Consensus Maximization with Linear Matrix Inequality Constraints

Pablo Speciale\textsuperscript{1}, Danda P. Paudel\textsuperscript{2}, Martin R. Oswald\textsuperscript{1}, Till Kroeger\textsuperscript{2}, Luc V. Gool\textsuperscript{2,4}, and Marc Pollefeys\textsuperscript{1,3}

\textsuperscript{1} Department of Computer Science, ETH Zürich.
\textsuperscript{3} Microsoft, Redmond, USA
\{pablo, moswald, marc.pollefeys\}@inf.ethz.ch
\textsuperscript{2} Computer Vision Laboratory, D-ITET, ETH Zürich
\textsuperscript{4} VISICS, ESAT/PSI, KU Leuven, Belgium
\{paudel, kroegert, vangool\}@vision.ee.ethz.ch

Abstract

Consensus maximization has proven to be a useful tool for robust estimation. While randomized methods like RANSAC are fast, they do not guarantee global optimality and fail to manage large amounts of outliers. On the other hand, global methods are commonly slow because they do not exploit the structure of the problem at hand.

In this paper, we show that the solution space can be reduced by introducing Linear Matrix Inequality (LMI) constraints. This leads to significant speed ups of the optimization time even for large amounts of outliers, while maintaining global optimality. We study several cases in which the objective variables have a special structure, such as rotation, scaled-rotation, and essential matrices, which are posed as LMI constraints. This is very useful in several standard computer vision problems, such as estimating Similarity Transformations, Absolute Poses, and Relative Poses, for which we obtain compelling results on both synthetic and real datasets. With up to 90 percent outlier rate, where RANSAC often fails, our constrained approach is consistently faster than the non-constrained one - while finding the same global solution.

1. Introduction

One of the major difficulties of many central computer vision problems – besides the proper handling of noise and incomplete data – is the robust detection of outliers. For many optimization methods, the number of outliers has a tremendous impact on the runtime or even on the solvability. A common approach for robust estimation is, therefore, to explicitly maximize the number of inliers for a given problem – also called Consensus Maximization. A large number of optimization methods for robust estimation have been proposed in the literature which can be roughly divided into local and global optimization methods.

Contributions. We argue that many computer vision problems have a special structure that can be leveraged in global robust estimation methods to make them much more com-

\textbf{Local Optimization Methods.} With currently more than 15K citations, RANSAC [13], is by far the most popular method. It has been used in numerous applications and many extensions have been proposed, e.g. [8, 33, 34] (see [7, 28] for an overview). The great advantage is its simplicity and effectiveness for various scenarios, but it also has a number of shortcomings: 1) it does not guarantee optimality and only finds a local optimum, 2) it cannot find the exact solution if it is not contained in the sampling set, and 3) its expected computation time grows exponentially with large amounts of outliers.

\textbf{Global Optimization Methods.} Global methods commonly have considerably larger computational costs, as they are mostly based on exhaustive search within the entire optimization domain. Almost every global method uses the Branch-and-Bound (BnB) strategy to make the search tractable, e.g. [1, 2, 16, 22, 38]. Similar to our approach, several methods use Mixed Integer Programming (MIP) [4, 9, 22, 36] within the BnB optimization in order to solve the overall problem faster. Recently [5] proposed to cast the problem of consensus maximization as a tree search problem which is then traversed with A*-search for faster optimization. This method does not need linearization of the residual and only traverses a small subset of the tree compared to exhaustive methods like [10, 26].

Application-wise, many related works are specialized to a particular problem class, like linear problems [22], pseudo-convex problems [5, 21], or, they are even more specialized to a specific type of geometric problem, for instance problems including rotations [2, 16], rotation+focal length [1], translation [14], rotation+translation+scale [27] or essential matrices [37]. Most of these methods specialize on a particular problem and their application to a different problem class is not necessarily straightforward.

In this paper, we propose a general optimization framework that covers all problems that can be expressed with LMI constraints and, therefore, tackles the majority of the aforementioned problem classes.

Acknowledgements. This project has received funding from the European Unions Horizon 2020 research and innovation programme, and European Research Council project VarCity, under grant agreements No. 637221 and 273940, respectively.
A Linear Matrix Inequality (LMI) is the constraint on \( y \in \mathbb{R}^n \) such that \( A(y) \succeq 0 \). A Semi-definite Program (SDP) consists of minimizing (or maximizing) a linear objective subject to LMI constraints. It is a convex optimization problem that can be efficiently solved using interior-point methods [3].

