Globally Optimal Estimates for Geometric Reconstruction Problems

Fredrik Kahl[1,2] Didier Henrion[3,4]
fredrik@maths.lth.se henrion@laas.fr

[1] Computer Science and Engineering, University of California San Diego, USA
[2] Centre for Mathematical Sciences, Lund University, Sweden
[3] LAAS-CNRS, Toulouse, France
[4] Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

Abstract

We introduce a framework for computing statistically optimal estimates of geometric reconstruction problems. While traditional algorithms often suffer from either local minima or non-optimality - or a combination of both - we pursue the goal of achieving global solutions of the statistically optimal cost-function.

Our approach is based on a hierarchy of convex relaxations to solve non-convex optimization problems with polynomials. These convex relaxations generate a monotone sequence of lower bounds and we show how one can detect whether the global optimum is attained at a given relaxation. The technique is applied to a number of classical vision problems: triangulation, camera pose, homography estimation and last, but not least, epipolar geometry estimation. Experimental validation on both synthetic and real data is provided. In practice, only a few relaxations are needed for attaining the global optimum.

1 Introduction

Minimizing globally a rational function of several variables is a difficult optimization problem in general. Multivariate polynomial minimization, a special case of rational minimization, is a hard problem already for degree 4 polynomials. For example [13], the problem of deciding whether an integer sequence $a_1, \dots, a_n$ can be partitioned, i.e. whether there exists $x \in \{\pm 1\}^n$ such that $\sum_{i=1}^n a_i x_i = 0$, or equivalently whether zero is the global minimum of polynomial $(\sum_i a_i x_i)^2 + \sum_i (x_i^2 - 1)^2$, is known to be NP-complete.

Many geometric computer vision problems can be formulated as a minimization problem where the objective function is a rational polynomial in the unknown variables. These rational polynomials arise due to the perspective mapping of the camera. In this paper, we show how such problems can be recast as a polynomial optimization problem using linear matrix inequalities (LMIs) and polynomial matrix inequalities (PMIs). Such problems have been under intense research during the last few years in the control community, e.g. [18, 8]. We leverage on these results in order to solve a number of geometric reconstruction problems, such as triangulation, camera pose and epipolar geometry estimation.

Our main contributions are:

(i) A general framework for computing globally optimal estimates for geometric vision problems is introduced. We apply this technique to problems for which current state-of-the-art uses local, iterative optimization techniques. Such methods are dependent on good initialization which is often hard to obtain in practice. Therefore they risk getting stuck in local minima.

(ii) We extend the theory of convex LMI relaxations by showing that even though only partial relaxations are employed, it is still possible to obtain global estimates, and more importantly, to detect if global optimality is achieved. As we shall see, this result makes it possible to avoid a combinatorial explosion of relaxation variables and many of the problems we consider become computationally tractable even when many unknown variables are involved.

1.1 Related work

Structure and motion problems in computer vision are core problems and have been studied for quite some time now. Many good algorithms can be found in recent textbooks, e.g. [7, 19]. It is well-known that local minima frequently occur and they have been analyzed in more detail in [25, 22, 21]. The reconstruction methods can be classified into three categories:

- *Non-optimal methods* use some simplied error criteria in order to obtain an estimate, often in closed-form. A classical example is the 8-point algorithm [7] or alternatively a minimal method [16], which was improved in [3] by enforcing the singularity of the fundamental matrix in the estimation process. These non-optimal schemes often serve as an initialization for a local method.

- *Local methods* such as gradient descent refinements, also known as bundle adjustment [27] do optimize the correct cost-function, but they are very sensitive to initialization point.

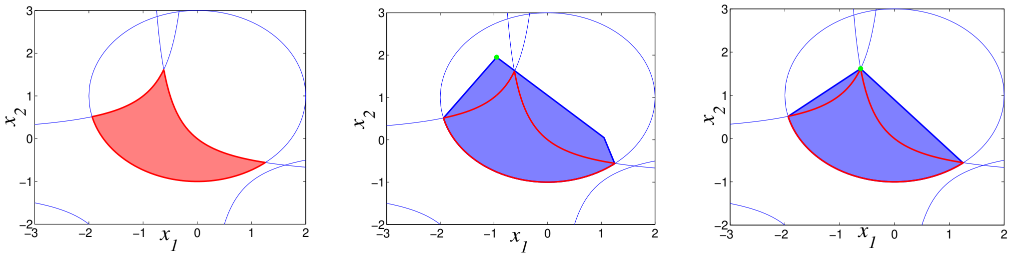
- *Global methods* are relatively rare in the vision literature. The triangulation problem for *two* views for different optimality criteria was solved in [6, 21, 20]. However, the problem is rather limited in complexity and it is hard to generalize even for three views (except in the case of L_∞-norm [5, 14]). For the statistically optimal L_2-norm and arbitrary number of views, this was previously an open problem.

  The factorization algorithm [26] computes a global estimate for both structure and motion with respect to the optimal L_2-norm, but this is unfortunately only valid for the affine camera model. Other interesting methods are graph-cuts which have successfully been applied to multi-view stereo matching [17] and interval analysis applied to auto-calibration [4]. However, one of the drawbacks of [4], which is also true for many other global methods is that they are computationally highly demanding.

We propose another strategy to achieve globally optimal estimates, which is still tractable from a computational point of view and which can handle harder problems than, for instance, the two-view triangulation problem. The method is global in the sense that it solves the problem when finite convergence occurs and it also provides a numerical *certificate* of global optimality. The formulation is based on the LMI formalism and we make extensive use of convex semidefinite programming (SDP). In particular, we rely on efficient SDP-solvers publicly already available, e.g. [24], which avoids the tedious and considerably difficult work of developing a specific optimizer. In addition, easy-to-use Matlab interfaces such as GloptiPoly [9] are also readily available.

The most closely related work we are aware of is [3]. A convexification scheme is also employed to solve a non-convex problem, namely the problem of estimating the fundamental matrix F subject to the cubic constraint $\det F = 0$. However, the objective function is the algebraic cost-function used in the 8-point algorithm. This problem can be simplified to a non-linear problem with two unknowns. Their approach of computing the solution involves solving a series of convex LMI problems via a bisection method.

