Inverse Game Theory: An Incenter-Based Approach (Supplementary Material)

1 Proof of Theorem 1

Proof. Since the angle in the objective is scale-invariant, we first assume that $\|\boldsymbol{\vartheta}\|_2 = \|\tilde{\boldsymbol{\vartheta}}\|_2 = 1$ for ease of the derivation. We later demonstrate that this constraint can be removed in the convex formulation. Under this assumption, we have

$$\langle \boldsymbol{\vartheta}, \tilde{\boldsymbol{\vartheta}} \rangle = -\frac{1}{2}\left\|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right\|_2^2 + 1. \tag{1}$$

The problem (5) in Section 4.2 can be rewritten as

$$\begin{aligned}
&\mathop{\arg\max}_{\|\boldsymbol{\vartheta}\|_2=1} \min_{\substack{\tilde{\boldsymbol{\vartheta}} \in \overline{\mathrm{int}(\mathbb{C})} \\ \|\tilde{\boldsymbol{\vartheta}}\|_2=1}} a\left(\boldsymbol{\vartheta}, \tilde{\boldsymbol{\vartheta}}\right) = \mathop{\arg\max}_{\|\boldsymbol{\vartheta}\|_2=1} \min_{\substack{\tilde{\boldsymbol{\vartheta}} \in \overline{\mathrm{int}(\mathbb{C})} \\ \|\tilde{\boldsymbol{\vartheta}}\|_2=1}} \arccos\langle \boldsymbol{\vartheta}, \tilde{\boldsymbol{\vartheta}} \rangle \\
&= \mathop{\arg\max}_{\|\boldsymbol{\vartheta}\|_2=1} \min_{\substack{\tilde{\boldsymbol{\vartheta}} \in \overline{\mathrm{int}(\mathbb{C})} \\ \|\tilde{\boldsymbol{\vartheta}}\|_2=1}} \arccos\left(-\frac{1}{2}\left\|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right\|_2^2 + 1\right) \\
&= \mathop{\arg\max}_{\|\boldsymbol{\vartheta}\|_2=1} \min_{\substack{\tilde{\boldsymbol{\vartheta}} \in \overline{\mathrm{int}(\mathbb{C})} \\ \|\tilde{\boldsymbol{\vartheta}}\|_2=1}} \left\|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right\|_2^2.
\end{aligned} \tag{2}$$

The above equation follows from the facts that the range of the function $\arccos$ is $[0, \pi]$, and $-\cos(\gamma)$ is monotone increasing for $\gamma \in [0, \pi]$.

In order to minimize the distance to any $\boldsymbol{\vartheta}$, the optimal $\tilde{\boldsymbol{\vartheta}}$ will always be in the boundary of $\mathbb{C}$. Hence, the inner minimization problem in equation (2) becomes

$$\min_{\substack{\tilde{\boldsymbol{\vartheta}} \in \overline{\mathrm{int}(\mathbb{C})} \\ \|\tilde{\boldsymbol{\vartheta}}\|_2=1}} \left\|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right\|_2^2 = \min_{\substack{\boldsymbol{x}^j \in \mathcal{X}^j \\ j \in [N]}} \left\{ \min_{\substack{\langle \boldsymbol{\Phi}^j_{\tilde{\boldsymbol{\vartheta}}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle = 0 \\ \|\tilde{\boldsymbol{\vartheta}}\|_2=1}} \left\|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right\|_2^2 \right\}. \tag{3}$$

Next, we first derive a solution to the following minimization problems which is related to the inside curly bracket in (3).

$$\begin{aligned}
\min_{\tilde{\boldsymbol{\vartheta}}} \quad & \left\|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right\|_2^2 \\
\text{s.t.} \quad & \langle \boldsymbol{\Phi}^j_{\tilde{\boldsymbol{\vartheta}}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle = 0.
\end{aligned} \tag{4}$$

The Lagrangian function of Problem (4) is

$$L\left(\tilde{\boldsymbol{\vartheta}}, v\right) = \left\|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right\|_2^2 + v\left(\langle \boldsymbol{\Phi}^j_{\tilde{\boldsymbol{\vartheta}}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle\right). \tag{5}$$

Correspondingly, we have the Lagrange dual function as

$$g(v) = \inf_{\tilde{\boldsymbol{\vartheta}}} L\left(\tilde{\boldsymbol{\vartheta}}, v\right) = \inf_{\tilde{\boldsymbol{\vartheta}}} \left\|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right\|_2^2 + v\left(\langle \boldsymbol{\Phi}^j_{\tilde{\boldsymbol{\vartheta}}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle\right). \tag{6}$$

To derive $g(v)$, we first derive $\frac{\partial L\left(\tilde{\boldsymbol{\vartheta}}, v\right)}{\partial \tilde{\boldsymbol{\vartheta}}} = 0$ as follows.

$$\begin{aligned}
\frac{\partial L\left(\tilde{\boldsymbol{\vartheta}}, v\right)}{\partial \tilde{\boldsymbol{\vartheta}}} &= -2\left(\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right) + v \cdot \frac{\partial \langle \boldsymbol{\Phi}^j_{\tilde{\boldsymbol{\vartheta}}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle}{\partial \tilde{\boldsymbol{\vartheta}}} \\
&= -2\left(\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\right) + v \cdot \boldsymbol{\phi}^j \odot \mathrm{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right) = 0.
\end{aligned} \tag{7}$$

Here, $\odot$ denotes Hadamard product. Then, we obtain

$$\tilde{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta} - \frac{v}{2}\boldsymbol{\phi}^j \odot \mathrm{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right), \tag{8}$$

where $\mathbb{1}_n \in \mathbb{R}^n$, $\otimes$ is Kronecker product, and vec means the vectorization of a matrix. Plugging (8) into (6) gives $g(v)$ as

$$\left\|\frac{v}{2}\boldsymbol{\phi}^j \odot \mathrm{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2 + v\left(\langle \tilde{\boldsymbol{\Phi}}^j_{\boldsymbol{\vartheta}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle\right), \tag{9}$$

where $\tilde{\boldsymbol{\Phi}}^j_{\boldsymbol{\vartheta}} = \left[\left(\boldsymbol{\theta}_i - \frac{v}{2}\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\left(\boldsymbol{x}^j_i - \hat{\boldsymbol{x}}^j_i\right)\right)^\top \boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_{i=1}^p \in \mathbb{R}^p$. Then, we derive the dual problem $\max_v g(v)$ of (4) as

$$\max_v \left\|\frac{v}{2}\boldsymbol{\phi}^j \odot \mathrm{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2 + v\left(\langle \tilde{\boldsymbol{\Phi}}^j_{\boldsymbol{\vartheta}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle\right). \tag{10}$$

Based on (10), we further compute $\frac{\partial g(v)}{\partial v}$ as follows.

$$\begin{aligned}
\frac{\partial g(v)}{\partial v} =& \frac{v}{2}\left\|\boldsymbol{\phi}^j \odot \mathrm{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2 \\
&+ \langle \tilde{\boldsymbol{\Phi}}^j_{\boldsymbol{\vartheta}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle + v\left(\langle \dot{\boldsymbol{\Phi}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle\right),
\end{aligned} \tag{11}$$

where $\dot{\boldsymbol{\Phi}}^j = \left[-\frac{1}{2}\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)^\top \left(\boldsymbol{x}^j_i - \hat{\boldsymbol{x}}^j_i\right)\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_{i=1}^p$, which belongs to $\mathbb{R}^p$.

To further derive (11), we rewrite $\tilde{\boldsymbol{\Phi}}^j_{\boldsymbol{\vartheta}}$ in (11) as

$$\begin{aligned}
&\left[\boldsymbol{\theta}_i{}^\top \boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right) - \frac{v}{2}\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)^\top \left(\boldsymbol{x}^j_i - \hat{\boldsymbol{x}}^j_i\right)\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_i \\
&= \left[\boldsymbol{\theta}_i{}^\top \boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_i - \frac{v}{2}\left[\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)^\top \left(\boldsymbol{x}^j_i - \hat{\boldsymbol{x}}^j_i\right)\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_i \\
&= \boldsymbol{\Phi}^j_{\boldsymbol{\vartheta}} + v\dot{\boldsymbol{\Phi}}^j.
\end{aligned} \tag{12}$$

Plugging (12) into (11), we can derive $\frac{\partial g(v)}{\partial v} = 0$ as

$$\frac{v}{2}\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2 + \left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j + v\dot{\boldsymbol{\Phi}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle + v\left\langle \dot{\boldsymbol{\Phi}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle = \frac{v}{2}\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2 + \left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle + 2v\left\langle \dot{\boldsymbol{\Phi}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle = 0.$$

Then, we can derive v as the following equation.

$$v = \frac{-2\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2 + 4\left\langle \dot{\boldsymbol{\Phi}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle}. \tag{13}$$

Plugging (13) into (8), we get the form of $\tilde{\boldsymbol{\vartheta}}$ as follows.

$$\begin{aligned}\tilde{\boldsymbol{\vartheta}} &= \boldsymbol{\vartheta} - \left\{ \frac{\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)}{2} \cdot \frac{2\left(-\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle\right)}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2 + 4\left\langle \dot{\boldsymbol{\Phi}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle} \right\} \\ &= \boldsymbol{\vartheta} - \frac{\left(-\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle\right)\left(\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right)}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2 + 4\left\langle \dot{\boldsymbol{\Phi}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle}.\end{aligned} \tag{14}$$

To simplify (14), we rewrite $\dot{\boldsymbol{\Phi}}^j \in \mathbb{R}^p$ in the following way.

$$\begin{aligned}\dot{\boldsymbol{\Phi}}^j &= \left[-\frac{1}{2}\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)^\top \left(\boldsymbol{x}_i^j - \hat{\boldsymbol{x}}_i^j\right)\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_{i=1}^p \\ &= -\frac{1}{2}\left[\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)^\top \boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_{i=1}^p \odot \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right).\end{aligned}$$

Then, we can derive $4\left\langle \dot{\boldsymbol{\Phi}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle$ as follows.

$$\begin{aligned}&4\left\langle \dot{\boldsymbol{\Phi}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle \\ &= -2\left\langle \left[\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)^\top \boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_i \odot \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right), \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle \\ &= -2\left\langle \left[\boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)^\top \boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_i, \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right) \odot \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right) \right\rangle \\ &= -2\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2.\end{aligned} \tag{15}$$

Plugging (15) into (14), we get $\tilde{\boldsymbol{\vartheta}}$ denoted by $\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)$:

$$\begin{aligned}\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j) &:= \underset{\left\langle \boldsymbol{\Phi}_{\tilde{\boldsymbol{\vartheta}}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle = 0}{\arg\min} \|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\|_2^2 \\ &= \boldsymbol{\vartheta} - \frac{\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle\left(\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right)}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2}.\end{aligned} \tag{16}$$

Then, we can obtain a solution to the minimization problem inside the curly brackets in (3) as follows.

$$\underset{\substack{\left\langle \boldsymbol{\Phi}_{\tilde{\boldsymbol{\vartheta}}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle = 0 \\ \|\tilde{\boldsymbol{\vartheta}}\|_2 = 1}}{\arg\min} \|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\|_2^2 = \frac{\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)}{\|\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\|_2}. \tag{17}$$

It can be easy to further prove (17) by contradiction.