2.3. LMI Constrained Quadratic Programming

A LMI constrained quadratic programming is a problem of the form:

\[
\begin{align*}
\text{minimize} & \quad y^T Q y + q^T y + r, \\
\text{subject to} & \quad A(y) \succeq 0,
\end{align*}
\]

where \( Q \) is a real symmetric matrix, \( q \) a real vector, and \( r \) a real scalar.

For \( Q \succeq 0 \), the problem of (2) can be optimally solved using SDP. Note that \( Q \succeq 0 \) can always be factorized into \( Q = M^T M \) for some matrix \( M \), using the Cholesky decomposition. Therefore (2) is equivalent to the following SDP:

\[
\begin{align*}
\text{minimize} & \quad y^T \theta, \\
\text{subject to} & \quad 
\begin{pmatrix}
1 & M^T \\
M & \theta - q^T y - r
\end{pmatrix} \succeq 0,
\end{align*}
\]

A(n) \( A(y) \geq 0 \).

3. Consensus Maximization with LMI

Consider a geometric transformation \( T(x) : U \to V \) that relates a pair of measurements \( \mathcal{P} = \{U, V\} \). Let \( \gamma(x) \) be the residual error for a known \( \mathcal{P} \) and the estimate \( x \). The problem of maximizing the measurements’ consensus (i.e. inlier set) under a LMI constraint is,

\[
\begin{align*}
\text{Problem 3.1 Given a set of measurement pairs} & \\
\text{maximize} & \quad \gamma_i(x), \\
\text{subject to} & \quad \gamma_i(x) \leq \epsilon, \quad \forall P_i \in \zeta, \\
& \quad A(x) \succeq 0.
\end{align*}
\]

In general, solving (4) exactly is non-trivial, as this is a NP-hard combinatorial optimization problem. Such problems are usually solved using sample-and-test techniques, such as RANSAC, with no guarantee on the optimality of the results. In contrast, exact methods are based on variations of tree search algorithms [5, 10, 22, 26, 38]. Our following proposition is concerned about the optimal solution search for a class of such problems.

Proposition 3.2 (Consensus with LMIs) Problem 3.1 can be solved optimally using a tree search method for linear \( \gamma_i(x) \) or quadratic residuals \( \gamma_i(x) = x^T Q_i x + q_i^T x + r_i \), with \( Q_i \succeq 0 \).
suggests that the general class of problems of the following form, is obtained, 

\[
\begin{align*}
\text{minimize} & \quad \max_i \gamma_i(x), \\
\text{subject to} & \quad A(x) \succeq 0.
\end{align*}
\]  

(5)

For linear \(\gamma_i(x)\), (5) is equivalent to the following SDP,

\[
\begin{align*}
\text{minimize} & \quad \theta, \\
\text{subject to} & \quad \gamma_i(x) \leq \theta, \quad \forall i, \\
& \quad A(x) \succeq 0.
\end{align*}
\]

(6)

Similarly, if \(\gamma_i(x) = x^T Q_i x + q_i^T x + r_i\) with \(Q_i \succeq 0\), then (5) can be solved using the following SDP, for \(Q_i = M_i^T M_i\),

\[
\begin{align*}
\text{minimize} & \quad \theta, \\
\text{subject to} & \quad \left(1 \quad M_i^T x \right) \theta - q_i^T x - r_i \succeq 0, \quad \forall i, \\
& \quad A(x) \succeq 0.
\end{align*}
\]

(7)

Alternatively, following a similar argument of convexity as in [11], one can show that (5) is a LP-type [24] (or generalized linear program). This concludes the proof.

