

# CDS

TECHNICAL MEMORANDUM NO. CIT-CDS 94-004  
February 15, 1994

**“Dynamic Estimation of Rigid Motion from  
Perspective Views via Recursive Identification of  
Exterior Differential Systems with Parameters on  
a Topological Manifold”**

Stefano Soatto, Ruggero Frezza and Pietro Perona

**Control and Dynamical Systems  
California Institute of Technology  
Pasadena, CA 91125**

# Dynamic Estimation of Rigid Motion From Perspective Views via Recursive Identification of Nonlinear Implicit Systems with Parameters on a Topological Manifold\*

Stefano Soatto† Ruggero Frezza‡ Pietro Perona†‡ Giorgio Picci‡

† California Institute of Technology 116-81, Pasadena-CA 91125

‡ Università di Padova, Dipartimento di Elettronica, Padova-Italy  
soatto@caltech.edu

**Keywords:** Dynamic vision, motion estimation, nonlinear identification, estimation on manifolds, Exterior Differential Systems, Differential Algebraic Equations

## Abstract

We formulate the problem of estimating the motion of a rigid object viewed under perspective projection as the identification of a dynamic model in Exterior Differential form with parameters on a topological manifold.

We first describe a general method for recursive identification of nonlinear implicit systems using prediction error criteria. The parameters are allowed to move slowly on some topological (not necessarily smooth) manifold. The basic recursion is solved in two different ways: one is based on a simple extension of the traditional Kalman Filter to nonlinear and implicit measurement constraints, the other may be regarded as a generalized “Gauss-Newton” iteration, akin to traditional Recursive Prediction Error Method techniques in linear identification. A derivation of the “Implicit Extended Kalman Filter” (IEKF) is reported in the appendix.

The ID framework is then applied to solving the visual motion problem: it indeed is possible to characterize it in terms of identification of an Exterior Differential System with parameters living on a  $C_0$  topological manifold, called the “essential manifold”. We consider two alternative estimation paradigms. The first is in the local coordinates of the essential manifold: we estimate the state of a nonlinear implicit model on a linear space. The second is obtained by a linear update on the (linear) embedding space followed by a projection onto the essential manifold. These schemes proved successful in performing the motion estimation task, as we show in experiments on real and noisy synthetic image sequences.

## 1 Introduction

The “visual motion estimation” is concerned with reconstructing the motion of the viewer relative to the environment from its projections onto the retina (or CCD surface). The task may be

---

\*Research funded by the California Institute of Technology, an AT&T Foundation Special Purpose grant, ONR grant N0014-93-1-0990 and grant ASI-RS-103 from the Italian Space Agency. This work is registered as Technical Report CIT-CDS 94-004, California Institute of Technology, 1994. Submitted to the invited session on “Dynamic Vision, System Theoretical Methods and Control Applications” at the 33rd IEEE conf. on Decision and Control, Florida, 1994.

separated into two steps: first establish which point on the retina corresponds to which across time (correspondence problem), and then estimate the motion of the viewer and the structure of the environment from the correspondence. This classification is rather arbitrary; it is convenient, however, to assume that the correspondence has been solved in order to concentrate on the geometric structure of the visual motion problem. For a review of the existing methods for addressing the correspondence problem, see for example [2].

Visual motion estimation is a key task in many control applications involving the interaction with the environment, such as autonomous robot navigation [14, 16, 15], visual-based tracking/servo [24, 25, 37], visual-based manipulation [4, 25], docking [28, 15], visual-based planning [11], active sensing [59]. In recent years the problem has been addressed using nonlinear estimation/identification techniques [40, 27, 45, 3, 57, 55].

In order to formalize the problem and cast it into a system-theoretic framework, we need to specify a “description” for the scene and for the motion of the viewer. Based on which scene descriptors are used, the existing methods for motion estimation may be classified as point-based, line-based, curve-based or model-based. We are interested in the simplest case when the scene is described by a number of feature points in the euclidean 3D space. For line-based schemes see [68, 62] and references therein. The curve-based approach has been addressed in [1, 61, 9].

The point-based methods may be further classified based on the camera model employed. The simplest cases assume either parallel projection [65, 64, 63, 51] or ideal perspective projection (pinhole model, see [18] for a review). More articulated camera models as general homographies allow parallel and perspective projection as a subcase [3, 60, 20, 54]. We will be mostly concerned with the classical pinhole model, although it is possible to generalize our schemes to more general camera representations and estimate the camera model along with visual motion (self-calibration, see [54, 20]). Recent schemes recover projective, non-metric structure and motion independent on the camera parameters [17, 51, 42].

Motion reconstruction schemes may be further classified based on the data processing technique as either 2-frames schemes (see for example [39, 29, 66]), multiframe-batch methods [65, 60] and recursive algorithms.

In the last decade a variety of schemes has been proposed for reconstructing recursively structure for known motion [40], motion for known structure [21, 22, 6] or both structure and motion [27, 3, 45, 57, 55]. In this paper we unify them into a common framework and highlight the limitations of the model employed, which motivate the new formalization in terms of identification (ID) of Exterior Differential Systems (EDS) [7] which we introduce.

We will then address a general technique for performing the ID of EDS using nonlinear prediction error criteria, and we will apply it to the visual motion problem. We will see how other problems in computational vision may be formulated in the framework of ID of EDS and solved with the technique presented in this paper.

## Organization of the paper

In section 2 we will cast the visual motion problem into a system-theoretic framework in terms of state estimation of a nonlinear dynamical systems with a differentiable state-manifold. Motivated by the structural limitation of the model which *defines* the visual motion problem in the case of feature points in the euclidean 3D space, we will develop a new formalization of the problem in terms of identification of an EDS with the parameters on a topological manifold, called the “essential manifold”. We will also present two examples of other problems in computer vision which may be cast as the identification of an EDS.

In section 4 we will analyze the estimation problem in general form and develop a suboptimal technique for recursive identification of nonlinear implicit systems nonlinear in the parameters using prediction error methods (PEM) [58]. The framework is that of approximate maximum likelihood or least squares identification using observers [33, 8, 31, 10, 48, 47, 32], extended to Differential Algebraic Equations with parameters on topological manifolds. We use a variation of the Extended Kalman Filter for implicit measurement constraints, called IEKF, which is derived in the appendix.

In section 5 we apply the method to the visual motion problem. We propose three schemes for performing the estimation task: the first consists of writing the estimator in the local coordinates of the parameter manifold, and then applying the IEKF. Alternatively we write the update in the embedding space and project it at each step onto the parameter manifold. A third scheme is based on a double iteration and corresponds to the extension of the usual least-squares PEM via Gauss-Newton iteration. The theoretical observability/identifiability of such schemes is addressed in [52].

In section 6 we compare the performance of the three schemes on real and noisy synthetic image sequences.

## 2 Visual motion estimation

In this section we formalize the visual motion problem when the structure of the scene is represented as a set of feature points in the euclidean 3D space. We restrict our attention to “static” environments, or equivalently to portions of the scene which are moving rigidly. In such a case the problem is “defined” by the rigid motion constraint and the perspective map.

### 2.1 Formulation in terms of state estimation

Consider a rigid set of feature points in 3D space with respect to some cartesian frame, for example the one moving with the observer. We call  $\mathbf{X}^i = \begin{bmatrix} X & Y & Z \end{bmatrix}_i^T \in \mathbb{R}^3$  the coordinates of the  $i^{TH}$  point, and we let  $i = 1 : N$ . As the camera moves between two discrete time instants, with rotation  $R$  and translation  $T$ , the coordinates change according to the rigid motion constraint:

$$\mathbf{X}^i(t+1) = R(t)\mathbf{X}^i(t) + T(t) \quad \forall i = 1 : N \quad (1)$$

where motion is represented by  $(R, T) \in SE(3)$  [43].

The camera (or eye) is represented by a map from the 3D space onto some 2D surface. We adopt for simplicity the ideal perspective projection model [5]:

$$\begin{aligned} \pi : \mathbb{R}^3 &\rightarrow \mathbb{R}P^2 \\ \mathbf{X} &\mapsto \pi(\mathbf{X}) \doteq \begin{bmatrix} x & y & 1 \end{bmatrix} = \mathbf{x} \doteq \begin{bmatrix} \frac{X}{Z} & \frac{Y}{Z} & 1 \end{bmatrix}^T; \end{aligned} \quad (2)$$

we measure  $\mathbf{x}$  up to some error:

$$\mathbf{y} = \mathbf{x} + n \quad n \in \mathcal{N}(0, R_n).$$

The representation thus proposed is the very simplest one can imagine; however, we will show that it is not the most appropriate for motion estimation.

The equations (1,2) may be regarded as a dynamical model describing the motion of points in 3D space, having a projection as measurement equation. Motion is the input to the system, and hence the estimator should “invert” the model in order to reconstruct motion from time varying

projection of feature points. Since the initial condition (structure at time zero) is not known, we have a combined “inversion/estimation” problem. It can be shown [52] that any inverse system for (1,2) is essentially instantaneous, hence it does not exploit recursiveness and its benefits. This is due to the fact that the model above is *driftless* [30, 44]; a common trick is then to use *dynamic extension*. We augment (1) with

$$\begin{cases} R(t+1) \doteq R(t) + n_R(t) \\ T(t+1) \doteq T(t) + n_T(t) \end{cases} \quad (3)$$

$n_R \in SO(3)$ ,  $n_T \in \mathbb{R}^3$ . Once inserted motion into the state dynamics we have transformed the motion problem into a state estimation task for a dynamical system with unknown inputs, since we do not know  $n_R$  and  $n_T$ .

If we have a dynamical model available, as for example when the camera is mounted on a moving vehicle, we may exploit it in place of (3). In lack of a mechanical model we may imply a statistical model, for example a fixed order random walk. The extreme case is  $n_T = 0$ ,  $n_R = 0$ , which corresponds to constant velocity motion.

A common model is a first order random walk, which describes a brownian motion. For instance we may assume  $n_T \in \mathcal{N}(0, R_T)$  and  $n_R \doteq e^{\tilde{n}_R \wedge} \in SO(3)$  with  $\tilde{n}_R \in \mathcal{N}(0, R_\Omega)$ . All of these are modeling assumptions, and they must be verified a posteriori.

A fundamental issue in deriving a state estimator (observer) is of course *observability* [30, 44, 34, 35, 36, 50]. The observability of the motion problem is addressed in [52]. The system under investigation (1,2,3) has the peculiarity of not only having a linearization which is not observable, but of also being non locally weakly observable. We need to impose metric constraints on the state manifold in order to achieve local observability; furthermore the observable manifold is covered with a high level of lie-differentiations, which makes the observer purely conditioned.

