



**HAL**  
open science

# Robust Estimation for an Inverse Problem Arising in Multiview Geometry

Arnak S. Dalalyan, Renaud Keriven

► **To cite this version:**

Arnak S. Dalalyan, Renaud Keriven. Robust Estimation for an Inverse Problem Arising in Multiview Geometry. *Journal of Mathematical Imaging and Vision*, 2012, 43 (1), pp.10-23. 10.1007/s10851-011-0281-3 . hal-00415603v2

**HAL Id: hal-00415603**

**<https://hal.science/hal-00415603v2>**

Submitted on 25 Mar 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Robust Estimation for an Inverse Problem Arising in Multiview Geometry

Arnak Dalalyan · Renaud Keriven

Received: date / Accepted: date

**Abstract** We propose a new approach to the problem of robust estimation for an inverse problem arising in multiview geometry. Inspired by recent advances in the statistical theory of recovering sparse vectors, we define our estimator as a Bayesian maximum a posteriori with multivariate Laplace prior on the vector describing the outliers. This leads to an estimator in which the fidelity to the data is measured by the  $L_\infty$ -norm while the regularization is done by the  $L_1$ -norm. The proposed procedure is fairly fast since the outlier removal is done by solving one linear program (LP). An important difference compared to existing algorithms is that for our estimator it is not necessary to specify neither the number nor the proportion of the outliers; only an upper bound on the maximal measurement error for the inliers should be specified. We present theoretical results assessing the accuracy of our procedure, as well as numerical examples illustrating its efficiency on synthetic and real data.

**Keywords** Structure from motion · Sparse recovery · Robust estimation ·  $L_1$ -relaxation

## 1 Introduction

In the present paper, we are concerned with a nonlinear inverse problem appearing in the structure from motion problem of multiview geometry. This problem,

that have received a great deal of attention by the computer vision community in last decade, consists in recovering a set of 3D points (structure) and a set of camera matrices (motion), when only 2D images of the aforementioned 3D points by some cameras are available. Throughout this work we assume that the internal parameters of cameras as well as their orientations are known. Thus, only the locations of camera centers and 3D points are to be estimated. In solving the structure from motion problem by state-of-the-art methods, it is customary to start by establishing correspondences between pairs of 2D data points. We will assume in the present study that these point correspondences have been already established.

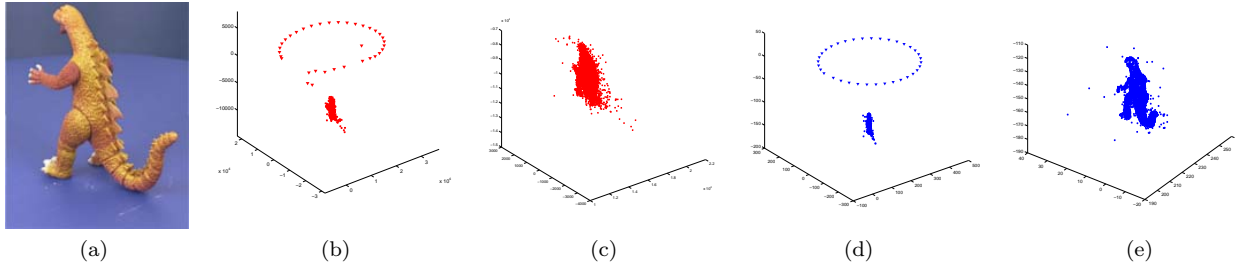
One can think of the structure from motion problem as the inverse problem of inverting the operator  $\mathcal{O}$  that takes as input the set of 3D points and the set of cameras, and produces as output the 2D images of the 3D points by the cameras. This approach will be further formalized in the next section. Generally, the operator  $\mathcal{O}$  is not injective, but in many situations (for example, when for each pair of cameras there are at least five 3D points in general position that are seen by these cameras [20]), there is only a small number of inputs, up to an overall similarity transform, having the same image by  $\mathcal{O}$ . In such cases, the solutions to the structure from motion problem can be found using algebraic arguments.

The main flaw of algebraic solutions is their sensitivity to the noise in the data: very often, because of the noise in the measurements, there is no input that could have generated the observed output. A natural approach to cope with such situations consists in searching for the input providing the closest possible output to the observed data. Then, a major issue is how to choose the metric in the output space. A standard ap-

---

A. Dalalyan  
Université Paris Est  
IMAGINE, Ecole des Ponts ParisTech  
Marne-la-Vallée, France  
E-mail: dalalyan@imagine.enpc.fr

R. Keriven  
E-mail: keriven@imagine.enpc.fr



**Fig. 1** (a) One image from the dinosaur sequence. (b)-(c) Camera locations and scene points estimated by the blind  $L_\infty$ -cost minimization. (d)-(e) Camera locations and scene points estimated by the proposed “outlier aware” procedure. This figure demonstrates that the estimator minimizing the  $L_\infty$ -cost is severely affected by the outliers.

proach [12] consists in measuring the distance between two elements of the output space in the Euclidean  $L_2$ -norm. In the structure from motion problem with more than two cameras, this leads to a hard non-convex optimization problem. A particularly elegant way of circumventing the non-convexity issues inherent to the use of  $L_2$ -norm consists in replacing it by the  $L_\infty$ -norm [11, 14, 22, 24, 26, 25, 23]. It has been shown that, for a number of problems,  $L_\infty$ -norm based estimators can be computed very efficiently using, for example, the iterative bisection method [14, Algorithm 1, p. 1608] that solves a convex program at each iteration. There is however an issue with the  $L_\infty$ -techniques that dampens the enthusiasm of practitioners: it is highly sensitive to outliers (*cf.* Fig. 1). In fact, among all  $L_q$ -metrics with  $q \geq 1$ , the  $L_\infty$ -metric is the most seriously affected by the outliers in the data. Two procedures have been introduced [26, 15] that make the  $L_\infty$ -estimator less sensitive to outliers. Although these procedures demonstrate satisfactory empirical performance, they suffer from a lack of sufficient theoretical support assessing the accuracy of produced estimates.

The purpose of the present work is to introduce and to theoretically investigate a new procedure of estimation in presence of noise and outliers. Our procedure combines  $L_\infty$ -norm for measuring the fidelity to the data and  $L_1$ -norm for regularization. It can be seen as a maximum a posteriori (MAP) estimator under uniformly distributed random noise and a sparsity favoring prior on the vector of outliers. Interestingly, this study bridges the work on the robust estimation in multiview geometry [9, 26, 15, 17] and the theory of sparse recovery in statistics and signal processing [7, 1, 3, 4]. Furthermore, since the estimator we propose solves the same convex program as that solved at each step of iteration of the procedure in [15], the theoretical arguments developed in the present work provide an explanation to the nice empirical performance of Kanade and Ke’s

procedure. Moreover, our procedure is complementary to that of Kanade and Ke, since the free parameter for our procedure is the maximal reprojection error of inliers and not the presumed number of outliers.

The rest of the paper is organized as follows. The next section gives the precise formulation of the translation estimation and triangulation problem to which the presented methodology can be applied. A brief review of the  $L_\infty$ -norm minimization algorithm is presented in Section 3. In Section 4, we introduce the statistical framework and derive a new procedure as a MAP estimator. Main results assessing the accuracy of this procedure are stated in Section 5, while Section 6 is devoted to a discussion on the relations of our methodology with some relevant recent studies. Section 7 contains numerical experiments supporting our theoretical results. The methodology of our study is summarized in Section 8 and the technical proofs are gathered in Section 9.

## 2 Translation estimation and triangulation

Let us start by presenting a problem of multiview geometry to which our approach can be successfully applied, namely the problem of translation estimation and triangulation in the case of known rotations. For rotation estimation algorithms, we refer the interested reader to [19, 10] and the references therein.

Let  $\mathbf{P}_i^*$ ,  $i = 1, \dots, m$ , be a sequence of  $m$  cameras that are known up to a translation. Recall that a camera is characterized by a  $3 \times 4$  matrix  $\mathbf{P}$  with real entries that can be written as  $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ , where  $\mathbf{K}$  is an invertible  $3 \times 3$  matrix called the camera calibration matrix,  $\mathbf{R}$  is a  $3 \times 3$  rotation matrix and  $\mathbf{t} \in \mathbb{R}^3$ . We will refer to  $\mathbf{t}$  as the translation of the camera  $\mathbf{P}$ . We can thus write  $\mathbf{P}_i^* = \mathbf{K}_i[\mathbf{R}_i|\mathbf{t}_i^*]$ ,  $i = 1, \dots, m$ . For a set of unknown scene points  $\mathbf{U}_j^*$ ,  $j = 1, \dots, n$ , expressed in homogeneous co-

ordinates (*i.e.*,  $\mathbf{U}_j^*$  is an element of the projective space  $\mathbb{P}^3$ ), we assume that noisy images of each  $\mathbf{U}_j^*$  by some cameras  $\mathbf{P}_i^*$  are observed. Thus, we have at our disposal the measurements

$$\mathbf{x}_{ij} = \frac{1}{\mathbf{e}_3^\top \mathbf{P}_i^* \mathbf{U}_j^*} \begin{bmatrix} \mathbf{e}_1^\top \mathbf{P}_i^* \mathbf{U}_j^* \\ \mathbf{e}_2^\top \mathbf{P}_i^* \mathbf{U}_j^* \end{bmatrix} + \boldsymbol{\xi}_{ij}, \quad j = 1, \dots, n, \quad i \in I_j, \quad (1)$$

where  $\mathbf{e}_\ell$ ,  $\ell = 1, 2, 3$ , stands for the unit vector of  $\mathbb{R}^3$  having one as the  $\ell^{\text{th}}$  coordinate and  $I_j$  is the set of indices of cameras for which the point  $\mathbf{U}_j^*$  is visible. We assume that the set  $\{\mathbf{U}_j^*\}$  does not contain points at infinity:  $\mathbf{U}_j^* = [\mathbf{X}_j^{*\top} | 1]^\top$  for some  $\mathbf{X}_j^* \in \mathbb{R}^3$  and for every  $j = 1, \dots, n$ .

We are now in a position to state the problem of translation estimation and triangulation in the context of multiview geometry. It consists in recovering the 3-vectors  $\{\mathbf{t}_i^*\}$  (translation estimation) and the 3D scene points  $\{\mathbf{X}_j^*\}$  (triangulation) from the noisy measurements  $\{\mathbf{x}_{ij}; j = 1, \dots, n; i \in I_j\} \subset \mathbb{R}^2$ . In what follows, we use the notation  $\boldsymbol{\theta}^* = (\mathbf{t}_1^{*\top}, \dots, \mathbf{t}_m^{*\top}, \mathbf{X}_1^{*\top}, \dots, \mathbf{X}_n^{*\top})^\top$ . Thus, we are interested in estimating  $\boldsymbol{\theta}^* \in \mathbb{R}^{3(m+n)}$ .

*Remark 1 (Cheirality)* It should be noted right away that if the point  $\mathbf{U}_j^*$  is in front of the camera  $\mathbf{P}_i^*$ , then  $\mathbf{e}_3^\top \mathbf{P}_i^* \mathbf{U}_j^* \geq 0$ . This is termed cheirality condition. Furthermore, we will assume that none of the true 3D points  $\mathbf{U}_j^*$  lies on the principal plane of a camera  $\mathbf{P}_i^*$ . This assumption implies that  $\mathbf{e}_3^\top \mathbf{P}_i^* \mathbf{U}_j^* > 0$  so that the quotients  $\mathbf{e}_\ell^\top \mathbf{P}_i^* \mathbf{U}_j^* / \mathbf{e}_3^\top \mathbf{P}_i^* \mathbf{U}_j^*$ ,  $\ell = 1, 2$ , are well defined.

