)$. Since for any vector x, it is always true that $\|x\|_1 \leq \sqrt{n}\|x\|_2, \forall x \in \mathbb{R}^n$. Based on the definition of r_2, we have $r_2 \leq \sqrt{n}r_3 \leq \sqrt{n}H(\hat{Q}) \cdot (\|\hat{Q}d^0 - v^0\| + \|min(0,(d^0,v^0))\|)$. This means $\sqrt{n}H(\hat{Q}) \cdot (\|\hat{Q}d^0 - v^0\| + \|min(0,(d^0,v^0))\|)$ can be considered as an upper bound of r_2. According to Theorem 1, if r_1 is greater than this upper bound, it can be stated that $\hat{d} \in K$ and the OD quasi-sparsity consistency holds. Unfortunately, calculating the Hoffman's constant is very challenging (Klatte and Thiere 1995). We could however estimate the Hoffman's constant via some algorithmic approaches based on the characterization of matrix $\hat{Q}$ (Güler, Hoffman, and Rothblum 1995, Wang and Lin 2014, Pena, Vera, and Zuluaga 2018). Details are omitted here to save space.

B. Proof of Theorem 2

Proof First, because the matrix $\hat{Q}$ in (3) is nonnegative, we can remove the constraint $v \geq 0$ there to rewrite it as

$$\begin{aligned} \min_{d\in\mathbb{R}^{|W|}, v\in\mathbb{R}^{|A|}} \quad & \|d - d^0\|_1 + \|v - v^0\|_1 \\ \text{s.t. } & \hat{Q}d - v = 0; \; d \geq 0. \end{aligned} \tag{44}$$

Since $(\hat{d}, \hat{v})$ is an extreme point to the set of optimal solutions of (3), it is also an extreme point to the set of optimal solutions of (44).

Next, we apply Lemma 1 to (44). The notation in Lemma 1 is related to notation here in the following way: $x = (d, \bar{v})$, $x^0 = (d^0, v^0)$,

$$M = \begin{bmatrix} \hat{Q} & -I \end{bmatrix}, \quad b = 0, \quad N = I \quad \text{and} \quad h = 0,$$

where the identity matrices I in M and N are of dimensions $|A| \times |A|$ and $|W| \times |W|$ respectively. The index set E in Lemma 1 is given by W_0. In addition, $J_P = W_P \cup A_P$, $J_Z = W_Z \cup A_Z$, and $J_N = W_N \cup A_N$. We further partition W_N and W_0 as

$$W_N = (W_N \cap W_0) \; \cup \; (W_N \setminus W_0),$$
$$W_0 = (W_Z \cap W_0) \; \cup \; (W_N \cap W_0);$$

recall that $W_P \cap W_0 = \emptyset$. Then, the matrix in Lemma 1

$$\begin{bmatrix} M_{:,J_P} & M_{:,J_N} \\ N_{E,J_P} & N_{E,J_N} \end{bmatrix}$$

is given by

$$\begin{array}{r|ccccc|}
 & |W_P| & |W_N \cap W_0| & |W_N \setminus W_0| & |A_P| & |A_N| \\ \hline
|A_P| & \hat{Q}_{A_P,W_P} & \hat{Q}_{A_P,W_N\cap W_0} & \hat{Q}_{A_P,W_N\setminus W_0} & -I & 0 \\
|A_Z| & \hat{Q}_{A_Z,W_P} & \hat{Q}_{A_Z,W_N\cap W_0} & \hat{Q}_{A_Z,W_N\setminus W_0} & 0 & 0 \\
|A_N| & \hat{Q}_{A_N,W_P} & \hat{Q}_{A_N,W_N\cap W_0} & \hat{Q}_{A_N,W_N\setminus W_0} & 0 & -I \\
|W_Z \cap W_0| & 0 & 0 & 0 & 0 & 0 \\
|W_N \cap W_0| & 0 & I & 0 & 0 & 0 \\ \hline
\end{array}$$

