Robust Fitting in Computer Vision: Easy or Hard?

Tat-Jun Chin, Zhipeng Cai, and Frank Neumann

School of Computer Science, The University of Adelaide

Abstract. Robust model fitting plays a vital role in computer vision, and research into algorithms for robust fitting continues to be active. Arguably the most popular paradigm for robust fitting in computer vision is *consensus maximisation*, which strives to find the model parameters that maximise the number of inliers. Despite the significant developments in algorithms for consensus maximisation, there has been a lack of fundamental analysis of the problem in the computer vision literature. In particular, whether consensus maximisation is "tractable" remains a question that has not been rigorously dealt with, thus making it difficult to assess and compare the performance of proposed algorithms, relative to what is theoretically achievable. To shed light on these issues, we present several computational hardness results for consensus maximisation. Our results underline the fundamental intractability of the problem, and resolve several ambiguities existing in the literature.

Keywords: Robust fitting, consensus maximisation, inlier set maximisation, computational hardness.

1 Introduction

Robustly fitting a geometric model onto noisy and outlier-contaminated data is a necessary capability in computer vision [1], due to the imperfectness of data acquisition systems and preprocessing algorithms (e.g., edge detection, keypoint detection and matching). Without robustness against outliers, the estimated geometric model will be biased, leading to failure in the overall pipeline.

In computer vision, robust fitting is typically performed under the framework of *inlier set maximisation*, a.k.a. *consensus maximisation* [2], where one seeks the model with the most number of inliers. For concreteness, say we wish to estimate the parameter vector $\mathbf{x} \in \mathbb{R}^d$ that defines the linear relationship $\mathbf{a}^T \mathbf{x} = b$ from a set of outlier-contaminated measurements $\mathcal{D} = \{(\mathbf{a}_i, b_i)\}_{i=1}^N$. The consensus maximisation formulation for this problem is as follows.

Problem 1 (MAXCON). Given input data $\mathcal{D} = \{(\mathbf{a}_i, b_i)\}_{i=1}^N$, where $\mathbf{a}_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$, and an inlier threshold $\epsilon \in \mathbb{R}_+$, find the $\mathbf{x} \in \mathbb{R}^d$ that maximises

$$\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}) = \sum_{i=1}^N \mathbb{I}\left(|\mathbf{a}_i^T \mathbf{x} - b_i| \le \epsilon\right), \tag{1}$$

where $\mathbb{I}$ returns 1 if its input predicate is true, and 0 otherwise.

The quantity $|\mathbf{a}_i^T \mathbf{x} - b_i|$ is the *residual* of the i-th measurement with respect to $\mathbf{x}$, and the value given by $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D})$ is the *consensus* of $\mathbf{x}$ with respect to $\mathcal{D}$. Intuitively, the consensus of $\mathbf{x}$ is the number of inliers of $\mathbf{x}$. For the robust estimate to fit the inlier structure well, the inlier threshold ϵ must be set to an appropriate value; the large number of applications that employ the consensus maximisation framework indicate that this is usually not an obstacle.

Developing algorithms for robust fitting, specifically for consensus maximisation, is an active research area in computer vision. Currently, the most popular algorithms belong to the class of randomised sampling techniques, i.e., RANSAC [2] and its variants [3, 4]. Unfortunately, such techniques do not provide certainty of finding satisfactory solutions, let alone optimal ones [5].

Increasingly, attention is given to constructing *globally optimal* algorithms for robust fitting, e.g., [6–14]. Such algorithms are able to deterministically calculate the best possible solution, i.e., the model with the highest achievable consensus. This mathematical guarantee is regarded as desirable, especially in comparison to the "rough" solutions provided by random sampling heuristics.

Recent progress in globally optimal algorithms for consensus maximisation seems to suggest that global solutions can be obtained efficiently or tractably [6–14]. Moreover, decent empirical performances have been reported. This raises hopes that good alternatives to the random sampling methods are now available. However, to what extent is the problem solved? Can we expect the global algorithms to perform well in general? Are there fundamental obstacles toward efficient robust fitting algorithms? What do we even mean by "efficient"?

1.1   Our contributions and their implications

Our contributions are *theoretical*. We resolve the above ambiguities in the literature, by proving the following computational hardness results. The implications of each result are also listed below.

> MAXCON is NP-hard (Section 2).

$\implies$ There are no algorithms that can solve MAXCON in time polynomial to the input size, which is proportional to N and d.

> MAXCON is W[1]-hard in the dimension d (Section 3.2).

$\implies$ There are no algorithms that can solve MAXCON in time $f(d)\mathrm{poly}(N)$, where $f(d)$ is an arbitrary function of d, and $\mathrm{poly}(N)$ is a polynomial of N.

> MAXCON is APX-hard (Section 4).

$\implies$ There are no polynomial time algorithms that can approximate MAXCON up to $(1-\delta)\psi^*$ for any known factor δ, where ψ^* is the maximum consensus.

As usual, the implications of the hardness results are subject to the standard complexity assumptions P$\neq$NP [15] and FPT$\neq$W[1]-hard [16].

Our analysis indicates the "extreme" difficulty of consensus maximisation. MAXCON is not only *intractable* (by standard notions of intractability [15, 16]), the W[1]-hardness result also suggests that any global algorithm will scale exponentially in a function of d, i.e., $N^{f(d)}$. In fact, if a conjecture of Erickson et al. [17] holds, MAXCON cannot be solved faster than N^d. Thus, the decent performances in [6–14] are unlikely to extend to the general cases in practical settings, where $N \geq 1000$ and $d \geq 6$ are common. More pessimistically, APX-hardness shows that MAXCON is impossible to approximate, in that there are no polynomial time approximation schemes (PTAS) [18] for MAXCON[1].

A slightly positive result is as follows.

> MAXCON is FPT (fixed parameter tractable) in the number of outliers o and dimension d (Section 3.3).

This is achieved by applying a special case of the algorithm of Chin et al. [13] on MAXCON to yield a runtime of $\mathcal{O}(d^o)\mathrm{poly}(N, d)$. However, this still scales exponentially in o, which can be large in practice (e.g., $o \geq 100$).

1.2   How are our theoretical results useful?

First, our results clarify the ambiguities on the efficiency and solvability of consensus maximisation alluded to above. Second, our analysis shows how the effort scales with the different input size parameters, thus suggesting more cogent ways for researchers to test/compare algorithms. Third, since developing algorithms for consensus maximisation is an active topic, our hardness results encourage researchers to consider alternative paradigms of optimisation, e.g., deterministically convergent heuristic algorithms [19–21] or preprocessing techniques [22–24].