*Remark 2 (Identifiability)* The parameter  $\boldsymbol{\theta}$  we have just defined is, in general, not identifiable from the measurements  $\{\mathbf{x}_{ij}\}$ . In fact, one easily checks that, for every  $\alpha \neq 0$  and for every  $\mathbf{t} \in \mathbb{R}^3$ , the parameters  $\{\mathbf{t}_i^*, \mathbf{X}_j^*\}$  and  $\{\alpha(\mathbf{t}_i^* - \mathbf{R}_i \mathbf{t}), \alpha(\mathbf{X}_j^* + \mathbf{t})\}$  generate the same measurements. To cope with this issue, we assume that  $\mathbf{t}_1^* = \mathbf{0}_3$  and that  $\min_{i,j} \mathbf{e}_3^\top \mathbf{P}_i^* \mathbf{U}_j^* = 1$ . Thus, in what follows we assume that  $\mathbf{t}_1^*$  is removed from  $\boldsymbol{\theta}^*$  and  $\boldsymbol{\theta}^* \in \mathbb{R}^{3(m+n-1)}$ . Further assumptions ensuring the identifiability of  $\boldsymbol{\theta}^*$  are given below.

### 3 Estimation by Sequential Convex Programming

This section presents results on the estimation of  $\boldsymbol{\theta}$  based on the reprojection error (RE) minimization. This material is essential for understanding the results that are at the core of the present work. In what follows, for every  $s \geq 1$ , we denote by  $\|\mathbf{x}\|_s$  the  $L_s$ -norm of a vector  $\mathbf{x}$ , *i.e.*,  $\|\mathbf{x}\|_s^s = \sum_j |x_j|^s$  if  $\mathbf{x} = (x_1, \dots, x_d)^\top$ . As usual, we extend this to  $s = +\infty$  by setting  $\|\mathbf{x}\|_\infty = \max_j |x_j|$ .

A classical method [12] for estimating the parameter  $\boldsymbol{\theta}$  is based on minimizing the sum of the squared

REs. This defines the estimator  $\widehat{\boldsymbol{\theta}}$  as a minimizer of the cost function  $\mathcal{C}_{2,2}(\boldsymbol{\theta}) = \sum_{i,j} \|\mathbf{x}_{ij} - \mathbf{x}_{ij}(\boldsymbol{\theta})\|_2^2$ , where  $\mathbf{x}_{ij}(\boldsymbol{\theta}) := [\mathbf{e}_1^\top \mathbf{P}_i \mathbf{U}_j; \mathbf{e}_2^\top \mathbf{P}_i \mathbf{U}_j]^\top / \mathbf{e}_3^\top \mathbf{P}_i \mathbf{U}_j$  is the 2-vector that we would obtain if  $\boldsymbol{\theta}$  were the true parameter. It can also be written as

$$\mathbf{x}_{ij}(\boldsymbol{\theta}) = \left[ \frac{\mathbf{e}_1^\top \mathbf{K}_i(\mathbf{R}_i \mathbf{X}_j + \mathbf{t}_i)}{\mathbf{e}_3^\top \mathbf{K}_i(\mathbf{R}_i \mathbf{X}_j + \mathbf{t}_i)}; \frac{\mathbf{e}_2^\top \mathbf{K}_i(\mathbf{R}_i \mathbf{X}_j + \mathbf{t}_i)}{\mathbf{e}_3^\top \mathbf{K}_i(\mathbf{R}_i \mathbf{X}_j + \mathbf{t}_i)} \right]^\top. \quad (2)$$

The minimization of  $\mathcal{C}_{2,2}$  is a hard nonconvex problem. In general, it does not admit closed-form solution and the existing iterative algorithms may often get stuck in local minima. An ingenious idea to overcome this difficulty [11,13] is based on the minimization of the  $L_\infty$  cost function

$$\mathcal{C}_{\infty,s}(\boldsymbol{\theta}) = \max_{j=1,\dots,n;i \in I_j} \|\mathbf{x}_{ij} - \mathbf{x}_{ij}(\boldsymbol{\theta})\|_s, \quad s \in [1, +\infty]. \quad (3)$$

Note that the substitution of the  $L_2$ -cost function by the  $L_\infty$ -cost function has been proved to lead to improved algorithms in other estimation problems as well, *cf.*, *e.g.*, [6]. This cost function has a clear practical advantage in that all its sublevel sets are convex. This property ensures that all minima of  $\mathcal{C}_{\infty,s}$  form a convex set and that an element of this set can be computed by solving a sequence of convex programs [14], *e.g.*, by the bisection algorithm. Note that for  $s = 1$  and  $s = +\infty$ , the minimization of  $\mathcal{C}_{\infty,s}$  can be recast in a sequence of LPs. The main idea behind the bisection algorithm can be summarized as follows. We aim to designate an algorithm computing  $\widehat{\boldsymbol{\theta}}_s \in \arg \min_{\boldsymbol{\theta}} \mathcal{C}_{\infty,s}(\boldsymbol{\theta})$ , for any pre-specified  $s \geq 1$ , over the set of all vectors  $\boldsymbol{\theta}$  satisfying the cheirality condition. Let us introduce the residuals  $\mathbf{r}_{ij}(\boldsymbol{\theta}) = \mathbf{x}_{ij} - \mathbf{x}_{ij}(\boldsymbol{\theta})$  that can be represented as

$$\mathbf{r}_{ij}(\boldsymbol{\theta}) = \left[ \frac{\mathbf{a}_{ij1}^\top \boldsymbol{\theta}}{\mathbf{c}_{ij}^\top \boldsymbol{\theta}}; \frac{\mathbf{a}_{ij2}^\top \boldsymbol{\theta}}{\mathbf{c}_{ij}^\top \boldsymbol{\theta}} \right]^\top, \quad (4)$$

for some vectors  $\mathbf{a}_{ij\ell}, \mathbf{c}_{ij} \in \mathbb{R}^{3(m+n-1)}$ . Furthermore, as presented in Remark 2, the cheirality conditions imply the set of linear constraints  $\mathbf{c}_{ij}^\top \boldsymbol{\theta} \geq 1$ . Thus, the problem of computing  $\widehat{\boldsymbol{\theta}}_s$  can be rewritten as

$$\text{minimize } \gamma \quad \text{subject to } \begin{cases} \|\mathbf{r}_{ij}(\boldsymbol{\theta})\|_s \leq \gamma, \\ \mathbf{c}_{ij}^\top \boldsymbol{\theta} \geq 1. \end{cases} \quad (5)$$

Note that the inequality  $\|\mathbf{r}_{ij}(\boldsymbol{\theta})\|_s \leq \gamma$  can be replaced by  $\|\mathbf{A}_{ij}\boldsymbol{\theta}\|_s \leq \gamma \mathbf{c}_{ij}^\top \boldsymbol{\theta}$  with  $\mathbf{A}_{ij} = [\mathbf{a}_{ij1}^\top; \mathbf{a}_{ij2}^\top]$ . Although (5) is not a convex problem, its solution can be well approximated by solving a sequence of convex feasibility problems of the form

$$P_{s,\gamma} : \quad \text{find } \boldsymbol{\theta} \quad \text{s.t.} \quad \begin{cases} \|\mathbf{A}_{ij}\boldsymbol{\theta}\|_s \leq \gamma \mathbf{c}_{ij}^\top \boldsymbol{\theta}, \\ \mathbf{c}_{ij}^\top \boldsymbol{\theta} \geq 1. \end{cases}$$

Given a small number  $\epsilon > 0$  controlling the accuracy of approximation, the bisection algorithm reads as follows:

**Step 1:** Compute a  $\hat{\boldsymbol{\theta}}$  satisfying the chirality conditions; for example, by solving a linear feasibility problem.

**Step 2:** Set  $\gamma_l = 0$  and  $\gamma_u = \mathcal{C}_{\infty, s}(\hat{\boldsymbol{\theta}})$ .

**Step 3:** Set  $\gamma = (\gamma_l + \gamma_u)/2$ .

**Step 4:** If  $P_{s, \gamma}$  has no solution, set  $\gamma_l = \gamma$ . Otherwise, replace the current value of  $\hat{\boldsymbol{\theta}}$  by a solution to  $P_{s, \gamma}$  and set  $\gamma_u = \mathcal{C}_{\infty, s}(\hat{\boldsymbol{\theta}})$ .

**Step 5:** If  $\gamma_u - \gamma_l < \epsilon$ , then assign to  $\hat{\boldsymbol{\theta}}_s$  the current value of  $\hat{\boldsymbol{\theta}}$  and terminate. Otherwise, go to Step 3.

## 4 Robust estimation by linear programming

This and the next sections contain the main theoretical contribution of the present work. We start with the precise formulation of the statistical model. We then exhibit a prior distribution on the unknown parameters of the model that leads to a MAP estimator.

### 4.1 The statistical model

Let us first observe that, in view of (1) and (4), the model we are considering can be rewritten as

$$\left[ \frac{\mathbf{a}_{ij1}^T \boldsymbol{\theta}^*}{\mathbf{c}_{ij}^T \boldsymbol{\theta}^*}; \frac{\mathbf{a}_{ij2}^T \boldsymbol{\theta}^*}{\mathbf{c}_{ij}^T \boldsymbol{\theta}^*} \right]^T = \xi_{ij}, \quad j = 1, \dots, n; \quad i \in I_j. \quad (6)$$

Let  $N = 2 \sum_{j=1}^n I_j$  be the total number of measurements and let  $M = 3(n + m - 1)$  be the size of the vector  $\boldsymbol{\theta}^*$ . Let us denote by  $\mathbf{A}$  (resp.  $\mathbf{C}$ ) the  $N \times M$  matrix formed by the concatenation of the row-vectors  $\mathbf{a}_{ij\ell}^T$  (resp.  $\mathbf{c}_{ij}^T$ ). Similarly, let us denote by  $\boldsymbol{\xi}$  the  $N$ -vector formed by concatenating the vectors  $\xi_{ij}$ . In these notation, Eq. (6) is equivalent to  $\mathbf{a}_p^T \boldsymbol{\theta}^* = (\mathbf{c}_p^T \boldsymbol{\theta}^*) \xi_p$ ,  $p = 1, \dots, N$ . This equation defines the statistical model in the case where there is no outlier. To extend this model to cover the situation where some outliers are present in the measurements, we introduce the vector  $\boldsymbol{\omega}^* \in \mathbb{R}^N$  defined by  $\omega_p^* = \mathbf{a}_p^T \boldsymbol{\theta}^* - (\mathbf{c}_p^T \boldsymbol{\theta}^*) \xi_p$  so that  $\omega_p^* = 0$  if the  $p^{\text{th}}$  measurement is an inlier and  $|\omega_p^*| > 0$  otherwise. This leads us to the model:

$$\mathbf{A} \boldsymbol{\theta}^* = \boldsymbol{\omega}^* + \text{diag}(\mathbf{C} \boldsymbol{\theta}^*) \boldsymbol{\xi}, \quad (7)$$

where  $\text{diag}(\mathbf{v})$  stands for the diagonal matrix having the components of  $\mathbf{v}$  as diagonal entries.

**Statement of the problem:** *Given the matrices  $\mathbf{A}$  and  $\mathbf{C}$ , estimate the parameter-vector  $\boldsymbol{\beta}^* = [\boldsymbol{\theta}^{*T}; \boldsymbol{\omega}^{*T}]^T$  based on the following prior information:*

$C_1$ : Eq. (7) holds with some small noise vector  $\boldsymbol{\xi}$ ,

$C_2$ :  $\min_p \mathbf{c}_p^T \boldsymbol{\theta}^* = 1$ ,

$C_3$ :  $\boldsymbol{\omega}^*$  is sparse, i.e., only a small number of coordinates of  $\boldsymbol{\omega}^*$  are different from zero.