which becomes

$$\begin{array}{r|ccccc|}
 & |W_P| & |W_N \cap W_0| & |W_N \setminus W_0| & |A_P| & |A_N| \\ \hline
|A_P| & \hat{Q}_{A_P,W_P} & 0 & \hat{Q}_{A_P,W_N\setminus W_0} & -I & 0 \\
|A_Z| & \hat{Q}_{A_Z,W_P} & 0 & \hat{Q}_{A_Z,W_N\setminus W_0} & 0 & 0 \\
|A_N| & \hat{Q}_{A_N,W_P} & 0 & \hat{Q}_{A_N,W_N\setminus W_0} & 0 & -I \\
|W_Z \cap W_0| & 0 & 0 & 0 & 0 & 0 \\
|W_N \cap W_0| & 0 & I & 0 & 0 & 0 \\ \hline
\end{array}$$

after elementary row operations. The rank of the entire matrix above is at most

$$|W_N \cap W_0| + |A_P| + |A_Z| + |A_N| = |W_N \cap W_0| + |A|.$$

By Lemma 1, we have

$$|J_Z| = |W_Z| + |A_Z| \geq (|W| + |A|) - (|W_N \cap W_0| + |A|).$$

Because $|W_N \cap W_0| + |W_Z| = |W_0 \cup W_Z|$, we obtain (12). □

C. Proof of Theorem 5

Proof First, since (d^*, v^*) satisfies Assumption 2 and is the unique optimal solution to (20), it is a strict local solution to (15) by Theorem 3.

Next, we analyze the solution sparsity of (d^*, v^*), by applying Lemma 1. Because rows in Q^* corresponding to indices $a \in A_0^*$ are zero, we can consider a reduced problem in which the links in A_0^* are removed. Let $\bar{v}$, $\bar{v}^*$ and $\bar{v}^0$ be subvectors of v, v^* and v^0 consisting of components corresponding to $a \notin A_0^*$, and $\bar{Q}^*$ be the submatrix of Q^* consisting of rows with indices not in A_0^*. Because (d^*, v^*) is the unique optimal solution to (20), $(d^*, \bar{v}^*)$ is the unique optimal solution to the following reduced problem:

$$\begin{aligned}
&\min_{d,\bar{v}} \quad \|d - d^0\|_1 + \|\bar{v} - \bar{v}^0\|_1 \\
&\text{s.t.} \quad (d, \bar{v}) \in \bar{S}^{**} = \left\{(d, \bar{v}) \;\middle|\; \bar{Q}^*(d - d^*) + \bar{v}^* = \bar{v};\; d \geq \epsilon,\; \bar{v} \geq 0\right\}.
\end{aligned} \tag{45}$$

The dimension of d is still $|W|$, the dimension of $\bar{v}$ is $|A| - |A_0^*|$ instead of $|A|$. By the way $\bar{v}^*$ is defined, the set $\{a \in A \setminus A_0^* : \bar{v}_a^* = 0\}$ is empty. The set $A \setminus A_0^*$ is partitioned into the following sets

$$\bar{A}_P^* = \{a \in A \setminus A_0^* :\; \bar{v}_a^* - \bar{v}_a^0 > 0\},$$

$$\bar{A}_Z^* = \{a \in A \setminus A_0^* :\; \bar{v}_a^* - \bar{v}_a^0 = 0\},$$

$$\bar{A}_N^* = \{a \in A \setminus A_0^* :\; \bar{v}_a^* - \bar{v}_a^0 < 0\},$$

which are related to A_P^*, A_Z^* and A_N^* by

$$\bar{A}_P^* = A_P^* \setminus A_0^* = A_P^*, \quad \bar{A}_Z^* = A_Z^* \setminus A_0^*, \quad \bar{A}_N^* = A_N^* \setminus A_0^*. \tag{46}$$

Because d^* and d^0 are the same as before, W is partitioned by the same sets W_P^*, W_Z^* and W_N^* defined in (28).

Next, we apply Lemma 1 to the problem (45). To relate the notation in Lemma 1 to the notation here, note that $x = (d, \bar{v})$, $x^0 = (d^0, \bar{v}^0)$,

$$M = \begin{bmatrix} \bar{Q}^* & -I \end{bmatrix}, \quad b = \bar{Q}^* d^* - \bar{v}^*, \quad N = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \quad \text{and} \quad h = \begin{bmatrix} \epsilon \\ 0 \end{bmatrix}.$$