1.3   What about non-linear models?

Our results are based specifically on MAXCON, which is concerned with fitting linear models. In practice, computer vision applications require the fitting of non-linear geometric models (e.g., fundamental matrix, homography, rotation). While a case-by-case treatment is ideal, it is unlikely that non-linear consensus maximisation will be easier than linear consensus maximisation [25–27].

1.4   Why not employ other robust statistical procedures?

Our purpose here is not to benchmark or advocate certain robust criteria. Rather, our primary aim is to establish the fundamental difficulty of consensus maximisation, which is widely used in computer vision. Second, it is unlikely that other robust criteria are easier to solve [28]. Although some that use differentiable robust loss functions (e.g., M-estimators) can be solved up to local optimality, it is unknown how far the local optima deviate from the global solution.

The rest of the paper is devoted to developing the above hardness results.

---

[1] Since RANSAC does not provide any approximation guarantees, it is not an "approximation scheme" by standard definition [18].

2   NP-hardness

The decision version of MAXCON is as follows.

Problem 2 (MAXCON-D). Given data $\mathcal{D} = \{(\mathbf{a}_i, b_i)\}_{i=1}^{N}$, an inlier threshold $\epsilon \in \mathbb{R}_+$, and a number $\psi \in \mathbb{N}_+$, does there exist $\mathbf{x} \in \mathbb{R}^d$ such that $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}) \geq \psi$?

Another well-known robust fitting paradigm is least median squares (LMS), where we seek the vector $\mathbf{x}$ that minimises the median of the residuals

$$\min_{\mathbf{x} \in \mathbb{R}^d} \quad \mathrm{med}\left(|\mathbf{a}_1^T \mathbf{x} - b_1|, \ldots, |\mathbf{a}_N^T \mathbf{x} - b_N|\right). \tag{2}$$

LMS can be generalised by minimising the k-th largest residual instead

$$\min_{\mathbf{x} \in \mathbb{R}^d} \quad \mathrm{kos}\left(|\mathbf{a}_1^T \mathbf{x} - b_1|, \ldots, |\mathbf{a}_N^T \mathbf{x} - b_N|\right), \tag{3}$$

where function kos returns its k-th largest input value.

Geometrically, LMS seeks the *slab* of the *smallest width* that contains *half* of the data points $\mathcal{D}$ in $\mathbb{R}^{d+1}$. A slab in $\mathbb{R}^{d+1}$ is defined by a normal vector $\mathbf{x}$ and width w as

$$h_w(\mathbf{x}) = \left\{(\mathbf{a}, b) \in \mathbb{R}^{d+1} \ \middle| \ |\mathbf{a}^T \mathbf{x} - b| \leq \frac{1}{2}w\right\}. \tag{4}$$

Problem (3) thus seeks the thinnest slab that contains k of the points. The decision version of (3) is as follows.

Problem 3 (k-SLAB). Given data $\mathcal{D} = \{(\mathbf{a}_i, b_i)\}_{i=1}^{N}$, an integer k where $1 \leq k \leq N$, and a number $w' \in \mathbb{R}_+$, does there exist $\mathbf{x} \in \mathbb{R}^d$ such that k of the members of $\mathcal{D}$ are contained in a slab $h_w(\mathbf{x})$ of width at most w'?

k-SLAB has been proven to be NP-complete in [17].

Theorem 1. *MAXCON-D is NP-complete.*

Proof. Let $\mathcal{D}$, k and w' define an instance of k-SLAB. This can be reduced to an instance of MAXCON-D by simply reusing the same $\mathcal{D}$, and setting $\epsilon = \frac{1}{2}w'$ and $\psi = k$. If the answer to k-SLAB is positive, then there is an $\mathbf{x}$ such that k points from $\mathcal{D}$ lie within vertical distance of $\frac{1}{2}w'$ from the hyperplane defined by $\mathbf{x}$, hence $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D})$ must be at least ψ and the answer to MAXCON-D is also positive. Conversely, if the answer to MAXCON-D is positive, then there is an $\mathbf{x}$ such that ψ points have vertical distance of less than ϵ to $\mathbf{x}$, hence a slab that is centred at $\mathbf{x}$ of width at most w' can enclose k of the points, and the answer to k-SLAB is also positive. □

The NP-completeness of MAXCON-D implies the NP-hardness of the optimisation version MAXCON. See Sec. 1.1 for the implications of NP-hardness.

3   Parametrised complexity

Parametrised complexity is a branch of algorithmics that investigates the inherent difficulty of problems with respect to structural parameters in the input [16]. In this section, we report several parametrised complexity results of MAXCON.

First, the *consensus set* $\mathcal{C}_\epsilon(\mathbf{x} \mid \mathcal{D})$ of $\mathbf{x}$ is defined as

$$\mathcal{C}_\epsilon(\mathbf{x} \mid \mathcal{D}) := \{i \in \{1, \ldots, N\} \mid |\mathbf{a}_i^T \mathbf{x} - b_i| \leq \epsilon\}. \tag{5}$$

An equivalent definition of consensus (1) is thus

$$\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}) = |\mathcal{C}_\epsilon(\mathbf{x} \mid \mathcal{D})|. \tag{6}$$

Henceforth, we do not distinguish between the integer subset $\mathcal{C} \subseteq \{1, \ldots, N\}$ that indexes a subset of $\mathcal{D}$, and the actual data that are indexed by $\mathcal{C}$.

3.1   XP in the dimension

The following is the *Chebyshev approximation* problem [29, Chapter 2] defined on the input data indexed by $\mathcal{C}$:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \quad \max_{i \in \mathcal{C}} \; |\mathbf{a}_i^T \mathbf{x} - b_i| \tag{7}$$

Problem (7) has the linear programming (LP) formulation

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^d, \gamma \in \mathbb{R}} \quad & \gamma \\ \text{s.t.} \quad & |\mathbf{a}_i^T \mathbf{x} - b_i| \leq \gamma, \; i \in \mathcal{C}, \end{aligned} \tag{LP[$\mathcal{C}$]}$$

which can be solved in polynomial time. Chebyshev approximation also has the following property.

Lemma 1. *There is a subset $\mathcal{B}$ of $\mathcal{C}$, where $|\mathcal{B}| \leq d + 1$, such that*

$$\min_{\mathbf{x} \in \mathbb{R}^d} \quad \max_{i \in \mathcal{B}} \; r_i(\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^d} \quad \max_{i \in \mathcal{C}} \; r_i(\mathbf{x}) \tag{8}$$

Proof. See [29, Section 2.3]. □

We call $\mathcal{B}$ a *basis* of $\mathcal{C}$. Mathematically, $\mathcal{B}$ is the set of active constraints to LP[$\mathcal{C}$], hence bases can be computed easily. In fact, LP[$\mathcal{B}$] and LP[$\mathcal{C}$] have the same minimisers. Further, for any subset $\mathcal{B}$ of size $d + 1$, a method by de la Vallée-Poussin can solve LP[$\mathcal{B}$] analytically in time polynomial to d; see [29, Chapter 2] for details.