Combining (2) and (17), we have the equation:

$$\min_{\substack{\tilde{\boldsymbol{\vartheta}} \in \operatorname{int}(\mathbb{C}) \\ \|\tilde{\boldsymbol{\vartheta}}\|_2 = 1}} \|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}\|_2^2 = \min_{\substack{\boldsymbol{x}^j \in \mathcal{X}^j \\ j \in [N]}} \left\|\boldsymbol{\vartheta} - \frac{\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)}{\|\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\|_2}\right\|_2^2.$$

Since we can easily derive that $\left\langle \boldsymbol{\vartheta}, \omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j) \right\rangle \geqslant 0$, i.e., $a\left(\boldsymbol{\vartheta}, \omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\right) \leqslant \frac{\pi}{2}$, and $\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j) \perp \left(\boldsymbol{\vartheta} - \omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\right)$, we then derive the following equation.

$$\begin{aligned}&\arg\min_{\boldsymbol{x}^j} \left\|\boldsymbol{\vartheta} - \frac{\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)}{\|\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\|_2}\right\|_2^2 = \arg\min_{\boldsymbol{x}^j} a\left(\boldsymbol{\vartheta}, \frac{\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)}{\|\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\|_2}\right) \\ &= \arg\min_{\boldsymbol{x}^j} a\left(\boldsymbol{\vartheta}, \omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\right) = \arg\min_{\boldsymbol{x}^j} \sin\left(a\left(\boldsymbol{\vartheta}, \omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\right)\right) \\ &= \arg\min_{\boldsymbol{x}^j} \left\|\boldsymbol{\vartheta} - \omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\right\|_2.\end{aligned} \tag{18}$$

Combining (17) and (18), we further derive that

$$\begin{aligned}&\arg\min_{\boldsymbol{x}^j} \left\|\boldsymbol{\vartheta} - \frac{\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)}{\|\omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\|_2}\right\|_2^2 = \arg\min_{\boldsymbol{x}^j} \left\|\boldsymbol{\vartheta} - \omega(\boldsymbol{\vartheta}, \boldsymbol{x}^j)\right\|_2 \\ &= \arg\min_{\boldsymbol{x}^j} \left\| \frac{\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle\left(\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right)}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2^2} \right\|_2 \\ &= \arg\min_{\boldsymbol{x}^j} \frac{\left|\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle\right|}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2}.\end{aligned}$$

Therefore, we can further express (2) as follows.

$$\begin{aligned}&\underset{\|\boldsymbol{\vartheta}\|_2 = 1}{\arg\max} \min_{\substack{\tilde{\boldsymbol{\vartheta}} \in \operatorname{int}(\mathbb{C}) \\ \|\tilde{\boldsymbol{\vartheta}}\|_2 = 1}} a\left(\boldsymbol{\vartheta}, \tilde{\boldsymbol{\vartheta}}\right) \\ &= \underset{\|\boldsymbol{\vartheta}\|_2 = 1}{\arg\max} \min_{\substack{\boldsymbol{x}^j \in \mathcal{X}^j \\ j \in [N]}} \frac{\left|\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle\right|}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2}.\end{aligned} \tag{19}$$

Next, by applying an epigraph reformulation for (19), we obtain the following equivalent optimization problem:

$$\begin{aligned}&\underset{\|\boldsymbol{\vartheta}\|_2 = 1}{\arg\max} \quad r \\ &\text{s.t.} \quad r = \min_{\substack{\boldsymbol{x}^j \in \mathcal{X}^j \\ j \in [N]}} \frac{\left|\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle\right|}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2}, \\ &\qquad \forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N].\end{aligned} \tag{20}$$

We rewrite the constraints in (20) as the following form.

$$\begin{aligned}&r \leqslant \frac{\left|\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle\right|}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2}, \forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N], \\ &\Rightarrow -\left|\left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle\right| + r\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2 \leqslant 0 \\ &\Rightarrow \left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle + r\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2 \leqslant 0.\end{aligned}$$

Based on the above form of the constraints, we can derive the equivalent optimization form for (19) as follows.

$$
\begin{aligned}
\underset{\boldsymbol{\vartheta}, r}{\arg\max} \quad & r \\
\text{s.t.} \quad & \langle \boldsymbol{\Phi}^j_{\boldsymbol{\vartheta}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle + r \left\| \boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right) \right\|_2 \leqslant 0, \\
& \forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N], \\
& \|\boldsymbol{\vartheta}\|_2 = 1.
\end{aligned} \tag{21}
$$

Next, we remove the constraint $\|\boldsymbol{\vartheta}\|_2 = 1$ in (21) and reformulate it as a convex optimization problem.

We first illustrate the assumption of a nonempty interior of $\mathbb{C}$ indicates that the optimal r in problem (21) is positive. To clarify this point, under the assumption, there exists some $\boldsymbol{\vartheta}$ such that $\langle \boldsymbol{\Phi}^j_{\boldsymbol{\vartheta}}, \hat{\boldsymbol{x}}^j - \boldsymbol{x}^j \rangle > 0$, , $\forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N]$. Hence, the following expression

$$
r = \frac{\min_{\boldsymbol{x}^j \in \mathcal{X}^j, j \in [N]} \langle \boldsymbol{\Phi}^j_{\boldsymbol{\vartheta}}, \hat{\boldsymbol{x}}^j - \boldsymbol{x}^j \rangle}{\max_{\boldsymbol{x}^j \in \mathcal{X}^j, j \in [N]} \left\| \boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right) \right\|_2} > 0
$$

is a feasible solution to (21), confirming that the optimal $r > 0$. Given $\|\boldsymbol{\vartheta}\|_2 = 1$, problem (21) can be further reformulated as the following minimization problem.

$$
\begin{aligned}
\min_{\boldsymbol{\vartheta}, r} \quad & \|\boldsymbol{\vartheta}\|_2 / r \\
\text{s.t.} \quad & \langle \boldsymbol{\Phi}^j_{\boldsymbol{\vartheta}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle + r \left\| \boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right) \right\|_2 \leqslant 0, \\
& \forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N], \\
& \|\boldsymbol{\vartheta}\|_2 = 1.
\end{aligned}
$$

Let $\tilde{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta}/r$, we can rewrite the above problem as

$$
\begin{aligned}
\min_{\tilde{\boldsymbol{\vartheta}}, r} \quad & \|\tilde{\boldsymbol{\vartheta}}\|_2 \\
\text{s.t.} \quad & \langle \boldsymbol{\Phi}^j_{\tilde{\boldsymbol{\vartheta}}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle + r \left\| \boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right) \right\|_2 \leqslant 0, \\
& \forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N], \\
& \|\tilde{\boldsymbol{\vartheta}}\|_2 = 1/r.
\end{aligned}
$$

Since r only appears in the constraint in above problem. Let $r = 1/\|\tilde{\boldsymbol{\vartheta}}\|_2$ for any $\tilde{\boldsymbol{\vartheta}}$, it will not affect the optimal value of $\tilde{\boldsymbol{\vartheta}}$. So we can simply drop the constraint $\|\tilde{\boldsymbol{\vartheta}}\|_2 = 1/r$. □

2 Robustness of Incenter to Perturbations

We provide additional details supporting the robustness interpretation of the incenter in Section 4.2. We formalize this robustness notion as follows. First, recall the definition of the set of consistent parameter vectors

$$
\mathbb{C} := \left\{ \boldsymbol{\vartheta} : \langle \boldsymbol{\Phi}^j_{\boldsymbol{\vartheta}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle \leqslant 0, \forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N] \right\}.
$$

Here, we can rewrite $\langle \boldsymbol{\Phi}^j_{\boldsymbol{\vartheta}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle$ as follows.

$$
\begin{aligned}
\langle \boldsymbol{\Phi}^j_{\boldsymbol{\vartheta}}, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle &= \langle \left[\boldsymbol{\theta}_i^\top \boldsymbol{\phi}_i\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right]_{i=1}^p, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle \\
&= \langle \boldsymbol{\vartheta}, \boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right) \rangle,
\end{aligned}
$$

Thus, the set $\mathbb{C}$ can be rewritten as the following form.

$$
\mathbb{C} := \left\{ \boldsymbol{\vartheta} : \langle \boldsymbol{\vartheta}, \boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right) \rangle \leqslant 0, \forall \boldsymbol{x}^j, \forall j \right\}.
$$

Next, we define vectors $\Delta\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j, \boldsymbol{x}^j\right)$ by normalizing the vectors $\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)$ within the set $\mathbb{C}$:

$$
\Delta\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j, \boldsymbol{x}^j\right) = \begin{cases} \frac{\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)}{\left\|\boldsymbol{\phi}^j \odot \operatorname{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)\right\|_2}, & \text{if } \boldsymbol{x}^j \neq \hat{\boldsymbol{x}}^j; \\ \mathbf{0}, & \text{Otherwise.} \end{cases}
$$

We notice that this normalization does not change the set $\mathbb{C}$, so we can further rewrite $\mathbb{C}$ using $\Delta\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j, \boldsymbol{x}^j\right)$ as

$$
\mathbb{C} := \left\{ \boldsymbol{\vartheta} : \langle \boldsymbol{\vartheta}, \Delta\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j, \boldsymbol{x}^j\right) \rangle \leqslant 0, \forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N] \right\}.
$$