A shorter version of this work has appeared in the conference paper [15]. In parallel to this work, some recent contributions have also appeared. In [23], the three-view triangulation problem is solved with a Gröbner basis technique. In [1], fractional programming using branch-and-bound is proposed to solve a subset of the problems considered in this paper.



(a) (b) (c)

Figure 1: **Problem (1)**. *(a) The feasible set (shaded region) is non-convex and delimited by circular and hyperbolic arcs. (b) Feasible set of the first convex LMI relaxation (blue region) is obtained by projecting the first-order moments onto the plane. The optimum is attained at the upper vertex (green dot). (c) Feasible set of the second convex LMI relaxation (blue region) is obtained by projecting the first-order moments onto the plane. The optimum equals the global optimum (green dot).*

2 Convex LMI Relaxations of Non-Convex Problems

2.1 Scalar Polynomial Optimization

There are several approaches to dealing with non-convexity in optimization problems. In [18] a technique is described which consists in building a hierarchy of nested convex linear matrix inequality (LMI) relaxations for non-convex optimization problems with scalar multivariate polynomial constraints. The relaxations are obtained by gradually adding lifting variables and constraints corresponding to linearizations of monomials up to a given degree. This is the technique we will adopt and we will exemplify the lifting idea below.

Under some mild assumptions akin to qualification constraints in mathematical programming, it is shown that the hierarchy of relaxations converges asymptotically to the global optimum. Convergence is proved using results of real algebraic geometry, namely the primal decomposition of a multivariate polynomial as a sum-of-squares, as well as the dual theory of moments of a measure localized on a semialgebraic set.

The LMI relaxation covering monomials up to a given even degree 2δ is referred to the LMI relaxation of order δ. It turns out that for many of the non-convex polynomial optimization problems described in the technical literature, global optima are reached at a given accuracy for a moderate number of lifting variables and constraints, hence for an LMI relaxation of moderate order. Standard routines of numerical linear algebra can be applied to provide a numerical certificate of global optimality based on computing ranks of moment matrices. In particular, *a sufficient condition* for reaching the global optimum is that the moment matrix has rank one.

As an illustrative example of the LMI relaxation technique, consider the non-convex optimization problem

$$\begin{array}{rcll} p^* & = & \max & x_2 \\ & & \text{s.t.} & g_1(x) = 3 + 2x_2 - x_1^2 - x_2^2 \geq 0 \\ & & & g_2(x) = -x_1 - x_2 - x_1 x_2 \geq 0 \\ & & & g_3(x) = 1 + x_1 x_2 \geq 0 \end{array} \tag{1}$$

where the linear objective function is maximized over a non-convex feasible set delimited by circular and hyperbolic arcs. The feasible region is shown in Figure 1(a). This example was originally described in [10].

The first LMI relaxation ($\delta = 1$) is

$$\begin{array}{ll} \max & y_{01} \\ \text{s.t.} & 3 + 2y_{01} - y_{20} - y_{02} \geq 0 \\ & -y_{10} - y_{01} - y_{11} \geq 0 \\ & 1 + y_{11} \geq 0 \\ & \begin{bmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{bmatrix} \succeq 0 \end{array}$$

with optimal value $p_1 = 2$. In this relaxation, the 3×3 positive semidefinite matrix (denoted by $\succeq 0$) is a moment matrix of order up to 2. Problem constraints are linearized with the help of lifting variables: a monomial $x_1^{k_1} x_2^{k_2}$ is replaced with $y_{k_1 k_2}$. Let $v_1(x) = [\,1, x_1, x_2\,]^\top$ be a basis for polynomials of degree 1. The moment matrix is obtained by linearizing the trivial relation $v_1(x)v_1(x)^\top \succeq 0$ valid for any $x \in \mathbf{R}^2$. Note that the matrix $v_1(x)v_1(x)^\top$ has rank one.

In Figure 1(b) we show the projection of the feasibility set of LMI relaxation onto the plane y_{10}, y_{01} of first-order moments. This convex feasibility set inscribes the original non-convex feasible set. We can see that the optimum of the LMI relaxation is achieved at a point that is infeasible for the original non-convex problem (1).

The second LMI relaxation ($\delta = 2$) is

$$\begin{array}{l} \max \quad y_{01} \\ \text{s.t.} \\ \begin{bmatrix} 3 + 2y_{01} - y_{20} - y_{02} & 3y_{10} + 2y_{11} - y_{30} - y_{12} & 3y_{01} + 2y_{02} - y_{21} - y_{03} \\ 3y_{10} + 2y_{11} - y_{30} - y_{12} & 3y_{20} + 2y_{21} - y_{40} - y_{22} & 3y_{11} + 2y_{12} - y_{31} - y_{13} \\ 3y_{01} + 2y_{02} - y_{21} - y_{03} & 3y_{11} + 2y_{12} - y_{31} - y_{13} & 3y_{02} + 2y_{03} - y_{22} - y_{04} \end{bmatrix} \succeq 0 \\ \begin{bmatrix} -y_{10} - y_{01} - y_{11} & -y_{20} - y_{11} - y_{21} & -y_{11} - y_{02} - y_{12} \\ -y_{20} - y_{11} - y_{21} & -y_{30} - y_{21} - y_{31} & -y_{21} - y_{12} - y_{22} \\ -y_{11} - y_{02} - y_{12} & -y_{21} - y_{12} - y_{22} & -y_{12} - y_{03} - y_{13} \end{bmatrix} \succeq 0 \\ \begin{bmatrix} 1 + y_{11} & y_{10} + y_{21} & y_{01} + y_{12} \\ y_{10} + y_{21} & y_{20} + y_{31} & y_{11} + y_{22} \\ y_{01} + y_{12} & y_{11} + y_{22} & y_{02} + y_{13} \end{bmatrix} \succeq 0 \\ \begin{bmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{bmatrix} \succeq 0 \end{array}$$

with optimal value $p_2 = 1.6180$, which is the global optimum p^* within numerical accuracy. In addition, first order moments $(y_{10}^*, y_{01}^*) = (-0.6180, 1.6180)$ provide an optimal solution of the original problem. This problem features a 6×6 moment matrix corresponding to moments of order up to 4, in addition to three 3×3 LMI constraints relaxing the polynomial inequality constraints. Let $v_2(x) = [\,1, x_1, x_2, x_1^2, x_1x_2, x_2^2\,]^\top$ be a basis for polynomials of degree 2. The 6×6 moment matrix constraint is obtained by linearizing the rank one matrix constraint $v_2(x)v_2(x)^\top \succeq 0$ (again, valid for any $x \in \mathbf{R}^2$) and the other LMI constraints correspond to linearizing $g_i(x)v_1(x)v_1(x)^\top \succeq 0$ for $i = 1, 2, 3$, where $g_i(x)$ denote the three inequality constraints in (1).

