Appendix

The appendix provides details on the construction of the cost and constraint matrices. In turn, the following subproblems are considered: The continuous subproblem, the discrete subproblem, and the combination of the two.

A. Continuous Problem

A homogenization variable $\mathbf{H} \in \{-\mathbf{1}_{2\times 2}, \mathbf{1}_{2\times 2}\}$, which can be enforced quadratically, is introduced to be able to write cost and constraint terms that are linear in $\mathcal{X}$ and $\boldsymbol{\theta}$. The impact of homogenization is discussed in [17]. To fix ideas, especially related to cost functions and constraints, $\mathbf{H} = \mathbf{1}_{2\times 2}$ can be assumed. The direction cosine matrix variables may be concatenated to yield $\mathbf{C} = \begin{bmatrix}\mathbf{C}_1 & \cdots & \mathbf{C}_i & \cdots & \mathbf{C}_N\end{bmatrix}$, while the position variables may be concatenated as $\mathbf{r} = \begin{bmatrix}\mathbf{r}_1 & \cdots & \mathbf{r}_i & \cdots & \mathbf{r}_N\end{bmatrix}$. The continuous problem variable is denoted $\boldsymbol{\Xi}$ and is obtained by horizontal concatenation of the robot states as well as of the homogenization variable

$$\boldsymbol{\Xi} = \begin{bmatrix}\mathbf{H} & \mathbf{C} & \mathbf{r}\end{bmatrix} \in \mathbb{R}^{2\times(2+3N)}. \tag{15}$$

A cost term that depends on the continuous variable may be expressed as $J_{\boldsymbol{\Xi}}(\boldsymbol{\Xi})$

$$J_{\boldsymbol{\Xi}}(\boldsymbol{\Xi}) = \left\langle \mathbf{Q}_{\boldsymbol{\Xi}}, \boldsymbol{\Xi}^\mathsf{T}\boldsymbol{\Xi} \right\rangle, \tag{16}$$

while a constraint may be expressed as

$$\left\langle \mathbf{A}_{\Xi}, \boldsymbol{\Xi}^\mathsf{T}\boldsymbol{\Xi} \right\rangle = b, \tag{17}$$

where $\mathbf{A}_{\Xi}$ is a matrix and b a scalar. While there are many cost terms and constraints in the problem, the cost term and constraint index is dropped for brevity of notation. The matrix inner product $\langle \cdot, \cdot \rangle$ is defined for matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ as $\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{tr}(\mathbf{A}\mathbf{B}^\mathsf{T}) = \sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}b_{ij}$. Importantly, the costs and constraints *can only be expressed using sums of dot products of the columns of* $\boldsymbol{\Xi}$ since, for a generic matrix $\mathbf{A}$,

$$\left\langle \mathbf{A}, \boldsymbol{\Xi}^\mathsf{T}\boldsymbol{\Xi} \right\rangle = \sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}(\boldsymbol{\Xi}^\mathsf{T}\boldsymbol{\Xi})_{ij} \tag{18}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}\boldsymbol{\xi}_i^\mathsf{T}\boldsymbol{\xi}_j. \tag{19}$$

It is worth discussing the main alternative to this approach, which consists of flattening the variable into a single column matrix $\boldsymbol{\xi} \in \mathbb{R}^{1+6N}$, containing all of the robot states and a scalar homogenization variable. This alternative leads to cost and constraint terms expressed using the quadratic expressions of the form $\boldsymbol{\xi}^\mathsf{T}\mathbf{A}\boldsymbol{\xi}$. This alternative leads to a larger problem dimension and can require a larger amount of redundant constraints on the relaxation. However, it is more flexible, since constraints and costs can be expressed using any scalar in the continuous variable $\boldsymbol{\Xi}$, unlike the approach used in this work where only dot products of the columns of $\boldsymbol{\Xi}$ may be used. A downstream consequence of only being able to use column dot products to express the cost is that the noise models used are less flexible.

