Certifiably Optimal Robust Angular Synchronization

Daniel Barath[1,2], **Keisuke Tateno**[2], **Marc Pollefeys**[1,3], **Federico Tombari**[2]
[1]ETH Zurich, [2]Google, [3]Microsoft
dbarath@ethz.ch

Abstract

Rotation averaging on the circle, equivalently robust angular synchronization, underlies a wide range of geometric estimation problems, such as gravity-aligned Structure-from-Motion, multi-way point-cloud registration, and in-plane alignment in single-particle cryo-electron microscopy. We present the first algorithm that *certifies* the bin-level global optimum of a smoothed maximum-consensus objective. Under a quantitative separation condition, the certified solution recovers the continuous max-consensus inlier set with high probability over the outlier draws. Our approach discretizes the circle into uniform bins, formulates a pairwise Markov Random Field and solves it via branch-and-bound with two key novelties: a label-pruning rule guided by Iterated Conditional Modes (ICM) that collapses the branching factor to a handful of candidates, and a node-constrained relaxation bound orders of magnitude tighter than the standard bound. Together, they reduce the search from millions of nodes to a few hundred on typical instances. On all tested problems, the algorithm terminates with a *proof* of global optimality of the MRF energy and returns the certified-optimal inlier set, typically in a few seconds. A binary graph-cut refinement then maps the discrete bin labeling to continuous sub-bin angles as a warm start for IRLS. Across synthetic graphs and four real-world domains – image-based SfM on IMC 2023/2024 and KITTI, multi-way LiDAR point-cloud registration on KITTI, cryo-EM angular synchronization on EMPIAR-10166, and planar pose-graph optimization with up to 10^4 poses – our method stays accurate at outlier rates where other baselines fail, reaching sub-degree accuracy on fully connected and sparse synthetic graphs. Code: `https://github.com/danini/optimal-angular-synchronization`.

1 Introduction

Globally optimal solutions to robust geometric estimation problems are extremely rare. Most formulations of interest, including maximum consensus [16] and its variants, are NP-hard [11] and even APX-hard: no polynomial-time approximation scheme exists for them unless P=NP. Practical approaches therefore rely on continuous surrogates – spectral relaxations [41], graduated non-convexity [46], semidefinite programming [38, 47], iteratively reweighted least squares [10] – although their tightness degrades precisely when robustness matters most: at high outlier rates. An algorithm that returns an estimate together with a *proof* that no better solution exists, even under extreme outlier contamination, represents a qualitative advance.

We target rotation averaging on the circle S^1, equivalently robust *angular synchronization* [41]: recover absolute angles $\theta_i \in [0, 2\pi)$ from noisy pairwise differences $\tilde{\theta}_{ij} \approx (\theta_j - \theta_i) \bmod 2\pi$, where a potentially large fraction η of the measurements are gross outliers. This 1D formulation is not a niche sub-problem: it is the same mathematical object that arises in multiple otherwise-separate fields, for example: (a) In image-based structure from motion (SfM), gravity readings from smartphones, drones, head-worn displays, and ground robots reduce the three-DoF camera rotation to a single yaw angle around the gravity axis [33, 32]. (b) In multi-way point-cloud registration [24],

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

the ground plane acts as the known direction and pairwise rigid alignments collapse to 1D rotation averaging on the residual yaw. (c) In single-particle cryo-electron microscopy, the in-plane rotation between projections that share a viewing direction is intrinsically a 1D synchronization problem, the original motivation of Singer's work on angular synchronization [41]. (d) In planar SLAM, 2D pose-graph optimization first synchronizes robot headings [9], and false loop closures act as gross outliers [42]. Rotation averaging on S^1 is therefore a unifying primitive, and an algorithm that is globally optimal and robust to high outlier rates pays off across different domains.

We present the first algorithm that *certifiably solves* the maximum-consensus MRF for 1D rotation averaging to global optimality at bin-level resolution. The method discretizes the circle into K uniform bins, formulates a Gaussian-weighted pairwise Markov random field that smooths the consensus indicator (Eq. (2)), and solves this MRF exactly via branch-and-bound with two novelties that make the search tractable. We prove that the discrete MRF optimum recovers the continuous max-consensus inlier set with high probability under a quantitative separation condition (Theorem 2). On every tested instance with up to 80% outlier contamination, the algorithm terminates with a proof of bin-level optimality of the MRF energy typically in seconds, achieving sub-degree accuracy at outlier rates where every baseline diverges.

Our contributions are as follows:

1. A branch-and-bound algorithm for the max-consensus MRF on S^1 with two novelties – label pruning guided by Iterated Conditional Modes (ICM) [5] and a node-constrained upper bound – that together reduce the search space by a factor of $> 10^4$ on the tested instances.
2. Formal guarantees that (i) the discrete optimum recovers the continuous max-consensus inlier set with high probability under a separation condition and (ii) the BnB bound is admissible.
3. A binary graph-cut refinement that narrows the certified solution from bin-level to sub-degree precision.

We validate the method across four otherwise-separate rotation-averaging applications – image-based SfM on IMC 2023/2024 and KITTI, multi-way LiDAR point-cloud registration on KITTI, cryo-EM in-plane synchronization on EMPIAR-10166, and 2D pose-graph optimization on six SLAM datasets – plus a full synthetic sweep over graph topology, outlier ratio, relative rotation and gravity noise. Our solver achieves the lowest median pairwise rotation error on every SfM, LiDAR and cryo-EM benchmark (and the lowest mean error on all but IMC 2024), reaches $\approx 1^\circ$ median error on four of six SLAM pose graphs at 50% corrupted loop closures, where ℓ_1-IRLS errs by 19–53°, and, at 80% outlier contamination on fully connected and sparse synthetic graphs, sub-degree accuracy that other baselines fail to reach.

2 Related Work

Rotation averaging and global SfM. Rotation averaging, the task of recovering absolute orientations from noisy pairwise relative rotations, is a core component of global structure-from-motion pipelines [19, 18, 39, 32]. Govindu [18] introduced the Lie-algebraic averaging framework, and Chatterjee and Govindu [10] proposed ℓ_1-IRLS, which remains the de facto standard in practice. Yang et al. [46] introduced graduated non-convexity (GNC) for robust spatial perception. Yang and Carlone [47] extended it to certifiable estimation via semidefinite relaxations, though the relaxation gap grows with the outlier rate. Eriksson et al. [15] proved strong duality for the SO(3) SDP relaxation under bounded noise. Dellaert et al. [14] achieved global optimality for the Riemannian (non-robust) cost by lifting to SO(p), and Rosen et al. [38] provided certifiable optimality for SE-Sync under moderate noise. Parra et al. [34] obtained $100\times$ speedups via coordinate descent. Lee and Civera [29] proposed hierarchical outlier rejection, Purkait et al. [36] learned averaging with NeuRoRA, and Zhang et al. [50] incorporated measurement uncertainty into the optimization. All of these methods operate on the full SO(3) group. We specialize to the circle S^1, exploiting the lower dimensionality for exact combinatorial optimization that would be intractable in higher dimensions.

Gravity-aligned rotation estimation. Modern consumer devices, including smartphones, head-worn displays, drones, and ground robots, carry inertial measurement units (IMUs) that provide reliable gravity estimates. When the gravity direction is known, the three-degree-of-freedom camera rotation reduces to a single yaw angle around the gravity axis, transforming SO(3) rotation averaging into angular synchronization on S^1. Exploiting a known vertical direction to simplify relative-rotation estimation goes back to Antone and Teller [1]. Pan et al. [33] exploited this reduction in a global SfM

pipeline, formulating gravity-aligned rotation averaging as robust circular regression and integrating it into GLOMAP [32], where it improved accuracy by 13 AUC points while running $8\times$ faster than the SO(3) baseline. In robotics, gravity priors from IMUs are routinely used in visual-inertial odometry [37] and SLAM systems to constrain the pose graph. Our work targets the same 1D rotation averaging problem as Pan et al. [33] but replaces the local circular regression, which is susceptible to local minima at high outlier rates, with a certifiably globally optimal discrete solver.

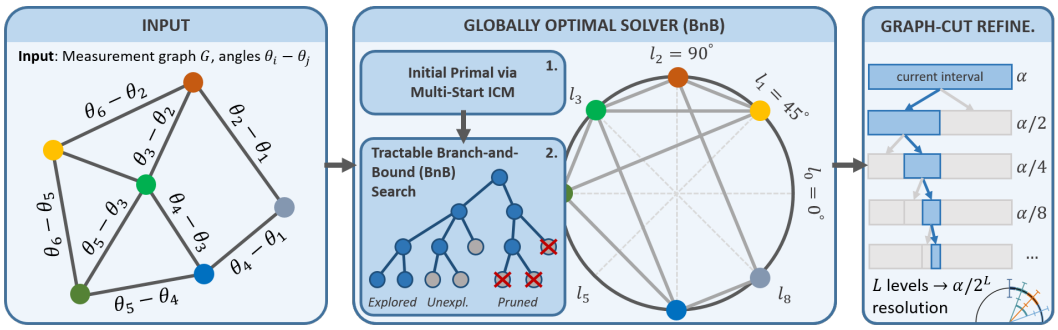
Angular synchronization across domains. Singer [41] formalized angular synchronization via the leading eigenvector of a Hermitian connection matrix and an SDP relaxation. Bandeira et al. [2] proved the SDP tight under Gaussian noise. Boumal [6] and Zhong and Boumal [51] sharpened the landscape and recovery analyses. Perry et al. [35] established computational–statistical gaps. Wang and Singer [45] gave ℓ_1-SDP exact-recovery guarantees under adversarial corruption, and Shi and Lerman [40] developed message-passing least squares for rotation synchronization. He et al. [21] introduced GNNSync, a neural approach for heterogeneous graphs. These results are strong under moderate noise but do not directly attack max-consensus under gross outliers. Beyond image-based SfM, the same 1D primitive drives three further applications: multi-way LiDAR/RGB-D registration, where a known ground-plane normal reduces pairwise SE(3) alignment to a yaw synchronization on a sparse scan graph [12, 23]; planar SLAM, where the headings of a pose graph are synchronized before its positions [9]; and single-particle cryo-EM, where projections sharing a viewing direction differ only in an in-plane rotation ψ and the ASPIRE pipeline [13] estimates these pairwise rotations via cross-correlation before synchronization. These settings routinely exhibit $> 40\%$ outlier rates where the continuous baselines above degrade sharply.

Robust estimation and maximum consensus. The maximum consensus framework, introduced by Fischler and Bolles [16] as the Random Sample Consensus (RANSAC) method, seeks the largest set of measurements consistent with a single model. Chin et al. [11] proved that maximum consensus is NP-hard and APX-hard for general geometric estimation problems, establishing fundamental computational barriers. For rotation synchronization specifically, cycle-consistency checking [49] and the cycle-edge message passing (CEMP) framework of Lerman and Shi [30] provide principled outlier detection with exact-recovery guarantees under sufficient connectivity. Yang et al. [48] introduced TEASER, which achieves certifiable robustness for point cloud registration via truncated least squares and SDP relaxation. Yang and Carlone [47] extend certifiable outlier rejection to general geometric estimation problems but rely on SDP relaxations whose tightness degrades at high outlier rates. Our work directly solves the maximum consensus problem on S^1 to certified global optimality via combinatorial optimization, providing not just a robust estimate but a proof that no assignment achieves higher consensus.

Discrete optimization and graph cuts. Szeliski et al. [43] provided an extensive comparison of MRF energy minimization methods. TRW-S [27, 44] provides LP-relaxation bounds but cannot certify integer optimality when a relaxation gap exists; our branch-and-bound closes this gap exactly. For the binary refinement stage, we exploit the classical result that binary MRFs with submodular potentials (supermodular in our maximization form) are exactly solvable via graph cuts [28, 7]. Boykov et al. [8] extended graph cuts to multi-label problems via α-expansion, and Besag [5] introduced ICM as a fast coordinate-ascent heuristic. Our pipeline combines all three: ICM for initialization, branch-and-bound for exact certification, and graph cuts for sub-degree refinement.

3 Method

Three properties make combinatorial certification tractable specifically for S^1: (i) the search space is one-dimensional, so a moderate grid ($K{=}64$, bin width 5.6°) places every continuous angle within $\alpha/2{\approx}2.8^\circ$ of a bin center; (ii) the pairwise MRF potential is translation-invariant in the bin difference, which lets a small ICM-pruned feasible set retain the kernel peak for almost every node pair; and (iii) at fixed grid resolution, BnB tree size is governed by the local conflict structure of inlier vs. outlier edges, not by the ambient K^N search space. The same recipe applied to the full SO(3) problem would require an $\Omega(K^3)$ grid and a node-constrained bound over a 3D label set, raising the per-node cost prohibitively. The algorithm has the following stages (Fig. 1): discretize the circle and formulate a pairwise MRF over bin labels (Sec. 3.1–3.2), solve it to certified global optimality via branch-and-bound (Sec. 3.3), and refine to sub-degree precision with binary graph cuts (Sec. 3.4).