The index set E is given by W_ϵ^*. In addition, $J_P = W_P^* \cup \bar{A}_P^*$, $J_Z = W_Z^* \cup \bar{A}_Z^*$, and $J_N = W_N^* \cup \bar{A}_N^*$. We further partition the index sets as

$$\begin{aligned}
W_P^* &= (W_P^* \cap W_\epsilon^*) \ \cup \ (W_P^* \setminus W_\epsilon^*), \\
W_N^* &= (W_N^* \cap W_\epsilon^*) \ \cup \ (W_N^* \setminus W_\epsilon^*), \\
W_\epsilon^* &= (W_P^* \cap W_\epsilon^*) \ \cup \ (W_Z^* \cap W_\epsilon^*) \ \cup \ (W_N^* \cap W_\epsilon^*).
\end{aligned}$$

Then, we write the matrix in Lemma 1

$$\begin{bmatrix} M_{:,J_P} & M_{:,J_N} \\ N_{E,J_P} & N_{E,J_N} \end{bmatrix}$$

in the following form:

$$\begin{array}{r|cccccc|}
 & |W_P^* \cap W_\epsilon^*| & |W_P^* \setminus W_\epsilon^*| & |W_N^* \cap W_\epsilon^*| & |W_N^* \setminus W_\epsilon^*| & |\bar{A}_P^*| & |\bar{A}_N^*| \\ \hline
|\bar{A}_P^*| & \bar{Q}^*_{\bar{A}_P^*, W_P^* \cap W_\epsilon^*} & \bar{Q}^*_{\bar{A}_P^*, W_P^* \setminus W_\epsilon^*} & \bar{Q}^*_{\bar{A}_P^*, W_N^* \cap W_\epsilon^*} & \bar{Q}^*_{\bar{A}_P^*, W_N^* \setminus W_\epsilon^*} & -I & 0 \\
|\bar{A}_Z^*| & \bar{Q}^*_{\bar{A}_Z^*, W_P^* \cap W_\epsilon^*} & \bar{Q}^*_{\bar{A}_Z^*, W_P^* \setminus W_\epsilon^*} & \bar{Q}^*_{\bar{A}_Z^*, W_N^* \cap W_\epsilon^*} & \bar{Q}^*_{\bar{A}_Z^*, W_N^* \setminus W_\epsilon^*} & 0 & 0 \\
|\bar{A}_N^*| & \bar{Q}^*_{\bar{A}_N^*, W_P^* \cap W_\epsilon^*} & \bar{Q}^*_{\bar{A}_N^*, W_P^* \setminus W_\epsilon^*} & \bar{Q}^*_{\bar{A}_N^*, W_N^* \cap W_\epsilon^*} & \bar{Q}^*_{\bar{A}_N^*, W_N^* \setminus W_\epsilon^*} & 0 & -I \\
|W_P^* \cap W_\epsilon^*| & I & 0 & 0 & 0 & 0 & 0 \\
|W_Z^* \cap W_\epsilon^*| & 0 & 0 & 0 & 0 & 0 & 0 \\
|W_N^* \cap W_\epsilon^*| & 0 & 0 & I & 0 & 0 & 0 \\ \hline
\end{array}$$

By subtracting appropriate multiples of the fourth row and the sixth row from the first, second, and third row, the matrix becomes

$$\begin{array}{r|cccccc|}
 & |W_P^* \cap W_\epsilon^*| & |W_P^* \setminus W_\epsilon^*| & |W_N^* \cap W_\epsilon^*| & |W_N^* \setminus W_\epsilon^*| & |\bar{A}_P^*| & |\bar{A}_N^*| \\ \hline
|\bar{A}_P^*| & 0 & \bar{Q}^*_{\bar{A}_P^*, W_P^* \setminus W_\epsilon^*} & 0 & \bar{Q}^*_{\bar{A}_P^*, W_N^* \setminus W_\epsilon^*} & -I & 0 \\
|\bar{A}_Z^*| & 0 & \bar{Q}^*_{\bar{A}_Z^*, W_P^* \setminus W_\epsilon^*} & 0 & \bar{Q}^*_{\bar{A}_Z^*, W_N^* \setminus W_\epsilon^*} & 0 & 0 \\
|\bar{A}_N^*| & 0 & \bar{Q}^*_{\bar{A}_N^*, W_P^* \setminus W_\epsilon^*} & 0 & \bar{Q}^*_{\bar{A}_N^*, W_N^* \setminus W_\epsilon^*} & 0 & -I \\
|W_P^* \cap W_\epsilon^*| & I & 0 & 0 & 0 & 0 & 0 \\
|W_Z^* \cap W_\epsilon^*| & 0 & 0 & 0 & 0 & 0 & 0 \\
|W_N^* \cap W_\epsilon^*| & 0 & 0 & I & 0 & 0 & 0 \\ \hline
\end{array}$$