**Mixed Integer Programming** Before entering into further details, we discuss the choice of solving (4) using Mixed Integer Semi-Definite Programming (MI-SDP) [9, 36]. MI-SDP framework can solve (4) optimally, which can be re-stated as:

\[
\begin{align*}
\text{minimize} & \quad \sum_i z_i, \\
\text{subject to} & \quad \gamma_i(x) \leq \epsilon + z_i M_i, \quad \forall i, \\
& \quad z_i \in \{0, 1\}, \\
& \quad A(x) \succeq 0.
\end{align*}
\]

(8)

where \(z = \{z_i\}_{i=1}^n\) are binary variables and \(M_i\) is a large enough positive constant. It is a common practice in optimization to ignore constraints by using a constant such as \(M_i\). See [6, Ch. 7] for guidelines on selecting this constant. Intuitively, the data pair generating the residual \(\gamma_i(x)\) will be considered as an outlier if \(z_i = 1\). Therefore, for the given optimal solution \(\mathbf{z}^*\) to (8), the maximum consensus set can be obtained by,

\[
\zeta^* = \{i \mid z_i^* = 0\}.
\]

(9)

**4. Transformation Equation**

Although Proposition 3.2 suggests that the general class of problems (4) can be optimally solved, in this work we are concerned with problems of the following form,

\[
\beta_i(x) u_i = B_i(x) u_i + b(x),
\]

(10)

where \(\{u_i, v_i\}\) are the measurement vectors of the pair \(P_i\), and \(B_i(x), b(x)\), and \(\beta_i(x)\) are the terms linear on \(x\) which form the transformation \(T(x)\). For a given problem, we wish to enforce the structural constraint of \(T(x)\) in terms of LMIs, while minimizing the residual error of (10).

**4.1. Residual with Noise Model**

We model the noise as a Gaussian process. The dissimilarity measure between two corresponding measurements is therefore expressed in terms of the generalized squared interpoint distance (also known as squared Mahalanobis distance). For a given pair \(P_i\), the residual error (or dissimilarity measure) is given by,

\[
\gamma_i(x) = \Delta_i(x)^T \Sigma^{-1} \Delta_i(x),
\]

(11)

\[
\Delta_i(x) = B_i(x) u_i + b(x) - \beta_i(x) v_i
\]

where \(\Sigma\) is the covariance matrix of a known distribution.

**Remark 4.1** The residual error (11) is a quadratic function of the form \(\gamma_i(x) = x^T Q_i x + q_i^T x + r_i\) with \(Q_i \succeq 0\).

**4.2. Residual Minimization**

Once the optimal inlier set \(\zeta\) for Problem 3.1 is obtained, the best estimate \(\mathbf{x}\) that minimizes the collective residual errors for all the inliers, while satisfying the LMI constraint, can be obtained using the following result. This can be considered as a refinement step.

**Result 4.2** The optimal estimate \(\mathbf{x}\) that minimizes the sum of residuals for all inlier pairs \(\zeta = \{I_j\}_{j=1}^m \subseteq \mathcal{Z}\) can be obtained using the LMI constrained quadratic programming,

\[
\begin{align*}
\text{minimize} & \quad x^T \left(\sum_{j=1}^m Q_j\right) x + \left(\sum_{j=1}^m d_j\right)^T x + \sum_{j=1}^m r_j, \\
\text{subject to} & \quad A(x) \succeq 0.
\end{align*}
\]

(12)

This is a convex problem, that can be solved efficiently using the SDP as discussed in Section 2.3. It follows that, \(Q_j \succeq 0 \implies \sum_{j=1}^m Q_j \succeq 0\).

**5. LMI Constraints**

In this section, we introduce four LMI constraints that we will use in our experiments. Two of these constraints were recently proposed in [15, 29]. The other two constraints are presented for the first time in this paper. First, we define the function \(\mathcal{L} : \mathbb{R}^{3 \times 3} \rightarrow \mathbb{R}^{4 \times 4}\) of the form,

\[
\mathcal{L}(A) = \begin{bmatrix}
0 & a_{12} & a_{13} & a_{21} & a_{22} & a_{23} & a_{31} & a_{32} & a_{33}
\end{bmatrix}.
\]

(13)
5.1. Orbitope and Rotation Matrix

Definition 5.1 (Orbitope [29]) An orbitope is the convex hull of an orbit of a compact algebraic group that acts linearly on a real vector space. The orbit has the structure of a real algebraic variety, and the orbitope is a convex semi-algebraic set.