Pairwise Measurement Graph $G = (V, E)$ | Globally Optimal Solver via Binner MRF on the Circle | Sub-degree refinement

Figure 1: **Overview** of our certified 1D rotation-averaging pipeline. *Input* (left): pairwise graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with noisy yaw measurements $\tilde{\theta}_{ij}$, a large portion of which may be outliers. *Globally optimal solver* (center): the circle is discretized into K uniform bins ($\alpha = 2\pi/K$) and the consensus objective is recast as a pairwise MRF over bin labels ℓ_i. Multi-start ICM seeds a strong primal $\underline{E}$; the ICM-guided feasible sets F_i collapse the branching factor from K to ∼3–5; the node-constrained bound (Eq. (6)) closes the search to a few hundred nodes, returning the certified-optimal labeling $\boldsymbol{\ell}^*$. *Graph-cut refinement* (right): each certified bin is recursively split and a binary graph cut picks the better half, reaching resolution $\alpha/2^L$.

3.1 Problem Formulation

We are given a measurement graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $N = |\mathcal{V}|$ nodes and $M = |\mathcal{E}|$ edges. Each node i carries an unknown angle $\theta_i \in [0, 2\pi)$ and each edge $(i, j) \in \mathcal{E}$ carries a measurement $\tilde{\theta}_{ij} \in [0, 2\pi)$. An edge is an *inlier* if $\tilde{\theta}_{ij} \approx (\theta_j - \theta_i) \bmod 2\pi$ up to bounded noise, and an *outlier* if $\tilde{\theta}_{ij}$ is drawn uniformly at random from $[0, 2\pi)$ and bears no relation to the true angles. We denote the outlier fraction by η. Let $d_\circ(a, b) = \min(|a - b|,\ 2\pi - |a - b|)$ denote the geodesic (wrap-around) distance on the circle. Given an inlier threshold $\tau > 0$, the *maximum consensus* objective seeks angles that maximize the number of geometrically consistent edges:

$$\max_{\theta_0, \ldots, \theta_{N-1} \in [0, 2\pi)} \sum_{(i,j) \in \mathcal{E}} \mathbf{1}\Big[d_\circ\big(\tilde{\theta}_{ij},\ (\theta_j - \theta_i) \bmod 2\pi\big) < \tau \Big]. \tag{1}$$

This problem is NP-hard in general [11]. All existing solvers for angular synchronization, including spectral methods [41], IRLS [10], and GNC [46], minimize continuous surrogate losses that are susceptible to local minima. Our approach directly attacks the combinatorial structure of (1).

Gauge freedom. The objective (1) is invariant to a global additive shift $\theta_i \mapsto \theta_i + \phi$. We fix this gauge by setting $\theta_0 = 0$ throughout.

3.2 Discrete MRF Formulation

We partition $[0, 2\pi)$ into K uniform bins of width $\alpha = 2\pi/K$; bin k has center $c_k = (k + \frac{1}{2})\alpha$. Each node i is assigned a discrete label $\ell_i \in \{0, \ldots, K-1\}$ indicating the bin containing θ_i. For each edge (i, j), the pairwise potential measures how well the labels (ℓ_i, ℓ_j) explain the measurement $\tilde{\theta}_{ij}$. A 0–1 (boxcar) indicator at threshold τ treats all compatible pairs identically, providing no localization signal. We instead use a truncated Gaussian centered at zero residual, which gives the BnB upper bounds (Sec. 3.3) something to discriminate with:

$$w_{ij}(\ell_i, \ell_j) = \begin{cases} \exp\left(-\dfrac{d_\circ\big(c_{\ell_j} - c_{\ell_i},\ \tilde{\theta}_{ij}\big)^2}{2\sigma^2} \right) & \text{if } d_\circ\big(c_{\ell_j} - c_{\ell_i},\ \tilde{\theta}_{ij}\big) < r, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where the kernel bandwidth defaults to $\sigma = \tau$ and the truncation radius to $r = \max(\alpha, 3\sigma)$. The MRF energy to be *maximized* is $E(\boldsymbol{\ell}) = \sum_{(i,j) \in \mathcal{E}} w_{ij}(\ell_i, \ell_j)$, with the gauge fixed by $\ell_0 = 0$. Under the correct labeling each inlier edge contributes ≈1 and each outlier contributes ≈0 (a uniform random outlier hits the kernel support with probability only $2r/(2\pi)$). Since the support $r \geq 3\sigma$ exceeds τ, E is a smoothed consensus count rather than the 0–1 count. Proposition 1 and Theorem 2 relate the two.

We first check that this discretization does not silently throw inliers away. Otherwise, optimizing E would not be meaningful for Eq. (1).

Proposition 1 (Discretization preserves inliers)**.** *Assume $\tau + \alpha \leq r$ (true for $\sigma = \tau$ and $\alpha \leq 2\tau$). Let $\boldsymbol{\theta}^*$ be any continuous solution to* (1) *and let $\hat{\ell}_i = \lfloor \theta_i^*/\alpha \rfloor$ be the nearest-bin labeling. Then every inlier edge under $\boldsymbol{\theta}^*$ (*i.e.*, every edge with $d_\circ(\tilde{\theta}_{ij},\, (\theta_j^* - \theta_i^*) \bmod 2\pi) < \tau$) satisfies $w_{ij}(\hat{\ell}_i, \hat{\ell}_j) > 0$ under the discrete potential* (2)*. In particular, let C^* denote the optimal consensus count and E_K^* the maximum of the MRF energy. Then $E_K^* \;\geq\; E(\hat{\boldsymbol{\ell}}) \;\geq\; C^* \cdot e^{-(\tau+\alpha)^2/(2\sigma^2)}$.*

Proof in Appendix A.1. In words, rounding every continuous max-consensus angle to its nearest bin produces a feasible discrete labeling under which every continuous-inlier edge still contributes positively, so the MRF optimum is at least as large as the continuous consensus count up to a constant Gaussian-decay factor. The following theorem shows that under a quantitative separation condition, the MRF optimum recovers the same inlier set as the continuous max-consensus solution.

Theorem 2 (Discrete-continuous equivalence, w.h.p.)**.** *Let $\boldsymbol{\theta}^*$ be a continuous max-consensus optimizer at threshold τ achieving C^* inliers, and let C_2^* denote the largest consensus count, at the same threshold τ, achievable by any* structurally distinct *solution – one whose inlier set differs from $\boldsymbol{\theta}^*$'s. Let $\delta_{\max} = \max_{\text{inlier } (i,j)} d_\circ(\tilde{\theta}_{ij},\, (\theta_j^* - \theta_i^*) \bmod 2\pi)$ be the largest inlier residual and $w_{\min} = e^{-(\delta_{\max}+\alpha)^2/(2\sigma^2)}$ the minimum MRF weight achievable by an inlier edge after discretization. Suppose all edges have classification margin $\gamma_{ij} = |d_\circ((\theta_j^* - \theta_i^*) \bmod 2\pi,\, \tilde{\theta}_{ij}) - \tau| > 0$. Let the M_{out} outlier edges carry i.i.d. uniform measurements on $[0, 2\pi)$, independent of the other $M_{\text{in}} = M - M_{\text{out}}$ edges, and write $c_\tau = e^{-\tau^2/(2\sigma^2)}$ and $B_\delta = M_{\text{out}} \frac{r}{\pi} + \sqrt{\frac{1}{2} M_{\text{out}} \big((N-1)\ln K + \ln(1/\delta)\big)}$ for $\delta \in (0,1)$. If $\alpha < \min_{(i,j)} \gamma_{ij}$ and $\delta_{\max} + \alpha < r$, then with probability at least $1 - \delta$ over the outlier draws,*

$$C^* \cdot w_{\min} \;>\; (1 - c_\tau)\, C_2^* \;+\; c_\tau \left(M_{\text{in}} + B_\delta\right) \tag{3}$$

implies that the MRF-optimal labeling $\boldsymbol{\ell}^*$ realizes the same inlier set as $\boldsymbol{\theta}^*$.

Proof in Appendix A.2. The term B_δ bounds the outlier hits uniformly over all K^{N-1} labelings. Note that the theorem characterizes the asymptotic fine-grid regime, whereas in practice $K = 64$ already succeeds since the truncated Gaussian kernel smoothly interpolates across bin boundaries. Also note that the discrete certificate itself is deterministic and per-instance. Only the bridge from the discrete optimum to the continuous max-consensus inlier set carries the w.h.p. qualifier.

3.3 Certified Global Optimization via Branch-and-Bound

Maximizing E over $\{0, \ldots, K-1\}^N$ is NP-hard for general graphs. We solve it exactly via branch-and-bound (BnB), a systematic enumeration that explores the space of partial assignments, pruning regions provably inferior to the current best solution. Unlike LP or SDP relaxations [15, 38, 47], which certify a solution only when the relaxation is tight, BnB certifies optimality over the *original discrete* problem: the returned labeling is provably the best among all K^N candidates. Two contributions make this tractable for our problem.

Strong Primal via Multi-Start ICM. The quality of the BnB depends critically on having a strong *primal* (lower bound on the optimal energy) from the outset, since a higher primal allows more aggressive pruning. We obtain this via Iterated Conditional Modes (ICM) [5]: starting from an initial labeling, each node is updated in turn to the label maximizing its local energy contribution, given the current labels of all neighbors:

$$\ell_i \;\leftarrow\; \arg \max_{k \in \{0,\ldots,K-1\}} \sum_{j:\ (i,j) \in \mathcal{E}} w_{ij}(k,\, \ell_j). \tag{4}$$

One sweep over all nodes costs $O(MK)$. Convergence to a coordinate-wise maximum (no single-node change can increase E) is guaranteed in finitely many steps, since the energy strictly increases at each update and is bounded. Empirically, a few sweeps suffice. To escape local maxima, we run R independent restarts with randomized initial labels and random node orderings, retaining the solution with the highest energy $\underline{E}$. The first restart uses a greedy breadth-first search (BFS) initialization from the gauge-fixed node: starting from $\ell_0 = 0$, each node's label is set to maximize the weight of its incoming BFS edge. This deterministic initialization already gets close to the global optimum in the majority of cases. The additional random restarts handle pathological configurations.

ICM-Guided Label Pruning. A naive BnB with K labels per node has branching factor K, making exhaustive search infeasible for even moderate K. Our key observation is that the ICM solution provides strong evidence about which labels are *plausible* for each node.

For each node i and candidate label k, we compute the *conditional reward*: the energy node i would receive if assigned label k while all its neighbors retain their ICM labels:

$$u_i(k) = \sum_{j\colon (i,j)\in\mathcal{E}} w_{ij}(k,\ \ell_j^{\mathrm{ICM}}). \tag{5}$$

When the ICM solution is correct, the correct label k^* aligns with all inlier neighbors, and its reward $u_i(k^*)$ is large because each inlier edge contributes a weight close to 1. A wrong label $k \neq k^*$ shifts the predicted difference by at least one bin width α on every inlier edge, reducing the Gaussian weight. Only outlier edges can accidentally match, each with probability $\approx 2r/(2\pi)$. We retain only labels satisfying $u_i(k) \geq \rho \cdot \max_{k'} u_i(k')$ for a pruning threshold ρ. The rest are *pruned* from the BnB search. We use $\rho = 0.3$, which offers a good tradeoff: smaller values (*e.g.*, $\rho = 0.1$) retain too many labels, slowing BnB, while larger values (*e.g.*, $\rho = 0.5$) occasionally prune the true label at very high outlier rates.

Pruning safety is conditional on the ICM warm-start being correct on the local neighborhood – exactly the regime that high-outlier rates threaten. We therefore quantify it w.r.t. a *measured* ICM agreement rate rather than a binary correctness flag. For node i with degree d_i, of which $(1-\eta)\,d_i$ edges are inliers, let $\rho_i^{\mathrm{ICM}} \in [0,1]$ be the fraction of inlier neighbors j whose multi-start ICM label ℓ_j^{ICM} matches the true bin (or any bin within ± 1 of it), and $\delta_{\mathrm{in}} \in (0,1]$ their mean weight $w_{ij}(k^*, \ell_j^{\mathrm{ICM}})$ at the correct label k^*. We obtain:

Proposition 3 (Pruning safety, conditional on ICM agreement)**.** *Treat the ICM labels as fixed and let $r = \max(\alpha, 3\sigma)$ be the kernel radius. For any $\delta \in (0,1)$, if*

$$(1-\eta)\,d_i\,\rho_i^{\mathrm{ICM}}\,\delta_{\mathrm{in}} \ >\ \rho\Big((1-\eta)\,d_i + \eta\,d_i\,\tfrac{r}{\pi} + \sqrt{\tfrac{1}{2}\eta\,d_i\,\ln(K/\delta)}\Big),$$

then with probability at least $1-\delta$ over the outlier draws, the correct label survives pruning: $k^* \in F_i$.

