A Quaternion-based Certifiably Optimal Solution to the Wahba Problem with Outliers

Supplementary Material

Heng Yang and Luca Carlone
Laboratory for Information & Decision Systems (LIDS)
Massachusetts Institute of Technology
{hankyang, lcarlone}@mit.edu

A. Proof of Proposition 3

Proof. Here we prove the equivalence between the mixed-integer program (13) and the optimization in (14) involving $N+1$ quaternions. To do so, we note that since $\theta_i \in \{+1,-1\}$ and $\frac{1+\theta_i}{2} \in \{0,1\}$, we can safely move $\frac{1+\theta_i}{2}$ inside the squared norm (because $0 = 0^2, 1 = 1^2$) in each summand of the cost function (13):

$$\sum_{i=1}^{N} \frac{1+\theta_i}{2} \frac{\|\hat{\boldsymbol{b}}_i - \boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}\|^2}{\sigma_i^2} + \frac{1-\theta_i}{2}\bar{c}^2 \tag{A1}$$
$$= \sum_{i=1}^{N} \frac{\|\hat{\boldsymbol{b}}_i - \boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1} + \theta_i\hat{\boldsymbol{b}}_i - \boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes(\theta_i\boldsymbol{q}^{-1})\|^2}{4\sigma_i^2} + \frac{1-\theta_i}{2}\bar{c}^2$$

Now we introduce N new unit quaternions $\boldsymbol{q}_i = \theta_i\boldsymbol{q}$, $i = 1,\dots,N$ by multiplying $\boldsymbol{q}$ by the N binary variables $\theta_i \in \{+1,-1\}$, a re-parametrization we called *binary cloning*. One can easily verify that $\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i = \theta_i(\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}) = \theta_i$. Hence, by substituting $\theta_i = \boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i$ into (A1), we can rewrite the mixed-integer program (13) as:

$$\min_{\substack{\boldsymbol{q}\in S^3 \\ \boldsymbol{q}_i\in\{\pm\boldsymbol{q}\}}} \sum_{i=1}^{N} \frac{\|\hat{\boldsymbol{b}}_i - \boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1} + \boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i - \boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1}\|^2}{4\sigma_i^2} + \frac{1-\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i}{2}\bar{c}^2, \tag{A2}$$

which is the same as the optimization in (14). ■

B. Proof of Proposition 4

Proof. Here we show that the optimization involving $N+1$ quaternions in (14) can be reformulated as the Quadratically-Constrained Quadratic Program (QCQP) in (15). Towards this goal, we prove that the objective function and the constraints in the QCQP are a re-parametrization of the ones in (14).

Equivalence of the objective functions. We start by developing the squared 2-norm term in (14):

$$\|\hat{\boldsymbol{b}}_i - \boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1} + \boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i - \boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1}\|^2$$
$$(\|\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i\|^2 = \|\hat{\boldsymbol{b}}_i\|^2 = \|\boldsymbol{b}_i\|^2,\ \hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i)\hat{\boldsymbol{b}}_i = \boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{b}_i\|^2)$$
$$(\|\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}\|^2 = \|\boldsymbol{R}\boldsymbol{a}_i\|^2 = \|\boldsymbol{a}_i\|^2)$$
$$(\|\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1}\|^2 = \|\theta_i\boldsymbol{R}\boldsymbol{a}_i\|^2 = \|\boldsymbol{a}_i\|^2)$$
$$((\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1})^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1}) = (\boldsymbol{R}\boldsymbol{a}_i)^{\mathsf{T}}(\theta_i\boldsymbol{R}\boldsymbol{a}_i) = \boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{a}_i\|^2)$$
$$= 2\|\boldsymbol{b}_i\|^2 + 2\|\boldsymbol{a}_i\|^2 + 2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{b}_i\|^2 + 2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{a}_i\|^2$$
$$-2\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}) - 2\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1})$$
$$-2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}) - 2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1}) \tag{A3}$$
$$(\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1}) = (\theta_i)^2\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}) = \hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1})\ )$$
$$(\ \hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1}) = \boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1})\ )$$
$$= 2\|\boldsymbol{b}_i\|^2 + 2\|\boldsymbol{a}_i\|^2 + 2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{b}_i\|^2 + 2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{a}_i\|^2$$
$$-4\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}) - 4\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}) \tag{A4}$$

where we have used multiple times the binary cloning equalities $\boldsymbol{q}_i = \theta_i\boldsymbol{q}, \theta_i = \boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i$, the equivalence between applying rotation to a homogeneous vector $\hat{\boldsymbol{a}}_i$ using quaternion product and using rotation matrix in eq. (10) from the main document, as well as the fact that vector 2-norm is invariant to rotation and homogenization (with zero padding).

Before moving to the next step, we make the following observation by combing eq. (8) and eq. (9):

$$\boldsymbol{\Omega}_1(\boldsymbol{q}^{-1}) = \boldsymbol{\Omega}_1^{\mathsf{T}}(\boldsymbol{q}), \quad \boldsymbol{\Omega}_2(\boldsymbol{q}^{-1}) = \boldsymbol{\Omega}_2^{\mathsf{T}}(\boldsymbol{q}) \tag{A5}$$

which states the linear operators $\boldsymbol{\Omega}_1(\cdot)$ and $\boldsymbol{\Omega}_2(\cdot)$ of $\boldsymbol{q}$ and its inverse $\boldsymbol{q}^{-1}$ are related by a simple transpose operation. In the next step, we use the equivalence between quaternion product and linear operators in $\boldsymbol{\Omega}_1(\boldsymbol{q})$ and $\boldsymbol{\Omega}_2(\boldsymbol{q})$ as defined in eq. (7)-(8) to simplify $\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1})$ in eq. (A4):

$$
\begin{aligned}
&\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}) \\
&{\scriptstyle(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i=\boldsymbol{\Omega}_1(\boldsymbol{q})\hat{\boldsymbol{a}}_i\ ,\ \boldsymbol{\Omega}_1(\boldsymbol{q})\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}=\boldsymbol{\Omega}_2(\boldsymbol{q}^{-1})\boldsymbol{\Omega}_1(\boldsymbol{q})\hat{\boldsymbol{a}}_i=\boldsymbol{\Omega}_2^{\mathsf{T}}(\boldsymbol{q})\boldsymbol{\Omega}_1(\boldsymbol{q})\hat{\boldsymbol{a}}_i)} \\
&=\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{\Omega}_2^{\mathsf{T}}(\boldsymbol{q})\boldsymbol{\Omega}_1(\boldsymbol{q})\hat{\boldsymbol{a}}_i) \qquad \text{(A6)}\\
&(\boldsymbol{\Omega}_2(\boldsymbol{q})\hat{\boldsymbol{b}}_i=\hat{\boldsymbol{b}}_i\otimes\boldsymbol{q}=\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{q}\ ,\ \boldsymbol{\Omega}_1(\boldsymbol{q})\hat{\boldsymbol{a}}_i=\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i=\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}) \\
&=\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1^{\mathsf{T}}(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}. \qquad \text{(A7)}
\end{aligned}
$$

Now we can insert eq. (A7) back to eq. (A4) and write:

$$
\begin{aligned}
&\|\hat{\boldsymbol{b}}_i-\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}+\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i-\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1}\|^2 \\
&=2\|\boldsymbol{b}_i\|^2+2\|\boldsymbol{a}_i\|^2+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{b}_i\|^2+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{a}_i\|^2 \\
&-4\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1})-4\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}) \qquad \text{(A8)}\\
&=2\|\boldsymbol{b}_i\|^2+2\|\boldsymbol{a}_i\|^2+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{b}_i\|^2+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{a}_i\|^2 \\
&-4\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1^{\mathsf{T}}(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}-4\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1^{\mathsf{T}}(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q} \qquad \text{(A9)}\\
&{\scriptstyle(\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1^{\mathsf{T}}(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}=\theta_i\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1^{\mathsf{T}}(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}=\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1^{\mathsf{T}}(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}_i)} \\
&=2\|\boldsymbol{b}_i\|^2+2\|\boldsymbol{a}_i\|^2+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{b}_i\|^2+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{a}_i\|^2 \\
&-4\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1^{\mathsf{T}}(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}-4\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1^{\mathsf{T}}(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}_i \qquad \text{(A10)}\\
&(-\boldsymbol{\Omega}_1^{\mathsf{T}}(\hat{\boldsymbol{b}}_i)=\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)) \\
&=2\|\boldsymbol{b}_i\|^2+2\|\boldsymbol{a}_i\|^2+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{b}_i\|^2+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\|\boldsymbol{a}_i\|^2 \\
&+4\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}+4\boldsymbol{q}^{\mathsf{T}}\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{q}_i, \qquad \text{(A11)}
\end{aligned}
$$

which is quadratic in $\boldsymbol{q}$ and $\boldsymbol{q}_i$. Substituting eq. (A11) back to (14), we can write the cost function as:

$$
\begin{aligned}
&\sum_{i=1}^{N}\frac{\|\hat{\boldsymbol{b}}_i-\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}^{-1}+\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i\hat{\boldsymbol{b}}_i-\boldsymbol{q}\otimes\hat{\boldsymbol{a}}_i\otimes\boldsymbol{q}_i^{-1}\|^2}{4\sigma_i^2}+\frac{1-\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}_i}{2}\bar{c}^2 \\
&=\sum_{i=1}^{N}\boldsymbol{q}_i^{\mathsf{T}}\left(\underbrace{\frac{(\|\boldsymbol{b}_i\|^2+\|\boldsymbol{a}_i\|^2)\mathbf{I}_4+2\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)}{2\sigma_i^2}+\frac{\bar{c}^2}{2}\mathbf{I}_4}_{:=\boldsymbol{Q}_{ii}}\right)\boldsymbol{q}_i \\
&+2\boldsymbol{q}^{\mathsf{T}}\left(\underbrace{\frac{(\|\boldsymbol{b}_i\|^2+\|\boldsymbol{a}_i\|^2)\mathbf{I}_4+2\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)}{4\sigma_i^2}-\frac{\bar{c}^2}{4}\mathbf{I}_4}_{:=\boldsymbol{Q}_{0i}}\right)\boldsymbol{q}_i, \qquad \text{(A12)}
\end{aligned}
$$

where we have used two facts: (i) $\boldsymbol{q}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{q}=\theta_i^2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{q}=\boldsymbol{q}_i^{\mathsf{T}}\boldsymbol{A}\boldsymbol{q}_i$ for any matrix $\boldsymbol{A}\in\mathbb{R}^{4\times4}$, (ii) $c=c\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}=\boldsymbol{q}^{\mathsf{T}}(c\mathbf{I}_4)\boldsymbol{q}$ for any real constant c, which allowed writing the quadratic forms of $\boldsymbol{q}$ and constant terms in the cost as quadratic forms of $\boldsymbol{q}_i$. Since we have not changed the decision variables $\boldsymbol{q}$ and $\{\boldsymbol{q}_i\}_{i=1}^N$, the optimization in (14) is therefore equivalent to the following optimization:

$$
\min_{\substack{\boldsymbol{q}\in\mathcal{S}^3\\ \boldsymbol{q}_i\in\{\pm\boldsymbol{q}\}}}\sum_{i=1}^{N}\boldsymbol{q}_i^{\mathsf{T}}\boldsymbol{Q}_{ii}\boldsymbol{q}_i+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{Q}_{0i}\boldsymbol{q}_i \qquad \text{(A13)}
$$

where $\boldsymbol{Q}_{ii}$ and $\boldsymbol{Q}_{0i}$ are the known 4×4 data matrices as defined in eq. (A12).