In Figure 1(c) we show the projection of the feasibility set of the second LMI relaxation onto the plane y_{10}, y_{01} of first-order moments. By construction, the feasibility set of the second LMI relaxation is included in the feasibility set of the first LMI relaxation. We can see that the feasibility set of the second LMI relaxation is the convex hull of the original non-convex feasible set, and the global optimum is now attained because the objective function is linear in the first-order moments.

2.2 Polynomial Matrix Optimization

In this paper we will face a particular class of polynomial optimization problems where the constraints are not scalar, but polynomial matrix inequalities (PMI). Moreover, only a limited subset of the decision variables enter in a non-linear and hence possibly non-convex way in the PMIs.

We can thus apply the same methodology as described in Section 2.1 to build up a hierarchy of nested LMI relaxations. The main difference however is that the LMI relaxations are obtained by linearizing the non-linear monomials only, a technique that we will refer to as a *partial relaxation*. On the one hand, this results in a dramatic drop in the number of LMI variables and constraints when compared with the full relaxation. On the other hand, in contrast with the scalar case, we are not able to ensure asymptotic convergence to the global optimum. However, if the moment matrix corresponding to this limited subset of non-convex variables has rank one, then we have a numerical certificate of global optimality just as in the scalar case.

Experimentally, it has been observed that minimizing the trace of the moment matrix generally results in a low rank moment matrix. In practice, in an LMI relaxation we add to the objective function the trace of the moment matrix weighted by a sufficiently small positive scalar ϵ. This heurestic is further elaborated on in Section 8.

3 Optimal Structure and Motion

Given image coordinates in one or several views, our goal is to infer the scene points and/or the camera motion. In this section, we formulate the problem and specify what is meant by an optimal estimate.

A perspective camera projects a point U in 3D space to a point u in the image plane as

$$\lambda u = PU. \tag{2}$$

Here the points are represented by homogeneous coordinates. In this expression, P is a rank-3 matrix of size 3×4 called the *camera matrix* and λ is a (positive) scalar accounting for depth.

Assume that the measured image points, denoted by $\hat{u}_i$, $i = 1, ..., N$, are corrupted by independent Gaussian noise, but otherwise, an ideal perspective camera model. Then the statistically optimal cost-function is the least-squares errors between measured and reprojected image points [7]. Hence, our goal is to solve the following optimization problem,

$$\min \sum_{i=1}^{N} d(\hat{u}_i, u_i(x))^2 \quad \text{s.t.} \quad \lambda_i(x) > 0, \tag{3}$$

where $d(\cdot, \cdot)$ is the Euclidean distance and x denotes the set of unknowns. Each term in the cost-function can be expressed as a rational function,

$$d(\hat{u}_i, u_i(x))^2 = \frac{f_{i1}(x)^2 + f_{i2}(x)^2}{\lambda_i(x)^2}, \tag{4}$$

where $f_{i1}(x)$, $f_{i2}(x)$ and $\lambda_i(x)$ are polynomials in x. In addition, one can require that $\lambda_i(x) > 0$, again, since all points should be in front of the cameras.

Minimizing the sum of rational functions can be achieved by reducing to the same denominator and applying the LMI relaxation technique described in [13], which is an extension of the technique of [18]. However, this approach is computationally demanding and not tractable if N is large, due to the high degree of the resulting denominator.

Instead, suppose each residual in (4) has an upper bound γ_i, that is, $(f_{i1}(x)^2 + f_{i2}(x)^2)/\lambda_i(x)^2 \leq \gamma_i$. Then, the formulation in (3) is equivalent to

$$\begin{array}{ll} \min & \gamma_1 + \gamma_2 + \ldots + \gamma_N \\ \text{s.t.} & f_{i1}(x)^2 + f_{i2}(x)^2 \leq \gamma_i \lambda_i(x)^2 \\ & \lambda_i(x) > 0 \qquad\qquad\qquad\qquad i = 1, ..., N. \end{array} \tag{5}$$

The hierarchy of LMI relaxations of [18] can be applied to this polynomial optimization problem. Denoting by δ the highest degree of x occurring in polynomials $f_{i1}(x)$, $f_{i2}(x)$, $\lambda_i(x)$, the constraints in the above problem have degree $2\delta+1$, meaning that the first LMI relaxation to be tried has order δ.

4 Schur Formulation

In this section we will show how the polynomial optimization problem in the previous section can be recast and relaxed using polynomial matrix inequalities.

Before we continue, we need to introduce a concept due to Schur [2]. Let

$$M = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}.$$

be a symmetric matrix and suppose that $A \succ 0$. Then, the following are equivalent:

$$M \succeq 0 \quad \Longleftrightarrow \quad C - B^\top A^{-1} B \succeq 0.$$

The matrix $C - B^\top A^{-1} B$ is called the *Schur complement* of M.

