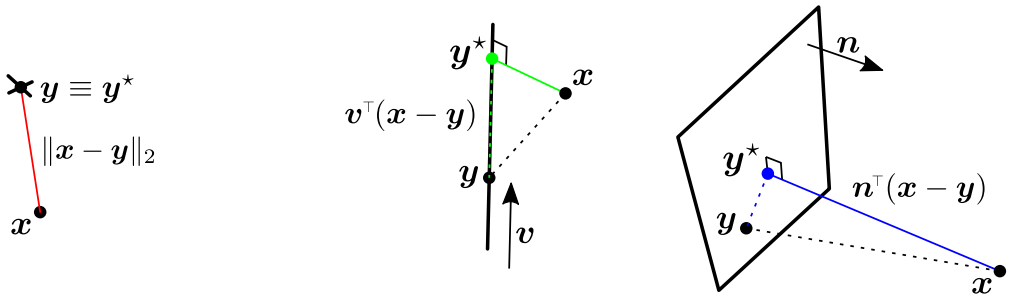
Supplementary material for
Convex Global 3D Registration with Lagrangian Duality

Jesus Briales Javier Gonzalez-Jimenez
MAPIR-UMA Group
University of Malaga, Spain
{jesusbriales,javiergonzalez}@uma.es

Contents

Table 1. Symbols used in the paper

Matrix operators	
$\text{vec}(\cdot)$	Column-wise vectorization
$\otimes$	Matrix Kronecker product
Problem variables	
$\boldsymbol{r} \in \mathbb{R}^9$	Column-wise vectorization of $\boldsymbol{R} \in \text{SO}(3)$
$\boldsymbol{\tau} \in \mathbb{R}^{12}$	Vectorization of $\boldsymbol{T} \in \text{SO}(3) \times \mathbb{R}^3$
Homogeneous variables	
$\tilde{\boldsymbol{a}} \equiv [\boldsymbol{a}; 1]/[\boldsymbol{a}; y]/[\boldsymbol{a}; \gamma]$	Augmented homogeneous vector
$\tilde{\boldsymbol{A}} \equiv \begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}^\top & c \end{bmatrix}$	Augmented matrix (for hom. coordinate)



(a) Point-to-point distance (b) Point-to-line distance (c) Point-to-plane distance

Figure 1. Given a 3D point $\boldsymbol{x}$, the distance to the closest point $\boldsymbol{y}^\star$ in the primitive is readily obtained using simple algebra for the case of a point (a), a line (b) or a plane (c).

1. Generalized distance function

This section provides the expression for the closest distance of a point to the different primitives considered in the paper (see Fig. 1), namely the square distance from a 3D point $\boldsymbol{x}$ to a 3D primitive P is typically defined by

$$\mathrm{d}_P^2(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{y}^\star\|_2^2 = \min_{\boldsymbol{y}' \in P} \|\boldsymbol{x} - \boldsymbol{y}'\|_2^2. \tag{1}$$

The considered primitives have simple closed-form solutions to (1) that are used in each case to characterize the parameters of a generalized distance function of the form

$$\mathrm{d}_P(\boldsymbol{x})^2 = \|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{C}}^2 = (\boldsymbol{x} - \boldsymbol{y})^\top \boldsymbol{C} (\boldsymbol{x} - \boldsymbol{y}). \tag{2}$$

A summary is available in Tab. 2.

Table 2. Generalized parameters (2) for each correspondence

Correspondence	$\boldsymbol{y}$	$\boldsymbol{C}$
Point-to-point	The point primitive	$\boldsymbol{I}_3$
Point-to-line	Any point in the line	$\boldsymbol{I}_3 - \boldsymbol{v}\boldsymbol{v}^\top$
Point-to-plane	Any point in the plane	$\boldsymbol{n}\boldsymbol{n}^\top$

1.1. Point-to-point

There is only one point, so the closest point $\boldsymbol{y}^\star$ is the proper primitive (Fig. 1(a)), and this is a trivial case with

$$\mathrm{d}_{\text{point}}^2(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 = (\boldsymbol{x} - \boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y}). \tag{3}$$

This is a particularization of the generalized metric (2) with

$$\boldsymbol{C} = \boldsymbol{I}_3. \tag{4}$$

1.2. Point-to-line

Given any point $\boldsymbol{y}$ in the line and its direction described by the unit vector $\boldsymbol{v}$, the signed distance from $\boldsymbol{y}$ to the point $\boldsymbol{y}^\star$ closest to $\boldsymbol{x}$ is $\boldsymbol{v}^\top(\boldsymbol{x} - \boldsymbol{y})$ (see Fig. 1(b)). As a result the seeked distance can be computed as

$$\begin{aligned}
\mathrm{d}_{\text{line}}^2(\boldsymbol{x}) &= \|\boldsymbol{x} - \boldsymbol{y}^\star\|_2^2 && (5)\\
&= \|\boldsymbol{x} - (\boldsymbol{y} + (\boldsymbol{v}^\top(\boldsymbol{x} - \boldsymbol{y}))\boldsymbol{v})\|_2^2 && (6)\\
&= \|(\boldsymbol{x} - \boldsymbol{y}) - \boldsymbol{v}\boldsymbol{v}^\top(\boldsymbol{x} - \boldsymbol{y}))\|_2^2 && (7)\\
&= \|(\boldsymbol{I}_3 - \boldsymbol{v}\boldsymbol{v}^\top)(\boldsymbol{x} - \boldsymbol{y})\|_2^2. && (8)
\end{aligned}$$

The matrix $\boldsymbol{I} - \boldsymbol{v}\boldsymbol{v}^\top$ is the orthogonal projection onto the line. An orthogonal projection matrix is symmetric,

$$\boldsymbol{I} - \boldsymbol{v}\boldsymbol{v}^\top = (\boldsymbol{I} - \boldsymbol{v}\boldsymbol{v}^\top)^\top \tag{9}$$

and idempotent,

$$(\boldsymbol{I} - \boldsymbol{v}\boldsymbol{v}^\top)^2 = \boldsymbol{I} - \boldsymbol{v}\boldsymbol{v}^\top. \tag{10}$$

As a result, the distance to the line (8) can be refactored as

$$\begin{aligned} \mathrm{d}^2_{\text{line}}(\boldsymbol{x}) &= (\boldsymbol{x}-\boldsymbol{y})^\top (\boldsymbol{I} - \boldsymbol{v}\boldsymbol{v}^\top)^\top (\boldsymbol{I} - \boldsymbol{v}\boldsymbol{v}^\top)(\boldsymbol{x}-\boldsymbol{y}) \\ &= (\boldsymbol{x}-\boldsymbol{y})^\top (\boldsymbol{I} - \boldsymbol{v}\boldsymbol{v}^\top)(\boldsymbol{x}-\boldsymbol{y}), \end{aligned} \tag{11}$$

and comparing to the generalized distance (2) this settles for the line case

$$\boldsymbol{C} = \boldsymbol{I} - \boldsymbol{v}\boldsymbol{v}^\top. \tag{12}$$

1.3. Point-to-plane

Given any point $\boldsymbol{y}$ in the plane and its unit normal vector $\boldsymbol{n}$, it is well known from basic geometry that the signed distance from $\boldsymbol{x}$ to the plane is $\boldsymbol{n}^\top(\boldsymbol{x}-\boldsymbol{y})$ (see Fig. 1(c)). The square of this distance yields

$$\mathrm{d}^2_{\text{plane}}(\boldsymbol{x}) = (\boldsymbol{n}^\top(\boldsymbol{x}-\boldsymbol{y}))_2^2 \tag{13}$$

$$= (\boldsymbol{x}-\boldsymbol{y})^\top (\boldsymbol{n}\boldsymbol{n}^\top)(\boldsymbol{x}-\boldsymbol{y}), \tag{14}$$

so generalized distance (2) for the plane case is given by

$$\boldsymbol{C} = \boldsymbol{n}\boldsymbol{n}^\top. \tag{15}$$

1.4. Discussion

Note that even though we have used here the 2-norm for the closest distance (1), the considered generalized distance could also cover other extended norms such as $\|\boldsymbol{x} - \boldsymbol{y}^\star\|_{\boldsymbol{A}}$, with $\boldsymbol{A}$ a positive semidefinite matrix condensing uncertainty weights.