Now it remains to prove that the above optimization (A13) is equivalent to the QCQP in (15). Recall that $\boldsymbol{x}$ is the column vector stacking all the $N+1$ quaternions, *i.e.*, $\boldsymbol{x}=[\boldsymbol{q}^{\mathsf{T}}\ \boldsymbol{q}_1^{\mathsf{T}}\ \dots\ \boldsymbol{q}_N^{\mathsf{T}}]^{\mathsf{T}}\in\mathbb{R}^{4(N+1)}$. Let us introduce symmetric matrices $\boldsymbol{Q}_i\in\mathbb{R}^{4(N+1)\times4(N+1)}, i=1,\dots,N$ and let the 4×4 sub-block of $\boldsymbol{Q}_i$ corresponding to sub-vector $\boldsymbol{u}$ and $\boldsymbol{v}$, be denoted as $[\boldsymbol{Q}_i]_{uv}$; each $\boldsymbol{Q}_i$ is defined as:

$$
[\boldsymbol{Q}_i]_{uv}=\begin{cases}\boldsymbol{Q}_{ii} & \text{if } \boldsymbol{u}=\boldsymbol{q}_i \text{ and } \boldsymbol{v}=\boldsymbol{q}_i\\ \boldsymbol{Q}_{0i} & {\scriptstyle\text{if } \boldsymbol{u}=\boldsymbol{q} \text{ and } \boldsymbol{v}=\boldsymbol{q}_i \text{ or } \boldsymbol{u}=\boldsymbol{q}_i \text{ and } \boldsymbol{v}=\boldsymbol{q}}\\ \boldsymbol{0}_{4\times4} & \text{otherwise}\end{cases} \qquad \text{(A14)}
$$

i.e., $\boldsymbol{Q}_i$ has the diagonal 4×4 sub-block corresponding to $(\boldsymbol{q}_i,\boldsymbol{q}_i)$ be $\boldsymbol{Q}_{ii}$, has the two off-diagonal 4×4 sub-blocks corresponding to $(\boldsymbol{q},\boldsymbol{q}_i)$ and $(\boldsymbol{q}_i,\boldsymbol{q})$ be $\boldsymbol{Q}_{0i}$, and has all the other 4×4 sub-blocks be zero. Then we can write the cost function in eq. (A13) compactly using $\boldsymbol{x}$ and $\boldsymbol{Q}_i$:

$$
\sum_{i=1}^{N}\boldsymbol{q}_i^{\mathsf{T}}\boldsymbol{Q}_{ii}\boldsymbol{q}_i+2\boldsymbol{q}^{\mathsf{T}}\boldsymbol{Q}_{0i}\boldsymbol{q}_i=\sum_{i=1}^{N}\boldsymbol{x}^{\mathsf{T}}\boldsymbol{Q}_i\boldsymbol{x} \qquad \text{(A15)}
$$

Therefore, we proved that the objective functions in (14) and the QCQP (15) are the same.

Equivalence of the constraints. We are only left to prove that (14) and (15) have the same feasible set, *i.e.*, the following two sets of constraints are equivalent:

$$
\begin{cases}\boldsymbol{q}\in\mathcal{S}^3\\ \boldsymbol{q}_i\in\{\pm\boldsymbol{q}\},\\ i=1,\dots,N\end{cases}\Leftrightarrow\begin{cases}\boldsymbol{x}_q^{\mathsf{T}}\boldsymbol{x}_q=1\\ \boldsymbol{x}_{q_i}\boldsymbol{x}_{q_i}^{\mathsf{T}}=\boldsymbol{x}_q\boldsymbol{x}_q^{\mathsf{T}},\\ i=1,\dots,N\end{cases} \qquad \text{(A16)}
$$

We first prove the ($\Rightarrow$) direction. Since $\boldsymbol{q}\in\mathcal{S}^3$, it is obvious that $\boldsymbol{x}_q^{\mathsf{T}}\boldsymbol{x}_q=\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}=1$. In addition, since $\boldsymbol{q}_i\in\{+\boldsymbol{q},-\boldsymbol{q}\}$, it follows that $\boldsymbol{x}_{q_i}\boldsymbol{x}_{q_i}^{\mathsf{T}}=\boldsymbol{q}_i\boldsymbol{q}_i^{\mathsf{T}}=\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}=\boldsymbol{x}_q\boldsymbol{x}_q^{\mathsf{T}}$. Then we proof the reverse direction ($\Leftarrow$). Since $\boldsymbol{x}_q^{\mathsf{T}}\boldsymbol{x}_q=\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}$, so $\boldsymbol{x}_q^{\mathsf{T}}\boldsymbol{x}_q=1$ implies $\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q}=1$ and therefore $\boldsymbol{q}\in\mathcal{S}^3$. On the other hand, $\boldsymbol{x}_{q_i}\boldsymbol{x}_{q_i}^{\mathsf{T}}=\boldsymbol{x}_q\boldsymbol{x}_q^{\mathsf{T}}$ means $\boldsymbol{q}_i\boldsymbol{q}_i^{\mathsf{T}}=\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}$. If we write $\boldsymbol{q}_i=[q_{i1},q_{i2},q_{13},q_{i4}]^{\mathsf{T}}$ and $\boldsymbol{q}=[q_1,q_2,q_3,q_4]$, then the following matrix equality holds:

$$
\begin{bmatrix} q_{i1}^2 & q_{i1}q_{i2} & q_{i1}q_{i3} & q_{i1}q_{i4}\\ \star & q_{i2}^2 & q_{i2}q_{i3} & q_{i2}q_{i4}\\ \star & \star & q_{i3}^2 & q_{i3}q_{i4}\\ \star & \star & \star & q_{i4}^2\end{bmatrix}=\begin{bmatrix} q_1^2 & q_1q_2 & q_1q_3 & q_1q_4\\ \star & q_2^2 & q_2q_3 & q_2q_4\\ \star & \star & q_3^2 & q_3q_4\\ \star & \star & \star & q_4^2\end{bmatrix} \qquad \text{(A17)}
$$

First, from the diagonal equalities, we can get $q_{ij}=\theta_jq_j,\theta_j\in\{+1,-1\},j=1,2,3,4$. Then we look at the off-diagonal equality: $q_{ij}q_{ik}=q_jq_k,j\neq k$, since $q_{ij}=\theta_jq_j$ and $q_{ik}=\theta_kq_k$, we have $q_{ij}q_{ik}=\theta_j\theta_kq_jq_k$, from which we can have $\theta_j\theta_k=1,\forall j\neq k$. This implies that all the binary values $\{\theta_j\}_{j=1}^4$ have the same sign, and therefore they are equal to each other. As a result, $\boldsymbol{q}_i=\theta_i\boldsymbol{q}=\{+\boldsymbol{q},-\boldsymbol{q}\}$, showing the two sets of constraints in eq. (A16) are indeed equivalent. Therefore, the QCQP in eq. (15) is equivalent to the optimization in (A13), and the original optimization in (14) that involves $N+1$ quaternions, concluding the proof. ∎

C. Proof of Proposition 5

Proof. Here we show that the non-convex QCQP written in terms of the vector $\boldsymbol{x}$ in Proposition 4 (and eq. (15)) is equivalent to the non-convex problem written using the matrix $\boldsymbol{Z}$ in Proposition 5 (and eq. (18)). We do so by showing that the objective function and the constraints in (18) are a re-parametrization of the ones in (15).

Equivalence of the objective function. Since $\boldsymbol{Z} = \boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}$, and denoting $\boldsymbol{Q} \doteq \sum_{i=1}^{N} \boldsymbol{Q}_i$, we can rewrite the cost function in (18) as:

$$\sum_{i=1}^{N} \boldsymbol{x}^{\mathsf{T}} \boldsymbol{Q}_i \boldsymbol{x} = \boldsymbol{x}^{\mathsf{T}} \left(\sum_{i=1}^{N} \boldsymbol{Q}_i \right) \boldsymbol{x} = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{Q} \boldsymbol{x} = \operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}} \right) = \operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{Z} \right) \tag{A18}$$

showing the equivalence of the objectives in (15) and (18).

Equivalence of the constraints. It is trivial to see that $\boldsymbol{x}_q^{\mathsf{T}} \boldsymbol{x}_q = \operatorname{tr}\left(\boldsymbol{x}_q \boldsymbol{x}_q^{\mathsf{T}} \right) = 1$ is equivalent to $\operatorname{tr}\left([\boldsymbol{Z}]_{qq} \right) = 1$ by using the cyclic property of the trace operator and inspecting the structure of $\boldsymbol{Z}$. In addition, $\boldsymbol{x}_{q_i} \boldsymbol{x}_{q_i}^{\mathsf{T}} = \boldsymbol{x}_q \boldsymbol{x}_q^{\mathsf{T}}$ also directly maps to $[\boldsymbol{Z}]_{q_i q_i} = [\boldsymbol{Z}]_{qq}$ for all $i = 1, \ldots, N$. Lastly, requiring $\boldsymbol{Z} \succeq 0$ and $\operatorname{rank}(\boldsymbol{Z}) = 1$ is equivalent to restricting $\boldsymbol{Z}$ to the form $\boldsymbol{Z} = \boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}$ for some vector $\boldsymbol{x} \in \mathbb{R}^{4(N+1)}$. Therefore, the constraint sets of eq. (15) and (18) are also equivalent, concluding the proof. ∎

D. Proof of Proposition 6

Proof. We show eq. (19) is a convex relaxation of (18) by showing that (i) eq. (19) is a relaxation (*i.e.*, the constraint set of (19) includes the one of (18)), and (ii) eq. (19) is convex. (i) is true because from (18) to (19) we have dropped the $\operatorname{rank}(\boldsymbol{Z}) = 1$ constraint. Therefore, the feasible set of (18) is a subset of the feasible set of (19), and the optimal cost of (19) is always smaller or equal than the optimal cost of (18). To prove (ii), we note that the objective function and the constraints of (19) are all linear in $\boldsymbol{Z}$, and $\boldsymbol{Z} \succeq 0$ is a convex constraint, hence (19) is a convex program. ∎

E. Proof of Theorem 7

Proof. To prove Theorem 7, we first use Lagrangian duality to derive the *dual* problem of the QCQP in (15), and draw connections to the naive SDP relaxation in (19) (Section E.1). Then we leverage the well-known *Karush-Kuhn-Tucker* (KKT) conditions [3] to prove a general sufficient condition for tightness, as shown in Theorem A6 (Section E.2). Finally, in Section E.3, we demonstrate that in the case of no noise and no outliers, we can provide a constructive proof to show the sufficient condition in Theorem A6 always holds.

E.1. Lagrangian Function and Weak Duality

Recall the expressions of $\boldsymbol{Q}_i$ in eq. (A14), and define $\boldsymbol{Q} = \sum_{i=1}^{N} \boldsymbol{Q}_i$. The matrix $\boldsymbol{Q}$ has the following block structure:

$$\boldsymbol{Q} = \sum_{i=1}^{N} \boldsymbol{Q}_i = \left[\begin{array}{c|ccc} \mathbf{0} & \boldsymbol{Q}_{01} & \ldots & \boldsymbol{Q}_{0N} \\ \hline \boldsymbol{Q}_{01} & \boldsymbol{Q}_{11} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Q}_{0N} & \mathbf{0} & \ldots & \boldsymbol{Q}_{NN} \end{array} \right]. \tag{A19}$$

With cost matrix $\boldsymbol{Q}$, and using the cyclic property of the trace operator, the QCQP in eq. (15) can be written compactly as in the following proposition.

Proposition A1 (Primal QCQP)**.** *The QCQP in eq. (15) is equivalent to the following QCQP:*

$$(P) \quad \min_{\boldsymbol{x} \in \mathbb{R}^{4(N+1)}} \quad \operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}} \right) \tag{A20}$$
$$\text{subject to} \quad \operatorname{tr}\left(\boldsymbol{x}_q \boldsymbol{x}_q^{\mathsf{T}} \right) = 1$$
$$\boldsymbol{x}_{q_i} \boldsymbol{x}_{q_i}^{\mathsf{T}} = \boldsymbol{x}_q \boldsymbol{x}_q^{\mathsf{T}}, \forall i = 1, \ldots, N$$

We call this QCQP the primal problem *(P).*

The proposition can be proven by inspection. We now introduce the *Lagrangian function* [3] of the primal (P).