Now, set $A = \mathrm{diag}(\lambda_i(x)^2, \lambda_i(x)^2)$, $B = [\, f_{i1}(x), f_{i2}(x)\,]^\top$ and $C = \gamma_i$. It follows immediately that the Schur complement condition $C - B^\top A^{-1} B \succeq 0$ is equivalent to the inequality in (5) and hence we have the following reformulation:

$$\begin{array}{ll} \min & \gamma_1 + \gamma_2 + \ldots + \gamma_N \\ \text{s.t.} & \begin{bmatrix} \lambda_i(x)^2 & 0 & f_{i1}(x) \\ 0 & \lambda_i(x)^2 & f_{i2}(x) \\ f_{i1}(x) & f_{i2}(x) & \gamma_i \end{bmatrix} \succeq 0 \\ & \lambda_i(x) > 0 \qquad\qquad\qquad\qquad i = 1, ..., N. \end{array} \tag{6}$$

We will refer to this as the *Schur form*. Note that γ_i appears only as a linear term and the only non-linearity remaining is due to $\lambda_i(x)^2$ if the polynomials $f_{i1}(x)$ and $f_{i2}(x)$ are of degree one. Thus in order to take advantage of the result in Section 2.2, it is enough to apply LMI relaxations on $x = [x_1, x_2, ...]^T$ and it is not necessary to relax γ_i, $i = 1, ..., N$ provided the globality check is fulfilled for some relaxation order. If we were to apply full relaxations to all variables, the problem would become intractable for small N.

5 Example: Triangulation

We will first illustrate the convexification schemes on a simple geometric reconstruction problem, namely the triangulation problem. Then, other applications will be given in the next section.

In the triangulation problem, the camera matrices P_i, $i = 1, ..., N$ are considered to be known and the goal is to recover the unknown scene point $U = [\, x, 1\,]^\top = [\, x_1, x_2, x_3, 1\,]^\top$. It is easy to verify that the polynomials $f_{i1}(x)$, $f_{i2}(x)$ and $\lambda_i(x)$ in (4) have degree one and that the coefficients are determined by the elements in the camera matrix *and* the measured image coordinates.

As an example, consider the following camera triplet,

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, P_2 = \begin{bmatrix} -1 & -1 & -1 & 0 \\ 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \; P_3 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & -1 & 0 & 1 \end{bmatrix}$$

and assume that the measured image point in each view is at the origin (which is no restriction since it can be accomplished by changing coordinate systems). What is the optimal 3D point in terms of minimal reprojection errors?

The polynomial functions defined in (4) for the first camera are particularly simple $f_{11}(x) = x_1$, $f_{12}(x) = x_2$, $\lambda_1(x) = 1$. Note that the first residual is actually a polynomial. Introduce the squared upper bounds γ_2^2 and γ_3^2 for the second and third residuals[1]. Then, the scalar formulation (5) can be stated as

$$\begin{array}{ll} \min & x_1^2 + x_2^2 + \gamma_2^2 + \gamma_3^2 \\ \text{s.t.} & (x_1 + x_2 + x_3)^2 + (1 + x_1 - x_3)^2 \leq \gamma_2^2 (1 + x_3)^2 \\ & x_2^2 + (1 - x_3)^2 \leq \gamma_3^2 (1 - x_1 - x_2)^2. \end{array}$$

[1] Experimentally, we have found that replacing γ_i with γ_i^2 works better. Note that this does not increase the LMI relaxation order, which is still δ.

Form	RMS	Order	Moments	Variables
Polynomial	.362	2	21×21	125
	.181	3	56×56	461
	.162	4	126×126	1286
Schur	.175	1	4×4	12
	.164	2	10×10	37
	.162	3	21×21	86
Bundle adj.	.161	n.a.	n.a.	3

Table 1: Data from the triangulation example, from left to right: problem formulation, Root Mean Squares (RMS) errors in pixels, LMI relaxation order δ, the size of the moment matrix and the total number of decision variables.

The inequality constraints are of degree four and hence the lowest possible relaxation order is two. Relaxation is required for all five variables $(x_1, x_2, x_3, \gamma_2, \gamma_3)$. We have ignored the positive depth constraint, though, it would be straightforward to incorporate.

The Schur formulation (6) of this problem is given by

$$\begin{array}{ll} \min & \gamma_1 + \gamma_2 + \gamma_3 \\ \text{s.t.} & \begin{bmatrix} 1 & 0 & x_1 \\ 0 & 1 & x_2 \\ x_1 & x_2 & \gamma_1 \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} (1+x_3)^2 & 0 & -x_1 - x_2 - x_3 \\ 0 & (1+x_3)^2 & 1 + x_1 - x_3 \\ -x_1 - x_2 - x_3 & 1 + x_1 - x_3 & \gamma_2 \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} (1 - x_1 - x_2)^2 & 0 & -x_2 \\ 0 & (1 - x_1 - x_2)^2 & 1 - x_3 \\ -x_2 & 1 - x_3 & \gamma_3 \end{bmatrix} \succeq 0. \end{array}$$

Similar to problem (1), the LMI relaxed formulation is obtained by the lifting procedure: a monomial $x_1^{k_1} x_2^{k_2} x_3^{k_3}$ is replaced with the lifting variable $y_{k_1 k_2 k_3}$ and adding that the moment matrix should be positive semidefinite.

The results of the two LMI problems are summarized in Table 1. The best estimate of the polynomial formulation was also refined using bundle adjustment[2], resulting in the following 3D point estimates for the three different schemes:

$$U_{poly} = \begin{bmatrix} \text{-.176} \\ \text{-.110} \\ .780 \\ 1 \end{bmatrix}, U_{Schur} = \begin{bmatrix} \text{-.182} \\ \text{-.138} \\ .826 \\ 1 \end{bmatrix}, U_{bundle} = \begin{bmatrix} \text{-.181} \\ \text{-.113} \\ .813 \\ 1 \end{bmatrix}.$$

Examining the moment matrix of the estimates, the ratio of the two largest singular values, σ_2/σ_1, is 0.03 and 0.3 for the polynomial method and the Schur form, respectively. Thus, the moment matrices are close to rank one which would guarantee global optimality, cf. Section 2. More simulations are given in Section 7.

Remark. It is important to note that in the polynomial scheme, the number of decision variables increases drastically while the increase in the Schur form is more moderate. The complexity of the polynomial approach for problems with (i) more than three views *or* (ii) more degrees of freedom is computationally very demanding. Therefore, we will focus on the more promising Schur method in the remaining part of the paper.

[2]Bundle adjustment also optimizes the sum-of-squares cost-function, but it is based on iterative, gradient descent minimization.