2. Linear transformation operator

The Euclidean transformation of a point $\boldsymbol{x}$ by $\boldsymbol{T}$ into $\boldsymbol{x}'$ is usually referred to symbolically as the operation $\boldsymbol{T} \oplus \boldsymbol{x}$. Depending on the specific parameterization chosen for the transformation $\boldsymbol{T} = (\boldsymbol{R}, \boldsymbol{t})$, this is a non-linear operator w.r.t. $\boldsymbol{T}$. However, if we choose a matrix representation for the rotation $\boldsymbol{R}$, the expression of the transformed point $\boldsymbol{x}'$ is linear in the elements of $\boldsymbol{R}$ and $\boldsymbol{t}$:

$$\boldsymbol{x}' = \boldsymbol{T} \oplus \boldsymbol{x} = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{t} \tag{16}$$

If we take a different perspective on the operator, this can be seen as the mapping of the transformation $(\boldsymbol{R}, \boldsymbol{t})$ into a point $\boldsymbol{x}'$ according to a fixed $\boldsymbol{x}$:

$$(\cdot) \oplus \boldsymbol{x} : \mathrm{SE}(3) \to \mathbb{R}^3 \tag{17}$$

$$\boldsymbol{T} \oplus \boldsymbol{x} \to \boldsymbol{x}' \tag{18}$$

It will be convenient for the formulation exploited in the paper to write this linear operator in a matrix form with the parameters in the transformation given in vector form:

$$(\cdot) \oplus \boldsymbol{x} : \mathrm{vec}(\mathrm{SE}(3)) \equiv \mathbb{R}^{12} \to \mathbb{R}^3 \tag{19}$$

Using the properties of the vectorization operator $\mathrm{vec}(\cdot)$ [2], namely $\mathrm{vec}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = (\boldsymbol{B}^\top \otimes \boldsymbol{A})\,\mathrm{vec}(\boldsymbol{X})$, we see that

$$\boldsymbol{R}\boldsymbol{x} = \mathrm{vec}(\boldsymbol{R}\boldsymbol{x}) = (\boldsymbol{x}^\top \otimes \boldsymbol{I}_3)\,\mathrm{vec}(\boldsymbol{R}). \tag{20}$$

Thus, with some simple block algebra the linear operator is

$$\boldsymbol{T} \oplus \boldsymbol{x} = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{t} = \underbrace{\begin{bmatrix}\boldsymbol{x}^\top \otimes \boldsymbol{I}_3 & \boldsymbol{I}_3\end{bmatrix}}_{\tilde{\boldsymbol{x}}^\top \otimes \boldsymbol{I}_3} \underbrace{\begin{bmatrix}\mathrm{vec}(\boldsymbol{R}) \\ \boldsymbol{t}\end{bmatrix}}_{\boldsymbol{\tau}}. \tag{21}$$

Here $\tilde{\boldsymbol{x}}^\top \otimes \boldsymbol{I}_3$ becomes the matrix defining the linear operator, with $\tilde{\boldsymbol{x}} = \begin{bmatrix}\boldsymbol{x}^\top, 1\end{bmatrix}^\top$ the homogeneous augmented version of $\boldsymbol{x}$, whereas $\boldsymbol{\tau}$ stands for a vectorization of the variables defining the transformation, $\boldsymbol{\tau} \equiv \mathrm{vec}(\boldsymbol{T})$.

3. Marginalization in quadratic forms

This section provides a detailed overview on how the minimization of a quadratic objective can be marginalized w.r.t. the unconstrained variables. In our particular case the problem is

$$f^\star = \min_{\substack{\boldsymbol{R}\in \mathrm{SO}(3)\\ \boldsymbol{t}_i\in\mathbb{R}^3}} \tilde{\boldsymbol{\tau}}^\top \tilde{\boldsymbol{M}} \tilde{\boldsymbol{\tau}}, \quad \tilde{\boldsymbol{\tau}} = \left[\mathrm{vec}(\boldsymbol{R})^\top, \boldsymbol{t}^\top, 1\right]^\top \tag{22}$$

From the method of Lagrange multipliers, the necessary conditions for a local minimum (and hence for a global mimum $(\boldsymbol{R}^\star, \boldsymbol{t}^\star)$) is that the gradient of the Lagrangian w.r.t. all the parameters vanishes:

$$\nabla \mathrm{L}(\boldsymbol{R}, \boldsymbol{t}, \boldsymbol{\lambda}) = \boldsymbol{0}. \tag{23}$$

For our problem (22) the Lagrangian takes the form

$$\mathrm{L}(\boldsymbol{R}, \boldsymbol{t}, \boldsymbol{\lambda}) = f(\boldsymbol{R}, \boldsymbol{t}) + \boldsymbol{\lambda} \cdot \boldsymbol{c}_{\mathrm{SO}(3)}(\boldsymbol{R}), \tag{24}$$

that is, the constraints in the problem affect only the rotation and are abstracted to some set of functions $\boldsymbol{c}_{\mathrm{SO}(3)}(\boldsymbol{R})$. These constraints are independent of $\boldsymbol{t}$ and taking the gradient w.r.t. $\boldsymbol{t}$ we get

$$\frac{\partial \mathrm{L}(\boldsymbol{R}, \boldsymbol{t}, \boldsymbol{\lambda})}{\partial \boldsymbol{t}} = \frac{\partial f(\tilde{\boldsymbol{\tau}})}{\partial \boldsymbol{t}} = \boldsymbol{0}_{3\times 1}. \tag{25}$$

Let us reorder the quadratic form as

$$f(\tilde{\boldsymbol{\tau}}) = \begin{bmatrix} !\boldsymbol{t} \\ \boldsymbol{t} \end{bmatrix}^\top \begin{bmatrix} \tilde{\boldsymbol{M}}_{!\boldsymbol{t},!\boldsymbol{t}} & \tilde{\boldsymbol{M}}_{!\boldsymbol{t},\boldsymbol{t}} \\ \tilde{\boldsymbol{M}}_{\boldsymbol{t},!\boldsymbol{t}} & \boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}} \end{bmatrix} \begin{bmatrix} !\boldsymbol{t} \\ \boldsymbol{t} \end{bmatrix} \tag{26}$$

$$= !\boldsymbol{t}^\top \tilde{\boldsymbol{M}}_{!\boldsymbol{t},!\boldsymbol{t}}\, !\boldsymbol{t} + 2\boldsymbol{t}^\top \tilde{\boldsymbol{M}}_{\boldsymbol{t},!\boldsymbol{t}}\, !\boldsymbol{t} + \boldsymbol{t}^\top \boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}\, \boldsymbol{t}, \tag{27}$$

where $!\boldsymbol{t}$ refers to the complementary set of variables in the quadratic form that are not $\boldsymbol{t}$:

$$!\boldsymbol{t} = \tilde{\boldsymbol{\tau}} \backslash \boldsymbol{t} = \begin{bmatrix} \mathrm{vec}(\boldsymbol{R}) \\ 1 \end{bmatrix} \equiv \tilde{\boldsymbol{r}}. \tag{28}$$

The zero gradient condition w.r.t. $\boldsymbol{t}$ (25) becomes

$$\frac{\partial f(\tilde{\boldsymbol{\tau}})}{\partial \boldsymbol{t}} = 2\tilde{\boldsymbol{M}}_{\boldsymbol{t},!\boldsymbol{t}}\, \tilde{\boldsymbol{r}} + 2\boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}\, \boldsymbol{t} = \boldsymbol{0}, \tag{29}$$

which is a linear system from which the optimal $\boldsymbol{t}^\star$ can be obtained as

$$\boldsymbol{t}^\star(\tilde{\boldsymbol{r}}) = -\boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}^{-1} \tilde{\boldsymbol{M}}_{\boldsymbol{t},!\boldsymbol{t}}\, \tilde{\boldsymbol{r}}. \tag{30}$$

Note that if $\boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}$ is not invertible, for any vector $\boldsymbol{u}$ in its nullspace we get $\boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}\boldsymbol{u} = \boldsymbol{0}$. If $\boldsymbol{t}^\star$ is a solution to the linear system (29), so is $\boldsymbol{t}^\star + \boldsymbol{u}$ as well:

$$\tilde{\boldsymbol{M}}_{\boldsymbol{t},!\boldsymbol{t}}\, \tilde{\boldsymbol{r}} + \boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}(\boldsymbol{t}^\star + \boldsymbol{u}) \tag{31}$$

$$= (\tilde{\boldsymbol{M}}_{\boldsymbol{t},!\boldsymbol{t}}\, \tilde{\boldsymbol{r}} + \boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}\, \boldsymbol{t}^\star) + \boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}\, \boldsymbol{u} = \boldsymbol{0}. \tag{32}$$

This case would leave us with infinitely many solutions and thus the problem would not be well-posed.