Proof in Appendix A.3. The bound charges every inlier edge the full weight 1 at any wrong label, so it holds for wrong labels at any distance from k^*. The proposition gives a sufficient condition on the ICM agreement rate ρ_i^{ICM}: at $\rho = 0.3$, $\eta = 0.5$, $K = 64$, $\delta = 0.05$ and $d_i = 20$–50 it requires $\rho_i^{\mathrm{ICM}}\delta_{\mathrm{in}} \gtrsim 0.44$–$0.50$; with the ~ 0.85 agreement we observe empirically (over all neighbors), it holds whenever $\delta_{\mathrm{in}} \gtrsim 0.6$. Pruning and the safeguards below together reduce the branching factor from K to $|F_i| \approx 3$–5, a $10\times$+ reduction.

To bridge the gap when ICM agreement is locally low, we add 2 safeguards: (i) ℓ_i^{ICM} is always retained in F_i; (ii) if $|F_i| < 3$ we expand F_i with the labels within ± 2 bins of ℓ_i^{ICM}.

Node-constrained upper bound. The per-edge bound $\bar{E}_{\mathrm{edge}} = \sum_{(i,j)} \max_{a,b} w_{ij}(a,b)$ is loose since each edge picks its best label pair independently, ignoring that each node carries a *single* label across all incident edges. Our *node-constrained* bound enforces that consistency. Let $\mathcal{F}$ be the free nodes with feasible sets F_i (frozen at the BnB root from the ICM solution). For each free node, we charge the full weight of every mixed (free–fixed) incident edge and *half* the optimistic weight of every free–free edge. The symmetric halving avoids double-counting:

$$\bar{E}_{\mathrm{node}} = \sum_{\substack{(i,j)\in\mathcal{E}\\ \text{both fixed}}} w_{ij}(\ell_i,\ell_j) \ +\ \sum_{i\in\mathcal{F}} \max_{k\in F_i}\Bigg[\sum_{\substack{j\in N(i)\\ j\notin\mathcal{F}}} w_{ij}(k,\ell_j) \ +\ \tfrac{1}{2}\sum_{\substack{j\in N(i)\\ j\in\mathcal{F}}} \max_{l\in F_j} w_{ij}(k,l)\Bigg]. \tag{6}$$

Theorem 4 (Admissible upper bound)**.** *For any labeling $\boldsymbol{\ell}$ extending the partial assignment with $\ell_i \in F_i$ for all $i \in \mathcal{F}$, $E(\boldsymbol{\ell}) \leq \bar{E}_{\mathrm{node}}$.*

Proof in Appendix A.4. The bound is admissible in the branch-and-bound sense: it never underestimates the best completion of a partial assignment, so pruning can never discard the optimum. The symmetric halving enforces that each free–free edge is claimed at most once across endpoints, forcing a single consistent label per node; empirically this drops the optimality gap from $\sim 45\%$

(per-edge) to $\sim$4–10%, and the bound remains admissible for arbitrary potentials and feasible sets (no translation-invariance assumption).

Search. At each BnB node, we compute the node-constrained bound, prune if dominated by the current primal, branch on the most-constrained free node and explore F_i best-first. Top-level subtrees run in parallel via OpenMP, and a tighter primal found by any thread benefits all others. Pseudocode: Appendix B. With $|F_i| \approx 4$, BnB typically explores $<$200 nodes for N=20–50 at up to 50% outliers (0.3 s at N=20, Table 4); at 80% on dense graphs the count rises to $\sim$30,000 nodes ($\sim$10 s).

3.4 Binary Graph-Cut Refinement

BnB returns the certified-optimal inlier set and per-node bins of width α. We then sharpen to sub-degree precision through L levels of binary graph-cut subdivision (right panel of Fig. 1), each splitting every node's current interval at its midpoint and deciding left vs. right by a binary MRF whose pairwise potential averages (2) over n_q^2 quadrature samples inside the two sub-intervals (with $\sigma_{\text{ref}} = \tau$). Binary MRFs with supermodular pairwise potentials (*i.e.*, $w(0,0)+w(1,1) \geq w(0,1)+w(1,0)$) admit an exact polynomial-time solution by reduction to s-t min-cut [28]. On average, 94% of edges are supermodular at every level (measured on N=50 dense graphs at $\eta \in \{0, 0.3, 0.6, 0.8\}$); for the remaining $\sim$6%, with violation $\Delta_{ij} = w(0,1) + w(1,0) - w(0,0) - w(1,1) > 0$, we add capacity $\Delta_{ij}/2$ to both arcs $i \leftrightarrow j$, a heuristic submodular approximation. The selected inlier set is unaffected – it is the certified-optimal set from BnB – so the consensus-maximization guarantee carries through. The split only relaxes optimality of the sub-bin warm start handed to IRLS. After L=6 levels (for τ=5°), $\alpha/2^L \approx 0.09°$, well below measurement noise.

4 Experiments

We evaluate on synthetic graphs and on four problems: image-based SfM, multi-way LiDAR point-cloud registration, single-particle cryo-EM, and 2D pose-graph optimization. We compare against six baselines spanning the main algorithmic families: Singer's spectral approach (**Spectral**) [41], cycle-edge message passing (**CEMP**) [30], graduated non-convexity (**GNC**) [46], the ℓ_1-IRLS pipeline of Chatterjee and Govindu (**ℓ_1-IRLS**) [10], the gravity-aligned circular regression of Pan et al. (**CircReg**) [33] used in GLOMAP, and a QUASAR-style truncated-least-squares SDP relaxation [47] adapted to angular synchronization on S^1 (**TLS-SDP**; iterative reweighted SDP with TLS thresholding). Unless stated otherwise, τ=5°, K=64; runtimes are in seconds. On the SfM and LiDAR benchmarks of Sec. 4.2, all methods except ℓ_1-IRLS (GLOMAP's 3-DoF default) solve the gravity-aligned 1D problem, and CircReg and ours share GLOMAP's IRLS polish, included in their runtimes. Elsewhere, baselines are reported as returned by their solvers.

4.1 Synthetic Graphs

We evaluate on three graph topologies at N=50 nodes: *fully connected*, *sparse* ($\sim$32 edges/node, an SfM-like covisibility graph), and *sequential* (10-frame sliding window, a streaming SLAM-like graph). For each topology, we sweep the outlier ratio $\eta \in \{0, 0.1, \ldots, 0.8\}$ at σ_{noise}=2°, the measurement-noise standard deviation $\sigma_{\text{noise}} \in [0.1°, 10°]$ at η=50%, and the gravity-noise standard deviation $\sigma_{\text{grav}} \in [0°, 10°]$ at η=50%, σ_{noise}=2°; gravity noise models the pitch/roll error from an imperfect IMU propagating into the 1D problem.

Fig. 2 reports per-run median error averaged over 50 seeds. On the *outlier-ratio sweep*, our method stays single-degree well past the breakdown point of every baseline: at 80% outliers, 0.55° on fully connected (vs. 8.2° CircReg, 26.9° TLS-SDP, $\geq$ 22.8° for ℓ_1-IRLS/GNC/CEMP/Spectral), 0.87° on sparse (vs. 59.9° TLS-SDP and $\geq$ 38.6° for every other baseline), and at 70% on sequential, 1.64° vs. $\geq$ 27° for every baseline. The TLS-SDP curve illustrates the SDP-relaxation breakdown highlighted in the introduction: it tracks ℓ_1-IRLS, CircReg and ours at $\eta \leq 0.5$ but degrades sharply from $\eta = 0.6$, jumping from 0.31° at η=0.5 to 26.9° at η=0.8 on dense graphs (from 1.10° to 59.9° on sparse), exactly the regime our certified BnB stays single-degree. At 90% (beyond the range of Fig. 2) no tested method recovers the angles. Every certified solution carries a proof of global optimality. The *measurement-noise* sweep shows graceful degradation for ℓ_1-IRLS, CircReg and ours, which track within 0.3° at σ_{noise}=5° on the dense graph (1.09°/1.35°/1.07°), with CEMP/Spectral 4–5$\times$ worse and GNC $\sim$1.5$\times$ worse. The *gravity-noise* sweep mirrors this: at σ_{grav}=5° on the dense graph, Ours, CircReg and ℓ_1-IRLS sit at 3.31°/3.36°/3.31°, with CEMP and Spectral con-

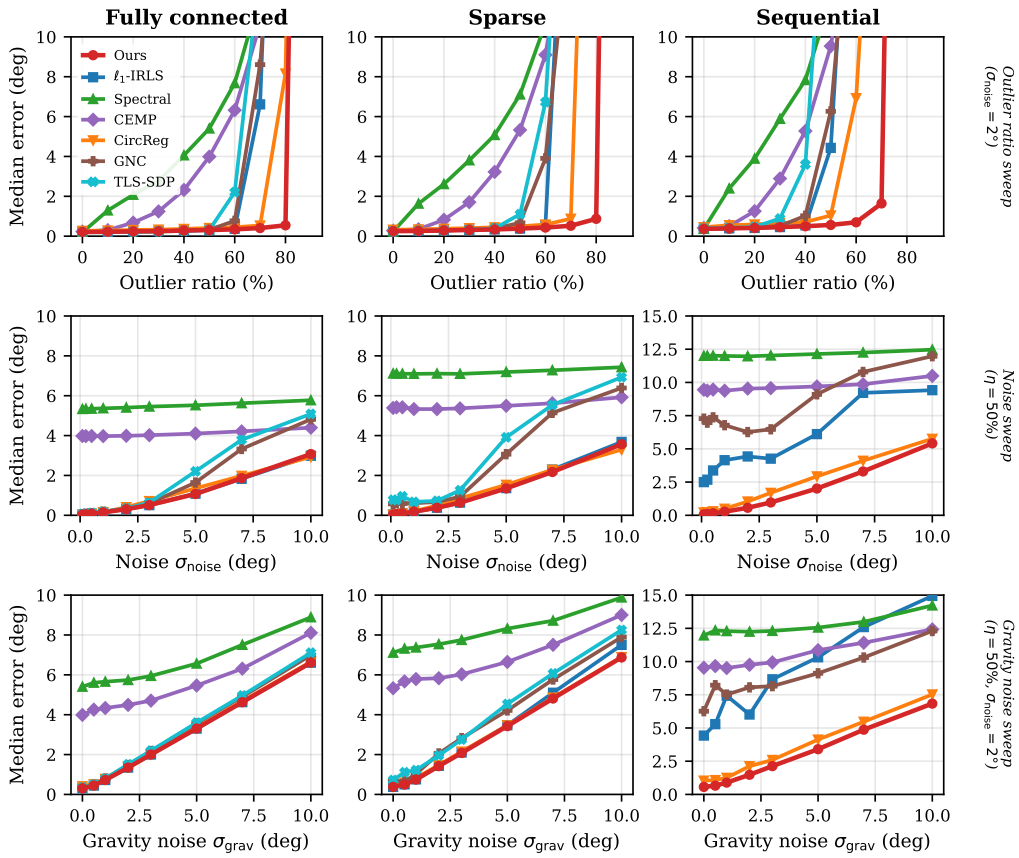


Figure 2: **Synthetic benchmarks** with N=50 nodes. *Columns:* graph topology (fully connected, sparse $\sim$32 edges/node, sequential window 10). *Rows:* outlier-ratio sweep at σ_{noise}=2° (top), measurement-noise sweep at η=50% (middle), and gravity-noise sweep at η=50%, σ_{noise}=2° (bottom). Mean of per-run median error over 50 seeds. Our method (red) has the lowest or near-lowest error in every configuration; the gap to the best baseline widens sharply at high outlier rates.

sistently 1.5–2× behind. The certified BnB therefore widens the robustness envelope by 10–20 outlier-percentage-points beyond the strongest competitor while matching it in the clean regime.

What makes BnB tractable. Two independent contributions drive the empirical speed-up. ICM-guided label pruning reduces the feasible-label set from K=64 to $|F_i| \approx$ 3–5 – a factor-10+ branching-factor cut – by removing labels whose conditional reward (Eq. (5)) is more than $\rho^{-1} = 3.3\times$ below the ICM optimum. The node-constrained upper bound then tightens the BnB relaxation gap from $\sim$45% (per-edge) to 4–10% on the same instances. Together, on N=50 fully connected graphs BnB explores fewer than 35,000 nodes at every tested outlier rate and runs in $\leq$12 s (Table 4); in the majority of runs ($\eta \leq 50\%$) the tree stays below 200 nodes, and N=20 graphs take 0.3 s at η=0.5. In runs we do not tabulate, removing either component pushes the search past $5 \cdot 10^6$ nodes, making certification effectively intractable. Appendix C.5 (Table 4) reports a runtime sweep over $N \in \{20, 50, 100, 200\}$. Our IRLS polish does not rescue the baselines at η=0.8: their inlier recall is 0.07–0.49 vs. 0.89–0.97 for ours (App. E).