It is clear that the rank of the entire matrix above is at most

$$|W_P^* \cap W_\epsilon^*| + |W_N^* \cap W_\epsilon^*| + |\bar{A}_P^*| + |\bar{A}_Z^*| + |\bar{A}_N^*| = |W_\epsilon^* \setminus W_Z^*| + |A| - |A_0^*|.$$

By Lemma 1, we have

$$|J_Z| = |W_Z^*| + |\bar{A}_Z^*| \geq (|W| + |A| - |A_0^*|) - (|W_\epsilon^* \setminus W_Z^*| + |A| - |A_0^*|).$$

Using (46), we rewrite the above inequality as

$$|W_\epsilon^* \setminus W_Z^*| + |W_Z^*| + |A_Z^* \setminus A_0^*| \geq |W|.$$

Because $|W_\epsilon^* \setminus W_Z^*| + |W_Z^*| = |W_\epsilon^* \cup W_Z^*|$, we obtain (29). $\square$

D. Differentiability of $\Psi(d)$

Recall that, under Assumption 1, for each $d \in \mathbb{R}_+^{|W|}$ there is a unique user-equilibrium link flow denoted by $\Psi(d)$, which is the unique solution to the following traffic user equilibrium problem

$$v \in G(d) \quad \text{and} \quad \langle c(v), v' - v \rangle \geq 0 \text{ for each } v' \in G(d). \tag{47}$$

Here, $c : \mathbb{R}_+^{|A|} \to \mathbb{R}^{|A|}$ is the link cost function, and $G(d)$ is the set of feasible link flows defined as

$$G(d) = \left\{ v \in \mathbb{R}^{|A|} \mid \ v = \Delta q \ \text{ for some } q \in \mathbb{R}_+^{|\mathcal{R}|} \ \text{ satisfying } \Lambda q = d \right\}, \tag{48}$$

where Λ and Δ are the OD-route incidence matrix and the link-route incidence matrix respectively. In this section, we consider a point $d^* \in \mathbb{R}^{|W|}$ with $d^* > 0$, and give a condition that guarantees Ψ to be continuously differentiable on an open neighborhood of d^* in $\mathbb{R}_{++}^{|W|}$. See Lu (2008) for more references on sensitivity analysis of the traffic user equilibrium problem.

The traffic user equilibrium problem can also be written in a route-based formulation

$$q \in H(d) \quad \text{and} \quad \langle C(q), q' - q \rangle \geq 0 \text{ for each } q' \in H(d), \tag{49}$$

where $C(q) = \Delta^T c(\Delta q)$ is the route cost function and $H(d)$ is the set of feasible route flows under d defined as

$$H(d) = \left\{ q \in \mathbb{R}_+^{|\mathcal{R}|} \mid \ \Lambda q = d \right\}.$$

A solution to (49) is called a user-equilibrium route flow under d. The formulation (49) is a characterization of the Wardrop traffic equilibrium, in the sense that $q \in H(d)$ solves (49) if and only if $q_p = 0$ for any non-user-optimal route p. Finally, (47) and (49) are related in that v is a solution to (47) if and only if there exists a solution q of (49) with $v = \Delta q$, and that $q \in H(d)$ solves (49) if and only if Δq solves (47).