It can be observed that the vectors $\Delta\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j, \boldsymbol{x}^j\right)$ are determined by the dataset $\widehat{\mathcal{D}} = \left\{\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right\}_{j=1}^N$. Hence, the vector $\boldsymbol{\vartheta} \in \mathbb{C}$ which is most robust to perturbations in the data (i.e., perturbation of $\Delta\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j, \boldsymbol{x}^j\right)$) can be found by solving

$$
\begin{aligned}
\underset{\substack{\|\boldsymbol{\vartheta}\|_2 = 1 \\ \boldsymbol{\vartheta} \in \mathbb{C}}}{\arg\max} \quad \min_{\substack{\boldsymbol{\gamma} \in \mathbb{R}^{np} \\ \left(\hat{\boldsymbol{x}}^k, \hat{\boldsymbol{s}}^k\right) \in \left\{\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right)\right\}_{j=1}^N \\ \boldsymbol{x}^j \in \mathcal{X}^j}} \quad & \|\boldsymbol{\gamma}\|_2^2 \\
\text{s.t.} \quad & \langle \boldsymbol{\vartheta}, \Delta\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j, \boldsymbol{x}^j\right) + \boldsymbol{\gamma} \rangle = 0.
\end{aligned} \tag{22}
$$

In problem (22), the outer maximization player optimizes $\boldsymbol{\vartheta} \in \mathbb{C}$ to require the largest perturbation $\boldsymbol{\gamma}$ for it to lie on the perturbed facet of $\mathbb{C}$. The inner minimization player seeks the most vulnerable facet of $\mathbb{C}$, which requires the smallest perturbation vector $\boldsymbol{\gamma}$ to satisfy $\langle \boldsymbol{\vartheta}, \Delta\left(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j, \boldsymbol{x}^j\right) + \boldsymbol{\gamma} \rangle = 0$. It can be easily illustrated that Problem (22) is equivalent to the incenter $\boldsymbol{\vartheta}^{\text{in}}$ reformulation in Theorem 1.

3 Convergence of Algorithm 1

We can derive the convergence rate of the mirror descent algorithm for solving problem (8) as follows. We assume that Bregman divergence function $\mathcal{B}_\omega(\boldsymbol{\vartheta}, \boldsymbol{\nu}) \leqslant R$ for some $R > 0$, and $\mathbb{E}\|g_t(\boldsymbol{\vartheta})\|_*^2 \leqslant G$ for all $\boldsymbol{\vartheta}, \boldsymbol{\nu}$. Let $f(\boldsymbol{\vartheta})$ be the objective of problem (8). With learning rate $\eta_t = \frac{c}{\sqrt{t}}$ for some constant $c > 0$, the learning algorithm guarantees that

$$
\mathbb{E}\left[f\left(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\vartheta}_t\right)\right] - \min_{\boldsymbol{\vartheta}} f(\boldsymbol{\vartheta}) \leqslant \frac{1}{\sqrt{T}}\left(\frac{R}{c} + cG\right). \tag{23}
$$

Proof. Let $\boldsymbol{\vartheta}^* \in \arg\min_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} f(\boldsymbol{\vartheta})$. By the definition of the subgradient, we have that

$$
\mathbb{E}\left[f(\boldsymbol{\vartheta}_t) - f(\boldsymbol{\vartheta}^*)\right] \leqslant \mathbb{E}\left[\langle g(\boldsymbol{\vartheta}_t), \boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}^* \rangle\right] \tag{24}
$$

Next, we apply the standard inequality for analyzing the mirror descent algorithm:

$$
\begin{aligned}
\langle g(\boldsymbol{\vartheta}_t), \boldsymbol{\vartheta}_t - \boldsymbol{\vartheta}^* \rangle \leqslant\ & \frac{1}{\eta_t}\left(\mathcal{B}_\omega(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta}_t) - \mathcal{B}_\omega(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta}_{t+1})\right) \\
& + \frac{\eta_t}{2}\|g(\boldsymbol{\vartheta}_t)\|_*^2.
\end{aligned} \tag{25}
$$

Then, we can further derive the expectation form of (25) as

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\langle g\left(\boldsymbol{\vartheta}_t\right),\boldsymbol{\vartheta}_t-\boldsymbol{\vartheta}^*\rangle\right]\\
&\leqslant \mathbb{E}\left[\sum_{t=1}^{T}\frac{1}{\eta_t}\left(\mathcal{B}_\omega\left(\boldsymbol{\vartheta}^*,\boldsymbol{\vartheta}_t\right)-\mathcal{B}_\omega\left(\boldsymbol{\vartheta}^*,\boldsymbol{\vartheta}_{t+1}\right)\right)+\frac{1}{2}\sum_{t=1}^{T}\eta_t\|g\|_*^2\right]\\
&\leqslant \frac{R}{\eta_1}+R\sum_{t=2}^{T}\left(\frac{1}{\eta_t}-\frac{1}{\eta_{t-1}}\right)+\frac{G}{2}\sum_{t=1}^{T}\eta_t\\
&=\frac{R}{\eta_T}+\frac{G}{2}\sum_{t=1}^{T}\eta_t\leqslant\left(\frac{R}{c}+cG\right)\sqrt{T}.
\end{aligned}
\tag{26}
$$

Finally, we use Jensen's inequality to further derive that

$$
\begin{aligned}
\mathbb{E}\left[f\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\vartheta}_t\right)-f\left(\boldsymbol{\vartheta}^*\right)\right]&\leqslant\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\left(f\left(\boldsymbol{\vartheta}_t\right)-f\left(\boldsymbol{\vartheta}^*\right)\right)\right]\\
&\leqslant\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\langle g\left(\boldsymbol{\vartheta}_t\right),\boldsymbol{\vartheta}_t-\boldsymbol{\vartheta}^*\rangle\right]\\
&\leqslant\frac{1}{\sqrt{T}}\left(\frac{R}{c}+cG\right).
\end{aligned}
\tag{27}
$$

□

4 Proof of Lemma 1

Proof. In the two-firm Bertrand-Nash competition with a linear demand function, the corresponding Jacobian matrix

$$\nabla\boldsymbol{F}\left(\boldsymbol{x}\right)=-\begin{bmatrix}2\theta_{11} & \theta_{12}\\ \theta_{21} & 2\theta_{22}\end{bmatrix}$$

is symmetric due to the constraint that $\theta_{21}=\theta_{12}=1$.

Let variables $\beth=-\begin{bmatrix}\theta_{11} & \theta_{21}/2\\ \theta_{12}/2 & \theta_{22}\end{bmatrix}$, $\tilde{\beth}=\begin{bmatrix}\theta_{13} & \theta_{23}\\ \theta_{14} & \theta_{24}\end{bmatrix}$, then problem (8) in Section 4.3 can be reformulated as

$$
\begin{aligned}
\min_{\beth,\tilde{\beth}}\quad & \frac{1}{N}\sum_{j=1}^{N}\ell_{\beth,\tilde{\beth}}\left(\hat{\boldsymbol{x}}^j,\hat{\boldsymbol{s}}^j\right)+\alpha\mathcal{R}\left(\beth,\tilde{\beth}\right)\\
\text{s.t.}\quad & \beth_{ii}\geqslant 0,\quad \beth_{i,-i}=-\frac{1}{2},\forall i=1,2,\\
& \beth\succeq 0,\quad \tilde{\beth}\succeq 0.
\end{aligned}
\tag{28}
$$

Here, $\beth_{ii}\geqslant 0$ for $i=1,2$ imply that θ_{11} and θ_{22} are non-positive, ensuring the concavity of each utility function U_i. The equality constraints $\beth_{i,-i}=-\frac{1}{2}$ for $i=1,2$ indicate that $\theta_{21}=\theta_{12}=1$, ensuring symmetry in the cross-effects between the firms. The conditions $\beth\succeq 0$ and $\tilde{\beth}\succeq 0$ ensure that both $\beth$ and $\tilde{\beth}$ are positive semidefinite matrices.

The loss function $\ell_{\beth,\tilde{\beth}}\left(\hat{\boldsymbol{x}}^j,\hat{\boldsymbol{s}}^j\right)$ defined in Definition 3 is $\max_{\boldsymbol{x}^j\in\mathcal{X}^j}\left\{-\mathrm{Tr}\left(\Psi^j\beth\right)+\mathrm{Tr}\left(\tilde{\Psi}^j\tilde{\beth}\right)+\left\|\boldsymbol{\phi}^j\odot\mathrm{vec}\left(\mathbb{1}_n\otimes\left(\boldsymbol{x}^j-\hat{\boldsymbol{x}}^j\right)\right)\right\|_2\right\}$, where the vector $\boldsymbol{v}^j=\boldsymbol{\phi}^j\odot\mathrm{vec}\left(\mathbb{1}_n\otimes\left(\boldsymbol{x}^j-\hat{\boldsymbol{x}}^j\right)\right)$, the matrices $\Psi^j=\begin{bmatrix}v_1^j & 2v_2^j\\ 2v_5^j & v_6^j\end{bmatrix}$ and $\tilde{\Psi}^j=\begin{bmatrix}v_3^j & v_4^j\\ v_7^j & v_8^j\end{bmatrix}$.

Since the constraints $\beth_{ii}\geqslant 0$ for all $i=1,2$ can be rewritten as $\mathrm{Tr}\left(\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right)\beth\right)\geqslant 0$, for $i=1,2$. Constraints $\beth_{i,-i}=-\frac{1}{2}$ for all $i=1,2$ are equivalent to $\mathrm{Tr}\left(A\beth\right)=-1$, we can rewrite (28) as follows.

$$
\begin{aligned}
\min_{\beth,\tilde{\beth}}\quad & \frac{1}{N}\sum_{j=1}^{N}\ell_{\beth,\tilde{\beth}}\left(\hat{\boldsymbol{x}}^j,\hat{\boldsymbol{s}}^j\right)+\alpha\mathcal{R}\left(\beth,\tilde{\beth}\right)\\
\text{s.t.}\quad & \mathrm{Tr}\left(\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right)\beth\right)\geqslant 0,\forall i=1,2,\quad \mathrm{Tr}\left(A\beth\right)=-1,\\
& \beth\succeq 0,\quad \tilde{\beth}\succeq 0.
\end{aligned}
\tag{29}
$$

In (29), matrix $A=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$ is an anti-diagonal matrix. The vector $\boldsymbol{e}_i$ has a one in the i-th position and zeros in all other positions. The loss function $\ell_{\beth,\tilde{\beth}}\left(\hat{\boldsymbol{x}}^j,\hat{\boldsymbol{s}}^j\right)$ is expressed as: $\max_{\boldsymbol{x}^j\in\mathcal{X}^j}\left\{-\mathrm{Tr}\left(\Psi_s^j\beth\right)+\mathrm{Tr}\left(\tilde{\Psi}_s^j\tilde{\beth}\right)+\left\|\boldsymbol{\phi}^j\odot\mathrm{vec}\left(\mathbb{1}_n\otimes\left(\boldsymbol{x}^j-\hat{\boldsymbol{x}}^j\right)\right)\right\|_2\right\}$, where $\Psi_s^j=\begin{bmatrix}v_1^j & v_2^j+v_5^j\\ v_2^j+v_5^j & v_6^j\end{bmatrix}$, $\tilde{\Psi}_s^j=\begin{bmatrix}v_3^j & \frac{v_4^j+v_7^j}{2}\\ \frac{v_4^j+v_7^j}{2} & v_8^j\end{bmatrix}$.