1) Cost Matrix Construction: Cost terms in the continuous variable are expressed using the quadratic form

$$J_{\boldsymbol{\Xi}}(\boldsymbol{\Xi}) = \left\langle \mathbf{Q}_{\boldsymbol{\Xi}}, \boldsymbol{\Xi}^\mathsf{T}\boldsymbol{\Xi} \right\rangle. \tag{20}$$

For the prior and odometry terms, the entries of $\mathbf{Q}_{\boldsymbol{\Xi}}$ are inserted directly into corresponding entries of the overall cost matrix. For the unknown data associations, they are inserted into entries of the overall cost matrix that involve corresponding discrete variables.

2) Lifting a Pose to Known Landmark Measurement: The error is the same as for the unknown landmark position measurement,

$$e(\mathbf{C}_{ab}, \mathbf{r}; \boldsymbol{\ell}, \mathbf{y}) = \boldsymbol{\ell}^\mathsf{T}\boldsymbol{\ell} + \mathbf{r}^\mathsf{T}\mathbf{r} + \mathbf{y}^\mathsf{T}\cancel{\mathbf{C}^\mathsf{T}\mathbf{C}}\mathbf{y} + 2\mathbf{r}^\mathsf{T}\mathbf{C}\mathbf{y} - 2\mathbf{r}^\mathsf{T}\boldsymbol{\ell} - 2\boldsymbol{\ell}^\mathsf{T}\mathbf{C}\mathbf{y} \tag{21}$$

$$= \boldsymbol{\ell}^\mathsf{T}\boldsymbol{\ell} + \mathbf{r}^\mathsf{T}\mathbf{r} + \mathbf{y}^\mathsf{T}\mathbf{y} + 2\mathbf{r}^\mathsf{T}\mathbf{C}\mathbf{y} - 2\mathbf{r}^\mathsf{T}\boldsymbol{\ell} - 2\boldsymbol{\ell}^\mathsf{T}\mathbf{C}\mathbf{y}, \tag{22}$$

where subscripts are dropped for the sake of brevity. The last term bears a bit more consideration to include it in the matrix inner product formulation such that

$$\boldsymbol{\ell}^\mathsf{T}\mathbf{C}\mathbf{y} = \operatorname{tr}(\mathbf{C}\mathbf{y}\boldsymbol{\ell}^\mathsf{T}) \tag{23}$$

$$= \operatorname{tr}(\mathbf{C}(\boldsymbol{\ell}\mathbf{y}^\mathsf{T})^\mathsf{T}) \tag{24}$$

$$= \langle \mathbf{C}, \boldsymbol{\ell}\mathbf{y}^\mathsf{T} \rangle. \tag{25}$$

However, now the landmark $\boldsymbol{\ell}$ is known, such that the error is a function of $\mathbf{C}_{ab}, \mathbf{r}$ and parametrized using $\boldsymbol{\ell}, \mathbf{y}$, where $\mathbf{y}$ is the measured relative position of the landmark. Using the state $\begin{bmatrix}\mathbf{1} & \mathbf{C} & \mathbf{r}\end{bmatrix}$, the error may be written

$$
\begin{aligned}
e(\mathbf{C}_{ab}, \mathbf{r}; \boldsymbol{\ell}, \mathbf{y}) &= \boldsymbol{\ell}^\mathsf{T}\boldsymbol{\ell} + \mathbf{r}^\mathsf{T}\mathbf{r} + \mathbf{y}^\mathsf{T}\mathbf{y} + 2\mathbf{r}^\mathsf{T}\mathbf{C}\mathbf{y} - 2\mathbf{r}^\mathsf{T}\boldsymbol{\ell} - 2\boldsymbol{\ell}^\mathsf{T}\mathbf{C}\mathbf{y} && (26)\\
&= \left\langle \mathbf{Q}, \mathbf{X}^\mathsf{T}\mathbf{X} \right\rangle && (27)\\
&= \left\langle \mathbf{Q}, \begin{bmatrix} \mathbf{1} & \mathbf{C} & \mathbf{r} \\ & \mathbf{1} & \mathbf{C}^\mathsf{T}\mathbf{r} \\ & & \mathbf{r}^\mathsf{T}\mathbf{r} \end{bmatrix} \right\rangle && (28)\\
&= \left\langle \begin{bmatrix} \operatorname{diag}([(\boldsymbol{\ell}^\mathsf{T}\boldsymbol{\ell} + \mathbf{y}^\mathsf{T}\mathbf{y})\mathbf{e}_1]) & -2\boldsymbol{\ell}\mathbf{y}^\mathsf{T} & -2\boldsymbol{\ell} \\ & & 2\mathbf{y} \\ & & 1 \end{bmatrix}, \begin{bmatrix} \mathbf{1} & \mathbf{C} & \mathbf{r} \\ & \mathbf{1} & \mathbf{C}^\mathsf{T}\mathbf{r} \\ & & \mathbf{r}^\mathsf{T}\mathbf{r} \end{bmatrix} \right\rangle && (29)
\end{aligned}
$$