The substitution of $\boldsymbol{t}^\star(\tilde{\boldsymbol{r}})$ into the complete quadratic form (26) leaves us with a reduced quadratic form $q(\tilde{\boldsymbol{r}})$ function of $\tilde{\boldsymbol{r}}$ only:

$$q(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{Q}} \tilde{\boldsymbol{r}}. \tag{33}$$

The *marginalized* quadratic form matrix $\tilde{\boldsymbol{Q}}$ has the expression

$$\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{M}}_{!\boldsymbol{t},!\boldsymbol{t}} - \tilde{\boldsymbol{M}}_{!\boldsymbol{t},\boldsymbol{t}} \boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}^{-1} \tilde{\boldsymbol{M}}_{\boldsymbol{t},!\boldsymbol{t}}, \tag{34}$$

which is equivalent to the Schur complement $\tilde{\boldsymbol{M}}/\boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}$ of the complete matrix $\tilde{\boldsymbol{M}}$ w.r.t. the translation diagonal subblock $\boldsymbol{M}_{\boldsymbol{t},\boldsymbol{t}}$. Thus, after marginalization the *marginalized* optimization problem (35) reduces to finding the optimal value for the remaining variable $\boldsymbol{r}$ so that

$$\boldsymbol{r}^\star = \arg\min_{\boldsymbol{R}\in \mathrm{SO}(3)} \underbrace{\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{Q}} \tilde{\boldsymbol{r}}}_{q(\tilde{\boldsymbol{r}})}, \quad \tilde{\boldsymbol{r}} = \begin{bmatrix} \mathrm{vec}(\boldsymbol{R}) \\ 1 \end{bmatrix}. \tag{35}$$

4. Relaxations and Lagrangian duality

The marginalized problem (35) is difficult only because of the non-convex constraints that apply to the rotation $\boldsymbol{R}$. In fact, if we drop the constraint $\boldsymbol{R} \in \mathrm{SO}(3)$ and perform the optimization in $\boldsymbol{R} \in \mathbb{R}^{3\times 3}$ the remaining *relaxed* problem

$$f_l = \min_{\boldsymbol{R}\in\mathbb{R}^{3\times 3}} \tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{Q}} \tilde{\boldsymbol{r}}, \tag{36}$$

is quadratic and straightforward to solve. However, the optimization domain in the unconstrained problem is bigger, $\dim(\mathbb{R}^{3\times 3}) = 9$ vs. $\dim(\mathrm{SO}(3)) = 3$, and its optimal solution will most probably lie out of $\mathrm{SO}(3)$, that is, it will not be a proper rotation. This "naive" relaxation can be improved if we penalize the violation of constraints in the optimized objective in order to favour solutions closer to the original feasible domain.

The Lagrangian function does this for any general constrained optimization problem of the form

$$f^\star = \inf_{\boldsymbol{x}} f(\boldsymbol{x}), \quad \text{s.t. } c_i(\boldsymbol{x}) = 0,\ \forall i \in \mathcal{C}, \tag{37}$$

(where $\mathcal{C}$ is a set indexing the constraints $c_i(\boldsymbol{x})$) by penalizing the original optimization objective with a weighted sum of all the constraints:

$$\mathrm{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i\in\mathcal{C}} \lambda_i c_i(\boldsymbol{x}). \tag{38}$$

Note the optimum $\boldsymbol{x}^\star$ for the original problem (37) fulfills all the constraints $c_i(\boldsymbol{x}^\star) = 0$, so $\mathrm{L}(\boldsymbol{x}^\star, \boldsymbol{\lambda}) = f^\star$ for any $\boldsymbol{\lambda}$. The unconstrained problem

$$d(\boldsymbol{\lambda}) = \min_{\boldsymbol{x}} \mathrm{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \leq f^\star \tag{39}$$

provides a family of relaxations parameterized by the Lagrange multipliers $\boldsymbol{\lambda}$.

The key insight of Lagrangian duality is that, by choosing the parameter $\boldsymbol{\lambda}$ appropiately, we can obtain better relaxations ($d(\boldsymbol{\lambda})$ closer to $f^\star$). The (Lagrangian) *dual problem* seeks the best possible relaxation by choosing the weights λ_i, henceforth *dual variables*, that maximize the lower bound obtained from the Lagrangian relaxation:

$$d^\star = \max_{\boldsymbol{\lambda}} d(\boldsymbol{\lambda}). \tag{40}$$

For consistency with this terminology, the original problem, its unknowns and objective are referred to as *primal*. From the definitions above it is easy to prove that

$$d^\star \leq f^\star. \tag{41}$$

This fundamental result, referred to as *weak (Lagrangian) duality*, holds for any optimization problem and the positive value $f^\star - d^\star \geq 0$ is called the *duality gap*. If the duality gap is zero, that is, $d^\star = f^\star$, we say that there is *strong duality* and the relaxation is *tight*.

The dual problem has very appealing properties [1, Sec. 5.2]:

1. It is *always* convex, so it can be solved globally using local search techniques, which are inherently faster that global search methods.

2. If there is *strong duality*, the primal optimum $\boldsymbol{x}^\star$ is also the minimizer of the Lagrangian evaluated at the dual optimum:

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \mathrm{L}(\boldsymbol{x}, \boldsymbol{\lambda}^\star). \tag{42}$$

This last property is highly useful, as in many cases these constraints suffice to fully recover the primal optimum $\boldsymbol{x}^\star$ in terms of the dual optimum $\boldsymbol{\lambda}^\star$. As a result, if there is strong duality we can avoid the non-convexity of the general primal problem and solve instead the dual problem, then recover the primal solution. This settles a very interesting course of action for global optimization.

5. Lagrangian derivation

This section shows the construction of the Lagrangian function for the *equivalent*, *homogeneous*, *strengthened* primal problem:

$$\min_{\boldsymbol{R},y} \underbrace{\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{Q}} \tilde{\boldsymbol{r}}}_{q(\tilde{\boldsymbol{r}})}, \quad \tilde{\boldsymbol{r}} = \begin{bmatrix} \mathrm{vec}(\boldsymbol{R}) \\ y \end{bmatrix} \tag{$\tilde{\mathcal{P}}$}$$

$$\text{s.t. } \boldsymbol{R}^\top \boldsymbol{R} = y^2 \boldsymbol{I}_3, \tag{43}$$

$$\boldsymbol{R}\boldsymbol{R}^\top = y^2 \boldsymbol{I}_3, \tag{44}$$

$$\boldsymbol{R}^{(i)} \times \boldsymbol{R}^{(j)} = y \boldsymbol{R}^{(k)}, \ (i,j,k) \in (123), \tag{45}$$

$$y^2 = 1. \tag{46}$$

We use the notation (123) to refer to the set of cyclic permutations: $(123) = \{(1,2,3),(2,3,1),(3,1,2)\}$. The Lagrangian of the primal problem ($\tilde{\mathcal{P}}$) is then built by introducing the equality constraints as penalization terms

$$\mathrm{L}(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{\lambda}}) = q(\tilde{\boldsymbol{r}}) \qquad \text{[objective]} \tag{47}$$

$$+ \langle \boldsymbol{\Lambda}_\mathrm{r}, y^2 \boldsymbol{I}_3 - \boldsymbol{R}\boldsymbol{R}^\top \rangle \qquad \text{[orthonormal rows]} \tag{48}$$

$$+ \langle \boldsymbol{\Lambda}_\mathrm{c}, y^2 \boldsymbol{I}_3 - \boldsymbol{R}^\top \boldsymbol{R} \rangle \qquad \text{[orthonormal columns]} \tag{49}$$

$$+ \sum_{(i,j,k)\in(123)} \langle \boldsymbol{\lambda}_{\mathrm{d}_{ijk}}, \boldsymbol{R}^{(i)} \times \boldsymbol{R}^{(j)} - y\boldsymbol{R}^{(k)} \rangle \qquad \text{[handedness]} \tag{50}$$

$$+ \langle \gamma, 1 - y^2 \rangle. \qquad \text{[homogeneization]} \tag{51}$$

Every penalization term is written as the inner product of a Lagrange multiplier with a set of constraints. If the constraint is scalar (51) the inner product is a simple scalar product $\langle a, b \rangle = ab$. For vector constraints (50) it becomes the dot product for vectors, $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^\top \boldsymbol{b}$. Finally if the constraints form a matrix (48)(49), we use the inner product for matrices $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{B})$.