### 4.2 Sparsity prior and MAP estimator

To derive an estimator of the parameter  $\boldsymbol{\beta}^*$ , we place ourselves in the Bayesian framework. To this end, we impose a probabilistic structure on the noise vector  $\boldsymbol{\xi}$  and introduce a prior distribution on the unknown vector  $\boldsymbol{\beta}$ .

Since the noise  $\boldsymbol{\xi}$  represents the difference (in pixels) between the measurements and the true image points, it is naturally bounded and, generally, does not exceed the level of a few pixels. Therefore, it is reasonable to assume that the components of  $\boldsymbol{\xi}$  are uniformly distributed in some compact set of  $\mathbb{R}^2$ , centered at the origin. We assume in what follows that the subvectors  $\xi_{ij}$  of  $\boldsymbol{\xi}$  are uniformly distributed in the square  $[-\sigma, \sigma]^2$  and are mutually independent. Note that this implies that all the coordinates of  $\boldsymbol{\xi}$  are independent. In practice, this assumption can be enforced by decorrelating the measurements using the empirical covariance matrix [16]. We define the prior on  $\boldsymbol{\theta}$  as the uniform distribution on the polytope  $\mathcal{P} = \{\boldsymbol{\theta} \in \mathbb{R}^M : \mathbf{C} \boldsymbol{\theta} \geq \mathbf{1}\}$ , where the inequality is understood componentwise. The density of this distribution is  $p_1(\boldsymbol{\theta}) \propto \mathbf{1}_{\mathcal{P}}(\boldsymbol{\theta})$ , where  $\propto$  stands for the proportionality relation and  $\mathbf{1}_{\mathcal{P}}(\boldsymbol{\theta}) = 1$  if  $\boldsymbol{\theta} \in \mathcal{P}$  and 0 otherwise. When  $\mathcal{P}$  is unbounded, this results in an improper prior, which is however not a problem for defining the Bayes estimator.

The task of choosing a prior on  $\boldsymbol{\omega}$  is more delicate in that it should reflect the information that  $\boldsymbol{\omega}$  is sparse. The most natural prior would be the one having a density which is a decreasing function of the  $L_0$ -norm of  $\boldsymbol{\omega}$ , i.e., of the number of its nonzero coefficients. However, the computation of estimators based on this type of priors is NP-hard. An approach for overcoming this difficulty relies on using the  $L_1$ -norm instead of the  $L_0$ -norm. Following this idea, we define the prior distribution on  $\boldsymbol{\omega}$  by the probability density  $p_2(\boldsymbol{\omega}) \propto f(\|\boldsymbol{\omega}\|_1)$ , where  $f$  is some decreasing function<sup>2</sup> defined on  $[0, \infty)$ . Assuming in addition that  $\boldsymbol{\theta}$  and  $\boldsymbol{\omega}$  are independent, we get the following prior on  $\boldsymbol{\beta}$ :

$$\pi(\boldsymbol{\beta}) = \pi(\boldsymbol{\theta}; \boldsymbol{\omega}) \propto \mathbf{1}_{\mathcal{P}}(\boldsymbol{\theta}) \cdot f(\|\boldsymbol{\omega}\|_1). \quad (8)$$

**Theorem 1** *Assume that the noise  $\boldsymbol{\xi}$  has independent entries which are uniformly distributed in  $[-\sigma, \sigma]$  for some  $\sigma > 0$ , then the MAP estimator  $\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\theta}}^T; \hat{\boldsymbol{\omega}}^T]^T$*

<sup>1</sup> To get a matrix of the same size as  $\mathbf{A}$ , in the matrix  $\mathbf{C}$  each row is duplicated twice.

<sup>2</sup> The most common choice is  $f(x) = e^{-x}$  corresponding to the multivariate Laplace density.

based on the prior  $\pi$  defined by Eq. (8) is the solution of the optimization problem:

$$\min \|\boldsymbol{\omega}\|_1 \quad \text{s.t.} \quad \begin{cases} |\mathbf{a}_p^\top \boldsymbol{\theta} - \omega_p| \leq \sigma \mathbf{c}_p^\top \boldsymbol{\theta}, \quad \forall p \\ \mathbf{c}_p^\top \boldsymbol{\theta} \geq 1, \quad \forall p. \end{cases} \quad (9)$$

The proof of this theorem is a simple exercise and is left to the reader.

*Remark 3 (Condition C<sub>2</sub>)* One easily checks that any solution of (9) satisfies condition C<sub>2</sub>. Indeed, if for some solution  $\hat{\boldsymbol{\beta}}$  it were not the case, then  $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} / \min_p \mathbf{c}_p^\top \hat{\boldsymbol{\theta}}$  would satisfy the constraints of (9) and  $\tilde{\boldsymbol{\omega}}$  would have a smaller  $L_1$ -norm than that of  $\hat{\boldsymbol{\omega}}$ , which is in contradiction with the fact that  $\hat{\boldsymbol{\beta}}$  solves (9).

*Remark 4 (The role of  $\sigma$ )* In the definition of  $\hat{\boldsymbol{\beta}}$ ,  $\sigma$  is a free parameter that can be interpreted as the level of separation of inliers from outliers. The proposed algorithm implicitly assumes that all the measurements  $\mathbf{x}_{ij}$  for which  $\|\boldsymbol{\xi}_{ij}\|_\infty > \sigma$  are outliers, while all the others are treated as inliers.

If  $\sigma$  is unknown, a reasonable way of acting is to impose a prior distribution on the possible values of  $\sigma$  and to define the estimator  $\hat{\boldsymbol{\beta}}$  as a MAP estimator based on the prior incorporating the uncertainty on  $\sigma$ . When there are no outliers and the prior on  $\sigma$  is decreasing, this approach leads to the estimator minimizing the  $L_\infty$  cost function. In the presence of outliers, the shape of the prior on  $\sigma$  becomes more important for the definition of the estimator. This is an interesting point for future investigation.

### 4.3 Two-step procedure

Building on the previous arguments, we introduce the following two-step algorithm.

**Input:**  $\{\mathbf{a}_p, \mathbf{c}_p; p = 1, \dots, N\}$  and  $\sigma$ .

**Step 1:** Compute  $[\hat{\boldsymbol{\theta}}^\top; \hat{\boldsymbol{\omega}}^\top]^\top$  as a solution to (9) and set  $J = \{p : \hat{\omega}_p = 0\}$ .

**Step 2:** Apply the bisection algorithm to the reduced data set  $\{\mathbf{x}_p; p \in J\}$ .

Two observations are in order. First, when applying the bisection algorithm at Step 2, we can use  $\mathcal{C}_{\infty, s}(\hat{\boldsymbol{\theta}})$  as the initial value of  $\gamma_u$ . The second observation is that a better way of acting would be to minimize the weighted  $L_1$ -norm of  $\boldsymbol{\omega}$ , where the weight assigned to  $\omega_p$  is inversely proportional to the depth  $\mathbf{c}_p^\top \boldsymbol{\theta}^*$ . Since  $\boldsymbol{\theta}^*$  is unknown, a reasonable strategy consists in adding a step in between Step 1 and Step 2, which performs the weighted minimization with weights  $\{(\mathbf{c}_p^\top \hat{\boldsymbol{\theta}})^{-1}; p = 1, \dots, N\}$ .

## 5 Accuracy of estimation

Let us introduce some additional notation. Recall the definition of  $\mathcal{P}$  and set  $\partial\mathcal{P} = \{\boldsymbol{\theta} : \min_p \mathbf{c}_p^\top \boldsymbol{\theta} = 1\}$ . For every subset of indices  $J \subset \{1, \dots, N\}$ , we denote by  $\mathbf{A}_J$  the  $N \times M$  matrix obtained from  $\mathbf{A}$  by replacing the rows that have an index outside  $J$  by zero. Furthermore, for every  $J \subset \{1, \dots, N\}$ , let us define

$$\delta_J(\boldsymbol{\theta}^*) = \sup_{\boldsymbol{\theta} \in \partial\mathcal{P}, \mathbf{A}\boldsymbol{\theta} \neq \mathbf{A}\boldsymbol{\theta}^*} \frac{\|\mathbf{A}_J(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|_2}{\|\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|_2}, \quad \boldsymbol{\theta}^* \in \partial\mathcal{P}. \quad (10)$$

One easily checks that for every  $\boldsymbol{\theta} \in \partial\mathcal{P}$ ,  $\delta_J(\boldsymbol{\theta}) \in [0, 1]$  and  $\delta_J(\boldsymbol{\theta}) \leq \delta_{J'}(\boldsymbol{\theta})$  if  $J \subset J'$ .

**Assumption A:** The real number  $\lambda(\boldsymbol{\theta}^*)$  defined by

$$\lambda(\boldsymbol{\theta}^*) = \min_{\boldsymbol{\theta} \in \partial\mathcal{P} \setminus \{\boldsymbol{\theta}^*\}} \frac{\|\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|_2}{\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2}$$

is strictly positive.

### 5.1 The noise free case

To evaluate the quality of estimation, we first place ourselves in the case where  $\sigma = 0$ . The estimator  $\hat{\boldsymbol{\beta}}$  of  $\boldsymbol{\beta}^*$  is then defined as a solution to the optimization problem

$$\min \|\boldsymbol{\omega}\|_1 \quad \text{over } \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\omega} \end{bmatrix} \quad \text{s.t.} \quad \begin{cases} \mathbf{A}\boldsymbol{\theta} = \boldsymbol{\omega} \\ \mathbf{C}\boldsymbol{\theta} \geq \mathbf{1} \end{cases}. \quad (11)$$

In this particular case the proposed procedure coincides with the well-known estimator that minimizes the  $L_1$ -norm of  $\mathbf{A}\boldsymbol{\theta}$  subject to  $\mathbf{C}\boldsymbol{\theta} \geq \mathbf{1}$ . Although this procedure was known, to the best of our knowledge the theoretical results of this section are new.

From now on, for every index set  $T$  and for every vector  $\mathbf{h}$ ,  $\mathbf{h}_T$  stands for the vector equal to  $\mathbf{h}$  on an index set  $T$  and zero elsewhere. The complementary set of  $T$  will be denoted by  $T^c$ .

**Theorem 2** Let Assumption A be fulfilled and let  $T_0$  (resp.  $T_1$ ) denote the index set corresponding to the locations of  $S$  largest entries<sup>3</sup> of  $\boldsymbol{\omega}^*$  (resp.  $(\boldsymbol{\omega}^* - \hat{\boldsymbol{\omega}})_{T_0^c}$ ). If  $\delta_{T_0}(\boldsymbol{\theta}^*) + \delta_{T_0 \cup T_1}(\boldsymbol{\theta}^*) < 1$  then, for some constant  $C_0$ , it holds:

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq C_0 \|\boldsymbol{\omega}^* - \boldsymbol{\omega}_S^*\|_1, \quad (12)$$

where  $\boldsymbol{\omega}_S^*$  stands for the vector  $\boldsymbol{\omega}^*$  with all but the  $S$ -largest entries set to zero. In particular, if  $\boldsymbol{\omega}^*$  has no more than  $S$  nonzero entries, then the estimation is exact:  $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^*$ .

<sup>3</sup> in absolute value

*Remark 5* The assumption  $\delta_{T_0}(\boldsymbol{\theta}^*) + \delta_{T_0 \cup T_1}(\boldsymbol{\theta}^*) < 1$  is close in spirit to the restricted isometry property (cf., e.g., [7, 4, 2] and the references therein). It is very likely that results similar to that of Theorem 2 hold under other kind of assumptions recently introduced in the theory of sparse recovery based on  $L_1$ -relaxation [8, 30, 1]. This investigation is left for future research.