3) Lifting a Relative Pose Measurement: Given a relative pose measurement such that

$$
\begin{aligned}
\mathbf{T}_{k+1} &= \mathbf{T}_k \Delta\mathbf{T} && (30)\\
\begin{bmatrix} \mathbf{C}_{k+1} & \mathbf{r}_{k+1} \\ \mathbf{0} & 1 \end{bmatrix} &= \begin{bmatrix} \mathbf{C}_k\Delta\mathbf{C} & \mathbf{r}_k + \mathbf{C}_k\Delta\mathbf{r} \\ \mathbf{0} & 1 \end{bmatrix}, && (31)
\end{aligned}
$$

the first error corresponds to the translation error and is given by

$$
\begin{aligned}
J_r &= \|\mathbf{r}_{k+1} - \mathbf{r}_k - \mathbf{C}_k\Delta\mathbf{r}\|_2^2 && (32)\\
&= (\mathbf{r}_{k+1} - \mathbf{r}_k - \mathbf{C}_k\Delta\mathbf{r})^\mathsf{T}(\mathbf{r}_{k+1} - \mathbf{r}_k - \mathbf{C}_k\Delta\mathbf{r}) && (33)\\
&= \mathbf{r}_{k+1}^\mathsf{T}\mathbf{r}_{k+1} + \mathbf{r}_k^\mathsf{T}\mathbf{r}_k + \Delta\mathbf{r}^\mathsf{T}\Delta\mathbf{r} + 2\mathbf{r}_k^\mathsf{T}\mathbf{C}_k\Delta\mathbf{r} - 2\mathbf{r}_{k+1}^\mathsf{T}\mathbf{C}_k\Delta\mathbf{r} - 2\mathbf{r}_{k+1}\mathbf{r}_k. && (34)
\end{aligned}
$$

For $\mathbf{X} = \begin{bmatrix}\mathbf{C}_k & \mathbf{r}_k & \mathbf{r}_{k+1}\end{bmatrix}$, this yields the following cost,