Proposition A2 (Lagrangian of Primal QCQP)**.** *The Lagrangian function of (P) can be written as:*

$$\mathcal{L}(\boldsymbol{x}, \mu, \boldsymbol{\Lambda}) = \operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}} \right) - \mu(\operatorname{tr}\left(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}} \right) - 1) - \sum_{i=1}^{N} \operatorname{tr}\left(\boldsymbol{\Lambda}_i \boldsymbol{x}\boldsymbol{x}^{\mathsf{T}} \right) \tag{A21}$$
$$= \operatorname{tr}\left((\boldsymbol{Q} - \mu\boldsymbol{J} - \boldsymbol{\Lambda})\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}} \right) + \mu. \tag{A22}$$

where $\boldsymbol{J}$ is a sparse matrix with all zeros except the first 4×4 diagonal block being identity matrix:

$$\boldsymbol{J} = \left[\begin{array}{c|ccc} \mathbf{I}_4 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right], \tag{A23}$$

and each $\boldsymbol{\Lambda}_i, i = 1, \ldots, N$ is a sparse Lagrangian multiplier matrix with all zeros except two diagonal sub-blocks $\pm\boldsymbol{\Lambda}_{ii} \in \mathrm{Sym}^{4\times4}$ (symmetric 4×4 matrices):

$$\boldsymbol{\Lambda}_i = \left[\begin{array}{c|ccccc} \boldsymbol{\Lambda}_{ii} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\boldsymbol{\Lambda}_{ii} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right], \tag{A24}$$

and $\mathbf{\Lambda}$ *is the sum of all* $\mathbf{\Lambda}_i$*'s,* $i = 1, \ldots, N$*:*

$$\mathbf{\Lambda} = \sum_{i=1}^{N} \mathbf{\Lambda}_i =$$

$$\left[\begin{array}{c|ccccc} \sum_{i=1}^{N} \mathbf{\Lambda}_{ii} & \mathbf{0} & \ldots & \mathbf{0} & \ldots & \mathbf{0} \\ \hline \mathbf{0} & -\mathbf{\Lambda}_{11} & \ldots & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & -\mathbf{\Lambda}_{ii} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \ldots & -\mathbf{\Lambda}_{NN} \end{array}\right]. \quad \text{(A25)}$$

Proof. The sparse matrix $\boldsymbol{J}$ defined in (A23) satisfies $\operatorname{tr}\left(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}\right) = \operatorname{tr}\left(\boldsymbol{x}_q\boldsymbol{x}_q^{\mathsf{T}}\right)$. Therefore $\mu(\operatorname{tr}\left(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}\right) - 1)$ is the same as $\mu(\operatorname{tr}\left(\boldsymbol{x}_q\boldsymbol{x}_q^{\mathsf{T}}\right) - 1)$, and μ is the Lagrange multiplier associated to the constraint $\operatorname{tr}\left(\boldsymbol{x}_q\boldsymbol{x}_q^{\mathsf{T}}\right) = 1$ in (P). Similarly, from the definition of the matrix $\mathbf{\Lambda}_i$ in (A25), it follows:

$$\operatorname{tr}\left(\mathbf{\Lambda}_i\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}\right) = \operatorname{tr}\left(\mathbf{\Lambda}_{ii}(\boldsymbol{x}_{q_i}\boldsymbol{x}_{q_i}^{\mathsf{T}} - \boldsymbol{x}_q\boldsymbol{x}_q^{\mathsf{T}})\right), \quad \text{(A26)}$$

where $\mathbf{\Lambda}_{ii}$ is the Lagrange multiplier (matrix) associated to each of the constraints $\boldsymbol{x}_{q_i}\boldsymbol{x}_{q_i}^{\mathsf{T}} = \boldsymbol{x}_q\boldsymbol{x}_q^{\mathsf{T}}$ in (P). This proves that (A21) (and eq. (A22), which rewrites (A21) in compact form) is the Lagrangian function of (P). ■

From the expression of the Lagrangian, we can readily obtain the Lagrangian *dual* problem.

Proposition A3 (Lagrangian Dual of Primal QCQP)**.** *The following SDP is the Lagrangian dual for the primal QCQP (P) in eq.* (A20)*:*

$$(D) \quad \max_{\substack{\mu\in\mathbb{R} \\ \mathbf{\Lambda}\in\mathbb{R}^{4(N+1)\times4(N+1)}}} \quad \mu \quad \text{(A27)}$$
$$\textit{subject to} \quad \boldsymbol{Q} - \mu\boldsymbol{J} - \mathbf{\Lambda} \succeq 0$$

where $\boldsymbol{J}$ *and* $\mathbf{\Lambda}$ *satisfy the structure in eq.* (A23) *and* (A25)*.*

Proof. By definition, the dual problem is [3]:

$$\max_{\mu,\mathbf{\Lambda}} \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \mu, \mathbf{\Lambda}), \quad \text{(A28)}$$

where $\mathcal{L}(\boldsymbol{x}, \mu, \mathbf{\Lambda})$ is the Lagrangian function. We observe:

$$\max_{\mu,\mathbf{\Lambda}} \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \mu, \mathbf{\Lambda}) = \begin{cases} \mu & \text{if } \boldsymbol{Q} - \mu\boldsymbol{J} - \mathbf{\Lambda} \succeq 0 \\ -\infty & \text{otherwise} \end{cases}. \quad \text{(A29)}$$

Since we are trying to maximize the Lagrangian (with respect to the dual variables), we discard the case leading to a cost of $-\infty$, obtaining the dual problem in (A27). ■

To connect the Lagrangian dual (D) to the naive SDP relaxation (19) in Proposition 6, we notice that the naive SDP relaxation is the dual SDP of the Lagrangian dual (D).

Proposition A4 (Naive Relaxation is the Dual of the Dual)**.** *The following SDP is the dual of the Lagrangian dual (D) in* (A27)*:*

$$(DD) \quad \min_{\boldsymbol{Z}\succeq 0} \quad \operatorname{tr}(\boldsymbol{Q}\boldsymbol{Z}) \quad \text{(A30)}$$
$$\textit{subject to} \quad \operatorname{tr}([\boldsymbol{Z}]_{qq}) = 1$$
$$[\boldsymbol{Z}]_{q_iq_i} = [\boldsymbol{Z}]_{qq}, \forall i = 1, \ldots, N$$

and (DD) is the same as the naive SDP relaxation in (19)*.*

Proof. We derive the Lagrangian dual problem of (DD) and show that it is indeed (D) (see similar example in [3, p. 265]). Similar to the proof of Proposition (A2), we can associate Lagrangian multiplier $\mu\boldsymbol{J}$ (eq. (A23)) to the constraint $\operatorname{tr}([\boldsymbol{Z}]_{qq}) = 1$, and associate $\mathbf{\Lambda}_i, i = 1, \ldots, N$ (eq. (A24)) to constraints $[\boldsymbol{Z}]_{q_iq_i} = [\boldsymbol{Z}]_{qq}, i = 1, \ldots, N$. In addition, we can associate matrix $\mathbf{\Theta} \in \mathrm{Sym}^{4(N+1)\times4(N+1)}$ to the constraint $\boldsymbol{Z} \succeq 0$. Then the Lagrangian of the SDP (DD) is:

$$\mathcal{L}(\boldsymbol{Z}, \mu, \mathbf{\Lambda}, \mathbf{\Theta})$$
$$= \operatorname{tr}(\boldsymbol{Q}\boldsymbol{Z}) - \mu(\operatorname{tr}(\boldsymbol{J}\boldsymbol{Z}) - 1) - \sum_{i=1}^{N}(\operatorname{tr}(\mathbf{\Lambda}_i\boldsymbol{Z})) - \operatorname{tr}(\mathbf{\Theta}\boldsymbol{Z})$$
$$= \operatorname{tr}((\boldsymbol{Q} - \mu\boldsymbol{J} - \mathbf{\Lambda} - \mathbf{\Theta})\boldsymbol{Z}) + \mu, \quad \text{(A31)}$$

and by definition, the dual problem is:

$$\max_{\mu,\mathbf{\Lambda},\mathbf{\Theta}} \min_{\boldsymbol{Z}} \mathcal{L}(\boldsymbol{Z}, \mu, \mathbf{\Lambda}, \mathbf{\Theta}). \quad \text{(A32)}$$

Because:

$$\max_{\mu,\mathbf{\Lambda},\mathbf{\Theta}} \min_{\boldsymbol{Z}} \mathcal{L} = \begin{cases} \mu & \text{if } \boldsymbol{Q} - \mu\boldsymbol{J} - \mathbf{\Lambda} - \mathbf{\Theta} \succeq 0 \\ -\infty & \text{otherwise} \end{cases}, \quad \text{(A33)}$$

we can get the Lagrangian dual problem of (DD) is:

$$\max_{\mu,\mathbf{\Lambda},\mathbf{\Theta}} \quad \mu \quad \text{(A34)}$$
$$\text{subject to} \quad \boldsymbol{Q} - \mu\boldsymbol{J} - \mathbf{\Lambda} \succeq \mathbf{\Theta}$$
$$\mathbf{\Theta} \succeq 0$$

Since $\mathbf{\Theta}$ is independent from the other decision variables and the cost function, we inspect that setting $\mathbf{\Theta} = \mathbf{0}$ actually maximizes μ and therefore can be removed. Removing $\mathbf{\Theta}$ from (A34) indeed leads to (D) in eq. (A27). ■

We can also verify weak duality by the following calculation. Denote $f_{DD} = \operatorname{tr}(\boldsymbol{Q}\boldsymbol{Z})$ and $f_D = \mu$. Recalling the structure of $\mathbf{\Lambda}$ from eq. (A25), we have $\operatorname{tr}(\mathbf{\Lambda}\boldsymbol{Z}) = 0$ because $[\boldsymbol{Z}]_{q_iq_i} = [\boldsymbol{Z}]_{qq}, \forall i = 1, \ldots, N$. Moreover, we have $\mu = \mu\operatorname{tr}(\boldsymbol{J}\boldsymbol{Z})$ due to the pattern of $\boldsymbol{J}$ from eq. (A23) and $\operatorname{tr}([\boldsymbol{Z}]_{qq}) = 1$. Therefore, the following inequality holds:

$$f_{DD} - f_D = \operatorname{tr}(\boldsymbol{Q}\boldsymbol{Z}) - \mu$$
$$= \operatorname{tr}(\boldsymbol{Q}\boldsymbol{Z}) - \mu\operatorname{tr}(\boldsymbol{J}\boldsymbol{Z}) - \operatorname{tr}(\mathbf{\Lambda}\boldsymbol{Z})$$
$$= \operatorname{tr}((\boldsymbol{Q} - \mu\boldsymbol{J} - \mathbf{\Lambda})\boldsymbol{Z}) \geq 0 \quad \text{(A35)}$$

where the last inequality holds true because both $\boldsymbol{Q}-\mu\boldsymbol{J}-\boldsymbol{\Lambda}$ and $\boldsymbol{Z}$ are positive semidefinite matrices. Eq. (A35) shows $f_{DD}\geq f_D$ always holds inside the feasible set and therefore by construction of (P), (D) and (DD), we have the following *weak duality* relation:

$$f_D^\star \leq f_{DD}^\star \leq f_P^\star. \tag{A36}$$

where the first inequality follows from eq. (A35) and the second inequality originates from the point that (DD) is a convex relaxation of (P), which has a larger feasible set and therefore the optimal cost of (DD) $(f_{DD}^\star)$ is always smaller than the optimal cost of (P) $(f_P^\star)$.

E.2. KKT conditions and strong duality

Despite the fact that weak duality gives lower bounds for objective of the primal QCQP (P), in this context we are interested in cases when *strong duality* holds, *i.e.*:

$$f_D^\star = f_{DD}^\star = f_P^\star, \tag{A37}$$

since in these cases solving any of the two convex SDPs (D) or (DD) will also solve the original non-convex QCQP (P) to *global optimality*.

Before stating the main theorem for strong duality, we study the *Karush-Kuhn-Tucker* (KKT) conditions [3] for the primal QCQP (P) in (A20), which will help pave the way to study strong duality.

Proposition A5 (KKT Conditions for Primal QCQP)**.** *If $\boldsymbol{x}^\star$ is an optimal solution to the primal QCQP (P) in* (A20) *(also* (15)*), and let $(\mu^\star,\boldsymbol{\Lambda}^\star)$ be the corresponding optimal dual variables (maybe not unique), then it must satisfy the following KKT conditions:*

(Stationary condition)

$$(\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star)\boldsymbol{x}^\star=\boldsymbol{0}, \tag{A38}$$

(Primal feasibility condition)

$$\boldsymbol{x}^\star \text{ satisfies the constraints in (A20)}. \tag{A39}$$

Using Propositions A1-A5, we state the following theorem that provides a sufficient condition for strong duality.

Theorem A6 (Sufficient Condition for Strong Duality)**.** *Given a stationary point $\boldsymbol{x}^\star$, if there exist dual variables $(\mu^\star,\boldsymbol{\Lambda}^\star)$ (maybe not unique) such that $(\boldsymbol{x}^\star,\mu^\star,\boldsymbol{\Lambda}^\star)$ satisfy both the KKT conditions in Proposition A5 and the dual feasibility condition $\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star\succeq 0$ in Proposition A3, then:*

(i) There is no duality gap between (P), (D) and (DD), i.e. $f_P^\star=f_D^\star=f_{DD}^\star$,

(ii) $\boldsymbol{x}^\star$ is a global minimizer for (P).

Moreover, if we have rank $(\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star)=4(N+1)-1$, i.e., *$\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star$ has $4(N+1)-1$ strictly positive eigenvalues and only one zero eigenvalue, then we have the following:*

(iii) $\pm\boldsymbol{x}^\star$ are the two unique global minimizers for (P),

(iv) The optimal solution to (DD), denoted as $\boldsymbol{Z}^\star$, has rank 1 and can be written as $\boldsymbol{Z}^\star=(\boldsymbol{x}^\star)(\boldsymbol{x}^\star)^\mathsf{T}$.