We emphasize that the constant  $C_0$  is rather small. For example, if  $\delta_{T_0}(\boldsymbol{\theta}^*) + \delta_{T_0 \cup T_1}(\boldsymbol{\theta}^*) = 0.5$ , then  $\|\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}^*\|_2 + \|\mathbf{A}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|_2 \leq (8/\sqrt{5})\|\boldsymbol{\omega}^* - \boldsymbol{\omega}_S^*\|_1$ .

## 5.2 The noisy case

The assumption  $\sigma = 0$  is an idealization of the reality that has the advantage of simplifying the mathematical derivations. While such a simplified setting is useful for conveying the main ideas behind the proposed methodology, it is of major practical importance to discuss the extensions to the more realistic *noisy* model.

**Theorem 3** *Let the assumptions of Theorem 2 be fulfilled and let the noise vector  $\boldsymbol{\xi}$  satisfy  $\|\boldsymbol{\xi}\|_\infty \leq \sigma$ , then*

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq C_0\|\boldsymbol{\omega}^* - \boldsymbol{\omega}_S^*\|_1 + C_1\sigma(\|\mathbf{C}\widehat{\boldsymbol{\theta}}\|_2 + \|\mathbf{C}\boldsymbol{\theta}^*\|_2),$$

where  $C_0$  and  $C_1$  are some constants.

The constants  $C_0$  and  $C_1$  have rather simple explicit forms which are given in Remark 6 of Section 9. These constants have very reasonable values provided that the parameter  $\lambda$  in Assumption A is not too small and  $\delta_{T_0}(\boldsymbol{\theta}^*) + \delta_{T_0 \cup T_1}(\boldsymbol{\theta}^*)$  is not too close to one.

It should also be noted that Theorem 3 covers also the case of random noise vector  $\boldsymbol{\xi}$ . Indeed, if  $\boldsymbol{\xi}$  is random with all its coordinates a.s. bounded by  $\sigma$ , then the conclusion of Theorem 3 holds with probability one. The situation is a bit different in the case of unbounded random errors, since even if all the assumptions are fulfilled, the result of Theorem 3 is guaranteed to hold only with probability  $p_\sigma$ , where  $p_\sigma = \mathbf{P}(\|\boldsymbol{\xi}\|_\infty \leq \sigma)$  is in general  $< 1$ . Fortunately, it is often possible to make this probability close to one by a proper choice of  $\sigma$  using well-known inequalities controlling the tails of random variables. For instance, if the coordinates of  $\boldsymbol{\xi}$  are independent centered Gaussian with variance  $v^2$ , then by choosing  $\sigma = v\sqrt{4\log N}$ , we can guarantee that  $p_\sigma \geq 1 - N^{-1}$ , which in typical cases is very close to one since  $N$  is large.

The result of Theorem 3 may appear not very conventional in that its right hand side contains a term depending on the estimator  $\widehat{\boldsymbol{\theta}}$ . All numerical experiments we did show that the term  $\|\mathbf{C}\widehat{\boldsymbol{\theta}}\|_2$  is not very large compared to  $\|\mathbf{C}\boldsymbol{\theta}^*\|_2$ , which is always larger than  $\sqrt{N}$ . However, for the sake of completeness we present another result that—at the price of a stronger assumption—leads

to an upper bound (on the accuracy of the estimator) which is independent of  $\widehat{\boldsymbol{\beta}}$ .

**Theorem 4** *Let us introduce the quantity*

$$\lambda_{\mathbf{C}/\mathbf{A}}(\boldsymbol{\theta}^*) = \sup_{\boldsymbol{\theta} \in \partial\mathcal{P} \setminus \{\boldsymbol{\theta}^*\}} \frac{\|\mathbf{C}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|_2}{\|\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|_2}.$$

*If all the assumptions of Theorem 3 are fulfilled and  $\delta_{T_0}(\boldsymbol{\theta}^*) + \delta_{T_0 \cup T_1}(\boldsymbol{\theta}^*) + \sqrt{2}\sigma\lambda_{\mathbf{C}/\mathbf{A}}(\boldsymbol{\theta}^*) < 1$ , then*

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq C'_0\|\boldsymbol{\omega}^* - \boldsymbol{\omega}_S^*\|_1 + C'_1\sigma\|\mathbf{C}\boldsymbol{\theta}^*\|_2$$

where  $C'_0$  and  $C'_1$  are some constants.

## 5.3 Discussion on assumptions

Assumption A is necessary for identifying the parameter vector  $\boldsymbol{\theta}^*$  even in the case without outliers. In fact, if  $\boldsymbol{\omega}^* = \mathbf{0}$ , and if Assumption A is not fulfilled, then<sup>4</sup>  $\exists \boldsymbol{\theta}^1 \in \partial\mathcal{P} \setminus \{\boldsymbol{\theta}^*\}$  such that  $\mathbf{A}\boldsymbol{\theta}^1 = \mathbf{A}\boldsymbol{\theta}^*$ . This obviously implies that the vector  $\boldsymbol{\theta}^*$  is not identifiable.

The main assumption in Theorems 2 and 3 is that  $\delta_{T_0}(\boldsymbol{\theta}^*) + \delta_{T_0 \cup T_1}(\boldsymbol{\theta}^*) < 1$ . While this assumption is by no means necessary, it should be recognized that it cannot be drastically relaxed. In fact, it is easy to give an example showing that the condition  $\delta_{T_0 \cup T_1}(\boldsymbol{\theta}^*) < 1$  is necessary. For instance, let  $S = 1$  and

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \\ 0 & 3 & 2 \end{bmatrix}, \quad \boldsymbol{\theta}^* = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{\omega}^* = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Then  $T_0 = \{3\}$  and it can be checked that  $\delta_{T_0 \cup \{2\}}(\boldsymbol{\theta}^*) = 1$ , since the sup is attained for  $\boldsymbol{\theta}' = [0 \ 1 \ 0]^T$ . For this example, consistent estimation of  $\boldsymbol{\theta}^*$  is impossible since there is no particular reason for choosing  $\boldsymbol{\theta}^*$  instead of  $\boldsymbol{\theta}'$ . This kind of situations are discarded thanks to the assumption on  $\delta_{T_0}(\boldsymbol{\theta}^*) + \delta_{T_0 \cup T_1}(\boldsymbol{\theta}^*)$ .

Note also that the mapping  $J \mapsto \delta_J(\boldsymbol{\theta})$  is subadditive, that is  $\delta_{J \cup J'}(\boldsymbol{\theta}) \leq \delta_J(\boldsymbol{\theta}) + \delta_{J'}(\boldsymbol{\theta})$ . Therefore, the condition of Thm. 2 is fulfilled as soon as  $\delta_J(\boldsymbol{\theta}^*) < 1/3$  for every index set  $J$  of cardinality  $\leq S$ . Thus, the condition  $\max_{J: |J| \leq S} \delta_J(\boldsymbol{\theta}^*) < 1/3$  is sufficient for identifying  $\boldsymbol{\theta}^*$  in presence of  $S$  outliers.

A simple upper bound on  $\delta_J$ , can be computed as follows.

**Proposition 1** *Let us denote by  $\mathbf{U}$  the  $N \times \text{Rank}(\mathbf{A})$  matrix having orthonormal columns spanning the image of  $\mathbf{A}$ . Then, for every index set  $J \subset \{1, \dots, N\}$  and for every  $\boldsymbol{\theta} \in \partial\mathcal{P}$ , we have  $\delta_J(\boldsymbol{\theta}) \leq \|\mathbf{U}_J\|$ , where the matrix norm is understood as the largest singular value.*

<sup>4</sup> We assume for simplicity that  $\partial\mathcal{P}$  is compact.

The proof of this proposition, obtained by replacing the sup over  $\partial\mathcal{P}$  by the sup over  $\mathbb{R}^M$ , is given in Section 9. Note that for a given  $J$ , the computation of  $\|\mathbf{U}_J\|$  is far easier than that of  $\delta_J(\boldsymbol{\theta})$ .

## 6 Relation to previous work

The present work is closely related and to some extent is complementary to that of Kanade and Ke [15], Sim and Hartley [26] and Candès and Randall [3]. The first two papers propose two different approaches for allowing the  $L_\infty$ -cost minimization procedure to handle outliers, while the third paper is the first one applying the  $L_1$  relaxation heuristic to the problem of robust estimation. After the submission of the present paper, the closely related work [21] has been published, in which the authors establish that the procedures of [26] and [15] are related through duality, and present additional numerical experiments comparing these methods.

In [15], the authors argue for minimizing the  $(m+1)$ th largest in absolute value reprojection error in order to handle (at most)  $m$  outliers. This problem being in general NP-hard, the authors propose an algorithm, hereafter referred to as KK-procedure, solving a sequence of convex problems leading to an estimator that, in some particular cases, minimizes the  $(m+1)$ th largest RE. It is however not clear how often the KK-procedure will really produce the  $(m+1)$ th largest RE minimizer. There is actually no theoretical investigation supporting the KK-procedure.

Since the estimator proposed in the present paper solves an optimization problem that coincides with that solved at each step of iteration of the KK-procedure, roughly speaking, the theoretical arguments presented in previous sections provide an explanation to the nice empirical performance of the KK-procedure. Moreover, our procedure is complementary to the KK-procedure, since the free parameter for our procedure is the precision of inliers, while the free parameter for the KK-procedure is the presumed number of outliers. The complexities of these algorithms are very comparable.

The procedure introduced by Sim and Hartley [26], hereafter referred to as SH-procedure, consists in iteratively computing the  $L_\infty$ -cost minimizer and in removing, at each cycle, the measurements that have maximal RE. For a fixed positive integer  $k$ —the free parameter of the procedure—the SH-procedure stops when the total number of removed measurements exceeds  $k$ . While the authors prove that at each cycle at least one outlier is removed, there is no theoretical result evaluating the number of inliers removed at each cycle. As for the KK-procedure, the SH-procedure requires the number of presumed outliers. (Note however that one can also

consider a stopping rule depending on the desired accuracy of the reconstruction; namely, one can decide to terminate iterations when the maximal reprojection error becomes smaller than a prescribed threshold  $\sigma$ .) The SH-procedure is substantially more time consuming than the KK-procedure, as well as the one proposed in the present work. In fact, in many situations, the number of outliers is of order of several hundreds while the average number of measurements removed at each iteration varies between 10 and 30 for different datasets. In such cases, the SH-procedure may require a large number of cycles being very time-expensive.

In the statistical literature, the approach consisting in  $L_1$ -relaxation for robust estimation in the presence of measurement errors has been recently considered in [3], see also Candès and Tao [5]. However, there is a key difference between the framework considered by the authors and the one of the present work. In fact, Candès and Randall are concerned by the problem of decoding linear codes in which the matrix  $\mathbf{A}$  can be chosen by the encoder/decoder. Therefore, their results require some conditions that are prohibitively restrictive in our context. For example, the columns of  $\mathbf{A}$  are assumed to be orthogonal.

## 7 Numerical experiments

The aim of this section is twofold. First, we show that the presented methodology can be effectively implemented and leads to estimators that are competitive with the state-of-the art procedures. Second, we provide an empirical evaluation of the quantities involved in our theoretical results for several real-world datasets. This gives an idea of the order of magnitude of the constants appearing in the theorems.

We implemented the algorithm in MatLab, using the SeDuMi package for solving LPs [28]. The Matlab code of our program can be freely downloaded from [imagine.enpc.fr/~dalalyan/3D.html](http://imagine.enpc.fr/~dalalyan/3D.html). To test our approach, we applied our algorithm to four datasets: the *dinosaur* and the *corridor* sequences<sup>5</sup>, as well as the *fountain-P11* and the *Herz-Jesu-P25* sequences<sup>6</sup>.