4.2 Real-World Rotation Averaging

Datasets and pipelines. We evaluate on three applications sharing the same downstream rotation-averaging problem. For *image-based SfM* (haiper and phototourism from IMC 2023; IMC 2024; KITTI [17]), the measurement graph is produced by GLOMAP [32]. For *multi-way LiDAR registration* (KITTI odometry at strides s=5, 10 [17]), we use Predator [23] correspondences and RANSAC for pairwise $SE(3)$ estimation, taking the yaw around the ground-plane normal. For *cryo-EM*, we follow standard cryo-EM angular-synchronization practice [41, 13]: project N=100 views of the

Table 1: **Rotation averaging** on real-world data, averaged per dataset. We report the Area Under the recall Curve (AUC) at $3°/5°/10°$ ↑, mean and median pairwise rotation error (°) ↓, and runtime (s). Below each dataset name we list the average per-scene graph size: $\bar{N}$ is the mean number of nodes (images / frames / scans / particles) and $|\bar{\mathcal{E}}|$ the mean number of pairwise edges. Global SfM uses GLOMAP [32]. Multi-way PC registration uses Predator [23] with RANSAC. EMPIAR-10166 uses cryo-EM projections of EMD-3295 with per-particle CTF drawn from the empirical EMPIAR-10166 defocus distribution. On SfM and PC registration, CircReg and Ours share GLOMAP's IRLS polish, included in their runtimes; ℓ_1-IRLS is GLOMAP's 3-DoF averager. Extended table with more baselines in Appendix C.

	Dataset	Method	AUC@3° ↑	AUC@5° ↑	AUC@10° ↑	mean ↓	median ↓	time (s)
Global SfM	Haiper $\bar{N}$=18, $\lvert\bar{\mathcal{E}}\rvert$=110	ℓ_1-IRLS [10]	26.2	34.2	53.2	6.6	4.0	**1.3**
		CircReg [33]	44.5	55.2	70.3	2.4	1.6	**1.3**
		Ours	**46.4**	**57.3**	**73.6**	**1.7**	**1.5**	1.7
	Phototourism $\bar{N}$=364, $\lvert\bar{\mathcal{E}}\rvert$=66k	ℓ_1-IRLS [10]	34.2	44.5	57.7	12.3	2.7	88.2
		CircReg [33]	44.5	53.1	64.5	5.3	2.1	**1.2**
		Ours	**44.8**	**53.7**	**65.3**	**5.1**	**1.9**	82.2
	IMC 2024 $\bar{N}$=499, $\lvert\bar{\mathcal{E}}\rvert$=8.5k	ℓ_1-IRLS [10]	5.1	9.6	19.3	38.1	28.2	8.1
		CircReg [33]	14.0	21.6	34.3	**32.5**	19.7	**0.3**
		Ours	**15.8**	**23.5**	**35.9**	33.0	**19.6**	15.6
	KITTI $\bar{N}$=212, $\lvert\bar{\mathcal{E}}\rvert$=605	ℓ_1-IRLS [10]	40.2	44.9	49.9	44.0	28.9	**1.7**
		CircReg [33]	53.8	58.9	64.4	30.6	20.6	**1.7**
		Ours	**59.6**	**68.2**	**77.3**	**5.9**	**4.6**	2.3
Cryo-EM	EMPIAR-10166 $\bar{N}$=100, $\lvert\bar{\mathcal{E}}\rvert$=400	ℓ_1-IRLS [10]	50.2	61.8	71.5	15.0	1.3	**0.0**
		CircReg [33]	52.0	64.6	75.4	11.1	1.2	1.0
		Ours	**57.9**	**71.9**	**83.4**	**6.4**	**1.1**	10.6
PC Registration	KITTI (s=5) $\bar{N}$=400, $\lvert\bar{\mathcal{E}}\rvert$=1790	ℓ_1-IRLS [10]	28.5	43.4	61.7	9.5	8.0	**0.1**
		CircReg [33]	60.1	67.4	75.1	8.4	5.6	**0.1**
		Ours	**60.6**	**68.4**	**76.6**	**6.7**	**2.4**	2.3
	KITTI (s=10) $\bar{N}$=201, $\lvert\bar{\mathcal{E}}\rvert$=720	ℓ_1-IRLS [10]	15.6	22.4	30.4	45.7	31.8	0.1
		CircReg [33]	31.8	36.5	41.6	33.9	29.6	**0.0**
		Ours	**34.1**	**39.2**	**44.9**	**27.4**	**20.3**	10.4

EMD-3295 map (human p97 at 2.3 Å [3]) at a fixed viewing direction with random in-plane rotations, apply a per-particle CTF whose defocus is drawn from the EMPIAR-10166 particle stack [20] (2.5–84 kÅ), add white noise at SNR 1 – generous compared to the SNR$\ll$1 regime of raw EMPIAR particles; we use it to isolate the synchronization-vs-CTF interaction and leave lower-SNR regimes for future work – and extract $\tilde{\psi}_{ij}$ by a 360-candidate cross-correlation search; we inject 40% extra uniform outliers and average over 10 seeds. Gravity (or, for cryo-EM, the common direction) comes from GT so that the rotation-averaging step is isolated. On SfM and LiDAR, CircReg and ours share GLOMAP's IRLS polish and include it in their runtime. On cryo-EM, every method runs standalone on the 1D problem.

Results. Table 1 reports rotation AUC at $[3°, 5°, 10°]$, mean and median pairwise rotation error, and per-sequence runtime. On *SfM*, our method attains the lowest median error on all four benchmarks (and the lowest mean on all but IMC 2024) and leads CircReg by $+0.3$–5.8 AUC@3°, with the biggest win on KITTI's long sequential graphs and a clear gain on haiper's short sparse sequences, where a single bad edge drags continuous regression off-optimum. On *PC registration*, the median error on KITTI s=10 is 20.3° (Ours) vs. 29.6° (CircReg) and 31.8° (ℓ_1-IRLS). LiDAR outliers under limited overlap are heavy-tailed, so local solvers follow the geometric median and get pulled off, whereas the global consensus solver finds the correct inlier set. On *cryo-EM* at 40% outliers we lead CircReg on every AUC threshold ($+5.9/+7.3/+8.0$) and on mean error ($6.4°$ vs. $11.1°$). The mean gap exceeding the median indicates certified BnB avoids the heavy-tail per-node failures of continuous methods. The certified step adds 0.1-2 s on moderate graphs (Haiper, KITTI SfM/LiDAR) and 15-81 s on large, dense collections (IMC 2024, Phototourism) – still modest compared to upstream feature matching. The results of other baselines are in App. C (Table 3).

Table 2: **2D pose-graph optimization** at 50% corrupted loop closures: median heading error ($^\circ$) and runtime (s). Odometry ignores all loops. TLS-SDP is not run ($N{>}200$). Full results in App. D.

	INTEL	CSAIL	FR079	MITb	M3500	City10000
Error ($^\circ$)						
Ours	**0.8**	**0.5**	**1.1**	**33.8**	**4.2**	**1.0**
ℓ_1-IRLS [10]	41.3	53.0	19.4	47.8	81.1	53.4
Spectral [41]	85.5	80.7	88.6	85.6	82.2	76.5
CEMP [30]	71.3	81.8	76.7	75.7	80.0	–
CircReg [33]	77.7	84.0	75.8	68.8	89.9	93.6
GNC [46]	60.6	84.4	83.2	91.0	86.7	98.9
Odometry	42.4	1.9	1.9	82.6	20.2	18.8
Runtime (s)						
Ours	5.2	2.9	4.4	1.5	19.8	237
ℓ_1-IRLS [10]	**0.1**	**0.1**	**0.1**	**0.02**	**0.5**	1.2
Spectral [41]	6.2	6.3	6.0	8.4	11.5	**0.1**
CEMP [30]	6.4	6.6	6.0	7.8	12.0	–
CircReg [33]	7.0	6.9	6.2	8.7	13.3	11.5
GNC [46]	6.6	6.5	6.0	8.3	12.2	10.4

4.3 2D Pose-Graph Optimization

Setup. We recover the headings of six standard planar pose graphs [9] (INTEL, CSAIL, FR079, MITb, M3500 [31] and City10000 [26], with 808–10^4 poses) from odometry, trusted as in robust-PGO work [42], and loop closures, a fraction f of which we replace with $\mathrm{Uniform}[0, 2\pi)$ draws (5 random seeds). With no ground-truth trajectories, we score gauge-aligned per-pose errors against the rotation-only solution of the uncorrupted graph, which our BnB certifies on four of the six graphs.

Results. At $f{=}0.5$ (Table 2), we stay near one degree on INTEL, CSAIL, FR079 and City10000, where ℓ_1-IRLS errs by 19–53° and Spectral, CEMP, CircReg and GNC by 60–99°: on near-tree pose graphs each loop closure has high leverage, and the small spectral gap weakens spectral and spectrally initialized methods. MITb, with 20 loop closures for 808 poses, is the hardest graph for our method. Our runtime grows from $1.5\,\mathrm{s}$ to $4\,\mathrm{min}$ at $N{=}10^4$, slower than ℓ_1-IRLS but small against the place-recognition front end.

5 Conclusion

We presented the first certifiably optimal algorithm for robust 1D rotation averaging on S^1, solving a smoothed pairwise max-consensus MRF by branch-and-bound with ICM-guided label pruning and a node-constrained upper bound that together collapse the search space by four orders of magnitude on the tested instances. On up to $N{=}50$ and 80% outliers the algorithm certifies the bin-level global optimum in about $10\,\mathrm{s}$. Across image SfM, multi-way point cloud registration, cryo-EM in-plane synchronization and 2D pose-graph optimization, it beats baselines at high outlier rates while matching them in the clean regime, with upstream-dominated runtime overhead – a practical demonstration that certified combinatorial optimization is tractable for a 1D geometric estimation problem of wide interest.

Limitations. The certificate concerns the smoothed MRF. Recovering the original max-consensus inlier set additionally requires the separation condition (3) of Theorem 2 and holds w.h.p., the only step where the i.i.d. outlier model enters. Graph-cut refinement is inexact on the $\sim 6\%$ non-supermodular edges, which affects only the IRLS warm start, never the optimal inlier set. BnB trees grow on sequential graphs beyond 40% outliers. Dense graphs beyond $N \approx 10^3$ become memory- and BnB-bound ($O(|\mathcal{E}|K^2)$ potential tables), whereas sparse SLAM-style graphs scale to at least $N{=}10^4$. On such graphs, certification – as opposed to solution quality – can degrade gracefully into a small optimality gap.

References

[1] Matthew E. Antone and Seth Teller. Automatic recovery of relative camera rotations for urban scenes. In *Proc. CVPR*, volume 2, pages 282–289, 2000.

[2] Afonso S Bandeira, Nicolas Boumal, and Amit Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, 163:145–167, 2017.

[3] Soojay Banerjee, Alberto Bartesaghi, Alan Merk, Prashant Rao, Stacie L. Bulfer, Yongzhao Yan, Neal Green, Barbara Mroczkowski, R. Jeffrey Neitz, Peter Wipf, Veronica Falconieri, Raymond J. Deshaies, Jacqueline L. S. Milne, Donna Huryn, Michelle Arkin, and Sriram Subramaniam. 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science*, 351(6275):871–875, 2016.

[4] Daniel Barath. SupeRANSAC: One RANSAC to rule them all. *arXiv preprint arXiv:2506.04803*, 2025.

[5] Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48(3):259–302, 1986.

[6] Nicolas Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.

[7] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE TPAMI*, 26(9):1124–1137, 2004.

[8] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE TPAMI*, 23(11):1222–1239, 2001.

[9] Luca Carlone and Andrea Censi. From angular manifolds to the integer lattice: Guaranteed orientation estimation with application to pose graph optimization. *IEEE Trans. Robotics*, 30(2):475–492, 2014.

[10] Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In *Proc. ICCV*, pages 521–528, 2013.

[11] Tat-Jun Chin, Zhipeng Cai, and Frank Neumann. Robust fitting in computer vision: Easy or hard? In *Proc. ECCV*, pages 715–730, 2018.

[12] Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In *Proc. CVPR*, pages 2511–2520, 2020.

[13] Computational CryoEM group. ASPIRE-Python: Algorithms for single particle reconstruction. `https://github.com/ComputationalCryoEM/ASPIRE-Python`, 2024. Princeton University.

[14] Frank Dellaert, David M Rosen, Jing Wu, Robert Mahony, and Luca Carlone. Shonan rotation averaging: Global optimality by surfing $SO(p)^n$. In *Proc. ECCV*, pages 292–308, 2020.

[15] Anders Eriksson, Carl Olsson, Fredrik Kahl, and Tat-Jun Chin. Rotation averaging and strong duality. In *Proc. CVPR*, pages 127–135, 2018.

[16] 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.

[17] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *Proc. CVPR*, pages 3354–3361, 2012.

[18] Venu Madhav Govindu. Combining two-view constraints for motion estimation. In *Proc. CVPR*, volume 2, pages 218–225, 2001.

[19] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *Int. J. Computer Vision*, 103(3):267–305, 2013.

[20] David Haselbach, Jil Schrader, Felix Lambrecht, Fabian Henneberg, Ashwin Chari, and Holger Stark. Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nature Communications*, 8:15578, 2017.

[21] Yixuan He, Gesine Reinert, David Wipf, and Mihai Cucuringu. Robust angular synchronization via directed graph neural networks. In *Proc. ICLR*, 2024.

[22] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[23] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. PREDATOR: Registration of 3D point clouds with low overlap. In *Proc. CVPR*, pages 4267–4276, 2021.