A 3-dimensional rotation matrix \( R \in SO(3) \) has dimension three. However, its tautological orbitope is a convex body of dimension nine. The following theorem is a key ingredient of this work:

Theorem 5.2 (SO(3) Orbitope [29]) The tautological orbitope \( \text{conv}(SO(3)) \) is a spectrahedron whose boundary is a quartic hypersurface. A 3×3 matrix \( A \) lies in \( \text{conv}(SO(3)) \) if and only if,

\[
I_{4 \times 4} + L(A) \succeq 0. \tag{14}
\]

For some applications, the desired rotation matrix must have a restricted rotation angle around an arbitrary rotation axis. For example, the relative rotation between two cameras cannot be too large for them to share a common field of view. Such rotation angle restrictions can be expressed as LMI constraints using the following Lemma.

Lemma 5.3 (Bounded SO(3) [15]) Consider \( R \in SO(3) \) expressed in the angle-axis form by a rotation angle \( \theta \) around an arbitrary axis. The eigen-decomposition of the symmetric matrix \( R + R^\top \) is of the form,

\[
R + R^\top = U \begin{bmatrix} 2 \cos \theta & 0 & 0 \\ 0 & 2 \cos \theta & 0 \\ 0 & 0 & 2 \end{bmatrix} U^{-1}. \tag{15}
\]

For \( |\theta| \leq 90^\circ \), the symmetric matrix \( R + R^\top \succeq 0 \). In fact, the rotation angle is within the given upper bound, \( |\theta| \leq \bar{\theta} \leq 90^\circ \), if and only if,

\[
R + R^\top \succeq 2 \cos \bar{\theta} I_{3 \times 3}. \tag{16}
\]

5.2. Transformations Coupled with Rotation

Some transformations involving both rotation and an extra scale (e.g. Metric Transformation) can be expressed with the help of a scaled-rotation matrix. The following definition deals with the structure of a scaled-rotation matrix.

Definition 5.4 (SSO(3)) Given a real, compact, linear algebraic group \( \mathcal{H} \), a 3-dimensional scaled-special orthogonal group is defined by,

\[
SSO(3) = \{ H \in \mathcal{H} : HH^\top = \alpha^2 I_{3 \times 3}, \det(H) = \alpha^3, \alpha > 0 \}. \tag{17}
\]

In the following proposition, we provide a convex relaxation for the scaled-rotation matrix as a LMI.

Proposition 5.5 (SSO(3) and SO(3) Orbitope)
\( \forall S \in SSO(3) \) there exists \( A \in \text{conv}(SO(3)) \) such that \( S = \alpha A \), if and only if \( \exists \alpha > 0 \):

\[
\alpha I_{4 \times 4} + L(S) \succeq 0. \tag{18}
\]

Proof If \( S \in SSO(3) \iff S = \alpha A \) for \( A \in SO(3) \) and \( \alpha > 0 \). Notice from \( (13) \) that \( \frac{1}{\alpha} L(S) = L(\frac{S}{\alpha}) \). Therefore, \( (18) \) is equivalent to \( I_{4 \times 4} + L(\frac{S}{\alpha}) \succeq 0 \). After replacing \( A = \frac{S}{\alpha} \), we get the same result as in Theorem 5.2. The backward proof for equivalence can be obtained in a very similar manner.

Remark 5.6 For a given \( \alpha > 0 \) and \( S = \alpha R \) with \( R \) bounded by \( |\theta| \leq \bar{\theta} \leq 90^\circ \), the following LMI must also hold true,

\[
S + S^\top \succeq 2 \alpha \cos \bar{\theta} I_{3 \times 3}. \tag{19}
\]

In the calibrated relative pose formulation, a rotation matrix appears with a skew-symmetric matrix, the so-called Essential matrix. More formally, the normalized Essential matrix is defined as follows.

Definition 5.7 (Normalized Essential Matrix) The set of normalized Essential matrices for two cameras related by rotation \( R \) and translation \( t \) is given by,

\[
\mathcal{E} = \{ E = [t]_x R : R \in SO(3), ||t|| = 1 \}. \tag{20}
\]

where \([t]_x\) is a 3×3 skew-symmetric matrix.

We will now show that the structural constraint of the Essential matrix can also be expressed as LMI using our following proposition.