Let $\mathbf{x}$ be an arbitrary candidate solution to MAXCON, and $(\hat{\mathbf{x}}, \hat{\gamma})$ be the minimisers to LP[$\mathcal{C}_\epsilon(\mathbf{x} \mid \mathcal{D})$], i.e., the Chebyshev approximation problem on the consensus set of $\mathbf{x}$. The following property can be established.

Lemma 2. $\Psi_\epsilon(\hat{\mathbf{x}} \mid \mathcal{D}) \geq \Psi_\epsilon(\mathbf{x} \mid \mathcal{D})$.

Proof. By construction, $\hat{\gamma} \leq \epsilon$. Hence, if $(\mathbf{a}_i, b_i)$ is an inlier to $\mathbf{x}$, i.e., $|\mathbf{a}_i^T \mathbf{x} - b_i| \leq \epsilon$, then $|\mathbf{a}_i^T \hat{\mathbf{x}} - b_i| \leq \hat{\gamma} \leq \epsilon$, i.e., $(\mathbf{a}_i, b_i)$ is also an inlier to $\hat{\mathbf{x}}$. Thus, the consensus of $\hat{\mathbf{x}}$ is no smaller than the consensus of $\mathbf{x}$. □

Lemmas 1 and 2 suggest a rudimentary algorithm for consensus maximisation that attempts to find the basis of the maximum consensus set, as encapsulated in the proof of the following theorem.

Theorem 2. *MAXCON is XP (slice-wise polynomial) in the dimension d.*

Proof. Let $\mathbf{x}^*$ be a witness to an instance of MAXCON-D with positive answer, i.e., $\Psi_\epsilon(\mathbf{x}^* \mid \mathcal{D}) \geq \psi$. Let $(\hat{\mathbf{x}}^*, \hat{\gamma}^*)$ be the minimisers to $\text{LP}[\mathcal{C}_\epsilon(\mathbf{x}^* \mid \mathcal{D})]$. By Lemma 2, $\hat{\mathbf{x}}^*$ is also a positive witness to the instance. By Lemma 1, $\hat{\mathbf{x}}^*$ can be found by enumerating all $(d+1)$-subsets of $\mathcal{D}$, and solving Chebyshev approximation (7) on each $(d+1)$-subset. There are a total of $\binom{N}{d+1}$ subsets to check; including the time to evaluate $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D})$ for each candidate, the runtime of this simple algorithm is $\mathcal{O}(N^{d+2}\text{poly}(d))$, which is polynomial in N for a fixed d. □

Theorem 2 shows that for a fixed dimension d, MAXCON can be solved in time polynomial in the number of measurements N (this is consistent with the results in [8, 12]). However, this does not imply that MAXCON is tractable (following the standard meaning of tractability in complexity theory [15, 16]). Moreover, in practical applications, d could be large (e.g., $d \geq 5$), thus the rudimentary algorithm above will not be efficient for large N.

3.2   W[1]-hard in the dimension

Can we remove d from the exponent of the runtime of a globally optimal algorithm? By establishing W[1]-hardness in the dimension, this section shows that it is not possible. Our proofs are inspired by, but extends quite significantly from, that of [30, Section 5]. First, the source problem is as follows.

Problem 4 (k-CLIQUE). Given undirected graph $G = (V, E)$ with vertex set V and edge set E and a parameter $k \in \mathbb{N}_+$, does there exist a clique in G with k vertices?

k-CLIQUE is W[1]-hard w.r.t. parameter k [31]. Here, we demonstrate an FPT reduction from k-CLIQUE to MAXCON-D with fixed dimension d.

Generating the input data Given input graph $G = (V, E)$, where $V = \{1, \ldots, M\}$, and size k, we construct a $(k+1)$-dimensional point set $\mathcal{D}_G = \{(\mathbf{a}_i, b_i)\}_{i=1}^N = \mathcal{D}_V \cup \mathcal{D}_E$ as follows:

− The set $\mathcal{D}_V$ is defined as

$$\mathcal{D}_V = \{(\mathbf{a}_\alpha^v, b_\alpha^v)\}_{\alpha=1,\ldots,k}^{v=1,\ldots,M}, \tag{9}$$

where

$$\mathbf{a}_\alpha^v = [0, \ldots, 0, 1, 0, \ldots, 0]^T \tag{10}$$

is a k-dimensional vector of 0's except at the α-th element where the value is 1, and

$$b_\alpha^v = v. \tag{11}$$

– The set $\mathcal{D}_E$ is defined as

$$
\begin{aligned}
\mathcal{D}_E = \{(\mathbf{a}_{\alpha,\beta}^{u,v}, b_{\alpha,\beta}^{u,v}) \mid \quad & u, v = 1, \ldots, M, \\
& \langle u, v \rangle \in E, \langle v, u \rangle \in E, \\
& \alpha, \beta = 1, \ldots, k, \\
& \alpha < \beta \},
\end{aligned}
\tag{12}
$$

where

$$\mathbf{a}_{\alpha,\beta}^{u,v} = [0, \ldots, 0, 1, 0, \ldots, 0, M, 0, \ldots, 0]^T \tag{13}$$

is a k-dimensional vector of 0's, except at the α-th element where the value is 1 and the β-th element where the value is M, and

$$b_{\alpha,\beta}^{u,v} = u + Mv. \tag{14}$$

The size N of $\mathcal{D}_G$ is thus $|\mathcal{D}_V| + |\mathcal{D}_E| = kM + 2|E|\binom{k}{2}$.