$$
\begin{aligned}
J_r &= \left\langle \mathbf{Q}, \mathbf{X}^\mathsf{T}\mathbf{X} \right\rangle && (35)\\
&= \left\langle \mathbf{Q}, \begin{bmatrix} \mathbf{C}_k^\mathsf{T} \\ \mathbf{r}_k^\mathsf{T} \\ \mathbf{r}_{k+1} \end{bmatrix} \begin{bmatrix}\mathbf{C}_k & \mathbf{r}_k & \mathbf{r}_{k+1}\end{bmatrix} \right\rangle && (36)\\
&= \left\langle \mathbf{Q}, \begin{bmatrix} \mathbf{C}_k^\mathsf{T}\mathbf{C}_k & \mathbf{C}_k^\mathsf{T}\mathbf{r}_k & \mathbf{C}_k^\mathsf{T}\mathbf{r}_{k+1} \\ \mathbf{r}_k^\mathsf{T}\mathbf{C}_k & \mathbf{r}_k^\mathsf{T}\mathbf{r}_k & \mathbf{r}_k^\mathsf{T}\mathbf{r}_{k+1} \\ \mathbf{r}_{k+1}^\mathsf{T}\mathbf{C}_k & \mathbf{r}_{k+1}^\mathsf{T}\mathbf{r}_k & \mathbf{r}_{k+1}^\mathsf{T}\mathbf{r}_{k+1} \end{bmatrix} \right\rangle && (37)\\
&= \left\langle \begin{bmatrix} \mathbf{0} & 2\Delta\mathbf{r} & -2\Delta\mathbf{r} \\ & 1 & -2 \\ & & 1 \end{bmatrix}, \begin{bmatrix} \mathbf{C}_k & \mathbf{r}_k & \mathbf{r}_{k+1} \\ \mathbf{r}_k^\mathsf{T}\mathbf{C}_k & \mathbf{r}_k^\mathsf{T}\mathbf{r}_k & \mathbf{r}_k^\mathsf{T}\mathbf{r}_{k+1} \\ \mathbf{r}_{k+1}^\mathsf{T}\mathbf{C}_k & \mathbf{r}_{k+1}^\mathsf{T}\mathbf{r}_k & \mathbf{r}_{k+1}^\mathsf{T}\mathbf{r}_{k+1} \end{bmatrix} \right\rangle. && (38)
\end{aligned}
$$

The second error corresponds to the rotation error and is given by

$$
\begin{aligned}
J_c &= \|\mathbf{C}_{k+1} - \mathbf{C}_k\Delta\mathbf{C}\|_F^2 && (39)\\
&= \operatorname{tr}\left((\mathbf{C}_{k+1} - \mathbf{C}_k\Delta\mathbf{C})\,(\mathbf{C}_{k+1} - \mathbf{C}_k\Delta\mathbf{C})^\mathsf{T} \right) && (40)\\
&= 2\operatorname{tr}(\mathbf{1}) - 2\operatorname{tr}(\mathbf{C}_{k+1}^\mathsf{T}\mathbf{C}_k\Delta\mathbf{C}) && (41)\\
&= 2\operatorname{tr}(\mathbf{1} - \Delta\mathbf{C}\mathbf{C}_{k+1}^\mathsf{T}\mathbf{C}_k) && (42)\\
& && (43)\\
& && (44)
\end{aligned}
$$

Making use of $\operatorname{tr}(\mathbf{A}\mathbf{B}^\mathsf{T}) = \sum_i\sum_j \mathbf{A}_{ij}\mathbf{B}_{ij} = \langle \mathbf{A}, \mathbf{B} \rangle$ allows to write

$$
\begin{aligned}
J_c &= 2\operatorname{tr}(\mathbf{1} - \Delta\mathbf{C}\mathbf{C}_{k+1}^\mathsf{T}\mathbf{C}_k) && (45)\\
&= 2\left(\operatorname{tr}\mathbf{1} - \left\langle \Delta\mathbf{C}, \mathbf{C}_k^\mathsf{T}\mathbf{C}_{k+1} \right\rangle\right) && (46)\\
&= \left\langle \mathbf{Q}, \begin{bmatrix} \mathbf{1} \\ \mathbf{C}_k^\mathsf{T} \\ \mathbf{C}_{k+1}^\mathsf{T} \end{bmatrix} \begin{bmatrix}\mathbf{1} & \mathbf{C}_k & \mathbf{C}_{k+1}\end{bmatrix} \right\rangle && (47)\\
&= \left\langle \begin{bmatrix} 2\mathbf{1} & & \\ & \mathbf{0} & -2\Delta\mathbf{C} \\ & & \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{1} & \mathbf{C}_k & \mathbf{C}_{k+1} \\ \mathbf{C}_k^\mathsf{T} & \mathbf{C}_k^\mathsf{T}\mathbf{C}_k & \mathbf{C}_k^\mathsf{T}\mathbf{C}_{k+1} \\ \mathbf{C}_{k+1}^\mathsf{T} & \mathbf{C}_{k+1}^\mathsf{T}\mathbf{C}_k & \mathbf{C}_{k+1}^\mathsf{T}\mathbf{C}_{k+1} \end{bmatrix} \right\rangle. && (48)
\end{aligned}
$$