□

5 Proof of Proposition 1

Proof. First, we rewrite the Lagrangian of the primal problem (14) in Lemma 1 of Section 4.5 as follows:

$$
\begin{aligned}
\mathcal{L}\left(\beth,\tilde{\beth},\boldsymbol{\lambda},\nu,\Xi,\tilde{\Xi}\right)=&\frac{1}{N}\sum_{j=1}^{N}\ell_{\beth,\tilde{\beth}}\left(\hat{\boldsymbol{x}}^j,\hat{\boldsymbol{s}}^j\right)+\alpha\mathcal{R}\left(\beth,\tilde{\beth}\right)-\\
&\sum_{i=1}^{2}\lambda_i\mathrm{Tr}\left(\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right)\beth\right)+\nu\left(\mathrm{Tr}\left(A\beth\right)+1\right)-\mathrm{Tr}(\Xi\beth)-\mathrm{Tr}\left(\tilde{\Xi}\tilde{\beth}\right).
\end{aligned}
\tag{30}
$$

Here, $\boldsymbol{\lambda}$, ν, Ξ, and $\tilde{\Xi}$ represent the dual variables. The terms $\mathrm{Tr}\left(\Xi\beth\right)$ and $\mathrm{Tr}\left(\tilde{\Xi}\tilde{\beth}\right)$ enforce the constraints $\beth\succeq 0$ and $\tilde{\beth}\succeq 0$, respectively. The regularization term is defined by the Frobenius norm, i.e., $\mathcal{R}\left(\beth,\tilde{\beth}\right)=\frac{1}{2}\left(\|\beth\|_F^2+\|\tilde{\beth}\|_F^2\right)$.

For ease of presentation, we denote $h\left(\boldsymbol{x}^j,\hat{\boldsymbol{x}}^j,\hat{\boldsymbol{s}}^j\right):=\left\|\boldsymbol{\phi}^j\odot\mathrm{vec}\left(\mathbb{1}_n\otimes\left(\boldsymbol{x}^j-\hat{\boldsymbol{x}}^j\right)\right)\right\|_2$, and the maximizer $\bar{\boldsymbol{x}}^j=\arg\max_{\boldsymbol{x}^j\in\mathcal{X}^j}\left\{-\mathrm{Tr}\left(\Psi_s^j\beth\right)+\mathrm{Tr}\left(\tilde{\Psi}_s^j\tilde{\beth}\right)+h\left(\boldsymbol{x}^j,\hat{\boldsymbol{x}}^j,\hat{\boldsymbol{s}}^j\right)\right\}$. Let the symmetric matrices $\Upsilon=\frac{1}{N}\sum_{j=1}^{N}\Psi_s^j\left(\bar{\boldsymbol{x}}^j\right)$ and $\tilde{\Upsilon}=\frac{1}{N}\sum_{j=1}^{N}\tilde{\Psi}_s^j\left(\bar{\boldsymbol{x}}^j\right)$. We can then rewrite the Lagrangian $\mathcal{L}\left(\beth,\tilde{\beth},\boldsymbol{\lambda},\nu,\Xi,\tilde{\Xi}\right)$, as given in equation (30), as follows

$$
\begin{aligned}
\mathcal{L}=&\frac{\alpha}{2}\|\beth\|_F^2-\langle\Upsilon+\sum_{i=1}^{2}\lambda_i\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right)+\Xi-\nu A,\beth\rangle\\
&+\frac{\alpha}{2}\|\tilde{\beth}\|_F^2+\langle\tilde{\Upsilon}-\tilde{\Xi},\tilde{\beth}\rangle+\frac{1}{N}\sum_{j=1}^{N}h\left(\bar{\boldsymbol{x}}^j,\hat{\boldsymbol{x}}^j,\hat{\boldsymbol{s}}^j\right)+\nu.
\end{aligned}
\tag{31}
$$

Based on (31), we derive the first order conditions as

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \beth} &= \alpha \beth - \Upsilon - \sum_{i=1}^{2} \lambda_i \left(\boldsymbol{e}_i \boldsymbol{e}_i^\top\right) - \Xi + \nu A = \mathbf{0}. \\ \frac{\partial \mathcal{L}}{\partial \tilde{\beth}} &= \alpha \tilde{\beth} + \tilde{\Upsilon} - \tilde{\Xi} = \mathbf{0}.\end{aligned} \tag{32}$$

Next, we derive the Lagrange dual problem for the primal problem (14) in Lemma 1 of Section 4.5. We begin by writing the corresponding Lagrange dual function as follows.

$$\mathcal{G}\left(\boldsymbol{\lambda}, \nu, \Xi, \tilde{\Xi}\right) = \inf_{\beth, \tilde{\beth}} \mathcal{L}\left(\beth, \tilde{\beth}, \boldsymbol{\lambda}, \nu, \Xi, \tilde{\Xi}\right).$$

By substituting (32) into (31), we can derive $\mathcal{G}\left(\boldsymbol{\lambda}, \nu, \Xi, \tilde{\Xi}\right)$ as follows. First, we express $\frac{\alpha}{2}\|\beth\|_F^2$ and $\frac{\alpha}{2}\|\tilde{\beth}\|_F^2$ as:

$$\begin{aligned}\frac{\alpha}{2}\|\beth\|_F^2 &= \frac{\alpha}{2}\mathrm{Tr}\left(\beth^\top \beth\right) = \frac{1}{2\alpha}\left\|\Upsilon + \sum_{i=1}^{2}\lambda_i\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right) + \Xi - \nu A\right\|_F^2. \\ \frac{\alpha}{2}\|\tilde{\beth}\|_F^2 &= \frac{\alpha}{2}\mathrm{Tr}\left(\tilde{\beth}^\top \tilde{\beth}\right) = \frac{1}{2\alpha}\left\| - \tilde{\Upsilon} + \tilde{\Xi}\right\|_F^2.\end{aligned}$$

Based on the above equations of $\frac{\alpha}{2}\|\beth\|_F^2$ and $\frac{\alpha}{2}\|\tilde{\beth}\|_F^2$, we can further derive the corresponding dual problem as

$$\begin{aligned}\max_{\Xi, \tilde{\Xi}, \boldsymbol{\lambda}, \nu} \quad & \mathcal{G}\left(\boldsymbol{\lambda}, \nu, \Xi, \tilde{\Xi}\right) \\ \text{s.t.} \quad & \lambda_i \geqslant 0, \forall i = 1, 2, \\ & \Xi \succeq 0, \tilde{\Xi} \succeq 0,\end{aligned} \tag{33}$$

where $\mathcal{G}\left(\boldsymbol{\lambda}, \nu, \Xi, \tilde{\Xi}\right)$ is the following equation:

$$\begin{aligned}\mathcal{G}\left(\boldsymbol{\lambda}, \nu, \Xi, \tilde{\Xi}\right) &= -\frac{1}{2\alpha}\left\|\Upsilon + \sum_{i=1}^{2}\lambda_i\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right) + \Xi - \nu A\right\|_F^2 \\ &\quad - \frac{1}{2\alpha}\left\| - \tilde{\Upsilon} + \tilde{\Xi}\right\|_F^2 + \frac{1}{N}\sum_{j=1}^{N} h\left(\bar{\boldsymbol{x}}^j, \hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right) + \nu.\end{aligned}$$

To employ the primal-dual interior-point algorithm, we need to derive the corresponding perturbed KKT conditions. The basic idea is to utilize a logarithmic barrier function to penalize the constraints in both the primal and dual problems. Specifically, using the logarithmic barrier, we first approximate the primal problem (14) in Lemma 1 as follows.

$$\begin{aligned}\min_{\beth, \tilde{\beth}} \; & \frac{\alpha}{2}\|\beth\|_F^2 - \langle \Upsilon, \beth \rangle + \frac{\alpha}{2}\|\tilde{\beth}\|_F^2 + \langle \tilde{\Upsilon}, \tilde{\beth} \rangle + \frac{1}{N}\sum_{j=1}^{N} h\left(\bar{\boldsymbol{x}}^j, \hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right) \\ & - \frac{1}{\mu}\left(\log\det(\beth) + \log\det(\tilde{\beth}) + \sum_{i=1}^{2}\log\left(\mathrm{Tr}\left(\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right)\beth\right)\right)\right) \\ \text{s.t.} \quad & \mathrm{Tr}\left(A\beth\right) = -1.\end{aligned} \tag{34}$$

Here, μ represents the logarithmic barrier parameter. By employing the penalized barrier function, the inequality constraints of the original problem are seamlessly integrated into the objective function, as shown in (34).