[24] Shengze Jin, Iro Armeni, Marc Pollefeys, and Daniel Barath. Multiway point cloud mosaicking with diffusion and global optimization. In *Proc. CVPR*, pages 20838–20849, 2024.

[25] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image matching across wide baselines: From paper to practice. *Int. J. Computer Vision*, 129:517–547, 2021.

[26] Michael Kaess, Ananth Ranganathan, and Frank Dellaert. iSAM: Incremental smoothing and mapping. *IEEE Trans. Robotics*, 24(6):1365–1378, 2008.

[27] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE TPAMI*, 28(10):1568–1583, 2006.

[28] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? *IEEE TPAMI*, 26(2):147–159, 2004.

[29] Seong Hun Lee and Javier Civera. HARA: A hierarchical approach for robust rotation averaging. In *Proc. CVPR*, pages 15777–15786, 2022.

[30] Gilad Lerman and Yunpeng Shi. Robust group synchronization via cycle-edge message passing. *Foundations of Computational Mathematics*, 22:1665–1741, 2022.

[31] Edwin Olson, John Leonard, and Seth Teller. Fast iterative alignment of pose graphs with poor initial estimates. In *Proc. IEEE ICRA*, pages 2262–2269, 2006.

[32] Linfei Pan, Daniel Barath, Marc Pollefeys, and Johannes L. Schönberger. Global structure-from-motion revisited. In *Proc. ECCV*, pages 58–77, 2024.

[33] Linfei Pan, Marc Pollefeys, and Daniel Barath. Gravity-aligned rotation averaging with circular regression. In *Proc. ECCV*, pages 97–116, 2024.

[34] Alvaro Parra, Shin-Fang Chng, Tat-Jun Chin, Anders Eriksson, and Ian Reid. Rotation coordinate descent for fast globally optimal rotation averaging. In *Proc. CVPR*, pages 4298–4307, 2021.

[35] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Message-passing algorithms for synchronization problems over compact groups. *Communications on Pure and Applied Mathematics*, 71(11):2275–2322, 2018.

[36] Pulak Purkait, Tat-Jun Chin, and Ian Reid. NeuRoRA: Neural robust rotation averaging. In *Proc. ECCV*, pages 137–154, 2020.

[37] Tong Qin, Peiliang Li, and Shaojie Shen. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robotics*, 34(4):1004–1020, 2018.

[38] 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. *Int. J. Robotics Research*, 38(2–3):95–125, 2019.

[39] Johannes L Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proc. CVPR*, pages 4104–4113, 2016.

[40] Yunpeng Shi and Gilad Lerman. Message passing least squares framework and its application to rotation synchronization. In *Proc. ICML*, pages 8796–8806, 2020.

[41] Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1):20–36, 2011.

[42] Niko Sünderhauf and Peter Protzel. Switchable constraints for robust pose graph SLAM. In *Proc. IEEE/RSJ IROS*, pages 1879–1884, 2012.

[43] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE TPAMI*, 30(6):1068–1080, 2008.

[44] Martin J Wainwright, Tommi S Jaakkola, and Alan S Willsky. MAP estimation via agreement on trees: Message-passing and linear programming. *IEEE Trans. Information Theory*, 51(11):3697–3717, 2005.

[45] Lanhui Wang and Amit Singer. Exact and stable recovery of rotations for robust synchronization. *Information and Inference*, 2(2):145–193, 2013.

[46] 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 Robotics and Automation Letters*, 5(2):1127–1134, 2020.

[47] Heng Yang and Luca Carlone. Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization. *IEEE TPAMI*, 45(3):2816–2834, 2023.

[48] Heng Yang, Jingnan Shi, and Luca Carlone. TEASER: Fast and certifiable point cloud registration. *IEEE Trans. Robotics*, 37(2):314–333, 2021.

[49] Christopher Zach, Manfred Klopschitz, and Marc Pollefeys. Disambiguating visual relations using loop constraints. In *Proc. CVPR*, pages 1426–1433, 2010.

[50] Ganlin Zhang, Viktor Larsson, and Daniel Barath. Revisiting rotation averaging: Uncertainties and robust losses. In *Proc. CVPR*, pages 17215–17224, 2023.

[51] Yiqiao Zhong and Nicolas Boumal. Near-optimal bounds for phase synchronization. *SIAM Journal on Optimization*, 28(2):989–1016, 2018.

A Proofs

This appendix contains the full proofs of all results stated in the main paper.

A.1 Proof of Proposition 1 (Discretization preserves inliers)

Proof. Each bin center $c_{\hat{\ell}_i}$ lies within $\alpha/2$ of θ_i^*. For an inlier edge under $\boldsymbol{\theta}^*$ the continuous residual is $d_\circ(\tilde{\theta}_{ij},\, (\theta_j^* - \theta_i^*) \bmod 2\pi) < \tau$, and the discretization perturbs the predicted relative angle $\theta_j^* - \theta_i^*$ by at most α (one $\alpha/2$ from each endpoint). By the triangle inequality on S^1, the discretized residual $d_\circ(c_{\hat{\ell}_j} - c_{\hat{\ell}_i},\, \tilde{\theta}_{ij})$ is therefore at most $\tau + \alpha$, which is $\leq r = \max(\alpha,\, 3\sigma)$ for the default choice $\sigma = \tau$ as long as $\alpha \leq 2\tau$ (satisfied by our default K). The Gaussian branch of (2) thus applies and gives $w_{ij}(\hat{\ell}_i, \hat{\ell}_j) \geq e^{-(\tau+\alpha)^2/(2\sigma^2)}$. Outlier edges contribute ≥ 0, so summing over the C^* inlier edges yields $E(\hat{\boldsymbol{\ell}}) \geq C^* \cdot e^{-(\tau+\alpha)^2/(2\sigma^2)}$, and $E_K^* \geq E(\hat{\boldsymbol{\ell}})$ because $\hat{\boldsymbol{\ell}}$ is one feasible labeling. □

A.2 Proof of Theorem 2 (Discrete-continuous equivalence)

Proof. **Lower bound on $E(\hat{\boldsymbol{\ell}})$.** The nearest-bin labeling $\hat{\ell}_i = \lfloor \theta_i^*/\alpha \rfloor$ places each bin center within $\alpha/2$ of the true angle, so under $\hat{\boldsymbol{\ell}}$ the discretized residual on any continuous-inlier edge is at most $\delta_{\max} + \alpha < r$. The Gaussian branch of (2) therefore applies and gives weight $\geq w_{\min}$ per inlier edge; all other edges contribute non-negatively, hence $E(\hat{\boldsymbol{\ell}}) \geq C^*\, w_{\min}$.

$\hat{\boldsymbol{\ell}}$ realizes $\boldsymbol{\theta}^*$'s inlier set. For every edge, the discretized residual under $\hat{\boldsymbol{\ell}}$ differs from the continuous residual under $\boldsymbol{\theta}^*$ by at most α (one $\alpha/2$ per endpoint). Since $\alpha < \min_{ij} \gamma_{ij}$, it stays on the same side of τ, so $\hat{\boldsymbol{\ell}}$ classifies every edge identically to $\boldsymbol{\theta}^*$.

Upper bound on $E(\boldsymbol{\ell}')$ for any structurally distinct $\boldsymbol{\ell}'$. Treat the bin centers of $\boldsymbol{\ell}'$ as a continuous angle vector with residuals ε_{ij}. Every weight satisfies $w_{ij} \leq (1 - c_\tau)\, \mathbf{1}[\varepsilon_{ij} < \tau] + c_\tau\, \mathbf{1}[\varepsilon_{ij} < r]$: it is at most 1 below τ, at most c_τ on $[\tau, r)$, and 0 beyond r. Since $\boldsymbol{\ell}'$ realizes an inlier set distinct from $\boldsymbol{\theta}^*$'s, at most C_2^* edges have $\varepsilon_{ij} < \tau$. At most M_{in} non-outlier edges have $\varepsilon_{ij} < r$, plus the $S(\boldsymbol{\ell}')$ outlier edges whose measurement lands inside the kernel support. Hence

$$E(\boldsymbol{\ell}') \;\leq\; (1 - c_\tau)\, C_2^* + c_\tau \big(M_{\text{in}} + S(\boldsymbol{\ell}')\big).$$

Uniform bound on S. For a fixed $\boldsymbol{\ell}'$, each outlier measurement lands within r of $c_{\ell'_j} - c_{\ell'_i}$ independently with probability $2r/(2\pi) = r/\pi$, so $S(\boldsymbol{\ell}')$ is a sum of M_{out} independent Bernoulli(r/π) variables, and Hoeffding's inequality [22] gives $\Pr[S(\boldsymbol{\ell}') > M_{\text{out}}\, r/\pi + t] \leq e^{-2t^2/M_{\text{out}}}$. Choosing $t = \sqrt{\frac{1}{2} M_{\text{out}} \ln(K^{N-1}/\delta)}$ and taking a union bound over all K^{N-1} gauge-fixed labelings, with probability at least $1 - \delta$ we have $S(\boldsymbol{\ell}') \leq B_\delta$ for every $\boldsymbol{\ell}'$ simultaneously, including labelings that depend on the data, such as $\boldsymbol{\ell}^*$.

Conclusion. On this event, (3) gives $E(\boldsymbol{\ell}') \leq (1 - c_\tau)\, C_2^* + c_\tau (M_{\text{in}} + B_\delta) < C^* w_{\min} \leq E(\hat{\boldsymbol{\ell}})$ for every structurally distinct $\boldsymbol{\ell}'$. The MRF optimum satisfies $E(\boldsymbol{\ell}^*) \geq E(\hat{\boldsymbol{\ell}})$, so it is not structurally distinct: it realizes the same inlier set as $\hat{\boldsymbol{\ell}}$, which equals $\boldsymbol{\theta}^*$'s. □

A.3 Proof of Proposition 3 (Pruning safety)

Proof. By the definition of ρ_i^{ICM} and δ_{in}, the agreeing inlier neighbors alone give $u_i(k^*) \geq (1 - \eta)\, d_i\, \rho_i^{\text{ICM}}\, \delta_{\text{in}}$. Fix a wrong label $k \neq k^*$. Each of the $(1 - \eta)\, d_i$ inlier neighbors contributes at most 1 to $u_i(k)$. An outlier neighbor contributes only if its measurement lands within r of $c_{\ell_j^{\text{ICM}}} - c_k$; with the ICM labels fixed, this happens independently with probability $2r/(2\pi) = r/\pi$. The number S_k of such hits is thus a sum of $\eta\, d_i$ independent Bernoulli(r/π) variables, and Hoeffding's inequality [22] gives

$$\Pr\Big[S_k \;>\; \eta\, d_i\, \tfrac{r}{\pi} + \sqrt{\tfrac{1}{2} \eta\, d_i\, \ln(K/\delta)}\Big] \;\leq\; \delta/K.$$

A union bound over the at most $K - 1$ wrong labels shows that, with probability at least $1 - \delta$, every wrong label satisfies $u_i(k) \leq (1 - \eta)\, d_i + \eta\, d_i\, r/\pi + \sqrt{\frac{1}{2} \eta\, d_i\, \ln(K/\delta)}$. Label k^* is pruned only

if $u_i(k^*) < \rho \max_{k'} u_i(k')$, and since $\rho < 1$ the maximum must then be attained at a wrong label. On the event above, the condition of the proposition gives $u_i(k^*) > \rho\, u_i(k)$ for every wrong k, so $k^* \in F_i$. □

A.4 Proof of Theorem 4 (Admissible upper bound)

Proof. Partition $\mathcal{E} = \mathcal{E}_{\text{fixed}} \cup \mathcal{E}_{\text{mixed}} \cup \mathcal{E}_{\text{free}}$ according to whether zero, one, or both endpoints lie in $\mathcal{F}$, and decompose $E(\boldsymbol{\ell}) = E_{\text{fixed}} + E_{\text{mixed}}(\boldsymbol{\ell}) + E_{\text{free}}(\boldsymbol{\ell})$ accordingly. By construction E_{fixed} equals the first sum on the right-hand side of (6), so it suffices to show

$$\sum_{i\in\mathcal{F}} \max_{k\in F_i} \Bigg[\sum_{\substack{j\in N(i)\\ j\notin\mathcal{F}}} w_{ij}(k,\ell_j) + \tfrac{1}{2} \sum_{\substack{j\in N(i)\\ j\in\mathcal{F}}} \max_{l\in F_j} w_{ij}(k,l) \Bigg] \;\geq\; E_{\text{mixed}}(\boldsymbol{\ell}) + E_{\text{free}}(\boldsymbol{\ell}). \tag{7}$$

Fix any feasible $\boldsymbol{\ell}$ with $\ell_i \in F_i$ for every $i \in \mathcal{F}$. For each free node i, evaluating the bracket on the left of (7) at the specific label $k = \ell_i$ instead of taking the maximum can only decrease it, so

$$\sum_{i\in\mathcal{F}} \max_{k\in F_i} \Big[\cdots \Big] \;\geq\; \sum_{i\in\mathcal{F}} \Bigg[\sum_{\substack{j\in N(i)\\ j\notin\mathcal{F}}} w_{ij}(\ell_i,\ell_j) + \tfrac{1}{2} \sum_{\substack{j\in N(i)\\ j\in\mathcal{F}}} \max_{l\in F_j} w_{ij}(\ell_i,l) \Bigg]. \tag{8}$$

The first inner sum, summed over $i \in \mathcal{F}$, equals $E_{\text{mixed}}(\boldsymbol{\ell})$ exactly: each mixed edge has a unique free endpoint and is counted once. For the second inner sum, since $\ell_j \in F_j$ for every free j, we have $\max_{l\in F_j} w_{ij}(\ell_i,l) \geq w_{ij}(\ell_i,\ell_j)$. Hence

$$\tfrac{1}{2} \sum_{i\in\mathcal{F}} \sum_{\substack{j\in N(i)\\ j\in\mathcal{F}}} \max_{l\in F_j} w_{ij}(\ell_i,l) \;\geq\; \tfrac{1}{2} \sum_{i\in\mathcal{F}} \sum_{\substack{j\in N(i)\\ j\in\mathcal{F}}} w_{ij}(\ell_i,\ell_j) \;=\; \tfrac{1}{2}\cdot 2 \sum_{(i,j)\in\mathcal{E}_{\text{free}}} w_{ij}(\ell_i,\ell_j) \;=\; E_{\text{free}}(\boldsymbol{\ell}),$$

where the factor of two arises because each free–free edge appears once from each of its two endpoints. Combining the two contributions establishes (7), and adding E_{fixed} to both sides gives $\bar{E}_{\text{node}} \geq E(\boldsymbol{\ell})$. The argument never invokes translation invariance, the structure of F_i, or the kernel form, so it remains valid for arbitrary pairwise potentials and arbitrary feasible sets. □

B BnB Pseudo-code

Algorithm 1 below summarizes the certified branch-and-bound procedure described in Section 3.3.