4) Lifting a Prior: Given a prior such that the error is given by

$$J_p = \|\mathbf{x} - \check{\mathbf{x}}\|_2^2 \tag{49}$$
$$= (\mathbf{x} - \check{\mathbf{x}})^\mathsf{T}(\mathbf{x} - \check{\mathbf{x}}) \tag{50}$$
$$= \mathbf{x}^\mathsf{T}\mathbf{x} - 2\check{\mathbf{x}}^\mathsf{T}\mathbf{x} + \check{\mathbf{x}}^\mathsf{T}\check{\mathbf{x}} \tag{51}$$

for $\mathbf{X} = \begin{bmatrix}\mathbf{1} & \mathbf{x}\end{bmatrix}$, this yields the following cost

$$J_p = \mathbf{x}^\mathsf{T}\mathbf{x} - 2\check{\mathbf{x}}^\mathsf{T}\mathbf{x} + \check{\mathbf{x}}^\mathsf{T}\check{\mathbf{x}} \tag{52}$$
$$= \langle \mathbf{Q}, \mathbf{X}^\mathsf{T}\mathbf{X}\rangle \tag{53}$$
$$= \left\langle \begin{bmatrix}\check{\mathbf{x}}^\mathsf{T}\check{\mathbf{x}} & -2\check{\mathbf{x}}^\mathsf{T} \\ & 1\end{bmatrix}, \begin{bmatrix}\mathbf{1} & \mathbf{x} \\ \mathbf{x} & \mathbf{x}^\mathsf{T}\mathbf{x}\end{bmatrix}\right\rangle. \tag{54}$$

5) Constraint Matrix Definition: The continuous variable constraints enforce orthonormality of the direction cosine matrix (DCMs) in the relaxation. For brevity, the constraints are given for the specific subvariables considered. In the overall problem, they are then inserted directly into the entries of the overall constraint matrix corresponding the subvariables. Symmetric constraint matrices are used in SDPs. Any non-symmetric constraint matrix may be symmetrized as $\mathbf{A}_{\text{sym}} = (\mathbf{A} + \mathbf{A}^\mathsf{T})/2$. Given a DCM $\mathbf{C}$, the continuous variable constraints are of the form

$$\left\langle \mathbf{A}, \begin{bmatrix}\mathbf{H} & \mathbf{C} \\ \mathbf{C}^\mathsf{T} & \mathbf{C}^\mathsf{T}\mathbf{C}\end{bmatrix}\right\rangle = 0, \tag{55}$$

where the homogenization variable $\mathbf{H}$ can be thought of as identity. The matrix $\mathbf{A}$ is used as a generic constraint matrix in this section. The continuous constraint matrices then correspond to

1) Orthonormality. For each column index combination $i, j \in (1,2) \times (1,2)$, a constraint matrix is added with the following structure.
   - If $i \neq j$, constraint matrix $\mathbf{A}$ has $\mathbf{A}[\text{slice}(\mathbf{c}_1), \text{slice}(\mathbf{c}_2)] = 1$ corresponding to orthogonality of different columns of $\mathbf{C}$.
   - If $i = j$, constraint matrix $\mathbf{A}$ has $\mathbf{A}[\text{slice}(\mathbf{c}_1), \text{slice}(\mathbf{c}_2)] = 1$, and $\mathbf{A}[\text{slice}(\mathbf{h}_1), \text{slice}(\mathbf{h}_1)] = -1$, corresponding to unit norm for each column of $\mathbf{C}$.
2) Structure of two-dimensional DCM. For the planar case, the DCM has the structure $\mathbf{C} = \begin{bmatrix} c & -s \\ s & c\end{bmatrix}$. Two additional constraints may then be added,
   - Constraint matrix $\mathbf{A}$ with $\mathbf{A}[\text{slice}(\mathbf{c}_1), \text{slice}(\mathbf{h}_1)] = 1$, $\mathbf{A}[\text{slice}(\mathbf{c}_2), \text{slice}(\mathbf{h}_2)] = -1$,
   - Constraint matrix $\mathbf{A}$ with $\mathbf{A}[\text{slice}(\mathbf{c}_1), \text{slice}(\mathbf{h}_2)] = 1$, $\mathbf{A}[\text{slice}(\mathbf{c}_2), \text{slice}(\mathbf{h}_1)] = 1$.

