

From Correspondences to Pose: Non-minimal Certifiably Optimal Relative Pose without Disambiguation

Javier Tirado-Garín Javier Civera
I3A, University of Zaragoza
{j.tiradog, jcivera}@unizar.es

Abstract

Estimating the relative camera pose from $n \geq 5$ correspondences between two calibrated views is a fundamental task in computer vision. This process typically involves two stages: 1) estimating the essential matrix between the views, and 2) disambiguating among the four candidate relative poses that satisfy the epipolar geometry. In this paper, we demonstrate a novel approach that, for the first time, bypasses the second stage. Specifically, we show that it is possible to directly estimate the correct relative camera pose from correspondences without needing a post-processing step to enforce the cheirality constraint on the correspondences. Building on recent advances in certifiable non-minimal optimization, we frame the relative pose estimation as a Quadratically Constrained Quadratic Program (QCQP). By applying the appropriate constraints, we ensure the estimation of a camera pose that corresponds to a valid 3D geometry and that is globally optimal when certified. We validate our method through exhaustive synthetic and real-world experiments, confirming the efficacy, efficiency and accuracy of the proposed approach. Code is available at <https://github.com/javrtg/C2P>.

1. Introduction

Finding the relative pose between two calibrated views is crucial in many computer vision applications. This task is particularly relevant, among others, in Structure from Motion (SfM) [44, 53], and Simultaneous Localization And Mapping (SLAM) [13, 46, 47]. In SfM, it serves to geometrically verify the correspondences as well as to provide pairwise constraints for pose averaging schemes [27, 40]. In SLAM, besides correspondence verification, it is also used for bootstrapping the odometry of the camera and computing an initial estimate of the 3D map.

The relative pose problem has five observable degrees of freedom: three for the relative rotation between the cameras, and two for the direction of the relative transla-

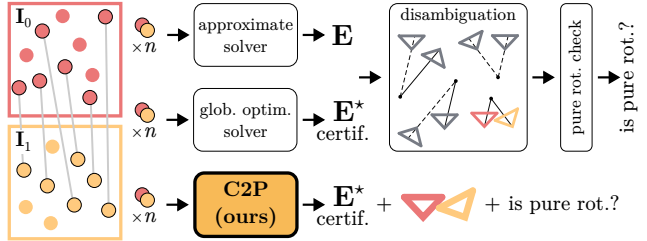


Figure 1. **Relative pose directly from matches, without posterior disambiguation and pure rotation checks.** Traditionally, estimating the relative pose involves two steps: 1) Estimating the essential matrix \mathbf{E} using an approximate or globally-optimal solver, and 2) disambiguating the unique geometrically valid pose among four candidate relative poses, with an additional step to determine if the motion is purely rotational. In this paper, we introduce C2P, a globally-optimal and certifiable approach that, for the first time, solves the relative pose problem in a single step.

tion. The standard approach for its computation [29] begins by considering a set of n pixel correspondences between the two images. These correspondences can be established through matching the descriptors of keypoints extracted from the images [1, 16, 41, 43], or more recently by estimating a (semi)dense 2D mapping between the views [17, 58, 61]. The pose is then computed by minimizing epipolar errors [29], requiring at least five correspondences. Solvers that handle $n = 5$ correspondences are termed *minimal* [48, 54] and those able to handle all the correspondences are called *non-minimal* [10, 69].

This paper focuses on non-minimal solvers. In a practical setup in which input correspondences may contain outliers, these solvers are essential within RANSAC [19, 49] and Graduated Non-Convexity (GNC) [66]. In RANSAC, non-minimal solvers are used to improve the accuracy of the *so-far-best* and final models (initially computed with minimal solvers) thanks to noise cancellation of the inliers [49]. In GNC, globally-optimal non-minimal solvers serve as fundamental building blocks to robustly solve a weighted instance of the problem in an iterative fashion.

Since the seminal paper by Longuet-Higgins [42] in

1981, computing this non-minimal estimate of the relative pose, involves two key steps: 1) estimating the *essential matrix*—which models the epipolar geometry of the problem—via an approximate (minimal or local optimization) method, or a certifiably globally-optimal method, and 2) disambiguating the true relative pose from those satisfying the same epipolar geometry. This ambiguity arises from ignoring the *cheirality constraints* which enforce a valid 3D geometry (mainly that the 3D points observed in the image must be in front of the camera), resulting in an additional overhead that scales with the number of points.

In this paper, we demonstrate that this two-step approach, gold standard for more than 40 years, is not a must, and present a method that, for the first time directly estimates the relative pose without requiring posterior disambiguation. For this purpose, we leverage convex optimization theory and demonstrate that we can obtain a certifiably globally-optimal solution that is geometrically valid. Additionally, our method can determine if the motion between the images is purely rotational. This information is relevant, since the translation vector is undefined under pure rotation and its estimation should not be trusted. Current methods require a posterior verification for the same purpose. We present a visual overview of both current approaches and our method in Figure 1. Our main contributions are: 1) We provide a non-minimal certifiably globally-optimal method that, for the first time, solves the relative pose problem without the need of disambiguation. We derive the sufficient and necessary conditions to recover the optimal solution, 2) for this purpose, we also derive a novel characterization of the normalized essential manifold that is needed to enforce a geometrically valid solution, and 3) besides our theoretical contributions, we show experimentally that our method scales better than the alternatives in the literature.

2. Related Work

Non-minimal epipolar geometry. Initial methods for estimating the essential matrix (or *fundamental matrix* in the uncalibrated case) [26, 42] rely on linear relaxations, which do not account for the nonlinearities arising from the constraints of the problem. As a result, these methods provide an approximate solution that needs to be projected onto the appropriate space. While several methods [21, 22, 30, 59, 60] refine an approximate estimate through local optimization on the manifold of the essential matrix, they, despite being certifiable [21, 22], cannot guarantee the global optimality of the solution. To achieve global optimality, various methods [28, 36, 68] employ Branch and Bound (BnB) techniques to explore the feasible parameter space and eliminate regions that are guaranteed not to contain the optimal solution. However, these methods can exhibit exponential time complexity in the worst case. More recently,

the use of relaxed Quadratically Constrained Quadratic Programs (QCQP) onto Semidefinite Programs (SDP) via the Shor’s relaxation [2, 6], has enabled methods to provide, and certify, globally-optimal solutions [10, 23, 34, 69, 70]. Briaes et al. [10] adopt an eigenvalue formulation of the problem [36] and show that a tight relaxation of this non-convex formulation can be achieved through redundant constraints. Zhao [69] achieves a significant increase in performance by optimizing with respect to a more efficient formulation, resulting in a reduced set of constraints and parameters. In addition, Garcia-Salguero et al. [23] derive redundant constraints to improve the general tightness of the work by Zhao [69]. Finally, Karimian and Tron [34] show that, under moderate levels of noise, a faster solution can be achieved using the Riemannian staircase algorithm [5, 11]. In this paper, we address a common limitation of all previous approaches. We demonstrate that it is possible to avoid the posterior disambiguation of the four relative poses that satisfy their solutions. We achieve this also by relaxing a QCQP formulation of the problem. We design and introduce the constraints that take into account the 3D geometric meaning of the estimates, and provide a fast globally optimal solution for the geometrically correct relative pose.

Related certifiably globally optimal methods. Similar relaxations to those used in previous methods are also applied in various closely related areas of computer vision. For example, applications of Shor’s relaxation are found in tasks such as solving Wahba’s problem [65], pointcloud and 3D registration [8, 67] and multiview triangulation [25]. Similarly, the Riemannian staircase algorithm, coupled with local manifold optimization, is used in tasks exploiting the low-rank nature of their solutions, including rotation averaging [15] and pose synchronization [9, 50].

3. Non-minimal solver for the relative pose

In this work, we assume a central camera model, making our method suitable for, *e.g.*, pinhole, fisheye and omnidirectional cameras. We consider that each correspondence, i , is parameterized by unit bearing vectors $\mathbf{f}_{0,i}, \mathbf{f}_{1,i} \in \mathcal{S}^2$. Given a set of $n > 5$ correspondences $\{\mathbf{f}_{0,i}, \mathbf{f}_{1,i}\}_{i=1}^n$, our goal is to directly estimate the relative pose, which we parameterize with a rotation matrix $\mathbf{R} \in \text{SO}(3)$ and a unit-norm translation vector $\mathbf{t} \in \mathcal{S}^2$. These parameters define the manifold of normalized essential matrices [30, 69]:

$$\mathcal{M}_{\mathbf{E}} := \{\mathbf{E} \mid \mathbf{E} = [\mathbf{t}]_{\times} \mathbf{R}, \mathbf{R} \in \text{SO}(3), \mathbf{t} \in \mathcal{S}^2\}^1. \quad (1)$$

However, Eq. (1) does not account for the geometric peculiarities and symmetry of the epipolar constraints [59, 60], which motivates us to investigate a necessary and sufficient characterization for directly estimating the relative pose. In

¹Where $[\mathbf{t}]_{\times}$ denotes the skew-symmetric matrix corresponding to \mathbf{t} .

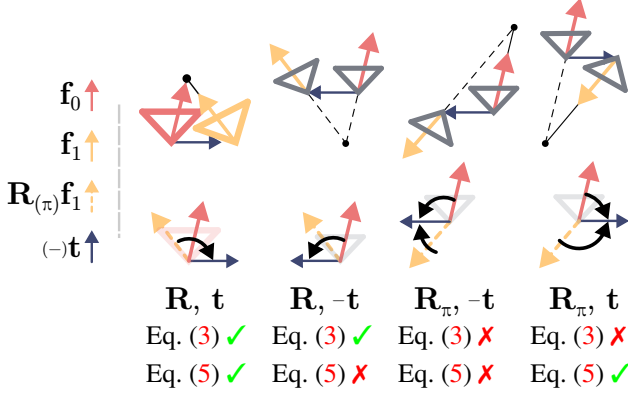


Figure 2. **Necessary geometric conditions.** When removing the rotational flow between the bearing vectors [37], *i.e.* considering \mathbf{f}_0 and $\mathbf{R}\mathbf{f}_1$, they must exhibit the same (counter-)clockwise turn w.r.t. the translation (Eq. (3)). Otherwise, the rotation must be a reflected version, \mathbf{R}_π , of the true rotation, \mathbf{R} . Considering the correct rotation, \mathbf{f}_0 must have greater projection onto the translation than $\mathbf{R}\mathbf{f}_1$ (Eq. (5)). Otherwise, the bearings would not meet along the direction of their beams (\rightarrow , \rightarrow), implying that the translation is flipped ($-\mathbf{t}$) w.r.t. the correct one, \mathbf{t} . Therefore, besides avoiding triangulation, these constraints completely disambiguate the relative pose and are generally applicable to central camera models since they do not rely on traditional positive-depth constraints.

Sec. 3.1, we derive such characterization with a focus on making it amenable for a QCQP formulation (Sec. 3.3). We demonstrate that the relaxed QCQP yields a tight and certifiably globally optimal solution for the relative pose (Sec. 4).

3.1. Necessary and sufficient constraints

It is well known in the literature [29, Sec. 9.6] that four relative poses satisfy the same epipolar geometry. However, only one of these poses is geometrically valid, in the sense that it leads to an estimation of the 3D points being in front of the cameras². A common approach to circumvent this involves triangulating the points and checking for the positive sign of their depths [29, 48], *a posteriori*. Preliminary experiments showed that enforcing this positive-depth constraint *during the optimization* is challenging and costly due to the complexities associated with the use of rotation matrices [10, 12], even after imposing convex hull constraints of $\text{SO}(3)$ [51, 52]. Consequently, we propose using simpler and more efficient constraints to disambiguate the pose during the optimization and visualize them in Figure 2.