Algorithm 1 Certified BnB for the 1D max-consensus MRF.

Input: pairwise weights $w_{ij}(\cdot,\cdot)$, K, pruning threshold ρ=0.3, tolerance ε
Run multi-start ICM (Eq. (4)); set $\boldsymbol{\ell}^{\text{ICM}}$ and primal $\underline{E} \leftarrow E(\boldsymbol{\ell}^{\text{ICM}})$
Build $F_i \leftarrow \{k : u_i(k) \geq \rho \max_{k'} u_i(k')\} \cup \{\ell_i^{\text{ICM}}\}$; if $|F_i| < 3$, add $\ell_i^{\text{ICM}} \pm 1, \pm 2$ (Sec. 3.3)
Push the root partial assignment $\{\ell_0 = 0\}$ to the BnB stack
while stack non-empty **do**
  Pop a partial assignment; compute $\bar{E}_{\text{node}}$ (Eq. (6))
  if $\bar{E}_{\text{node}} \leq \underline{E} + \varepsilon$ **then**
    prune this branch
  end if
  if all nodes are fixed **then**
    update primal: $\underline{E} \leftarrow \max(\underline{E}, E(\boldsymbol{\ell}))$
  else
    pick the free node i^* with most free neighbors; for each $k \in F_{i^*}$ in best-first order, push $\{\ell_{i^*} = k\}$
  end if
end while
Return: certified-optimal labeling $\boldsymbol{\ell}^*$ with $E(\boldsymbol{\ell}^*) = \underline{E}$

Defaults. We run $R = \max(20, \min(200, |\mathcal{E}|))$ ICM restarts (200 on the N=50 benchmarks), each for at most 100 sweeps, with $\rho = 0.3$ and $\varepsilon = 10^{-6}$. BnB stops after 10 s or $5 \cdot 10^6$ nodes and then returns the incumbent with its gap. Refinement uses $L = 6$ levels for $\tau = 5^\circ$ ($\alpha/2^L \approx 0.09^\circ$), $n_q = 4$ quadrature samples per sub-interval and $\sigma_{\text{ref}} = \tau$. Our final polish runs five circular-ℓ_1 (ADMM) iterations and then Geman–McClure IRLS with $\sigma = \tau$ on the classified inliers; on the SfM and LiDAR benchmarks, our output instead initializes GLOMAP's IRLS (Appendix C).

C Real-World Experiments: Detailed Protocol

This appendix gives the full per-application protocol behind Table 1 of the main paper. Across all three application families (image SfM, multi-way LiDAR registration, cryo-EM) the rotation-averaging step receives a list of pairwise yaw measurements $\{(\tilde{\theta}_{ij},\,(i,j))\}$ together with a per-node gravity (or common viewing direction) extracted from the ground-truth poses, and outputs a per-node absolute angle. The only component that changes between methods is this rotation-averaging step; all upstream feature extraction, two-view geometry, and pose estimation are shared. On SfM and LiDAR, CircReg and ours run GLOMAP's gravity-aware IRLS polish, included in their runtime. Spectral, CEMP, GNC and TLS-SDP solve the same gravity-aligned 1D problem standalone, and ℓ_1-IRLS is GLOMAP's 3-DoF averager without gravity. On cryo-EM, every method solves the 1D in-plane problem standalone.

C.1 Image-Based Structure from Motion

Datasets. We use three benchmarks from the Image Matching Challenge [25] that span small repeatable-symmetry scenes to large crowd-sourced phototourism collections: *haiper* from IMC 2023 (3 scenes: `bike`, `chairs`, `fountain`), *phototourism* from IMC 2023 (15 landmark scenes including `sacre_coeur`, `brandenburg_gate`, `taj_mahal`, etc.; `temple_nara_japan` lacks GT), and IMC 2024 (4 scenes: `church`, `dioscuri`, `lizard`, `pond`), plus KITTI [17] (11 scenes; with stride $s = 10$; using a sequential graph).

Pipeline. We use the GLOMAP [32] pipeline as implemented in the official release. SIFT features and exhaustive two-view matching produce the COLMAP database; GLOMAP's mapper estimates pairwise two-view geometries. The relative rotations R_{ij} written out by GLOMAP's `--relpose_output_path` are the input. Per-image gravity g_i is taken as the projection of the world-frame $[0, 1, 0]$ axis through the GT camera rotation; the pairwise yaw $\tilde{\theta}_{ij}$ is the gravity-aligned in-plane component of R_{ij} (a Wahba-style decomposition aligning g_i and g_j to a common vertical and reading off the residual rotation around that axis). The standalone `glomap rotation_averager` is then invoked with the appropriate combination of `--relpose_path`, `--gravity_path`, and (for our method) `--initial_rotation_path`, so CircReg and ours share GLOMAP's IRLS polish stage.

Metric. For every pair of GT-registered images (i, j) we compute the geodesic angular error of the estimated relative rotation against the GT relative rotation; we then average per-scene across all pairs and report rotation-only AUC@$3^\circ/5^\circ/10^\circ$, the mean and median pairwise error, and the wall-clock time of the rotation-averaging step (excluding GLOMAP's feature extraction and matching, which dominates total compute).

C.2 Multi-Way LiDAR Point-Cloud Registration

Datasets. We evaluate on *KITTI* odometry [17] sequences 00–10, sub-sampled at strides s=5 and s=10 frames (the LiDAR is recorded at 10 Hz, so s=5 corresponds to $\sim$0.5 s spacing and s=10 to $\sim$1 s). The two strides yield two qualitatively different graphs: the denser low-stride graph (s=5) and the lower-overlap graph (s=10).

Pipeline. Velodyne scans feed into Predator [23], producing per-pair 3D-3D correspondences with overlap and saliency scores. We sample correspondences proportional to overlap $\times$ saliency, then fit a rigid $SE(3)$ transform per pair via SupeRANSAC [4] with $50{,}000$ iterations and an inlier threshold matched to Predator's KITTI default. For KITTI we additionally pass each estimated R_{ij}^{velo} through the velodyne-to-camera extrinsic (from `calib_velo_to_cam.txt`) so the rotation-averaging output lives in the same frame as the GT poses. Gravity is the $-y$-axis of the camera frame on KITTI (both extracted from the GT poses). The rotation-averaging step is then identical to the SfM case: `glomap rotation_averager` reading the same edges file, with the same IRLS polish and the same per-method settings.

Metric. Per-pair geodesic rotation error against GT, averaged within sequence, then averaged across sequences (11 KITTI sequences per stride).

C.3 Cryo-EM In-Plane Synchronization

Dataset. We project the EMD-3295 map (human p97 bound to the inhibitor UPCDC30245, 2.3 Å [3]) and draw all imaging parameters from the EMPIAR-10166 particle stack (human 26S proteasome [20]). We follow the standard cryo-EM angular-synchronization evaluation [41, 13]: the EMPIAR-10166 raw-particle dataset does not provide per-particle ground-truth in-plane angles for the synchronization problem in isolation; the projection-based protocol below closes that gap while keeping every imaging parameter, including the per-particle CTF, drawn from the actual EMPIAR-10166 metadata.

Pipeline. We start from the EMD-3295 map ($312\times312\times312$ voxels, 0.637 Å pixel size) and downsample it to $64\times64\times64$ via Fourier-truncation. We then generate $N{=}100$ projections at a single fixed viewing direction (Euler angles 30° and 45°) with random ground-truth in-plane rotations $\psi_i \sim \mathrm{Uniform}[0, 2\pi)$. Each projection is convolved with a per-particle CTF (`aspire.operators.RadialCTFFilter`) whose defocus is sampled from the empirical EMPIAR-10166 distribution – range 2,479–84,147 Å, mean 26,626 Å, drawn from the `rlnDefocusU/V` columns of the EMPIAR-10166 `particles_0.star` file – with the EMPIAR-10166 optics block (300 kV, $C_s{=}2.7$ mm, amplitude contrast 0.1). White Gaussian noise is added at SNR 1 (signal variance measured on the clean CTF-applied stack, noise variance set to match). Pairwise in-plane angle estimates $\tilde{\psi}_{ij}$ come from a 360-candidate normalized cross-correlation rotation search (one rotation candidate per degree) after a circular mask at radius 0.45 of the image side and a soft cosine low-pass at 0.25 of Nyquist. The measurement graph is a random spanning path plus uniformly random pairs, up to $4N = 400$ edges (mean degree 8). We additionally inject a 40% uniform-random outlier rate on top of the naturally noisy pairwise estimates and run the rotation averaging on the resulting graph. AUCs and the mean error are averaged over 10 seeds (5 for TLS-SDP); the median column is the median of the per-seed medians.

Metric. For each particle we recover an estimated absolute $\hat{\psi}_i$, gauge-align (median-of-offsets) to the ground-truth ψ_i^{GT}, and report rotation AUC@$3^\circ/5^\circ/10^\circ$, mean and median absolute angular error.

C.4 Extended Real-World Comparison

Table 3 below is the full version of Table 1 of the main paper, including Spectral [41], CEMP [30], GNC [46] and TLS-SDP [47], which were omitted from the main-paper table for space. Spectral, CEMP and GNC solve the same gravity-aligned 1D problem as CircReg and ours, yet are dominated by CircReg or our method on every dataset.

Table 3: Extended real-world rotation-averaging results, including Spectral [41], CEMP [30], GNC [46], and a QUASAR-style TLS-SDP relaxation [47] (omitted from the main-paper Table 1 for space). Metrics and protocol are identical to Table 1. TLS-SDP (cvxpy+SCS) runs only on datasets whose scenes all have $N \leq 200$, with 5 seeds on EMPIAR-10166. On IMC 2024, GNC's linear solves fail on graphs with isolated nodes and it returns its Spectral initialization; on EMPIAR-10166, most edges lie in no 3-cycle, so CEMP keeps every edge and reduces to Spectral.