6 More Applications

6.1 Camera Pose

In the problem of camera pose, also known as camera resectioning or absolute orientation, the camera matrix is the object of interest. Given a set of known 3D points $\{U_i\}_{i=1}^N$ in the scene, the goal is to reconstruct the camera matrix P. Let

$$P = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ x_5 & x_6 & x_7 & x_8 \\ x_9 & x_{10} & x_{11} & 1 \end{bmatrix}.$$

The polynomials $f_{i1}(x)$ $f_{i2}(x)$ and $\lambda_i(x)$ will again be affine functions (that is, having degree one) and the coefficients are determined by the scene points U_i and the measured points $\hat{u}_i$. While $f_{i1}(x)$ and $f_{i2}(x)$ will generally depend on all 11 variables in x, the depth function $\lambda_i(x)$ depends only on three variables, namely $\lambda_i(x) = \lambda_i(x_9, x_{10}, x_{11})$. Thus, in the optimization process, it will suffice to do partial relaxations on these three variables, as described in Section 2.2. If all 11 variables were to be relaxed, then the problem would have become computationally impossible, or at least hard, already for relaxation orders greater than two.

6.2 Homography Estimation

A homography is a projective transformation from $\mathcal{P}^n$ to $\mathcal{P}^n$. The problem of estimating a homography is similar to that of camera pose. For example, suppose we are given a collection of 3D points $\{U_i\}_{i=1}^N$ on a plane, then there exists a homography, which can be represented by a 3×3 matrix H, mapping these points to the image plane, as $\lambda_i u_i = H U_i$ where U_i are homogeneous plane coordinates. Hence, by setting

$$H = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & 1 \end{bmatrix},$$

the problem can be put in the standard form (6) with LMI relaxations on x_7 and x_8.

More generally, if we are given two collections of points in $\mathcal{P}^n$, then we can compute the homography $H : \mathcal{P}^n \mapsto \mathcal{P}^n$, mapping one set to the other. All measurement errors will be assumed to be in one of the point sets and it will suffice to relax the variables appearing in the last row of the homography matrix in accordance with the principle of partial relaxation, cf. Section 2.2.

In the case of a plane-to-image homography as described above, it makes sense to speak of the optimal homography. However, for other problems involving projective transformations, it may not be the best choice to optimize the L_2-norm. For example, for an inter-image homography it would be better to use a symmetrical cost-function. We do not pursue the topic of general homographies from $\mathcal{P}^n$ to $\mathcal{P}^n$ any further.

6.3 Epipolar Geometry

Given corresponding image points in two views, we could in principle follow the same strategy as before in order to reconstruct both camera matrices and scene structure. But this would unfortunately lead to an intractable problem since there are simply too many variables that would appear in non-linear polynomials.

Therefore, we will reformulate the problem once again. Given corresponding points u and u' in two images, the epipolar constraint should be fulfilled:

$$u'^T F u = 0,$$

where F is a 3×3 matrix of rank two. Given F, one can recover uniquely two camera matrices modulo projective coordinate system [7]. In [28], the following optimization criterion was analyzed:

$$\begin{array}{ll} \min & \sum_i \frac{(\hat{u}_i'^T F \hat{u}_i)^2}{(F\hat{u}_i)_1^2 + (F\hat{u}_i)_2^2 + (F^\top \hat{u}_i')_1^2 + (F^\top \hat{u}_i')_2^2} \\ \text{s.t.} & \det F = 0. \end{array}$$

It was shown that the criterion can be regarded as a first order approximation of the optimal two-view structure and motion problem. Moreover, for certain motion configurations, it is even equivalent. From a practical perspective, the reconstructed motion using the criterion was very close to the statistically optimal one.

Analogously to the derivation in Section 4, let γ_i be an upper bound on the ith residual term. Using a Schur complement argument, the problem can be reformulated as,

$$\begin{array}{ll} \min & \gamma_1 + \gamma_2 + \ldots + \gamma_N \\ \text{s.t.} & \begin{bmatrix} (F\hat{u}_i)_1^2 + (F\hat{u}_i)_2^2 + (F^\top \hat{u}'_i)_1^2 + (F^\top \hat{u}'_i)_2^2 & \hat{u}'^T_i F \hat{u}_i \\ \hat{u}'^T_i F \hat{u}_i & \gamma_i \end{bmatrix} \succeq 0 \\ & \det F = 0 \qquad\qquad\qquad\qquad i = 1, ..., N. \end{array}$$

Finally, by parametrizing the fundamental matrix by

$$F = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & 1 \end{bmatrix},$$

the problem is given in a Schur form. As the determinant constraint is cubic in x, a relaxation order of at least two is required. All elements in x appear in non-linear expressions and hence all eight variables need to be relaxed. In addition to $\det F = 0$ one can add $x_i \det F = 0, i = 1, ..., 8$ without increasing the complexity and to tighten the LMI relaxations of the non-convex problem.

7 Experimental Validation

The proposed approach for geometric reconstruction problems has been validated on both simulated and real data. The goal has been to determine if a global estimate is actually obtained and if so, at what accuracy. This is not an easy task, however, since there are no other independent methods to compute the global estimate.

We have compared our algorithms with standard linear techniques as well as bundle adjustment [7]. In all experiments, the bundle adjustment has been initialized with (i) our method, (ii) the linear algorithm and when available, (iii) the synthetically generated ground truth[3]. Out of these bundle results, only the one with lowest reprojection error is reported. For easier comparison, the Root Mean Squares (RMS) errors are given instead of the sum of squares errors.

7.1 Implementation Details

All the described reconstruction algorithms have been implemented under the Matlab environment in the publicly available package GloptiPoly [9] using the conic programming solver SeDuMi [24]. The computation times (i.e. the cputime for the solver SeDuMi) vary from .23 s for three-view triangulation to 3.4 s for epipolar geometry with 104 correspondences on a Pentium 4 with 2.8 GHz. In all experiments but the last one, an LMI relaxation order equal to 3 has been used with $\epsilon = 0.001$ for the trace of the moment matrix, cf. Section 2.2. For the estimation of epipolar geometry, the corresponding numbers are LMI order 2 and $\epsilon = 0.00001$. These settings have been found empirically. There are many tuning parameters in the algorithm that can significantly impact on the performance and accuracy (including all the tuning parameters of the SeDuMi solver). It is out of the scope of this paper to give a comprehensive description of the respective and relative influences of all these parameters.