Now, let $v^* = \Psi(d^*)$, and let $q^* \in \mathbb{R}_+^{|\mathcal{R}|}$ be a user-equilibrium route flow under d^*. Since v^* is the only user-equilibrium link flow under Assumption 1, we have $v^* = \Delta q^*$. It is possible to have multiple user-equilibrium route flows corresponding to v^*, so the choice of q^* is not unique. Since $d^* > 0$, for each OD pair w there exists at least one route p connecting w with $q_p^* > 0$; we choose one such route for each w and denote it by $p(w)$. We then divide the set $\mathcal{R}$ of all routes into the following four subsets:

$$\begin{aligned}
&\mathcal{R}^{00} \text{ is the set of all user-optimal routes } p \text{ with } q_p^* = 0,\\
&\mathcal{R}^{01} \text{ is the set of all non-user-optimal routes},\\
&\mathcal{R}^{1} = \{p(w), w \in W\},\\
&\mathcal{R}^{2} \text{ is the set of all routes } p \text{ with } q_p^* > 0 \text{ not in } \mathcal{R}^1.
\end{aligned} \tag{50}$$

We denote the numbers of routes in $\mathcal{R}^{00}$, $\mathcal{R}^{01}$, $\mathcal{R}^1$ and $\mathcal{R}^2$ by $\pi_{00}, \pi_{01}, \pi_1$, and π_2 respectively; note that $\pi_1 = |W|$. We then partition the matrices Λ and Δ by column as

$$\begin{bmatrix} \Lambda \\ \Delta \end{bmatrix} = \begin{bmatrix} \Lambda_{00} & \Lambda_{01} & \Lambda_1 & \Lambda_2 \\ \Delta_{00} & \Delta_{01} & \Delta_1 & \Delta_2 \end{bmatrix} \tag{51}$$

corresponding to routes in $\mathcal{R}^{00}, \mathcal{R}^{01}, \mathcal{R}^1$, and $\mathcal{R}^2$ respectively. While different choices of q^* may result in different $\mathcal{R}^{00}, \mathcal{R}^1, \mathcal{R}^2$, they all lead to the same $\mathcal{R}^{01}$ and $\mathcal{R}^{00} \cup \mathcal{R}^1 \cup \mathcal{R}^2$ because the route cost $C(q^*)$ is the same under different choices of q^*.

The following theorem provides a condition under which Ψ is continuously differentiable in a neighborhood of d^*, and a formula to compute the Jacobian matrix $\nabla\Psi(d)$ for d in this neighborhood. The theorem assumes the link cost function c to be continuously differentiable on an open neighborhood of v^* in $\mathbb{R}^{|A|}$, while the domain of the link cost function c here is $\mathbb{R}^{|A|}_+$. There are two ways to handle situations in which some entries in v^* are zero. One way is to extend the domain of c to include an open neighborhood of v^* in $\mathbb{R}^{|A|}$, to evaluate the Jacobian matrix $\nabla c(v^*)$. Another way is to first remove the links $a \in A$ with v^*_a from the network to compute $\nabla\Psi(d)$ using (53) for the reduced network, and then add zero rows to $\nabla\Psi(d)$ corresponding to those a back; this works because the links with zero flow remain to have zero flow under small perturbation of d under condition (2) in the theorem. As long as c can be extended to a continuously differentiable function on a neighborhood of v^* in $\mathbb{R}^{|A|}$, these two treatments will give the same Jacobian matrix $\nabla\Psi(d)$ for the original network, because the rows in $\tilde{A}$ and Δ_1 corresponding to links a with $v^*_a = 0$ are all zero.

Theorem 6. *Suppose Assumption 1 holds. Let $d^* > 0$, $v^* = \Psi(d^*)$, and q^* be a user-equilibrium route flow under d^*. Suppose that the link cost function c is continuously differentiable on an open neighborhood V of v^* in $\mathbb{R}^{|A|}$, and write $\tilde{L} = \nabla c(v^*) \in \mathbb{R}^{|A|\times|A|}$. Let $\tilde{A}$ be a matrix whose columns form a basis for the column space of the matrix $[\Delta_2 - \Delta_1\Lambda_2, \Delta_{00} - \Delta_1\Lambda_{00}]$. Suppose that the following two conditions hold:*

(1) The matrix $\tilde{A}^T\tilde{L}\tilde{A}$ is nonsingular.