Proof. Recall from eq. (A36) that we already have weak duality by construction of (P), (D) and (DD). Now since $(\boldsymbol{x}^\star,\mu^\star,\boldsymbol{\Lambda}^\star)$ satisfies the KKT conditions (A39) and (A38), we have:

$$(\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star)\boldsymbol{x}^\star=\boldsymbol{0}\Rightarrow$$
$$(\boldsymbol{x}^\star)^\mathsf{T}(\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star)(\boldsymbol{x}^\star)=0\Rightarrow \tag{A40}$$
$$(\boldsymbol{x}^\star)^\mathsf{T}\boldsymbol{Q}(\boldsymbol{x}^\star)=\mu^\star(\boldsymbol{x}^\star)^\mathsf{T}\boldsymbol{J}(\boldsymbol{x}^\star)+(\boldsymbol{x}^\star)^\mathsf{T}\boldsymbol{\Lambda}^\star(\boldsymbol{x}^\star)\Rightarrow \tag{A41}$$

($\boldsymbol{x}^\star$ satisfies the constraints in (P) by KKT (A39))

(Recall structural partition of $\boldsymbol{J}$ and $\boldsymbol{\Lambda}$ in (A23) and (A25))

$$\operatorname{tr}\left(\boldsymbol{Q}(\boldsymbol{x}^\star)(\boldsymbol{x}^\star)^\mathsf{T}\right)=\mu^\star, \tag{A42}$$

which shows the cost of (P) is equal to the cost of (D) at $(\boldsymbol{x}^\star,\mu^\star,\boldsymbol{\Lambda}^\star)$. Moreover, since $\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star\succeq 0$ means $(\mu^\star,\boldsymbol{\Lambda}^\star)$ is actually dual feasible for (D), hence we have strong duality between (P) and (D): $f_P^\star=f_D^\star$. Because $f_{DD}^\star$ is sandwiched between $f_P^\star$ and $f_D^\star$ according to (A36), we have indeed strong duality for all of them:

$$f_D^\star=f_{DD}^\star=f_P^\star, \tag{A43}$$

proving (i). To prove (ii), we observe that for any $\boldsymbol{x}\in\mathbb{R}^{4(N+1)}$, $\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star\succeq 0$ means:

$$\boldsymbol{x}^\mathsf{T}(\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star)\boldsymbol{x}\geq 0. \tag{A44}$$

Specifically, let $\boldsymbol{x}$ be any vector that lies inside the feasible set of (P), *i.e.*, $\operatorname{tr}\left(\boldsymbol{x}_q\boldsymbol{x}_q^\mathsf{T}\right)=1$ and $\boldsymbol{x}_{q_i}\boldsymbol{x}_{q_i}^\mathsf{T}=\boldsymbol{x}_q\boldsymbol{x}_q^\mathsf{T},\forall i=1,\ldots,N$, then we have:

$$\boldsymbol{x}^\mathsf{T}(\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star)\boldsymbol{x}\geq 0\Rightarrow$$
$$\boldsymbol{x}^\mathsf{T}\boldsymbol{Q}\boldsymbol{x}\geq\mu^\star\boldsymbol{x}^\mathsf{T}\boldsymbol{J}\boldsymbol{x}+\boldsymbol{x}^\mathsf{T}\boldsymbol{\Lambda}^\star\boldsymbol{x}\Rightarrow \tag{A45}$$
$$\operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\right)\geq\mu^\star=\operatorname{tr}\left(\boldsymbol{Q}(\boldsymbol{x}^\star)(\boldsymbol{x}^\star)^\mathsf{T}\right), \tag{A46}$$

showing that the cost achieved by $\boldsymbol{x}^\star$ is no larger than the cost achieved by any other vectors inside the feasible set, which means $\boldsymbol{x}^\star$ is indeed a global minimizer to (P).

Next we use the additional condition of rank $(\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star)=4(N+1)-1$ to prove $\pm\boldsymbol{x}^\star$ are the two unique global minimizers to (P). Denote $\boldsymbol{M}^\star=\boldsymbol{Q}-\mu^\star\boldsymbol{J}-\boldsymbol{\Lambda}^\star$, since $\boldsymbol{M}^\star$ has only one zero eigenvalue with associated eigenvector $\boldsymbol{x}^\star$ (*cf.* KKT condition (A38)), its nullspace is defined by $\ker(\boldsymbol{M}^\star)=\{\boldsymbol{x}\in\mathbb{R}^{4(N+1)}:\boldsymbol{x}=a\boldsymbol{x}^\star,a\in\mathbb{R}\}$. Now denote the feasible set of (P) as $\Omega(P)$. It is clear to see that any vector in $\Omega(P)$ is a vertical stacking of $N+1$ unit quaternions and thus must have 2-norm equal to $\sqrt{N+1}$. Since $\boldsymbol{x}^\star \in \Omega(P)$ is already true, in order for any vector $\boldsymbol{x} = a\boldsymbol{x}^\star$ in $\ker(\boldsymbol{M}^\star)$ to be in $\Omega(P)$ as well, it must hold $|a|\|\boldsymbol{x}\| = \sqrt{N+1}$ and therefore $a = \pm 1$, *i.e.*, $\ker(\boldsymbol{M}^\star) \cap \Omega(P) = \{\pm\boldsymbol{x}^\star\}$. With this observation, we can argue that for any $\boldsymbol{x}$ inside $\Omega(P)$ that is not equal to $\{\pm\boldsymbol{x}^\star\}$, $\boldsymbol{x}$ cannot be in $\ker(\boldsymbol{M}^\star)$ and therefore:

$$\boldsymbol{x}^\mathsf{T}(\boldsymbol{M}^\star)\boldsymbol{x} > 0 \Rightarrow \tag{A47}$$
$$\boldsymbol{x}^\mathsf{T}\boldsymbol{Q}\boldsymbol{x} > \mu^\star\boldsymbol{x}^\mathsf{T}\boldsymbol{J}\boldsymbol{x} + \boldsymbol{x}^\mathsf{T}\boldsymbol{\Lambda}^\star\boldsymbol{x} \Rightarrow \tag{A48}$$
$$\operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\right) > \mu^\star = \operatorname{tr}\left(\boldsymbol{Q}(\boldsymbol{x}^\star)(\boldsymbol{x}^\star)^\mathsf{T}\right), \tag{A49}$$

which means for any vector $\boldsymbol{x} \in \Omega(P)/\{\pm\boldsymbol{x}^\star\}$, it results in strictly higher cost than $\pm\boldsymbol{x}^\star$. Hence $\pm\boldsymbol{x}^\star$ are the two unique global minimizers to (P) and (iii) is true.

To prove (iv), notice that since strong duality holds and $f^\star_{DD} = f^\star_D$, we can write the following according to eq. (A35):

$$\operatorname{tr}\left((\boldsymbol{Q} - \mu^\star\boldsymbol{J} - \boldsymbol{\Lambda}^\star)\boldsymbol{Z}^\star\right) = 0. \tag{A50}$$

Since $\boldsymbol{M}^\star = \boldsymbol{Q} - \mu^\star\boldsymbol{J} - \boldsymbol{\Lambda}^\star \succeq 0$ and has rank $4(N+1)-1$, we can write $\boldsymbol{M}^\star = \bar{\boldsymbol{M}}^\mathsf{T}\bar{\boldsymbol{M}}$ with $\bar{\boldsymbol{M}} \in \mathbb{R}^{(4(N+1)-1)\times 4(N+1)}$ and $\operatorname{rank}\left(\bar{\boldsymbol{M}}\right) = 4(N+1)-1$. Similarly, we can write $\boldsymbol{Z}^\star = \bar{\boldsymbol{Z}}\bar{\boldsymbol{Z}}^\mathsf{T}$ with $\bar{\boldsymbol{Z}} \in \mathbb{R}^{4(N+1)\times r}$ and $\operatorname{rank}\left(\bar{\boldsymbol{Z}}\right) = r = \operatorname{rank}\left(\boldsymbol{Z}^\star\right)$. Then from (A50) we have:

$$\begin{aligned} &\operatorname{tr}\left(\boldsymbol{M}^\star\boldsymbol{Z}^\star\right) = \operatorname{tr}\left(\bar{\boldsymbol{M}}^\mathsf{T}\bar{\boldsymbol{M}}\bar{\boldsymbol{Z}}\bar{\boldsymbol{Z}}^\mathsf{T}\right) \\ &= \operatorname{tr}\left(\bar{\boldsymbol{Z}}^\mathsf{T}\bar{\boldsymbol{M}}^\mathsf{T}\bar{\boldsymbol{M}}\bar{\boldsymbol{Z}}\right) = \operatorname{tr}\left((\bar{\boldsymbol{M}}\bar{\boldsymbol{Z}})^\mathsf{T}(\bar{\boldsymbol{M}}\bar{\boldsymbol{Z}})\right) \\ &= \|\bar{\boldsymbol{M}}\bar{\boldsymbol{Z}}\|_F^2 = 0, \end{aligned} \tag{A51}$$

which gives us $\bar{\boldsymbol{M}}\bar{\boldsymbol{Z}} = \mathbf{0}$. Using the rank inequality $\operatorname{rank}\left(\bar{\boldsymbol{M}}\bar{\boldsymbol{Z}}\right) \geq \operatorname{rank}\left(\bar{\boldsymbol{M}}\right) + \operatorname{rank}\left(\bar{\boldsymbol{Z}}\right) - 4(N+1)$, we have:

$$\begin{aligned} 0 \geq 4(N+1) - 1 + r - 4(N+1) \Rightarrow \\ r \leq 1. \end{aligned} \tag{A52}$$

Since $\bar{\boldsymbol{Z}} \neq \mathbf{0}$, we conclude that $\operatorname{rank}\left(\boldsymbol{Z}^\star\right) = \operatorname{rank}\left(\bar{\boldsymbol{Z}}\right) = r = 1$. As a result, since $\operatorname{rank}\left(\boldsymbol{Z}^\star\right) = 1$, and the rank constraint was the only constraint we dropped when relaxing the QCQP (P) to SDP (DD), we conclude that the relaxation is indeed tight. In addition, the rank 1 decomposition $\bar{\boldsymbol{Z}}$ of $\boldsymbol{Z}^\star$ is also the global minimizer to (P). However, from (iii), we know there are only two global minimizers to (P): $\boldsymbol{x}^\star$ and $-\boldsymbol{x}^\star$, so $\bar{\boldsymbol{Z}} \in \{\pm\boldsymbol{x}^\star\}$. Since the sign is irrelevant, we can always write $\boldsymbol{Z}^\star = (\boldsymbol{x}^\star)(\boldsymbol{x}^\star)^\mathsf{T}$, concluding the proof for (iv). ■

E.3. Strong duality in noiseless and outlier-free case

Now we are ready to prove Theorem 7 using Theorem A6. To do so, we will show that in the noiseless and outlier-free case, it is always possible to construct $\mu^\star$ and $\boldsymbol{\Lambda}^\star$ from $\boldsymbol{x}^\star$ and $\boldsymbol{Q}$ such that $(\boldsymbol{x}^\star, \mu^\star, \boldsymbol{\Lambda}^\star)$ satisfies the KKT conditions, and the dual matrix $\boldsymbol{M}^\star = \boldsymbol{Q} - \mu^\star\boldsymbol{J} - \boldsymbol{\Lambda}^\star$ is positive semidefinite and has only one zero eigenvalue.