Since the constraints in the primal problem ($\tilde{\mathcal{P}}$) are (homogeneous) quadratic functions of $\boldsymbol{R}$, a linear combination of them will still be quadratic w.r.t. $\boldsymbol{R}$ and the Lagrangian can be refactored into

$$\mathrm{L}(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{\lambda}}) = \gamma + \tilde{\boldsymbol{r}}^\top (\tilde{\boldsymbol{Q}} + \tilde{\boldsymbol{P}}(\tilde{\boldsymbol{\lambda}})) \tilde{\boldsymbol{r}}. \tag{52}$$

The offset γ originates from the homogeneization constraint (51), the only constraint with a non-homogeneous term. The augmented vector of dual variables $\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^{22}$ is formed from the concatenation of all the dual variables, and the *penalization matrix* $\tilde{\boldsymbol{P}}$ is the sum of the different components coming from the constraints in ($\tilde{\mathcal{P}}$), namely orthonormality of rows (r) (48), orthonormality of columns (c) (49), positive determinant or handedness (d) (50) and homogeneization (h) (51):

$$\tilde{\boldsymbol{P}}(\tilde{\boldsymbol{\lambda}}) = \tilde{\boldsymbol{P}}_\mathrm{r}(\boldsymbol{\Lambda}_\mathrm{r}) + \tilde{\boldsymbol{P}}_\mathrm{c}(\boldsymbol{\Lambda}_\mathrm{c}) + \tilde{\boldsymbol{P}}_d(\underbrace{\{\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\}}_{\boldsymbol{\lambda}_\mathrm{d}}) + \tilde{\boldsymbol{P}}_\mathrm{h}(\gamma). \tag{53}$$

Next we will fully characterize these penalization terms. A complete summary of the constraints, associated dual variables and penalizations can be seen in Tab. 5. A visualization of the pattern for the resulting matrices $\tilde{\boldsymbol{P}}_{(\cdot)}$ is available in Fig. 2.

5.1. Orthonormality of rotation rows (48)

The matrix constraint for the orthonormality of rows, $\boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{I}_3$, provides the penalization term

$$\langle \boldsymbol{\Lambda}_\mathrm{r}, y^2 \boldsymbol{I}_3 - \boldsymbol{R}\boldsymbol{R}^\top \rangle = \mathrm{tr}(\boldsymbol{\Lambda}_\mathrm{r}^\top (y^2 \boldsymbol{I}_3 - \boldsymbol{R}\boldsymbol{R}^\top)) \tag{54}$$

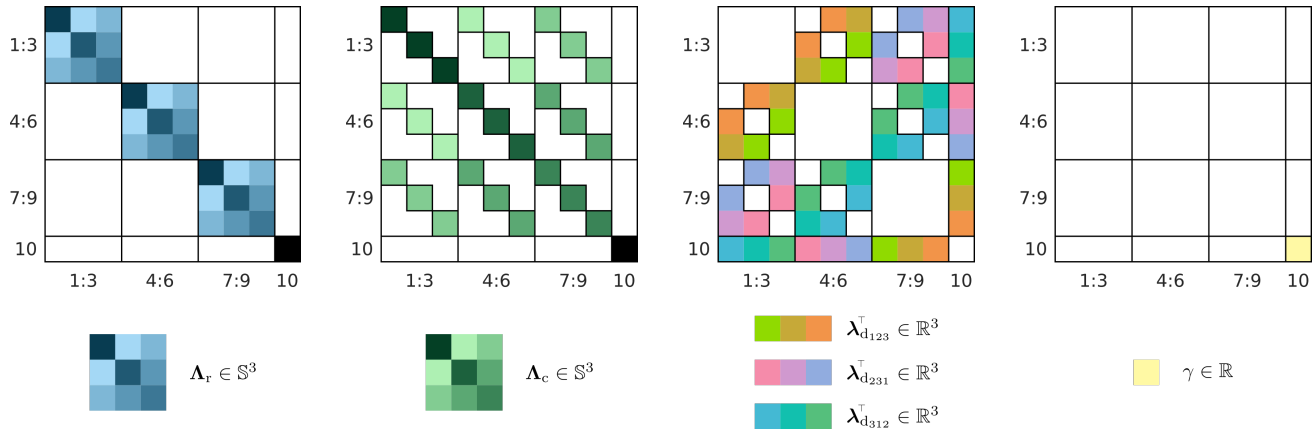
$$[\text{using } \boldsymbol{\Lambda}_\mathrm{r} = \boldsymbol{\Lambda}_\mathrm{r}^\top \text{ and cyclic property of } \mathrm{tr}(\cdot)] = y^2 \,\mathrm{tr}(\boldsymbol{\Lambda}_\mathrm{r}) - \mathrm{tr}(\boldsymbol{R}^\top \boldsymbol{\Lambda}_\mathrm{r} \boldsymbol{R}) \tag{55}$$

$$[\text{using } \mathrm{tr}(\boldsymbol{I}\boldsymbol{X}^\top \boldsymbol{B}\boldsymbol{Y}) \text{ (81)}] = y^2 \,\mathrm{tr}(\boldsymbol{\Lambda}_\mathrm{r}) - \boldsymbol{r}^\top (\boldsymbol{I}_3 \otimes \boldsymbol{\Lambda}_\mathrm{r}) \boldsymbol{r} \tag{56}$$

$$[\text{reordering into block matrix}] = \tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{P}}_\mathrm{r}(\boldsymbol{\Lambda}_\mathrm{r}) \tilde{\boldsymbol{r}}. \tag{57}$$

Table 3. Table of constraints and penalizations for ($\tilde{\mathcal{P}}$)

Constraint type	Constraint equation	Dual variable	Penalization term
Orthonormal rows	$y^2\boldsymbol{I}_3 - \boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{0}$	$\boldsymbol{\Lambda}_\mathrm{r} = \begin{bmatrix} \lambda_1 & \lambda_6 & \lambda_5 \\ \lambda_6 & \lambda_2 & \lambda_4 \\ \lambda_5 & \lambda_4 & \lambda_3 \end{bmatrix} \in \mathbb{S}^3$	$\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{P}}_\mathrm{r}(\boldsymbol{\Lambda}_\mathrm{r})\tilde{\boldsymbol{r}}$, see (58)
Orthonormal columns	$y^2\boldsymbol{I}_3 - \boldsymbol{R}^\top\boldsymbol{R} = \boldsymbol{0}$	$\boldsymbol{\Lambda}_\mathrm{c} = \begin{bmatrix} \lambda_7 & \lambda_{12} & \lambda_{11} \\ \lambda_{12} & \lambda_8 & \lambda_{10} \\ \lambda_{11} & \lambda_{10} & \lambda_9 \end{bmatrix} \in \mathbb{S}^3$	$\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{P}}_\mathrm{c}(\boldsymbol{\Lambda}_\mathrm{c})\tilde{\boldsymbol{r}}$, see (63)
Handedness	$\boldsymbol{R}^{(1)} \times \boldsymbol{R}^{(2)} - y\boldsymbol{R}^{(3)} = \boldsymbol{0}$ $\boldsymbol{R}^{(2)} \times \boldsymbol{R}^{(3)} - y\boldsymbol{R}^{(1)} = \boldsymbol{0}$ $\boldsymbol{R}^{(3)} \times \boldsymbol{R}^{(1)} - y\boldsymbol{R}^{(2)} = \boldsymbol{0}$	$\boldsymbol{\lambda}_{\mathrm{d}_{123}} = \begin{bmatrix} \lambda_{13} & \lambda_{14} & \lambda_{15} \end{bmatrix}^\top \in \mathbb{R}^3$ $\boldsymbol{\lambda}_{\mathrm{d}_{231}} = \begin{bmatrix} \lambda_{16} & \lambda_{17} & \lambda_{18} \end{bmatrix}^\top \in \mathbb{R}^3$ $\boldsymbol{\lambda}_{\mathrm{d}_{312}} = \begin{bmatrix} \lambda_{19} & \lambda_{20} & \lambda_{21} \end{bmatrix}^\top \in \mathbb{R}^3$	$\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{P}}_{\mathrm{d}_{ijk}}(\boldsymbol{\lambda}_{\mathrm{d}_{ijk}})\tilde{\boldsymbol{r}}$, see (74)
Homogeneization	$1 - y^2 = 0$	$\gamma \equiv \lambda_{22} \in \mathbb{R}$	$\gamma + \tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{P}}_\mathrm{h}(\gamma)\tilde{\boldsymbol{r}}$, see (77)