Rotation disambiguation. As shown in [37], regardless of the sign of the translation \mathbf{t} , the two normals of the epipolar

plane of a correspondence $(\mathbf{f}_0, \mathbf{f}_1)$: $\mathbf{t} \times \mathbf{f}_0$ and $\mathbf{t} \times \mathbf{R}\mathbf{f}_1$ satisfy:

$$(\mathbf{t} \times \mathbf{f}_0) \cdot (\mathbf{t} \times \mathbf{R}\mathbf{f}_1) > 0, \quad (2)$$

Intuitively, considering \mathbf{f}_0 and $\mathbf{R}\mathbf{f}_1$ in the same coordinate system, the smallest in-plane rotations required to align \mathbf{f}_0 with \mathbf{t} and $\mathbf{R}\mathbf{f}_1$ with \mathbf{t} must be both clockwise or counter-clockwise. This condition is not met when the 3D point does not lie along one of the bearing vectors' beams, which is the case for the (incorrect) reflected rotation matrix. However, besides involving \mathbf{R} , Eq. (2) is cubic on the unknowns, requiring a re-formulation with additional parameters to adapt it for a QCQP. A more straightforward solution arises upon realizing that $\mathbf{t} \times \mathbf{R}\mathbf{f}_1 = [\mathbf{t}]_\times \mathbf{R}\mathbf{f}_1 = \mathbf{E}\mathbf{f}_1$, leading to:

$$(\mathbf{E}\mathbf{f}_1) \cdot ([\mathbf{t}]_\times \mathbf{f}_0) = \mathbf{f}_1^\top \mathbf{E}^\top [\mathbf{t}]_\times \mathbf{f}_0 > 0, \quad (3)$$

which now depends quadratically on the unknowns and is thus suitable for a QCQP formulation.

Translation disambiguation. Building on the previous intuition that both \mathbf{f}_0 and $\mathbf{R}\mathbf{f}_1$ require a (counter-)clockwise in-plane rotation to be aligned with \mathbf{t} , it follows that:

$$\mathbf{f}_0^\top \mathbf{t} - (\mathbf{R}\mathbf{f}_1)^\top \mathbf{t} \geq 0. \quad (4)$$

If this condition is not met, the bearing vectors would not intersect along their beams, implying that the translation vector is the (incorrect) negative of \mathbf{t} . The impact of this constraint on restricting the space of possible unit translations is shown in Fig. 3. However, Eq. (4) again involves globally optimizing \mathbf{R} . A more efficient approach is to optimize the rotated translation vector $\mathbf{q} := \mathbf{R}^\top \mathbf{t}$, $\mathbf{q} \in \mathcal{S}^2$, resulting in:

$$\mathbf{f}_0^\top \mathbf{t} - \mathbf{f}_1^\top \mathbf{q} \geq 0. \quad (5)$$

which is linear in the unknowns and can be incorporated in a QCQP with an homogenization variable [10, 24]. However, this introduces the challenge of ensuring that $\mathbf{q} = \mathbf{R}^\top \mathbf{t}$ still holds without optimizing \mathbf{R} . We address this next by deriving a novel definition of the normalized essential manifold.

Manifold constraints. To ensure that $\mathbf{q} = \mathbf{R}^\top \mathbf{t}$ holds during the optimization, we need to mutually constrain \mathbf{t} and \mathbf{q} . Previous definitions of the normalized essential manifold, such as those involving the left, right and quintessential matrix sets [23, 34, 69], do not include this kind of constraints. The most suitable constraints for our purposes are the ones involving the adjugate matrix of \mathbf{E} : $\text{Adj}(\mathbf{E}) = \mathbf{q}\mathbf{t}^\top$, used in [23] as redundant constraints. We show in Th. 3.1 that these, along with norm constraints, suffice to define $\mathcal{M}_{\mathbf{E}}$, and to perform an efficient joint optimization of \mathbf{E} , \mathbf{t} and \mathbf{q} .

Theorem 3.1. A real 3×3 matrix, \mathbf{E} , is an element of $\mathcal{M}_{\mathbf{E}}$ if and only if it satisfies:

$$(i) \quad \text{tr}(\mathbf{E}\mathbf{E}^\top) = 2 \quad \text{and} \quad (ii) \quad \text{Adj}(\mathbf{E}) = \mathbf{q}\mathbf{t}^\top, \quad (6)$$

for two vectors $\mathbf{t}, \mathbf{q} \in \mathcal{S}^2$ and where $\text{Adj}(\mathbf{E})$ represents the *adjugate matrix* [32] of \mathbf{E} .

²In practice, issues such as noise and small-scale translation relative to the observed scene can cause some 3D points to appear behind the cameras, despite a correct pose estimate. Consequently, cheirality is often verified for all points and then aggregated to robustly select the correct pose [7, 53].

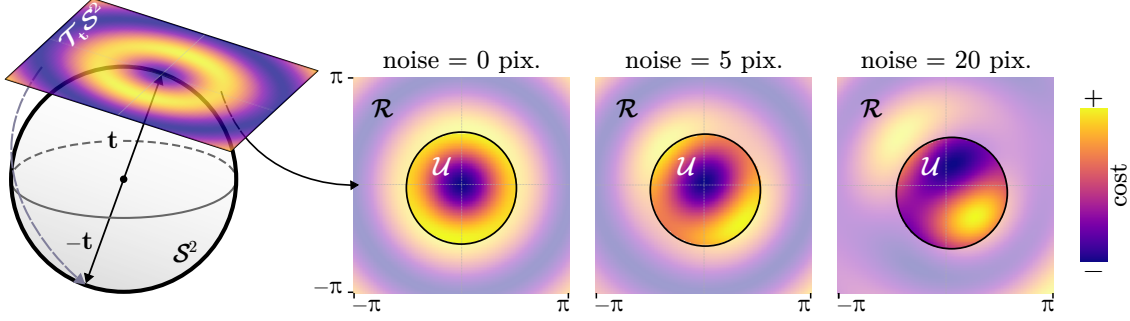


Figure 3. **Automatic disambiguation of the relative pose.** Our method restricts the set of possible rotations and translations (unit vectors, due to scale ambiguity) for solving the relative pose, by incorporating cheirality constraints in the optimization. We visualize this for the translation with *cost* maps of squared epipolar errors in the tangent space at the ground-truth translation, $\mathcal{T}_t \mathcal{S}^2$, and for different levels of noise. Elements in $\mathcal{T}_t \mathcal{S}^2$ are mapped to the sphere along geodesics using the exponential map, which is a bijective mapping for $\|\mathbf{v}\| \leq \pi$ with $\mathbf{v} \in \mathcal{T}_t \mathcal{S}^2$ [4]. This enables us to show, on the right, the space not satisfying the constraint of Eq. (5) with lower opacity, named \mathcal{R} . As can be seen, the global minimum corresponding to \mathbf{t} , always lies within the unrestricted space, named \mathcal{U} , while it excludes the global minima corresponding to $-\mathbf{t}$. Therefore the solver is able to automatically select the translation with the correct sign as the solution.

Proof. For the *if* direction, assume $\text{Adj}(\mathbf{E}) = \mathbf{q}\mathbf{t}^\top$ and $\text{tr}(\mathbf{E}\mathbf{E}^\top) = 2$. The outer product $\mathbf{q}\mathbf{t}^\top$ is a rank-1 matrix, implying that $\text{rank}(\mathbf{E}) = 2$. Thus, the singular value decomposition (SVD) of \mathbf{E} is $\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, with $\mathbf{D} := \text{diag}(\sigma_0, \sigma_1, 0)$, $\sigma_0, \sigma_1 \in \mathbb{R}^+$ and $\mathbf{U}, \mathbf{V} \in \text{O}(3)$. The adjugate of \mathbf{E} can then be expressed as³:

$$\text{Adj}(\mathbf{E}) = \text{Adj}(\mathbf{U}\mathbf{D}\mathbf{V}^\top), \quad (7)$$

$$= \text{Adj}(\mathbf{V}^\top) \text{Adj}(\mathbf{D}) \text{Adj}(\mathbf{U}), \quad (8)$$

$$= \pm \mathbf{V} \text{diag}(0, 0, \sigma_0\sigma_1) \mathbf{U}^\top, \quad (9)$$

$$= \pm \sigma_0\sigma_1 \mathbf{v}_2 \mathbf{u}_2^\top, \quad (10)$$

where \mathbf{v}_2 and \mathbf{u}_2 are the third columns of \mathbf{V} and \mathbf{U} , respectively. Since the vectors involved in both outer products, $\mathbf{v}_2 \mathbf{u}_2^\top$ and $\mathbf{q}\mathbf{t}^\top$, are unit vectors, this implies that $\sigma_0\sigma_1 = 1$. Furthermore, since $\text{tr}(\mathbf{E}\mathbf{E}^\top) = 2$, it follows⁴ that $\sigma_0^2 + \sigma_1^2 = 2$. These two equations lead to the bi-quadratic equations:

$$\sigma_i^4 - 2\sigma_i^2 + 1 = (\sigma_i - 1)^2(\sigma_i + 1)^2 = 0, \quad i \in \{0, 1\}, \quad (11)$$

which have $\sigma_i = 1$ as positive roots, meaning that \mathbf{E} has two non-zero singular values, both equal to 1. Hence, \mathbf{E} is an essential matrix that belongs to $\mathcal{M}_{\mathbf{E}}$.

For the *only if* direction, assume \mathbf{E} is an essential matrix in $\mathcal{M}_{\mathbf{E}}$. Thus, condition (i) is satisfied, and there exist two unit vectors $\mathbf{q}, \mathbf{t} \in \mathcal{S}^2$ and a rotation matrix $\mathbf{R} \in \text{SO}(3)$, satisfying that $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R} = \mathbf{R}[\mathbf{q}]_\times$ and $\mathbf{q} = \mathbf{R}^\top \mathbf{t}$. We

can show then that

$$\text{Adj}(\mathbf{E}) = \text{Adj}([\mathbf{t}]_\times \mathbf{R}), \quad (12)$$

$$= \text{Adj}(\mathbf{R}) \text{Adj}([\mathbf{t}]_\times), \quad (13)$$

$$= \mathbf{R}^\top \mathbf{t} \mathbf{t}^\top = \mathbf{q} \mathbf{t}^\top. \quad (14)$$

Thus, condition (ii) is also satisfied. \square

3.2. Recovery of the rotation

Assuming tight solutions for $\mathbf{E} \in \mathcal{M}_{\mathbf{E}}$ and $\mathbf{t}, \mathbf{q} \in \mathcal{S}^2$, we can directly recover the rotation $\mathbf{R} \in \text{SO}(3)$ without disambiguation. A (normalized) essential matrix $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$, depends linearly on \mathbf{R} and since $\text{rank}(\mathbf{E}) = 2$, this provides six independent equations for solving \mathbf{R} (its nine elements). Thus, three additional independent equations are needed. Notably, \mathbf{t} lies in the nullspace of $[\mathbf{t}]_\times$, allowing us to find the remaining equations in the definition $\mathbf{q} = \mathbf{R}^\top \mathbf{t}$. Hence, \mathbf{R} can be determined as the solution to this linear system. Since our method is empirically tight, our estimates \mathbf{E}, \mathbf{t} and \mathbf{q} belong to their respective spaces, implying that the resulting \mathbf{R} belongs to $\text{SO}(3)$. While not theoretically necessary, for better numerical accuracy, we can: 1) project the resulting \mathbf{R} onto $\text{SO}(3)$ by classical means [31], and 2) consider the linearly dependent equations stemming from the equivalent definitions: $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R} = \mathbf{R}[\mathbf{q}]_\times$, $\mathbf{q} = \mathbf{R}^\top \mathbf{t}$ and $\mathbf{t} = \mathbf{R}\mathbf{q}$. The corresponding normal equations have as their LHS and RHS terms $2\mathbf{I}_{9 \times 9}$ and $2\mathbf{t}\mathbf{q}^\top - [\mathbf{t}]_\times \mathbf{E} - \mathbf{E}[\mathbf{q}]_\times$, respectively (with the RHS expressed in vectorized form). Thus, \mathbf{R} can be solved in closed-form as

$$\mathbf{R} = \mathbf{t}\mathbf{q}^\top - \frac{1}{2}([\mathbf{t}]_\times \mathbf{E} + \mathbf{E}[\mathbf{q}]_\times). \quad (15)$$

3.3. QCQP

The quadratic nature of our constraints motivates us to formulate the relative pose problem as a Quadratically Con-

³We use the adjugate matrix property $\text{Adj}(\mathbf{A}\mathbf{B}) = \text{Adj}(\mathbf{B}) \text{Adj}(\mathbf{A})$, for any $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$. We also use that $\text{Adj}(\mathbf{Q}) = \det(\mathbf{Q})\mathbf{Q}^\top = (\pm 1)\mathbf{Q}^\top$ for any orthogonal matrix $\mathbf{Q} \in \text{O}(n)$.