Normalization of the input coordinates is essential, both for the linear algorithms and for the Schur method. The moment matrices contain lifting variables of high-order monomials making the optimization sensitive to data scaling. This preprocessing step has been done by changing coordinate systems in order to get coordinates with magnitude around one, see [7]. In the Schur formulation, only transformations invariant to the cost-function are applied. Numerical experiments reveal that

[3]The ground truth gives only zero reprojection error when no image noise is added.

the numerical behavior of conic programming solvers is significantly improved (in terms of convergence speed and quality of the computed optimizers) when moment matrices features entries with magnitude less than one [11].

In all parametrizations of the unknowns, one element in a homogeneous vector is dehomogenized, normally the last element. This may cause numerical problems when the last element happens to be close to zero. On the other hand, the situation is detectable via, for instance, the global optimality check and one can then dehomogenize another element in the vector and rerun the algorithm.

7.2 Synthetic Data

All simulated data was generated in the following manner. Uniformly random 3D points with coordinates within $[-1, 1]$ units were projected to cameras with focal lengths of 1 pixel. The camera centers were (randomly) chosen at distances of 5 units from the origin in average. The cameras' viewing directions were also random, though biased towards the origin. In addition, the image coordinates were corrupted by zero-mean Gaussian noise with varying levels of s.t.d. This procedure typically gives coordinates with absolute values less than a pixel.

We have tested the Schur formulations for triangulation and camera pose on simulated data. The results are presented in Figure 2 and the graphs show the average result of 500 repetitions. The behavior of the two schur algorithms relative the other methods are similar when noise levels and the number of points/views, respectively, are varied.

In Figures 2(a) and (e), one can see that the errors for the Schur formulations follow very close to that of the best obtained with bundle adjustment. Recall that the bundle adjustment is initialized with the Schur estimate and the linear estimate as well as the ground truth. The result with the lowest RMS error is kept. The linear algorithm yields worse estimates, as expected.

In Figures 2(b) and (f), the percentage of times the refined estimates (Schur or linear) equal the estimate of the (best) optimum, obtained by applying bundle adjustment. Generally, the Schur estimates attain the optimum and the moment matrices are close to rank one. Hence, it is likely that the global optimum is retrieved. The refined linear estimates risk getting stuck in local optima.

When the number of views or points are varied, the Schur algorithms also closely follow the bundle result, see Figures 2(c) and (g), and generally the optimum is attained, cf. Figures 2(d) and (h). As more views or points are taken into account, the differences between the linear and the Schur methods decrease.

7.3 Real Data

We have worked with two publicly available sequences with given point correspondences[4] to test the performance on real image data. The first one is a corridor sequence with a forward-moving camera motion consisting 104 correspondences visible in all 11 images. The second one is a turn-table sequence of a dinosaur with 36 images and in total 328 points with lots of occlusions, cf. Figure 3.

Out of the 104 correspondences in the corridor sequence, 23 points lie on the left, frontal wall and hence the points should be coplanar in space. These points were used to compute inter-image homographies between consecutive views[5]. In Figure 4(a), the errors are shown. The linear algorithm performs well, but generally the Schur algorithm performs better and it has similar performance as bundle adjustment.

The two-view epipolar geometries for consecutive images in the corridor and the dinosaur sequences have also been computed. In the corridor sequence, all 104 correspondences were used and the RMS errors are presented in Figure 4(b). The epipolar lines for the first image pair are illustrated in Figure 3. The performance of the Schur method is again comparable to the result of bundle adjustment.

We have also computed the epipolar geometries for consecutive views in the dinosaur sequence, cf. Figure 3. The average RMS errors for the 35 image pairs are .209, .209 and .201 for our method,

[4] Available at http://www.robots.ox.ac.uk/~vgg/data.html.

[5] Our optimization criterion is not the best choice in this case, since it is implicitly assumed that there is no noise in one image.