Note that the model described above is “block-diagonal” with respect to the structure parameters  $\mathbf{X}_i$ , and any observable motion combination can be observed regardless the number of visible points. Indeed it is strongly intuitive that the more points are available, the more robust the perception (estimate) of motion should be. Also note that, once motion has been estimated, structure is *linearly observable* [52] from the model (1,2), and hence a simple EKF will suffice to estimate it, provided that we keep an explicit representation of the second order statistics of the motion estimation errors [45, 57].

These considerations motivate the introduction of a new model, based upon a motion representation which dates back to Longuet-Higgins [39].

## 2.2 Formulation as identification of an exterior differential system

A rigid scene is moving with  $T(t), R(t)$  between two time instants. Then it is immediate to see (fig. 1) that the vector  $\mathbf{X}$ , describing the coordinates of the generic point at time  $t$ , the vector  $\mathbf{X}'$  of coordinates at time  $t+1$  and  $T$ , are coplanar, and therefore their triple product is zero. This is true of course also for  $\mathbf{x}, \mathbf{x}'$  and  $T$ , since  $\mathbf{x}$  is the projective coordinate of  $\mathbf{X}$  and therefore the two are coincident in  $\mathbb{R}^3$ , interpreted as the “ray-space” model [49]. When expressed with respect to a common reference, for example that at time  $t$ , we may write the triple product as<sup>1</sup>

$$\mathbf{x}'^T R(T \wedge \mathbf{x}_i) = 0 \quad \forall i = 1 : N. \quad (4)$$

---

<sup>1</sup>Note that we model rigid motion with  $T, R$  s.t.  $\mathbf{X}' = R(\mathbf{X} - T)$ , for consistency with the notation of [39].

The Essential Constraint for rigid Motion

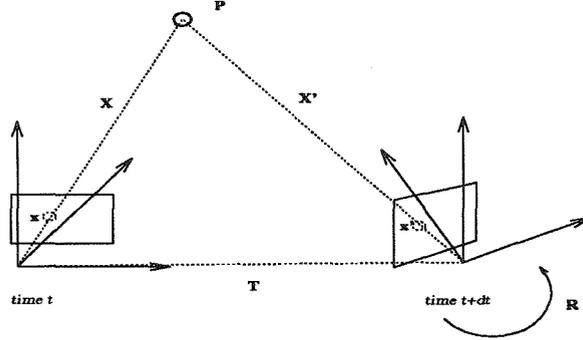


Figure 1: The essential constraint

It turns out that the above constraint is not only a consequence of rigid motion, but it also suffices to characterize it, once eight or more constraints are given [41, 39]. The operator

$$T\wedge = \begin{bmatrix} 0 & -T_3 & T_2 \\ T_3 & 0 & -T_1 \\ -T_2 & T_1 & 0 \end{bmatrix} \doteq S.$$

belongs to  $so(3)$  [43]. Following Longuet-Higgins [39] we call

$$\mathbf{Q} \doteq R(T\wedge)$$

so that the above constraint, which we now call the “essential constraint”, becomes

$$\mathbf{x}'_i{}^T \mathbf{Q} \mathbf{x}_i = 0 \quad \forall i = 1 \dots N. \tag{5}$$

Estimating motion corresponds to identifying the model

$$\begin{cases} (\mathbf{Q} \mathbf{x}_i)^T \mathbf{x}'_i = 0 & \mathbf{Q} \in E \\ \mathbf{y}_i = \mathbf{x}_i + \mathbf{n}_i & \mathbf{n}_i \in \mathcal{N}(0, R_{n_i}), \end{cases} \quad \forall i = 1 \dots N$$

which is in the form of an Exterior Differential System [7]; the parameters  $T, R$  are encoded in  $\mathbf{Q}$ .

Since the constraint (5) is linear in  $\mathbf{Q}$ , we use the (improper) notation

$$\chi_{\mathbf{x}'(t), \mathbf{x}(t)} \mathbf{Q}(t) = 0$$

where  $\chi$  is an  $N \times 9$  matrix combining  $x_i, x'_i$  and  $\mathbf{Q}$  is interpreted as a nine dimensional vector. The generic row of  $\chi$  has the form  $[x_1 x'_1 \ x_2 x'_2 \ x_1 x'_1 \ x_1 x'_2 \ x_2 x'_2 \ x'_2 \ x_1 \ x_2 \ 1]$ .

### 2.2.1 The Essential Space

A rigid motion may be represented as an element of the Lie group  $SE(3)$ , which is naturally embedded in  $\mathbb{R}^{4 \times 4}$  via homogeneous coordinates:

$$\begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \in SE(3) \subset \mathbb{R}^{16}.$$

We have indeed seen that rigid motion may be encoded using the essential constraint (5) based on the  $3 \times 3$  matrix  $\mathbf{Q} \doteq R(T\wedge) \in \mathbb{R}^9$ . Since we can reconstruct translation only up to a scale factor, we may restrict  $\mathbf{Q}$  to belong to  $\mathbb{R}P^8$  instead than  $\mathbb{R}^9$ . It is customary to set the norm of translation to be unitary; this can be done without loss of generality, as long as translation is not zero. The zero-norm translation case can be dealt with separately, and we will discuss it later. Now for simplicity we assume  $\|\mathbf{Q}\|_2 = \|T\| = 1$ . The matrix  $\mathbf{Q}$  belongs to the space

$$\tilde{E} \doteq \{RS | R \in SO(3), S \doteq T\wedge \in so(3), \|T\| = 1\} \subset \mathbb{R}^9$$

which is called the *essential space*. The essential space encodes rigid motion in a more compact way than  $SE(3)$ , the price being that we loose the smooth group structure. Indeed, as shown in [55, 52], a slight modification of  $\tilde{E}$  proves to have the structure of a topological manifold. For let  $d_{x,x'}(\mathbf{Q})$  be the triangulation function which gives the depth of a point from its motion  $\mathbf{Q}$  and its projective coordinates  $\mathbf{x}, \mathbf{x}'$ . Then  $E \doteq \tilde{E} \cap d_{x,x'}^{-1}(\mathbb{R}_+^2)$  is a topological manifold called the “essential manifold” [55]. Call

$$\begin{aligned} \Phi : E &\rightarrow \mathbb{R}^5 \\ \mathbf{Q} &\mapsto [V, \Omega]^T \end{aligned}$$

a chart of the local coordinates atlas of the essential manifold ( see [55] for an explicit characterization of  $\Phi$ );  $[V, \Omega]^T$  represent the canonical (exponential) local coordinates of  $(T, R) \in SE(3)$  via

$$R \doteq e^{(\Omega\wedge)} \tag{6}$$

$$T \doteq \frac{1}{\|\Omega\|} \left[ (I - e^{(\Omega\wedge)}) \Omega \wedge + \Omega \Omega^T \right] V. \tag{7}$$

$E$  also has the structure of an algebraic variety [41], which we will not discuss in this paper.

### 2.2.2 Motion representation on the essential space

A rigid motion with unit norm translation may be represented as an element of the essential manifold  $E$ . For non-unit translations (but still positive norm), it is sufficient to scale  $\mathbf{Q}$  to  $\mathbf{Q}/\|T\|$ , since the singular values of  $\mathbf{Q}$  are  $\{\|T\|, \|T\|, 0\}$  (see appendix B). At each time instant we have a set of  $N$  constraints in the form

$$\chi_{x'(t),x(t)}(\mathbf{Q}(t)) = 0,$$

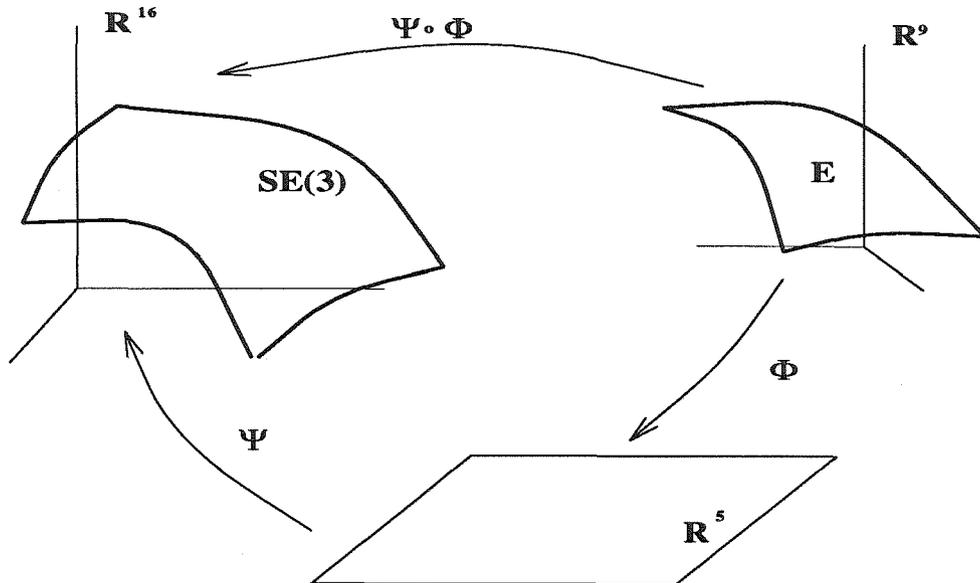
therefore  $\mathbf{Q}$  lies at the intersection between the essential manifold and the linear variety  $\chi_{x'(t),x(t)}^{-1}(0)$  (see fig. 2).

Note that even after imposing unit norm there is still a sign indeterminacy in  $\mathbf{Q}$ , which accounts for the two solutions  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$ . These solutions become four when transformed to local coordinates. These ambiguities can be overcome by imposing the positive depth constraint as it is done in the definition of the essential manifold [52, 55]: in fact, out of the four different combinations of  $R$  and  $T$ , only one corresponds to points which are in front of the observer [67, 23, 46, 19].