(a) Orthonormal rows, see (58) (b) Orthonormal columns, see (63) (c) Handedness, see (74) (d) Homogeneization, see (77)

Figure 2. Sparsity pattern of the penalization matrices (top row) and dual variables (bottom row) due to the different sets of constraints. A coloured cell indicates its value depends (linearly) only on the corresponding dual variable λ_i. Black cells (in $(10, 10)$) stand for values involving a linear combination of several dual variables.

The matrix Lagrange multiplier $\boldsymbol{\Lambda}_\mathrm{r}$ is symmetric because only 6 effective different scalar constraints stand here. The quadratic penalization matrix $\tilde{\boldsymbol{P}}_\mathrm{r}$ is shown in Fig. 2(a) and takes the expression

$$\tilde{\boldsymbol{P}}_\mathrm{r}(\boldsymbol{\Lambda}_\mathrm{r}) = \left[\begin{array}{c|c} -\boldsymbol{I}_3 \otimes \boldsymbol{\Lambda}_\mathrm{r} & \boldsymbol{0}_{9\times 1} \\ \hline \boldsymbol{0}_{1\times 9} & \mathrm{tr}(\boldsymbol{\Lambda}_\mathrm{r}) \end{array}\right]. \tag{58}$$

5.2. Orthonormality of rotation columns (49)

Similar to the previous orthonormality constraint but with exchanged transposes, $\boldsymbol{R}^\top\boldsymbol{R} = \boldsymbol{I}_3$, the orthonormality of rotation columns provides the penalization

$$\begin{aligned} \langle \boldsymbol{\Lambda}_\mathrm{c}, y^2\boldsymbol{I}_3 - \boldsymbol{R}^\top\boldsymbol{R}\rangle &= \mathrm{tr}(\boldsymbol{\Lambda}_\mathrm{c}^\top(y^2\boldsymbol{I}_3 - \boldsymbol{R}^\top\boldsymbol{R})) && (59)\\ &= y^2\,\mathrm{tr}(\boldsymbol{\Lambda}_\mathrm{c}) - \mathrm{tr}(\boldsymbol{\Lambda}_\mathrm{c}^\top\boldsymbol{R}^\top\boldsymbol{R}) && (60)\\ \text{[using } \mathrm{tr}(\boldsymbol{A}^\top\boldsymbol{X}^\top\boldsymbol{I}\boldsymbol{Y}) \text{ (81)]} &= y^2\,\mathrm{tr}(\boldsymbol{\Lambda}_\mathrm{c}) - \boldsymbol{r}^\top(\boldsymbol{\Lambda}_\mathrm{c}\otimes\boldsymbol{I}_3)\boldsymbol{r} && (61)\\ \text{[reordering into block matrix]} &= \tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{P}}_\mathrm{r}(\boldsymbol{\Lambda}_\mathrm{c})\tilde{\boldsymbol{r}}. && (62) \end{aligned}$$

Again, there are only 6 different constraints, so $\mathbf{\Lambda}_{\mathrm{c}} \in \mathbb{S}^3$. The quadratic penalization matrix $\tilde{\boldsymbol{P}}_{\mathrm{c}}$ is shown in Fig. 2(b) and takes the expression

$$\tilde{\boldsymbol{P}}_{\mathrm{c}}(\mathbf{\Lambda}_{\mathrm{c}}) = \left[\begin{array}{c|c} -\mathbf{\Lambda}_{\mathrm{c}} \otimes \boldsymbol{I}_3 & \mathbf{0}_{9\times 1} \\ \hline \mathbf{0}_{1\times 9} & \mathrm{tr}(\mathbf{\Lambda}_{\mathrm{c}}) \end{array}\right]. \tag{63}$$

5.3. Determinant constraints (50)

The right-hand rule applied to the columns of the rotation matrix $\boldsymbol{R}$ takes the form

$$\boldsymbol{R}^{(i)} \times \boldsymbol{R}^{(j)} = y\boldsymbol{R}^{(k)}, \tag{64}$$

where $\boldsymbol{R}^{(u)}$ refers to the u-th column in the matrix $\boldsymbol{R}$. Let us introduce the *canonical vector* $\boldsymbol{e}_u \in \mathbb{R}^3$, defined as the vector that is zero everywhere except for the u-th entry which is 1. It will be useful as well to define the *canonical matrix* $\boldsymbol{e}_{uv}$ as the matrix which has entry 1 at the position (u, v) and is zero everywhere else. Note $\boldsymbol{e}_{uv} = \boldsymbol{e}_u \boldsymbol{e}_v^\top$. Then the handedness constraint can be written in a more convenient form as

$$\begin{aligned}
\langle \boldsymbol{\lambda}_{\mathrm{d}_{ijk}}, \boldsymbol{R}^{(i)} \times \boldsymbol{R}^{(j)} - y\boldsymbol{R}^{(k)} \rangle &= \boldsymbol{\lambda}_{\mathrm{d}_{ijk}}^\top ((\boldsymbol{R}\boldsymbol{e}_i) \times (\boldsymbol{R}\boldsymbol{e}_j) - y(\boldsymbol{R}\boldsymbol{e}_k)) = && (65)\\
[\text{using } \boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c}) = -\boldsymbol{b}\cdot(\boldsymbol{a}\times\boldsymbol{c})] &= -(\boldsymbol{R}\boldsymbol{e}_i)^\top (\boldsymbol{\lambda}_{\mathrm{d}_{ijk}} \times (\boldsymbol{R}\boldsymbol{e}_j)) - y\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}^\top (\boldsymbol{R}\boldsymbol{e}_k) && (66)\\
[\text{using } \boldsymbol{a}\times\boldsymbol{b} = [\boldsymbol{a}]_\times \boldsymbol{b}] &= -\boldsymbol{e}_i^\top \boldsymbol{R}^\top \left[\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\right]_\times \boldsymbol{R}\boldsymbol{e}_j - y\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}^\top \boldsymbol{R}\boldsymbol{e}_k && (67)\\
[\text{using } \mathrm{tr}(a) = a, a\in\mathbb{R}] &= -\mathrm{tr}(\boldsymbol{e}_i^\top \boldsymbol{R}^\top \left[\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\right]_\times \boldsymbol{R}\boldsymbol{e}_j) - y\,\mathrm{tr}(\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}^\top \boldsymbol{R}\boldsymbol{e}_k) && (68)\\
[\text{using cyclic property of } \mathrm{tr}(\cdot)] &= -\mathrm{tr}(\boldsymbol{e}_j\boldsymbol{e}_i^\top \boldsymbol{R}^\top \left[\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\right]_\times \boldsymbol{R}) - y\,\mathrm{tr}(\boldsymbol{e}_k\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}^\top \boldsymbol{R}) && (69)\\
[\text{reordering transposes}] &= -\mathrm{tr}(\boldsymbol{e}_{ij}^\top \boldsymbol{R}^\top \left[\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\right]_\times \boldsymbol{R}) - y\,\mathrm{tr}((\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\boldsymbol{e}_k^\top)^\top \boldsymbol{R}) && (70)\\
[\text{using } \mathrm{tr}(\boldsymbol{A}^\top\boldsymbol{X}^\top\boldsymbol{I}\boldsymbol{Y}) \text{ (81) and } \mathrm{tr}(\boldsymbol{A}^\top\boldsymbol{B}) \text{ (80)}] &= -\boldsymbol{r}^\top(\boldsymbol{e}_{ij}\otimes\left[\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\right]_\times)\boldsymbol{r} - y\,\mathrm{vec}(\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\boldsymbol{e}_k^\top)^\top\boldsymbol{r} && (71)\\
[\text{using } \mathrm{vec}(\boldsymbol{a}\boldsymbol{b}^\top) \text{ (79)}] &= -\boldsymbol{r}^\top(\boldsymbol{e}_{ij}\otimes\left[\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\right]_\times)\boldsymbol{r} - y(\boldsymbol{e}_k\otimes\boldsymbol{\lambda}_{\mathrm{d}_{ijk}})^\top\boldsymbol{r} && (72)\\
[\text{reordering into block matrix}] &= \tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{P}}_{\mathrm{d}_{ijk}}(\boldsymbol{\lambda}_{\mathrm{d}_{ijk}})\,\tilde{\boldsymbol{r}}. && (73)
\end{aligned}$$