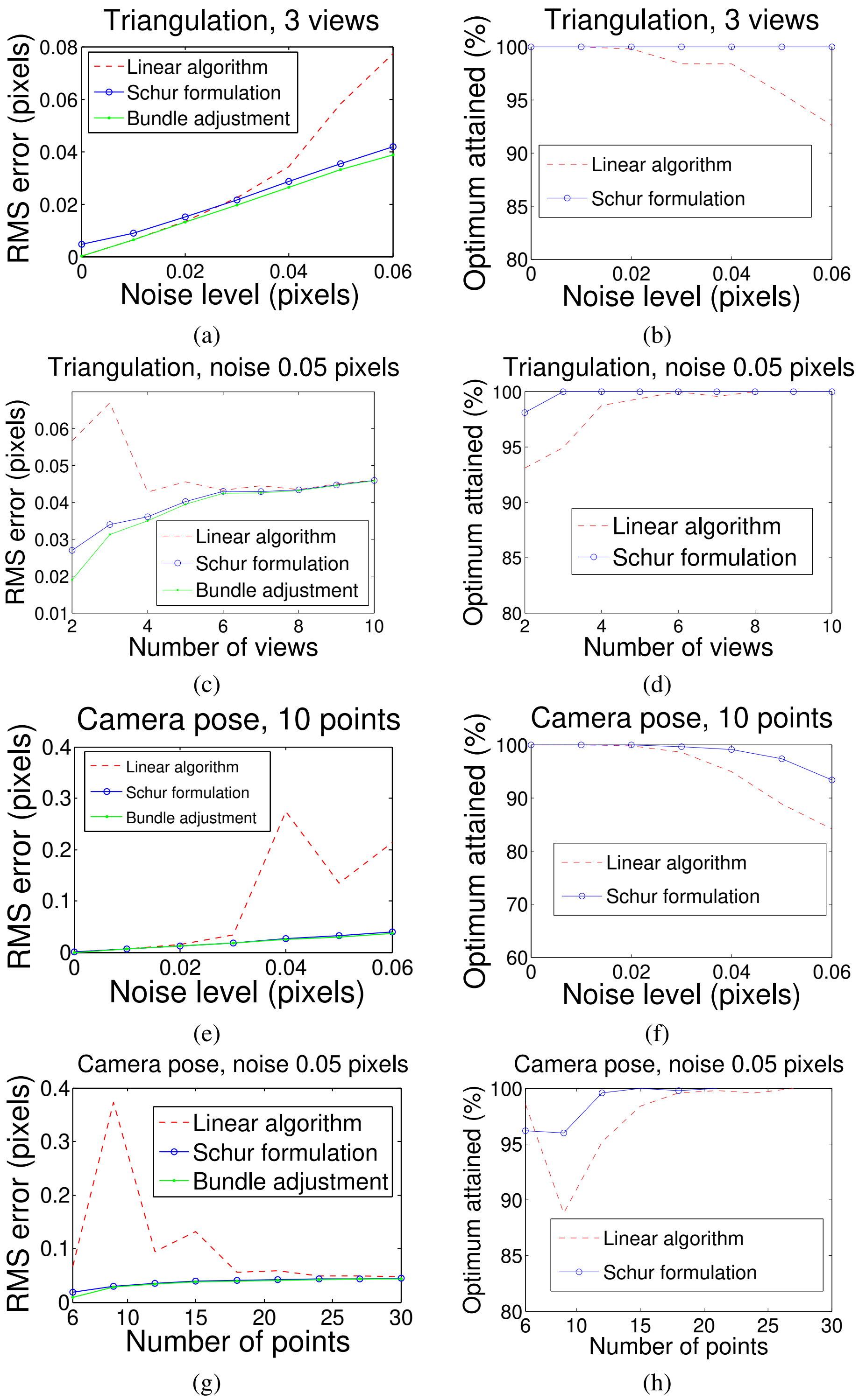
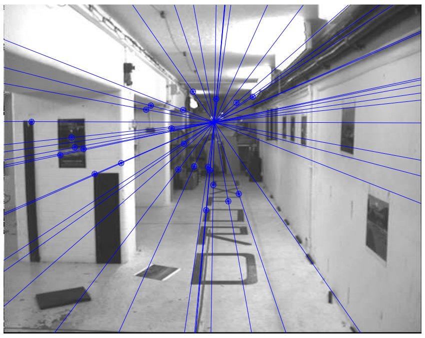
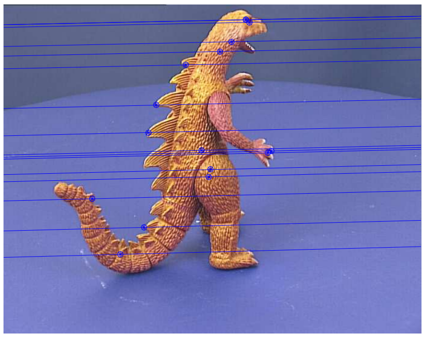


Figure 2: Triangulation: (a)-(d). Camera pose: (e)-(h). See text for details.

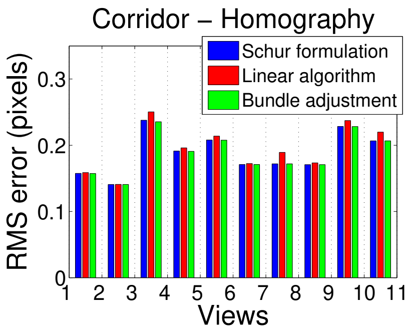


(a) Corridor

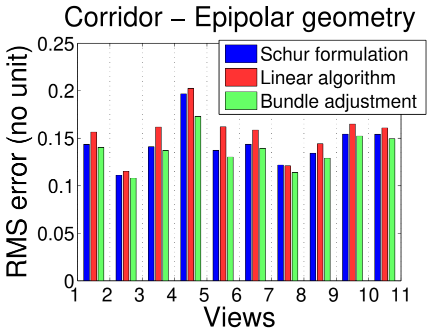


(b) Dinosaur

Figure 3: First images of the corridor (a) and dinosaur (b) sequences with epipolar lines with respect to the second view. While the first camera is moving forwards (or backwards), the other one is moving sideways.



(a)



(b)

Figure 4: Estimation results of homographies (a) and two-view epipolar geometries (b) for consecutive images in the corridor sequence.

the 8-point algorithm and bundle adjustment, respectively. Hence, the three methods are similar in performance for this scene.

We have noticed that the Schur algorithm for epipolar geometry estimation is more sensitive to data scaling than for the other tasks. Also, it is usually not good practice to dehomogenize the last element in the fundamental matrix since it will vanish if the optical axes intersect. Due to this sensitivity, one may have to rerun the Schur algorithm several times if an element in the fundamental matrix is dehomogenized which has small magnitude. A priori, it may be hard to say which element has the largest magnitude, or more importantly, which element gives the most accurate result. Such numerical aspects need to be further investigated. A possible reason for this sensitivity is that there are more variables involved in non-linear expressions compared to the other applications.

8 Discussion

The area of geometric reconstruction problems is a mature field and state-of-the-art methods are quite sophisticated. For example, linear methods with proper data normalization perform often satisfactorily for low noise levels. Still, our approach gives generally better estimates, particularly, for higher noise levels. In fact, our experiments indicate that the estimates are very close to the global optimum and that the risk of getting stuck in a local minima is small.

We have extended the existing theory of LMI relaxations of scalar polynomial optimization problems to matrix polynomial optimization problems. Instead of linearizing all the monomials, we used partial relaxations, i.e. we considered only a limited subset of variables corresponding to non-linear, hence potentially non-convex terms in the PMIs. In general, these structure-exploiting partial relaxations are of smaller size than the full relaxations described in [18]. On the negative side, contrary to the full relaxations, we cannot guarantee asymptotic convergence to the global optimum. However, practice reveals that the moment matrix has numerical rank close to one for LMI relaxations of moderate order, which ensures global optimality in most cases. Recently, the asymptotically converging hierarchy of LMI relaxations - originally proposed for scalar polynomial optimization in [18] - has been extended to PMI problems, using sum-of-squares decompositions of polynomial matrices [12]. Better convergence properties and a more favorable numerical behavior of the conic programming solver are expected since these new relaxations exploit the particular matrix structure. Note however that these new relaxations have not yet been applied to the PMI problems described in our paper.