(2) There exist $u_{\pi_{00}} \in \mathbb{R}^{\pi_{00}}$ and $v_{\pi_2} \in \mathbb{R}^{\pi_2}$ such that $u_{\pi_{00}} > 0$ and

$$(\Delta_{00} - \Delta_1\Lambda_{00})u_{\pi_{00}} + (\Delta_2 - \Delta_1\Lambda_2)v_{\pi_2} = 0. \tag{52}$$

Then, Ψ is continuously differentiable on an open neighborhood $\mathcal{O}$ of d^* in $\mathbb{R}^{|W|}_{++}$ with

$$\nabla\Psi(d) = \tilde{A}\left(\tilde{A}^T\nabla c(\Psi(d))\tilde{A}\right)^{-1}\tilde{A}^T\left[-\nabla c(\Psi(d))\Delta_1\right] + \Delta_1 \quad \textit{for each } d \in \mathcal{O}. \tag{53}$$

Proof By Lu (2008, Section 2.5), condition (2) here implies that the critical cone K defined in that paper (the critical cone K is different from the set K in this paper) is the column space of $\tilde{A}$. Condition (1) here implies Condition 1.1 in Lu (2008) holds, with the matrix $\tilde{A}$ here in place of the matrix A in that paper; Condition 1.2 in Lu (2008) holds because the matrix B there is empty. Thus, by Lu (2008, Theorem 2.3), Ψ is differentiable at d^* under assumptions given in this theorem, and the Jacobian matrix $\nabla\Psi(d^*)$ is given by (53) with $\tilde{L} = \nabla c(\Psi(d^*))$. It remains to show that Ψ is continuously differentiable on an open neighborhood $\mathcal{O}$ of d^* in $\mathbb{R}^{|W|}_{++}$. By Lu (2008, Proposition

2.2), condition (2) here guarantees the existence of a user-equilibrium route flow $q^{**} \in H(d^*)$ with $\Delta q^{**} = v^*$ and $q^{**}_p > 0$ for all user-optimal routes $p \in \mathcal{R}^{00} \cup \mathcal{R}^1 \cup \mathcal{R}^2$. On the other hand, by Lu (2008, Theorem 2.3), there exists a neighborhood D' of d^* in $\mathbb{R}^{|W|}_{++}$ such that $\Psi(\cdot)$ is continuous on D' and $\Psi(d) \in V$ for each $d \in D'$. Now apply Facchinei and Pang (2007, Corollary 3.2.5(a)) to the set-valued map

$$(d, v) \to \{q \in \mathbb{R}^{|\mathcal{R}|} : \ q \geq 0, \ \Lambda q = d, \ \Lambda q = v\}$$

and the base point (d^*, v^*, q^{**}), to find an open neighborhood D'' of d^* in D', such that for each $d \in D''$ there exists a user-equilibrium route flow $q(d)$ with $q(d^*) = q^{**}$ and $q_p(d) > 0$ for all $p \in \mathcal{R}^{00} \cup \mathcal{R}^1 \cup \mathcal{R}^2$. Moreover, with the continuity of Ψ, we can shrink D'' further to ensure that non-user-optimal routes under d^* continue to be non-user-optimal under any $d \in D''$. It follows that $q_p(d) = 0$ for all $p \in \mathcal{R}^{01}$ and $d \in D''$. Lastly, by shrinking D'' further if necessary, we can ensure that $\tilde{A}^T \nabla c(\Psi(d))\tilde{A}$ is nonsingular for each $d \in D''$.

Now, apply (Lu 2008, Theorem 2.3) to an arbitrary $d \in D''$ and the corresponding user-equilibrium arc flow $\Psi(d)$, using $q(d)$ to partition the route set as $\mathcal{R} = \bar{\mathcal{R}}^{00} \cup \bar{\mathcal{R}}^{01} \cup \bar{\mathcal{R}}^1 \cup \bar{\mathcal{R}}^2$. Because $q_p(d) > 0$ for $p \in \mathcal{R}^{00} \cup \mathcal{R}^1 \cup \mathcal{R}^2$, we can let $\bar{\mathcal{R}}^1 = \mathcal{R}^1$. We have $\bar{\mathcal{R}}^{00} = \emptyset$ because all user-optimal routes are used under $q(d)$, $\bar{\mathcal{R}}^{01} = \mathcal{R}^{01}$ because non-user-optimal (user-optimal) routes under d^* remain to be non-user-optimal (user-optimal) under d, and $\bar{\mathcal{R}}^2 = \mathcal{R}^2 \cup \mathcal{R}^{00}$ consists of routes p with $q_p(d) > 0$ not in $\mathcal{R}^1$. Let $\bar{\Lambda}_2$ and $\bar{\Delta}_2$ be submatrices of Λ and Δ corresponding to routes in $\bar{\mathcal{R}}^2$; the matrix $[\bar{\Delta}_2 - \Delta_1 \bar{\Lambda}_2]$ contains the same columns as $[\Delta_2 - \Delta_1 \Lambda_2, \Delta_{00} - \Delta_1 \Lambda_{00}]$, so columns of the same matrix $\tilde{A}$ defined in the statement of the present theorem form a basis for the column space of $[\bar{\Delta}_2 - \Delta_1 \bar{\Lambda}_2]$. The first condition for Ψ to be differentiable at d holds because $\tilde{A}^T \nabla c(\Psi(d))\tilde{A}$ is nonsingular for each $d \in D''$ by the choice of D'', and the second condition also holds because $\bar{\mathcal{R}}^{00}$ is empty. Consequently, Ψ is differentiable at d with