B. Discrete Problem

The discrete variables arise from the unknown data association cost given by

$$J_{\text{uda, lndmrk}}(\mathcal{X}) = \sum_{i=1}^{N}\sum_{k_i=1}^{K_i}\sum_{j=1}^{n_\ell} \theta_{ikj} r(\mathbf{C}_i, \mathbf{r}_i; \boldsymbol{\ell}_j, \mathbf{y}_{k_i}). \tag{56}$$

The outer loop is over the time indices i. The middle loop is over the k'th received measurement at timestep i, $\mathbf{y}_{k_i}$. The inner loop is over the possible data associations. The discrete variables are constrained to be boolean, $\theta_{ik_ij} \in \{0, 1\}$. Furthermore, they are constrained by

$$\sum_{j=1}^{n_\ell} \theta_{ikj} = 1, \quad \forall i, k. \tag{57}$$

meaning that every measurement comes from a single landmark. The overall problem variable involves both discrete and continuous variables, and requires redundant constraints. These constraints in turn require quadratic constraints on the discrete variables of the form

$$\begin{bmatrix}1 & \boldsymbol{\theta}^\mathsf{T}\end{bmatrix} \mathbf{A}_\theta \begin{bmatrix}1 \\ \boldsymbol{\theta}\end{bmatrix} = 0. \tag{58}$$

Constraint matrices $\mathbf{A}_\theta$ may be constructed from the following observations. Since they are used to construct *redundant* constraints, it is insufficient to only use the boolean and sum constraints, and a few more are required.

- $\sum_{j=1}^{n_\ell} \theta_{i,k,j} = 1$. This corresponds to $\mathbf{A}_\theta$ with $\mathbf{A}[\text{slice}(1), \text{slice}(\theta_{i,k,j})] = 1$ for every j, as well as $\mathbf{A}[\text{slice}(1), \text{slice}(1)] = -1$.
- $\theta_{i,k,j}^2 - \theta_{i,k,j} = 0$. This corresponds to $\mathbf{A}_\theta$ with $\mathbf{A}[\text{slice}(\theta_{i,k,j}), \text{slice}(\theta_{i,k,j})] = 1$, $\mathbf{A}[\text{slice}(1), \text{slice}(\theta_{i,k,j})] = -1$.

- $\theta_{i,k,j_1}\theta_{i,k,j_2} = 0$, for $j_1 \neq j_2$. This corresponds to $\mathbf{A}_\theta$ with $\mathbf{A}[\text{slice}(\theta_{i,k,j_1}), \text{slice}(\theta_{i,k,j_2})] = 1$.
- $\theta_{i,k_2,j_2} \sum_{j=1}^{n_\ell} \theta_{i,k_1,j} - \theta_{i,k_2,j_2} = 0$, for $k_1 \neq k_2$, one constraint for each applicable j_2. This is a premultiplication of the sum constraint for one unknown data association measurement with discrete variables corresponding to other unknown data association measurements, from the same timestep. This corresponds to $\mathbf{A}_\theta$ constraints, one for each $k_1 \neq k_2$ and applicable j_2. For given k_1, k_2, j_2, the constraint matrix $\mathbf{A}_\theta$ has $\mathbf{A}[\text{slice}(\theta_{i,k_2,j_2}), \text{slice}(\theta_{i,k_1,j})] = 1$ and $\mathbf{A}[\text{slice}(\theta_{i,k_2,j_2}), \text{slice}(1)] = -1$.