Setting the inlier threshold Under our reduction, $\mathbf{x} \in \mathbb{R}^d$ is responsible for "selecting" a subset of the vertices V and edges E of G. First, we say that $\mathbf{x}$ selects vertex v if a point $(\mathbf{a}_\alpha^v, b_\alpha^v) \in \mathcal{D}_V$, for some α, is an inlier to $\mathbf{x}$, i.e., if

$$|(\mathbf{a}_\alpha^v)^T \mathbf{x} - b_\alpha^v| \le \epsilon \equiv x_\alpha \in [v - \epsilon, v + \epsilon], \tag{15}$$

where x_α is the α-th element of $\mathbf{x}$. The key question is how to set the value of the inlier threshold ϵ, such that $\mathbf{x}$ selects no more than k vertices, or equivalently, such that $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_V) \le k$ for all $\mathbf{x}$.
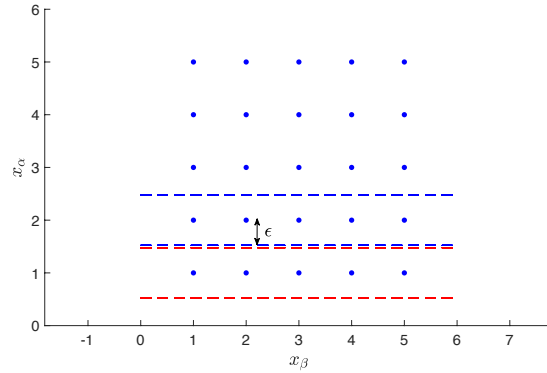
Lemma 3. *If $\epsilon < \frac{1}{2}$, then $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_V) \le k$, with equality achieved if and only if $\mathbf{x}$ selects k vertices of G.*

Proof. For any u and v, the ranges $[u - \epsilon, u + \epsilon]$ and $[v - \epsilon, v + \epsilon]$ cannot overlap if $\epsilon < \frac{1}{2}$. Hence, x_α lies in at most one of the ranges, i.e., each element of $\mathbf{x}$ selects at most one of the vertices; see Fig. 1. This implies that $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_V) \le k$. $\square$

Second, a point $(\mathbf{a}_{\alpha,\beta}^{u,v}, b_{\alpha,\beta}^{u,v})$ from $\mathcal{D}_E$ is an inlier to $\mathbf{x}$ if

$$|(\mathbf{a}_{\alpha,\beta}^{u,v})^T \mathbf{x} - b_{\alpha,\beta}^{u,v}| \le \epsilon \equiv |(x_\alpha - u) + M(x_\beta - v)| \le \epsilon. \tag{16}$$

As suggested by (16), the pairs of elements of $\mathbf{x}$ are responsible for selecting the edges of G. To prevent each element pair x_α, x_β from selecting more than one edge, or equivalently, to maintain $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_E) \le \binom{k}{2}$, the setting of ϵ is crucial.

Fig. 1. The blue dots indicate the integer values in the dimensions x_α and x_β. If $\epsilon < \frac{1}{2}$, then the ranges defined by (15) for all $v = 1, \ldots, M$ do not overlap. Hence, x_α can select at most one vertex of the graph.

Lemma 4. *If $\epsilon < \frac{1}{2}$, then $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_E) \leq \binom{k}{2}$, with equality achieved if and only if $\mathbf{x}$ selects $\binom{k}{2}$ edges of G.*

Proof. For each α, β pair, the constraint (16) is equivalent to the two linear inequalities

$$
\begin{aligned}
x_\alpha + Mx_\beta - u - Mv &\leq \epsilon, \\
x_\alpha + Mx_\beta - u - Mv &\geq -\epsilon,
\end{aligned}
\tag{17}
$$

which specify two opposing half-planes (i.e., a slab) in the space (x_α, x_β). Note that the slopes of the half-plane boundaries do not depend on u and v. For any two unique pairs (u_1, v_1) and (u_2, v_2), we have the four linear inequalities

$$
\begin{aligned}
x_\alpha + Mx_\beta - u_1 - Mv_1 &\leq \epsilon, \\
x_\alpha + Mx_\beta - u_1 - Mv_1 &\geq -\epsilon, \\
x_\alpha + Mx_\beta - u_2 - Mv_2 &\leq \epsilon, \\
x_\alpha + Mx_\beta - u_2 - Mv_2 &\geq -\epsilon.
\end{aligned}
\tag{18}
$$

The system (18) can be simplified to

$$
\begin{aligned}
\frac{1}{2}\left[u_2 - u_1 + M(v_2 - v_1)\right] &\leq \epsilon, \\
\frac{1}{2}\left[u_1 - u_2 + M(v_1 - v_2)\right] &\leq \epsilon.
\end{aligned}
\tag{19}
$$

Setting $\epsilon < \frac{1}{2}$ ensures that the two inequalities (19) cannot be consistent for all unique pairs (u_1, v_1) and (u_2, v_2). Geometrically, with $\epsilon < \frac{1}{2}$, the two slabs defined by (17) for different (u_1, v_1) and (u_2, v_2) pairs do not intersect; see Fig. 2 for an illustration.

Hence, if $\epsilon < \frac{1}{2}$, each element pair x_α, x_β of $\mathbf{x}$ can select at most one of the edges. Cumulatively, $\mathbf{x}$ can select at most $\binom{k}{2}$ edges, thus $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_E) \leq \binom{k}{2}$. □
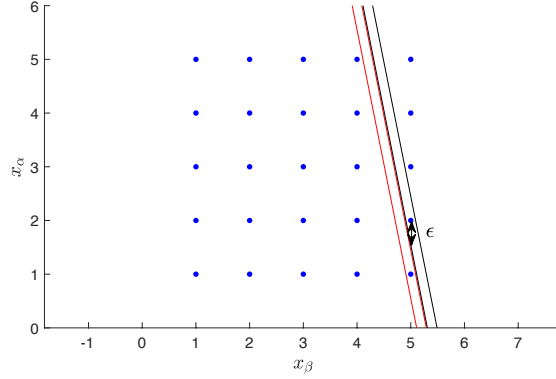
Fig. 2. The blue dots indicate the integer values in the dimensions x_α and x_β. If $\epsilon < \frac{1}{2}$, then any two slabs defined by (17) for different (u_1, v_1) and (u_2, v_2) pairs do not intersect. The figure shows two slabs corresponding to $u_1 = 1$, $v_1 = 5$, $u_2 = 2$, $v_2 = 5$.

Up to this stage, we have shown that if $\epsilon < \frac{1}{2}$, then $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_G) \leq k + \binom{k}{2}$, with equality achievable if there is a clique of size k in G. To establish the FPT reduction, we need to establish the reverse direction, i.e., if $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_G) = k + \binom{k}{2}$, then there is a k-clique in G. The following lemma shows that this can be assured by setting $\epsilon < \frac{1}{M+2}$.