As time goes by, the point  $\mathbf{Q}(t)$ , corresponding to the actual motion, describes a trajectory on  $E$  (and a corresponding one in local coordinates):

$$\mathbf{Q}(t) \mapsto \mathbf{Q}(t+1) \doteq \mathbf{Q}(t) + n_{\mathbf{Q}}(t).$$

The last equation is in fact just a *definition* of the right-hand side, since we do not know  $n_{\mathbf{Q}}(t)$ . For now we will consider the previous equation as a discrete time dynamical model for  $\mathbf{Q}$  on the



essential manifold, having  $n_{\mathbf{Q}}$  as *unknown* input. If we accompany it with the essential constraint, we get

$$\begin{cases} \mathbf{Q}(t+1) = \mathbf{Q}(t) + n_{\mathbf{Q}}(t) & \mathbf{Q} \in E \\ 0 = \chi_{y'(t), y(t)} \mathbf{Q}(t) \\ \mathbf{y}_i = \mathbf{x}_i + n_i & \forall i = 1 \dots N \end{cases} \quad (8)$$

where  $n(t) \in \mathcal{N}(0, R_n)$ . Note that now the visual motion problem is defined as the estimation of the state of the above model, which is defined on the essential manifold. As it can be seen the system is “linear” (both the state equation and the essential constraint are linear in  $\mathbf{Q}$ ); however,  $E$  is not a linear space. In the section 4 we will develop a general tool for addressing the identification problem.

### 3 Other problems which may be formulated as ID of EDS

In the present section we consider, as an example, two additional problems which may be cast in the framework of identification of exterior differential systems with parameters on a manifold.

The first problem is “camera self calibration”, which consists of the dynamic estimation of the camera model along with the motion parameters. It has been shown [54] that the problem may be formulated as an extension of the scheme derived in the previous section when the essential manifold is substituted by the space of the “fundamental matrices” [20, 54].

The second problem is the recovery of the direction of translation using subspace methods: [53] provides a way of formalizing the problem as identification of an exterior differential system with parameters on a sphere [26, 53].

#### 3.1 Dynamic self-calibration

In so far we have represented the camera as an ideal central projection of unit focal length. When the camera model is a more general projective transformation in  $\mathbb{R}P^2$ , eq. (5) does not hold.

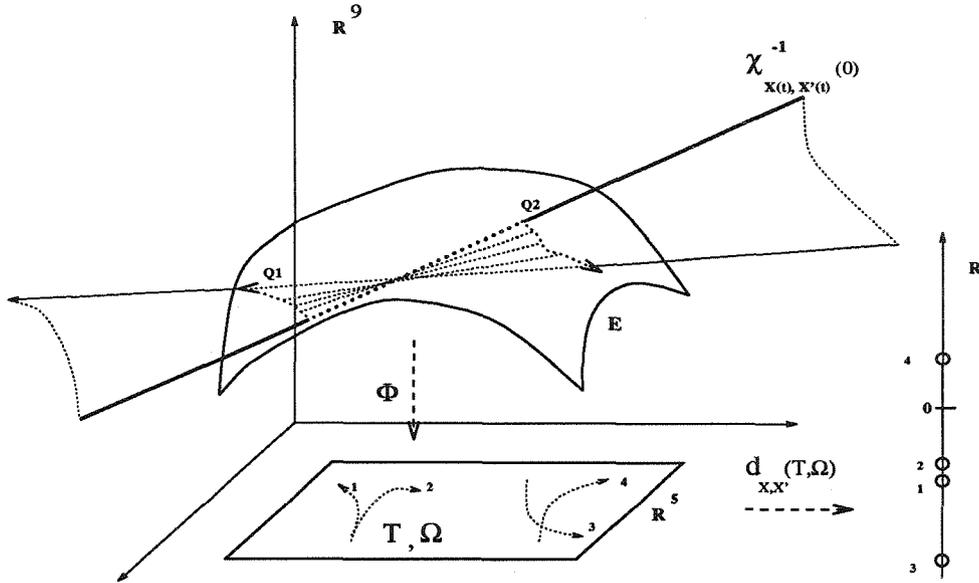


Figure 2: Structure of the motion problem on the Essential Space

However, a similar constraint may be derived based on the epipolar geometry as

$$\mathbf{x}_i'^T F \mathbf{x}_i = 0 \quad \forall i = 1 \dots N. \quad (9)$$

The matrix  $F$  is called “fundamental matrix”; it defines the relation between each point  $i$  and its corresponding epipolar line [20]. If the camera is represented (in homogeneous coordinates) as a  $3 \times 4$  matrix  $\begin{bmatrix} \mathbf{A} & | & 0 \end{bmatrix}$  where

$$\mathbf{A} \doteq \begin{bmatrix} fs_x & 0 & -i_0 \\ 0 & fs_y & -j_0 \\ 0 & 0 & 1 \end{bmatrix}$$

is the internal parameter matrix<sup>2</sup>, then it can be shown that

$$\mathbf{Q} \doteq \mathbf{A}^T F \mathbf{A} \in E \quad (10)$$

is an *essential matrix*.

Faugeras et al. [20] propose to estimate the matrix  $F$  from the constraint (9), and then *impose the structure of the fundamental matrix* (10) *a posteriori* by solving a set of polynomial equations known as Kruppa equations. Such equations are indeed ill conditioned, and the scheme is very sensitive to noise. Furthermore, temporal coherence of the camera model is not exploited.

If we substitute (10) into (9) we have a model

$$\begin{cases} (\mathbf{A}^{-T} \mathbf{Q} \mathbf{A}^{-1} \mathbf{x}_i)^T \mathbf{x}_i = 0 & \mathbf{Q} \in E \\ \mathbf{y}_i = \mathbf{x}_i + \mathbf{n}_i & \forall i. \end{cases}$$

<sup>2</sup> $f$  is the focal length,  $(i_0, j_0)$  the coordinates of the optical center and  $(s_x, s_y)$  the pixel sizes along the image plane coordinates. The deviation from  $90^\circ$  of the angle between the optical axis and the CCD surface is usually on the order of  $1^\circ$ , and we may therefore neglect it.

Estimating the camera parameters, along with rigid motion, may then be formulated as *identification of the above exterior differential model*, where the parameters are on the manifold  $E \times AF$ , and  $AF$  is the set of affine transformations of  $\mathbb{R}^2$  represented in homogeneous coordinates. This formulation has been derived in [54].

### 3.2 Recovering rigid motion using subspace methods

Consider the following expression of the motion field:  $\dot{\mathbf{x}}_i(t) = [\tilde{\mathcal{A}}_i(\mathbf{x}_i, V(\theta, \phi)) | \mathcal{B}_i(\mathbf{x})] \begin{bmatrix} \frac{1}{z_i(t)} \\ \Omega_i(t) \end{bmatrix}$ ,

where  $\tilde{\mathcal{A}}(\mathbf{x}, V) \doteq \begin{bmatrix} V_1 - xV_3 \\ V_2 - yV_3 \end{bmatrix}$  and  $V \in \mathbf{S}^2$  is represented in local coordinates as  $V(\theta, \phi)$ .  $\mathcal{B} \doteq \begin{bmatrix} -xy & 1+x^2 & -y \\ -1-y^2 & xy & x \end{bmatrix}$ . If we observe  $N$  points we may write  $\dot{\mathbf{x}} = \tilde{\mathcal{C}}(V, \mathbf{x})[\frac{1}{z_1}, \dots, \frac{1}{z_N}, \Omega]^T \doteq \tilde{\mathcal{C}}\mathbf{d}^T$ , where

$$\tilde{\mathcal{C}}(V, \mathbf{x}) \doteq \begin{bmatrix} \tilde{\mathcal{A}}_1 & & \mathcal{B}_1 \\ & \ddots & \vdots \\ & & \tilde{\mathcal{A}}_N & \mathcal{B}_N \end{bmatrix}.$$

Under the usual rank conditions, we may compute the least squares approximation of  $\mathbf{d}$  as

$$\hat{\mathbf{d}} = \tilde{\mathcal{C}}^\dagger \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \doteq \tilde{\mathcal{C}}^\dagger \dot{\mathbf{x}}$$

and therefore the motion field specifies the constraint [26]

$$\tilde{\mathcal{C}}^\perp(V, \mathbf{x})\dot{\mathbf{x}} = 0$$

where  $\perp$  indicates the orthogonal complement. Heeger and Jepson [26] propose to estimate the direction of translation by minimizing the two norm of the above constraint over  $V \in \mathbf{S}^2$ . They perform such a minimization by extensive search over all possible directions  $\theta, \phi$ .

Indeed it is immediate to see [53] that the problem of estimating the direction of translation may be rephrased as the problem of *identifying the following exterior differential system, with parameters  $V$  on a sphere*:

$$\begin{cases} \tilde{\mathcal{C}}^\perp(V, \mathbf{x})\dot{\mathbf{x}} = 0 & V \in \mathbf{S}^2 \\ \mathbf{y}_i = \mathbf{x}_i + n_i & \forall i. \end{cases}$$

This problem can be solved in a principled manner using the results of the next section, without requiring any extensive search or sampling of the sphere. See [53] for more details.

## 4 Identification of nonlinear implicit systems with prediction error criteria

Suppose  $\{x(t)\} \in \mathbb{R}^N$  is a trajectory on a linear state-space, which is subject to an implicit dynamic constraint of the form

$$h[x(t), dx(t), a] = 0 \quad x(0) = x_0 \quad (11)$$

where  $a \in M$  are some unknown parameters which can move (slowly) on a topological (not necessarily smooth) manifold. Call  $\alpha \doteq \psi(a) \in \mathbb{R}^m$  the local coordinates correspondent of  $a$ . Suppose we are able to measure  $x$  up to some white, zero-mean gaussian noise:

$$y(t) = x(t) + n(t) \quad n \in \mathcal{N}(0, R_n).$$

We are interested in identifying the parameters  $a$  recursively from the measurements  $\{y(t)\}$  based on the minimization of some cost function of the prediction error (for a classical treatment of PEM for linear explicit models see for example [58]).

A common paradigm for PEM identification consists in forcing a Kalman Filter to work as a parameter estimator. The state of the filter is augmented with the unknown parameters, which are described using a low order random walk model. The sequel of modeling operations is described as

$$\text{from } \{x(t)\} \text{ and } \{\dot{x} = ax\} \text{ identify } a \quad \text{via observing} \quad \begin{cases} a(t+1) = a(t) + n_a \\ y' = ay + \tilde{n} \end{cases}$$

where  $y, y'$  are noisy measurements of  $x, \dot{x}$  and  $\tilde{n}$  is a residual which can be characterized in terms of the noise  $n$ . Our aim is to generalize this paradigm to nonlinear implicit dynamics and parameters living in topological manifolds. In the following we will consider discrete time dynamics, which fall in the same scheme described above, once we substitute  $\dot{x}, y'$  with  $x(t+1), y(t+1)$ .