The (symmetric) quadratic penalization matrix $\tilde{\boldsymbol{P}}_{\mathrm{d}_{ijk}}(\boldsymbol{\lambda}_{\mathrm{d}_{ijk}})$ takes the expression

$$\tilde{\boldsymbol{P}}_{\mathrm{d}_{ijk}} = \left[\begin{array}{c|c} -\boldsymbol{e}_{ij}\otimes\left[\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}\right]_\times & -(\boldsymbol{e}_k\otimes\boldsymbol{\lambda}_{\mathrm{d}_{ijk}}) \\ \hline \mathbf{0}_{1\times 9} & 0 \end{array}\right]. \tag{74}$$

Note this expression is generic for $(i, j, k) \in (123) = \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\}$. The complete penalization matrix $\tilde{\boldsymbol{P}}_d(\boldsymbol{\lambda}_{\mathrm{d}}) = \tilde{\boldsymbol{P}}_{\mathrm{d}_{123}}(\boldsymbol{\lambda}_{\mathrm{d}_{123}}) + \tilde{\boldsymbol{P}}_{\mathrm{d}_{231}}(\boldsymbol{\lambda}_{\mathrm{d}_{231}}) + \tilde{\boldsymbol{P}}_{\mathrm{d}_{312}}(\boldsymbol{\lambda}_{\mathrm{d}_{312}})$ from all the determinant constraints, after taking an equivalent symmetric matrix for the representation, is shown in Fig. 2(c).

5.4. Homogeneization constraint (51)

Finally, the simple scalar constraint $y^2 = 1$ yields

$$\begin{aligned}
\langle \gamma, 1 - y^2 \rangle &= \gamma - \gamma y^2 && (75)\\
&= \gamma + \tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{P}}_{\mathrm{h}}(\gamma)\tilde{\boldsymbol{r}} && (76)
\end{aligned}$$

For compatibility with the rest of quadratic expressions this has been written in matrix form too, where $\tilde{\boldsymbol{P}}_{\mathrm{h}}(\gamma) = -\gamma\boldsymbol{e}_{10,10}$ is the matrix with all zeros but the lower right corner, where it contains the value $-\gamma$:

$$\tilde{\boldsymbol{P}}_{\mathrm{h}}(\gamma) = \left[\begin{array}{c|c} \mathbf{0}_{9\times 9} & \mathbf{0}_{9\times 1} \\ \hline \mathbf{0}_{1\times 9} & -\gamma \end{array}\right]. \tag{77}$$

6. Additional experimental results

Due to the lack of space, only the most relevant experimental results and metrics were shown in the paper. In this section we include all the obtained results comparing the `BnB` solution [4], and the dual-based relaxations, `Ours` and `Olsson` [3]. The results support the claims in the main document.

Evaluation metrics

We used several metrics to evaluate the performance of the methods. Next we describe them and give some notes on the representation used for the results. All the statistics were obtained from 200 samples.

Suboptimality gap The suboptimality gap is defined as $\Delta = f - f^\star$, where $f^\star$ is the minimum objective value attained in the global solution. Since our solution always provided a globally optimal solution (in a certifiable way, since $f = d^\star \Rightarrow f = f^\star$), it is possible to plot this value for all the methods giving a clear reference for the performance: A solution is optimal if $\Delta = 0$, suboptimal otherwise.

We displayed the suboptimality gap in the oncoming figures using boxplots. To show as much information as possible, we also superimposed a sorted set of points displaying the original values underlying each boxplot. Note that for the globally optimal methods, `Ours` and `BnB`, the boxplots and points degenerate to a single line in $\Delta = 0$.

In order to cope with the high variability of the suboptimality values for the different methods, we applied a custom scale in the Y axis, where the $\log(1 + x)$ of the values is displayed in order to better visualize different orders of magnitude while keeping the lower threshold $\Delta = 0$ in the origin.

Global optimality ratio The global optimality ratio provides a refined view of suboptimality metric, where we directly show a bar plot with the percentage of cases in which a globally optimal solution ($\Delta = 0$) was attained.

Time The computational performance of the different algorithms is evaluated by measuring the CPU time necessary to solve each problem. Note these times are just for orientation, as the implementations were not highly optimized. Nevertheless, the conclusions when comparing the different methods remain valid. The obtained values are represented with a shaded error bar plot displaying the median values plus a band defined by the 1st and 3rd quartiles to reflect the distribution of the values. Note that we always used a logarithmic scale for the time values, in order to better compare all the considered methods.