Preliminaries. When there are no noise and outliers in the measurements, *i.e.*, $\boldsymbol{b}_i = \boldsymbol{R}\boldsymbol{a}_i, \forall i = 1, \dots, N$, we have $\|\boldsymbol{b}_i\|^2 = \|\boldsymbol{a}_i\|^2, \forall i = 1, \dots, N$. Moreover, without loss of generality, we assume $\sigma_i^2 = 1$ and $\bar{c}^2 > 0$. With these assumptions, we simplify the blocks $\boldsymbol{Q}_{0i}$ and $\boldsymbol{Q}_{ii}$ in the matrix $\boldsymbol{Q}$, *cf.* (A19) and (A12):

$$\boldsymbol{Q}_{0i} = \tfrac{\|\boldsymbol{a}_i\|^2}{2}\mathbf{I}_4 + \tfrac{\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)}{2} - \tfrac{\bar{c}^2}{4}\mathbf{I}_4, \tag{A53}$$
$$\boldsymbol{Q}_{ii} = \|\boldsymbol{a}_i\|^2\mathbf{I}_4 + \boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i) + \tfrac{\bar{c}^2}{2}\mathbf{I}_4. \tag{A54}$$

Due to the primal feasibility condition (A39), we know $\boldsymbol{x}^\star$ can be written as $N+1$ quaternions (*cf.* proof of (A16)): $\boldsymbol{x}^\star = [(\boldsymbol{q}^\star)^\mathsf{T}\ \theta_1^\star(\boldsymbol{q}^\star)^\mathsf{T}\ \dots\ \theta_N^\star(\boldsymbol{q}^\star)^\mathsf{T}]^\mathsf{T}$, where each $\theta_i^\star$ is a binary variable in $\{-1, +1\}$. Since we have assumed no noise and no outliers, we know $\theta_i^\star = +1$ for all i's and therefore $\boldsymbol{x}^\star = [(\boldsymbol{q}^\star)^\mathsf{T}\ (\boldsymbol{q}^\star)^\mathsf{T}\ \dots\ (\boldsymbol{q}^\star)^\mathsf{T}]^\mathsf{T}$. We can write the KKT stationary condition in matrix form as:

$$\overbrace{\left[\begin{array}{c|ccc} -\mu^\star\mathbf{I}_4 - \sum_{i=1}^{N}\boldsymbol{\Lambda}^\star_{ii} & \boldsymbol{Q}_{01} & \dots & \boldsymbol{Q}_{0N} \\ \hline \boldsymbol{Q}_{01} & \boldsymbol{Q}_{11} + \boldsymbol{\Lambda}^\star_{11} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Q}_{0N} & \mathbf{0} & \dots & \boldsymbol{Q}_{NN} + \boldsymbol{\Lambda}^\star_{NN} \end{array}\right]}^{\boldsymbol{M}^\star = \boldsymbol{Q} - \mu^\star\boldsymbol{J} - \boldsymbol{\Lambda}^\star} \overbrace{\left[\begin{array}{c} \boldsymbol{q}^\star \\ \hline \boldsymbol{q}^\star \\ \vdots \\ \boldsymbol{q}^\star \end{array}\right]}^{\boldsymbol{x}^\star} = \mathbf{0} \tag{A55}$$

and we index the block rows of $\boldsymbol{M}^\star$ from top to bottom as $0, 1, \dots, N$. The first observation we make is that eq. (A55) is a (highly) under-determined linear system with respect to the dual variables $(\mu^\star, \boldsymbol{\Lambda}^\star)$, because the linear system has $10N+1$ unknowns (each symmetric 4×4 matrix $\boldsymbol{\Lambda}^\star_{ii}$ has 10 unknowns, plus one unknown from $\mu^\star$), but only has $4(N+1)$ equations. To expose the structure of the linear system, we will apply a similarity transformation to the matrix $\boldsymbol{M}^\star$. Before we introduce the similarity transformation, we need additional properties about quaternions, described in the Lemma below. The properties can be proven by inspection.

Lemma A7 (More Quaternion Properties)**.** *The following properties about unit quaternions, involving the linear operators $\boldsymbol{\Omega}_1(\cdot)$ and $\boldsymbol{\Omega}_2(\cdot)$ introduced in eq. (8) hold true:*

(i) Commutative: for any two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^4$, The following equalities hold:

$$\boldsymbol{\Omega}_1(\boldsymbol{x})\boldsymbol{\Omega}_2(\boldsymbol{y}) = \boldsymbol{\Omega}_2(\boldsymbol{y})\boldsymbol{\Omega}_1(\boldsymbol{x}); \tag{A56}$$
$$\boldsymbol{\Omega}_1(\boldsymbol{x})\boldsymbol{\Omega}_2^\mathsf{T}(\boldsymbol{y}) = \boldsymbol{\Omega}_2^\mathsf{T}(\boldsymbol{y})\boldsymbol{\Omega}_1(\boldsymbol{x}); \tag{A57}$$
$$\boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{x})\boldsymbol{\Omega}_2(\boldsymbol{y}) = \boldsymbol{\Omega}_2(\boldsymbol{y})\boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{x}); \tag{A58}$$
$$\boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{x})\boldsymbol{\Omega}_2^\mathsf{T}(\boldsymbol{y}) = \boldsymbol{\Omega}_2^\mathsf{T}(\boldsymbol{y})\boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{x}). \tag{A59}$$

(ii) Orthogonality: for any unit quaternion $\boldsymbol{q} \in \mathcal{S}^3$, $\boldsymbol{\Omega}_1(\boldsymbol{q})$ and $\boldsymbol{\Omega}_2(\boldsymbol{q})$ are orthogonal matrices:

$$\boldsymbol{\Omega}_1(\boldsymbol{q})\boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q}) = \boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q})\boldsymbol{\Omega}_1(\boldsymbol{q}) = \mathbf{I}_4; \tag{A60}$$

$$\boldsymbol{\Omega}_2(\boldsymbol{q})\boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q}) = \boldsymbol{\Omega}_2^\mathsf{T}(\boldsymbol{q})\boldsymbol{\Omega}_1(\boldsymbol{q}) = \mathbf{I}_4. \tag{A61}$$

(iii) For any unit quaternion $\boldsymbol{q} \in \mathcal{S}^3$, the following equalities hold:

$$\boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q})\boldsymbol{q} = \boldsymbol{\Omega}_2^\mathsf{T}(\boldsymbol{q})\boldsymbol{q} = [0,0,0,1]^\mathsf{T}. \tag{A62}$$

(iv) For any unit quaternion $\boldsymbol{q} \in \mathcal{S}^3$, denote $\boldsymbol{R}$ as the unique rotation matrix associated with $\boldsymbol{q}$, then the following equalities hold:

$$\boldsymbol{\Omega}_1(\boldsymbol{q})\boldsymbol{\Omega}_2^\mathsf{T}(\boldsymbol{q}) = \boldsymbol{\Omega}_2^\mathsf{T}(\boldsymbol{q})\boldsymbol{\Omega}_1(\boldsymbol{q}) = \begin{bmatrix} \boldsymbol{R} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \doteq \tilde{\boldsymbol{R}}; \tag{A63}$$

$$\boldsymbol{\Omega}_2(\boldsymbol{q})\boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q}) = \boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q})\boldsymbol{\Omega}_2(\boldsymbol{q}) = \begin{bmatrix} \boldsymbol{R}^\mathsf{T} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \doteq \tilde{\boldsymbol{R}}^\mathsf{T}. \tag{A64}$$

Rewrite dual certificates using similarity transform. Now we are ready to define the similarity transformation. We define the matrix $\boldsymbol{D} \in \mathbb{R}^{4(N+1)\times 4(N+1)}$ as the following block diagonal matrix:

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{\Omega}_1(\boldsymbol{q}^\star) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_1(\boldsymbol{q}^\star) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Omega}_1(\boldsymbol{q}^\star) \end{bmatrix}. \tag{A65}$$

It is obvious to see that $\boldsymbol{D}$ is an orthogonal matrix from (ii) in Lemma A7, *i.e.*, $\boldsymbol{D}^\mathsf{T}\boldsymbol{D} = \boldsymbol{D}\boldsymbol{D}^\mathsf{T} = \mathbf{I}_{4(N+1)}$. Then we have the following Lemma.

Lemma A8 (Similarity Transformation)**.** *Define $\boldsymbol{N}^\star \doteq \boldsymbol{D}^\mathsf{T}\boldsymbol{M}^\star\boldsymbol{D}$, then:*

(i) $\boldsymbol{N}^\star$ and $\boldsymbol{M}^\star$ have the same eigenvalues, and

$$\boldsymbol{M}^\star \succeq 0 \Leftrightarrow \boldsymbol{N}^\star \succeq 0,\ \operatorname{rank}(\boldsymbol{M}^\star) = \operatorname{rank}(\boldsymbol{N}^\star). \tag{A66}$$

(ii) Define $\boldsymbol{e} = [0,0,0,1]^\mathsf{T}$ and $\boldsymbol{r} = [\boldsymbol{e}^\mathsf{T}\ \boldsymbol{e}^\mathsf{T}\ \ldots\ \boldsymbol{e}^\mathsf{T}]^\mathsf{T}$ as the vertical stacking of $N+1$ copies of $\boldsymbol{e}$, then:

$$\boldsymbol{M}^\star\boldsymbol{x}^\star = \mathbf{0} \Leftrightarrow \boldsymbol{N}^\star\boldsymbol{r} = \mathbf{0}. \tag{A67}$$

Proof. Because $\boldsymbol{D}^\mathsf{T}\boldsymbol{D} = \mathbf{I}_{4(N+1)}$, we have $\boldsymbol{D}^\mathsf{T} = \boldsymbol{D}^{-1}$ and $\boldsymbol{N}^\star = \boldsymbol{D}^\mathsf{T}\boldsymbol{M}^\star\boldsymbol{D} = \boldsymbol{D}^{-1}\boldsymbol{M}^\mathsf{T}\boldsymbol{D}$ is similar to $\boldsymbol{M}^\star$. Therefore, by matrix similarity, $\boldsymbol{M}^\star$ and $\boldsymbol{N}^\star$ have the same eigenvalues, and $\boldsymbol{M}^\star$ is positive semidefinite if and only if $\boldsymbol{N}^\star$ is positive semidefinite [4, p. 12]. To show (ii), we start by pre-multiplying both sides of eq. (A55) by $\boldsymbol{D}^\mathsf{T}$:

$$\boldsymbol{M}^\star\boldsymbol{x}^\star = \mathbf{0} \Leftrightarrow \boldsymbol{D}^\mathsf{T}\boldsymbol{M}^\star\boldsymbol{x}^\star = \boldsymbol{D}^\mathsf{T}\mathbf{0} \Leftrightarrow \tag{A68}$$

$$(\boldsymbol{D}\boldsymbol{D}^\mathsf{T} = \mathbf{I}_{4(N+1)})$$

$$\boldsymbol{D}^\mathsf{T}\boldsymbol{M}^\star(\boldsymbol{D}\boldsymbol{D}^\mathsf{T})\boldsymbol{x}^\star = \mathbf{0} \Leftrightarrow \tag{A69}$$

$$(\boldsymbol{D}^\mathsf{T}\boldsymbol{M}^\star\boldsymbol{D})(\boldsymbol{D}^\mathsf{T}\boldsymbol{x}^\star) = \mathbf{0} \Leftrightarrow \tag{A70}$$

$$(\boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q}^\star)\boldsymbol{q}^\star = \boldsymbol{e} \text{ from (iii) in Lemma A7})$$

$$\boldsymbol{N}^\star\boldsymbol{r} = \mathbf{0}, \tag{A71}$$

concluding the proof. ■

Lemma A8 suggests that constructing $\boldsymbol{M}^\star \succeq 0$ and $\operatorname{rank}(\boldsymbol{M}^\star) = 4(N+1)-1$ that satisfies the KKT conditions (A38) is equivalent to constructing $\boldsymbol{N}^\star \succeq 0$ and $\operatorname{rank}(\boldsymbol{N}^\star) = 4(N+1)-1$ that satisfies (A71). We then study the structure of $\boldsymbol{N}^\star$ and rewrite the KKT stationary condition.

Rewrite KKT conditions. In noiseless and outlier-free case, the KKT condition $\boldsymbol{M}^\star\boldsymbol{x}^\star = \mathbf{0}$ is equivalent to $\boldsymbol{N}^\star\boldsymbol{r} = \mathbf{0}$. Formally, after the similarity transformation $\boldsymbol{N}^\star = \boldsymbol{D}^\mathsf{T}\boldsymbol{M}^\star\boldsymbol{D}$, the KKT conditions can be explicitly rewritten as in the following proposition.