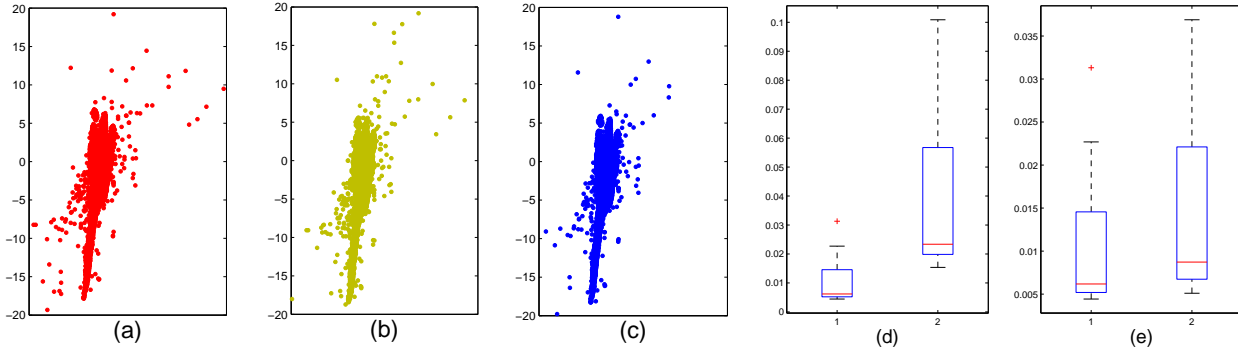
### 7.1 Dinosaur data

The *dinosaur* sequence consists of 36 images of a dinosaur on a turntable, see Fig. 1 (a) for one example. The 2D image points which are tracked across the image sequence and the projection matrices of 36 cameras

<sup>5</sup> [www.robots.ox.ac.uk/~vgg/data1.html](http://www.robots.ox.ac.uk/~vgg/data1.html)

<sup>6</sup> [cvlab.epfl.ch/~streacha/multiview/](http://cvlab.epfl.ch/~streacha/multiview/)





**Fig. 2** DINOSAUR DATA: (a)-(c) Upper view of the 3D points estimated by the KK-procedure (a), by the SH-procedure (b) and by our procedure. (d) Boxplots of the errors when estimating the camera centers by our procedure (left) and by the KK-procedure (right). (e) Boxplots of the errors when estimating the camera centers by our procedure (left) and by the SH-procedure (right).

are provided as well. There are 16,432 image points corresponding to 4,983 scene points. This data is severely affected by outliers which results in a very poor accuracy of the “blind”  $L_\infty$ -cost minimization procedure. Its maximal RE equals 63 pixel and, as shown in Fig. 1, the estimated camera centers are not on the same plane and the scatter plot of scene points is inaccurate.

We ran our procedure with  $\sigma = 0.5$  pixel. If for  $p$ th measurement  $|\omega_p/\mathbf{c}_p^T\boldsymbol{\theta}|$  was larger than  $\sigma/4$ , then it has been considered as an outlier and removed from the dataset. The corresponding 3D scene point was also removed if, after the step of outlier removal, it was seen by only one camera. This resulted in removing 1,306 image points and 297 scene points. The plots (d) and (e) of Fig. 1 show the estimated camera centers and estimated scene points. We see, in particular, that the camera centers are almost coplanar. Note that in this example, the second step of the procedure described in Section 4.3 does not improve on the estimator computed at the first step. Thus, an accurate estimate is obtained by solving only one linear program.

We compared our procedure with the SH-procedure [26] and the KK-procedure [15]. For the SH-procedure, we iteratively computed the  $L_\infty$ -cost minimizer by removing, at each cycle  $j$ , the measurements that had a RE larger than  $E_{max,j} - 0.5\epsilon$ , where  $E_{max,j}$  was the largest RE at the cycle  $j$ . We have stopped the SH-procedure when the number of removed measurements exceeded 1,500. This number has been attained after 53 cycles. Therefore, the execution time was approximately 50 times larger than for our procedure. The estimator obtained by SH-procedure has a maximal RE equal to 1.33 pixel, whereas the maximal RE for our estimator is of 0.62 pixel. Concerning the KK-procedure, we run it with the parameter value  $m = N - N_I$  with  $N_I = 15,000$ , which is approximately the number of inliers detected by our method. Recall that the KK-

procedure aims at minimizing the  $m$ th largest RE. As shown in Fig. 2, our procedure is quite competitive with those of [15, 26].

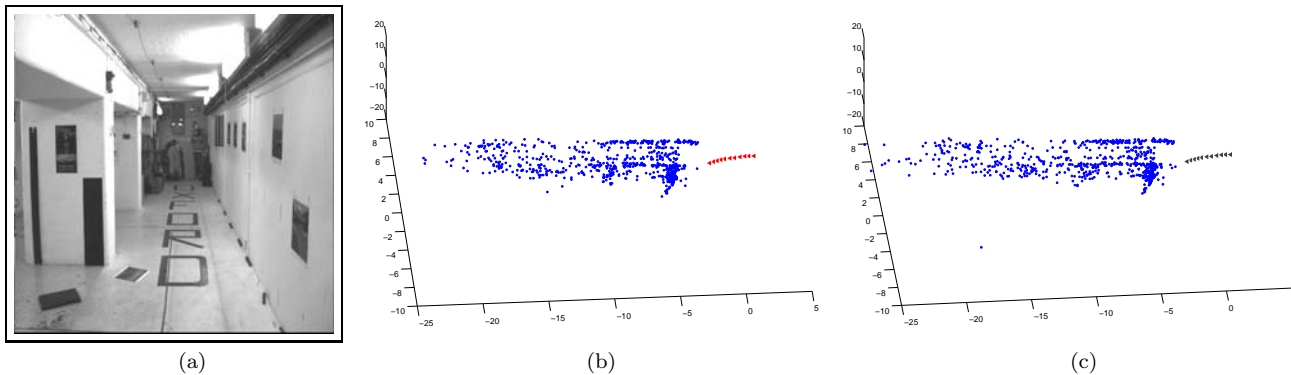
## 7.2 Corridor data

Let us turn to the corridor sequence, consisting of 11 frames one of which is shown in Fig. 3. Matched 2D image points, true 3D scene points and the camera matrices are also provided. There are 737 scene points and 4,035 image points. Thus in average, to each scene point correspond nearly 5.5 image points. We have first ran the original  $L_\infty$  cost minimization algorithm to evaluate the influence of outliers. It has produced an estimator having maximal reprojection error equal to 1.7 pixel, with a very accurate estimator of the locations of cameras. We then ran our algorithm with  $\sigma = 0.5$  pixel. It classified 214 image points and 8 scene points as outliers. This also lead to an improvement of the accuracy of estimation of the camera locations by a factor larger than three. The resulting estimators of camera locations and scene points are shown in Figure 3.

To do more experiments, we removed the outliers detected by our procedure from the dataset. This resulted in a “clean” dataset with 3,813 image points and 729 scene points. We then artificially added outliers in order to study the impact of the number of outliers on the performance of the algorithm. Thus, we have randomly chosen—among 3,813 measurements present in the dataset— $S$  measurements that served as outliers. To these  $S$  measurements, we have added independent 2D vectors with independent coordinates having the same distribution as  $\zeta$ , where

$$\mathbf{P}(\zeta \in dx) = \frac{1}{2}e^{-(|x|-a)}\mathbf{1}_{[a,+\infty)}(|x|) dx.$$

In other terms,  $\zeta$  is the symmetrized version of an exponential random variable translated by  $a$ , where  $a$  is a



**Fig. 3** CORRIDOR DATA: (a) One image out of 11. (b) Camera locations and scene points estimated by our method. (c) True cameras and scene points.

positive parameter corresponding to the magnitude of the RE for outliers. We have run our procedure with  $\sigma = 1$  pixel on this dataset for several values of  $S$  and  $a$ ; namely,  $S = 20; 200; 500; 1,000$ ; and  $a = 5; 10$ . For each value of  $S$  and  $a$ , the experiment has been repeated 500 times. We also did the same experiment for the SH-procedure which has been stopped as soon as the number of removed observations exceeded  $2S$ .

The results of these experiments are summarized in Table 1, where we present the average values and the standard deviations for the number of true positives and false positives, as well as for the accuracy of estimating the camera locations and for the number of cycles for SH-procedure, where each cycle comprises one realization of the bisection algorithm. To compute the accuracy, we have beforehand normalized the camera locations so that they are centered and the average distance to the origin is equal to one.

The results reported in Table 1 demonstrate the complementarity of the SH-procedure and the one proposed in the present work. In fact, the SH-procedure outperforms our procedure in terms of the accuracy of estimating the camera locations when  $S \geq 500$  or  $a = 10$ . This improvement is achieved at the cost of much larger execution times. For example, when  $S = 500$  and  $a = 5$ , the average execution time for our algorithm is more than 25 times smaller than that for the SH-procedure and the results of our algorithm have very acceptable accuracy. It is also noteworthy that this synthetic dataset is particularly well suited for the use of the SH-procedure, since more than 95% of the outliers have a RE lying in the interval  $[a, a + 3]$ , which leads to the removal of a large number of measurements at each cycle. In the case of the dataset of the next subsection, for instance, the REs of outliers are much more spread out and, as a consequence, the average number of measurements removed at each cycle is more than 5

times smaller than in the case of the synthetic data of this subsection.

### 7.3 Herz-Jesu data

This is one of benchmark datasets of [27]; it contains 25 frames which are corrected for distortion. Each frame is of size  $2048 \times 3072$ . The authors provide the camera matrices as well. We first established pairwise correspondences between different frames using SIFT descriptors [18, 29]. The resulting correspondance matrix is available at <http://imagine.enpc.fr/~dalalyan/3D.html>. It consists of 15,323 scene points and 87,968 measurements. The naive  $L_\infty$ -cost minimization leads to an estimator with maximal RE larger than 1,000 pixels.

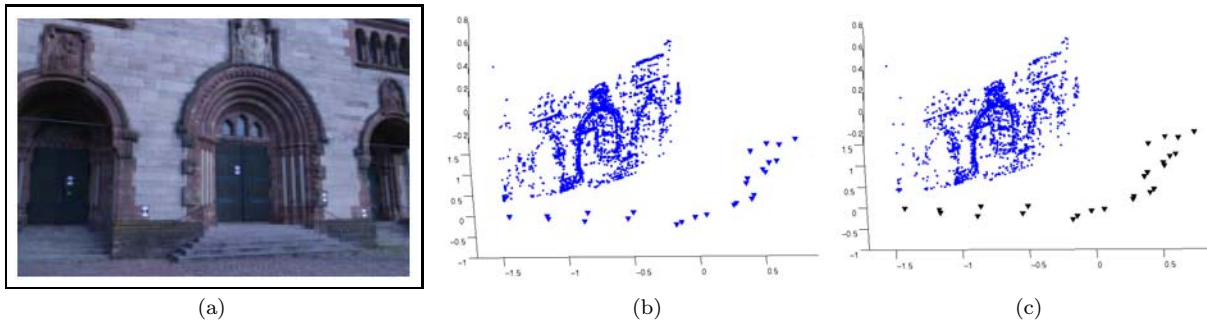
We applied our procedure, with the parameter  $\sigma = 0.5$  pixel, to this dataset. It classified 32,093 image points and 10,702 scene points as outliers and resulted in an estimator that has a maximal RE bounded by 0.25 pixel. The estimated camera locations and 3D points classified as inliers are shown in Figure 4. The accuracy for estimating camera locations was equal to 0.037. We have also tried to apply the SH-procedure on this dataset. After 100 cycles the number of removed measurements were slightly larger than 1,000 and the maximal RE was still on the order of 350 pixels. Finally, we applied the KK-procedure with  $m = 25,000$ , which is the approximate number of inliers detected by our method. The accuracy for the resulted estimator of camera locations was 0.058. The boxplots of errors for different cameras are presented in Figure 5.