Proposition 5.8 (Essential Matrix and Orbitope) A 3×3 matrix \( E \) belongs to the set of normalized Essential matrices, \( \mathcal{E} \) of \( (20) \), only if,

\[
2 I_{4 \times 4} + L(E) \succeq 0. \tag{21}
\]

Proof For any \( E \in \mathcal{E} \), its singular value decomposition is given by \( E = U \text{diag}(1,1,0) V^\top \). This can further be decomposed into \( E = A + B \), with \( A \in SO(3) \) and \( B = U \text{diag}(0,0,||UV^\top||) V^\top \). Using \( (13) \), one can show that \( L(B) \succeq -I_{4 \times 4} \). Furthermore, from \( (14) \), \( L(A) \succeq I_{4 \times 4} \). Therefore, \( L(E) = L(A) + L(B) \preceq -2 I_{4 \times 4} \), which leads to \( (21) \).

6. Multiple View Geometry Problems

In this section, we present three examples of multiple view geometry problems formulated as in Problem 3.1 (Consensus Maximization with LMIs). Different problems specify different variable terms, in reference to \( (10) \), and their LMI constraints are summarized in Table 1.
6.1. Similarity Transformation

We consider a set of images acquired by a collection of cameras which observe the same scene. These images are then fed into a SfM pipeline \([17, 30]\) to obtain a 3D Reconstruction. Let \(\{u_i, v_i\} \in \mathbb{R}^3, i = 1, \ldots, n\), with \(n \geq 4\), be the Cartesian coordinate vector pairs of the SfM-induced camera centers \(U_i\) and their real world positions \(V_i\) (e.g. GPS measurements). Since the SfM reconstruction is metric, SfM-induced cameras and their world measurements are related by

\[
v_i = S(x) u_i + t(x),
\]

where \(S(x)\) is a \(3 \times 3\) scaled-rotation matrix, \(t(x)\) a \(3 \times 1\) translation vector, and \(x \in \mathbb{R}^{12}\). Notice that (22) is analogous to (10), hence its residual error can be written as in (11). On the other hand, a direct application of Proposition 5.5 provides a convex relaxation, as a LMI, to the scaled-rotation matrix constraint, i.e. \(S(x) \in SSO(3)\).

\[S(x) = \alpha A \text{ such that } A \in \text{conv}(SO(3))\]

\[
\mathbf{K}_s = \{ \alpha I_{4 \times 4} + \mathcal{L}(S(x)) \} \succeq 0.
\]

6.2. Absolute Pose

We consider measurement vector pairs \(\{u_i, v_i\} \in \mathbb{R}^3, i = 1, \ldots, n\), with \(n \geq 5\), where \(u_i\) is the Cartesian representation of the scene point \(U_i\) in the world frame and \(v_i\) is the homogeneous representation of the image point \(V_i\) in the camera frame of a calibrated camera. If \([R|t]\) is the camera pose w.r.t. the world frame, then the scene and image points are related by,

\[
(r_3(x)^T u_i + t_3(x)) v_i = R(x) u_i + t(x),
\]

where \(x \in \mathbb{R}^{12}\), \(r_i\) are the row vectors of \(R(x)\), and \(t_i\) are \(i^{th}\) elements of \(t(x)\). Notice again that (24) is analogous to (10), hence its residual error can be written as in (11). On the other hand, a convex relaxation of the constraint \(R(x) \in SO(3)\) can be expressed as a LMI, using Theorem 5.2.

\[R(x) \in \text{conv}(SO(3))\]

\[
\mathbf{K}_a = \{ I_{4 \times 4} + \mathcal{L}(R(x)) \} \succeq 0.
\]

6.3. Relative Pose

We consider the homogeneous vector pairs \(\{u_i, v_i\} \in \mathbb{R}^3, i = 1, \ldots, n\), with \(n \geq 8\), which are the measurements of the image points \(\{U_i, V_i\}\) of two calibrated cameras. For an essential matrix \(E\), the relationship between two image points can be expressed as,

\[
([n_i]_x e_2(x) - (n_i)_x e_1(x)]^T u_i = [n_i]_x E(x) u_i,
\]

where \(x \in \mathbb{R}^9\), \([[n_i]_x\] is a \(3 \times 3\) skew symmetric matrix for any \(n_i \in \text{Null}(v_i^T)\), and \(e(X)\) are the row vectors of \(E\). Notice again that (26) is analogous to (10), hence its residual error can be written as in (11). On the other hand, a direct application of Proposition 5.8 provides a convex relaxation, as a LMI, to the Essential matrix constraint, i.e. \(E(x) \in \mathcal{E}\).