Proposition A9 (KKT conditions after similarity transformation)**.** *The KKT condition $\boldsymbol{N}^\star\boldsymbol{r} = \mathbf{0}$ (which is equivalent to eq. (A55)) can be written in matrix form:*

$$\left[\begin{array}{c|ccc} -\sum_{i=1}^{N}\bar{\boldsymbol{\Lambda}}_{ii}^\star & \bar{\boldsymbol{Q}}_{01} & \ldots & \bar{\boldsymbol{Q}}_{0N} \\ \hline \bar{\boldsymbol{Q}}_{01} & \bar{\boldsymbol{Q}}_{11}+\bar{\boldsymbol{\Lambda}}_{11}^\star & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\boldsymbol{Q}}_{0N} & \mathbf{0} & \ldots & \bar{\boldsymbol{Q}}_{NN}+\bar{\boldsymbol{\Lambda}}_{NN}^\star \end{array}\right] \begin{bmatrix} \boldsymbol{e} \\ \hline \boldsymbol{e} \\ \vdots \\ \boldsymbol{e} \end{bmatrix} = \mathbf{0}. \tag{A72}$$

with $\mu^\star = 0$ being removed compared to eq. (A55) and $\bar{\boldsymbol{Q}}_{0i}$ and $\bar{\boldsymbol{Q}}_{ii}, i = 1,\ldots,N$ are the following sparse matrices:

$$\bar{\boldsymbol{Q}}_{0i} \doteq \boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q}^\star)\boldsymbol{Q}_{0i}\boldsymbol{\Omega}_1(\boldsymbol{q}^\star) = \begin{bmatrix} \left(\frac{\|\boldsymbol{a}_i\|^2}{2} - \frac{\bar{c}^2}{4}\right)\mathbf{I}_3 - \frac{[\boldsymbol{a}_i]_\times^2}{2} - \frac{\boldsymbol{a}_i\boldsymbol{a}_i^\mathsf{T}}{2} & \mathbf{0} \\ \mathbf{0} & -\frac{\bar{c}^2}{4} \end{bmatrix}; \tag{A73}$$

$$\bar{\boldsymbol{Q}}_{ii} \doteq \boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q}^\star)\boldsymbol{Q}_{ii}\boldsymbol{\Omega}_1(\boldsymbol{q}^\star) = \begin{bmatrix} \left(\|\boldsymbol{a}_i\|^2 + \frac{\bar{c}^2}{2}\right)\mathbf{I}_3 - [\boldsymbol{a}_i]_\times^2 - \boldsymbol{a}_i\boldsymbol{a}_i^\mathsf{T} & \mathbf{0} \\ \mathbf{0} & \frac{\bar{c}^2}{2} \end{bmatrix}, \tag{A74}$$

and $\bar{\boldsymbol{\Lambda}}_{ii}^\star \doteq \boldsymbol{\Omega}_1^\mathsf{T}(\boldsymbol{q}^\star)\boldsymbol{\Lambda}_{ii}^\star\boldsymbol{\Omega}_1(\boldsymbol{q}^\star)$ has the following form:

$$\bar{\boldsymbol{\Lambda}}_{ii}^\star = \begin{bmatrix} \boldsymbol{E}_{ii} & \boldsymbol{\alpha}_i \\ \boldsymbol{\alpha}_i^\mathsf{T} & \lambda_i \end{bmatrix}, \tag{A75}$$

where $\boldsymbol{E}_{ii} \in \mathrm{Sym}^{3\times 3}$, $\boldsymbol{\alpha}_i \in \mathbb{R}^3$ and $\lambda_i \in \mathbb{R}$.

Proof. We first prove that $\bar{\boldsymbol{Q}}_{0i}$ and $\bar{\boldsymbol{Q}}_{ii}$ have the forms in (A73) and (A74) when there are no noise and outliers in the measurements. Towards this goal, we examine the similar matrix to $\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)$ (as it is a common part to $\boldsymbol{Q}_{0i}$ and $\boldsymbol{Q}_{ii}$):

$$\boldsymbol{\Omega}_1^{\mathsf{T}}(\boldsymbol{q}^\star)\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\boldsymbol{\Omega}_1(\boldsymbol{q}^\star)$$

(Commutative property in Lemma A7 (i))

$$= \boldsymbol{\Omega}_1^{\mathsf{T}}(\boldsymbol{q}^\star)\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_1(\boldsymbol{q}^\star)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i) \tag{A76}$$

(Orthogonality property in Lemma A7 (ii))

$$= \boldsymbol{\Omega}_1^{\mathsf{T}}(\boldsymbol{q}^\star)\boldsymbol{\Omega}_2(\boldsymbol{q}^\star)\boldsymbol{\Omega}_2^{\mathsf{T}}(\boldsymbol{q}^\star)\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_1(\boldsymbol{q}^\star)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i) \tag{A77}$$

(Lemma A7 (i) and (iv))

$$= (\tilde{\boldsymbol{R}}^\star)^{\mathsf{T}}\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2^{\mathsf{T}}(\boldsymbol{q}^\star)\boldsymbol{\Omega}_1(\boldsymbol{q}^\star)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i) \tag{A78}$$

(Lemma A7 (iv))

$$= (\tilde{\boldsymbol{R}}^\star)^{\mathsf{T}}\boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)(\tilde{\boldsymbol{R}}^\star)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i) \tag{A79}$$

$$= \begin{bmatrix} (\boldsymbol{R}^\star)^{\mathsf{T}}[\boldsymbol{b}_i]_\times \boldsymbol{R}^\star & (\boldsymbol{R}^\star)^{\mathsf{T}}\boldsymbol{b}_i \\ -\boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{R}^\star & 0 \end{bmatrix}\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i) \tag{A80}$$

$$= \begin{bmatrix} [\boldsymbol{a}_i]_\times & \boldsymbol{a}_i \\ -\boldsymbol{a}_i^{\mathsf{T}} & 0 \end{bmatrix}\begin{bmatrix} -[\boldsymbol{a}_i]_\times & \boldsymbol{a}_i \\ -\boldsymbol{a}_i^{\mathsf{T}} & 0 \end{bmatrix} \tag{A81}$$

$$= \begin{bmatrix} -[\boldsymbol{a}_i]_\times^2 - \boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}} & \boldsymbol{0} \\ \boldsymbol{0} & -\|\boldsymbol{a}_i\|^2 \end{bmatrix}. \tag{A82}$$

Using this property, and recall the definition of $\boldsymbol{Q}_{0i}$ and $\boldsymbol{Q}_{ii}$ in eq. (A53) and (A54), the similar matrices to $\boldsymbol{Q}_{0i}$ and $\boldsymbol{Q}_{ii}$ can be shown to have the expressions in (A73) and (A74) by inspection.

Showing $\bar{\boldsymbol{\Lambda}}_{ii}$ having the expression in (A75) is straightforward. Since $\boldsymbol{\Lambda}_{ii}^\star$ is symmetric, $\bar{\boldsymbol{\Lambda}}_{ii}^\star = \boldsymbol{\Omega}_1^{\mathsf{T}}(\boldsymbol{q}^\star)\boldsymbol{\Lambda}_{ii}^\star\boldsymbol{\Omega}_1(\boldsymbol{q}^\star)$ must also be symmetric and therefore eq. (A75) must be true for some $\boldsymbol{E}_{ii}$, $\boldsymbol{\alpha}_i$ and λ_i.

Lastly, in the noiseless and outlier-free case, $\mu^\star$ is zero due to the following:

$$\mu^\star = \operatorname{tr}\left(\boldsymbol{Q}(\boldsymbol{x}^\star)(\boldsymbol{x}^\star)^{\mathsf{T}}\right)$$

(Recall $\boldsymbol{Q}$ from eq. (A19))

$$= (\boldsymbol{q}^\star)^{\mathsf{T}}\left(\textstyle\sum_{i=1}^N(\boldsymbol{Q}_{ii} + 2\boldsymbol{Q}_{0i})\right)(\boldsymbol{q}^\star) \tag{A83}$$

$$= (\boldsymbol{q}^\star)^{\mathsf{T}}\left(2\textstyle\sum_{i=1}^N\|\boldsymbol{a}_i\|^2 + \boldsymbol{\Omega}_1(\hat{\boldsymbol{b}}_i)\boldsymbol{\Omega}_2(\hat{\boldsymbol{a}}_i)\right)(\boldsymbol{q}^\star) \tag{A84}$$

(Recall eq. (A7))

$$= 2\textstyle\sum_{i=1}^N\|\boldsymbol{a}_i\|^2 - \hat{\boldsymbol{b}}_i^{\mathsf{T}}(\boldsymbol{q}^\star \otimes \hat{\boldsymbol{a}}_i \otimes (\boldsymbol{q}^\star)^{-1}) \tag{A85}$$

$$= 2\textstyle\sum_{i=1}^N\|\boldsymbol{a}_i\|^2 - \boldsymbol{b}_i^{\mathsf{T}}(\boldsymbol{R}^\star\boldsymbol{a}_i) \tag{A86}$$

$$= 2\textstyle\sum_{i=1}^N\|\boldsymbol{a}_i\|^2 - \|\boldsymbol{b}_i\|^2 = 0. \tag{A87}$$

concluding the proof. ■

From KKT condition to sparsity pattern of dual variable. From the above proposition about the rewritten KKT condition (A72), we can claim the following sparsity pattern on the dual variables $\bar{\boldsymbol{\Lambda}}_{ii}$.

Lemma A10 (Sparsity Pattern of Dual Variables)**.** *The KKT condition eq. (A72) holds if and only if the dual variables* $\{\bar{\boldsymbol{\Lambda}}_{ii}\}_{i=1}^N$ *have the following sparsity pattern:*

$$\bar{\boldsymbol{\Lambda}}_{ii}^\star = \begin{bmatrix} \boldsymbol{E}_{ii} & \boldsymbol{0}_3 \\ \boldsymbol{0}_3 & -\frac{\bar{c}^2}{4} \end{bmatrix}, \tag{A88}$$

i.e., $\boldsymbol{\alpha}_i = 0$ *and* $\lambda_i = -\frac{\bar{c}^2}{4}$ *in eq. (A75) for every* $i = 1, \ldots, N$*.*

Proof. We first proof the trivial direction ($\Leftarrow$). If $\bar{\boldsymbol{\Lambda}}_{ii}^\star$ has the sparsity pattern in eq. (A88), then the product of the i-th block row of $\boldsymbol{N}^\star$ ($i = 1, \ldots, N$) and $\boldsymbol{r}$ writes (*cf.* eq. (A72)):

$$\left(\bar{\boldsymbol{Q}}_{0i} + \bar{\boldsymbol{Q}}_{ii} + \bar{\boldsymbol{\Lambda}}_{ii}^\star\right)\boldsymbol{e}$$

(Recall $\bar{\boldsymbol{Q}}_{0i}$ and $\bar{\boldsymbol{Q}}_{ii}$ from eq. (A73) and (A74))

$$= \begin{bmatrix} \star & \boldsymbol{0}_3 \\ \boldsymbol{0}_3 & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{0}_3 \\ 1 \end{bmatrix} = \boldsymbol{0}_4, \tag{A89}$$

which is equal to $\boldsymbol{0}_4$ for sure. For the product of the 0-th block row (the very top row) of $\boldsymbol{N}^\star$ and $\boldsymbol{r}$, we get:

$$\left(\textstyle\sum_{i=1}^N \bar{\boldsymbol{Q}}_{0i} - \bar{\boldsymbol{\Lambda}}_{ii}^\star\right)\boldsymbol{e}$$

$$= \begin{bmatrix} \star & \boldsymbol{0}_3 \\ \boldsymbol{0}_3 & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{0}_3 \\ 1 \end{bmatrix} = \boldsymbol{0}_4, \tag{A90}$$

which vanishes as well. Therefore, $\bar{\boldsymbol{\Lambda}}_{ii}^\star$ having the sparsity pattern in eq. (A88) provides a sufficient condition for KKT condition in eq. (A72). To show the other direction ($\Rightarrow$), first notice that eq. (A72) implies eq. (A89) holds true for all $i = 1, \ldots, N$ and in fact, eq. (A89) provides the following equation for constraining the general form of $\bar{\boldsymbol{\Lambda}}_{ii}^\star$ in eq. (A75):

$$\left(\bar{\boldsymbol{Q}}_{0i} + \bar{\boldsymbol{Q}}_{ii} + \bar{\boldsymbol{\Lambda}}_{ii}^\star\right)\boldsymbol{e}$$

$$= \begin{bmatrix} \star & \boldsymbol{\alpha}_i \\ \boldsymbol{\alpha}_i^{\mathsf{T}} & \lambda_i + \frac{\bar{c}^2}{4} \end{bmatrix}\begin{bmatrix} \boldsymbol{0}_3 \\ 1 \end{bmatrix} = \boldsymbol{0}_4, \tag{A91}$$

which directly gives rise to:

$$\begin{cases} \boldsymbol{\alpha}_i = \boldsymbol{0}_3 \\ \lambda_i + \frac{\bar{c}^2}{4} = 0 \end{cases}. \tag{A92}$$

and the sparsity pattern in eq. (A88), showing that $\bar{\boldsymbol{\Lambda}}_{ii}^\star$ having the sparsity pattern in eq. (A88) is also a necessary condition. ■

Find the dual certificate. Lemma A10 further suggests that the linear system resulted from KKT conditions (A72) (also (A55)) is highly under-determined in the sense that we have full freedom in choosing the $\boldsymbol{E}_{ii}$ block of the dual variable $\bar{\boldsymbol{\Lambda}}_{ii}^\star$. Therefore, we introduce the following proposition.