155 136 Let $\mathcal{G} = \mathcal{G}\left(\boldsymbol{\lambda}, \nu, \Xi, \tilde{\Xi}\right)$ as defined in (33). By applying the logarithmic barrier, (33) can also be approximated as

$$\max_{\Xi, \tilde{\Xi}, \boldsymbol{\lambda}, \nu} \mathcal{G} + \frac{1}{\mu}\left(\sum_{i=1}^{2}\log\lambda_i + \log\left(\det\left(\Xi\right)\right) + \log\left(\det\left(\tilde{\Xi}\right)\right)\right). \tag{35}$$

Next, we derive the perturbed KKT conditions based on (34) and (35). We start by expressing the Lagrangian $\mathcal{L}_\mu\left(\beth, \tilde{\beth}, \nu\right)$ for the primal barrier problem (34) as follows.

$$\begin{aligned}\mathcal{L}_\mu\left(\beth, \tilde{\beth}, \nu\right) &= \frac{\alpha}{2}\|\beth\|_F^2 - \langle \Upsilon, \beth \rangle + \frac{\alpha}{2}\|\tilde{\beth}\|_F^2 + \langle \tilde{\Upsilon}, \tilde{\beth} \rangle \\ & - \frac{1}{\mu}\left(\log\det(\beth) + \log\det(\tilde{\beth}) + \sum_{i=1}^{n}\log\left(\mathrm{Tr}\left(\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right)\beth\right)\right)\right) \\ & + \nu\left(\mathrm{Tr}\left(A\beth\right) + 1\right) + \frac{1}{N}\sum_{j=1}^{N} h\left(\bar{\boldsymbol{x}}^j, \hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j\right).\end{aligned}$$

Setting $\frac{\partial \mathcal{L}_\mu(\beth, \tilde{\beth}, \nu)}{\partial \beth} = \mathbf{0}$, $\frac{\partial \mathcal{L}_\mu(\beth, \tilde{\beth}, \nu)}{\partial \tilde{\beth}} = \mathbf{0}$, we can obtain

$$\begin{aligned}& \alpha\beth - \Upsilon - \frac{1}{\mu}\left(\beth^{-1} + \sum_{i=1}^{n}\frac{\boldsymbol{e}_i\boldsymbol{e}_i^\top}{\mathrm{Tr}\left(\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right)\beth\right)}\right) + \nu A^\top = \mathbf{0}. \\ & \alpha\tilde{\beth} + \tilde{\Upsilon} - \frac{1}{\mu}\left(\tilde{\beth}^{-1}\right) = \mathbf{0}.\end{aligned} \tag{36}$$

Let $\mathcal{G}_\mu\left(\boldsymbol{\lambda}, \nu, \Xi, \tilde{\Xi}\right)$ represent the objective function in problem (35). By setting $\frac{\partial \mathcal{G}_\mu\left(\boldsymbol{\lambda}, \nu, \Xi, \tilde{\Xi}\right)}{\partial \Xi} = \mathbf{0}$, we then have

$$-\frac{1}{\alpha}\left(\Upsilon + \sum_{i=1}^{2}\lambda_i\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right) + \Xi - \nu A\right) + \frac{1}{\mu}\Xi^{-1} = \mathbf{0}. \tag{37}$$

Since the primal and dual variables that satisfy the KKT conditions are optimal solutions, they must satisfy all the equations in (32), (36) and (37). Accordingly, we can further derive the following three equations.

$$\beth = \frac{1}{\mu}\Xi^{-1}, \tilde{\beth} = \frac{1}{\mu}\tilde{\Xi}^{-1}, \lambda_i = \frac{1}{\mu\mathrm{Tr}\left(\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right)\beth\right)}, \forall i = 1, 2. \tag{38}$$

Hence, we derive the equations in Proposition 1, which stem from (32) and (38). Specifically, $\mathrm{Tr}\left(\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right)\beth^*\right) \geqslant 0$ for all $i = 1, 2$, $\mathrm{Tr}\left(A\beth^*\right) = -1$, and $\beth^*, \tilde{\beth}^* \succeq 0$ represent primal feasibility. Additionally, $\lambda_i^* \geqslant 0, \forall i = 1, 2$, and $\Xi^*, \tilde{\Xi}^* \succeq 0$ ensure dual feasibility. The equations in (32) and (38) correspond to stationarity and complementarity, respectively.

□

6 Newton Updates in Algorithm 2

In the primal-dual interior-point method, a sequence of iterates is generated that approximates the central path, ultimately converging to the optimal primal and dual solutions. In this context, the basic iterative step is derived using Newton's method. Ideally, each Newton step updates the points, progressively refining the solution until the optimal points for both the primal and dual problems are found. The detailed Newton updates are presented as follows.

For problem (14) in Lemma 1, points $\left(\beth, \tilde{\beth}, \Xi, \tilde{\Xi}, \boldsymbol{\lambda}, \nu\right)$ on the central path satisfy the nonlinear equations in Proposition 1. To find such points, the essence of the Newton step lies in linearizing these nonlinear equations along the central path. We begin by listing these equations as follows.

$$
\begin{gathered}
\operatorname{Tr}(A\beth) + 1 = 0. \\
\lambda_i \operatorname{Tr}\left(\left(\boldsymbol{e}_i \boldsymbol{e}_i^\top\right)\beth\right) - \frac{1}{\mu} = 0, \forall i = 1, 2. \\
\Xi\beth - \frac{1}{\mu}\boldsymbol{I} = \mathbf{0}, \\
\tilde{\Xi}\tilde{\beth} - \frac{1}{\mu}\boldsymbol{I} = \mathbf{0}. \\
\alpha\tilde{\beth} - \tilde{\Xi} + \frac{1}{N}\sum_{j=1}^{N} \tilde{\Psi}_s^j\left(\bar{\boldsymbol{x}}^j\right) = \mathbf{0}. \\
\alpha\beth - \sum_{i=1}^{2} \lambda_i\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right) - \Xi + \nu A = \frac{1}{N}\sum_{j=1}^{N}\Psi_s^j\left(\bar{\boldsymbol{x}}^j\right).
\end{gathered}
$$

To solve these equations, we seek to find a zero of the vector-valued function $\boldsymbol{F}\left(\beth, \tilde{\beth}, \Xi, \tilde{\Xi}, \boldsymbol{\lambda}, \nu\right)$, which is defined as

$$
\begin{pmatrix}
\operatorname{Tr}(A\beth) + 1 \\
\lambda_1 \operatorname{Tr}\left(\left(\boldsymbol{e}_1\boldsymbol{e}_1^\top\right)\beth\right) - \frac{1}{\mu} \\
\lambda_2 \operatorname{Tr}\left(\left(\boldsymbol{e}_2\boldsymbol{e}_2^\top\right)\beth\right) - \frac{1}{\mu} \\
\Xi\beth - \frac{1}{\mu}\boldsymbol{I} \\
\tilde{\Xi}\tilde{\beth} - \frac{1}{\mu}\boldsymbol{I} \\
\alpha\beth - \sum_{i=1}^{2}\lambda_i\left(\boldsymbol{e}_i\boldsymbol{e}_i^\top\right) - \Xi - \frac{1}{N}\sum_{j=1}^{N}\Psi_s^j\left(\bar{\boldsymbol{x}}^j\right) + \nu A \\
\alpha\tilde{\beth} - \tilde{\Xi} + \frac{1}{N}\sum_{j=1}^{N}\tilde{\Psi}_s^j\left(\bar{\boldsymbol{x}}^j\right)
\end{pmatrix}.
$$

However, solving this problem directly is challenging, so we employ Newton's method to first linearize it and then solve it approximately. Specifically, we aim to find its Jacobian $J\left(\beth, \tilde{\beth}, \Xi, \tilde{\Xi}, \boldsymbol{\lambda}, \nu\right)$, and then iteratively solve the following system for update direction $\left(\Delta\beth, \Delta\tilde{\beth}, \Delta\Xi, \Delta\tilde{\Xi}, \Delta\boldsymbol{\lambda}, \Delta\nu\right)$:

$$
J\left(\beth, \tilde{\beth}, \Xi, \tilde{\Xi}, \boldsymbol{\lambda}, \nu\right)
\begin{pmatrix}
\Delta\beth \\ \Delta\tilde{\beth} \\ \Delta\Xi \\ \Delta\tilde{\Xi} \\ \Delta\boldsymbol{\lambda} \\ \Delta\nu
\end{pmatrix}
= -\boldsymbol{F}\left(\beth, \tilde{\beth}, \Xi, \tilde{\Xi}, \boldsymbol{\lambda}, \nu\right).
$$

By finding the Jacobian, we derive the following system:

$$
\begin{pmatrix}
\langle A, \Delta\beth\rangle \\
\lambda_1\langle \boldsymbol{e}_1\boldsymbol{e}_1^\top, \Delta\beth\rangle + \operatorname{Tr}\left(\left(\boldsymbol{e}_1\boldsymbol{e}_1^\top\right)\beth\right)\Delta\lambda_1 \\
\lambda_2\langle \boldsymbol{e}_2\boldsymbol{e}_2^\top, \Delta\beth\rangle + \operatorname{Tr}\left(\left(\boldsymbol{e}_2\boldsymbol{e}_2^\top\right)\beth\right)\Delta\lambda_2 \\
\Xi\Delta\beth + \Delta\Xi\beth \\
\tilde{\Xi}\Delta\tilde{\beth} + \Delta\tilde{\Xi}\tilde{\beth} \\
\alpha\Delta\beth - \Delta\Xi - \left(\boldsymbol{e}_1\boldsymbol{e}_1^\top\right)\Delta\lambda_1 - \left(\boldsymbol{e}_2\boldsymbol{e}_2^\top\right)\Delta\lambda_2 + A\Delta\nu \\
\alpha\Delta\tilde{\beth} - \Delta\tilde{\Xi}
\end{pmatrix}
= -\boldsymbol{F}. \tag{39}
$$

Next, we solve the equation in (39) to derive the Newton update directions, i.e., $\left(\Delta\beth, \Delta\tilde{\beth}, \Delta\Xi, \Delta\tilde{\Xi}, \Delta\boldsymbol{\lambda}, \Delta\nu\right)$. These directions are subsequently utilized to iteratively refine the primal and dual variables during each step of Algorithm 2.

7 Multi-Player Bertrand-Nash Competition

We first explain the constraint $\theta_{12} = \theta_{21} = 1$ in Lemma 1 for the two-firm Bertrand-Nash competition. This constraint accounts for the flexibility of the utility parameters. To be more specific, let $\boldsymbol{\theta}_i^o$ be the unknown parameters for all $i = 1, 2$. In this two-firm competition, it is common for θ_{12}^o to differ from θ_{21}^o. However, as shown in the proof of Lemma 1, the matrix $\nabla \boldsymbol{F}$ needs to be symmetric to formulate the semidefinite programming problem. To ensure symmetry, we normalize each parameter vector $\boldsymbol{\theta}_i^o$ by dividing it by $\theta_{i,-i}^o$, where $\theta_{i,-i}^o = \theta_{1,2}^o$ if $i = 1$, and $\theta_{i,-i}^o = \theta_{2,1}^o$ if $i = 2$. This normalization operator does not change the corresponding equilibrium solutions. Consequently, the new parameters satisfy $\theta_{12} = \theta_{21} = 1$. Therefore, the parameters $\boldsymbol{\theta}_i$ in Section 4.5 can be interpreted as those derived from the original parameters $\boldsymbol{\theta}_i^o$, where $\theta_{1,2}^o \neq \theta_{2,1}^o$.