\[E(x) \in \text{conv}(SSO(3))\]

\[
\mathbf{K}_r = \{ 2 I_{4 \times 4} + \mathcal{L}(E(x)) \} \succeq 0.
\]

7. Experiments & Results

We performed experiments for the three problems described in Section 6, both on synthetic and real data. Our approach was implemented in MATLAB2016a using the Yalmip toolbox and Mosek as SDP solver. All experiments were carried out on an Intel Core i7 CPU 2.60GHz with 12GB RAM. Although there is still room for improvement by correctly modeled covariance matrix of individual applications, we used the Euclidean distance, i.e. \(\Sigma = I\) in Section 4.1. The error measurement metrics used for evaluating the quality of the results are: the errors in rotation \(R\), translation \(T\), scale \(S\), and the RMS 3D error. For each experiment, we compute the errors \(\Delta r = ||r - r_{gt}||\), \(\Delta t = ||t - t_{gt}||\), and \(\Delta s = ||s - s_{gt}||\). The errors reported as \(\Delta R, \Delta T,\) and \(\Delta S\) are the RMS values of \(N\) experiments. Here, \(r\) is a vector obtained by stacking three rotation angles in degrees, and \(t_{gt}, t_{gt}\) and \(s_{gt}\) are the ground truth values.

1https://yalmip.github.io/
2https://www.mosek.com/
7.1. Similarity Transformation

In this section, we show the general properties of the proposed method before and after adding the LMI constraints. For all our experiments, we restrict the reconstruction scale within [0.2, 5.0], recall \( \alpha \) in Proposition 5.5.

**Synthetic data.** We synthetically generated Similarity Transformations, which were applied to \( N \) points (uniformly generated) in order to obtain Ground Truth correspondences. We then introduced outliers by adding a high amount of noise to a particular subset of these correspondences, until the desired outlier ratio was obtained.

In Table 2, a time comparison between MI-SDP (without the LMI constraints from equation (8)) and Ours (with constraints) has been presented for a small number of points. In the presence of high amount of outliers, the speedups are very significant. Fig. 1 shows the runtime of our method with the constraints. In Fig. 2, we have compared our method against RANSAC. We fixed the number of points to \( N = 100 \) for all cases, and ran different instances while increasing the outlier ratio. For Fig. 2c the outlier ratio was set to 75% and the experiments were conducted for 100 times. Notice that the average performance of RANSAC (in this figure) corresponds approximately to its performance for 75% in Fig. 2b.

**Real data.** Images from the Yahoo Flickr Creative Commons dataset \([18, 32]\) were used to obtain 3D Reconstructions with COLMAP\(^3\), an open-source Structure-from-Motion (SfM) pipeline \([30]\). The SfM Reconstructions acquired correspond to: Colosseum (2060 images), Notre Dame (3743 images), Pantheon (1385 images) and Trevi Fountain (2909 images). Approximately only 10% of the images had GPS information; the numbers of valid GPS Tag found for each dataset is shown in Table 3. This table also provides additional quantitative results. Due to the lack of Ground Truth registration parameters, the reported quantitative results were visually evaluated. As expected, the quality of the results improves with increasing numbers of inliers. This can be observed with a large number of inliers in Colosseum, in contrast to the Pantheon dataset (where the number of inliers is only 14). For the qualitative evaluation, all 3D Point Clouds were registered to Open Street Maps\(^4\) and are shown in Fig. 3.

Fig. 4 provides a comparison between RANSAC and our method for the Colosseum dataset. In the same figure, the GPS elevation measurements plot is also provided. It can be observed that the GPS data exhibit huge deviations along the vertical axis, affecting RANSAC in particular.

---

\(^3\)https://colmap.github.io/

\(^4\)http://openstreetmap.org
Table 3: Similarity Transformation (real data): Quantitative results. Visually evaluated quality of registration.