Several numerical aspects related with these LMI relaxations deserve to be studied in further detail. Since the LMI relaxations are built on moment matrices which may feature monomials of relatively high order, the use of alternative polynomial bases (Chebyshev, orthogonal polynomials) may be interesting. Appropriate definitions of numerical analysis concepts such as conditioning or scaling of these moment matrices would also be required in this context. Moreover, evaluating the rank of a numerical matrix is a difficult task. The underlying numerical analysis problem is ill-posed, in the sense that the rank function maps a continuous set (reals) onto a discrete set (integers) and a vanishing perturbation can affect the output. Evaluating the rank requires to set up an arbitrary threshold on the eigenvalues (absolute or relative) - a sensitive task.

Interior-point algorithms used in most of the conic programming solvers (e.g. in SeDuMi) to solve LMI problems consist in applying a Newton scheme (on a Lagrangian built via a suitably defined barrier function) yielding iterates in the interior of the feasibility region, i.e. in the interior of the cone of positive semidefinite matrices. Only at the optimum, the iterate may reach the boundary of the cone, resulting in rank-deficient positive semidefinite matrices. However, the final iterate has typically maximum rank amongst all possible optimizers. This may be in conflict with our expectations of a low-rank (ideally rank-one) moment matrix allowing to guarantee global optimality [11]. Unfortunately, it is known that rank-minimization under LMI constraints is a difficult, non-convex optimization problem in general. As already mentioned, numerical experiments reveal that minimizing the trace of the moment matrix generally results in low-rank optimizers, but this is only a heuristic.

Acknowledgements. This work has benefited from many discussions with J. B. Lasserre, Manmohan Chandraker and Sameer Agarwal. Financial support for F. Kahl was provided by the U.C. Micro Program, the Swedish Research Council (VR 2004-4579) and European Commission (Grant 011838, SMERobot). D. Henrion's research was partly supported by project MOGA NT05-3-41612 of the French National Research Agency (ANR), by projects 102/05/0011 and 102/06/0652 of the Grant Agency of the Czech Republic and project ME 698/2003 of the Ministry of Education of the Czech Republic.

References

[1] S. Agarwal, M. Chandraker, F. Kahl, S. Belongie, and D.J. Kriegman. Practical global optimization for multiview geometry. In *European Conf. Computer Vision*, Graz, Austria, 2006.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] G. Chesi, A. Garullli, A. Vicino, and R. Cipolla. Estimating the fundamental matrix via constrained least-squares: A convex approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(3):397–401, 2002.

[4] A. Fusiello, A. Benedetti, M. Farenzena, and A. Busti. Globally convergent autocalibration using interval analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(12):1633–1638, 2004.

[5] R. Hartley and F. Schaffalitzky. L-∞ minimization in geometric reconstruction problems. In *Conf. Computer Vision and Pattern Recognition*, volume I, pages 504–509, Washington DC, USA, 2004.

[6] R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. Second Edition.

[8] D. Henrion and A. Garulli, editors. *Positive Polynomials in Control*. Lecture Notes in Control and Information Sciences. Springer-Verlag, 2005.

[9] D. Henrion and J. B. Lasserre. GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi. *ACM Trans. Math. Soft.*, 29(2):165–194, 2003.

[10] D. Henrion and J. B. Lasserre. Solving nonconvex optimization problems - how GloptiPoly is applied to problems in robust and nonlinear control. *IEEE Control Systems Magazine*, 24(3):72–83, 2004.

[11] D. Henrion and J. B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. In *Positive Polynomials in Control*. Springer-Verlag, 2005.

[12] D. Henrion and J. B. Lasserre. Convergent relaxations of polynomial matrix inequalities and static output feedback. *IEEE Trans. Automatic Control*, 51(2):192–202, 2006.

[13] D. Jibetean and E. de Klerk. Global optimization of rational functions: a semidefinite programming approach. *Mathematical Programming*, 106(1):93–109, 2006.

[14] F. Kahl. Multiple view geometry and the L_∞-norm. In *Int. Conf. Computer Vision*, pages 1002–1009, Beijing, China, 2005.

[15] F. Kahl and D. Henrion. Globally optimal estimates for geometric reconstruction problems. In *Int. Conf. Computer Vision*, pages 978–985, Beijing, China, 2005.

[16] F. Kahl, A. Heyden, and L. Quan. Minimal projective reconstruction including missing data. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(4):418–424, 2001.

[17] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *European Conf. Computer Vision*, volume III, pages 82–96, Copenhagen, Denmark, 2002.

[18] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optimization*, 11:796–817, 2001.

[19] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer Verlag, 2003.

[20] D. Nistér. *Automatic dense reconstruction from uncalibrated video sequences*. PhD thesis, Royal Institute of Technology KTH, Sweden, 2001.

[21] J. Oliensis. Exact two-image structure from motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(12):1618–1633, 2002.

[22] S. Soatto and R. Brockett. Optimal structure from motion: Local ambiguities and global estimates. In *Conf. Computer Vision and Pattern Recognition*, Santa Barbara, USA, 1998.

[23] H. Stewénius, F. Schaffalitzky, and D. Nistér. How hard is three-view triangulation really? In *Int. Conf. Computer Vision*, pages 686–693, Beijing, China, 2005.

[24] J.F. Sturm. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11-12:625–653, 1999.

[25] R. Szeliski and S. B. Kang. Shape ambiguities in structure from motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(5), May 1997.

[26] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. Journal Computer Vision*, 9(2):137–154, 1992.

[27] B. Triggs, P.F. McLauchlan, Hartley R.I, and A.W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Vision Algorithms'99*, pages 298–372, in conjunction with ICCV'99, Kerkyra, Greece, 1999.

[28] Z. Zhang. On the optimization criteria used in two-view motion analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(7):717–729, 1998.