First we proceed in analogy with the linear-explicit case: we describe the local coordinates of the parameters as first order random walk, and use the dynamic constraint as an implicit measurement constraint:

$$\begin{cases} \alpha(t+1) = \alpha(t) + n_\alpha(t) & \alpha(0) = \alpha_0 \\ h[y(t) - n(t), y(t-1) - n(t-1), \psi^{-1}(\alpha(t))] = 0 \end{cases} \quad (12)$$

where we have substituted the index  $t$  with  $t-1$  in the measurements  $\{y\}$  (or equivalently the estimator runs with one step delay). The noise process  $\{n(t)\}$  induces a residual in the measurement equation: if we approximate  $x(t)$  with  $y(t)$ , in general we will observe  $h[y(t), y(t-1), a] = \tilde{n}$ , where  $\tilde{n}$  depends on  $n(t), n(t-1), y(t), y(t-1)$  and  $a$ . This residual is exactly the prediction error (or pseudo-innovation) when choosing a least-squares criterion in the PEM.

Let us collect the measurements into a vector  $\bar{y}^T(t) \doteq [y^T(t) \ y^T(t-1)]^T$ , and idem with  $\bar{n}(t) \doteq [n^T(t) \ n^T(t-1)]^T$ . Our task is to estimate  $\alpha$  from the model

$$\begin{cases} \alpha(t+1) = \alpha(t) + n_\alpha(t) & \alpha(0) = \alpha_0 \\ h[\bar{y}(t) - \bar{n}(t), \psi^{-1}(\alpha(t))] = 0. \end{cases} \quad (13)$$

In order to follow the course of the linear-explicit case, we have to solve a number of problems:

1. the noise  $\bar{n}$  is not white, since  $E[\bar{n}(t)\bar{n}^T(s)] = \begin{bmatrix} R_n\delta(t-s) & R_n\delta(t-s+1) \\ R_n\delta(t-s-1) & R_n\delta(t-s) \end{bmatrix}$
2. the error  $\bar{n}$  does not appear additively in the measurement equation
3. the measurement equation is nonlinear and implicit.

The Extended Kalman Filter (EKF) [33, 8, 31] is a general-purpose extension to nonlinear systems of the traditional Kalman Filter. It is based on a variational model about the best current trajectory. The systems is linearized at each step around the current estimate in order to calculate a correcting gain; the update of the previous estimate is then performed on the original (nonlinear) equations. In order to solve step 3. we need to further extend the EKF to cope with the implicit measurement

constraint. This is done in appendix A. We call the result Implicit Extended Kalman Filter (IEKF); some variations of the scheme have been used in different applications in the last years, see for example [13]. The derivation is based on the simple fact that the variational model about the current trajectory is *linear and explicit*, so that the a pseudo-innovation process may be defined analogously to the explicit case. Note that the local coordinates chart of the parameter manifold  $\psi$  enters into the measurement equation, and therefore it is differentiated in order to compute the gain. However, the update equation is calculated on the actual nonlinear model, so that discontinuities of the derivative of  $\psi^{-1}$ , which may happen when switching from one chart to another, are well tolerated.

The derivation of the IEKF in appendix A also solves step 2. The residual of the measurement equation  $\tilde{n}$ , which is in fact the pseudo-innovation of the filter, is characterized in terms of  $\bar{n}$ , provided that the last is white, zero-mean and uncorrelated with  $n_\alpha$ .

In the following section we will show how to whiten  $\bar{n}$  and therefore reduce the problem to a form suitable for using the IEKF. Later we will see how the problem simplifies by assuming that  $\bar{n}$  is white.

#### 4.1 Uncorrelating the model from the measurements

Consider a first order expansion of the measurement equation about the point  $\bar{y}(t), \alpha(t)$ :

$$h[\bar{y}(t), \psi^{-1}(\alpha(t))] - D_+(t)n(t) - D_-(t)n(t-1) = \mathcal{O}(\|\bar{n}\|^2) \cong 0$$

where we have defined

$$D_+(t) \doteq \left( \frac{\partial h[x(t), x(t-1), a]}{\partial x(t)} \right)_{|\bar{y}(t), \psi^{-1}(\alpha(t))} \quad (14)$$

$$D_-(t) \doteq \left( \frac{\partial h[x(t), x(t-1), a]}{\partial x(t-1)} \right)_{|\bar{y}(t), \psi^{-1}(\alpha(t))} \quad (15)$$

Here the residual  $\tilde{n}(t) = -D_+(t)n(t) - D_-(t)n(t-1)$  is clearly correlated. In order to estimate the dynamics of  $n(t)$ , we may insert it into the state dynamics: call  $z(t) \doteq n(t-1)$ .

$$\begin{cases} \alpha(t+1) = \alpha(t) + n_\alpha(t) & \alpha(0) = \alpha_0 \\ z(t+1) = n(t) & z(0) = 0 \\ 0 = h[\bar{y}(t), \psi^{-1}(\alpha(t))] - D_-(t)z(t) + w(t) \end{cases} \quad (16)$$

where we have defined  $w(t) \doteq -D_+(t)n(t)$ . Now the measurement error  $w$  is white; however, it is correlated with the model error  $v \doteq [n_\alpha^T, n^T]^T$ . We may therefore project the model error onto the span of the measurement error  $H(w)$  in order to make the two orthogonal. We define  $\tilde{v}(t) \doteq v(t) - E[v(t)|H(w)]$ . Since  $w(t)$ ,  $n(t)$  and  $n_\alpha(t)$  are white, it is easily seen that  $E[v(t)|H(w)] = E[v(t)|w(t)] = \Sigma_{vw}\Sigma_w^{-1}w(t)$ .  $\Sigma_{vw}$  and  $\Sigma_n$  are variance/covariance matrices. If we define

$$Q(t) \doteq \begin{bmatrix} R_\alpha & 0 \\ 0 & R_n \end{bmatrix} \quad (17)$$

$$R(t) \doteq D_+(t)R_n(t)D_+^T(t) \quad (18)$$

$$S(t) \doteq \begin{bmatrix} 0 \\ -R_n(t)D_+^T(t) \end{bmatrix} \quad (19)$$

it is easy to see that  $\Sigma_{vw}\Sigma_w^{-1} = S(t)R^{-1}(t)$ ; furthermore  $\Sigma_{\tilde{v}} \doteq \tilde{Q}(t) = Q(t) + S(t)R^{-1}(t)S^T(t)$ . Now  $\tilde{v} \doteq v - SR^{-1}w$  is by construction orthogonal (uncorrelated) to  $w$ .

## 4.2 A model for PEM identification of nonlinear implicit models

In the previous paragraph we have derived a substitution for the model error which is by construction uncorrelated with the measurement error. Therefore we may write a new model which satisfies the conditions of appendix A:

$$\begin{cases} \alpha(t+1) = \alpha(t) & \alpha(0) = \alpha_0 \\ z(t+1) = K(t) (h[\bar{y}(t), \psi^{-1}(\alpha(t))] - D_-(t)z(t)) & z(0) = 0 \\ 0 = h[\bar{y}(t), \psi^{-1}(\alpha(t))] - D_-(t)z(t) + w(t) \end{cases} \quad (20)$$

where we have defined

$$K(t) \doteq R_n(t)D_+(t)(D_+(t)R_n(t)D_+^T(t))^{-1} \quad (21)$$

$$w(t) \doteq -D_+(t)n(t). \quad (22)$$

By applying the results of appendix A, we have a pseudo-optimal PEM identification scheme described by the following iteration:

### Prediction step

$$\begin{cases} \hat{\alpha}(t+1|t) = \hat{\alpha}(t|t) & \hat{\alpha}(0|0) = \alpha_0 \\ \hat{z}(t+1|t) = K(t) (h[\bar{y}(t), \hat{\alpha}(t|t)] - D_-(t)\hat{z}(t|t)) & \hat{z}(0|0) = 0 \\ P(t+1|t) = F(t)P(t|t)F^T(t|t) + \hat{Q}(t) & P(0|0) = P_0 \end{cases} \quad (23)$$

where  $F \doteq \begin{bmatrix} I & 0 \\ K(t)(C(t) - [0 \quad D_-(t)]) \end{bmatrix}$  and  $C(t) \doteq \left( \frac{\partial h[\bar{y}, \psi^{-1}(\alpha)]}{\partial \alpha} \right)_{|\hat{\alpha}(t|t), \bar{y}(t)}$ .

### Update step

$$\begin{cases} \begin{bmatrix} \hat{\alpha}(t+1|t+1) \\ \hat{z}(t+1|t+1) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}(t+1|t) \\ z(t+1|t+1) \end{bmatrix} + L(t+1) (h[\bar{y}(t), \hat{\alpha}(t+1|t)] - D_-(t)\hat{z}(t+1|t)) \\ P(t+1|t+1) = \Gamma(t+1)P(t+1|t)\Gamma^T(t+1) + L(t+1)D_+(t)R_n(t+1)D_+^T(t)L^T(t+1) \end{cases} \quad (24)$$

where

$$L \doteq PC^T\Lambda^{-1} \quad (25)$$

$$\Lambda \doteq CPC^T + D_+(t)R_n(t+1)D_+^T(t) \quad (26)$$

$$\Gamma \doteq I - LC \quad (27)$$

Note that we are trying to estimate a process  $\{z(t)\}$  which is nearly white noise ( $n(t)$  is correlated only within one step). Furthermore if we expect a large number of measurement components  $n$ , the cost in updating a large state and tuning a large number of model-variance parameters may be relevant. In practical applications the approximation  $\tilde{n}$  as white noise are often best conditioned. In the following section we show how the structure of the filter simplifies in such a case.

## 4.3 A simplified version: approximate Least Squares PEM identification

In this section we report the equations of the parameter estimator which are obtained supposing that the residual  $\tilde{n}$  is white. This correspond to applying the results of appendix A directly to the model of eq. (13):

### Prediction step

$$\begin{cases} \hat{\alpha}(t+1|t) = \hat{\alpha}(t|t) \\ P(t+1|t) = P(t|t) + R_\alpha(t) \end{cases} \quad \begin{cases} \hat{\alpha}(0|0) = \alpha_0 \\ P(0|0) = P_0 \end{cases} \quad (28)$$

### Update step

$$\begin{cases} \hat{\alpha}(t+1|t+1) = \hat{\alpha}(t+1|t) + L(t+1)h[\bar{y}(t), \psi^{-1}(\hat{\alpha}(t+1|t))] \\ P(t+1|t+1) = \Gamma(t+1)P(t+1|t)\Gamma^T(t+1) + L(t+1)D_+(t)R_n(t+1)D_+^T(t)L^T(t+1) \end{cases} \quad (29)$$

where now we the quantities  $L, \Lambda$  and  $\Gamma$  are defined according to appendix A. Note that we have reduced the size of the state from  $n + m$  down to  $m$ .