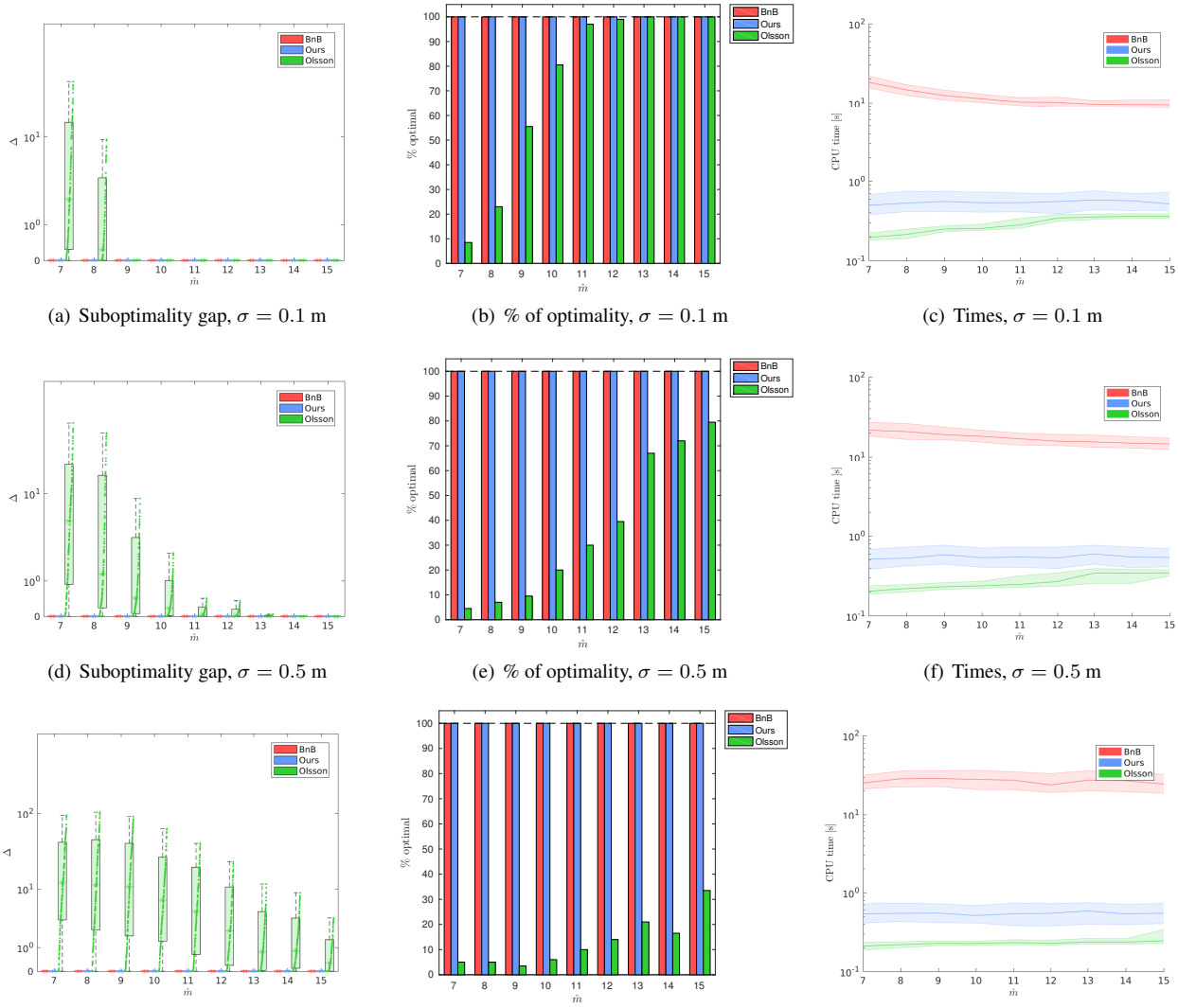
Problem parameters

As discussed in the paper, the main problem parameters defining the overall difficulty of solving the optimization problem globally are the *effective number of correspondences* $\hat{m}$ and the *level of the noise* σ affecting the measurements. A more detailed explanation of these is available in the main document.

6.1. Evaluation on synthetic data

6.1.1 Challenging range of $\hat{m}$

The evaluation in a challenging range of $\hat{m}$ values, going from the almost minimal case $\hat{m} = 7$ up to $\hat{m} = 15$, shows that as the number of correspondences increases the level of difficulty of the problem decreases (see Fig. 3). This is consistent with the claim in [3]. This explains why the looser relaxation `Olsson` has better performance as $\hat{m}$ grows. This trend holds for different levels of noise σ, although as expected the general difficulty keeps higher when the noise is larger, making the performance of `Olsson` worse. In all cases, as the provably optimal method `BnB`, our method provided a globally optimal solution. However, the computational burden for the `BnB` method was notably higher, and in the *easier* scenario ($\sigma = 0.1$) it increased with the decrease in the number of correspondences. This is consistent with the remarks in [4], where the near-minimal cases were tagged as harder to explore in the Branch and Bound framework. The time for the dual-based algorithms (`Ours` and `Olsson`) remained practically constant in all cases and was notably lower than for `BnB` (around 2 orders of magnitude).



(a) Suboptimality gap, $\sigma = 0.1$ m (b) % of optimality, $\sigma = 0.1$ m (c) Times, $\sigma = 0.1$ m

(d) Suboptimality gap, $\sigma = 0.5$ m (e) % of optimality, $\sigma = 0.5$ m (f) Times, $\sigma = 0.5$ m

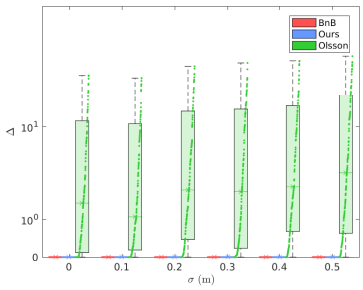
(g) Suboptimality gap, $\sigma = 1$ m (h) % of optimality, $\sigma = 1$ m (i) Times, $\sigma = 1$ m

Figure 3. Evaluation in a challenging range of effective number of correspondences, from the almost minimal case $\hat{m} = 7$ to $\hat{m} = 15$. Different levels of noise were considered: A low value of $\sigma = 0.1$ m (top row), a middel value of $\sigma = 0.5$ m (middle row), and a more challenging scenario with $\sigma = 1$ m (bottom row).

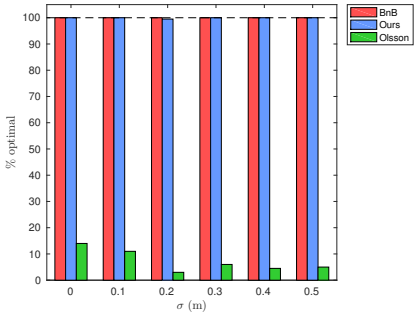
6.1.2 Evaluation w.r.t. noise level

As in [3], we provide a continuous analysis on how the performance of the methods are affected for different levels of noise σ within usual application values, going from the ideal noise-free case with $\sigma = 0$ to the significantly high value of $\sigma = 0.5$ m. This evaluation is repeated several times under a fixed effective number of correspondences, namely $\hat{m} = \{7, 10, 14, 21\}$. The whole set of results can be seen in Fig. 4.

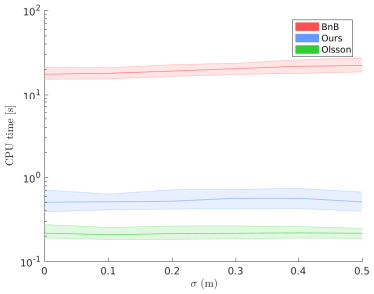
As expected, the general trend is that the difficulty of the problem increases with the noise level σ, making it harder for `Olsson` to reach the global solution, and also rendering the resolution with `BnB` slower. In all cases our solution provided the globally optimal solution, at a constant low resolution time.