Lemma 5. *If $\epsilon < \frac{1}{M+2}$, then $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_G) \leq k + \binom{k}{2}$, with equality achievable if and only if there is a clique of size k in G.*

Proof. The 'only if' direction has already been proven. To prove the 'if' direction, we show that if $\epsilon < \frac{1}{M+2}$ and $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_G) = k + \binom{k}{2}$, the subgraph $S(\mathbf{x}) = \{\lfloor x_1 \rceil, ..., \lfloor x_k \rceil\}$ is a k-clique, where each $\lfloor x_\alpha \rceil$ represents a vertex index in G. Since $\epsilon < \frac{1}{2}$, $\lfloor x_\alpha \rceil = u$ if and only if $(\mathbf{a}_\alpha^u, b_\alpha^u)$ is an inlier. Therefore, $S(\mathbf{x})$ consists of all vertices selected by $\mathbf{x}$. From Lemma 3 and Lemma 4, when $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_G) = k + \binom{k}{2}$, $\mathbf{x}$ is consistent with k points in $\mathcal{D}_V$ and $\binom{k}{2}$ points in $\mathcal{D}_E$. The inliers in $\mathcal{D}_V$ specifies the k vertices in $S(\mathbf{x})$. The 'if' direction is true if all selected $\binom{k}{2}$ edges are only edges in $S(\mathbf{x})$, i.e., for each inlier point $(\mathbf{a}_{\alpha,\beta}^{u,v}, b_{\alpha,\beta}^{u,v}) \in \mathcal{D}_E$, $(\mathbf{a}_\alpha^u, b_\alpha^u)$ and $(\mathbf{a}_\beta^v, b_\beta^v)$ are also inliers w.r.t. $\mathbf{x}$. The prove is done by contradiction:

If $\epsilon < \frac{1}{M+2}$, given an inlier $(\mathbf{a}_{\alpha,\beta}^{u,v}, b_{\alpha,\beta}^{u,v})$, from (16) we have:

$$|(x_\alpha - u) + M(x_\beta - v)| =$$

$$|[(\lfloor x_\alpha \rceil - u) + M(\lfloor x_\beta \rceil - v)] + [(x_\alpha - \lfloor x_\alpha \rceil) + M(x_\beta - \lfloor x_\beta \rceil)]| < \frac{1}{M+2}. \quad (20)$$

Assume at least one of $(\mathbf{a}_\alpha^u, b_\alpha^u)$ and $(\mathbf{a}_\beta^v, b_\beta^v)$ is not an inlier, from (15) and $\epsilon < \frac{1}{M+2}$, we have $\lfloor x_\alpha \rceil \neq u$ or $\lfloor x_\beta \rceil \neq v$, which means that at least one of $(\lfloor x_\alpha \rceil - u)$ and $(\lfloor x_\beta \rceil - v)$ is not zero. Since all elements of $\mathbf{x}$ satisfy (15), both $(\lfloor x_\alpha \rceil - u)$ and $(\lfloor x_\beta \rceil - v)$ are integers between $[-(M-1), (M-1)]$. If only one of $(\lfloor x_\alpha \rceil - u)$
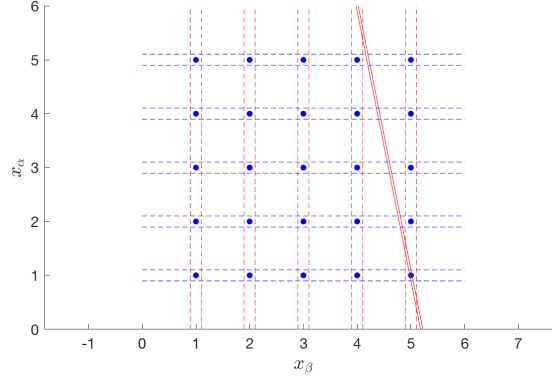
Fig. 3. If $\epsilon < \frac{1}{M+2}$, then the slab (17) that contains a point $(\mathbf{a}_{\alpha,\beta}^{u,v}, b_{\alpha,\beta}^{u,v}) \in \mathcal{D}_E$, where (u,v) is an edge in $\mathcal{G}$, does not intersect with any grid region besides the one formed by $(\mathbf{a}_\alpha^u, b_\alpha^u)$ and $(\mathbf{a}_\beta^v, b_\beta^v)$. In this figure, $u = 1$ and $v = 5$.

and $(\lfloor x_\beta \rceil - v)$ is not zero, then $|(\lfloor x_\alpha \rceil - u) + M(\lfloor x_\beta \rceil - v)| \geq |1 + M \cdot 0| = 1$. If both are not zero, then $|(\lfloor x_\alpha \rceil - u) + M(\lfloor x_\beta \rceil - v)| \geq |(M-1) + M \cdot 1| = 1$ Therefore, we have

$$|(\lfloor x_\alpha \rceil - u) + M(\lfloor x_\beta \rceil - v)| \geq 1. \tag{21}$$

Also due to (15), we have

$$|(x_\alpha - \lfloor x_\alpha \rceil) + M(x_\beta - \lfloor x_\beta \rceil)| \leq (M+1) \cdot \epsilon = \frac{M+1}{M+2}. \tag{22}$$

Combining (21) and (22), we have

$$
\begin{aligned}
&|[(\lfloor x_\alpha \rceil - u) + M(\lfloor x_\beta \rceil - v)] + [(x_\alpha - \lfloor x_\alpha \rceil) + M(x_\beta - \lfloor x_\beta \rceil)]| \geq \\
&|[(\lfloor x_\alpha \rceil - u) + M(\lfloor x_\beta \rceil - v)]| - |[(x_\alpha - \lfloor x_\alpha \rceil) + M(x_\beta - \lfloor x_\beta \rceil)]| \geq \\
&1 - \frac{M+1}{M+2} = \frac{1}{M+2},
\end{aligned}
\tag{23}
$$

which contradicts (20). It is obvious that $S(\mathbf{x})$ can be computed within linear time. Hence, the 'if' direction is true when $\epsilon < \frac{1}{M+2}$. $\qquad\square$

To illustrate Lemma 5, Fig. 3 depicts the value of $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_G)$ in the subspace (x_α, x_β) for $\epsilon < \frac{1}{M+2}$. Observe that $\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_G)$ attains the highest value of 3 in this subspace if and only if x_α and x_β select a pair of vertices that are connected by an edge in G.

Completing the reduction We have demonstrated a reduction from k-CLIQUE to MAXCON-D, where the main work is to generate data $\mathcal{D}_G$ which has number of measurements $N = k|V| + 2|E|\binom{k}{2}$ that is linear in $|G|$ and polynomial in k, and dimension $d = k$. In other words, the reduction is FPT in k. Setting $\epsilon < \frac{1}{M+2}$ and $\psi = k + \binom{k}{2}$ completes the reduction.

Theorem 3. *MAXCON is W[1]-hard w.r.t. the dimension d.*

Proof. Since k-CLIQUE is W[1]-hard w.r.t. k, by the above FPT reduction, MAXCON is W[1]-hard w.r.t. d. □

The implications of Theorem 3 have been discussed in Sec. 1.1.