⁴Since $\mathbf{E}\mathbf{E}^\top = \mathbf{U}\mathbf{D}^2\mathbf{U}^\top$, it follows that $\text{tr}(\mathbf{E}\mathbf{E}^\top) = \text{tr}(\mathbf{U}\mathbf{D}^2\mathbf{U}^\top) = \text{tr}(\mathbf{D}^2) = \sigma_0^2 + \sigma_1^2$.

strained Quadratic Program (QCQP). Since we parameterize the problem using the essential matrix \mathbf{E} (along with \mathbf{t}, \mathbf{q}), we draw inspiration from the efficient optimization strategy in [69]. We optimize the same cost function. However, we differentiate from [69] in that our approach automatically disambiguates the pose during the optimization, thanks to efficiently constraining the parameter space.

Quadratic optimization In a noise-free scenario, an essential matrix \mathbf{E} , satisfies the epipolar constraints: $\mathbf{f}_{0,i}^\top \mathbf{E} \mathbf{f}_{1,i} = 0$, for any correspondence i [29]. In practice, this does not happen and we instead optimize \mathbf{E} by minimizing the sum of squared normalized epipolar errors [37, 39, 45]:

$$\min \sum_{i=1}^n (\mathbf{f}_{0,i}^\top \mathbf{E} \mathbf{f}_{1,i})^2. \quad (16)$$

Since $\text{vec}(\mathbf{f}_{0,i}^\top \mathbf{E} \mathbf{f}_{1,i}) = (\mathbf{f}_{0,i} \otimes \mathbf{f}_{1,i})^\top \text{vec}(\mathbf{E}^\top)$ [62], this implies that Eq. (16) can be reformulated as:

$$\min \mathbf{e}^\top \mathbf{C} \mathbf{e}, \quad (17)$$

$$\mathbf{e} := \text{vec}(\mathbf{E}^\top), \quad \mathbf{C} := \sum_i (\mathbf{f}_{0,i} \otimes \mathbf{f}_{1,i})(\mathbf{f}_{0,i} \otimes \mathbf{f}_{1,i})^\top, \quad (18)$$

where \otimes is the Kronecker product [62] and $\mathbf{e} \in \mathbb{R}^9$ represents the vector resulting from stacking the rows of \mathbf{E} .

Problem QCQP. We formulate the relative pose problem as the following QCQP:

$$\min_{\mathbf{e}, \mathbf{t}, \mathbf{q}} \mathbf{e}^\top \mathbf{C} \mathbf{e}, \quad (19)$$

$$\text{s.t.} \quad \text{tr}(\mathbf{E} \mathbf{E}^\top) = 2, \quad \text{Adj}(\mathbf{E}) = \mathbf{q} \mathbf{t}^\top, \quad (20)$$

$$\mathbf{t}^\top \mathbf{t} = 1, \quad \mathbf{q}^\top \mathbf{q} = 1, \quad (21)$$

$$\bar{\mathbf{f}}_1 \mathbf{E}^\top [\mathbf{t}]_\times \bar{\mathbf{f}}_0 - s_r^2 = 0, \quad (22)$$

$$h \bar{\mathbf{f}}_0^\top \mathbf{t} - h \bar{\mathbf{f}}_1^\top \mathbf{q} - s_t^2 = 0, \quad (23)$$

$$h^2 = 1. \quad (24)$$

Both the minimization term and the constraints are quadratic. Eqs. (20) and (21) correspond to the constraints presented in Theorem 3.1, and Eqs. (22) to (24) to those presented at the beginning of Sec. 3.1. Thus, we have $d = 18$ parameters and $m = 15$ constraints. As commented in Sec. 3.1, these constraints are necessary and sufficient to disambiguate the relative pose during optimization.

Introduction of inequalities. To transform the proposed inequalities of Eqs. (3) and (5) into compact quadratic equalities that facilitate the use of off-the-shelf SDP solvers (Sec. 4), we use two techniques: 1) we multiply Eq. (23) with an *homogenization variable* h , restricted to the values $\{-1, 1\}$. This introduces a spurious (negative) solution if $h = -1$, but this is trivially checked and corrected [10, 24]. 2) we introduce *slack variables* [20] $s_r^2, s_t^2 \in \mathbb{R}$. Since s_r^2 and s_t^2 are non-negative, this ensures the fulfillment of the

inequalities. Additionally, s_t^2 offers an interesting advantage, as we show in Sec. 5, since it enables the detection of pure rotational motions. Finally, note that we have adopted a modified notation for the bearings: $\bar{\mathbf{f}}_i$, to denote that instead of selecting a random correspondence, we average at runtime the scalar coefficients from Eqs. (22) and (23), stemming from all the correspondences. This approach is motivated to average potential inlier noise in the correspondences. We detail this averaging in Supplementary A.1.

4. SDP relaxation and optimization

Generally, optimizations of QCQPs like the one in Prob. (QCQP) are (nonconvex) NP-hard. However, semidefinite programming (SDP) relaxations, have shown to be a powerful tool to tackle global optimization in computer vision [14]. We draw inspiration from this, and relax the QCQP to an SDP. First, we write Prob. (QCQP) in general form as:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \mathbf{x}^\top \mathbf{C}_0 \mathbf{x}, \quad (25)$$

$$\text{s.t.} \quad \mathbf{x}^\top \mathbf{A}_i \mathbf{x} = b_i, \quad i \in \{1, \dots, m\}, \quad (26)$$

where $\mathbf{x} = [\mathbf{e}^\top, \mathbf{t}^\top, \mathbf{q}^\top, h, s_r, s_t]^\top$, \mathbf{C}_0 is the adaptation of \mathbf{C} in Eq. (18) to the formulation in Eq. (25), *i.e.* $\mathbf{C}_0 \in \mathcal{S}_+^{d,5}$ is a block-diagonal matrix whose only nonzero block is \mathbf{C} , and $\mathbf{A}_i \in \mathcal{S}^d$ are symmetric matrices formed by the reformulation of Eqs. (20) to (24) to the form of Eq. (26)⁶.

To relax the QCQP, we use the cyclic property of the trace to realize that $\text{tr}(\mathbf{x}^\top \mathbf{C}_0 \mathbf{x}) = \text{tr}(\mathbf{C}_0 \mathbf{x} \mathbf{x}^\top) = \text{tr}(\mathbf{C}_0 \mathbf{X})$, where $\mathbf{X} := \mathbf{x} \mathbf{x}^\top \in \mathcal{S}_+^d$ constitutes a lifting of the parameters from \mathbb{R}^d to \mathcal{S}_+^d . Doing similarly for Eq. (26), allows us to define the following SDP:

Problem SDP

$$\min_{\mathbf{X} \in \mathcal{S}_+^d} \text{tr}(\mathbf{C}_0 \mathbf{X}), \quad (27)$$

$$\text{s.t.} \quad \text{tr}(\mathbf{A}_i \mathbf{X}) = b_i, \quad i \in \{1, \dots, m\}, \quad (28)$$

$$\mathbf{X} \succeq 0. \quad (29)$$

SDPs are convex, being solvable globally-optimally in practice [63]. The relaxation comes from not imposing any constraint on \mathbf{X} that ensures that the feasible set of Prob. (SDP) matches that of Prob. (QCQP). If the global minimum of Probs. (SDP) and (QCQP) match, then we say that the relaxation is *tight*. To obtain the solution estimate \mathbf{x}^* we need to recover it from the globally optimal \mathbf{X}^* . Interestingly, unlike other SDP relaxations in the literature [25, 65], in the relative pose problem we cannot assume that tightness is achieved when $\text{rank}(\mathbf{X}^*) = 1$, *i.e.*, that \mathbf{x}^* can be recovered by the factorization $\mathbf{X}^* = \mathbf{x}^* (\mathbf{x}^*)^\top$. Several works

⁵We denote as \mathcal{S}_+^d the set of $d \times d$ positive semidefinite (PSD) matrices.

⁶In practice, off-the-shelf solvers like SDPA [64] allow the use of a sparse representation of the constraints and only store the nonzero values of our constraints and minimization term.

have noted this empirically [10, 34, 69]. For instance, the formulation of [69] can be tight when $\text{rank}(\mathbf{X}^*) = 2$. We prove this in Supplementary B. Inspired by [23], to increase the overall tightness of our solutions, we also include the following redundant quadratic constraints in our method:

$$\mathbf{E}\mathbf{E}^\top = [\mathbf{t}]_\times [\mathbf{t}]_\times^\top, \quad \mathbf{E}^\top \mathbf{E} = [\mathbf{q}]_\times [\mathbf{q}]_\times^\top. \quad (30)$$

We will refer to our method as C2P, and as C2P-fast when the redundant constraints of Equation (30) are not used.

4.1. Relative pose recovery

Given the optimal SDP solution, \mathbf{X}^* , our goal is to recover the optimal relative pose. We define $\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^* := \mathbf{X}_{[1:16,1:16]}^*$ as the top-left 16×16 submatrix of \mathbf{X}^* . Empirically, $\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^*$ exhibits three nonzero singular values of varying magnitudes, while others are close to zero ($\sim 10^{-6}$). This observation aligns with our parameterization, existing three linearly independent vectors that equally minimize Eq. (19) (shown in Theorem 4.1). However, only the vector corresponding to the largest singular value meets the proposed cheirality constraints (Sec. 3.1). The reasoning for this is as follows: \mathbf{X}^* being a PSD matrix, can be decomposed (ignoring the singular values close to 0) as $\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^* = \sum_i^3 \sigma_i \mathbf{v}_i \mathbf{v}_i^\top$ when tight, and with (σ_i, \mathbf{v}_i) being a singular value-vector pair. This implies that 1) Each \mathbf{v}_i minimizes Eq. (19) (see Theorem 4.1), and 2) to satisfy the cheirality constraints, the singular vector \mathbf{v}_i corresponding to the correct solution, must have the biggest singular value in order to satisfy Eqs. (22) and (23). If we are strict, we should take into account the norms of \mathbf{E} , \mathbf{t} and \mathbf{q} to see which singular pair contributes more positively in Eqs. (22) and (23). However, in practice, the solution is consistently found in the dominant singular vector (with biggest σ_i). We show some experiments corroborating this in Figure 4. Given the dominant vector \mathbf{v}_0 , we extract the parameters from its decomposition $\mathbf{v}_0 = [(\mathbf{e}^*)^\top, (\mathbf{t}^*)^\top, (\mathbf{q}^*)^\top, h]^\top$. We then appropriately scale them, and follow the rotation recovery approach of Sec. 3.2. The complete method is outlined in Algorithm 1. For more details, please refer to Supp. A.

Sufficient and necessary condition for global optimality. To alleviate the notation in the following proof, let us define the following auxiliary variables:

$$\mathbf{X}_{\mathbf{e}}^* := \mathbf{X}_{[1:9,1:9]}^*, \quad \mathbf{X}_{\mathbf{tq}}^* := \mathbf{X}_{[10:15,10:15]}^*, \quad (31)$$

$$\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^* := \mathbf{X}_{[1:16,1:16]}^*, \quad \mathbf{v}_{\mathbf{tq}} := [\mathbf{t}^\top, \mathbf{q}^\top]^\top, \quad (32)$$

where $\mathbf{X}_{[i:j,i:j]}^*$ represents the submatrix of \mathbf{X}^* extracted by selecting the rows and columns ranging from index i through index j .