### 7.4 Fountain data

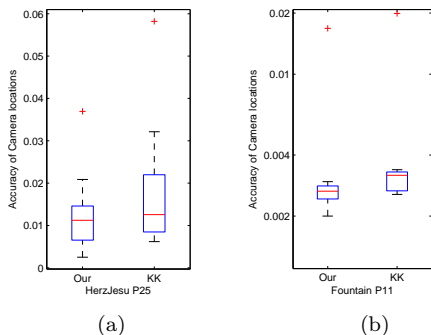
Fountain-P11 is another dataset presented in [27], which contains eleven frames of a fountain. One of these frames

|     |       | Our procedure |      |         |       |          |      | SH-procedure |      |         |       |          |      |        |      |
|-----|-------|---------------|------|---------|-------|----------|------|--------------|------|---------|-------|----------|------|--------|------|
| $a$ | $S$   | TP            |      | FP      |       | Accuracy |      | TP           |      | FP      |       | Accuracy |      | Cycles |      |
|     |       | Mean          | StD  | Mean    | StD   | Mean     | StD  | Mean         | StD  | Mean    | StD   | Mean     | StD  | Mean   | StD  |
| 5   | 20    | 19.2          | 0.93 | 12.8    | 4.15  | .016     | .006 | 16.7         | 2.00 | 40.4    | 11.59 | .079     | .013 | 3.7    | 0.48 |
| 5   | 200   | 192.1         | 2.76 | 197.2   | 17.37 | .018     | .006 | 192.4        | 5.93 | 225.9   | 15.53 | .050     | .020 | 15.2   | 2.02 |
| 5   | 500   | 474.6         | 4.50 | 747.9   | 29.86 | .049     | .015 | 490.9        | 3.15 | 537.3   | 21.89 | .013     | .003 | 28.8   | 1.82 |
| 5   | 1,000 | 935.5         | 8.49 | 1,186.9 | 37.87 | .173     | .046 | 972.1        | 4.81 | 1,061.1 | 20.45 | .012     | .004 | 46.0   | 1.41 |
| 10  | 20    | 19.3          | 0.77 | 30.4    | 8.28  | .013     | .004 | 17.2         | 2.10 | 37.1    | 8.54  | .131     | .043 | 3.9    | 0.74 |
| 10  | 200   | 192.7         | 2.14 | 672.0   | 40.42 | .024     | .009 | 192.7        | 2.50 | 210.7   | 2.99  | .011     | .033 | 15.5   | 0.56 |
| 10  | 500   | 484.0         | 3.74 | 1,252.8 | 50.53 | .080     | .039 | 492.5        | 2.39 | 534.36  | 19.80 | .014     | .006 | 32.6   | 1.85 |
| 10  | 1,000 | 966.4         | 5.95 | 1,541.0 | 33.23 | .211     | .073 | 976.4        | 5.09 | 1,056.8 | 24.79 | .010     | .002 | 53.5   | 2.69 |

**Table 1** TP is the number of true positives, that is the number of correctly classified real outliers. FP is the number of false positives, that is the number of inliers classified as outliers. Accuracy is the largest of 11 reals measuring the distances between the estimated and true camera locations for each camera. Cycles is the number of cycles required by the SH-procedure to achieve the desired number of removed measurements. **Mean** is the average over 500 replications, while **StD** is the standard deviation.



**Fig. 4** HERZ-JESU-P25 DATA [27]: (a) One image out of 25. (b) Camera locations and scene points estimated by our method. (c) True cameras and estimated scene points.



**Fig. 5** Boxplots of the errors when estimating the camera centers by our procedure (left) and by the KK-procedure (right) for (a) HerzJesu P25 sequence and (b) Fountain P11 sequence.

is presented in Figure 6 (a). We established the correspondances in the same way as for the previous dataset and got 10,455 scene points and 31,714 image points. We applied our procedure with the same parameter value  $\sigma = 0.5$ . It has classified as outliers 2,531 scene points and 8,217 image points. The result for estimated camera locations and scene points is shown in Figure 6. We see that the camera locations are quite close to the true camera locations provided by [27]. The accuracy for estimating camera locations by our procedure was

equal to 0.017, while that of the KK-procedure with  $m = 50,000$  was 0.02. The boxplots of errors are shown in Figure 5.

### 7.5 Evaluation of constants

In this subsection, we attempt to make a numerical evaluation of different constants involved in our theoretical results. To this end, we consider the Fountain P11 dataset (experiments on the three other datasets led to very similar results) and choose at random 1000 scene-points, estimated by our procedure. These scene points are considered as ground truth and each of them is projected onto 5 image planes, chosen at random in such a way that the resulting image point lies inside the image box. Since the ground truth cameras are known, these projections are computable. We further add to all measurements a random noise drawn from the uniform distribution  $\mathcal{U}([-σ, σ])$ . We then choose at random  $N_o$  measurements, considered as outliers, and add to these measurements a random noise uniformly distributed in  $[-3σ, -2σ] \cup [2σ, 3σ]$ . The worst-case results over 10 replications obtained for different values of  $N_o$  and  $\sigma$  are reported in Table 2. Note here that the “worst-case results” means the largest values for  $\delta_{T_0}$ ,  $\delta_{T_0 \cup T_1}$  and

| $\sigma$ | $N_o$ | $\lambda$ | $\delta_{T_0}$ | $\delta_{T_0 \cup T_1}$ | $\lambda_{C/A}$ |
|----------|-------|-----------|----------------|-------------------------|-----------------|
| 0.5      | 50    | 1.89      | 0.122          | 0.128                   | 2.17            |
| 0.5      | 100   | 2.04      | 0.199          | 0.206                   | 2.01            |
| 0.5      | 400   | 2.65      | 0.389          | 0.413                   | 1.55            |
| 0.5      | 800   | 3.32      | 0.527          | 0.561                   | 1.24            |
| 1.0      | 50    | 3.73      | 0.113          | 0.117                   | 1.11            |
| 1.0      | 100   | 3.95      | 0.183          | 0.191                   | 1.04            |
| 1.0      | 400   | 5.61      | 0.397          | 0.432                   | 0.78            |
| 1.0      | 800   | 6.63      | 0.515          | 0.558                   | 0.62            |
| 4.0      | 50    | 14.66     | 0.100          | 0.106                   | 0.28            |
| 4.0      | 100   | 17.02     | 0.214          | 0.221                   | 0.24            |
| 4.0      | 400   | 21.72     | 0.395          | 0.412                   | 0.19            |
| 4.0      | 800   | 26.47     | 0.519          | 0.564                   | 0.15            |

**Table 2** Numerical evaluation of constants appearing in theorems on a synthetic data described in Section 7.5. One can remark that the main theoretical assumption  $\delta_{T_0} + \delta_{T_0 \cup T_1} < 1$  is fulfilled when the number of outliers  $N_o$  is not larger than 400, whereas it fails when  $N_o = 800$ .

$\lambda_{C/A}$  and the smallest values for  $\lambda$ . It should also be noted that all mathematical derivations hold true when the quantities  $\delta_T(\boldsymbol{\theta}^*)$  (for  $T = T_0$  and  $T = T_1$ ),  $\lambda(\boldsymbol{\theta}^*)$  and  $\lambda_{C/A}(\boldsymbol{\theta}^*)$  are replaced by  $\delta_T(\boldsymbol{\theta}^*, \hat{\boldsymbol{\theta}}) = \frac{\|\mathbf{A}_J(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|_2}{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2}$ ,  $\lambda(\boldsymbol{\theta}^*, \hat{\boldsymbol{\theta}}) = \frac{\|\mathbf{A}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|_2}{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2}$  and  $\lambda_{C/A}(\boldsymbol{\theta}^*, \hat{\boldsymbol{\theta}}) = \frac{\|\mathbf{C}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|_2}{\|\mathbf{A}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|_2}$ . It is the values of these quantities that are reported in Table 2.

These numerical evaluation reveals that the assumptions of Theorem 2 and 3 are fulfilled when the number of outliers is less than or equal to 400 and are not fulfilled when this number is equal to 800. Of course, this does not mean that the proposed method fails when the number of outliers exceeds 800. This evaluation merely shows the limits of our theoretical results: they are not able to capture the properties of the estimation procedure for large number of outliers. However, to the best of our knowledge, there are no theoretical results in the literature assessing the quality of robust estimators applicable to our context and holding under weaker assumption.

One can also observe that the value of  $\lambda_{C/A}$  is always prohibitively large. Interestingly, in all the numerical experiments we did, the entries of the vector  $\mathbf{C}\hat{\boldsymbol{\theta}}$  were bounded by the respective entries of  $\mathbf{C}\boldsymbol{\theta}^*$ . Unfortunately, we did not succeed to give a mathematical proof to this curious phenomenon. However, this shows that in most situations the term  $\|\mathbf{C}\hat{\boldsymbol{\theta}}\|_2$  is smaller than the term  $\|\mathbf{C}\boldsymbol{\theta}^*\|$  and, therefore, one does not need to resort to Theorem 4.

## 8 Conclusion

In this paper, we presented a rigorous Bayesian framework for the problem of translation estimation and triangulation that have led to a new robust estimation procedure. We have formulated the problem under consideration as a nonlinear inverse problem with a high-dimensional unknown parameter-vector. This vector encapsulates the information on the scene points and the camera locations, as well as the information on the location of outliers in the data. The proposed estimator exploits the sparse nature of the vector of outliers through  $L_1$ -norm minimization.

Although we focused in the present paper on the problem of translation estimation and triangulation, the proposed approach applies to other problems of computer vision such as homography estimation, camera resectioning and 3D reconstruction using a reference plane. (More details on the relation of these problems and the inverse problem considered in this work can be found in [14].)

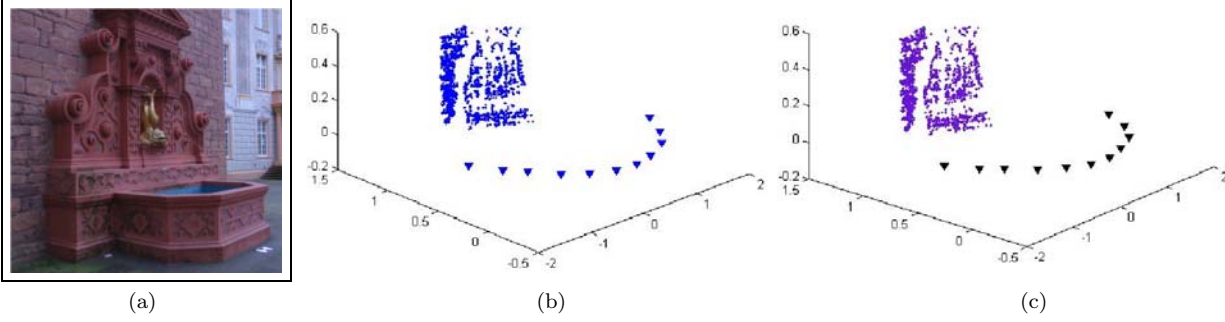
We have given the mathematical proof of the result demonstrating the efficiency of the proposed estimator under mild assumptions. Unfortunately, the verification of these assumptions is impossible, since it requires the knowledge of the true parameter. It is an interesting open problem to find possibly stronger but verifiable assumptions that allow to theoretically assess the accuracy of the estimation.

We applied our procedure to four real-world and synthetic datasets and compared to some recently proposed procedures. The results of these experiments support our theoretical results and demonstrate the complementarity of our procedure to those previously proposed in the literature. In particular, our procedure is in general much faster than that of Sim and Hartley [26] in terms of execution times and is more theoretically justified than the method of Kanade and Ke [15].