3.3   FPT in the number of outliers and dimension

Let $f(\mathcal{C})$ and $\hat{\mathbf{x}}_{\mathcal{C}}$ respectively indicate the minimised objective value and minimiser of LP[$\mathcal{C}$]. Consider two subsets $\mathcal{P}$ and $\mathcal{Q}$ of $\mathcal{D}$, where $\mathcal{P} \subseteq \mathcal{Q}$. The statement

$$f(\mathcal{P}) \leq f(\mathcal{Q}) \qquad (24)$$

follows from the fact that LP[$\mathcal{P}$] contains only a subset of the constraints of LP[$\mathcal{Q}$]; we call this property *monotonicity*.

Let $\mathbf{x}^*$ be a global solution of an instance of MAXCON, and let $\mathcal{I}^* := \mathcal{C}_\epsilon(\mathbf{x}^* \mid \mathcal{D}) \subset \mathcal{D}$ be the maximum consensus set. Let $\mathcal{C}$ index a subset of $\mathcal{D}$, and let $\mathcal{B}$ be the basis of $\mathcal{C}$. If $f(\mathcal{C}) > \epsilon$, then by Lemma 1

$$f(\mathcal{D}) \geq f(\mathcal{C}) = f(\mathcal{B}) > \epsilon. \qquad (25)$$

The monotonicity property affords us further insight.

Lemma 6. *At least one point in $\mathcal{B}$ do not exist in $\mathcal{I}^*$.*

Proof. By monotonicity,

$$\epsilon < f(\mathcal{B}) \leq f(\mathcal{I}^* \cup \mathcal{B}). \qquad (26)$$

Hence, $\mathcal{I}^* \cup \mathcal{B}$ cannot be equal to $\mathcal{I}^*$, for if they were equal, then $f(\mathcal{I}^* \cup \mathcal{B}) = f(\mathcal{I}^*) \leq \epsilon$ which violates (26). □

The above observations suggest an algorithm for MAXCON that recursively removes basis points to find a consensus set, as summarised in Algorithm 1. This algorithm is a special case of the technique of Chin et al. [13]. Note that in the worst case, Algorithm 1 finds a solution with consensus d (i.e., the minimal case to fit $\mathbf{x}$), if there are no solutions with higher consensus to be found.

Theorem 4. *MAXCON is FPT in the number of outliers and dimension.*

Proof. Algorithm 1 conducts a depth-first tree search to find a recursive sequence of *basis* points to remove from $\mathcal{D}$ to yield a consensus set. By Lemma 6, the longest sequence of basis points that needs to be removed is $o = N - |\mathcal{I}^*|$, which is also the maximum tree depth searched by the algorithm (each descend of the tree removes one point). The number of nodes visited is of order $(d+1)^o$, since the branching factor of the tree is $|\mathcal{B}|$, and by Lemma 1, $|\mathcal{B}| \leq d + 1$.

At each node, LP[$\mathcal{C}$] is solved, with the largest of these LPs having $d + 1$ variables and N constraints. Algorithm 1 thus runs in $\mathcal{O}(d^o \text{poly}(N, d))$ time, which is exponential only in the number of outliers o and dimension d. □

Using [32, Theorem 2.3] and the repeated basis detection and avoidance procedure in [13, Sec. 3.1], the complexity of Algorithm 1 can be improved to $\mathcal{O}((o + 1)^d \text{poly}(N, d))$. See [33, Sec. 3.5] for details.

Algorithm 1 FPT algorithm for MAXCON.

Require: $\mathcal{D} = \{\mathbf{a}_i, b_i\}_{i=1}^N$, threshold ϵ.
 1: $\mathcal{C} \leftarrow \{1, \ldots, N\}$, $\hat{\mathbf{x}} \leftarrow$ NULL, $\hat{\psi} \leftarrow 0$.
 2: $[\hat{\mathbf{x}}, \hat{\psi}] \leftarrow$ fitRem$(\mathcal{C}, \epsilon, \hat{\mathbf{x}}, \hat{\psi})$.
 3: **return** $\hat{\mathbf{x}}$.

$[\hat{\mathbf{x}}, \hat{\psi}] =$ fitRem$(\mathcal{C}, \epsilon, \hat{\mathbf{x}}, \hat{\psi})$
 1: Solve LP$[\mathcal{C}]$ to obtain $f(\mathcal{C})$, $\hat{\mathbf{x}}_{\mathcal{C}}$, and the basis $\mathcal{B}$ of $\mathcal{C}$.
 2: **if** $f(\mathcal{C}) \leq \epsilon$ **then**
 3:     **if** $|\mathcal{C}| > \hat{\psi}$ **then**
 4:         $\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}}_{\mathcal{C}}$, $\hat{\psi} \leftarrow |\mathcal{C}|$. //Found a better consensus set.
 5:     **end if**
 6: **else**
 7:     **for** each $i \in \mathcal{B}$ **do**
 8:         $[\hat{\mathbf{x}}, \hat{\psi}] \leftarrow$ fitRem$(\mathcal{C} \setminus \mathrm{i}, \epsilon, \hat{\mathbf{x}}, \hat{\psi})$. //Remove points from basis and refit.
 9:     **end for**
10: **end if**
11: **return** $\hat{\mathbf{x}}$ and $\hat{\psi}$.

---

4   Approximability

Given the inherent intractability of MAXCON, it is natural to seek recourse in approximate solutions. However, this section shows that it is not possible to construct PTAS [18] for MAXCON.

Our development here is inspired by [34, Sec. 3.2]. First, we define our source problem: given a set of k Boolean variables $\{v_j\}_{j=1}^k$, a *literal* is either one of the variables, e.g., v_j, or its negation, e.g., $\neg v_j$. A *clause* is a disjunction over a set of literals, i.e., $v_1 \vee \neg v_2 \vee v_3$. A *truth assignment* is a setting of the values of the k variables. A clause is *satisfied* if it evaluates to true.

Problem 5 (MAX-2SAT). Given M clauses $\mathcal{K} = \{\mathcal{K}_i\}_{i=1}^M$ over k Boolean variables $\{v_j\}_{j=1}^k$, where each clause has exactly two literals, what is the maximum number of clauses that can be satisfied by a truth assignment?

MAX-2SAT is APX-hard [35], meaning that there are no algorithms that run in polynomial time that can approximately solve MAX-2SAT up to a desired error ratio. Here, we show an L-reduction [36] from MAX-2SAT to MAXCON, which unfortunately shows that MAXCON is also APX-hard.