Theorem 4.1. *The semidefinite relaxation of Prob. (QCQP) is tight if and only if $\text{rank}(\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^*) \in [1, 3]$ and its submatrices $\mathbf{X}_{\mathbf{e}}^*$, $\mathbf{X}_{\mathbf{tq}}^*$ are rank-1.*

Proof. For the *only if* part, assume the relaxation is tight. Then, \mathbf{X}^* is contained in the convex hull of the linearly independent rank-1 solutions to the relative pose problem [10]. We can form at most three feasible linearly-independent solutions that minimize the cost equally⁷, e.g.:

$$\mathbf{x}_0 := \begin{bmatrix} \mathbf{e} \\ \mathbf{v}_{\mathbf{tq}} \\ h \end{bmatrix}, \quad \mathbf{x}_1 := \begin{bmatrix} \mathbf{e} \\ -\mathbf{v}_{\mathbf{tq}} \\ h \end{bmatrix}, \quad \mathbf{x}_2 := \begin{bmatrix} \mathbf{e} \\ -\mathbf{v}_{\mathbf{tq}} \\ -h \end{bmatrix}, \quad (33)$$

Note that \mathbf{t} and \mathbf{q} must share the same sign to be feasible. Thus, the convex combination $(\alpha_i \in \mathbb{R}^+, \sum_i \alpha_i = 1, i \in \{0, 1, 2\})$ that forms the globally optimal matrix can be expressed as:

$$\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^* = \alpha_0 \mathbf{x}_0 \mathbf{x}_0^\top + \alpha_1 \mathbf{x}_1 \mathbf{x}_1^\top + \alpha_2 \mathbf{x}_2 \mathbf{x}_2^\top, \quad (34)$$

$$= \begin{bmatrix} \mathbf{e}\mathbf{e}^\top & a_0 \mathbf{e}\mathbf{v}_{\mathbf{tq}}^\top & a_1 h \mathbf{e} \\ a_0 \mathbf{v}_{\mathbf{tq}} \mathbf{e}^\top & \mathbf{v}_{\mathbf{tq}} \mathbf{v}_{\mathbf{tq}}^\top & a_2 h \mathbf{v}_{\mathbf{tq}} \\ a_1 h \mathbf{e}^\top & a_2 h \mathbf{v}_{\mathbf{tq}}^\top & 1 \end{bmatrix}, \quad (35)$$

where $a_0 := (\alpha_0 - \alpha_1 - \alpha_2)$, $a_1 := (\alpha_0 + \alpha_1 - \alpha_2)$, $a_2 := (\alpha_0 - \alpha_1 + \alpha_2)$. Therefore, $\text{rank}(\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^*) \in [1, 3]$ and $\text{rank}(\mathbf{X}_{\mathbf{e}}^*) = \text{rank}(\mathbf{X}_{\mathbf{tq}}^*) = 1$.

For the *if* part, we build upon [69, Theorem 2]. Since \mathbf{X}^* is a positive semidefinite (PSD) matrix, $\mathbf{X}_{\mathbf{e}}^*$ and $\mathbf{X}_{\mathbf{tq}}^*$ are also PSD as they are principal submatrices of \mathbf{X}^* [55]. Given that $\mathbf{X}_{\mathbf{e}}^*$ and $\mathbf{X}_{\mathbf{tq}}^*$ are both rank-1 matrices, it follows that there exist two vectors $\mathbf{e}^* \in \mathbb{R}^9$ and $\mathbf{v}_{\mathbf{tq}}^* \in \mathbb{R}^6$ that fulfill the primal problem's constraints and satisfy $\mathbf{e}^*(\mathbf{e}^*)^\top = \mathbf{X}_{\mathbf{e}}^*$ and $\mathbf{v}_{\mathbf{tq}}^*(\mathbf{v}_{\mathbf{tq}}^*)^\top = \mathbf{X}_{\mathbf{tq}}^*$.

Regarding the rank of $\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^*$, since it is PSD, it can be factorized as $\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^* = \mathbf{L}\mathbf{L}^\top$, where $\mathbf{L} \in \mathbb{R}^{16 \times r}$ and $r := \text{rank}(\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^*)$. Thus, to satisfy the rank-1 property of $\mathbf{X}_{\mathbf{e}}^*$ and $\mathbf{X}_{\mathbf{tq}}^*$, each column k of \mathbf{L} must be given by: $[a_k \mathbf{e}^\top, b_k \mathbf{v}_{\mathbf{tq}}^\top, c_k h]^\top$, for some scalars $a_k, b_k, c_k \in \mathbb{R}$. This constraint limits the rank of $\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^*$ to at most 3, as any additional column in \mathbf{L} would be a linear combination of the existing ones. Therefore, since $\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^*$ must be feasible, this implies that $\text{rank}(\mathbf{X}_{\mathbf{E},\mathbf{tq},h}^*) \in [1, 3]$ and that it is a convex combination of the three linearly independent solutions, stemming from our parameterization of the problem, and thus the relaxation is tight. \square

5. Experiments

To solve Prob. (SDP), we use SDPA [20, 64], a solver for general SDPs based on primal-dual interior-point methods. For the experiments we use an Intel Core i5 CPU at 3.00 GHz. For comparison with non-minimal works, we consider the globally optimal methods of [23, 69].

⁷Recall that the correct solution, besides minimizing the cost term, satisfies the cheirality constraints, and is the singular vector associated with the highest singular value, as commented in Sec. 4.1.

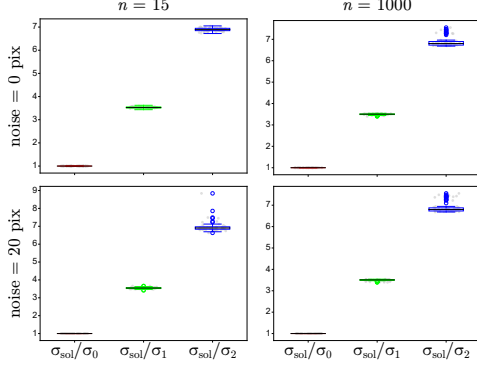


Figure 4. **The solution is found in the dominant singular vector.** We show, across different levels of noise and number of correspondences (we repeat each experiment 100 times), boxplots for the ratios $\sigma_{\text{sol}}/\sigma_0$ (left box), $\sigma_{\text{sol}}/\sigma_1$ (middle box), $\sigma_{\text{sol}}/\sigma_2$ (right box). σ_{sol} corresponds to the singular value whose vector contains the estimate closest to the ground-truth. $\sigma_0, \sigma_1, \sigma_2$ represent, in order ($\sigma_0 > \sigma_1 > \sigma_2$) the top-three singular values (the rest are close to zero). As can be seen, the solution vector consistently corresponds to the dominant singular vector *i.e.* $\sigma_{\text{sol}}/\sigma_0 = 1$. Therefore, we can directly select the dominant vector to recover the solution.

Algorithm 1 C2P: Relative pose without disambiguation

Input: List of correspondences $\{\mathbf{f}_{0,i}, \mathbf{f}_{1,i}\}_{i=1}^n$, threshold $\varepsilon_t \in \mathbb{R}^+$, optional weights $\{w_i\}_{i=1}^n$

Output: $\mathbf{E}^* \in \mathcal{M}_{\mathbf{E}}$, geometrically correct relative pose ($\mathbf{R}^* \in \text{SO}(3), \mathbf{t}^* \in \mathcal{S}^2$), `certif`, `is_pure_rot`

- 1: # Precomputed constraint elements in Eq. (28) form
 - 2: $\{(\mathbf{A}_i, \mathbf{b}_i)\}_{i=1}^{m-2} \leftarrow$ Eqs. (20), (21), (24) and (30)
 - 3: # Cost matrix and data-dependent constraint matrices
 - 4: $\mathbf{C}, \mathbf{A}_r, \mathbf{A}_t \leftarrow$ Eqs. (18), (22) and (23)
 - 5: # Solve SDP and recover geometrically-valid rel. pose
 - 6: $\mathbf{X}^* \leftarrow \text{SDP}(\mathbf{C}, \{(\mathbf{A}_i, \mathbf{b}_i)\}_{i=1}^{m-2}, \mathbf{A}_r, \mathbf{A}_t)$ # *e.g.* [64]
 - 7: $\mathbf{E}^*, \mathbf{t}^*, \mathbf{q}^*, s_t \leftarrow \text{EIG}(\mathbf{X}^*)$ # Sec. 4.1 and Supp. A.2
 - 8: # Global-optimality and (near) pure-rotation certificates
 - 9: `certif` $\leftarrow \text{BOOL}(\mathbf{X}^* \text{ meets Th. 4.1 rank conditions})$
 - 10: `is_pure_rot` $\leftarrow \text{BOOL}(s_t^2 < \varepsilon_t)$
 - 11: # Improve numerical accuracy—if needed (Supp. A.3)
 - 12: **if** $s_t^2 < 10^{-4}$ **then**
 - 13: $\mathbf{E}^*, \mathbf{t}^*, \mathbf{q}^* \leftarrow \text{SIGNED EIG}(\mathbf{X}^*_{[:,15:,15:]})$ # Supp. A.3
 - 14: **end if**
 - 15: $\mathbf{R}^* \leftarrow \text{RECOVERROTATION}(\mathbf{E}^*, \mathbf{t}^*, \mathbf{q}^*)$ # Sec. 3.2
 - 16: **return** $\mathbf{E}^*, \mathbf{R}^*, \mathbf{t}^*, \text{certif}, \text{is_pure_rot}$
-

Synthetic data. Following [10, 23, 34, 69] we test our method with synthetic experiments. To simulate the scenes and cameras, we follow the procedure of [69]. Specifically, we set the absolute pose of one camera to the identity. For the other camera, the direction of its relative translation is uniformly sampled with a maximum magnitude of 2, and the relative rotation is generated with random Euler angles

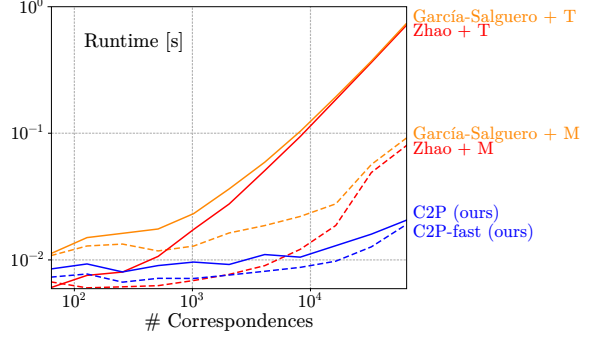
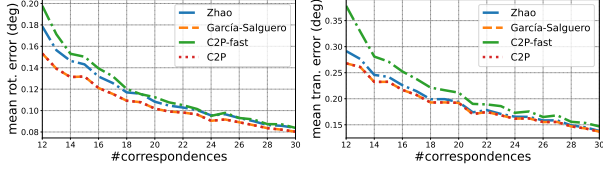


Figure 5. **Run time vs number of correspondences.** We compare the execution time (in sec.) of C2P against Zhao [69] and Garcia-Salguero et al. [23]. Unlike C2P, [23, 69] need a post-processing step to disambiguate the four valid candidate poses. For this, we use two methods: (T) the classic cheirality check [29], that triangulates the points and checks for positive-depths, and (M) A faster alternative, that avoids triangulation and instead checks Eq. (36). C2P-fast and [69] + M, are the fastest when the number of correspondences is low, and there exist small difference (< 2 ms) if we use redundant constraints (C2P). However, for $n > 10^3$ (common in dense matchers [17, 18, 61]) the disambiguation step starts dominating the runtime of [69] + M, while both versions of our method (C2P and C2P-fast) present up to 4x and 35x times better runtimes w.r.t. the fastest, and slowest alternative, respectively.

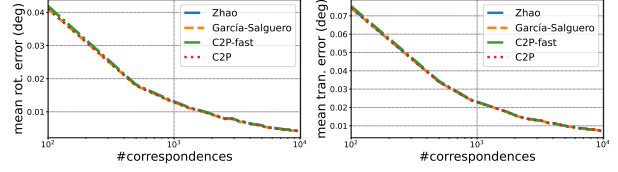
bounded to 0.5 radians in absolute value. This generates random relative poses as they would appear in practical situations. We uniformly sample point correspondences around the origin with a distance varying between 4 and 8 and then transform each to the reference frames of the cameras to obtain the unit bearing vectors. We add noise to the bearings assuming a spherical camera *i.e.* we extract the tangent plane of each bearing and add uniformly distributed random noise expressed in pixels inside this plane.