### Detecting outliers

Note that each component of the pseudo-innovation is a measure of the consistency of each datum with the current parameter estimates. This proves useful when applied to the motion problem because it allows us to easily segment the scene into a number of independently moving objects [56].

## 4.4 An iterative scheme for computing the update

The IEKF update seen in the previous section may be substituted with a Gauss-Newton iteration, as it is customary in recursive ID of linear models:

$$\hat{\alpha}(k+1) = \hat{\alpha}(k) - L_{NR}(k)h(\hat{\alpha}(k))$$

where  $L_{NR} = J_h^\dagger(\hat{\alpha}(k))$  and  $J_h$  is the jacobian of  $h$ .

Note that at each fixed time we could perform a Newton-Raphson iteration on the function  $h(\bar{y}, \alpha)$ , for which local convergence results are known as well as bounds on the convergence rate. This suggests, as an alternative to the IEKF, to fix  $t$  and perform a Newton-Raphson iteration along the  $k$  coordinate. Once this is done we propagate the estimate across time with an iteration which now is *linear*, and has all the desirable asymptotic properties.

### 4.4.1 Iteration at each fixed time

At each time instant a new set of measurements  $\bar{y}$  becomes available. The dynamic constraint imposes

$$h[\bar{y}(t), \alpha] = 0 \quad \forall t$$

Define  $T_\alpha h : \mathbb{R}^m \rightarrow \mathbb{R}^N$  to be the derivative of the map  $h$  and  $J_h(\alpha)$  the Jacobian matrix calculated at the point  $\alpha$ . Suppose that there exists some  $\alpha^*$  such that  $h(\bar{y}(t), \alpha^*) = 0$  for our particular (fixed)  $t$ . Then we may write a first order expansion around the point  $\alpha^*$ , starting from some point  $\alpha_0$  (we neglect time indices for the remainder of this section); the resulting iteration, which is obtained by neglecting the second order term of the expansion, is defined by

$$h[\alpha_k] \doteq J_h(\alpha_k)[\alpha_{k+1} - \alpha_k].$$

At each iteration we solve for  $Y$  the linear problem

$$J_h(\alpha_k)Y = h[\alpha_k]$$

and then define  $\alpha_{k+1} \doteq \alpha_k + Y$ . In general, also due to noise, we can expect  $h[\alpha_k] \notin \text{Im}(J_h(\alpha_k))$ , so that we will be seeking for  $Y$  such that  $J_h(\alpha_k)Y$  is the projection of  $h[\alpha_k]$  onto the range space of  $J_h(\alpha_k)$ :

$$\alpha_{k+1} \doteq \alpha_k - L_{NR}(k)h[\alpha_k].$$

where  $L_{NR}(k) \doteq \left( J_h^T(\alpha_k)J_h(\alpha_k) \right)^{-1} J_h^T(\alpha_k)$ . The map defined by the right-hand side of the above equation is contractive as long as  $J_h(\alpha_k)$  has full rank, in which case the scheme is guaranteed to converge to some (possibly local) minimum.

At each time the scheme will converge to some  $\alpha^*$ , which best explains the noisy measurements  $y_i, y'_i$ ; hence we have  $\alpha^* = \alpha + n_\alpha$  where  $n_\alpha$  is a noise term whose variance can be inferred from the variance of  $n_i$  and a linearization of the scheme about zero-noise. The estimate obtained at each fixed time, together with its variance, is fed to a time-integration step, which we describe next.

#### 4.4.2 Propagation along time: disambiguation of local minima

At each fixed time the iteration along  $k$  converges to a fixed point  $\alpha^*(t)$ , then we may propagate the information across time with a similar iteration:

$$\hat{\alpha}(t+1) = \hat{\alpha}(t) + L(t)[\alpha^*(t) - \hat{\alpha}(t)]$$

which implements a linear Kalman filter based upon the model

$$\alpha(t+1) = \alpha(t) + n_0(t) \tag{30}$$

$$\alpha^*(t) = \alpha(t) + n_\alpha(t) \tag{31}$$

where  $n_0$  is the error of the random walk model for motion, which we assume to be white zero-mean and gaussian, and  $n_\alpha$  is the error made by the fixed-time iteration.  $L$  is the usual linear Kalman gain [33, 31]. The above model has all the desirable properties, as it satisfies the conditions of the fundamental theorem of the asymptotic theory of Kalman Filtering.

Suppose now that the  $k$ -iteration has converged to a local minimum, which is compatible with the current observations. At the next step the  $t$ -iteration will predict an estimate which is in general no longer compatible with the current observations. This should help to disambiguate local minima as the measurements accumulate in time.

## 5 Application to the visual motion estimation problem

We have seen in the previous sections that motion estimation may be regarded as estimation of the state of a system of a difference equations on the essential manifold having unknown inputs.

The first approach we describe consists in composing equations (8) with the local coordinate chart  $\Phi$ , ending up with a *nonlinear* dynamical model for motion in  $\mathbb{R}^5$ . At this point we have to make some assumptions about motion: since we do not have any dynamical model, we will assume a statistical model. In particular we will assume that motion is a *first order random walk* (brownian motion) in  $\mathbb{R}^5$  (see fig. 3 left). The problem then becomes that of estimating the state of a nonlinear system driven by white, zero-mean gaussian noise. This will be done using the technique developed in the previous section.

In the second approach we change the model for motion: in particular we assume motion to be a *first order random walk in  $\mathbb{R}^9$  projected onto the essential manifold* (see fig. 3 left). We will see that this leads to a method for estimating motion via solving at each step a *linear estimation* problem

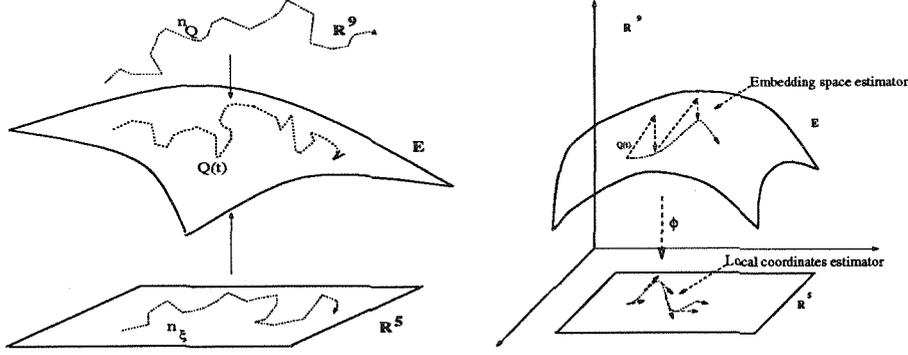


Figure 3: (Left) Model of motion as a random walk in  $\mathbb{R}^5$  lifted to the manifold or as a random walk in  $\mathbb{R}^9$  projected onto the manifold. (Right) Estimation on the Essential Space

in the linear embedding space and then “projecting” the estimate onto the essential manifold (see fig. 3 right).

It is very important to understand that these are modeling assumptions about motion which can be validated only a posteriori. In general we observe that the first method solves a strongly nonlinear problem with techniques which are based upon linearization of the system about the current reference trajectory, so that the linearization error may be relevant. The second method does not involve any linearization, while it imposes the constraint of belonging to the essential manifold in a weaker way. This approach has indeed a very transparent structure which can be studied in full detail.

The third method is based upon splitting the iteration according to the recursive Gauss-Newton scheme illustrated in the previous section.

The next three sections are devoted to describing these three techniques. Note that each method produces, together with the motion estimates, the variance of the estimation error, which is to be used by the subsequent modules of the structure and motion estimation scheme [57].

## 5.1 Local coordinates estimator

Consider composing the system (8) with the map  $\Phi$ :

$$\begin{aligned} \Phi : E &\rightarrow S^2 \times \mathbb{R}^3 \sim \mathbb{R}^5 \\ \mathbf{Q} &\mapsto \xi \doteq \begin{bmatrix} V \\ \Omega \end{bmatrix} \end{aligned}$$

where  $T$  is expressed in spherical coordinates for radius one, for convenience of representation. Then the system in local coordinate becomes

$$\xi(t+1) = \xi(t) + n_\xi(t); \quad \xi(t_0) = \xi_0 \quad (32)$$

$$0 = \chi_{y(t), y'(t)} \mathbf{Q}(\xi(t)) + \tilde{n}(t). \quad (33)$$

As we said we model motion  $\{\xi\}$  as a first order random walk:  $n_\xi(t) \in \mathcal{N}(0, R_\xi)$  for some  $R_\xi$  which is referred to as variance of the model error. While the above assumption is rather arbitrary and can be validated only a posteriori, it is often safe to assume that the noise in the measurements  $y(t)$  is a white zero-mean gaussian process with variance  $R_n$ .

The system above is now in a form suitable for using an Implicit Extended Kalman Filter (EKF). Finally the equations of the estimator can be summarized: call  $C \doteq \left( \frac{\partial \chi \mathbf{Q}}{\partial \xi} \right)$  and  $D \doteq \left( \frac{\partial \tilde{\chi} \mathbf{Q}}{\partial \mathbf{x}} \right)$ .

**Prediction step:**

$$\hat{\xi}(t+1|t) = \hat{\xi}(t|t) ; \hat{\xi}(0|0) = \xi_0 \quad (34)$$

$$P(t+1|t) = P(t|t) + R_\xi ; P(0|0) = P_0 \quad (35)$$

**Update step:**

$$\hat{\xi}(t+1|t+1) = \hat{\xi}(t+1|t) - L(t+1)\chi\mathbf{Q}(\hat{\xi}(t+1|t), t) \quad (36)$$

$$P(t+1|t+1) = \Gamma(t+1)P(t+1|t)\Gamma^T(t+1) + L(t+1)R_{\tilde{n}}(t+1)L^T(t+1) \quad (37)$$

**Gain:**

$$L(t+1) = P(t+1|t)C^T(t+1)\Lambda^{-1}(t+1) \quad (38)$$

$$\Lambda(t+1) = C(t+1)P(t+1|t)C^T(t+1) + R_{\tilde{n}}(t+1) \quad (39)$$

$$\Gamma(t+1) = I - L(t+1)C(t+1) \quad (40)$$

**Innovation variance:**

$$R_{\tilde{n}}(t+1) = D(t+1)R_n D^T(t+1) \quad (41)$$

Note that  $P(t|t)$  is the variance of the motion estimation error which is used as variance of measurement error by the subsequent modules of the motion and structure estimation scheme. This formulation was first introduced by Di Bernardo et al. [13] in a slightly different form.