$$\nabla \Psi(d) = \tilde{A}\left(\tilde{A}^T \nabla c(\Psi(d))\tilde{A}\right)^{-1} \tilde{A}^T \left[-\nabla c(\Psi(d))\Delta_1 \right] + \Delta_1,$$

which is continuous with respect to d due to the continuous differentiability of c and the continuity of Ψ. Letting $\mathcal{O} = D''$ completes the proof. $\square$

E. Definition of F1-score, accuracy, and RMSE

To define F1-score, it is necessary to introduce the confusion matrix that is widely used in statistical classification. As shown in Table 13, for a given network, we define the set of insignificant/significant OD pairs based on the true OD demand matrix (denoted as W_1 and $W \setminus W_1$ in this paper) as the ground truth, while the set of the insignificant/significant OD pairs under the estimated OD demands as the prediction.

Table 13 Confusion matrix of OD quasi-sparsity consistency

		Partitions in the true OD matrix (true)	
		Insignificant(+)	Significant(-)
Partitions in the estimated OD matrix (prediction)	Insignificant(+)	True positives (TP)	False positives (FP)
	Significant(-)	False negatives (FN)	True negatives (TN)

The confusion matrix can be used to define the True Positive Rate (TPR, or recall/sensitivity) and precision, which can then be used to define F1-score. We give their formulas as follows.

$$TPR = \frac{TP}{TP+FN} \tag{54}$$

$$precision = \frac{TP}{TP+FP} \tag{55}$$

$$F_1 = \frac{2 \times precision \times TPR}{precision + TPR} \tag{56}$$

TPR indicates the portion of insignificant OD pairs in the true OD demands that still keep insignificant in the estimated OD demands, as calculated in (54). Precision means the portion of insignificant OD pairs in the estimated OD demands that are insignificant in the true OD demands, as calculated in (55). F1-score is the harmonic mean of TPR and precision, which is widely used as the performance measure for classification algorithms since it is less sensitive to imbalanced dataset by considering both TPR and precision. As the OD demand dataset is highly imbalanced (with a large portion of insignificant OD pairs and a small portion of significant OD pairs), F1-score is an ideal measure to evaluate the quasi-sparsity consistency of the proposed QSOD models. The classification becomes better as the F1-score gets closer to 1, indicating more OD pairs in the estimated OD demand matrix will remain insignificant or significant as the true OD matrix.

Accuracy is defined in (57) based on the confusion matrix.

$$Accuracy = \frac{TP+TN}{Population} \tag{57}$$

Lastly, Root Mean Square Error (RMSE) is also used to evaluate how close the estimated value is to the "ground truth" value:

$$RMSE(d,\bar{d}) = \sqrt{\frac{\sum_{w\in W}(d_w - \bar{d}_w)^2}{|W|}} \tag{58}$$

where d is the estimated OD demands and $\bar{d}$ is the true OD demands.