Real data. Following [23, 69], we additionally consider the six sequences from the dataset [56]. We generate 97 wide-baseline image pairs by grouping adjacent images. For each image pair, we extract correspondences with SIFT features [43]. Then, we use RANSAC to filter out wrong correspondences. These results are consistent with the synthetic scenarios and we provide them in Supplementary D due to space limitations.

Execution time Although our main contributions are theoretical, our method has the advantage of scaling better with the number of correspondences, thanks to not requiring posterior disambiguation. We compare our runtimes with those of [23, 69] in Fig. 5 for varying number of points. Since [23, 69] need to disambiguate the four candidate poses satisfying the same epipolar geometry, we consider two techniques: (T) classic triangulation of the correspondences, followed by a positive-depth check [29] (we use OpenCV’s `recoverPose` [7] for this) and, for fairer comparison, we



(a) Regime R1: $n \in [12, 30]$



(b) Regime R2: $n \in [10^2, 10^4]$

Figure 6. **Accuracy vs number of correspondences.** (a) For small n , both C2P and the method of Garcia-Salguero et al. [23] perform the best, while C2P is faster than [23]. We argue that redundant constraints, such as those leveraged in these methods, are particularly helpful in this regime. (b) As n increases (right) both our C2P and C2P-fast scale better than [23, 69] while having the same accuracy.

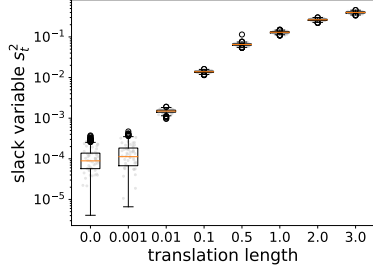


Figure 7. s_t^2 as pure rot. metric. As the translation magnitude decreases, the accuracy of the estimated \mathbf{t} , declines due to its diminished impact on the cost function. Our slack variable, s_t^2 , allows us to directly identify when the magnitude of \mathbf{t} is $\leq 10^{-3}$ of relative to the scene’s scale through simple thresholding on s_t^2 .

also consider a method (M) that avoids triangulation and is based on checking the estimated sign of the norms computed with the midpoint method [3, 38]. Specifically, we select the camera pose that satisfies the most:

$$(\mathbf{R}\mathbf{f}_1 \times \mathbf{f}_0) \cdot (\mathbf{f}_0 \times \mathbf{t}) > 0, \quad (\mathbf{R}\mathbf{f}_1 \times \mathbf{f}_0) \cdot (\mathbf{R}\mathbf{f}_1 \times \mathbf{t}) > 0, \quad (36)$$

for all correspondences. A geometric derivation of Eq. (36) is found in [38]. In Supplementary C we provide an algebraic derivation. As can be seen, C2P-fast and [69] + M, are the fastest methods when the number of correspondences is low. Additionally, there exist small difference (< 2 ms) w.r.t. C2P which includes redundant constraints. However, for $n > 10^3$ (common in dense matchers [17, 18, 61]) the disambiguation starts dominating the runtime of [69] + M, while both C2P and C2P-fast present better scaling.

Accuracy vs number of correspondences We first test the accuracy of the methods w.r.t. the number of correspondences (n). To better visualize their behavior, we consider two regimes: R1) $n \in [12, 30]$ and R2) $n \in [10^2, 10^4]$. We fix the noise to 1 pixel (the same conclusions hold at different noise levels, as shown in Supp. D). For both regimes we repeat the experiments 1000 times. We set the step size of n to 1 in R1 and 400 in R2. In Fig. 6, we report the mean rotation: $\arccos(0.5(\text{tr}(\mathbf{R}_{\text{true}}^\top \mathbf{R}) - 1))$ and translation errors $\arccos(\mathbf{t}_{\text{true}}^\top \mathbf{t})$ in degrees, where $\mathbf{R}_{\text{true}}, \mathbf{t}_{\text{true}}$ is the

ground-truth pose. From this experiment, we conclude that redundant constraints help to improve the accuracy when n is low, as in R1), both C2P and [23], outperform [69], while our C2P is faster (Fig. 5). In R2) all methods are equally accurate, while C2P and C2P-fast start becoming the fastest methods. In practice, we can easily switch between C2P and C2P-fast based on n , thus achieving a good balance in speed and accuracy when compared to the alternatives.

Pure rotations In this experiment, we verify our method’s efficacy under near-pure rotational motions, which is known to be challenging [10, 37]. Besides, the slack variable s_t^2 , corresponding to Eq. (23), enables the detection of such motions through simple thresholding. Intuitively, Eq. (23) corresponds to Eq. (4) and this inequality becomes 0 under pure rotations ($\mathbf{f}_0 = \mathbf{R}\mathbf{f}_1$ in this case). In this experiment, we vary the translation magnitude and do 1000 repetitions for each, setting the noise to 0.5 pixels. Results in Supplementary D, show that the rotation accuracy is unaffected by the translation magnitude, but as this decreases, the estimate \mathbf{t} worsens since the minimization of the epipolar errors become more insensitive to \mathbf{t} . C2P behaves similarly as [23, 69], while C2P has the advantage of directly identifying near-pure rotations by just checking s_t^2 , thus avoiding extra steps such as the required in [69] (see Fig. 7).

6. Conclusion and Limitations

In this paper, we introduced C2P, a novel method for estimating the relative pose that, for the first time in the literature, does not need posterior disambiguation. Our approach efficiently constrains the parameter space during optimization using the necessary and sufficient geometric and manifold constraints, resulting in better runtime with more correspondences and a better balance between accuracy and efficiency compared to the alternatives. Additionally, our method is certifiably globally optimal and can directly detect near-pure rotational motions. C2P, however, as of now, cannot deal with outliers, which is left for future work.

Acknowledgements. This work was supported by the Ministerio de Universidades Scholarship FPU21/04468, the Spanish Government (PID2021-127685NB-I00 and TED2021-131150B-I00) and the Aragón Government (DGA T45_23R).

References

- [1] Pablo Alcantarilla, Jesus Nuevo, and Adrien Bartoli. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *BMVC*. British Machine Vision Association, 2013. [1](#)
- [2] Xiaowei Bao, Nikolaos V Sahinidis, and Mohit Tawarmalani. Semidefinite relaxations for quadratically constrained quadratic programming: A review and comparisons. *Mathematical programming*, 129:129–157, 2011. [2](#)
- [3] Paul A Beardsley, Andrew Zisserman, and David William Murray. Navigation using affine structure from motion. In *ECCV*. Springer, 1994. [8](#), [3](#)
- [4] Nicolas Boumal. *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023. [4](#)
- [5] Nicolas Boumal, Vlad Voroninski, and Afonso Bandeira. The non-convex burer-monteiro approach works on smooth semidefinite programs. In *NeurIPS*. Curran Associates, Inc., 2016. [2](#)
- [6] Stephen P Boyd and Lieven Vandenbergh. *Convex optimization*. Cambridge university press, 2004. [2](#)
- [7] G. Bradski. The OpenCV Library. *Dr. Dobbs's Journal of Software Tools*, 2000. [3](#), [7](#)
- [8] Jesus Briales and Javier Gonzalez-Jimenez. Convex global 3d registration with lagrangian duality. In *CVPR*, 2017. [2](#)
- [9] Jesus Briales and Javier Gonzalez-Jimenez. Cartan-sync: Fast and global se(d)-synchronization. *IEEE RA-L*, 2(4): 2127–2134, 2017. [2](#)
- [10] Jesus Briales, Laurent Kneip, and Javier Gonzalez-Jimenez. A Certifiably Globally Optimal Solution to the Non-Minimal Relative Pose Problem. In *CVPR*, 2018. [1](#), [2](#), [3](#), [5](#), [6](#), [7](#), [8](#)
- [11] Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical programming*, 95(2): 329–357, 2003. [2](#)
- [12] Samuel Burer and Kyungchan Park. A strengthened sdp relaxation for quadratic optimization over the stiefel manifold. *Journal of optimization theory and applications*, pages 1–20, 2023. [3](#)
- [13] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. *IEEE T-RO*, 37(6):1874–1890, 2021. [1](#)
- [14] Diego Cifuentes, Sameer Agarwal, Pablo A Parrilo, and Rekha R Thomas. On the local stability of semidefinite relaxations. *Mathematical Programming*, pages 1–35, 2022. [5](#)
- [15] Frank Dellaert, David M Rosen, Jing Wu, Robert Mahony, and Luca Carlone. Shonan rotation averaging: Global optimality by surfing so(p)ⁿ so(p)ⁿ. In *ECCV*, pages 292–308. Springer, 2020. [2](#)
- [16] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *CVPRW*, 2018. [1](#)
- [17] Johan Edstedt, Ioannis Athanasiadis, Mårten Wadenbäck, and Michael Felsberg. DKM: Dense Kernelized Feature Matching for Geometry Estimation. In *CVPR*, 2023. [1](#), [7](#), [8](#)
- [18] Johan Edstedt, Qiyu Sun, Georg Bökman, Mårten Wadenbäck, and Michael Felsberg. RoMa: Revisiting Robust Losses for Dense Feature Matching. *arXiv preprint arXiv:2305.15404*, 2023. [7](#), [8](#)
- [19] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. [1](#)
- [20] Katsuki Fujisawa, Masakazu Kojima, Kazuhide Nakata, and Makoto Yamashita. Sdpa (semidefinite programming algorithm) user’s manual—version 6.2. 0. *Department of Mathematical and Computing Sciences, Tokyo Institute of Technology. Research Reports on Mathematical and Computing Sciences Series B: Operations Research*, 2002. [5](#), [6](#)
- [21] Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez. Fast and robust certifiable estimation of the relative pose between two calibrated cameras. *Journal of Mathematical Imaging and Vision*, 63(8):1036–1056, 2021. [2](#)
- [22] Mercedes Garcia-Salguero, Jesus Briales, and Javier Gonzalez-Jimenez. Certifiable relative pose estimation. *Image and Vision Computing*, 109:104142, 2021. [2](#)
- [23] Mercedes Garcia-Salguero, Jesus Briales, and Javier Gonzalez-Jimenez. A tighter relaxation for the relative pose problem between cameras. *Journal of Mathematical Imaging and Vision*, 64(5):493–505, 2022. [2](#), [3](#), [6](#), [7](#), [8](#), [1](#), [4](#)
- [24] Matthew Peter Giamou. *Semidefinite Relaxations for Geometric Problems in Robotics*. PhD thesis, University of Toronto (Canada), 2023. [3](#), [5](#)
- [25] Linus Härenstam-Nielsen, Niclas Zeller, and Daniel Cremers. Semidefinite relaxations for robust multiview triangulation. In *CVPR*, 2023. [2](#), [5](#)
- [26] R.I. Hartley. In defense of the eight-point algorithm. *IEEE TPAMI*, 19(6):580–593, 1997. [2](#)
- [27] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *IJCV*, 103:267–305, 2013. [1](#)
- [28] Richard I Hartley and Fredrik Kahl. Global optimization through rotation space search. *IJCV*, 82(1):64–79, 2009. [2](#)
- [29] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. [1](#), [3](#), [5](#), [7](#)
- [30] Uwe Helmke, Knut Hüper, Pei Yean Lee, and John Moore. Essential matrix estimation using gauss-newton iterations on a manifold. *IJCV*, 74:117–136, 2007. [2](#)
- [31] Nicholas J Higham. Matrix nearness problems and applications. *Applications of matrix theory*, 22, 1989. [4](#)
- [32] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge university press, 2012. [3](#)
- [33] M. J. Todd K. C. Toh and R. H. Tütüncü. SDPT3 — A Matlab software package for semidefinite programming, Version 1.3. *Optimization Methods and Software*, 11(1-4):545–581, 1999. [2](#)
- [34] Arman Karimian and Roberto Tron. Essential matrix estimation using convex relaxations in orthogonal space. In *ICCV*, pages 17142–17152, 2023. [2](#), [3](#), [6](#), [7](#)
- [35] Laurent Kneip and Paul Furgale. Opengv: A unified and generalized approach to real-time calibrated geometric vision. In *IEEE ICRA*, pages 1–8, 2014. [3](#)