## 5.2 The essential estimator

Suppose that motion, instead of being a random walk in  $\mathbb{R}^5$ , is represented in the essential manifold as the “projection” of a random walk through  $\mathbb{R}^9$  (see fig. 3). The “projection” operator onto the space  $E$  is denoted by  $pr_{\langle E \rangle}(\cdot)$ :

$$\begin{aligned} pr_{\langle E \rangle} : \mathbb{R}^3 \times 3 &\rightarrow E \\ M &\mapsto \mathbf{U} \text{diag}\{1, 1, 0\} \mathbf{V}^T \end{aligned} \quad (42)$$

where  $\mathbf{U}, \mathbf{V}$  are defined by the Singular Value Decomposition of  $M = \mathbf{U}\Sigma\mathbf{V}^T$ . The fact that this operator maps onto the essential manifold is proved in appendix B. Note that the projection minimizes the Frobenius norm and the 2-norm of the distance from a point in  $\mathbb{R}^{3 \times 3}$  to the essential manifold [23, 41, 69].

Now we define the operator  $\oplus$  that takes two elements in  $\mathbb{R}^{3 \times 3}$ , sums them and then projects the result onto the essential manifold:

$$\begin{aligned} \oplus : \mathbb{R}^{3 \times 3} \times \mathbb{R}^{3 \times 3} &\rightarrow E \\ M1, M2 &\mapsto \mathbf{Q} = pr_{\langle E \rangle}(M1 + M2) \end{aligned}$$

where the symbol  $+$  is the usual sum in  $\mathbb{R}^{3 \times 3}$ . With the above definitions our model for motion becomes simply

$$\mathbf{Q}(t+1) = \mathbf{Q}(t) \oplus n_{\mathbf{Q}}(t) \quad (43)$$

where  $n_{\mathbf{Q}}(t) \in \mathcal{N}(0, R_{n_{\mathbf{Q}}})$  is represented by a white zero-mean gaussian noise in  $\mathbb{R}^9$ . If we couple the above equation with (8) we have again a dynamical model on an euclidean space (in our case  $\mathbb{R}^9$ ) driven by white noise. The Essential Estimator is the least variance filter built for the above model, and corresponds to a linear Kalman filter update in the embedding space, followed by a projection onto the essential manifold. Note that in principle the gain could be precomputed offline, for each possible configuration of motion and feature positions.

**Prediction step:**

$$\hat{\mathbf{Q}}(t+1|t) = \hat{\mathbf{Q}}(t|t); \hat{\mathbf{Q}}(0|0) = \mathbf{Q}_0 \quad (44)$$

$$P(t+1|t) = P(t|t) + R_{\mathbf{Q}}; P(0|0) = P_0 \quad (45)$$

**Update step:**

$$\hat{\mathbf{Q}}(t+1|t+1) = \hat{\mathbf{Q}}(t+1|t) \oplus L(t+1)\chi(t)\hat{\mathbf{Q}}(t+1|t) \quad (46)$$

$$P(t+1|t+1) = \Gamma(t+1)P(t+1|t)\Gamma^T(t+1) + L(t+1)R_{\tilde{n}}(t+1)L^T(t+1) \quad (47)$$

**Gain:**

$$L(t+1) = -P(t+1|t)\chi(t)\Lambda^{-1}(t+1) \quad (48)$$

$$\Lambda(t+1) = \chi(t)P(t+1|t)\chi(t) + R_{\tilde{n}}(t+1) \quad (49)$$

$$\Gamma(t+1) = I - L(t+1)\chi(t) \quad (50)$$

$$R_{\tilde{n}}(t+1) = D(t+1)R_n D^T(t+1) \quad (51)$$

### 5.3 2-D fixed-point estimator

At each time instant a new set of measurements becomes available in the form of position of projected points onto the image plane, encoded in  $\chi(t)$ . The essential constraint imposes

$$\chi(t)\mathbf{Q}(\xi(t)) = 0 \quad \forall t.$$

The Gauss-Newton method generates the iteration

$$\chi(\xi_k) \doteq J_{\chi}(\xi_k)[\xi_{k+1} - \xi_k].$$

At each iteration we solve for  $Y$  the linear problem

$$J_{\chi}(\xi_k)Y = \chi(\xi_k)$$

and then define  $\xi_{k+1} \doteq \xi_k + Y$ .

### 5.4 Identifiability of rigid motion

The theoretical observability/identifiability of the models thus refined is addressed in [52]. It is proved that the model is globally observable once the viewer does not move on a quadric surface which contains all the visible points. Note that such a condition is satisfied almost always due to noise in the measurements.

## 5.5 Further issues

Insofar we have assumed that  $\|T\| \neq 0$ . It may be shown that there is no loss of generality in this assumption [55]. In fact, due to the noise in the measurements, there will be always a translation compatible (in least squares sense) with the observations. The scheme automatically scales translation to unit norm and the inverse depth. The issue is discussed in [55].

The scheme may be further extended to more general camera models in order to estimate camera internal parameters along with visual motion (camera self-calibration). See [54] for details.

The essential filters are used in deriving a scene segmentation method based on 3D motion, which proved successful in extreme experiments such as the segmentation of the Ullmann scene of two transparent cylinders counter-rotating one inside the other [56].

## 6 Experimental assessment

We have tested the described algorithms on a variety of motion and structure configurations. We report the simulations performed on the same data sets of [57]. These consist of views of a cloud of points under a discontinuous motion with singular regions (zero-translation and non-zero rotation) and are described in [57]. Gaussian noise with 1 pixel Std has been added to the measurements. Simulations have been performed with a variable number of points down to 1 point for constant velocity motion, and show consistent performance.

### The local coordinates estimator

In fig. 4,5 we show the three components of translational and rotational velocity as estimated by the local coordinate estimator. Convergence is reached in less than 20 steps. Tuning has been performed, as with the other schemes, within an order of magnitude, and the Std of the estimation error are reported in the tables below. It must be pointed out that we have observed a better behavior by increasing the variance of the pseudoinnovation. This is due to the fact that the EKF relies on the hypothesis that the measurement noise is white and the linearization error is negligible, while in this case it is not. Initialization is performed with one step of the traditional Longuet-Higgins algorithm. The computational cost of one iteration is of about 100 Kflops for 20 points.

Note that if we have available some dynamical model for motion we may easily insert it into the state model.

### The Essential estimator

In fig. 8 we show the 9 components of the essential matrix as estimated by the essential estimator. convergence is 4 times slower than the local coordinate version, but each step is 10 times faster. Note that in principle the gains may be precomputed offline, for each possible configuration of points in the image plane. We have noted step-like convergence with plateaus followed by switching regions. These correspond to switching of the first two eigenspaces of the SVD of  $\mathbf{Q}$ . When brought to local coordinates we have estimates for rotation and translation 6,7. It is noted that the homeomorphism  $\Phi$  may have singularities due to noise when the last eigenspace is changed with one of the other two. This causes the spikes observed in the estimates of motion. However, note that there is no transient to recover, since the errors do not occur in the estimation step, but in transferring to local coordinates. The switching can be avoided by a higher level control on the continuity of the singular values. The computational cost amounts to circa 41 Kflops per each step for 20 points.

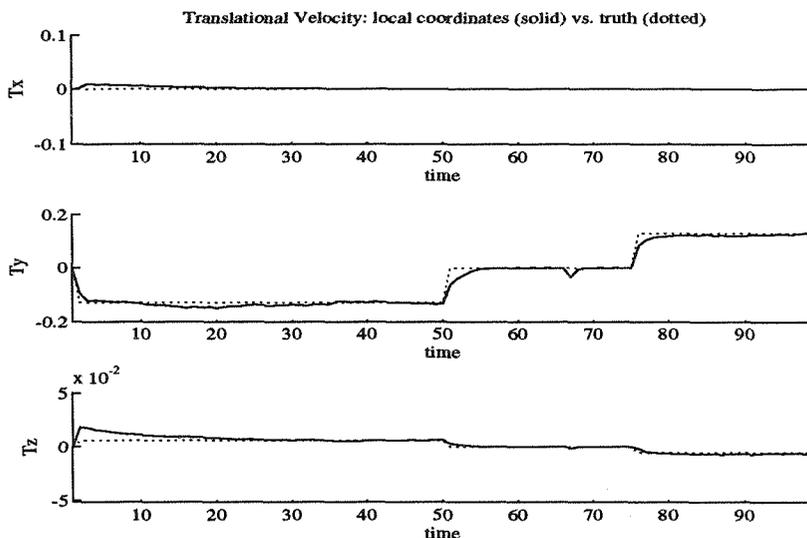


Figure 4: Components of translational velocity as estimated by the local coordinates estimator. The ground truth is shown in dotted lines.

### The 2-D iteration

The performance of the 2-D iteration is reported in fig. 9,10. This scheme proved very accurate after proper initialization, even though the error analysis used for calculating the variance of the estimates at each fixed time was approximate. Speed may be adjusted by varying the number of iterations at each fixed time. We have noticed that this converges after a number of steps between 3 and 7. The cost of the scheme for 7 iterations and 20 points is 100 Kflops. The simulations reported were done using a constant variance of the error of the k-iteration, and hence show larger errors than the other schemes.

We now summarize the performance of the three schemes: mean and Std are computed between time 30 and 50 for the local coordinate scheme and the 2-D iteration, while between time 150 and 200 for the essential estimator.

Scheme	$T_x$	$T_y$	$T_z$
Local	M: .0002 Std:.0004	M:-.0015 Std: .0048	M: .0002 Std: .0004
Essential	M: 3.9754E-5 Std: .0001	M: .0017 Std: .0013	M: .0002 Std: .0001
2-D	M: .376E-3 Std: .0009	M: -.0835E-3 Std: .0071	M: .2851E-3 Std: .0009

Scheme	$\Omega_x$	$\Omega_y$	$\Omega_z$
Local	M:.0008 Std: .0022	M:.0002 Std:.0002	M:-.0002 Std:.0008
Essential	M:-.0008 Std: .0004	M: 3.9949E-6 Std: .0002	M: -1.6107E-5 Std: .0004
2-D	M: .2156E-3 Std: .0034	M: .2261E-3 Std: .0006	M:.0073E-3 Std:.0006

### Experiments on real image sequences

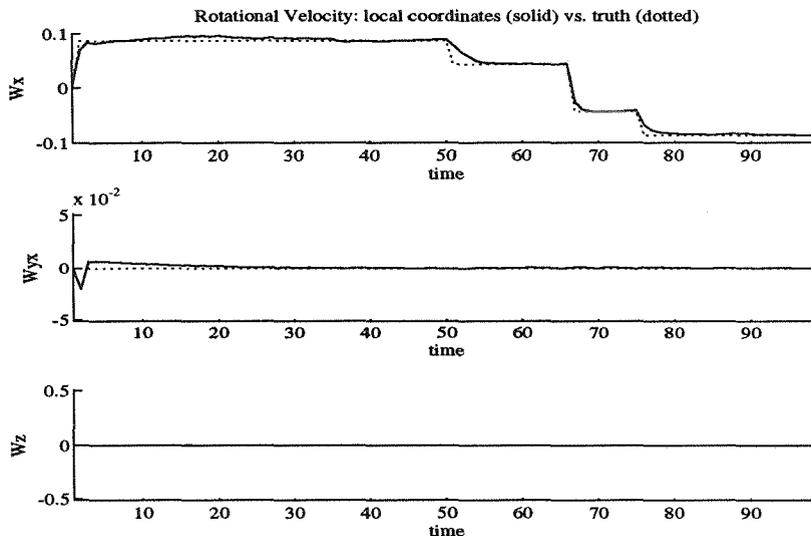


Figure 5: Components of rotational velocity as estimated by the local coordinates estimator.