Next, we provide the proofs for the semidefinite programming formulation in three-firm Bertrand-Nash competitions. We first present their parameter structures. Specifically, the linear demand function U_i for $i = 1, 2, 3$ is defined as

$$
U_i\left(\boldsymbol{x}, \boldsymbol{s}; \boldsymbol{\theta}_i^o\right) = \theta_{i1}^o x_1 + \theta_{i2}^o x_2 + \theta_{i3}^o x_3 + \theta_{i4}^o s + \theta_{i5}^o. \tag{40}
$$

Here, $\theta_{i,i}^o \leqslant 0$, and $\boldsymbol{\theta}_{i,-i}^o \geqslant 0$, where, for instance, if $i = 1$, then the vector $\boldsymbol{\theta}_{i,-i}^o = \left(\theta_{1,2}^o, \theta_{1,3}^o\right)$.

Let each parameter vector $\boldsymbol{\theta}_i^o = \left(\theta_{i1}^o, \theta_{i2}^o, \theta_{i3}^o, \theta_{i4}^o, \theta_{i5}^o\right) \in \mathbb{R}^5$. We can derive its Jacobian matrix $\nabla F(\boldsymbol{x})$ as

$$
\nabla \boldsymbol{F}(\boldsymbol{x}) = -\begin{bmatrix}
2\theta_{11}^o & \theta_{12}^o & \theta_{13}^o \\
\theta_{21}^o & 2\theta_{22}^o & \theta_{23}^o \\
\theta_{31}^o & \theta_{32}^o & 2\theta_{33}^o
\end{bmatrix}.
$$

To ensure the symmetry of $\nabla \boldsymbol{F}(\boldsymbol{x})$, we construct each $\boldsymbol{\theta}_i$ by multiplying each row by a non-negative number a_1, a_2, a_3:

$$
\begin{bmatrix}
a_1 * 2\theta_{11}^o & a_1 * \theta_{12}^o & a_1 * \theta_{13}^o \\
a_2 * \theta_{21}^o & a_2 * 2\theta_{22}^o & a_2 * \theta_{23}^o \\
a_3 * \theta_{31}^o & a_3 * \theta_{32}^o & a_3 * 2\theta_{33}^o
\end{bmatrix}.
$$

If we require this matrix to be symmetric, we need to impose three constraints, one for each off-diagonal element. This results in the following system of three equations:

$$\begin{cases} a_1 * \theta_{12}^o & = a_2 * \theta_{21}^o; \\ a_1 * \theta_{13}^o & = a_3 * \theta_{31}^o; \\ a_2 * \theta_{23}^o & = a_3 * \theta_{32}^o. \end{cases} \tag{41}$$

In (41), there are three equations with three unknowns, i.e., a_1, a_2, and a_3. This system typically has a solution, except in degenerate cases such as when $\boldsymbol{\theta}_i^o = 0$. Once the values of a_1, a_2, and a_3 are determined, we can construct $\boldsymbol{\theta}_i$ for all $i = 1, 2, 3$, for example, $\theta_{12} = a_1 * \theta_{12}^o$. The resulting Jacobian $\nabla \boldsymbol{F}(\boldsymbol{x})$ will be symmetric. Then, we can formulate the estimation of all $\boldsymbol{\theta}_i$ as a similar SDP as in Lemma 1.

In a four-firm Bertrand competition, the Jacobian $\nabla \boldsymbol{F}(\boldsymbol{x})$ will be a 4×4 matrix. Enforcing symmetry requires six constraints, but only four coefficients, i.e., a_1, a_2, a_3, and a_4, are available. This issue also arises with more than four players. In such cases, an additional constraint must be imposed, requiring the matrix of parameters to be symmetric from the outset. This will restrict the search to symmetric matrices.

8 Our Methods for Cournot Competition

In the Cournot competition, each parameter vector $\boldsymbol{\theta}_i$ can be represented in the form $(a, b, d, c_i) \in \mathbb{R}^4$. Given that some elements are shared across each $\boldsymbol{\theta}_i$, we redefine the parameter vector $\boldsymbol{\vartheta}$ and the corresponding loss function as follows.

Let $\boldsymbol{\vartheta} = (a, b, d, c_1, c_2, \ldots, c_p) \in \mathbb{R}^{p+3}$ be the parameter vector to be estimated. In the definition of consistent parameter vectors $\mathbb{C}$ in Section 4.2, we can further express vectors $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j \in \mathbb{R}^p$ and $\boldsymbol{\phi}_i(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j) \in \mathbb{R}^{p+3}$ as follows.

$$\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j = \left(\boldsymbol{\theta}^\top \boldsymbol{\phi}_1(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j), \boldsymbol{\theta}^\top \boldsymbol{\phi}_2(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j), \cdots, \boldsymbol{\theta}^\top \boldsymbol{\phi}_p(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j)\right)^\top;$$

$$\boldsymbol{\phi}_i(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j) = \left(-\left(\textstyle\sum_k \hat{\boldsymbol{x}}_k^j + \hat{\boldsymbol{x}}_i^j\right), 1, \hat{\boldsymbol{s}}^j, 0, \ldots, -1, \ldots, 0\right)^\top.$$

According to the proof of Theorem 1, we derive that computing the incenter of the set $\mathbb{C}$ is equivalent to solving the following convex optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\vartheta}} \quad & \|\boldsymbol{\vartheta}\|_2 \\ \text{s.t.} \quad & \langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle + \left\| \sum_{i=1}^p \boldsymbol{\phi}_i(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j) \left(\boldsymbol{x}_i^j - \hat{\boldsymbol{x}}_i^j\right) \right\|_2 \leqslant 0, \\ & \forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N]. \end{aligned} \tag{42}$$

The main difference from Theorem 1 is the inequality constraint, which stems from a different definition of $\boldsymbol{\vartheta}$. Correspondingly, we define loss function $\ell_{\boldsymbol{\vartheta}}(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j)$ of $\boldsymbol{\vartheta}$ in this competition, similar to the one in Definition 3, as

$$\max_{\boldsymbol{x}^j \in \mathcal{X}^j} \left\{ \langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle + \left\| \sum_{i=1}^p \phi_i(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j) \left(\boldsymbol{x}_i^j - \hat{\boldsymbol{x}}_i^j\right) \right\|_2 \right\}.$$

Based on this loss function, we then leverage Algorithm 1 in Section 4.4 to estimate $\boldsymbol{\vartheta}$ for the Cournot competition.

Next, we present the prior incorporation of monotonicity of $\boldsymbol{F}$ using the semidefinite programming framework. Specifically, in this Cournot game, Jacobian $\nabla \boldsymbol{F}(\boldsymbol{x})$

$$\nabla \boldsymbol{F}(\boldsymbol{x}) = -\begin{bmatrix} -2a & -a & \ldots & -a \\ -a & -2a & \ldots & -a \\ \vdots & \vdots & \vdots & \vdots \\ -a & -a & \ldots & -2a \end{bmatrix}$$

is a symmetric matrix.

Let $\daleth = \begin{bmatrix} 2a & a & \ldots & a \\ a & 2a & \ldots & a \\ \vdots & \vdots & \vdots & \vdots \\ a & a & \ldots & 2a \end{bmatrix}$, $\Lambda_1 = \begin{bmatrix} c_1 & 0 & \ldots & 0 \\ 0 & c_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & c_p \end{bmatrix}$,

and $\Lambda_2 = \begin{bmatrix} b & 0 \\ 0 & d \end{bmatrix}$, estimating $\boldsymbol{\vartheta}$ can be reformulated as

$$\begin{aligned} \min_{\daleth, \Lambda_1, \Lambda_2} \quad & \frac{1}{N} \sum_{j=1}^N \ell_{\daleth, \Lambda_1, \Lambda_2}(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j) + \alpha \mathcal{R}(\daleth, \Lambda_1, \Lambda_2) \\ \text{s.t.} \quad & \daleth \succeq 0, \Lambda_1 \succeq 0, \Lambda_2 \succeq 0. \end{aligned} \tag{43}$$

Here, loss function $\ell_{\daleth, \Lambda_1, \Lambda_2}(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j)$ is defined as the form:

$$\begin{aligned} \ell_{\daleth, \Lambda_1, \Lambda_2}(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j) = \max_{\boldsymbol{x}^j \in \mathcal{X}^j} \Big\{ & \frac{1}{2p} \mathrm{Tr}\left(\Psi_0^j \daleth\right) + \mathrm{Tr}\left(\Psi_1^j \Lambda_1\right) \\ & + \mathrm{Tr}\left(\Psi_2^j \Lambda_2\right) + \Big\| \sum_{i=1}^p \boldsymbol{\phi}_i(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j) \left(\boldsymbol{x}_i^j - \hat{\boldsymbol{x}}_i^j\right) \Big\|_2 \Big\}, \end{aligned}$$

where diagonal matrices $\Psi_0^j = \mathrm{diag}\left(v_1^j, \ldots, v_1^j\right) \in \mathbb{R}^{p \times p}$, $\Psi_1^j = \mathrm{diag}\left(v_4^j, \ldots, v_{p+3}^j\right) \in \mathbb{R}^{p \times p}$, $\Psi_2^j = \mathrm{diag}\left(v_2^j, v_3^j\right) \in \mathbb{R}^2$, and the vector $\boldsymbol{v}^j = \sum_{i=1}^p \boldsymbol{\phi}_i(\hat{\boldsymbol{x}}^j, \hat{\boldsymbol{s}}^j)\left(\boldsymbol{x}_i^j - \hat{\boldsymbol{x}}_i^j\right)$.

To solve the problem (43), we leverage the primal-dual interior-point approach as in Section 4.5. We begin by deriving the perturbed KKT conditions in this game as follows.

$$\daleth^*, \Xi^* \succeq 0; \quad \alpha \daleth + \frac{1}{2pN} \sum_{j=1}^N \Psi_0^j(\bar{\boldsymbol{x}}^j) - \Xi = \boldsymbol{0}.$$

$$\Lambda_1^*, \Lambda_2^* \succeq 0; \quad \alpha \Lambda_1 + \frac{1}{N} \sum_{j=1}^N \Psi_1^j(\bar{\boldsymbol{x}}^j) - \Gamma_1 = \boldsymbol{0}.$$

$$\Gamma_1^*, \Gamma_2^* \succeq 0; \quad \alpha \Lambda_2 + \frac{1}{N} \sum_{j=1}^N \Psi_2^j(\bar{\boldsymbol{x}}^j) - \Gamma_2 = \boldsymbol{0}.$$

$$\Gamma_1^* \Lambda_1{}^* - \frac{1}{\mu}\boldsymbol{I} = \boldsymbol{0}; \Xi^* \daleth^* - \frac{1}{\mu}\boldsymbol{I} = \boldsymbol{0}; \Gamma_2^* \Lambda_2{}^* - \frac{1}{\mu}\boldsymbol{I} = \boldsymbol{0}.$$

Here, $\daleth^*$, Λ_1^*, and Λ_2^* are the optimal solutions to the primal problem (43). Ξ^*, Γ_1^*, and Γ_2^* are the optimal solutions to the corresponding dual problem. And μ is the barrier parameter.