- [36] Laurent Kneip and Simon Lymn. Direct optimization of frame-to-frame rotation. In *ICCV*, 2013. 2
- [37] Laurent Kneip, Roland Siegwart, and Marc Pollefeys. Finding the exact rotation between two images independently of the translation. In *ECCV*, pages 696–709. Springer, 2012. 3, 5, 8
- [38] Seong Hun Lee and Javier Civera. Triangulation: Why optimize? In *BMVC*, 2019. 8
- [39] Seong Hun Lee and Javier Civera. Geometric interpretations of the normalized epipolar error. *arXiv preprint arXiv:2008.01254*, 2020. 5
- [40] Seong Hun Lee and Javier Civera. Hara: A hierarchical approach for robust rotation averaging. In *CVPR*, pages 15777–15786, 2022. 1
- [41] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. In *ICCV*, 2023. 1
- [42] H Christopher Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(5828):133–135, 1981. 1, 2
- [43] David G Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004. 1, 7, 3
- [44] Pierre Moulon, Pascal Monasse, Romuald Perrot, and Renaud Marlet. OpenMVG: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*, pages 60–74. Springer, 2016. 1
- [45] Dominik Muhle, Lukas Koestler, Nikolaus Demmel, Florian Bernard, and Daniel Cremers. The probabilistic normal epipolar constraint for frame-to-frame rotation optimization under uncertain feature positions. In *CVPR*, 2022. 5
- [46] Raúl Mur-Artal and Juan D. Tardós. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE T-RO*, 33(5):1255–1262, 2017. 1
- [47] Raúl Mur-Artal, J. M. M. Montiel, and Juan D. Tardós. Orb-slam: A versatile and accurate monocular slam system. *IEEE T-RO*, 31(5):1147–1163, 2015. 1
- [48] D. Nister. An efficient solution to the five-point relative pose problem. *IEEE TPAMI*, 26(6):756–770, 2004. 1, 3
- [49] Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm. Usac: A universal framework for random sample consensus. *IEEE TPAMI*, 35(8):2022–2038, 2012. 1
- [50] David M Rosen, Luca Carlone, Afonso S Bandeira, and John J Leonard. Se-sync: A certifiably correct algorithm for synchronization over the special euclidean group. *The International Journal of Robotics Research*, 38(2-3):95–125, 2019. 2
- [51] Raman Sanyal, Frank Sottile, and Bernd Sturmfels. Orbitopes. *Mathematika*, 57(2):275–314, 2011. 3
- [52] J. Saunderson, P. A. Parrilo, and A. S. Willsky. Semidefinite descriptions of the convex hull of rotation matrices. *SIAM Journal on Optimization*, 25(3):1314–1343, 2015. 3
- [53] Johannes L. Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, 2016. 1, 3
- [54] Henrik Stewénius, Christopher Engels, and David Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284–294, 2006. 1
- [55] Gilbert Strang and Linear Algebra. its applications. *Academic Press, New York*, 14:181208, 1980. 6, 2
- [56] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *CVPR*, 2008. 7, 3, 4
- [57] Jos F. Sturm. Using sedumi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11(1-4):625–653, 1999. 2
- [58] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. Loftr: Detector-free local feature matching with transformers. In *CVPR*, 2021. 1
- [59] Roberto Tron and Kostas Daniilidis. On the quotient representation for the essential manifold. In *CVPR*, 2014. 2
- [60] Roberto Tron and Kostas Daniilidis. The space of essential matrices as a riemannian quotient manifold. *SIAM Journal on Imaging Sciences*, 10(3):1416–1445, 2017. 2, 3
- [61] Prune Truong, Martin Danelljan, Radu Timofte, and Luc Van Gool. PDC-Net+: Enhanced Probabilistic Dense Correspondence Network. *IEEE TPAMI*, 45(8):10247–10266, 2023. 1, 7, 8
- [62] Charles F Van Loan. The ubiquitous kronecker product. *Journal of computational and applied mathematics*, 123(1-2):85–100, 2000. 5
- [63] Lieven Vandenbergh and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996. 5
- [64] Makoto Yamashita, Katsuki Fujisawa, Kazuhide Nakata, Maho Nakata, Mituhiro Fukuda, Kazuhiro Kobayashi, and Kazushige Goto. A high-performance software package for semidefinite programs: Sdpa 7. Technical report, 2010. 5, 6, 7, 2
- [65] Heng Yang and Luca Carlone. A quaternion-based certifiably optimal solution to the wahba problem with outliers. In *ICCV*, 2019. 2, 5
- [66] Heng Yang, Pasquale Antonante, Vasileios Tzoumas, and Luca Carlone. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. *IEEE RA-L*, 5(2):1127–1134, 2020. 1
- [67] Heng Yang, Jingnan Shi, and Luca Carlone. Teaser: Fast and certifiable point cloud registration. *IEEE T-RO*, 37(2):314–333, 2021. 2
- [68] Jiaolong Yang, Hongdong Li, and Yunde Jia. Optimal essential matrix estimation via inlier-set maximization. In *ECCV*. Springer, 2014. 2
- [69] Ji Zhao. An efficient solution to non-minimal case essential matrix estimation. *IEEE TPAMI*, 44(4):1777–1792, 2022. 1, 2, 3, 5, 6, 7, 8, 4
- [70] Ji Zhao, Wanting Xu, and Laurent Kneip. A certifiably globally optimal solution to generalized essential matrix estimation. In *CVPR*, 2020. 2

From Correspondences to Pose:

Non-minimal Certifiably Optimal Relative Pose without Disambiguation

Supplementary Material

In this supplementary material, we provide additional details of our method in Supplementary A, proofs in Supplementaries B and C and experiments in Supplementary D.

A. Additional details

A.1. Averaging of data-dependent constraints

We provide here compact expressions for averaging the data-dependent coefficients of the quadratic terms stemming from Eqs. (22) and (23). We will use the notation $\mathbf{a}^{(j)}$ to refer to the j -th element of a vector \mathbf{a} .

Rotation. For clarity, we recall Eq. (22):

$$\bar{\mathbf{f}}_1 \mathbf{E}^\top [\mathbf{t}]_\times \bar{\mathbf{f}}_0 - s_r^2 = 0, \quad (37)$$

The quadratic terms for one correspondence $(\mathbf{f}_0, \mathbf{f}_1)$ are:

$$\begin{aligned} & \mathbf{f}_0^{(0)} \mathbf{f}_1^{(0)} \mathbf{e}^{(3)} \mathbf{t}^{(2)} + \mathbf{f}_0^{(0)} \mathbf{f}_1^{(1)} \mathbf{e}^{(4)} \mathbf{t}^{(2)} + \mathbf{f}_0^{(0)} \mathbf{f}_1^{(2)} \mathbf{e}^{(5)} \mathbf{t}^{(2)} \\ & + \mathbf{f}_0^{(1)} \mathbf{f}_1^{(0)} \mathbf{e}^{(6)} \mathbf{t}^{(0)} + \mathbf{f}_0^{(1)} \mathbf{f}_1^{(1)} \mathbf{e}^{(7)} \mathbf{t}^{(0)} + \mathbf{f}_0^{(1)} \mathbf{f}_1^{(2)} \mathbf{e}^{(8)} \mathbf{t}^{(0)} \\ & + \mathbf{f}_0^{(2)} \mathbf{f}_1^{(0)} \mathbf{e}^{(0)} \mathbf{t}^{(1)} + \mathbf{f}_0^{(2)} \mathbf{f}_1^{(1)} \mathbf{e}^{(1)} \mathbf{t}^{(1)} + \mathbf{f}_0^{(2)} \mathbf{f}_1^{(2)} \mathbf{e}^{(2)} \mathbf{t}^{(1)} \\ & - \mathbf{f}_0^{(0)} \mathbf{f}_1^{(0)} \mathbf{e}^{(6)} \mathbf{t}^{(1)} - \mathbf{f}_0^{(0)} \mathbf{f}_1^{(1)} \mathbf{e}^{(7)} \mathbf{t}^{(1)} - \mathbf{f}_0^{(0)} \mathbf{f}_1^{(2)} \mathbf{e}^{(8)} \mathbf{t}^{(1)} \quad (38) \\ & - \mathbf{f}_0^{(1)} \mathbf{f}_1^{(0)} \mathbf{e}^{(0)} \mathbf{t}^{(2)} - \mathbf{f}_0^{(1)} \mathbf{f}_1^{(1)} \mathbf{e}^{(1)} \mathbf{t}^{(2)} - \mathbf{f}_0^{(1)} \mathbf{f}_1^{(2)} \mathbf{e}^{(2)} \mathbf{t}^{(2)} \\ & - \mathbf{f}_0^{(2)} \mathbf{f}_1^{(0)} \mathbf{e}^{(3)} \mathbf{t}^{(0)} - \mathbf{f}_0^{(2)} \mathbf{f}_1^{(1)} \mathbf{e}^{(4)} \mathbf{t}^{(0)} - \mathbf{f}_0^{(2)} \mathbf{f}_1^{(2)} \mathbf{e}^{(5)} \mathbf{t}^{(0)} \\ & - s_r^2 = 0, \end{aligned}$$

where, $\mathbf{e} := \text{vec}(\mathbf{E}^\top)$, as defined in the main paper.

With this ordering, the coefficients of the first nine terms of Eq. (38) (from $\mathbf{f}_0^{(0)} \mathbf{f}_1^{(0)}$ to $\mathbf{f}_0^{(2)} \mathbf{f}_1^{(2)}$) can be computed as $\text{vec}(\mathbf{f}_1 \mathbf{f}_0^\top)$, which is a nine-dimensional vector (one element per coefficient). Thus, the averaging of the terms across n correspondences $\{\mathbf{f}_{0,i}, \mathbf{f}_{1,i}\}_{i=1}^n$ can be expressed as:

$$\text{vec} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{f}_{1,i} \mathbf{f}_{0,i}^\top \right). \quad (39)$$

Furthermore, the subsequent nine quadratic terms have the same (but negated) coefficients. Thus, the values of Eq. (39) can be reused for these coefficients.

Translation. For clarity, we recall Eq. (23):

$$h \bar{\mathbf{f}}_0^\top \mathbf{t} - h \bar{\mathbf{f}}_1^\top \mathbf{q} - s_t^2 = 0, \quad (40)$$

The quadratic terms for one correspondence $(\mathbf{f}_0, \mathbf{f}_1)$ are:

$$\begin{aligned} & h \mathbf{f}_0^{(0)} \mathbf{t}^{(0)} + h \mathbf{f}_0^{(1)} \mathbf{t}^{(1)} + h \mathbf{f}_0^{(2)} \mathbf{t}^{(2)} \\ & - h \mathbf{f}_1^{(0)} \mathbf{q}^{(0)} - h \mathbf{f}_1^{(1)} \mathbf{q}^{(1)} - h \mathbf{f}_1^{(2)} \mathbf{q}^{(2)} \quad (41) \\ & - s_t^2 = 0 \end{aligned}$$

In this case the coefficients of the quadratic terms are directly given by the bearings. Thus, the average for the first three coefficients can be computed as $1/n \sum_i \mathbf{f}_0$, and as $-1/n \sum_i \mathbf{f}_1$ for the subsequent three coefficients.