F. Discussion about the solution closeness of QSOD and GLS

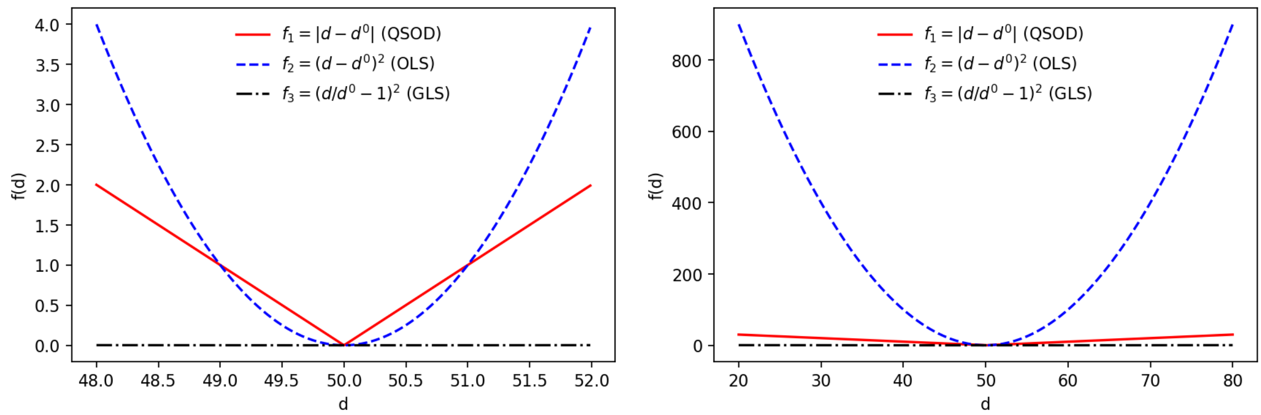
Here we discuss, for the fixed-mapping models, why the solution of GLS gets much closer to the QSOD solution after adding proper weights, when compared with the OLS solution (with no weight added). Denote (d_1,v_1) and (d_2,v_2) as the unique optimal solution and f_1 and f_2 as the objective function for the fixed-mapping QSOD model and OLS model, respectively. According to (Robinson and Lu 2008, Theorem 2), if both f_1 and f_2 are strongly monotone with modulus m and Lipschitzian with modules M, we have the following inequality:

$$||(d_1,v_1)-(d_2,v_2)|| \le \theta^{-1}\mu||f_1(d_1,v_1)-f_2(d_1,v_1)|| \tag{59}$$

where $\mu = M^{-2}m$ and $\theta = (1-[1-(m/M)^2]^{1/2})^{-1}$. (59) in fact gives an upper bound for the difference between the optimal solutions of the two models (QSOD and OLS), and such bound is only determined by (d_1,v_1), f_1, f_2, θ and μ, where the last two parameters rely on the collection of single-valued functions F which contain f_1 and f_2. Assume such function collection F also includes the GLS objective f_3 and denote the optimal solution for the GLS model as (d_3, v_3). Similarly, we will have:

$$||(d_1, v_1) - (d_3, v_3)|| \leq \theta^{-1}\mu||f_1(d_1, v_1) - f_3(d_1, v_1)|| \tag{60}$$

Notice that the RHS of (59) and (60) only differs in the OLS and GLS objective function (f_2 and f_3 respectively). If we only illustrate the problem in one dimension (as well as ignore the link flow deviations as they are similar to the OD demand deviations), then the QSOD, OLS, and GLS objective will have the following form respectively: $f_1 = |d - d^0|$, $f_2 = (d - d^0)^2$, and $f_3 = (d - d^0)^2/(\epsilon_1 d^0)^2$. Further we can ignore ϵ_1 in f_3 as ϵ_1 is a constant for all OD pairs. Therefore f_3 can be written as $(d/d^0 - 1)^2$. Taking $d^0 = 50$, Figure 7 shows the plots of the three functions in a range of d that include d^0. It can be seen that in the range that closes to d^0 (e.g. $d \in (48, 52)$), $|f_1 - f_2|$ is less than or equal to $|f_1 - f_3|$. When d deviates a lot from d^0 (as shown in Figure 7b), $|f_1 - f_2|$ will be much greater than $|f_1 - f_3|$. In the OD estimation problem, once the estimate d (or v) goes far from its prior value d^0 (or v^0), the RHS of (60) will be much less than that of (59), which indicates that $||(d_1, v_1) - (d_3, v_3)||$ has a small upper bound compared with $||(d_1, v_1) - (d_2, v_2)||$. This can partially explain the solution closeness between QSOD and GLS (compared with QSOD and OLS).



(a) $d \in (48, 52)$ (b) $d \in (20, 80)$

Figure 7 **Comparison across QSOD, OLS, and GLS objective function (one-dimension, $d^0 = 50$)**