We have tested our schemes on a sequence of 10 images of the rocket scene (see fig 11). There are 22 feature points visible, and the standard deviation of the location error on the image plane is about one pixel. The local coordinates estimator has a transient of about 20 steps to converge from arbitrary initial condition. Hence we have run the local estimator on the 10 images starting from zero initial condition, and we have used the final estimate as initial condition for a new run, whose results we report in figures 12-14. We did not perform any ad hoc tuning, and the setting was the same used in the simulations described at the previous paragraphs. In fig. 12 we report the 6 motion components as estimated from the local coordinates estimator and the corresponding ground truth (in dotted lines); the estimation error is plotted in figure 13. As it can be seen the estimates are within 5% error, and the final estimate is less than 1% off the true motion. Finally in fig.14 we report the norm of the pseudo-innovation of the filter, which converges to a value of about  $10^{-3}$  in less than 10 + 5 steps.

In this experiment we have used the true norm of translation as the scale factor. We have also run simulations in which the scale factor was calculated by updating the estimate of the distance between the two closest features, as in the experiments described in the previous paragraphs. In this case, however, convergence is slower, as the innovation norm reaches regime in about 20-25 steps.

## 7 Conclusions

We have proposed a novel formulation of the visual motion problem in a system theoretic framework as the identification of an exterior differential system with parameters on a topological manifold. We have first addressed the general identification problem using nonlinear prediction error criteria, and then applied the results to the visual motion problem. We have shown that other tasks in computer vision may be formulated as the identification of exterior an differential system with parameters on a manifold and solved in a principled manner.

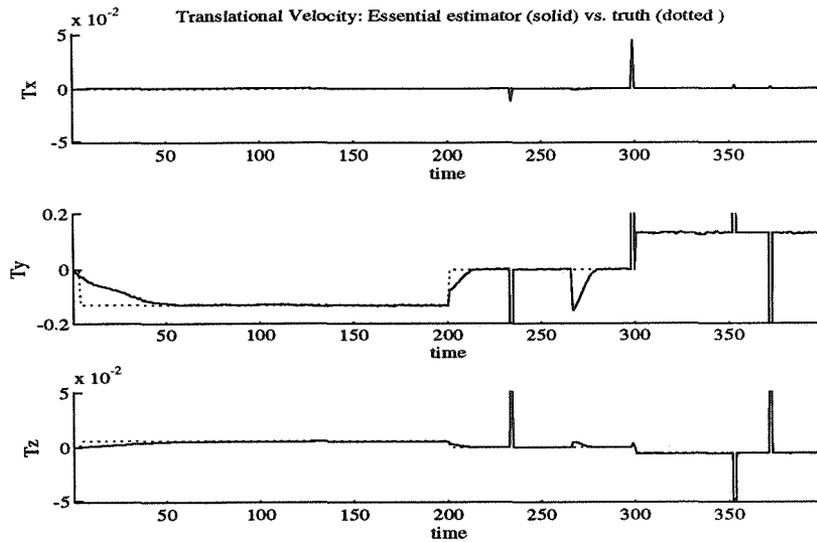


Figure 6: Components of translational velocity as estimated by the essential estimator. Note the spikes due to the local coordinates transformation. Note also that they do not affect convergence since they do not occur in the estimation process, but while transferring to local coordinates.

The proposed schemes prove very accurate and robust, as well as computationally light, as we show in the experimental section. Easy extensions of the schemes allow solving the camera self-calibration problem and the 3D motion-based segmentation.

## Acknowledgements

We wish to thank Prof. J.K. Åström for his discussions on implicit Kalman filtering, Prof. Richard Murray and Prof. Shankar Sastry for their observations and useful suggestions. Also discussions with Michiel van Nieuwstadt and Andrea Mennucci were helpful, as well as the suggestions of Prof. John Doyle and Prof. Manfred Morari.

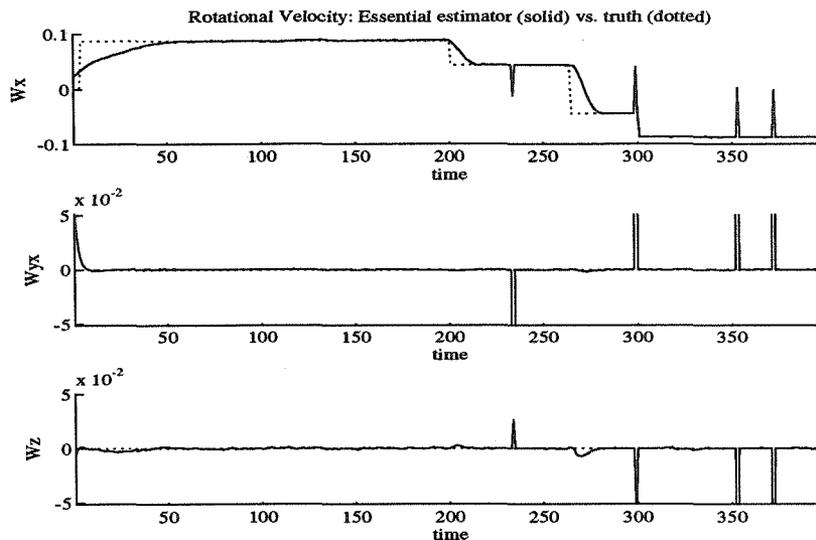


Figure 7: Components of rotational velocity as estimated by the local coordinates estimator. The ground truth is shown in dotted lines. Note the spikes due to the local coordinates transformation. Note also that there is no transient to recover since they do not occur in the estimation process.

## References

- [1] E. Arbogast and R. Mohr. An egomotion algorithm based on the tracking of arbitrary curves. *Proc. of the 2nd Europ. Conf. on Computer Vision*, 1992.
- [2] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. RPL-TR 9107, Queen's University Kingston, Ontario, Robotics and perception laboratory, 1992. Also in *Proc. CVPR 1992*, pp 236-242.
- [3] A. Azarbayejani, B. Horowitz, and A. Pentland. Recursive estimation of structure and motion using relative orientation constraints. *Proc. CVPR*, New York, 1993.
- [4] A. Blake, M. Taylor, and A. Cox. Grasping visual symmetry. *Proc. of the ICCV*, 1993.
- [5] W. Boothby. *Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1986.
- [6] T. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. *IEEE trans. PAMI*, Jan. 1986.
- [7] Bryant, Chern, Goldberg, and Goldsmith. *Exterior Differential Systems*. Mathematical Research Institute. Springer Verlag, 1992.
- [8] R.S. Bucy. Non-linear filtering theory. *IEEE Trans. A.C. AC-10*, 198, 1965.
- [9] R. Cipolla and A. Blake. Surface orientation and time to crash from image divergence and deformation. *Proc. of the European Conf. on Comp. Vision*, 1992.
- [10] H. Cox. On the estimation of state variables and parameters from noisy dynamical systems. *IEEE Trans. A.C. AC-9*, 5-12, 1964.
- [11] R. Curwen, A. Blake, and A. Zisserman. Real-time visual tracking for surveillance and path planning. *Proc. of the ECCV*, 1992.

- [12] Darmon. A recursive method to apply the hough transform to a set of moving objects. *Proc. IEEE, CH 1746 7/82*, 1982.
- [13] E. Di-Bernardo, L. Toniutti, R. Frezza, and G. Picci. Stima del moto dell'osservatore e della struttura della scena mediante visione monoculare. *Tesi di Laurea-Università di Padova*, 1993.
- [14] E. D. Dickmanns and Th. Christians. Relative 3d-state estimation for autonomous visual guidance of road vehicles. In *Intelligent autonomous system 2 (IAS-2)*, Amsterdam, 11-14 December 1989.
- [15] E. D. Dickmanns and V. Graefe. Applications of dynamic monocular machine vision. *Machine Vision and Applications*, 1:241-261, 1988.
- [16] E. D. Dickmanns and V. Graefe. Dynamic monocular machine vision. *Machine Vision and Applications*, 1:223-240, 1988.
- [17] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig?, In G. Sandini Ed. *Proc. ECCV92*, Lecture Notes in Computer Science, **588**, G. Sandini Ed. pp.563-578, Springer-Verlag, 1992.
- [18] O. Faugeras. *Three dimensional vision, a geometric viewpoint*. MIT Press, 1993.
- [19] O. D. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. *Int. J. of Computer Vision*, 1990.
- [20] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self-calibration: theory and experiments. *Proc. of the ECCV92, Vol. 588 of LNCS, Springer Verlag*, 1992.
- [21] D.B. Gennery. Tracking known 3-dimensional object. In *Proc. AAAI 2nd Natl. Conf. Artif. Intell.*, pages 13-17, Pittsburg, PA, 1982.
- [22] D.B. Gennery. Visual tracking of known 3-dimensional object. *Int. J. of Computer Vision*, 7(3):243-270, 1992.
- [23] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. 2nd Europ. Conf. Comput. Vision, G. Sandini (Ed.), LNCS-Series Vol. 588, Springer-Verlag*, 1992.
- [24] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura. Image-based dynamic visual servo for a hand-eye manipulator. In Kodama Kimura, editor, *Recent advances in mathematical theory of systems, control, networks, and signal processing II*, pages 609-614. Proceedings of the international symposium of MTNS, Mita Press, 1991.
- [25] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura. Manipulator control with image-based visual servo. In *IEEE intl' Conference on Robotics and Automation*, pages 2267-2272, 1991.
- [26] D. Heeger and A. Jepson. Subspace methods for recovering rigid motion i: algorithm and implementation. *Int. J. Comp. Vision vol. 7 (2)*, 1992.
- [27] J. Heel. Direct estimation of structure and motion from multiple frames. *AI Memo 1190, MIT AI Lab*, March 1990.
- [28] C.C. Ho and N.H. McClamrock. Autonomous spacecraft docking using a computer vision system. *Proc. of the 31st CDC - Tucson, AZ*, 1992.
- [29] B.K.P. Horn. Relative orientation. *Int. J. of Computer Vision*, 4:59-78, 1990.
- [30] A. Isidori. *Nonlinear Control Systems*. Springer Verlag, 1989.
- [31] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [32] J.K. Åström and P. Eykhoff. System identification, a survey. *Automatica, vol. 7*, 1971.
- [33] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. of the ASME-Journal of basic engineering.*, 35-45, 1960.