To solve the above equations, we apply Newton's method by finding Jacobian $J(\daleth, \Lambda_1, \Lambda_2, \Xi, \Gamma_1, \Gamma_2)$, and repeatedly solving the system for $(\Delta\daleth, \Delta\Lambda_1, \Delta\Lambda_2, \Delta\Xi, \Delta\Gamma_1, \Delta\Gamma_2)$:

$$J(\daleth, \Lambda_1, \Lambda_2, \Xi, \Gamma_1, \Gamma_2) \begin{pmatrix} \Delta\daleth \\ \Delta\Lambda_1 \\ \Delta\Lambda_2 \\ \Delta\Xi \\ \Delta\Gamma_1 \\ \Delta\Gamma_2 \end{pmatrix} = -\boldsymbol{F}(\daleth, \Lambda_1, \Lambda_2, \Xi, \Gamma_1, \Gamma_2).$$

Algorithm 1 Primal-Dual Interior-Point for Problem (43)

1: **function** PD-IP($\widehat{\mathcal{D}}, \alpha, \epsilon, \beth^0, \Lambda_1^0, \Lambda_2^0, \Xi^0, \Gamma_1^0, \Gamma_2^0$).
2: $\quad \frac{1}{\mu^0} \leftarrow \frac{\mathrm{Tr}\left(\beth^0 \Xi^0\right)+\mathrm{Tr}\left(\Lambda_1^0 \Gamma_1^0\right)}{3p} + \frac{\mathrm{Tr}\left(\Lambda_2^0 \Gamma_2^0\right)}{6}$;
3: $\quad k \leftarrow 0$.
4: $\quad$ **while** $\frac{1}{\mu^k} > \epsilon$ **do**
5: $\qquad$ Compute $\Psi_0^j\left(\bar{\boldsymbol{x}}_k^j\right)$, $\Psi_1^j\left(\bar{\boldsymbol{x}}_k^j\right)$, and $\Psi_2^j\left(\bar{\boldsymbol{x}}_k^j\right)$;
6: $\qquad$ Compute $\left(\Delta\beth^k, \Delta\Lambda_1^k, \Delta\Lambda_2^k, \Delta\Xi^k, \Delta\Gamma_1^k, \Delta\Gamma_2^k\right)$;
7: $\qquad$ Backtracking line search for step-sizes η_p^k, η_d^k;
8: $\qquad$ Compute $\left(\beth^{k+1}, \Lambda_1^{k+1}, \Lambda_2^{k+1}, \Xi^{k+1}, \Gamma_1^{k+1}, \Gamma_2^{k+1}\right)$;
9: $\qquad$ Compute $\frac{1}{\mu^{k+1}}$;
10: $\qquad$ $k \leftarrow k+1$;
11: $\quad$ **end while**
12: $\quad$ **return** $\left(\beth^k, \Lambda_1^k, \Lambda_2^k, \Xi^k, \Gamma_1^k, \Gamma_2^k\right)$.
13: **end function**

Here, $\boldsymbol{F}$ is the vector-valued function composed of the six equations derived in the perturbed KKT condition. By finding the Jacobian J, we rewrite the above equation as follows.

$$\begin{pmatrix} \Xi\Delta\beth + \Delta\Xi\beth \\ \Gamma_1\Delta\Lambda_1 + \Delta\Gamma_1\Lambda_1 \\ \Gamma_2\Delta\Lambda_2 + \Delta\Gamma_2\Lambda_2 \\ \alpha\Delta\beth - \Delta\Xi \\ \alpha\Delta\Lambda_1 - \Delta\Gamma_1 \\ \alpha\Delta\Lambda_2 - \Delta\Gamma_2 \end{pmatrix} = - \begin{pmatrix} \Xi\beth - \frac{1}{\mu} I \\ \Gamma_1\Lambda_1 - \frac{1}{\mu} I \\ \Gamma_2\Lambda_2 - \frac{1}{\mu} I \\ \alpha\beth + \frac{1}{2pN}\sum_{j=1}^N \Psi_0^j\left(\bar{\boldsymbol{x}}^j\right) - \Xi \\ \alpha\Lambda_1 + \frac{1}{N}\sum_{j=1}^N \Psi_1^j\left(\bar{\boldsymbol{x}}^j\right) - \Gamma_1 \\ \alpha\Lambda_2 + \frac{1}{N}\sum_{j=1}^N \Psi_2^j\left(\bar{\boldsymbol{x}}^j\right) - \Gamma_2 \end{pmatrix}. \tag{44}$$

In Algorithm 1, we present the primal-dual interior-point algorithm for solving (43). In line 5, we compute the matrices $\Psi_0^j(\bar{\boldsymbol{x}}_k^j)$, $\Psi_1^j(\bar{\boldsymbol{x}}_k^j)$, and $\Psi_2^j(\bar{\boldsymbol{x}}_k^j)$ for each data point using current parameters. In line 6, the Newton update direction is determined. In line 7, we utilize the backtracking line search to compute current step sizes η_p^k and η_d^k for the primal and dual variables, respectively. In lines 8 to 9, the current variables and the barrier parameter are updated for the next iteration.

9 Our Methods for Traffic Games

In the traffic game of Section 5.1, let $\boldsymbol{\vartheta} = \left(\boldsymbol{\theta}_1; \cdots; \boldsymbol{\theta}_{|\mathcal{A}|}\right)$. Given an observed Wardrop equilibrium $\hat{\boldsymbol{x}}^j$, we define the loss function $\ell_{\boldsymbol{\vartheta}}\left(\hat{\boldsymbol{x}}^j\right)$ of $\boldsymbol{\vartheta}$ as the following form.

$$\max_{\boldsymbol{x}^j \in \mathcal{X}^j} \left\{ -\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \rangle + \left\| \boldsymbol{\phi}^j \odot \mathrm{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right) \right\|_2 \right\}.$$

This function is derived from Theorem 1, differing by a minus sign in the first term. Here, $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j$, and $\boldsymbol{\phi}^j$ are expressed as $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j = \left(\boldsymbol{\theta}_1{}^\top \boldsymbol{\phi}_1\left(\hat{\boldsymbol{x}}^j\right), \cdots, \boldsymbol{\theta}_{|\mathcal{A}|}{}^\top \boldsymbol{\phi}_{|\mathcal{A}|}\left(\hat{\boldsymbol{x}}^j\right)\right)^\top \in \mathbb{R}^{|\mathcal{A}|}$, $\boldsymbol{\phi}^j = \left[\boldsymbol{\phi}_1^\top\left(\hat{\boldsymbol{x}}^j\right), \cdots, \boldsymbol{\phi}_{|\mathcal{A}|}^\top\left(\hat{\boldsymbol{x}}^j\right)\right]^\top \in \mathbb{R}^{2|\mathcal{A}|}$.

Next, we formulate the estimation of $\boldsymbol{\vartheta}$ using the semidefinite programming framework. Under the cost function in Section 5.1, we derive the corresponding Jacobian matrix as

$$\nabla \boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} \frac{\gamma}{(C_1)^\gamma}\theta_1^1 x_1^{\gamma-1} & 0 & \dots & 0 \\ 0 & \frac{\gamma}{(C_2)^\gamma}\theta_2^1 x_2^{\gamma-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{\gamma}{(C_{|\mathcal{A}|})^\gamma}\theta_{|\mathcal{A}|}^1 x_{|\mathcal{A}|}^{\gamma-1} \end{bmatrix}.$$

We observe that the matrix $\nabla \boldsymbol{F}(\boldsymbol{x})$ is positive semidefinite when all elements are non-negative. Let variables $\Lambda_0 = \begin{bmatrix} \theta_1^0 & 0 & \dots & 0 \\ 0 & \theta_2^0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \theta_{|\mathcal{A}|}^0 \end{bmatrix}$, and $\Lambda_1 = \begin{bmatrix} \theta_1^1 & 0 & \dots & 0 \\ 0 & \theta_2^1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \theta_{|\mathcal{A}|}^1 \end{bmatrix}$. Note that the positive semidefiniteness of the variables Λ_0 and Λ_1 is equivalent to the positive semidefiniteness of $\nabla \boldsymbol{F}(\boldsymbol{x})$. Similar to Cournot games, we can reformulate our estimation problem using the proposed loss function as follows.

$$\begin{aligned} \min_{\Lambda_0, \Lambda_1} \quad & \frac{1}{N}\sum_{j=1}^N \ell_{\Lambda_0,\Lambda_1}\left(\hat{\boldsymbol{x}}^j\right) + \alpha\mathcal{R}\left(\Lambda_0, \Lambda_1\right) \\ \text{s.t.} \quad & \Lambda_0 \geq 0, \Lambda_1 \geq 0. \end{aligned} \tag{45}$$

Here, the loss function $\ell_{\Lambda_0,\Lambda_1}\left(\hat{\boldsymbol{x}}^j\right)$ is defined as

$$\begin{aligned} \ell_{\Lambda_0,\Lambda_1}\left(\hat{\boldsymbol{x}}^j\right) = \max_{\boldsymbol{x}^j \in \mathcal{X}^j} \Big\{ & -\mathrm{Tr}\left(\Psi_0^j \Lambda_0\right) - \mathrm{Tr}\left(\Psi_1^j \Lambda_1\right) \\ & + \left\| \boldsymbol{\phi}^j \odot \mathrm{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right) \right\|_2 \Big\}, \end{aligned}$$

where the matrices $\Psi_0^j = \mathrm{diag}\left(v_1^j, v_3^j, \dots, v_{2|\mathcal{A}|-1}^j\right) \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$, $\Psi_1^j = \mathrm{diag}\left(v_2^j, v_4^j, \dots, v_{2|\mathcal{A}|}^j\right) \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$, and vector $\boldsymbol{v}^j$ is defined as: $\boldsymbol{v}^j = \boldsymbol{\phi}^j \odot \mathrm{vec}\left(\mathbb{1}_n \otimes \left(\boldsymbol{x}^j - \hat{\boldsymbol{x}}^j\right)\right)$.