A.2. Appropriate scaling of the solution estimates

As explained in Sec. 4.1, we extract the solution estimates from the dominant singular vector, denoted as \mathbf{v}_0 , of $\mathbf{X}_{\mathbf{E}, \mathbf{t}, \mathbf{q}, h}^* := \mathbf{X}_{[1:16, 1:16]}^*$. Following the ordering and notation of the main paper, this corresponds to $\mathbf{v}_0 = [(\mathbf{e}^*)^\top, (\mathbf{t}^*)^\top, (\mathbf{q}^*)^\top, h]^\top$. However, the norm constraints enforced during the optimization, namely $\mathbf{t}^\top \mathbf{t} = 1$, $\mathbf{q}^\top \mathbf{q} = 1$ and $\text{tr}(\mathbf{E} \mathbf{E}^\top) = 2$, apply to \mathbf{X}^* and not to \mathbf{v}_0 . Consequently, we cannot assume that the elements of this dominant vector will be scaled appropriately even after multiplying it with its singular value. The solution to this is straightforward: we separately normalize the vectors \mathbf{t} and \mathbf{q} to make them unit vectors, and scale $\mathbf{e} := \text{vec}(\mathbf{E}^\top)$ such that its nonzero singular values equal 1 (in practice, we use the SVD of \mathbf{E} for greater precision). Finally, we leverage the absence of products between the slack variable s_t and the rest of the parameters in Prob. (QCQP) to directly read s_t^2 from its corresponding diagonal entry in \mathbf{X}^* , thus avoiding the need to factorize \mathbf{X}^* to obtain its value.

A.3. Pure rotations and numerical accuracy

Under pure rotations, considering an optimal essential matrix, \mathbf{E}^* , any pair of translation vectors $\mathbf{t}, \mathbf{q} \in \mathcal{S}^2$ satisfying the definition $\mathbf{q} := \mathbf{R}^\top \mathbf{t}$ will minimize the sum of squared epipolar errors. Here, $\mathbf{R} \in \text{SO}(3)$ represents one of the two rotation matrices corresponding to \mathbf{E}^* . Given that both \mathbf{t}, \mathbf{q} belong to \mathcal{S}^2 , one might expect to find two additional singular vectors corresponding to nonzero singular values, in addition to the three singular vectors mentioned in Sec. 4.1. However, we empirically verified that four additional singular vectors appear instead. We observed the same phenomenon in [23, 69]. This phenomenon likely occurs because the constraints apply only to the optimal matrix \mathbf{X}^* . Therefore, the elements in the singular vectors of \mathbf{X}^* do not need to satisfy the norm constraints (e.g. that \mathbf{t}, \mathbf{q} belong to \mathcal{S}^2) to still minimize the cost function. This may explain the similar behavior noted in [10] regarding pure rotations.

Importantly, in our case, the correct solution can be extracted from the dominant singular vector thanks to Eq. (22), which enforces a larger singular value corresponding to the vector containing the solution. However, for pure rotations and in absence of noise, the component in

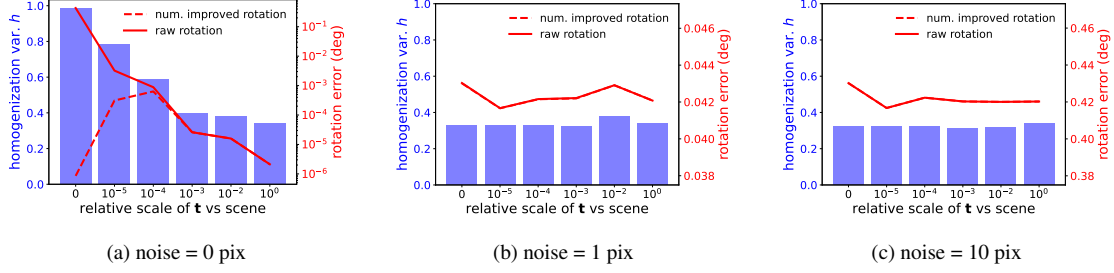


Figure 8. **Noise-free pure rotation scenarios.** For our method, an edge case consists of noise-free pure rotational motions. In noise-free scenarios (a) the estimate of h within the dominant singular vector—which contains the rest of the solution estimates—approaches 1 in near-pure rotational motions (when the relative scale is $< 10^{-4}$), negatively affecting the numerical accuracy of the other estimates. As can be seen, we effectively address this by using the dominant singular vector from the submatrix excluding h . However, despite the effectiveness of this solution, this numerical issue is not present in practical scenarios (b), (c), where noise affects the observations. These visualizations depict results averaged across 1000 different random instances of the same synthetic scenarios considered in the main paper.

the (unit) singular vector corresponding to the homogenization variable, h , dominates the rest, being close to ~ 1 . This predominance reduces the numerical accuracy of the other estimates (\mathbf{e} , \mathbf{t} and \mathbf{q}). Since this behavior is only present in a noise-free scenario, we can use a strict threshold in the slack variable s_t^2 (we use 10^{-4}) to detect such scenario. Consequently, only in this case, we extract the solution from the dominant singular vector of the submatrix corresponding only to \mathbf{e} , \mathbf{t} and \mathbf{q} , leveraging the previous numerically-inaccurate solution to just correct the sign of this new numerically-accurate solution, if necessary. This behavior is shown in Fig. 8.

B. Tightness of [69] when the SDP solution is rank-2

In [69, Eq. 11] the following QCQP is considered:

Problem QCQP-Z

$$\min_{\mathbf{E}, \mathbf{t}} \quad \mathbf{e}^\top \mathbf{C} \mathbf{e}, \quad (42)$$

$$\text{s.t.} \quad \mathbf{E} \mathbf{E}^\top = [\mathbf{t}]_\times [\mathbf{t}]_\times^\top, \quad \mathbf{t}^\top \mathbf{t} = 1. \quad (43)$$

The tightness conditions in [69, Th. 2] assume that tightness of the semidefinite relaxation imply $\text{rank}(\mathbf{X}^*) = 1$, where \mathbf{X}^* represents the optimal solution of the SDP. In this section, we adapt Theorem 4.1 to extend [69, Th. 2] and show that Prob. (QCQP-Z) can also be tight when $\text{rank}(\mathbf{X}^*) = 2$.

With a similar notation as in the main paper, let us define the following auxiliary variables:

$$\mathbf{X}_e^* := \mathbf{X}_{[1:9,1:9]}^*, \quad \mathbf{X}_t^* := \mathbf{X}_{[10:12,10:12]}^*. \quad (44)$$

Theorem B.1. *The semidefinite relaxation of Prob. (QCQP-Z) is tight if and only if $\text{rank}(\mathbf{X}^*) \in [1, 2]$, and its submatrices \mathbf{X}_e^* and \mathbf{X}_t^* are rank-1.*

Proof. For the *only if* direction, assume the relaxation is tight. Then, following [10], we can find \mathbf{X}^* in the convex

hull of the linearly independent rank-1 solutions to the relative pose problem⁸:

$$\mathbf{X}^* := \alpha_0 \mathbf{x}_0 \mathbf{x}_0^\top + \alpha_1 \mathbf{x}_1 \mathbf{x}_1^\top, \quad (45)$$

$$\mathbf{x}_0 := \begin{bmatrix} \mathbf{e} \\ \mathbf{t} \end{bmatrix}, \quad \mathbf{x}_1 := \begin{bmatrix} \mathbf{e} \\ -\mathbf{t} \end{bmatrix}, \quad (46)$$

where α_0, α_1 are non-negative scalars such that $\alpha_0 + \alpha_1 = 1$. This last condition ensures that the cost is optimal, *i.e.*, $\text{tr}(\mathbf{C}_0 \mathbf{X}^*) = \mathbf{e}^\top \mathbf{C} \mathbf{e}$, and that the resulting matrix \mathbf{X}^* is feasible. To see this, we can expand Eq. (45):

$$\mathbf{X}^* = \begin{bmatrix} (\alpha_0 + \alpha_1) \mathbf{e} \mathbf{e}^\top & (\alpha_0 - \alpha_1) \mathbf{e} \mathbf{t}^\top \\ (\alpha_0 - \alpha_1) \mathbf{t} \mathbf{e}^\top & (\alpha_0 + \alpha_1) \mathbf{t} \mathbf{t}^\top \end{bmatrix}, \quad (47)$$

to verify that $\alpha_0 + \alpha_1 = 1$ is needed to satisfy the norm constraint $\mathbf{t}^\top \mathbf{t} = 1$ (the rest of the constraints are satisfied for any valid combination of α_0 and α_1). This reveals that when the semidefinite relaxation is tight, the diagonal (upper-left and bottom-right) block matrices are rank-1 and that $\text{rank}(\mathbf{X}^*) \in \{1, 2\}$. Specifically⁹, $\text{rank}(\mathbf{X}^*) = 1$ when $\alpha_0 = 0$ and $\alpha_1 = 1$ or when $\alpha_0 = 1$ and $\alpha_1 = 0$. Otherwise $\text{rank}(\mathbf{X}^*) = 2$.

For the *if* part, we build upon [69, Theorem 2]. Since \mathbf{X}^* is a positive semidefinite (PSD) matrix, \mathbf{X}_e^* and \mathbf{X}_t^* are also PSD as they are principal submatrices of \mathbf{X}^* [55]. Given that \mathbf{X}_e^* and \mathbf{X}_t^* are both rank-1 matrices, it follows that there exist two vectors $\mathbf{e}^* \in \mathbb{R}^9$ and $\mathbf{t}^* \in \mathbb{R}^3$ that fulfill the primal problem's constraints and satisfy $\mathbf{e}^* (\mathbf{e}^*)^\top = \mathbf{X}_e^*$ and $\mathbf{t}^* (\mathbf{t}^*)^\top = \mathbf{X}_t^*$.

Regarding the rank of \mathbf{X}^* , since it is PSD, it can be factorized as $\mathbf{X}^* = \mathbf{L} \mathbf{L}^\top$, where $\mathbf{L} \in \mathbb{R}^{12 \times r}$ and $r := \text{rank}(\mathbf{X}^*)$. Thus, to satisfy the rank-1 property of \mathbf{X}_e^* and

⁸Note that the outer products of the negative counterparts, $[-\mathbf{e}^\top, -\mathbf{t}^\top]$ and $[-\mathbf{e}^\top, \mathbf{t}^\top]$, are not included, as they yield the same outer product.

⁹In practice, off-the-shelf SDP solvers [33, 57, 64] return a rank-2 block-diagonal solution [23], which corresponds to setting $\alpha_0 = \alpha_1 = 0.5$ in Eqs. (45) and (47).

\mathbf{X}_t^* , each column k of \mathbf{L} must be given by: $[a_k \mathbf{e}^\top, b_k \mathbf{t}^\top]^\top$, for some scalars $a_k, b_k \in \mathbb{R}$. This constraint limits the rank of \mathbf{X}^* to at most 2, as any additional column in \mathbf{L} would be a linear combination of the existing ones. Therefore, since \mathbf{X}^* must be feasible, this implies that $\text{rank}(\mathbf{X}^*) \in [1, 2]$ and that it is a convex combination of the two linearly independent solutions, stemming from Eq. (45), and thus the relaxation is tight. \square

C. Algebraic derivation of Equation (36)

Given estimates of the relative rotation and translation (\mathbf{R}, \mathbf{t}) , and a correspondence $(\mathbf{f}_0, \mathbf{f}_1)$, the midpoint method triangulates the corresponding 3D point $\mathbf{p} \in \mathbb{R}^3$. It identifies this point as the midpoint (mean) of the common perpendicular to the two rays originating from the bearings [3]. Specifically, it determines the norms $\lambda_0, \lambda_1 \in \mathbb{R}$ of the 3D points, $\mathbf{p}_0 := \lambda_0 \mathbf{f}_0, \mathbf{p}_1 := \lambda_1 \mathbf{f}_1$, in each camera reference system, that minimize the squared error $\|\mathbf{p}_0 - (\mathbf{R}\mathbf{p}_1 + \mathbf{t})\|^2$:

$$\lambda_0, \lambda_1 = \arg \min_{\lambda_0, \lambda_1} \|\lambda_0 \mathbf{f}_0 - (\lambda_1 \mathbf{R}\mathbf{f}_1 + \mathbf{t})\|^2. \quad (48)$$

If the 3D points and their midpoint (mean) satisfy the cheirality constraints, both norms λ_0 and λ_1 will be positive. Otherwise, at least one of the norms will be estimated as negative [60]. As will be shown, it is not necessary to explicitly compute λ_0 and λ_1 to estimate their signs.