<table>
<thead>
<tr>
<th>Scene</th>
<th>$\Delta \theta$ (Yaw)</th>
<th>$\Delta \theta$ (Pitch)</th>
<th>$\Delta \theta$ (Roll)</th>
<th>$\Delta T$</th>
<th>Height</th>
<th>Scale</th>
<th></th>
<th>/ N</th>
<th>Time [sec]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colosseum</td>
<td>$&lt; 1^\circ$</td>
<td>$&lt; 1^\circ$</td>
<td>$&lt; 1^\circ$</td>
<td>$&lt; 1,\text{m}$</td>
<td>$&lt; 1,\text{m}$</td>
<td>$&lt; 1,%$</td>
<td>$117/147$</td>
<td>88.26 s</td>
<td></td>
</tr>
<tr>
<td>Notre Dame</td>
<td>$&lt; 3^\circ$</td>
<td>$&lt; 2^\circ$</td>
<td>$&lt; 1^\circ$</td>
<td>$&lt; 2,\text{m}$</td>
<td>$&lt; 1,\text{m}$</td>
<td>$&lt; 1,%$</td>
<td>$103/144$</td>
<td>43.17 s</td>
<td></td>
</tr>
<tr>
<td>Pantheon</td>
<td>$&lt; 3^\circ$</td>
<td>$&lt; 5^\circ$</td>
<td>$&lt; 2^\circ$</td>
<td>$&lt; 3,\text{m}$</td>
<td>$&lt; 2,\text{m}$</td>
<td>$&lt; 7,%$</td>
<td>$14/ 47$</td>
<td>16.12 s</td>
<td></td>
</tr>
<tr>
<td>Trevi Fontain</td>
<td>$&lt; 2^\circ$</td>
<td>$&lt; 1^\circ$</td>
<td>$&lt; 3^\circ$</td>
<td>$&lt; 1,\text{m}$</td>
<td>$&lt; 1,\text{m}$</td>
<td>$&lt; 3,%$</td>
<td>$104/140$</td>
<td>65.68 s</td>
<td></td>
</tr>
</tbody>
</table>

$\Delta \theta$ [degree]: rotation error (Yaw, Pitch and Roll). $\Delta T$ [meters]: translation error. $\zeta^*$: maximum consensus set. $N$: number of available GPS tags.

Figure 4: Similarity Transformation (real data): comparison between RANSAC (purple) vs. Ours. The bottom plot shows the GPS elevation measurements, varying from 20 to 90 meters, where 42 (straight line) is the true elevation of Colosseum.

For this experiment, we used a SfM Reconstruction from an unordered set of 30 images. The ground truth absolute pose $[R|t]$ for 6 query images was also provided in the dataset. We follow an established [19] image-to-SfM localization pipeline to find the absolute pose of the query images: we compute SIFT [23] features in all query images, and match descriptors against the database of SIFT features from all 30 images used in the reconstruction. Since descriptors used in the SfM Reconstruction are associated with 3D points, we obtain a list of potential 3D-2D correspondences $\{u_i, v_i\}$. The 2D image points $v_i$ are transformed into normalized (homogeneous) image coordinates: $v_i = K^{-1} [v_i^t 1]^T$, where $K$ denotes internal camera calibration matrix. The list of correspondences $\{u_i, v_i\}$ for each query image contains on average an outlier percentage of 44.25 %. We aim at recovering the absolute pose $[R|t]$ of the query images, given the list of potential matches, as described in Section 6.2.

Table 4 complements this information with the execution time and the numbers of inliers. Here, we provide

Direct Linear Transform (DLT) followed by Levenberg-Marquardt (LM) minimization of the reprojection error.
### 7.3. Relative Pose

We conducted experiments with two different datasets – Fountain and Herz-Jesu – form [31]. The details of the datasets and the obtained results are shown in Table 5. The reported results obtained by our method in presence of the LMI constraint (27) is compared against RANSAC conducted on 5-point algorithm [25]. One can observe that the global search method with LMI constraints finds a larger set of inliers than that of RANSAC, for both datasets. The quality of relative pose estimation improves with increasing inliers in both cases.

### 8. Conclusion

We proposed a general global optimization framework for consensus maximization with linear matrix inequality constraints. We derived several LMI constraints and demonstrated that a number of central computer vision problems can be cast into this form. In particular, we successfully conducted experiments on problems of similarity transformation, absolute pose, and relative pose estimation. Experiments demonstrated also a significant speedup in computation time due to the addition of the LMI constraints, under a globally optimal framework.

As future work it is worth to explore the use of LMI constraints in combination with other exact methods [5, 10, 22, 26, 38], since the LMI constraints have the potential to improve their effectiveness. We therefore see these methods as complementary work, rather than competitors. Apart from combining it with other optimization methods, we further look forward to explore other computer vision problems that fit into the proposed optimization framework.
References