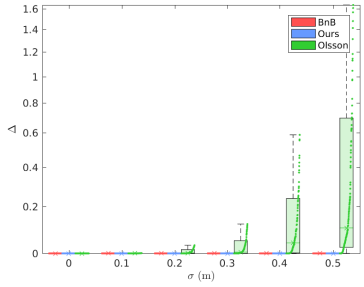
(a) Suboptimality gap, $\hat{m} = 7$



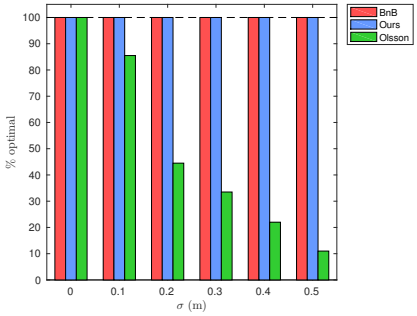
(b) % of optimality, $\hat{m} = 7$



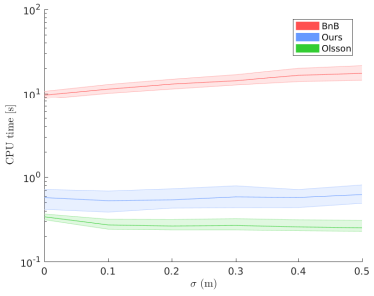
(c) Times, $\hat{m} = 7$



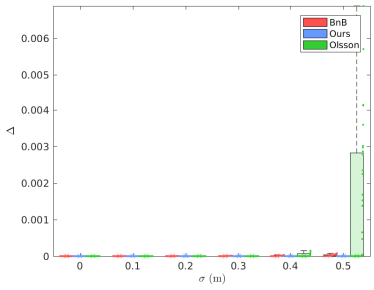
(d) Suboptimality gap, $\hat{m} = 10$



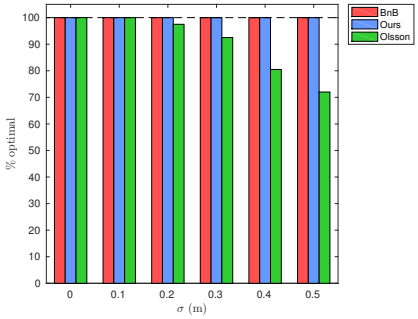
(e) % of optimality, $\hat{m} = 10$



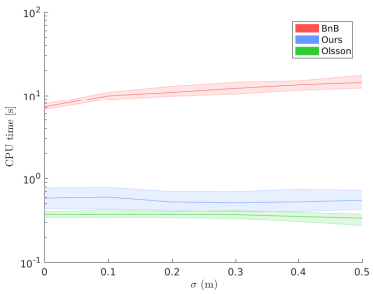
(f) Times, $\hat{m} = 10$



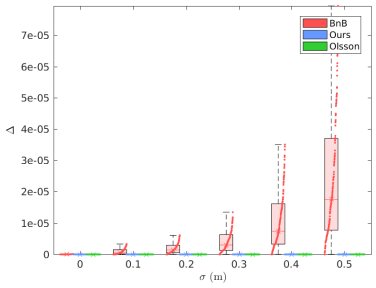
(g) Suboptimality gap, $\hat{m} = 14$



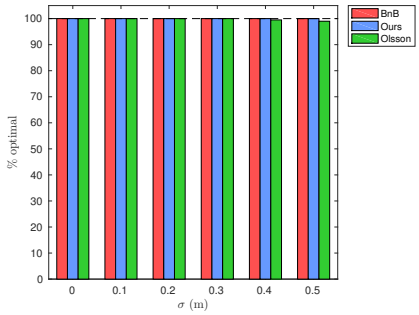
(h) % of optimality, $\hat{m} = 14$



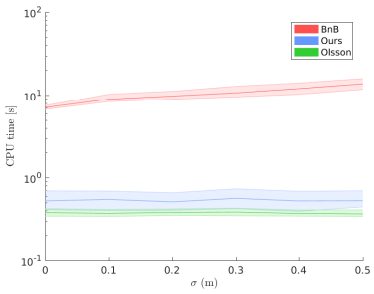
(i) Times, $\hat{m} = 14$



(j) Suboptimality gap, $\hat{m} = 21$



(k) % of optimality, $\hat{m} = 21$



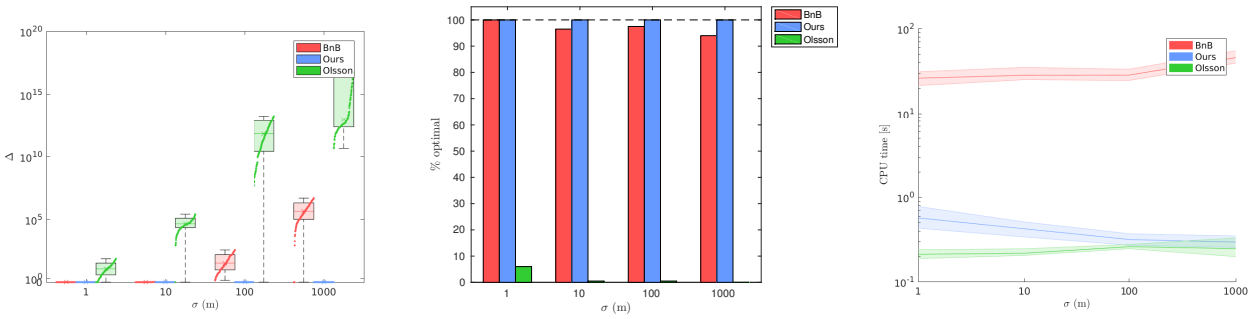
(l) Times, $\hat{m} = 21$

Figure 4. Evaluation w.r.t. noise level from $\sigma = 0$ to $\sigma = 0.5$ m, for different effective numbers of measurements: $\hat{m} = \{7, 10, 14, 21\}$. Note the decreasing scale in the suboptimality gap (first column of figures) as the effective number of correspondences $\hat{m}$ increases.

6.1.3 Evaluation in extreme conditions

Finally, we used the synthetic framework to test our method in a substantially difficult scenario. For that purpose, we took the most challenging conditions for the registration problem, that is almost minimal number of correspondences $\hat{m} = 7$ and *very* high noise level (far beyond any expectable value found in a real problem). Specifically, we increased the noise level σ exponentially from $\sigma = 1$ m to $\sigma = 1000$ m. With such a high level of noise the results of the problem would of little use in a real application, but again we do this just with the main aim of testing the behaviour and robustness of all the algorithms from a purely mathematical point of view.

The results, displayed in Fig. 5, show that under these severe conditions the relaxation used by `Olsson` never worked in the higher regimes of noise, and even the `BnB` method incurred in some suboptimalities probably due to numerical issues and exit conditions in the algorithm. Meanwhile, `Ours` worked in 100% of the cases without exception.



(a) Suboptimality gap, $\hat{m} = 7$ (b) % of optimality, $\hat{m} = 7$ (c) Times, $\hat{m} = 7$

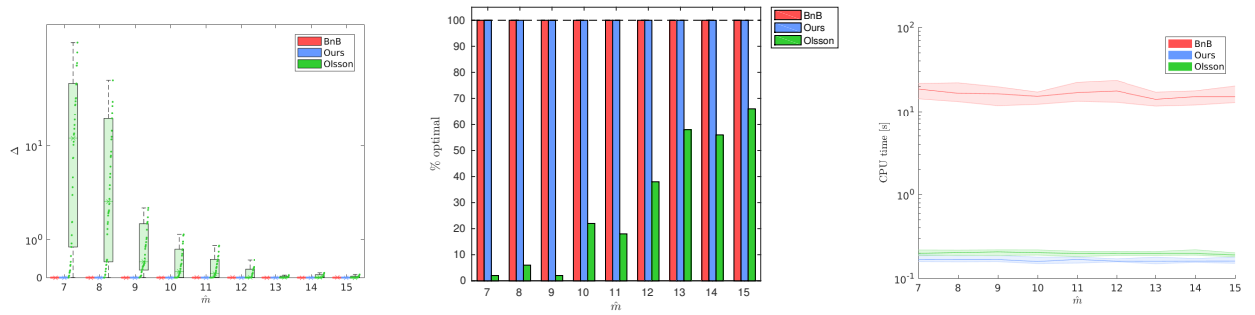
Figure 5. Evaluation in the most severe regime of parameters: Lowest possible number of correspondences $\hat{m} = 7$ and significantly high measurement noise. Note the different scales in the axes.

6.2. Evaluation on real data

We use the same real dataset used in [3]. Following the procedure explained in the main document, we were able to generate a large set of different registration problems by taking different effective numbers of correspondences $\hat{m}$. The noise level was intrinsic to the sensor used for the measurements, so it is out of our control.

As a result, as in Section 6.1.1, the evaluation on real data considers an increasing value for $\hat{m}$. Note however that, even though the sensor noise should be well below the synthetic noise considered in the top row of Fig. 3, $\sigma = 0.1$ m, the results are closer to those of a much higher synthetic noise (compare to $\sigma = 0.5$ m in Fig. 3). This may be a consequence of the non-Gaussian noise in the real measurements, which might be turning the problem even harder.

In any case, our method still reached a globally optimal solution in all cases, regardless of the effective number of measurements $\hat{m}$, and at a fraction of the time for `BnB`.



(a) Suboptimality gap (b) % of optimality (c) Times

Figure 6. Evaluation in real data taken from the Space Station dataset [3]. Different subsets of the complete dataset are sampled to get more challenging registration problems with a variable effective number of correspondences $\hat{m}$.

A. Some matrix calculus

Next we provide some properties related to the vectorization of matrix expressions. These relations were taken from [2]:

$$\mathrm{vec}(\boldsymbol{AXB}) = (\boldsymbol{B}^\top \otimes \boldsymbol{A})\,\mathrm{vec}(\boldsymbol{X}) \tag{78}$$

$$\mathrm{vec}(\boldsymbol{ab}^\top) = \boldsymbol{b} \otimes \boldsymbol{a} \tag{79}$$

$$\mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{B}) = \mathrm{vec}(\boldsymbol{A})^\top \mathrm{vec}(\boldsymbol{B}) \tag{80}$$

$$\mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{X}^\top \boldsymbol{BY}) = \mathrm{vec}(\boldsymbol{X})^\top (\boldsymbol{A} \otimes \boldsymbol{B})\,\mathrm{vec}(\boldsymbol{Y}) \tag{81}$$

References

[1] S. Boyd and L. Vandenberghe. *Convex optimization*. 2004. 5

[2] P. L. Fackler. Notes on Matrix Calculus. pages 1–14, 2005. 3, 13

[3] C. Olsson and A. Eriksson. Solving quadratically constrained geometrical problems using lagrangian duality. In *Pattern Recognition, 2008. ICPR 2008. 19th Int. Conf.*, pages 1–5. IEEE, 2008. 8, 9, 10, 12

[4] C. Olsson, F. Kahl, and M. Oskarsson. Branch-and-Bound Methods for Euclidean Registration Problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(5):783–794, 2009. 8, 9