Generating the input data Given an instance of MAX-2SAT with clauses $\mathcal{K} = \{\mathcal{K}_i\}_{i=1}^M$ over variables $\{v_j\}_{j=1}^k$, let each clause $\mathcal{K}_i$ be represented as $(\pm v_{\alpha_i}) \vee (\pm v_{\beta_i})$, where $\alpha_i, \beta_i \in \{1, \ldots, k\}$ index the variables that exist in $\mathcal{K}_i$, and $\pm$ here indicates either a "blank" (no negation) or $\neg$ (negation). Define

$$\mathrm{sgn}(\alpha_i) = \begin{cases} +1 & \text{if } v_{\alpha_i} \text{ occurs without negation in } \mathcal{K}_i, \\ -1 & \text{if } v_{\alpha_i} \text{ occurs with negation in } \mathcal{K}_i; \end{cases} \qquad (27)$$

similarly for $\text{sgn}(\beta_i)$. Construct the input data for MAXCON as

$$\mathcal{D}_{\mathcal{K}} = \{(\mathbf{a}_i^p, b_i^p)\}_{i=1,\ldots,M}^{p=1,\ldots,6}, \tag{28}$$

where there are six measurements for each clause. Namely, for each clause $\mathcal{K}_i$,

- $\mathbf{a}_i^1$ is a k-dimensional vector of zeros, except at the α_i-th and β_i-th elements where the values are respectively $\text{sgn}(\alpha_i)$ and $\text{sgn}(\beta_i)$, and $b_i^1 = 2$.
- $\mathbf{a}_i^2 = \mathbf{a}_i^1$ and $b_i^2 = 0$.
- $\mathbf{a}_i^3$ is a k-dimensional vector of zeros, except at the α_i-th element where the value is $\text{sgn}(\alpha_i)$, and $b_i^3 = -1$.
- $\mathbf{a}_i^4 = \mathbf{a}_i^3$ and $b_i^4 = 1$.
- $\mathbf{a}_i^5$ is a k-dimensional vector of zeros, except at the β_i-th element where the value is $\text{sgn}(\beta_i)$, and $b_i^5 = -1$.
- $\mathbf{a}_i^6 = \mathbf{a}_i^5$ and $b_i^6 = 1$.

The number of measurements N in $\mathcal{D}_{\mathcal{K}}$ is $6M$.

Setting the inlier threshold Given a solution $\mathbf{x} \in \mathbb{R}^k$ for MAXCON, the six input measurements associated with $\mathcal{K}_i$ are inliers under these conditions:

$$
\begin{aligned}
(\mathbf{a}_i^1, b_i^1) \text{ is an inlier} &\iff |\text{sgn}(\alpha_i)x_{\alpha_i} + \text{sgn}(\beta_i)x_{\beta_i} - 2| \leq \epsilon, \\
(\mathbf{a}_i^2, b_i^2) \text{ is an inlier} &\iff |\text{sgn}(\alpha_i)x_{\alpha_i} + \text{sgn}(\beta_i)x_{\beta_i}| \leq \epsilon,
\end{aligned} \tag{29}
$$

$$
\begin{aligned}
(\mathbf{a}_i^3, b_i^3) \text{ is an inlier} &\iff |\text{sgn}(\alpha_i)x_{\alpha_i} + 1| \leq \epsilon, \\
(\mathbf{a}_i^4, b_i^4) \text{ is an inlier} &\iff |\text{sgn}(\alpha_i)x_{\alpha_i} - 1| \leq \epsilon,
\end{aligned} \tag{30}
$$

$$
\begin{aligned}
(\mathbf{a}_i^5, b_i^5) \text{ is an inlier} &\iff |\text{sgn}(\beta_i)x_{\beta_i} + 1| \leq \epsilon, \\
(\mathbf{a}_i^6, b_i^6) \text{ is an inlier} &\iff |\text{sgn}(\beta_i)x_{\beta_i} - 1| \leq \epsilon,
\end{aligned} \tag{31}
$$

where x_α is the α-th element of $\mathbf{x}$. Observe that if $\epsilon < 1$, then at most one of (29), one of (30), and one of (31) can be satisfied. The following result establishes an important condition for L-reduction.

Lemma 7. *If $\epsilon < 1$, then*

$$\text{OPT}(\textit{MAXCON}) \leq 6 \cdot \text{OPT}(\textit{MAX-2SAT}), \tag{32}$$

$\text{OPT}(\textit{MAX-2SAT})$ *is the maximum number of clauses that can be satisfied for a given MAX-2SAT instance, and* $\text{OPT}(\textit{MAXCON})$ *is the maximum achievable consensus for the MAXCON instance generated under our reduction.*

Proof. If $\epsilon < 1$, for all $\mathbf{x}$, at most one of (29), one of (30), and one (31), can be satisfied, hence $\text{OPT}(\text{MAXCON})$ cannot be greater than $3M$. For any MAX-2SAT instance with M clauses, there is an algorithm [37] that can satisfy at least $\lceil \frac{M}{2} \rceil$ of the clauses, thus $\text{OPT}(\text{MAX-2SAT}) \geq \lceil \frac{M}{2} \rceil$. This leads to (32). □

Note that, if $\epsilon < 1$, rounding $\mathbf{x}$ to its nearest bipolar vector (i.e,, a vector that contains only -1 or 1) cannot decrease the consensus w.r.t. $\mathcal{D}_{\mathcal{K}}$. It is thus sufficient to consider $\mathbf{x}$ that are bipolar in the rest of this section.

Intuitively, $\mathbf{x}$ is used as a proxy for truth assignment: setting $x_j = 1$ implies setting $v_j = true$, and vice versa. Further, if one of the conditions in (29) holds for a given $\mathbf{x}$, then the clause $\mathcal{K}_i$ is satisfied by the truth assignment. Hence, for $\mathbf{x}$ that is bipolar and $\epsilon < 1$,

$$\Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_{\mathcal{K}}) = 2M + \sigma, \tag{33}$$

where σ is the number of clauses satisfied by $\mathbf{x}$. This leads to the final necessary condition for L-reduction.

Lemma 8. *If $\epsilon < 1$, then*

$$\left| \mathrm{OPT}(\textit{MAX-2SAT}) - \mathrm{SAT}(\mathbf{t}(\mathbf{x})) \right| = \left| \mathrm{OPT}(\textit{MAXCON}) - \Psi_\epsilon(\mathbf{x} \mid \mathcal{D}_{\mathcal{K}}) \right|, \tag{34}$$

where $\mathbf{t}(\mathbf{x})$ returns the truth assignment corresponding to $\mathbf{x}$, and $\mathrm{SAT}(\mathbf{t}(\mathbf{x}))$ returns the number of clauses satisfied by $\mathbf{t}(\mathbf{x})$.