## 9 Proofs

This section contains the proofs of the main theoretical claims. We begin with an auxiliary result and then present the proofs of Theorem 2, Theorem 3 and Theorem 4. To ease notation, since there is no confusion, we write  $\delta_J$ ,  $\lambda$  and  $\lambda_{C/A}$  instead of  $\delta_J(\boldsymbol{\theta}^*)$ ,  $\lambda(\boldsymbol{\theta}^*)$  and  $\lambda_{C/A}(\boldsymbol{\theta}^*)$ , respectively.

**Lemma 1** *Let  $\mathbf{v} \in \mathbb{R}^d$  be some vector and let  $S \leq d$  be a positive integer. If we denote by  $T$  the indices of  $S$  largest entries of the vector  $|\mathbf{v}|$ , then  $\|\mathbf{v}_{T^c}\|_2 \leq S^{-1/2} \|\mathbf{v}\|_1$ .*



**Fig. 6** FOUNTAIN-P11 DATA [27]: (a) One image out of 11. (b) Camera locations and scene points estimated by our method. (c) True cameras and estimated scene points.

*Proof* Let us denote by  $T_1$  the index set of  $S$  largest entries of  $|\mathbf{v}_{T^c}|$ , by  $T_2$  the index set of next  $S$  largest entries of  $|\mathbf{v}_{T^c}|$ , and so on. By triangle inequality, one has  $\|\mathbf{v}_{T^c}\|_2 \leq \sum_{j \geq 1} \|\mathbf{v}_{T_j}\|_2$ . On the other hand, one easily checks that  $|v_\ell|^2 \leq |v_\ell| \cdot \|\mathbf{v}_{T_{j-1}}\|_1 / S$  for every  $\ell \in T_j$  with the convention  $T_0 = T$ . This implies that  $\|\mathbf{v}_{T_j}\|_2^2 \leq \|\mathbf{v}_{T_j}\|_1 \|\mathbf{v}_{T_{j-1}}\|_1 / S$ , for every  $j \geq 1$ . After taking the square root of these inequalities and summing up over  $j$ , we get the desired result in view of the obvious inequality  $\|\mathbf{v}_{T_j}\|_1 \leq \|\mathbf{v}_{T_{j-1}}\|_1$ .

*Proof of Theorem 2* We set  $\mathbf{h} = \boldsymbol{\omega}^* - \widehat{\boldsymbol{\omega}}$  and  $\mathbf{g} = \boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}$ . Applying Lemma 1 to the vector  $\mathbf{v} = \mathbf{h}_{T_0^c}$  and to the index set  $T = T_1$ , we get

$$\|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2 \leq S^{-1/2} \|\mathbf{h}_{T_0^c}\|_1. \quad (13)$$

On the other hand, summing up the inequalities

$$\|\mathbf{h}_{T_0^c}\|_1 \leq \|(\boldsymbol{\omega}^* - \mathbf{h})_{T_0^c}\|_1 + \|\boldsymbol{\omega}_{T_0^c}^*\|_1$$

and  $\|\boldsymbol{\omega}_{T_0^c}^*\|_1 \leq \|(\boldsymbol{\omega}^* - \mathbf{h})_{T_0}\|_1 + \|\mathbf{h}_{T_0}\|_1$ , and using the relation  $\|(\boldsymbol{\omega}^* - \mathbf{h})_{T_0}\|_1 + \|(\boldsymbol{\omega}^* - \mathbf{h})_{T_0^c}\|_1 = \|\boldsymbol{\omega}^* - \mathbf{h}\|_1 = \|\widehat{\boldsymbol{\omega}}\|_1$ , we get

$$\|\mathbf{h}_{T_0^c}\|_1 + \|\boldsymbol{\omega}_{T_0^c}^*\|_1 \leq \|\widehat{\boldsymbol{\omega}}\|_1 + \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \|\mathbf{h}_{T_0}\|_1. \quad (14)$$

Since  $\boldsymbol{\beta}^*$  satisfies the constraints of the optimization problem (11) a solution of which is  $\widehat{\boldsymbol{\beta}}$ , we have  $\|\widehat{\boldsymbol{\omega}}\|_1 \leq \|\boldsymbol{\omega}^*\|_1$ . This inequality, in conjunction with (13) and (14), implies

$$\begin{aligned} \|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2 &\leq S^{-1/2} \|\mathbf{h}_{T_0}\|_1 + 2S^{-1/2} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 \\ &\leq \|\mathbf{h}_{T_0}\|_2 + 2S^{-1/2} \|\boldsymbol{\omega}_{T_0^c}^*\|_1, \end{aligned} \quad (15)$$

where the last step follows from the Cauchy-Schwartz inequality. Using once again the fact that both  $\widehat{\boldsymbol{\beta}}$  and  $\boldsymbol{\beta}^*$  satisfy the constraints of (11), we get  $\mathbf{h} = \mathbf{A}\mathbf{g}$ . Therefore,

$$\begin{aligned} \|\mathbf{h}\|_2 &\leq \|\mathbf{h}_{T_0 \cup T_1}\|_2 + \|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2 \\ &\leq \|\mathbf{h}_{T_0 \cup T_1}\|_2 + \|\mathbf{h}_{T_0}\|_2 + 2S^{-1/2} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 \\ &= \|\mathbf{A}_{T_0 \cup T_1} \mathbf{g}\|_2 + \|\mathbf{A}_{T_0} \mathbf{g}\|_2 + 2S^{-1/2} \|\boldsymbol{\omega}_{T_0^c}^*\|_1. \end{aligned}$$

The right side can be further bounded using (eq:13):

$$\begin{aligned} \|\mathbf{h}\|_2 &\leq (\delta_{T_0} + \delta_{T_0 \cup T_1}) \|\mathbf{A}\mathbf{g}\|_2 + 2S^{-1/2} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 \\ &= (\delta_{T_0} + \delta_{T_0 \cup T_1}) \|\mathbf{h}\|_2 + 2S^{-1/2} \|\boldsymbol{\omega}_{T_0^c}^*\|_1. \end{aligned} \quad (16)$$

Since  $\boldsymbol{\omega}_{T_0^c}^* = \boldsymbol{\omega}^* - \boldsymbol{\omega}_S^*$ , the last inequality yields  $\|\mathbf{h}\|_2 \leq (2S^{-1/2} / (1 - \delta_{T_0} - \delta_{T_0 \cup T_1})) \|\boldsymbol{\omega}^* - \boldsymbol{\omega}_S^*\|_1$ . To complete the proof, it suffices to observe that

$$\begin{aligned} \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 &\leq \|\mathbf{g}\|_2 + \|\mathbf{h}\|_2 \leq \lambda^{-1} \|\mathbf{A}\mathbf{g}\|_2 + \|\mathbf{h}\|_2 \\ &= (\lambda^{-1} + 1) \|\mathbf{h}\|_2 \leq C_0 \|\boldsymbol{\omega}^* - \boldsymbol{\omega}_S^*\|_1. \quad \square \end{aligned}$$

*Proof of Theorem 3* Let us define  $\boldsymbol{\eta} = \text{diag}(\mathbf{C}\boldsymbol{\theta}^*)\boldsymbol{\xi}$  and  $\widehat{\boldsymbol{\eta}} = \mathbf{A}\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\omega}}$ . It is clear that these vectors satisfy

$$\|\boldsymbol{\eta}\|_2 \leq \sigma \|\mathbf{C}\boldsymbol{\theta}^*\|_2 \quad \text{and} \quad \|\widehat{\boldsymbol{\eta}}\|_2 \leq \sigma \|\widehat{\mathbf{C}}\boldsymbol{\theta}\|_2 \quad (17)$$

thanks to the condition  $\|\boldsymbol{\xi}\|_\infty \leq \sigma$  and the constraints  $|\mathbf{a}_p^T \boldsymbol{\theta} - \omega_p| \leq \sigma \mathbf{c}_p^T \boldsymbol{\theta}$ ,  $\forall p$ , which are fulfilled by  $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\omega}})$ . Furthermore, since under the assumption  $\|\boldsymbol{\xi}\|_\infty \leq \sigma$  the vector  $\boldsymbol{\beta}^*$  satisfies the constraints of the LP (12), in view of (15), we have

$$\|\mathbf{h}_{T_{01}^c}\|_2 \leq \|\mathbf{h}_{T_0}\|_2 + 2S^{-1/2} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 \quad (18)$$

with  $\mathbf{h} = \boldsymbol{\omega}^* - \widehat{\boldsymbol{\omega}}$  and  $T_{01} = T_0 \cup T_1$ . On the other hand, since  $\mathbf{h} = \mathbf{A}\mathbf{g} + \widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}$ , we have

$$\|\mathbf{h}_{T_{01}^c}\|_2 \geq \|\mathbf{A}_{T_{01}^c} \mathbf{g}\|_2 - \|\widehat{\boldsymbol{\eta}}_{T_{01}^c}\|_2 - \|\boldsymbol{\eta}_{T_{01}^c}\|_2$$

and

$$\|\mathbf{h}_{T_0}\|_2 \leq \|\mathbf{A}_{T_0} \mathbf{g}\|_2 + \|\widehat{\boldsymbol{\eta}}_{T_0}\|_2 + \|\boldsymbol{\eta}_{T_0}\|_2.$$

Combining last three displays, we get

$$\begin{aligned} \|\mathbf{A}_{T_{01}^c} \mathbf{g}\|_2 &\leq \|\mathbf{h}_{T_{01}^c}\|_2 + \|\widehat{\boldsymbol{\eta}}_{T_{01}^c}\|_2 + \|\boldsymbol{\eta}_{T_{01}^c}\|_2 \\ &\leq \|\mathbf{h}_{T_0}\|_2 + \frac{2}{\sqrt{S}} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \|\widehat{\boldsymbol{\eta}}_{T_{01}^c}\|_2 + \|\boldsymbol{\eta}_{T_{01}^c}\|_2 \\ &\leq \|\mathbf{A}_{T_0} \mathbf{g}\|_2 + \frac{2}{\sqrt{S}} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 \\ &\quad + \|\widehat{\boldsymbol{\eta}}_{T_0}\|_2 + \|\boldsymbol{\eta}_{T_0}\|_2 + \|\widehat{\boldsymbol{\eta}}_{T_{01}^c}\|_2 + \|\boldsymbol{\eta}_{T_{01}^c}\|_2. \end{aligned}$$

Using the elementary inequality  $a + b \leq \sqrt{2(a^2 + b^2)}$ , the last inequality can be simplified to

$$\begin{aligned} \|A_{T_0^c} \mathbf{g}\|_2 &\leq \|A_{T_0} \mathbf{g}\|_2 + \frac{2}{\sqrt{S}} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \sqrt{2}(\|\widehat{\boldsymbol{\eta}}\|_2 + \|\boldsymbol{\eta}\|_2) \\ &\leq \|A_{T_0} \mathbf{g}\|_2 + \frac{2}{\sqrt{S}} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \sqrt{2} \epsilon, \end{aligned}$$

where for keeping formulae short we denoted by  $\epsilon$  the expression  $\sigma(\|\mathbf{C}\boldsymbol{\theta}^*\|_2 + \|\widehat{\mathbf{C}}\boldsymbol{\theta}\|_2)$ . Therefore,

$$\begin{aligned} \|\mathbf{A}\mathbf{g}\|_2 &\leq \|A_{T_0} \mathbf{g}\|_2 + \|A_{T_0^c} \mathbf{g}\|_2 \\ &\leq \|A_{T_0} \mathbf{g}\|_2 + \|A_{T_0} \mathbf{g}\|_2 + \frac{2}{\sqrt{S}} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \sqrt{2} \epsilon \\ &\leq (\delta_{T_0} + \delta_{T_0^c}) \|\mathbf{A}\mathbf{g}\|_2 + \frac{2}{\sqrt{S}} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \sqrt{2} \epsilon. \end{aligned}$$