Proposition A11 (Construction of Dual Variable)**.** *In the noiseless and outlier-free case, choosing $\boldsymbol{E}_{ii}$ as:*

$$\boldsymbol{E}_{ii} = [\boldsymbol{a}_k]_\times^2 - \frac{1}{4}\bar{c}^2\mathbf{I}_3, \forall i = 1, \dots, N \tag{A93}$$

and choosing $\bar{\boldsymbol{\Lambda}}_{ii}^\star$ as having the sparsity pattern in (A88) will not only satisfy the KKT conditions in (A72) but also make $\boldsymbol{N}^\star \succeq 0$ and $\operatorname{rank}(\boldsymbol{N}^\star) = 4(N+1) - 1$. Therefore, by Theorem A6, the naive relaxation in Proposition 6 is always tight and Theorem 7 is true.

Proof. By Lemma A10, we only need to prove the choice of $\boldsymbol{E}_{ii}$ in (A93) makes $\boldsymbol{N}^\star \succeq 0$ and $\operatorname{rank}(\boldsymbol{N}^\star) = 4(N+1) - 1$. Towards this goal, we will show that for any vector $\boldsymbol{u} \in \mathbb{R}^{4(N+1)}$, $\boldsymbol{u}^\top \boldsymbol{N}^\star \boldsymbol{u} \geq 0$ and the nullspace of $\boldsymbol{N}^\star$ is $\ker(\boldsymbol{N}^\star) = \{\boldsymbol{u} : \boldsymbol{u} = a\boldsymbol{r}, a \in \mathbb{R} \text{ and } a \neq 0\}$ ($\boldsymbol{N}^\star$ has a single zero eigenvalue with associated eigenvector $\boldsymbol{r}$). We partition $\boldsymbol{u} = [\boldsymbol{u}_0^\top\ \boldsymbol{u}_1^\top\ \dots\ \boldsymbol{u}_N^\top]^\top$, where $\boldsymbol{u}_i \in \mathbb{R}^4, \forall i = 0, \dots, N$. Then using the form of $\boldsymbol{N}^\star$ in (A72), we can write $\boldsymbol{u}^\top \boldsymbol{N}^\star \boldsymbol{u}$ as:

$$\begin{aligned}\boldsymbol{u}^\top \boldsymbol{N}^\star \boldsymbol{u} = &-\textstyle\sum_{i=1}^N \boldsymbol{u}_0^\top \bar{\boldsymbol{\Lambda}}_{ii}\boldsymbol{u}_0 + \\ &2\textstyle\sum_{i=1}^N \boldsymbol{u}_0^\top \bar{\boldsymbol{Q}}_{0i}\boldsymbol{u}_i + \sum_{i=1}^N \boldsymbol{u}_i^\top(\bar{\boldsymbol{Q}}_{ii} + \bar{\boldsymbol{\Lambda}}_{ii})\boldsymbol{u}_i \end{aligned} \tag{A94}$$

$$= \sum_{i=1}^N \underbrace{\boldsymbol{u}_0^\top(-\bar{\boldsymbol{\Lambda}}_{ii})\boldsymbol{u}_0 + \boldsymbol{u}_0^\top(2\bar{\boldsymbol{Q}}_{0i})\boldsymbol{u}_i + \boldsymbol{u}_i^\top(\bar{\boldsymbol{Q}}_{ii} + \bar{\boldsymbol{\Lambda}}_{ii})\boldsymbol{u}_i}_{m_i} \tag{A95}$$

Further denoting $\boldsymbol{u}_i = [\bar{\boldsymbol{u}}_i^\top\ u_i]^\top$ with $\bar{\boldsymbol{u}}_i \in \mathbb{R}^3$, m_i can be written as:

$$\begin{aligned} m_i = &\begin{bmatrix}\bar{\boldsymbol{u}}_0 \\ u_0\end{bmatrix}^\top \begin{bmatrix} -\boldsymbol{E}_{ii} & \mathbf{0} \\ \mathbf{0} & \frac{\bar{c}^2}{4}\end{bmatrix} \begin{bmatrix}\bar{\boldsymbol{u}}_0 \\ u_0\end{bmatrix} + \\ &\begin{bmatrix}\bar{\boldsymbol{u}}_0 \\ u_0\end{bmatrix}^\top \left[\begin{array}{c|c} \begin{array}{c}\left(\|\boldsymbol{a}_i\|^2 - \frac{\bar{c}^2}{2}\right)\mathbf{I}_3 \\ -[\boldsymbol{a}_i]_\times^2 - \boldsymbol{a}_i\boldsymbol{a}_i^\top\end{array} & \mathbf{0} \\ \hline \mathbf{0} & -\frac{\bar{c}^2}{2}\end{array}\right] \begin{bmatrix}\bar{\boldsymbol{u}}_i \\ u_i\end{bmatrix} + \\ &\begin{bmatrix}\bar{\boldsymbol{u}}_i \\ u_i\end{bmatrix}^\top \left[\begin{array}{c|c} \begin{array}{c}\left(\|\boldsymbol{a}_i\|^2 + \frac{\bar{c}^2}{2}\right)\mathbf{I}_3 \\ -[\boldsymbol{a}_i]_\times^2 - \boldsymbol{a}_i\boldsymbol{a}_i^\top + \boldsymbol{E}_{ii}\end{array} & \mathbf{0} \\ \hline \mathbf{0} & \frac{\bar{c}^2}{4}\end{array}\right] \begin{bmatrix}\bar{\boldsymbol{u}}_i \\ u_i\end{bmatrix} \end{aligned} \tag{A96}$$

$$\begin{aligned} = &\tfrac{\bar{c}^2}{4}\underbrace{\left(u_0^2 - 2u_0u_i + u_i^2\right)}_{=(u_0 - u_i)^2 \geq 0} + \bar{\boldsymbol{u}}_0^\top(-\boldsymbol{E}_{ii})\bar{\boldsymbol{u}}_0 + \\ &\bar{\boldsymbol{u}}_0^\top\left(\|\boldsymbol{a}_i\|^2\mathbf{I}_3 - \tfrac{\bar{c}^2}{2}\mathbf{I}_3 - [\boldsymbol{a}_i]_\times^2 - \boldsymbol{a}_i\boldsymbol{a}_i^\top\right)\bar{\boldsymbol{u}}_i + \\ &\bar{\boldsymbol{u}}_i^\top\left(\|\boldsymbol{a}_i\|^2\mathbf{I}_3 + \tfrac{\bar{c}^2}{2}\mathbf{I}_3 - [\boldsymbol{a}_i]_\times^2 - \boldsymbol{a}_i\boldsymbol{a}_i^\top + \boldsymbol{E}_{ii}\right)\bar{\boldsymbol{u}}_i, \end{aligned} \tag{A97}$$

where equality holds only when $u_i = u_0$ in the underbraced inequality in (A97). Now we insert the choice of $\boldsymbol{E}_{ii}$ in eq. (A93) to m_i to get the following inequality:

$$\begin{aligned} m_i \geq &\bar{\boldsymbol{u}}_0^\top(-[\boldsymbol{a}_i]_\times^2)\bar{\boldsymbol{u}}_0 + \tfrac{1}{4}\bar{c}^2\bar{\boldsymbol{u}}_0^\top\bar{\boldsymbol{u}}_0 + \\ &\|\boldsymbol{a}_i\|^2\bar{\boldsymbol{u}}_0^\top\bar{\boldsymbol{u}}_i - \tfrac{\bar{c}^2}{2}\bar{\boldsymbol{u}}_0^\top\bar{\boldsymbol{u}}_i + \bar{\boldsymbol{u}}_0^\top(-[\boldsymbol{a}_i]_\times^2)\bar{\boldsymbol{u}}_i - \bar{\boldsymbol{u}}_0^\top\boldsymbol{a}_i\boldsymbol{a}_i^\top\bar{\boldsymbol{u}}_i + \\ &\|\boldsymbol{a}_i\|^2\bar{\boldsymbol{u}}_i^\top\bar{\boldsymbol{u}}_i + \tfrac{1}{4}\bar{c}^2\bar{\boldsymbol{u}}_i^\top\bar{\boldsymbol{u}}_i - \bar{\boldsymbol{u}}_i^\top\boldsymbol{a}_i\boldsymbol{a}_i^\top\bar{\boldsymbol{u}}_i. \end{aligned} \tag{A98}$$

Using the following facts:

$$\|\boldsymbol{a}_i\|^2\bar{\boldsymbol{u}}_0^\top\bar{\boldsymbol{u}}_i = \bar{\boldsymbol{u}}_0^\top(-[\boldsymbol{a}_i]_\times^2 + \boldsymbol{a}_i\boldsymbol{a}_i^\top)\bar{\boldsymbol{u}}_i; \tag{A99}$$

$$\|\boldsymbol{a}_i\|^2\bar{\boldsymbol{u}}_i^\top\bar{\boldsymbol{u}}_i = \bar{\boldsymbol{u}}_i^\top(-[\boldsymbol{a}_i]_\times^2 + \boldsymbol{a}_i\boldsymbol{a}_i^\top)\bar{\boldsymbol{u}}_i, \tag{A100}$$

eq. (A98) can be simplified as:

$$\begin{aligned} m_i \geq &\bar{\boldsymbol{u}}_0^\top(-[\boldsymbol{a}_i]_\times^2)\bar{\boldsymbol{u}}_0 + \tfrac{1}{4}\bar{c}^2\bar{\boldsymbol{u}}_0^\top\bar{\boldsymbol{u}}_0 + \\ &2\bar{\boldsymbol{u}}_0^\top(-[\boldsymbol{a}_i]_\times^2)\bar{\boldsymbol{u}}_i - \tfrac{\bar{c}^2}{2}\bar{\boldsymbol{u}}_0^\top\bar{\boldsymbol{u}}_i + \\ &\bar{\boldsymbol{u}}_i^\top(-[\boldsymbol{a}_i]_\times^2)\bar{\boldsymbol{u}}_i + \tfrac{1}{4}\bar{c}^2\bar{\boldsymbol{u}}_i^\top\bar{\boldsymbol{u}}_i \end{aligned} \tag{A101}$$

$$\begin{aligned} = &\left([\boldsymbol{a}_i]_\times\bar{\boldsymbol{u}}_0 + [\boldsymbol{a}_i]_\times\bar{\boldsymbol{u}}_i\right)^\top\left([\boldsymbol{a}_i]_\times\bar{\boldsymbol{u}}_0 + [\boldsymbol{a}_i]_\times\bar{\boldsymbol{u}}_i\right) + \\ &\tfrac{\bar{c}^2}{4}\left(\bar{\boldsymbol{u}}_0^\top\bar{\boldsymbol{u}}_0 - 2\bar{\boldsymbol{u}}_0^\top\bar{\boldsymbol{u}}_i + \bar{\boldsymbol{u}}_i^\top\bar{\boldsymbol{u}}_i\right) \end{aligned} \tag{A102}$$

$$\begin{aligned} = &\|\boldsymbol{a}_i \times (\bar{\boldsymbol{u}}_0 + \bar{\boldsymbol{u}}_i)\|^2 + \tfrac{\bar{c}^2}{4}\|\bar{\boldsymbol{u}}_0 - \bar{\boldsymbol{u}}_i\|^2 \\ \geq &\ 0. \end{aligned} \tag{A103}$$

Since each m_i is nonnegative, $\boldsymbol{u}^\top\boldsymbol{N}^\star\boldsymbol{u} \geq 0$ holds true for any vector $\boldsymbol{u}$ and therefore $\boldsymbol{N}^\star \succeq 0$ is true. To see $\boldsymbol{N}^\star$ only has one zero eigenvalue, we notice that $\boldsymbol{u}^\top\boldsymbol{N}^\star\boldsymbol{u} = 0$ holds only when:

$$\begin{cases} u_i = u_0, \forall i = 1, \dots, N \\ \bar{\boldsymbol{u}}_0 = \bar{\boldsymbol{u}}_i, \forall i = 1, \dots, N \\ \boldsymbol{a}_i \times (\bar{\boldsymbol{u}}_0 + \bar{\boldsymbol{u}}_i) = \mathbf{0}_3, \forall i = 1, \dots, N \end{cases} \tag{A104}$$

because we have more than two $\boldsymbol{a}_i$'s that are not parallel to each other, eq. (A104) leads to $\bar{\boldsymbol{u}}_0 = \bar{\boldsymbol{u}}_i = \mathbf{0}_3, \forall i = 1, \dots, N$. Therefore the only set of nonzero vectors that satisfy the above conditions are $\{\boldsymbol{u} \in \mathbb{R}^{4(N+1)} : \boldsymbol{u} = a\boldsymbol{r}, a \in \mathbb{R} \text{ and } a \neq 0\}$. Therefore, $\boldsymbol{N}^\star$ has only one zero eigenvalue and $\operatorname{rank}(\boldsymbol{N}^\star) = 4(N+1) - 1$. ∎

Proposition A11 indeed proves the original Theorem 7 by giving valid constructions of dual variables under which strong duality always holds in the noiseless and outlier-free case. ∎

F. Proof of Proposition 8

Proof. Here we prove that eq. (20) is a convex relaxation of eq. (18) and that the relaxation is always tighter, *i.e.* the optimal objective of (20) is always closer to the optimal objective of (15), when compared to the naive relaxation (19).