Proof. For any bipolar $\mathbf{x}$ with consensus $2M + \sigma$, the truth assignment $\mathbf{t}(\mathbf{x})$ satisfies exactly σ clauses. Since the value of OPT(MAXCON) must take the form $2M + \sigma^*$, then OPT(MAX-2SAT) $= \sigma^*$. The condition (34) is immediately seen to hold by substituting the values into the equation. $\square$

We have demonstrated an L-reduction from MAX-2SAT to MAXCON, where the main work is to generate $\mathcal{D}_{\mathcal{K}}$ in linear time. The function $\mathbf{t}$ also takes linear time to compute. Setting $\epsilon < 1$ completes the reduction.

Theorem 5. *MAXCON is APX-hard.*

Proof. Since MAX-2SAT is APX-hard, by the above L-reduction, MAXCON is also APX-hard. $\square$

See Sec. 1.1 for the implications of Theorem 5.

5   Conclusions and future work

Given the fundamental difficulty of consensus maximisation as implied by our results (see Sec. 1.1), it would be prudent to consider alternative paradigms for optimisation, e.g., deterministically convergent heuristic algorithms [19–21] or preprocessing techniques [22–24].

Acknowledgements

References

1. Meer, P.: Robust techniques for computer vision. In Medioni, G., Kang, S.B., eds.: Emerging topics in computer vision. Prentice Hall (2004)
2. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6) (1981) 381–395
3. Choi, S., Kim, T., Yu, W.: Performance evaluation of RANSAC family. In: British Machine Vision Conference (BMVC). (2009)
4. Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.M.: USAC: a universal framework for random sample consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(8) (2013) 2022–2038
5. Tran, Q.H., Chin, T.J., Chojnacki, W., Suter, D.: Sampling minimal subsets with large spans for robust estimation. International Journal of Computer Vision (IJCV) **106**(1) (2014) 93–112
6. Li, H.: Consensus set maximization with guaranteed global optimality for robust geometry estimation. In: IEEE International Conference on Computer Vision (ICCV). (2009)
7. Zheng, Y., Sugimoto, S., Okutomi, M.: Deterministically maximizing feasible subsystems for robust model fitting with unit norm constraints. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). (2011)
8. Enqvist, O., Ask, E., Kahl, F., Åström, K.: Robust fitting for multiple view geometry. In: European Conference on Computer Vision (ECCV). (2012)
9. Bazin, J.C., Li, H., Kweon, I.S., Demonceaux, C., Vasseur, P., Ikeuchi, K.: A branch-and-bound approach to correspondence and grouping problems. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(7) (2013) 1565–1576
10. Yang, J., Li, H., Jia, Y.: Optimal essential matrix estimation via inlier-set maximization. In: European Conference on Computer Vision (ECCV). (2014)
11. Parra Bustos, A., Chin, T.J., Suter, D.: Fast rotation search with stereographic projections for 3d registration. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). (2014)
12. Enqvist, O., Ask, E., Kahl, F., Åström, K.: Tractable algorithms for robust model estimation. International Journal of Computer Vision **112**(1) (2015) 115–129
13. Chin, T.J., Purkait, P., Eriksson, A., Suter, D.: Efficient globally optimal consensus maximisation with tree search. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). (2015)
14. Campbell, D., Petersson, L., Kneip, L., Li, H.: Globally-optimal inlier set maximisation for simultaneous camera pose and feature correspondence. In: IEEE International Conference on Computer Vision (ICCV). (2017)
15. Garey, M.R., Johnson, D.S.: Computers and intractability: a guide to the theory of NP-completeness. W H Freeman & Co (1990)
16. Downey, R.G., Fellows, M.R.: Parametrized complexity. Springer-Verlag New York (1999)
17. Erickson, J., Har-Peled, S., Mount, D.M.: On the least median square problem. Discrete & Computational Geometry **36**(4) (2006) 593–607
18. Vazirani, V.: Approximation algorithms. Springer-Verlag Berlin (2001)
19. Le, H., Chin, T.J., Suter, D.: An exact penalty method for locally convergent maximum consensus. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). (2017)

20. Purkait, P., Zach, C., Eriksson, A.: Maximum consensus parameter estimation by reweighted L1 methods. In: Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR). (2017)
21. Cai, Z., Chin, T.J., Le, H., Suter, D.: Deterministic consensus maximization with biconvex programming. In: European Conference on Computer Vision (ECCV). (2018)
22. Svärm, L., Enqvist, O., Oskarsson, M., Kahl, F.: Accurate localization and pose estimation for large 3d models. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). (2014)
23. Parra Bustos, A., Chin, T.J.: Guaranteed outlier removal for rotation search. In: IEEE International Conference on Computer Vision (ICCV). (2015)
24. Chin, T.J., Kee, Y.H., Eriksson, A., Neumann, F.: Guaranteed outlier removal with mixed integer linear programs. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). (2016)
25. Johnson, D.S., Preparata, F.P.: The densest hemisphere problem. Theoretical Computer Science **6** (1978) 93–107
26. Ben-David, S., Eiron, N., Simon, H.: The computational complexity of densest region detection. Journal of Computer and System Sciences **64**(1) (2002) 22–47
27. Aronov, B., Har-Peled, S.: On approximating the depth and related problems. SIAM Journal on Computing **38**(3) (2008) 899–921
28. Bernholt, T.: Robust estimators are hard to compute. Technical Report 52, Technische Universität Dortmund (2005)
29. Cheney, E.W.: Introduction to Approximation Theory. McGraw-Hill (1966)
30. Giannopoulos, P., Knauer, C., Rote, G.: The parameterized complexity of some geometric problems in unbounded dimension. In: International Workshop on Parameterized and Exact Computation (IWPEC). (2009)
31. https://en.wikipedia.org/wiki/Parameterized_complexity
32. Matoušek, J.: On geometric optimization with few violated constraints. Discrete and Computational Geometry **14**(4) (1995) 365–384
33. Chin, T.J., Purkait, P., Eriksson, A., Suter, D.: Efficient globally optimal consensus maximisation with tree search. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **39**(4) (2017) 758–772
34. Amaldi, E., Kann, V.: The complexity and approximability of finding maximum feasible subsystems of linear relations. Theoretical Computer Science **147** (1995) 181–210
35. https://en.wikipedia.org/wiki/2-satisfiability
36. https://en.wikipedia.org/wiki/L-reduction
37. Johnson, D.S.: Approximation algorithms for combinatorial problems. J. Comput. System Sci. **9** (1974) 256–278