The rays of ideal, noise-free correspondences meet in a 3D point, satisfying $\lambda_0 \mathbf{f}_0 - \lambda_1 \mathbf{R}\mathbf{f}_1 = \mathbf{t}$, or in matrix form:

$$\underbrace{\begin{bmatrix} \mathbf{f}_0 & -\mathbf{R}\mathbf{f}_1 \end{bmatrix}}_{\mathbf{A} \in \mathbb{R}^{3 \times 2}} \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} = \mathbf{t} \quad (49)$$

In practice, we minimize the squared errors. As such, an equivalent solution to Eq. (48) is given as the solution to the system $\mathbf{A}^\top \mathbf{A} [\lambda_0, \lambda_1]^\top = \mathbf{A}^\top \mathbf{t}$:

$$\begin{bmatrix} 1 & -\mathbf{f}_0^\top \mathbf{R}\mathbf{f}_1 \\ -\mathbf{f}_0^\top \mathbf{R}\mathbf{f}_1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_0^\top \\ -(\mathbf{R}\mathbf{f}_1)^\top \end{bmatrix} \mathbf{t}, \quad (50)$$

where we have used that $(\mathbf{f}_k)^\top \mathbf{f}_k = 1, k \in \{0, 1\}$ since $\mathbf{f}_0, \mathbf{f}_1 \in \mathcal{S}^2$. Expanding Equation (50) leads to:

$$\lambda_0 - \lambda_1 \mathbf{f}_0^\top \mathbf{R}\mathbf{f}_1 = \mathbf{f}_0^\top \mathbf{t}, \quad (51)$$

$$\lambda_1 - \lambda_0 \mathbf{f}_0^\top \mathbf{R}\mathbf{f}_1 = -(\mathbf{R}\mathbf{f}_1)^\top \mathbf{t}. \quad (52)$$

which leads to the equivalent equations:

$$s^2 \lambda_1 = -(\mathbf{R}\mathbf{f}_1)^\top \mathbf{t} + (\mathbf{f}_0^\top \mathbf{R}\mathbf{f}_1)(\mathbf{f}_0^\top \mathbf{t}), \quad (53)$$

$$s^2 \lambda_0 = \mathbf{f}_0^\top \mathbf{t} - (\mathbf{f}_0^\top \mathbf{R}\mathbf{f}_1)((\mathbf{R}\mathbf{f}_1)^\top \mathbf{t}). \quad (54)$$

where

$$s^2 := 1 - (\mathbf{f}_0^\top \mathbf{R}\mathbf{f}_1)^2 = \sin^2 \angle(\mathbf{f}_0, \mathbf{R}\mathbf{f}_1). \quad (55)$$

Since $s^2 \geq 0$, this implies that the RHS of Eqs. (53) and (54) must be positive for λ_0, λ_1 to be positive too:

$$-(\mathbf{R}\mathbf{f}_1)^\top \mathbf{t} + (\mathbf{f}_0^\top \mathbf{R}\mathbf{f}_1)(\mathbf{f}_0^\top \mathbf{t}) > 0, \quad (56)$$

$$\mathbf{f}_0^\top \mathbf{t} - (\mathbf{f}_0^\top \mathbf{R}\mathbf{f}_1)((\mathbf{R}\mathbf{f}_1)^\top \mathbf{t}) > 0, \quad (57)$$

Lastly, to express Eqs. (56) and (57) in compact form, we can use the property of the cross product:

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}), \quad (58)$$

for any $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^3$, and with $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^\top \mathbf{b}$ representing the dot product between any vectors \mathbf{a}, \mathbf{b} . With this property, we reach the inequalities:

$$(\mathbf{R}\mathbf{f}_1 \times \mathbf{f}_0) \cdot (\mathbf{f}_0 \times \mathbf{t}) > 0, \quad (59)$$

$$(\mathbf{R}\mathbf{f}_1 \times \mathbf{f}_0) \cdot (\mathbf{R}\mathbf{f}_1 \times \mathbf{t}) > 0, \quad (60)$$

corresponding thus to Eq. (36).

D. Additional experiments

D.1. Accuracy vs noise and translation magnitude

In Figure 9, we show additional synthetic experiments. In Fig. 9(a)-(e), we verify that the conclusions drawn in Fig. 6 are consistent across different levels of noise. In regimes with a small number of points (Fig. 9(a),(c)), our C2P and [23] perform the best. Notably, C2P is faster than [23] (see Fig. 5) and slightly outperforms it in estimating translation (Fig. 9(c)). In regimes with a large number of points, the accuracy of our faster version of C2P is on-par with C2P itself and [23]. In Fig. 9(e), we fix the number of points at 1000 while varying the noise level and observe the same behavior. Finally, in Fig. 9(f), we demonstrate that C2P performs as well as [23, 69] when varying the scale of the translation w.r.t. the scene, and unlike them, C2P is also capable of directly detecting near-pure rotational motions and does not need posterior disambiguation step.

D.2. Real-data

Following [23, 69], we test our method on all the sequences from Strecha et al. [56]. We generate 97 wide-baseline image pairs by grouping adjacent images. For each image pair, correspondences are extracted using DoG + SIFT [43]. We then use the RANSAC implementation of OpenGV [35] to filter out wrong correspondences, setting the inlier threshold to 5 pix, which we found sufficient given the images resolution of 3072×2048 pix. The performance of C2P(-fast) and [23, 69] is shown in Fig. 10. The results align with those from the synthetic experiments. C2P-fast is the fastest among all the methods. However, C2P-fast is not always tight, resulting in a slight loss of accuracy with respect to the alternatives. On the other hand, our C2P is significantly more accurate than [69] and is on-par with [23]. Additionally, our C2P is 40% faster than [23] on average.

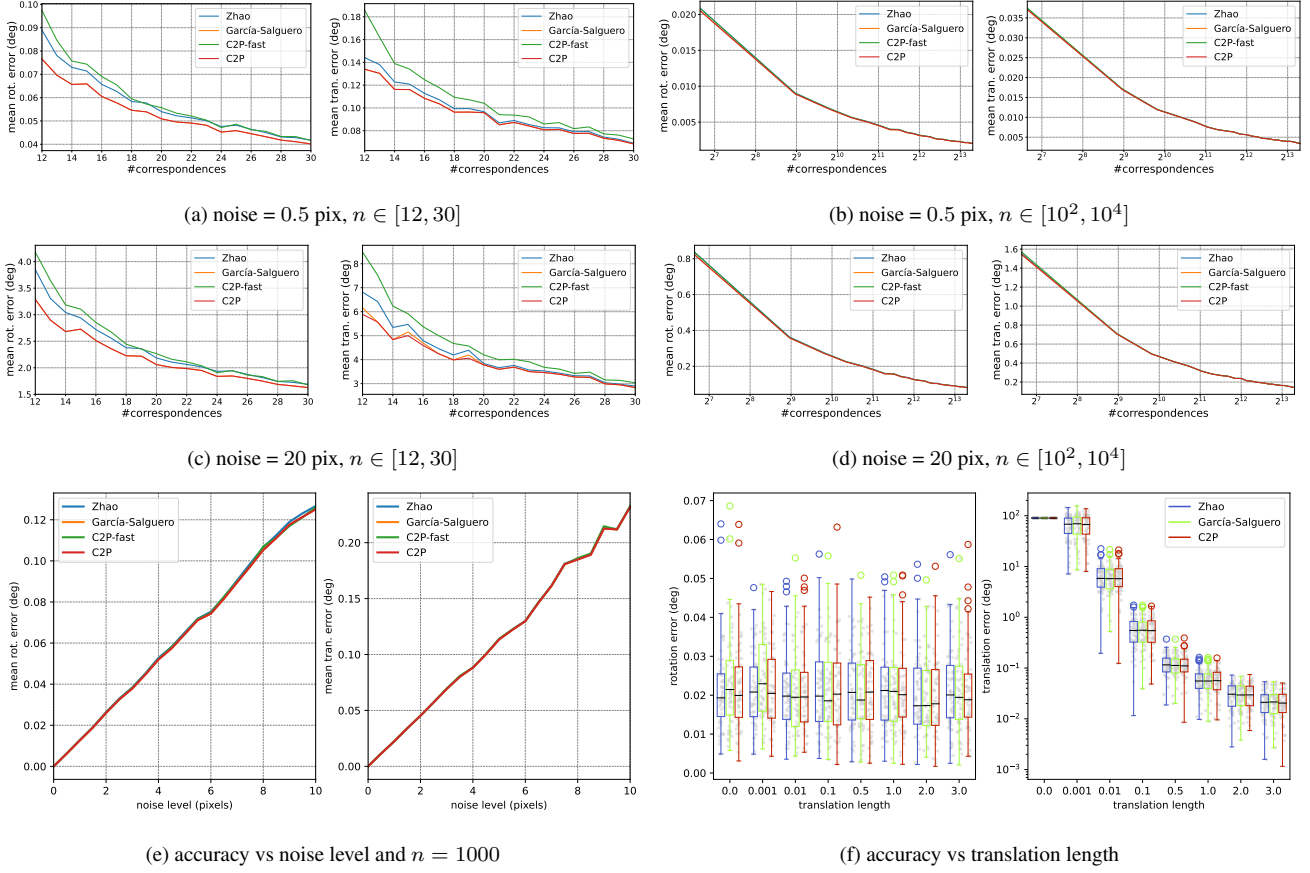


Figure 9. **Additional synthetic experiments.** We evaluate our proposed C2P and C2P-fast under various conditions: (a)-(d) number of correspondences, (a)-(e) noise levels, and (f) relative translation scale w.r.t. scene. As shown in (b) and (d), C2P-fast is well-suited for scenarios where $n > 10^3$, performing on-par with C2P and [23, 69], while being faster (Fig. 5). With fewer correspondences, as shown in (a) and (c), C2P outperforms [69], slightly surpassing the accuracy of [23] in estimating the translation, while also being faster. The same conclusions are reached when varying the noise levels (e). Finally, in (f) we show that C2P performs as well as [23, 69] when varying the scale of the translation relative to the scene, and unlike them, C2P is capable of directly detecting near-pure rotational motions.

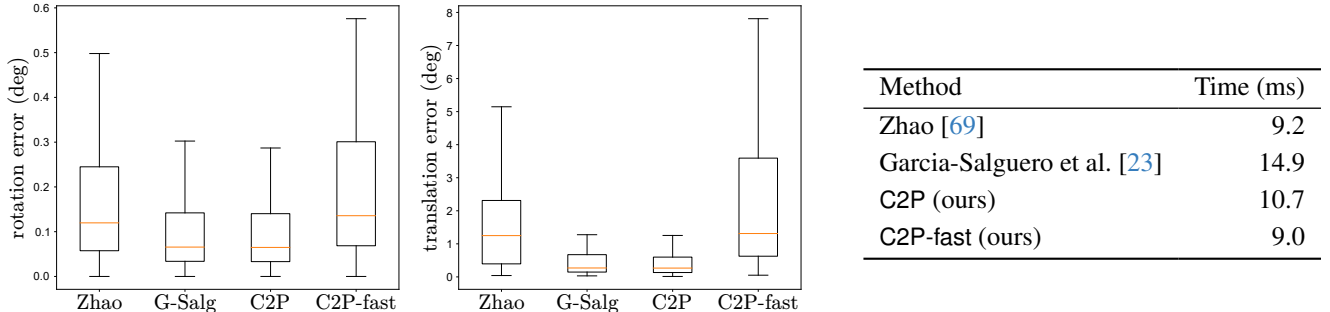


Figure 10. **Performance across all sequences (97 pairs) from Strecha et al. [56].** (left) Relative rotation and translation errors (in degrees) for all image pairs. (right) Averaged execution times for computing the relative pose for each method. As can be seen, C2P-fast is the fastest among all methods. However, C2P-fast is not always tight, resulting in a slight loss of accuracy when compared to the alternatives. On the other hand, our C2P is significantly more accurate than Zhao [69] and is on-par with Garcia-Salguero et al. [23] (labeled on the left as G-Salg.). Additionally, our C2P is, on average, 40% faster than [23].