To solve the problem (45), we employ the primal-dual interior-point approach. We begin by deriving the perturbed KKT conditions in this scenario. Let $(\Lambda_0^*, \Lambda_1^*)$ and (Γ_0^*, Γ_1^*) be the primal and dual optimal solutions for problem (45), respectively. The optimality conditions for the corresponding logarithmic barrier centering problem are then given by

$$\Gamma_0^*\Lambda_0{}^* - \frac{1}{\mu}\boldsymbol{I} = \boldsymbol{0}; \quad \Gamma_1^*\Lambda_1{}^* - \frac{1}{\mu}\boldsymbol{I} = \boldsymbol{0}.$$

$$\Lambda_0^*, \Lambda_1^* \geq 0; \quad \alpha\Lambda_0^* - \frac{1}{N}\sum_{j=1}^N \Psi_0^j\left(\bar{\boldsymbol{x}}^j\right) - \Gamma_0^* = \boldsymbol{0}.$$

$$\Gamma_0^*, \Gamma_1^* \geq 0; \quad \alpha\Lambda_1^* - \frac{1}{N}\sum_{j=1}^N \Psi_1^j\left(\bar{\boldsymbol{x}}^j\right) - \Gamma_1^* = \boldsymbol{0}.$$

Next, we utilize Newton's method to solve the four equations, following a process similar to Cournot games. Algorithm 2 summarizes the whole approach for solving (45).

10 Numerical Experiments

10.1 Feasibility for Utility Estimation

This approach selects a $\boldsymbol{\vartheta}$ from the set of consistent parameter vectors, i.e., $\mathbb{C}\setminus\{\boldsymbol{0}\}$, by solving the following feasibility

Algorithm 2 Primal-Dual Interior-Point for Problem (45)

function PD-IP($\widehat{\mathcal{D}}, \alpha, \epsilon, \Lambda_0^0, \Lambda_1^0, \Gamma_0^0, \Gamma_1^0$).
  $\frac{1}{\mu^0} \leftarrow \frac{\mathrm{Tr}\left(\Lambda_0^0\Gamma_0^0\right)+\mathrm{Tr}\left(\Lambda_1^0\Gamma_1^0\right)}{2p}$;
  $k \leftarrow 0$.
  while $\frac{1}{\mu^k} > \epsilon$ **do**
    Compute $\Psi_0^j\left(\bar{\boldsymbol{x}}_k^j\right), \Psi_1^j\left(\bar{\boldsymbol{x}}_k^j\right)$ using $\hat{\mathcal{D}}$, Λ_0^k, and Λ_1^k;
    Compute $\left(\Delta\Lambda_0^k, \Delta\Lambda_1^k, \Delta\Gamma_0^k, \Delta\Gamma_1^k\right)$;
    Backtracking line search for step-sizes η_p^k, η_d^k;
    Compute $\left(\Lambda_0^{k+1}, \Lambda_1^{k+1}, \Gamma_0^{k+1}, \Gamma_1^{k+1}\right)$;
    Compute $\frac{1}{\mu^{k+1}}$;
    $k \leftarrow k+1$;
  end while
  return $\left(\Lambda_0^k, \Lambda_1^k, \Gamma_0^k, \Gamma_1^k\right)$.
end function

optimization problem (*feasibility program*):

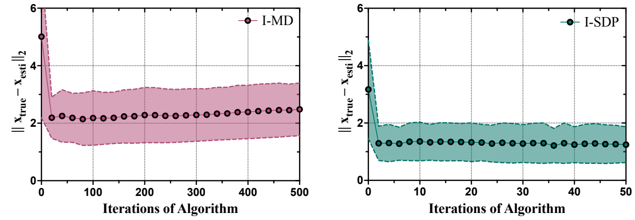
$$\begin{aligned} \min_{\boldsymbol{\vartheta}} \quad & 0 \\ \text{s.t.} \quad & \left\langle \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^j, \boldsymbol{x}^j - \hat{\boldsymbol{x}}^j \right\rangle \leqslant 0, \forall \boldsymbol{x}^j \in \mathcal{X}^j, \forall j \in [N]. \end{aligned} \tag{46}$$

Here, the constraints represent the variational inequalities that parameters $\boldsymbol{\vartheta}$ should satisfy based on observed dataset $\widehat{\mathcal{D}}$. To implement this program, we discretize each $\mathcal{X}^j$ to obtain discrete actions, especially when each $\mathcal{X}^j$ is infinite.

10.2 Bertsimas for Utility Estimation

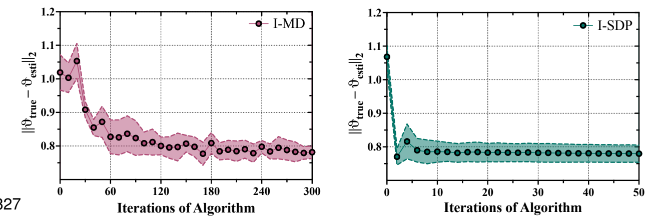
The **Bertsimas** estimation method is applicable to both two-firm Bertrand-Nash competition and three-firm Cournot games. In this subsection, we focus on the Bertrand competition to illustrate the approach. While the approach can be similarly extended to Cournot games, the details are omitted here for brevity due to their similarity. Specifically, in the two-firm Bertrand-Nash competition, this approach targets to solve the following inverse optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\vartheta},\boldsymbol{\epsilon},\boldsymbol{y}} \quad & \|\boldsymbol{\epsilon}\|_\infty \\ \text{s.t.} \quad & M_i\left(\hat{x}_1^j, \hat{x}_2^j, \hat{s}^j; \boldsymbol{\theta}_i\right) \leqslant y_i^j, \forall i = 1,2, \forall j = 1, \cdots, N, \\ & \sum_{i=1}^{2}\left(\bar{x}^j y_i^j - \hat{x}_i^j M_i\left(\hat{x}_1^j, \hat{x}_2^j, \hat{s}^j; \boldsymbol{\theta}_i\right)\right) \leqslant \epsilon_j, j=1,\cdots,N, \\ & \boldsymbol{y}^j \geqslant \mathbf{0}, \forall j = 1, \cdots, N, \\ & M_1\left(\hat{x}_1^j, \hat{x}_2^{\text{med}}, \hat{s}^{\text{med}}; \boldsymbol{\theta}_1\right) \geqslant M_1\left(\hat{x}_1^k, \hat{x}_2^{\text{med}}, \hat{s}^{\text{med}}; \boldsymbol{\theta}_1\right), \\ & \hat{x}_1^j \leqslant \hat{x}_1^k, \forall j,k = 1,\ldots,N, \\ & M_2\left(\hat{x}_1^{\text{med}}, \hat{x}_2^j, \hat{s}^{\text{med}}; \boldsymbol{\theta}_2\right) \geqslant M_2\left(\hat{x}_1^{\text{med}}, \hat{x}_2^k, \hat{s}^{\text{med}}; \boldsymbol{\theta}_2\right), \\ & \hat{x}_2^j \leqslant \hat{x}_2^k, \forall j,k = 1,\ldots,N, \\ & M_1\left(\underline{\hat{x}}_1, \hat{x}_2^{\text{med}}, \hat{s}^{\text{med}}; \boldsymbol{\theta}_1\right) = M_1^*\left(\underline{\hat{x}}_1, \hat{x}_2^{\text{med}}, \hat{s}^{\text{med}}; \boldsymbol{\theta}_1^*\right), \\ & M_2\left(\hat{x}_1^{\text{med}}, \underline{\hat{x}}_2, \hat{s}^{\text{med}}; \boldsymbol{\theta}_2\right) = M_2^*\left(\hat{x}_1^{\text{med}}, \underline{\hat{x}}_2, \hat{s}^{\text{med}}; \boldsymbol{\theta}_2^*\right). \end{aligned} \tag{47}$$



(a) Convergence of **I-MD**. (b) Convergence of **I-SDP**.

Figure 1: Convergence on the Cournot Testing Dataset.



(a) Convergence of **I-MD**. (b) Convergence of **I-SDP**.

Figure 2: Convergence Results on the Traffic Dataset.

Here, $\bar{x}$ represents the upper bound of the price, M_i denotes firm i's marginal revenue function, and $\boldsymbol{\theta}_i^*$ is its true parameter vector. Let $\hat{s}^{\text{med}}$ be the median value of $\hat{s}$ over the dataset. Breaking ties arbitrarily, $\hat{s}^{\text{med}}$ corresponds to some observation $j = j^{\text{med}}$. Then, $\hat{x}_1^{\text{med}}$ and $\hat{x}_2^{\text{med}}$ are the corresponding prices at time $j = j^{\text{med}}$. The prices $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_2$ represent the minimum prices observed in the dataset $\widehat{\mathcal{D}}$.

10.3 Implementation Details

We implement Algorithm 1 in Section 4.4 with $\omega(\boldsymbol{\vartheta}) = \frac{1}{2}\|\boldsymbol{\vartheta}\|_2^2$, $\mathcal{R}(\boldsymbol{\vartheta}) = \|\boldsymbol{\vartheta}\|_2$, and $\alpha = 0.01$. For Algorithm 2 in Section 4.5, we set $\epsilon = 0.001$ and $\alpha = 0.01$. All experiments are conducted on an Apple M1 Pro with 10-core CPU, 14-core GPU, 16-core Neural Engine, and 32GB of RAM. The datasets and codes are included with this material.

10.4 Convergence on Cournot Competition

In Figure 1, we depict the convergence performance of our **I-MD** and **I-SDP** methods on the Cournot testing dataset. The figures show the average error between the true equilibrium $\boldsymbol{x}_{\text{true}}$ and the equilibrium $\boldsymbol{x}_{\text{esti}}$ computed using $\boldsymbol{\vartheta}_{\text{esti}}$. Our **I-SDP** converges to lower errors with much smaller variances compared to **I-MD**, which can be attributed to the incorporation of priors and the primal-dual interior-point algorithm.

10.5 Convergence on Traffic Game

In Figure 2, we depict the convergence performance of our two methods, evaluated using the metric $\|\boldsymbol{\vartheta}_{\text{esti}} - \boldsymbol{\vartheta}_{\text{true}}\|_2$ over 30 trials. In Figure 2a and Figure 2b, after fewer iterations, our **I-SDP** converges to a lower error than **I-MD**.