To prove the first claim, we first show that the additional constraints in the last two lines of (20) are *redundant* for (18), *i.e.*, they are trivially satisfied by any feasible solution of (18). Towards this goal we note that eq. (18) is equivalent to (15), where $\boldsymbol{x}$ is a column vector stacking $N+1$ quaternions: $\boldsymbol{x} = [\boldsymbol{q}^\top\ \boldsymbol{q}_1^\top\ \dots \boldsymbol{q}_N^\top]^\top$, and where each $\boldsymbol{q}_i = \theta_i\boldsymbol{q}, \theta_i \in \{\pm 1\}, \forall i = 1, \dots, N$. Therefore, we have:

$$[\boldsymbol{Z}]_{qq_i} = \boldsymbol{q}\boldsymbol{q}_i^\top = \theta_i\boldsymbol{q}\boldsymbol{q}^\top = \boldsymbol{q}_i\boldsymbol{q}^\top = (\boldsymbol{q}\boldsymbol{q}_i^\top)^\top = [\boldsymbol{Z}]_{qq_i}^\top$$

$$[\boldsymbol{Z}]_{q_iq_j} = \boldsymbol{q}_i\boldsymbol{q}_j^\top = \theta_i\theta_j\boldsymbol{q}\boldsymbol{q}^\top = \boldsymbol{q}_j\boldsymbol{q}_i^\top = (\boldsymbol{q}_i\boldsymbol{q}_j^\top)^\top = [\boldsymbol{Z}]_{q_iq_j}^\top$$

This proves that the constraints $[\boldsymbol{Z}]_{qq_i} = [\boldsymbol{Z}]_{qq_i}^{\mathsf{T}}$ and $[\boldsymbol{Z}]_{q_iq_j} = [\boldsymbol{Z}]_{q_iq_j}^{\mathsf{T}}$ are redundant for (18). Therefore, problem (18) is equivalent to:

$$
\begin{aligned}
\min_{\boldsymbol{Z} \succeq 0} \quad & \operatorname{tr}(\boldsymbol{QZ}) \\
\text{subject to} \quad & \operatorname{tr}([\boldsymbol{Z}]_{qq}) = 1 \\
& [\boldsymbol{Z}]_{q_iq_i} = [\boldsymbol{Z}]_{qq}, \forall i = 1, \dots, N \\
& [\boldsymbol{Z}]_{qq_i} = [\boldsymbol{Z}]_{qq_i}^{\mathsf{T}}, \forall i = 1, \dots, N \\
& [\boldsymbol{Z}]_{q_iq_j} = [\boldsymbol{Z}]_{q_iq_j}^{\mathsf{T}}, \forall 1 \leq i < j \leq N \\
& \operatorname{rank}(\boldsymbol{Z}) = 1
\end{aligned}
\tag{A105}
$$

where we added the redundant constraints as they do not alter the feasible set. At this point, proving that (20) is a convex relaxation of (A105) (and hence of (18)) can be done with the same arguments of the proof of Proposition 6: in (20) we dropped the rank constraint (leading to a larger feasible set) and the remaining constraints are convex.

The proof of the second claim is straightforward. Since we added more constraints in (20) compared to the naive relaxation (19), the optimal cost of (20) always achieves a higher objective than (19), and since they are both relaxations, their objectives provide a lower bound to the original non-convex problem (18). ■

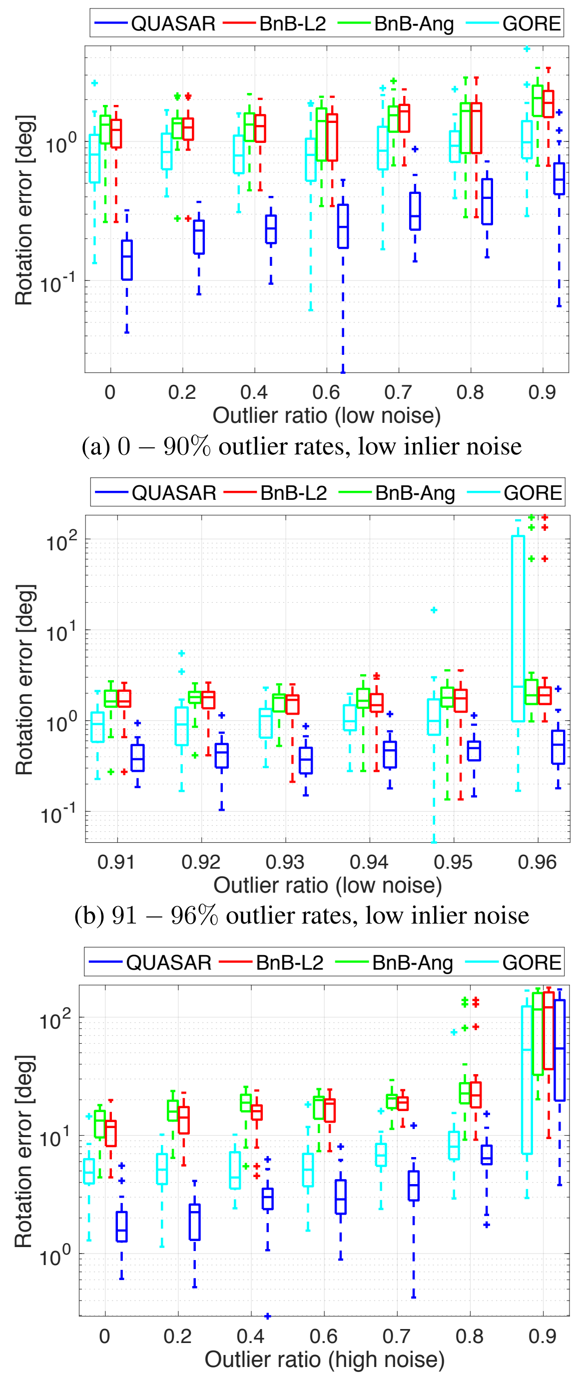
G. Benchmark against BnB

We follow the same experimental setup as in Section 6.1 of the main document, and benchmark QUASAR against (i) *Guaranteed Outlier Removal* [7] (label: GORE); (ii) BnB with *L-2 distance threshold* [2] (label: BnB-L2); and (iii) BnB with *angular distance threshold* [7] (label: BnB-Ang). Fig. A1 boxplots the distribution of rotation errors for 30 Monte Carlo runs in different combinations of outlier rates and noise corruptions. In the case of low inlier noise ($\sigma = 0.01$), QUASAR is robust against 96% outliers and achieves significantly better estimation accuracy compared to GORE, BnB-L2 and BnB-Ang, all of which experience failures at 96% outlier rates (Fig. A1(a,b)). In the case of high inlier noise ($\sigma = 0.1$), QUASAR is still robust against 80% outlier rates and has lower estimation error compared to the other methods.

H. Image Stitching Results

Here we provide extra image stitching results. As mentioned in the main document, we use the *Lunch Room* images from the PASSTA dataset [6], which contains 72 images in total. We performed pairwise image stitching for 12 times, stitching image pairs $(6i+1, 6i+7)$ for $i = 0, \dots, 11$ (when $i = 11$, $6i + 7 = 73$ is cycled to image 1). The reason for not doing image stitching between consecutive image pairs is to reduce the relative overlapping area so that SURF [1] feature matching is more prone to output outliers, creating a more challenging benchmark for QUASAR.

QUASAR successfully performed all 12 image stitching tasks and Table A1 reports the statistics. As we mentioned



(a) $0 - 90\%$ outlier rates, low inlier noise

(b) $91 - 96\%$ outlier rates, low inlier noise

(c) $0 - 90\%$ outlier rates, high inlier noise

Figure A1. Rotation estimation error by QUASAR, GORE, BnB-L2 and BnB-Ang on (a) $0 - 90\%$ outlier rates with low inlier noise, (b) $91 - 96\%$ outlier rates with low inlier noise and (c) $0 - 90\%$ outlier rates with high inlier noise.

	Mean	SD
SURF Outlier Ratio	14%	18.3%
Relative Duality Gap	1.40e−09	2.18e−09
Rank	1	0
Stable Rank	$1 + 8.33\text{e}{-17}$	2.96e−16

Table A1. Image stitching statistics (mean and standard deviation (SD)) of QUASAR on the *Lunch Room* dataset [6].

in the main document, the stitching of image pair (7,13) was the most challenging, due to the high outlier ratio of 66%, and the RANSAC-based stitching method [8, 5] as implemented by the Matlab "estimateGeometricTransform" function failed in that case. We show the failed example from RANSAC in Fig. A2.

References

[1] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: speeded up robust features. In *European Conf. on Computer Vision (ECCV)*, 2006. 10

[2] Jean-Charles Bazin, Yongduek Seo, and Marc Pollefeys. Globally optimal consensus set maximization through rotation search. In *Asian Conference on Computer Vision*, pages 539–551. Springer, 2012. 10

[3] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. 3, 4, 5

[4] Stephen Boyd and Lieven Vandenberghe. Subgradients. Notes for EE364b, 2006. 7

[5] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. 11

[6] Giulia Meneghetti, Martin Danelljan, Michael Felsberg, and Klas Nordberg. Image alignment for panorama stitching in sparsely structured environments. In *Scandinavian Conference on Image Analysis*, pages 428–439. Springer, 2015. 10, 11

[7] Álvaro Parra Bustos and Tat-Jun Chin. Guaranteed outlier removal for rotation search. In *Intl. Conf. on Computer Vision (ICCV)*, pages 2165–2173, 2015. 10

[8] Philip H. S. Torr and Andrew Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Comput. Vis. Image Underst.*, 78(1):138–156, 2000. 11

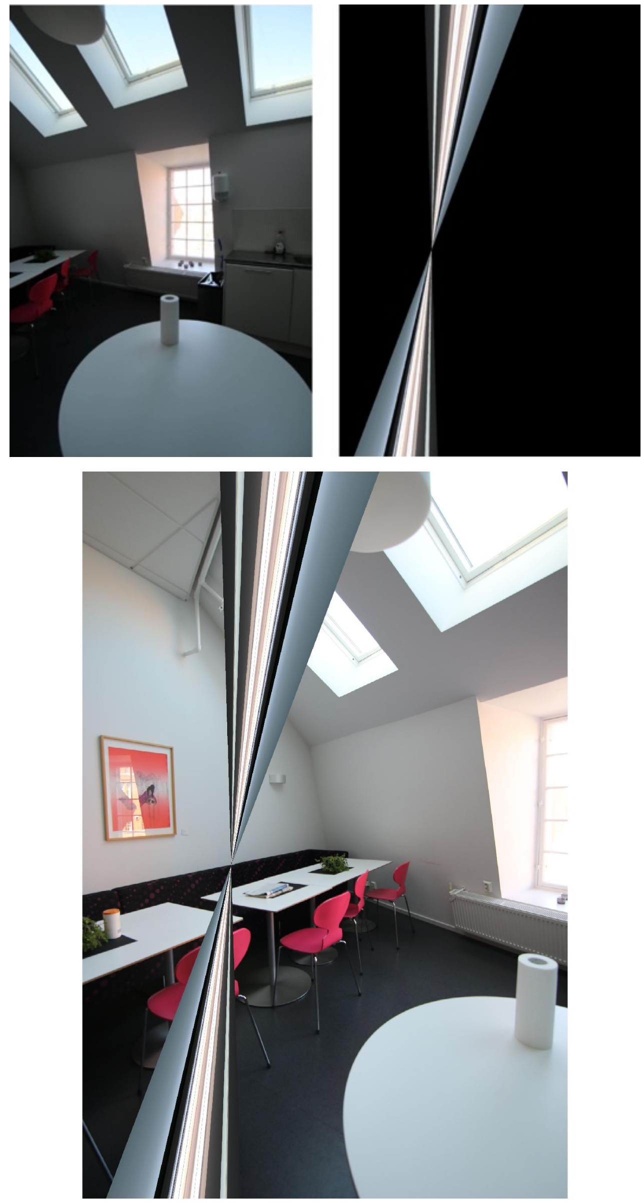


Figure A2. RANSAC-based image stitching algorithm as implemented in Matlab failed when merging image pair (7,13) of the *Lunch Room* dataset [6]. Top row, left: the original image 13; top row, right: image 13 after applying the (wrong) homography matrix estimated by Matlab's RANSAC-based estimateGeometricTransform algorithm; bottom row: failed stitching of image 13 and image 7 due to the incorrect estimation of the homography matrix.