	Dataset	Method	AUC@3° ↑	AUC@5° ↑	AUC@10° ↑	mean ↓	median ↓	time (s)
Global Structure-from-Motion	Haiper	Spectral [41]	28.0	36.3	49.1	8.3	6.7	**0.1**
		CEMP [30]	34.5	42.6	58.3	9.0	3.5	**0.1**
		GNC [46]	35.1	43.3	55.5	6.8	5.0	**0.1**
		TLS-SDP [47]	34.2	42.3	54.0	14.8	6.4	0.5
		ℓ_1-IRLS [10]	26.2	34.2	53.2	6.6	4.0	1.3
		CircReg [33]	44.5	55.2	70.3	2.4	1.6	1.3
		Ours	**46.4**	**57.3**	**73.6**	**1.7**	**1.5**	1.7
	Phototourism	Spectral [41]	37.2	47.4	60.0	10.0	2.5	10.3
		CEMP [30]	42.7	51.5	62.6	11.4	2.2	680.4
		GNC [46]	43.1	51.5	62.6	10.0	2.1	17.9
		TLS-SDP [47]	Not run: scenes with $N > 200$.					
		ℓ_1-IRLS [10]	34.2	44.5	57.7	12.3	2.7	88.2
		CircReg [33]	44.5	53.1	64.5	5.3	2.1	**1.2**
		Ours	**44.8**	**53.7**	**65.3**	**5.1**	**1.9**	82.2
	IMC 2024	Spectral [41]	7.8	12.4	22.0	42.8	29.8	3.1
		CEMP [30]	8.4	13.0	22.5	42.2	29.0	14.7
		GNC [46]	7.8	12.4	22.0	42.8	29.8	8.2
		TLS-SDP [47]	Not run: scenes with $N > 200$.					
		ℓ_1-IRLS [10]	5.1	9.6	19.3	38.1	28.2	8.1
		CircReg [33]	14.0	21.6	34.3	**32.5**	19.7	**0.3**
		Ours	**15.8**	**23.5**	**35.9**	33.0	**19.6**	15.6
	KITTI	Spectral [41]	4.0	6.2	11.2	60.5	53.6	3.9
		CEMP [30]	15.0	18.6	25.2	44.9	37.1	3.0
		GNC [46]	15.5	17.1	20.5	52.7	47.0	2.8
		TLS-SDP [47]	Not run: scenes with $N > 200$.					
		ℓ_1-IRLS [10]	40.2	44.9	49.9	44.0	28.9	**1.7**
		CircReg [33]	53.8	58.9	64.4	30.6	20.6	**1.7**
		Ours	**59.6**	**68.2**	**77.3**	**5.9**	**4.6**	2.3
Cryo-EM	EMPIAR-10166	Spectral [41]	7.6	11.1	19.8	22.7	14.4	1.1
		CEMP [30]	7.6	11.1	19.8	22.7	14.4	0.9
		GNC [46]	14.8	20.0	27.8	25.5	11.5	0.8
		TLS-SDP [47]	4.3	7.3	11.9	76.3	52.2	33.7
		ℓ_1-IRLS [10]	50.2	61.8	71.5	15.0	1.3	**0.0**
		CircReg [33]	52.0	64.6	75.4	11.1	1.2	1.0
		Ours	**57.9**	**71.9**	**83.4**	**6.4**	**1.1**	10.6
Multi-way Point Cloud Registration	KITTI (s=5)	Spectral [41]	39.9	48.6	59.1	20.0	12.1	**0.0**
		CEMP [30]	52.4	58.6	65.0	21.6	12.8	0.6
		GNC [46]	56.2	63.1	70.4	14.4	11.1	0.5
		TLS-SDP [47]	Not run: scenes with $N > 200$.					
		ℓ_1-IRLS [10]	28.5	43.4	61.7	9.5	8.0	0.1
		CircReg [33]	60.1	67.4	75.1	8.4	5.6	0.1
		Ours	**60.6**	**68.4**	**76.6**	**6.7**	**2.4**	2.3
	KITTI (s=10)	Spectral [41]	13.0	16.4	22.9	58.0	45.3	**0.0**
		CEMP [30]	15.4	19.2	25.4	51.9	41.1	0.1
		GNC [46]	20.1	24.1	29.1	53.9	39.5	0.1
		TLS-SDP [47]	Not run: scenes with $N > 200$.					
		ℓ_1-IRLS [10]	15.6	22.4	30.4	45.7	31.8	0.1
		CircReg [33]	31.8	36.5	41.6	33.9	29.6	**0.0**
		Ours	**34.1**	**39.2**	**44.9**	**27.4**	**20.3**	10.4

C.5 Scaling Behavior of Certified BnB

The main-paper synthetic experiments fix N=50. Table 4 sweeps $N \in \{20, 50, 100, 200\}$ at $\eta \in \{0.5, 0.8\}$ for fully connected (degree $N-1$) and sparse ($\sim$5 edges/node) graphs, 5 instances per cell, σ_{noise}=2°.

Table 4: Certified BnB scaling on synthetic graphs of varying size N and outlier rate η ($\sigma_{\text{noise}} = 2°$, $\tau = 5°$, $K = 64$, 5 seeds per cell). *full* is the complete graph (degree $N-1$); *sparse* samples $\sim$5 edges per node. We report mean total wall-clock time and mean median rotation error. Cells with median error replaced by "–" are below the information-theoretic threshold for angular synchronization: the expected number of inlier neighbors per node, $(1-\eta)\,d$, falls under $\ln N$ (3.0–5.3 for the tested N), so even the global optimum cannot recover accurate angles, and the runtime entry only reports BnB termination on a degenerate problem. The full N=20, η=0.8 cell is discussed in the text. The main-paper "sparse" protocol (Sec. 4.1) uses a denser $\sim$32 edges/node and remains above this threshold at all tested η.

		$N = 20$		$N = 50$		$N = 100$		$N = 200$	
Graph	η	total	err	total	err	total	err	total	err
full	0.5	0.31 s	0.40	5.2 s	0.29	19.5 s	0.22	1.2 min	0.15
full	0.8	10.4 s	–	12.2 s	0.54	23.1 s	0.36	1.1 min	0.25
sparse	0.5	1.1 s	–	10.3 s	–	10.6 s	–	11.3 s	–
sparse	0.8	0.74 s	–	10.3 s	–	10.6 s	–	11.2 s	–

Information-theoretic floor. Angular synchronization on a noisy graph requires the inlier subgraph to be (well) connected; under $i.i.d.$ inlier/outlier sampling, that needs the expected inlier degree per node $\bar{d}_{\text{in}} = (1-\eta)\,d$ to exceed roughly $\ln N$ (Erdős–Rényi connectivity threshold). With this appendix's $\sim$5 edges/node, $\bar{d}_{\text{in}} = 2.5$ at $\eta = 0.5$ and $\bar{d}_{\text{in}} = 1.0$ at $\eta = 0.8$ – both below the threshold for any of the tested N, so even the globally optimal labeling cannot recover accurate angles and we mark the error column "–". The smallest fully connected graph at η=0.8 ($\bar{d}_{\text{in}}$=3.8, just above $\ln 20 \approx 3.0$) is also marked "–": BnB hits its 10 s budget there and the error is 18.8°. The main-paper sparse experiments use $\sim$32 edges/node, giving $\bar{d}_{\text{in}} \approx 6.4$ even at η=0.8, which sits comfortably above the threshold and explains why our solver attains 0.87° there.

Scaling regime. Where the inlier graph is connected, cost grows roughly linearly with $|\mathcal{E}|$ at η=0.5 and is dominated by BnB-tree exploration at η=0.8. On N=200 fully-connected graphs at 80% outliers the per-instance certified time is about a minute, and accuracy improves with N (mean median error from 0.54° at N=50 to 0.25° at N=200).

D 2D Pose-Graph Optimization: Protocol and Full Results

This appendix details the protocol behind Table 2 and reports the full results.

Datasets. We use the planar pose graphs distributed with [9]: INTEL, CSAIL, FR079, MITb, M3500 [31], and City10000 [26] (g2o or TORO format). We read only the relative heading $\Delta\tilde{\theta}_{ij} \approx \theta_j - \theta_i$ of each edge. Edges between consecutive poses ($|i-j|$=1) are odometry; all other edges are loop closures. Graph sizes are listed in Table 5.

Corruption protocol. Odometry edges stay clean. For a corruption level $f \in \{0, 0.1, 0.25, 0.5, 0.75, 0.9\}$, a uniformly random subset of $\lfloor f \cdot \#\text{loops} \rceil$ loop closures is replaced by draws from $\text{Uniform}[0, 2\pi)$. We use 5 seeds per level and one instance at f=0. On City10000, where one run of our solver takes about 4 min, we evaluate only $f \in \{0.1, 0.25, 0.5\}$.

Reference and metric. The datasets ship no ground-truth trajectories. We therefore compute a reference by running our solver, with BnB, on the uncorrupted graph. It is certified on CSAIL, FR079, M3500 and City10000 (where all 64,311 edges are recovered as inliers in 265 s); on INTEL and MITb, BnB leaves a gap of 1.9% and 1.2%. As an independent check, ℓ_1-IRLS on the uncor-

rupted graph reproduces this reference within 0.01° on INTEL, CSAIL, FR079 and M3500. We report the median, over poses, of the angular error after gauge alignment, averaged over seeds.

Baselines. Spectral, CEMP, CircReg and GNC use the same algorithms as in Sec. 4; we only replace dense eigendecompositions and linear solves with sparse ones (`eigsh`, `spsolve`), without which the reference implementations already need 51–104 s at $N{=}1228$. We do not run CEMP on City10000, where its implementation is too slow.

Results. Table 5 extends Table 2 to every corruption level. On the clean graphs ($f{=}0$), ℓ_1-IRLS matches the reference (computed by our solver) to within 0.01°, except on MITb (13.5°), while Spectral, CEMP, CircReg and GNC already err by up to 73°. ℓ_1-IRLS degrades from $f{=}0.25$ on, whereas our error stays near one degree up to $f{=}0.5$ on INTEL, CSAIL, FR079 and City10000; at $f{\geq}0.75$, most methods fail, and the drift-limited Odometry baseline is often the best. Table 6 shows the inlier sets at $f{=}0.5$: our estimate keeps precision and recall of at least 0.96 on every graph, ℓ_1-IRLS keeps a recall of 0.85–0.97, and Spectral, CEMP, CircReg and GNC reach a recall of only 0.11–0.83. Table 7 splits our runtime by stage: Phase 1 takes over 90% of it on every graph, whereas graph-cut refinement and IRLS together take below 0.7 s even at $N{=}10^4$.

Table 5: **2D PGO, full results.** Median heading error (°) vs. the fraction f of corrupted loop closures, and mean runtime (s) at $f{=}0.5$. Best error per column in bold; the reference comes from our solver, so Ours at $f{=}0$ is exact by design. City10000 is evaluated at $f \in \{0.1, 0.25, 0.5\}$ only.

Method	$f{=}0$	$f{=}0.1$	$f{=}0.25$	$f{=}0.5$	$f{=}0.75$	$f{=}0.9$	time (s)
INTEL ($N{=}1{,}228$, $\lvert\mathcal{E}\rvert{=}1{,}483$, 256 loop closures)							
Ours	**0.00**	0.47	**0.51**	**0.84**	**5.05**	87.0	5.2
ℓ_1-IRLS [10]	0.01	**0.39**	2.54	41.3	80.1	80.9	0.1
Spectral [41]	1.65	20.1	74.8	85.5	86.4	82.5	6.2
CEMP [30]	1.64	21.9	70.9	71.3	85.1	84.3	6.4
CircReg [33]	1.52	17.6	43.9	77.7	89.5	81.2	7.0
GNC [46]	0.59	9.95	46.1	60.6	85.1	85.1	6.6
Odometry	42.4	42.4	42.4	42.4	42.4	**42.4**	0.00
CSAIL ($N{=}1{,}045$, $\lvert\mathcal{E}\rvert{=}1{,}172$, 128 loop closures)							
Ours	**0.00**	**0.05**	**0.35**	**0.50**	**0.60**	3.76	2.9
ℓ_1-IRLS [10]	**0.00**	0.06	9.05	53.0	60.3	81.9	0.05
Spectral [41]	62.8	73.7	75.8	80.7	88.3	85.0	6.3
CEMP [30]	72.7	78.3	74.9	81.8	85.6	87.6	6.6
CircReg [33]	73.4	74.1	78.2	84.0	87.9	85.2	6.9
GNC [46]	35.7	43.3	62.1	84.4	81.1	87.6	6.5
Odometry	1.90	1.90	1.90	1.90	1.90	**1.90**	0.00
FR079 ($N{=}989$, $\lvert\mathcal{E}\rvert{=}1{,}217$, 229 loop closures)							
Ours	**0.00**	0.13	**0.43**	**1.06**	40.5	82.4	4.4
ℓ_1-IRLS [10]	**0.00**	**0.08**	3.65	19.4	42.1	58.9	0.1
Spectral [41]	60.0	58.4	83.0	88.6	86.4	87.4	6.0
CEMP [30]	66.4	67.7	65.8	76.7	86.9	87.7	6.0
CircReg [33]	44.8	58.9	72.0	75.8	88.6	92.2	6.2
GNC [46]	31.9	65.4	68.7	83.2	85.0	96.2	6.0
Odometry	1.90	1.90	1.90	1.90	**1.90**	**1.90**	0.00
MITb ($N{=}808$, $\lvert\mathcal{E}\rvert{=}827$, 20 loop closures)							
Ours	**0.00**	**9.54**	**12.9**	**33.8**	57.3	61.9	1.5
ℓ_1-IRLS [10]	13.5	25.0	32.5	47.8	**53.9**	**49.1**	0.02
Spectral [41]	47.0	57.7	53.7	85.6	94.3	88.4	8.4
CEMP [30]	55.0	57.7	60.8	75.7	91.3	95.5	7.8
CircReg [33]	50.1	54.0	54.8	68.8	91.2	87.5	8.7
GNC [46]	38.7	65.4	42.3	91.0	94.0	77.0	8.3
Odometry	82.6	82.6	82.6	82.6	82.6	82.6	0.00
M3500 ($N{=}3{,}500$, $\lvert\mathcal{E}\rvert{=}5{,}453$, 1,954 loop closures)							
Ours	**0.00**	**0.62**	**1.59**	**4.25**	59.0	73.5	19.8
ℓ_1-IRLS [10]	**0.00**	0.97	23.8	81.1	87.1	87.2	0.5
Spectral [41]	14.1	48.4	77.5	82.2	90.5	89.4	11.5
CEMP [30]	10.2	39.1	76.6	80.0	90.2	87.5	12.0
CircReg [33]	10.1	30.5	64.8	89.9	82.5	89.9	13.3
GNC [46]	11.1	36.0	53.3	86.7	95.3	89.8	12.2
Odometry	20.2	20.2	20.2	20.2	**20.2**	**20.2**	0.00
City10000 ($N{=}10k$, $\lvert\mathcal{E}\rvert{=}64k$, 54k loop closures)							
Ours	–	**0.24**	**0.40**	**0.98**	–	–	237
ℓ_1-IRLS [10]	–	1.07	57.2	53.4	–	–	1.2
Spectral [41]	–	86.6	79.5	76.5	–	–	0.1
CircReg [33]	–	90.2	84.2	93.6	–	–	11.5
GNC [46]	–	67.8	91.4	98.9	–	–	10.4
Odometry	–	18.8	18.8	18.8	–	–	0.01

Table 6: **2D PGO, inlier sets** at $f=0.5$: precision/recall of the edges each method's estimate classifies as inliers ($\tau=5^\circ$), against the true corruption labels.