C. Combination of Discrete and Continuous Problems

The overall problem contains both discrete and continuous variables with corresponding costs and constraints. The optimization variable of interest is given by

$$\mathbf{X} = \begin{bmatrix}\mathbf{H} & \boldsymbol{\theta}^\mathsf{T} \otimes \mathbf{1}^{2\times 2} & \boldsymbol{\theta}^\mathsf{T} \otimes \boldsymbol{\Xi} & \boldsymbol{\Xi}\end{bmatrix}. \tag{59}$$

The localization problem is expressed as

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \sum_{i=1}^{n_\theta} \left\langle \mathbf{X}^\mathsf{T}\mathbf{X}, \mathbf{Q} \right\rangle \\ \text{subject to} \quad & \mathbf{X} \in \{-\mathbf{1}, \mathbf{1}\} \times \{\mathbf{0}, \mathbf{1}\}^{n_\theta} \times \\ & \quad \left(SO(2)^N \times (\mathbb{R}^2)^N\right)^{(1+n_\theta)}. \end{aligned} \tag{Reform. Prob}$$

The cost matrix $\mathbf{Q}$ is the overall cost matrix for the problem. The loss may be split into two parts, one solely depending on the continuous robot states $\mathcal{X}$, and one depending on both $\mathcal{X}$ and the discrete variables $\boldsymbol{\theta}$.

$$J(\mathcal{X}, \boldsymbol{\theta}) = J_1(\mathcal{X}) + J_2(\mathcal{X}, \boldsymbol{\theta}). \tag{60}$$

The part of the loss $J_1(\mathcal{X})$ depending solely on $\mathcal{X}$ may be written as

$$J_1(\mathcal{X}) = \sum_{i=1}^{n_{\text{f, cont}}} \left\langle \mathbf{Q}_{\boldsymbol{\Xi},i}, \boldsymbol{\Xi}^\mathsf{T}\boldsymbol{\Xi} \right\rangle, \tag{61}$$

where $n_{\text{f, cont}}$ is the number of error terms that depend solely on the continuous variable. The continuous variable cost matrices $\mathbf{Q}_{\boldsymbol{\Xi},i}$ are inserted directly into the corresponding entries of the overall cost matrix $\mathbf{Q}$. The discrete-variable dependent loss $J_2(\mathcal{X}, \boldsymbol{\theta})$ takes the form

$$J_2(\mathcal{X}, \boldsymbol{\theta}) = \sum_{i=1}^{n_\theta} \theta_i \left\langle \mathbf{Q}_{\boldsymbol{\Xi},i}, \boldsymbol{\Xi}^\mathsf{T}\boldsymbol{\Xi} \right\rangle, \tag{62}$$

where i now loops over the discrete variables, each of which has an associated relative landmark measurement error expressed using $\left\langle \mathbf{Q}_{\boldsymbol{\Xi},i}, \boldsymbol{\Xi}^\mathsf{T}\boldsymbol{\Xi} \right\rangle$. Each summand may be written as

$$\theta_i \left\langle \mathbf{Q}_{\boldsymbol{\Xi},i}, \boldsymbol{\Xi}^\mathsf{T}\boldsymbol{\Xi} \right\rangle = \left\langle \mathbf{Q}_{\boldsymbol{\Xi},i}, \boldsymbol{\Xi}^\mathsf{T}\left(\theta_i \boldsymbol{\Xi}\right) \right\rangle. \tag{63}$$

Therefore, for each pair of columns $\boldsymbol{\xi}_j, \boldsymbol{\xi}_k$ of $\boldsymbol{\Xi}$, the corresponding parts of the overall cost matrix $\mathbf{Q}$ may be set as $\mathbf{Q}[\text{slice}(\boldsymbol{\xi}_j), \text{slice}(\theta_i \boldsymbol{\xi}_k)] = \mathbf{Q}_{\boldsymbol{\Xi},i}[\text{slice}(\boldsymbol{\xi}_j), \text{slice}(\boldsymbol{\xi}_k)]$. The discrete variable constraints of Sec. B may be imposed on the overall optimization variable $\mathbf{X}$ by creating a constraint matrix $\mathbf{A}$ with $\mathbf{A}[\text{slice}(\boldsymbol{\Theta}_{i,1}), \text{slice}(\boldsymbol{\Theta}_{j,1})] = \mathbf{A}_{\boldsymbol{\Theta}}[\text{slice}(\theta_i), \text{slice}(\theta_j)]$. The column $\boldsymbol{\Theta}_{i,1}$ corresponds to the first column of the $\theta_i \mathbf{1}$ matrix present in the $\boldsymbol{\theta}^\mathsf{T} \otimes \mathbf{1}^{2\times 2}$ section of the optimization variable (64). Similarly, the column $\boldsymbol{\Theta}_{j,1}$ corresponds to the first column of the $\theta_j \mathbf{1}$ matrix present in the $\boldsymbol{\theta}^\mathsf{T} \otimes \mathbf{1}^{2\times 2}$ section of the optimization variable (64). Furthermore, redundant constraints are created by combining the discrete and continuous variable constraints, the details for which are present in the paper itself.