Finally, the chain of inequalities

$$\begin{aligned} \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 &\leq \|\mathbf{h}\|_2 + \|\mathbf{g}\|_2 \leq \|\mathbf{A}\mathbf{g}\|_2 + \|\mathbf{g}\|_2 + \epsilon \\ &\leq (1 + \lambda^{-1}) \|\mathbf{A}\mathbf{g}\|_2 + \epsilon \\ &\leq \frac{1 + \lambda^{-1}}{1 - \delta_{T_0} - \delta_{T_0^c}} \left( \frac{2}{\sqrt{S}} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \sqrt{2} \epsilon \right) + \epsilon \end{aligned}$$

completes the proof of the theorem.  $\square$

*Remark 6* The values of constants  $C_0$  and  $C_1$  can be easily deduced from the proof of Theorem 3. Indeed, one can see that

$$C_0 = \frac{2(1 + \lambda^{-1})}{\sqrt{S}(1 - \delta_{T_0} - \delta_{T_0 \cup T_1})}, \quad C_1 = 1 + \frac{\sqrt{2}(1 + \lambda^{-1})}{1 - \delta_{T_0} - \delta_{T_0 \cup T_1}}.$$

*Proof of Theorem 4* Repeating the arguments of the proof of Theorem 3, we get

$$\|A_{T_0^c} \mathbf{g}\|_2 \leq \|A_{T_0} \mathbf{g}\|_2 + \frac{2}{\sqrt{S}} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \sqrt{2} \epsilon,$$

with  $\epsilon = \sigma(\|\mathbf{C}\boldsymbol{\theta}^*\|_2 + \|\widehat{\mathbf{C}}\boldsymbol{\theta}\|_2)$  and  $\mathbf{g} = \boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}$ . The triangle inequality implies that  $\epsilon \leq \sigma(2\|\mathbf{C}\boldsymbol{\theta}^*\|_2 + \|\mathbf{C}\mathbf{g}\|_2)$  and, setting  $\mu_S = \frac{2}{\sqrt{S}} \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \sqrt{8}\sigma\|\mathbf{C}\boldsymbol{\theta}^*\|_2$ , we get

$$\|A_{T_0^c} \mathbf{g}\|_2 \leq \|A_{T_0} \mathbf{g}\|_2 + \mu_S + \sqrt{2} \sigma \|\mathbf{C}\mathbf{g}\|_2. \quad (19)$$

Using the definition of the matrix norm, one checks that

$$\|\mathbf{C}\mathbf{g}\|_2 = \|\mathbf{C}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A}\mathbf{g}\|_2 \leq \|\mathbf{C}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T\| \cdot \|\mathbf{A}\mathbf{g}\|_2.$$

This yields

$$\begin{aligned} \|\mathbf{A}\mathbf{g}\|_2 &\leq \|A_{T_0} \mathbf{g}\|_2 + \|A_{T_0^c} \mathbf{g}\|_2 \\ &\leq \|A_{T_0} \mathbf{g}\|_2 + \|A_{T_0} \mathbf{g}\|_2 + \mu_S + \sqrt{2} \sigma \|\mathbf{C}\mathbf{g}\|_2 \\ &\leq \|\mathbf{A}\mathbf{g}\|_2 (\delta_{T_0} + \delta_{T_0^c} + \sqrt{2} \sigma \lambda_{C/A}) + \mu_S. \end{aligned}$$

The last inequality can be rewritten as

$$\|\mathbf{A}\mathbf{g}\|_2 \leq \mu_S / (1 - \delta_{T_0} - \delta_{T_0^c} - \sqrt{2} \sigma \lambda_{C/A}) \quad (20)$$

provided that the denominator of the right hand side is strictly positive. Therefore,

$$\begin{aligned} \|\mathbf{g}\|_2 &\leq \lambda^{-1} \|\mathbf{A}\mathbf{g}\|_2 \leq \frac{\mu_S}{\lambda(1 - \delta_{T_0} - \delta_{T_0^c} - \sqrt{2} \sigma \lambda_{C/A})}, \\ \|\mathbf{h}\|_2 &\leq \|\mathbf{A}\mathbf{g}\|_2 + 2\sigma \|\mathbf{C}\boldsymbol{\theta}^*\|_2 + \sigma \|\mathbf{C}\mathbf{g}\|_2 \\ &\leq (1 + \sigma \lambda_{C/A}) \|\mathbf{A}\mathbf{g}\|_2 + 2\sigma \|\mathbf{C}\boldsymbol{\theta}^*\|_2 \\ &\leq \frac{(1 + \sigma \lambda_{C/A}) \mu_S}{1 - \delta_{T_0} - \delta_{T_0^c} - \sqrt{2} \sigma \lambda_{C/A}} + 2\sigma \|\mathbf{C}\boldsymbol{\theta}^*\|_2. \end{aligned}$$

These inequalities, combined with  $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq \|\mathbf{g}\|_2 + \|\mathbf{h}\|_2$ , complete the proof of the theorem.  $\square$

*Remark 7* The values of constants  $C'_0$  and  $C'_1$  can be deduced from the proof of Theorem 4. One easily checks that

$$\begin{aligned} C'_0 &= \frac{2(1 + \lambda^{-1} + \sigma \lambda_{C/A})}{\sqrt{S}(1 - \delta_{T_0} - \delta_{T_0 \cup T_1} - \sqrt{2} \sigma \lambda_{C/A})}, \\ C'_1 &= \frac{2\sqrt{2}(1 + \lambda^{-1} + \sigma \lambda_{C/A})}{1 - \delta_{T_0} - \delta_{T_0 \cup T_1} - \sqrt{2} \sigma \lambda_{C/A}} + 2. \end{aligned}$$

*Proof of Proposition 1* Let us denote by  $I_{N \times N, J}$  the  $N \times N$  matrix obtained from the identity matrix  $I_{N \times N}$  by zeroing all the rows with indices in  $J$ . Then, it holds

$$\begin{aligned} \delta_J(\boldsymbol{\theta}) &\leq \sup_{\mathbf{u} \in \mathbb{R}^M} \frac{\|I_{N \times N, J} \mathbf{A}\mathbf{u}\|_2}{\|\mathbf{A}\mathbf{u}\|_2} = \sup_{\mathbf{v} \in \text{Im}(\mathbf{A})} \frac{\|I_{N \times N, J} \mathbf{v}\|_2}{\|\mathbf{v}\|_2} \\ &= \sup_{\mathbf{w} \in \mathbb{R}^{\text{Rank}(\mathbf{A})}} \frac{\|I_{N \times N, J} \mathbf{U}\mathbf{w}\|_2}{\|\mathbf{U}\mathbf{w}\|_2}. \end{aligned}$$

On the one hand, it is clear that  $I_{N \times N, J} \mathbf{U} = \mathbf{U}_J$ . On the other hand, the fact that the columns of  $\mathbf{U}$  are orthonormal implies that  $\|\mathbf{U}\mathbf{w}\|_2^2 = \mathbf{w}^T \mathbf{U}^T \mathbf{U} \mathbf{w} = \mathbf{w}^T \mathbf{w} = \|\mathbf{w}\|_2^2$ . Therefore,

$$\delta_J(\boldsymbol{\theta}) \leq \sup_{\mathbf{w}} \frac{\|\mathbf{U}_J \mathbf{w}\|_2}{\|\mathbf{U}\mathbf{w}\|_2} = \sup_{\mathbf{w}} \frac{\|\mathbf{U}_J \mathbf{w}\|_2}{\|\mathbf{w}\|_2} = \|\mathbf{U}_J\|$$

and the desired result follows.  $\square$

**Acknowledgements** The work of the first author was partially supported by ANR under grants Callisto and Parcimonie.

## References

1. Bickel, P.J., Ritov, Y., Tsybakov, A.B.: Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37**(4), 1705–1732 (2009)
2. Candès, E., Tao, T.: The Dantzig selector: statistical estimation when  $p$  is much larger than  $n$ . *Ann. Statist.* **35**(6), 2313–2351 (2007)
3. Candès, E.J., Randall, P.A.: Highly robust error correction by convex programming. *IEEE Trans. Inform. Theory* **54**(7), 2829–2840 (2008)

4. Candès, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59**(8), 1207–1223 (2006)
5. Candès, E.J., Tao, T.: Decoding by linear programming. *IEEE Trans. Inform. Theory* **51**(12), 4203–4215 (2005)
6. Dalalyan, A.S., Juditsky, A., Spokoiny, V.: A new algorithm for estimating the effective dimension-reduction subspace. *J. Mach. Learn. Res.* **9**, 1648–1678 (2008)
7. Donoho, D., Elad, M., Temlyakov, V.: Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory* **52**(1), 6–18 (2006)
8. Donoho, D.L., Huo, X.: Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory* **47**(7), 2845–2862 (2001)
9. Enqvist, O., Kahl, F.: Robust optimal pose estimation. In: *European Conference on Computer Vision*, pp. I: 141–153 (2008)
10. Hartley, R., Kahl, F.: Global optimization through rotation space search. *The International Journal of Computer Vision* **82**(1), 64 – 79 (2009)
11. Hartley, R.I., Schaffalitzky, F.:  $L_\infty$  minimization in geometric reconstruction problems. In: *Conference on Computer Vision and Pattern Recognition*, pp. I: 504–509 (2004)
12. Hartley, R.I., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2004)
13. Kahl, F.: Multiple view geometry and the  $L_\infty$ -norm. In: *International Conference on Computer Vision*, pp. 1002–1009. IEEE Computer Society (2005)
14. Kahl, F., Hartley, R.I.: Multiple-view geometry under the  $L_\infty$  norm. *IEEE Trans. Pattern Analysis and Machine Intelligence* **30**(9), 1603–1617 (2008)
15. Kanade, T., Ke, Q.: Quasiconvex optimization for robust geometric reconstruction. In: *International Conference on Computer Vision*, pp. II: 986–993 (2005)
16. Ke, Q., Kanade, T.: Uncertainty models in quasiconvex optimization for geometric reconstruction. In: *Conference on Computer Vision and Pattern Recognition*, pp. I: 1199–1205 (2006)
17. Li, H.D.: A practical algorithm for  $L_\infty$  triangulation with outliers. In: *Conference on Computer Vision and Pattern Recognition* (2007)
18. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60**(2), 91–110 (2004)
19. Martinec, D., Pajdla, T.: Robust rotation and translation estimation in multiview reconstruction. In: *Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2007)
20. Nistér, D.: An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell* **26**(6), 756–777 (2004)
21. Olsson, C., Eriksson, A., Hartley, R.: Outlier removal using duality (2010)
22. Olsson, C., Eriksson, A.P., Kahl, F.: Efficient optimization for  $L_\infty$  problems using pseudoconvexity. In: *International Conference on Computer Vision*, pp. 1–8 (2007)
23. Seo, Y., Lee, H., Lee, S.W.: Outlier removal by convex optimization for l-infinity approaches. In: *PSIVT '09: Proceedings of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology*, pp. 203–214 (2009)
24. Seo, Y.D., Hartley, R.I.: A fast method to minimize  $L_\infty$  error norm for geometric vision problems. In: *International Conference on Computer Vision*, pp. 1–8 (2007)
25. Seo, Y.D., Lee, H.J., Lee, S.W.: Sparse structures in L-infinity norm minimization for structure and motion reconstruction. In: *European Conference on Computer Vision*, pp. I: 780–793 (2008)
26. Sim, K., Hartley, R.: Removing outliers using the  $L_\infty$  norm. In: *Conference on Computer Vision and Pattern Recognition*, pp. I: 485–494 (2006)
27. Strecha, C., von Hansen, W., Gool, L.V., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: *Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2009)
28. Sturm, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.* **11/12**(1-4), 625–653 (1999)
29. Vedaldi, A., Fulkerson, B.: *VLFeat: An open and portable library of computer vision algorithms* (2008)
30. Zhao, P., Yu, B.: On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7**, 2541–2563 (2006)