Method	INTEL	CSAIL	FR079	MITb	M3500	City10000
Ours	0.99/1.00	0.99/0.99	0.99/1.00	0.99/0.99	0.97/0.96	0.98/0.99
ℓ_1-IRLS [10]	0.98/0.94	0.98/0.96	0.97/0.94	0.99/0.97	0.90/0.85	0.96/0.93
Spectral [41]	0.97/0.62	0.95/0.17	0.95/0.22	0.98/0.37	0.87/0.18	0.82/0.11
CEMP [30]	0.96/0.61	0.96/0.17	0.95/0.21	0.98/0.37	0.88/0.18	–
CircReg [33]	0.96/0.83	0.96/0.63	0.95/0.64	0.99/0.81	0.87/0.69	0.90/0.67
GNC [46]	0.96/0.73	0.97/0.54	0.96/0.50	0.98/0.65	0.87/0.36	0.85/0.34
Odometry	1.00/0.92	1.00/0.95	1.00/1.00	1.00/0.99	0.99/0.88	0.96/0.46

Table 7: **2D PGO, our per-stage runtime** (s, mean at $f=0.5$). Phase 1 computes the initial labeling and bounds before BnB (multi-start ICM, Sec. 3.3); classification takes <1 ms.

Dataset	N	$\lvert\mathcal{E}\rvert$	Phase 1	graph-cut	IRLS	total
INTEL	1,228	1,483	4.86	0.28	0.03	5.17
CSAIL	1,045	1,172	3.08	0.27	0.02	3.37
FR079	989	1,217	4.04	0.29	0.02	4.34
MITb	808	827	1.39	0.09	0.00	1.49
M3500	3,500	5,453	18.19	0.11	0.06	18.36
City10000	10k	64k	236.11	0.09	0.59	236.79

E IRLS Parity and Inlier Sets

Is our gain due to a better inlier set, or to the final IRLS polish? To answer this, we apply the identical circular-ℓ_1 plus Geman–McClure IRLS polish that our pipeline uses (Appendix B) to every baseline: we classify each edge by the residual of the method's own estimate (same threshold for all methods), and run IRLS on that inlier set, warm-started at the method's solution. Table 8 reports the median error before and after the polish and, in parentheses, the precision/recall of the inlier set the polish receives, measured against the true inlier labels. At $\eta=0.8$ the baselines recover only 7–49% of the true inliers, so the polish cannot rescue them, while our method without any IRLS is already at $0.75^\circ/1.23^\circ$ (dense/sparse) because its inlier set is right. These runs use 20 seeds per cell, so our numbers differ slightly from Fig. 2, which averages 50.

Table 8: **IRLS parity.** Median error ($^\circ$) raw $\rightarrow$ after the identical IRLS polish, with the inlier precision/recall of each method's own estimate in parentheses. $N=50$, 20 seeds.

Method	$\eta=0.4$	$\eta=0.6$	$\eta=0.8$
Fully connected			
Spectral [41]	4.19 $\rightarrow$ 0.28 (0.96/0.43)	7.70 $\rightarrow$ 0.43 (0.84/0.24)	23.50 $\rightarrow$ 22.47 (0.36/0.09)
CEMP [30]	2.43 $\rightarrow$ 0.28 (0.97/0.66)	6.38 $\rightarrow$ 0.38 (0.86/0.30)	21.32 $\rightarrow$ 20.51 (0.41/0.10)
CircReg [33]	0.34 $\rightarrow$ 0.27 (0.98/0.97)	0.43 $\rightarrow$ 0.33 (0.96/0.95)	7.18 $\rightarrow$ 7.32 (0.61/0.49)
GNC [46]	0.26 $\rightarrow$ 0.27 (0.98/0.96)	0.55 $\rightarrow$ 0.42 (0.92/0.70)	22.85 $\rightarrow$ 20.25 (0.36/0.15)
ℓ_1-IRLS [10]	0.27 $\rightarrow$ 0.27 (0.98/0.99)	0.38 $\rightarrow$ 0.38 (0.95/0.94)	36.76 $\rightarrow$ 37.01 (0.36/0.22)
Ours w/o IRLS	0.42 (0.98/0.98)	0.48 (0.96/0.98)	0.75 (0.87/0.96)
Ours (full)	0.28 (0.98/0.99)	0.33 (0.96/0.99)	0.53 (0.88/0.97)
Sparse ($\sim$32 edges/node)			
Spectral [41]	4.90 $\rightarrow$ 0.54 (0.95/0.37)	10.57 $\rightarrow$ 1.78 (0.80/0.19)	37.96 $\rightarrow$ 37.47 (0.29/0.07)
CEMP [30]	3.21 $\rightarrow$ 0.35 (0.96/0.53)	9.04 $\rightarrow$ 1.13 (0.81/0.22)	41.65 $\rightarrow$ 46.81 (0.30/0.08)
CircReg [33]	0.45 $\rightarrow$ 0.34 (0.98/0.96)	0.60 $\rightarrow$ 0.45 (0.95/0.93)	32.74 $\rightarrow$ 32.95 (0.35/0.26)
GNC [46]	0.39 $\rightarrow$ 0.38 (0.98/0.90)	2.94 $\rightarrow$ 1.86 (0.86/0.54)	40.03 $\rightarrow$ 42.25 (0.28/0.14)
ℓ_1-IRLS [10]	0.34 $\rightarrow$ 0.34 (0.98/0.98)	0.52 $\rightarrow$ 0.52 (0.93/0.88)	51.65 $\rightarrow$ 51.55 (0.27/0.19)
Ours w/o IRLS	0.49 (0.98/0.98)	0.63 (0.95/0.98)	1.23 (0.78/0.89)
Ours (full)	0.34 (0.98/0.98)	0.43 (0.95/0.98)	0.85 (0.79/0.89)

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Section 1 state three contributions: (i) a branch-and-bound algorithm with ICM-guided label pruning and a node-constrained upper bound for the smoothed max-consensus MRF on S^1 (Sec. 3.3); (ii) formal guarantees on discretization (Prop. 1), discrete–continuous equivalence under a separation condition (Thm. 2), pruning safety (Prop. 3), and bound admissibility (Thm. 4); and (iii) an exact-when-supermodular binary graph-cut refinement that maps the certified bin labeling to sub-bin angles (Sec. 3.4). The evaluation supporting these claims is in Sec. 4 (synthetic graphs and four application domains, including 2D pose-graph optimization), with extended results, a runtime-vs-N scaling sweep, full pose-graph results and an IRLS-parity study in Appendices C–E. Quantitative claims (search-space reduction, sub-degree accuracy) are qualified as holding on the tested instances.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The Conclusion's *Limitations* paragraph and Sec. 4.1 jointly discuss three limits: (i) the certified inlier set matches the smoothed-MRF optimum, with recovery of the original max-consensus optimum additionally requiring the separation condition Eq. (3); (ii) the sub-bin graph-cut refinement that warm-starts IRLS is non-exact on the $\sim 6\%$ of edges that violate supermodularity (the inlier set itself is unaffected); and (iii) above $\sim 90\%$ outliers no tested method recovers the angles. Appendix C.5 characterizes the runtime envelope as a function of N, and Tab. 3 states where TLS-SDP was not run. The Limitations paragraph also states the scaling envelope: dense graphs beyond $N \approx 10^3$ become memory- and BnB-bound, while sparse pose graphs scale to at least $N{=}10^4$ (Sec. 4.3).

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Every theorem and proposition states its hypotheses in place, and every proof is given in full in Appendix A.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results?

   Answer: [Yes]

   Justification: The full algorithm is described in Sec. 3 with pseudo-code in Appendix B. Hyperparameters (K, τ, σ, r, pruning ρ) are stated in Secs. 3–4, and the remaining solver defaults (restarts, BnB budget, refinement levels, quadrature, final polish) in Appendix B. Per-application protocols (feature extractor, two-view geometry, gravity source, outlier-injection rate, seed counts) are documented in Appendix C for image SfM (GLOMAP+SIFT on IMC 2023/2024 and KITTI), multi-way LiDAR (Predator+RANSAC on KITTI $s{=}5, 10$), cryo-EM (synthetic projections of EMD-3295 with EMPIAR-10166 CTFs and SNR$=$1), and 2D pose-graph optimization (corruption protocol, reference construction and baseline implementations in Appendix D). Datasets are publicly available; code is released at `https://github.com/danini/optimal-angular-synchronization`.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results?

Answer: [Yes]

Justification: Code and data-generation scripts are released at `https://github.com/danini/optimal-angular-synchronization`. All datasets used are publicly available (IMC 2023/2024, KITTI, EMPIAR-10166, EMD-3295, and the planar pose graphs of [9]).

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details necessary to understand the results?

   Answer: [Yes]

   Justification: Secs. 3–4 specify graph sizes, noise models, outlier injection protocol, and solver hyperparameters (including the discretization K, pruning threshold $\rho = 0.3$). The work does not involve learning, so there are no training details.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: We report means over seeds and scenes without error bars; the released scripts regenerate the per-seed results. Synthetic results (Fig. 2) are mean-of-per-seed-medians over 50 random seeds per cell. Real-world tables aggregate per-scene metrics across all scenes in each dataset (Haiper n=3, Phototourism n=15, IMC 2024 n=4, KITTI SfM n=11, KITTI s=5, 10 n=11, EMPIAR-10166 n=10 seeds for the main-paper baselines, n=5 for TLS-SDP). The runtime sweep in Appendix C.5 averages each cell over 5 seeds; the IRLS-parity study (App. E) over 20 seeds; the pose-graph benchmark (Sec. 4.3) over 5 seeds per corruption level. The EMPIAR-10166 median column is the median of per-seed medians; all other aggregates are means.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Our solver and all baselines run on a single workstation CPU (Intel Core i9-10900K, 10 cores); Predator feature extraction for the LiDAR experiments uses one NVIDIA RTX 4090 GPU; BnB is OpenMP-parallelized across top-level subtrees (Sec. 3.3). Per-dataset wall-clock runtimes are reported in Table 1 (column "time") and Table 3; synthetic BnB tree size and certified-runtime are discussed in Sec. 4.1 and the full $N \in \{20, 50, 100, 200\}$ scaling sweep is in Appendix C.5 (Tab. 4). Pose-graph runtimes up to N=10^4 are in Table 2, with a per-stage split in Appendix D. The TLS-SDP baseline uses cvxpy+SCS and is intractable past $N \approx 200$.

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics?

   Answer: [Yes]

   Justification: The work uses only public datasets (photo-tourism imagery, LiDAR scans, anonymized cryo-EM particle stacks) and involves no human subjects.

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The work is a foundational algorithmic contribution to geometric estimation with no direct societal deployment; downstream applications (SfM, SLAM, cryo-EM reconstruction) are well-established and their societal considerations are orthogonal to this contribution.

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse?

Answer: [N/A]

Justification: No data or models with misuse risk are released.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets used in the paper properly credited, and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All cited datasets (IMC 2023/2024, KITTI, EMPIAR-10166, EMD-3295, and the planar pose graphs of [9, 31, 26]) and software dependencies (GLOMAP [32], ASPIRE [13]) are credited with original references. KITTI is released under CC BY-NC-SA 3.0, EMPIAR and EMDB entries under CC0, and the IMC data under the Kaggle competition terms; GLOMAP is BSD-3-Clause and ASPIRE GPL-3.0; the pose graphs are public research datasets. All permit academic use.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The new asset is a C++ solver ("pyoptrot1d") with Python bindings. It is released at `https://github.com/danini/optimal-angular-synchronization` with a README, reproduction scripts for every table and figure in this paper, and parameter documentation. The license is BSD-3-Clause, which permits academic and commercial reuse.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: No crowdsourcing or human-subject experiments were conducted.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: No human-subject research.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research?

Answer: [N/A]

Justification: The proposed method (BnB on a smoothed max-consensus MRF) contains no LLM component. LLM-based assistance was used only for routine writing polish and code editing, which the NeurIPS 2026 guidelines do not require to be declared on this checklist.