D. Redundant Constraints Corresponding to Problem Structure

1) Moment Constraints: Moment constraints are added that reflect the particular matrix structure. For the moment constraint presentation only, $\theta_{i,1}$ and $\theta_{i,2}$ are used to refer to the first and second columns of the matrix subblock of $\mathbf{X}$ that corresponds to θ_i.

- Moment constraint form 1
  - $\mathbf{A}[\text{slice}_\mathbf{X}(\mathbf{H}_i), \text{slice}_\mathbf{X}(\theta_i \boldsymbol{\xi}_k)] = 1$,
  - $\mathbf{A}[\text{slice}_\mathbf{X}(\theta_i), \text{slice}_\mathbf{X}(\boldsymbol{\xi}_k)] = -1$.
- Moment constraint form 2
  - $\mathbf{A}[\text{slice}_\mathbf{X}(\theta_{i,1} \boldsymbol{\xi}_k), \text{slice}_\mathbf{X}(\theta_{i,2})] = 1$,
  - $\mathbf{A}[\text{slice}_\mathbf{X}(\theta_{i,2} \boldsymbol{\xi}_k), \text{slice}_\mathbf{X}(\theta_{i,1})] = -1$.

- Moment constraint form 3
  - $\mathbf{A}[\text{slice}_{\mathbf{X}}(\theta_{i,1}\boldsymbol{\xi}_{k,1}), \text{slice}_{\mathbf{X}}(\theta_{i,2}\boldsymbol{\xi}_{k,2})] = 1$,
  - $\mathbf{A}[\text{slice}_{\mathbf{X}}(\theta_{i,1}\boldsymbol{\xi}_{k,2}), \text{slice}_{\mathbf{X}}(\theta_{i,2}\boldsymbol{\xi}_{k,1})] = -1$.

2) Column Structure of Optimization Variable: Due to the matrix structure of the optimization variable,

$$\mathbf{X} = \begin{bmatrix}\mathbf{H} & \boldsymbol{\theta}^{\mathsf{T}} \otimes \mathbf{1}^{2\times 2} & \boldsymbol{\theta}^{\mathsf{T}} \otimes \boldsymbol{\Xi} & \boldsymbol{\Xi}\end{bmatrix}, \tag{64}$$

where diagonal matrices are used to encode the discrete variables, extra redundant constraints need to be imposed on the relaxation. Let θ_{i,d_1} denote the matrix column name corresponding to the d_1'th column of the diagonal matrix corresponding to θ_i. Similarly, let θ_{j,d_2} denote the matrix column name corresponding to the d_2'th column of the diagonal matrix corresponding to the discrete variable θ_j, with $d_1 \neq d_2$. Then, a redundant constraint is given by $\mathbf{Z}[\text{slice}_{\mathbf{X}}(\theta_{i,d_1}), \text{slice}_{\mathbf{X}}(\theta_{j,d_2})] = 0$. This redundant constraint may be encoded using a quadratic constraint as $\mathbf{A}[\text{slice}_{\mathbf{X}}(\theta_{i,d_1}), \text{slice}_{\mathbf{X}}(\theta_{j,d_2})] = 1$. Furthermore, additional constraints are used to impose the diagonal structure of the blocks of the continuous variable (8) that are diagonal, in the same fashion as in [13].