- [34] A. J. Krener and A. Isidori. Linearization by output injection and nonlinear observers. *Systems and Control Letters* vol. 3, 1983.
- [35] A. J. Krener and W. Respondek. Nonlinear observers with linearizable error dynamics. *SIAM J. Control and Optimization*, 1985.
- [36] W. Lee and K. Nam. Observer design for autonomous discrete-time nonlinear systems. *Systems and Control Letters* vol. 17, 1991.
- [37] Ming Lei and Bijoy K. Ghosh. A new nonlinear feedback controller for visually-guided robotic motion tracking. *To appear in Proc. of the ECC '93*, Oct. 1992.
- [38] Y. Liu and O.D. Faugeras T.S. Huang. Determination of camera location from 2d to 3d line and point correspondences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(1):28–37, 1990.
- [39] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
- [40] L. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. of computer vision*, 1989.
- [41] S. Maybank. *Theory of reconstruction from image motion*. Springer Verlag, 1992.
- [42] R. Mohr. Projective reconstruction. in *Geometric Invariance in Computer Vision*, 1992.
- [43] M.Spivak. *A comprehensive introduction to differential geometry– Voll.I-V*. Publish or perish, 1970-75.
- [44] H. Nijmeijer and A.J.Van Der Shaft. *Nonlinear Dynamical Control Systems*. Springer Verlag, 1990.
- [45] J. Oliensis and J. Inigo-Thomas. Recursive multi-frame structure from motion incorporating motion error. *Proc. DARPA Image Understanding Workshop*, 1992.
- [46] P. Perona and S. Soatto. Motion and structure from 2 perspective views of  $p$  points – algorithm and error analysis. *Technical Report CIT/CNS 23-93 – California Institute of Technology*, Oct. 1992.
- [47] P.Eykhoff and P.C.Parks (eds). Special issue on identification and system parameters estimati on. *Automatica*, vol. 26, 1990.
- [48] R.E.Kopp and R.J. Orford. Linear regression applied to system identification for adaptive control systems. *AIAA Journal* , 2300-2306, 1963.
- [49] J.G. Semple and G.J. Kneebone. *Algebraic Projective Geometry*. Oxford, 1952.
- [50] A. J. Van Der Shaft. Observability and controllability for smooth nonlinear systems. *SIAM J. Control and Optim.* vol. 20 (3), 1982.
- [51] A. Shashua. Projective structure reconstruction. *AI Memo MIT AI Lab*, March 1993.
- [52] S. Soatto , P. Perona. Observability/Identifiability of rigid motion under perspective projection. *Technical Report CIT-CDS 94-001, California Institute of Technology*, 1994. Submitted to the 33rd CDC
- [53] S. Soatto. Subspace methods for recovering ridig motion realized by identifying exterior differential systems. *Technical Report CIT-CDS 94-005, California Institute of Technology*, 1994.
- [54] S. Soatto, R. Frezza, and P. Perona. Recursive estimation of camera motion from uncalibrated image sequences. *Technical Report CIT-CDS 94-003, California Institute of Technology*, 1994.
- [55] S. Soatto, R. Frezza, and P. Perona. Motion estimation on the essential manifold. *Proc. of the ECCV 94 – To appear in “Lecture Notes in Computer Sciences”, Springer Verlag*, May 1994.
- [56] S. Soatto and P. Perona. Three dimensional transparent structure segmentation and multiple 3d motion estimation from monocular perspective image sequences. *Technical Report CIT-CDS 93-022, California Institute of Technology*, 1993.

- [57] S. Soatto, P. Perona, R. Frezza, and G. Picci. Recursive motion and structure estimation with complete error characterization. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, pages 428–433, New York, June 1993.
- [58] T. Soderstrom and P. Stoica. *System Identification*. Prentice Hall, 1989.
- [59] M. Swain and M. Stricker (editors). Promising directions in active vision. Technical Report T.R. CS 91-27, University of Chicago, November 1991. Written by the attendees of the NSF Active Vision Workshop – August 5-7 1991.
- [60] R. Szeliski. *Technical Report*, 1993.
- [61] G. Taubin. Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1991.
- [62] C. J. Taylor and D. J. Kriegman. Structure and motion from line segments in multiple views. *Technical Report 9402, Yale University, YEAR = 1994*.
- [63] C. Tomasi and T. Kanade. Shape and motion from image streams: a factorization method – 3. detection and tracking of point features. CMU-CS 91-132, School of CS – CMU, April 1991.
- [64] C. Tomasi and T. Kanade. Shape and motion from image streams: a factorization method – 2. point features in 3d motion. CMU-CS 91-105, School of CS – CMU, January 1991.
- [65] C. Tomasi and T. Kanade. Shape and motion from image streams: a factorization method – 1. planar motion. CMU-CS 90-166, School of CS – CMU, September 1990.
- [66] J. Weng, T. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(5):451–476, 1989.
- [67] J. Weng, T.S. Huang, and N. Ahuja. Motion and structure from line correspondences: closed-form solution, uniqueness and optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):318–336, 1992.
- [68] Z. Zhang and O. Faugeras. Estimation of displacement from two 3d frames obtained from stereo. *TR 1440 – INRIA*, 1991.
- [69] Z. Zhang and O. Faugeras. *3D dynamic scene analysis*, volume 27 of *Information Sciences*. Springer-Verlag, 1992.

## A Extended Kalman Filtering for Implicit Measurement Constraints

We are interested in building an estimator for a process  $\{\alpha\}$  which is described by a stochastic difference equation

$$\alpha(t+1) = f(\alpha(t)) + v(t); \alpha(t_0) = \alpha_0$$

where  $v(t) \in \mathcal{N}(0, Q_v)$ . Suppose there is a measurable quantity  $x(t)$  which is linked to  $\alpha$  by the constraint

$$h[\alpha(t), x(t)] = 0 \quad \forall t. \quad (52)$$

We will assume throughout  $f, h \in C^r$ ;  $r \geq 1$ . Usually  $x$  is known via some noisy measurement:

$$x(t) = y(t) + w(t) : w(t) \in \mathcal{N}(0, R_w) \quad (53)$$

where the variance/covariance matrix  $R_w$  is derived from knowledge of the measurement device. The model we consider is hence of the form

$$\begin{cases} \alpha(t+1) = f(\alpha(t)) + v(t); \alpha(t_0) = \alpha_0 \\ h[\alpha(t), y(t) + w(t)] = 0 \end{cases} \quad (54)$$

### Construction of the variational model about the reference trajectory

Consider at each time sample  $t$  a reference trajectory  $\bar{\alpha}(t)$  which solves the difference equation

$$\bar{\alpha}(t+1) = f(\bar{\alpha}(t))$$

and the jacobian matrix

$$F(\bar{\alpha}(t)) \doteq F(t) = \left( \frac{\partial f}{\partial \alpha} \right)_{|\alpha(t)}$$

The linearization of the measurement equation about the point  $(\bar{\alpha}(t), y(t))$  is

$$h[\alpha(t), x(t)] = h[\bar{\alpha}(t), y(t)] + C(\bar{\alpha}, y)(\alpha(t) - \bar{\alpha}(t)) + D(\bar{\alpha}, y)(x(t) - y(t)) + \mathcal{O}(\mathcal{E}^2)$$

where

$$\begin{aligned} C(\bar{\alpha}, y) &\doteq \left( \frac{\partial h}{\partial \alpha} \right)_{|\bar{\alpha}(t), y(t)} \\ D(\bar{\alpha}, y) &\doteq \left( \frac{\partial h}{\partial x} \right)_{|\bar{\alpha}(t), y(t)} \\ \mathcal{E}^2 &\doteq \{ \|\alpha - \bar{\alpha}\|^2, \|x - y\|^2 \}. \end{aligned}$$

Exploiting the fact that  $h[\alpha, x] = 0$ , calling  $\delta\alpha(t) \doteq \alpha(t) - \bar{\alpha}(t)$  and neglecting the arguments in  $C$  and  $D$ , we have, up to second order terms

$$h[\bar{\alpha}(t), y(t)] = -C\delta\alpha(t) - Dw(t).$$

## Prediction Step

Suppose at some time  $t$  we have available the best estimate  $\hat{\alpha}(t|t)$ ; we may write the variational model about the trajectory  $\bar{\alpha}(t)$  defined such that

$$\bar{\alpha}(t+1) = f(\bar{\alpha}(t)) ; \bar{\alpha}(t) = \hat{\alpha}(t|t).$$

For small displacements we may write

$$\delta\alpha(t+1) = F(\bar{\alpha}(t))\delta\alpha(t) + \tilde{v}(t) \quad (55)$$

where the noise term  $\tilde{v}(t)$  may include a linearization error component.

Note that with such a choice we have  $\delta\hat{\alpha}(t|t) = 0$  and  $\delta\hat{\alpha}(t+1|t) = F(\bar{\alpha}(t))\delta\hat{\alpha}(t|t) = 0$ , from which we can conclude

$$\hat{\alpha}(t+1|t) = \bar{\alpha}(t+1) = f(\bar{\alpha}(t)) = f(\hat{\alpha}(t|t)). \quad (56)$$

The variance of the prediction error  $\delta\hat{\alpha}(t+1|t)$  is

$$P(t+1|t) = F(t)P(t|t)F^T(t) + \tilde{Q} \quad (57)$$

where  $\tilde{Q} = \text{var}(\tilde{v})$ . The last two equations represent the prediction step for the estimator and are equal, as expected, to the prediction of the explicit EKF [33, 31, 8].

## Update Step

At time  $t+1$  a new measurement becomes available  $y(t+1)$ , together with the prediction  $\hat{\alpha}(t+1|t)$  and its error variance  $P(t+1|t)$ . Exploiting the linearization of the measurement equation about  $\bar{\alpha}(t+1) = \hat{\alpha}(t+1|t)$ , we obtain, letting  $\hat{\alpha} \doteq \hat{\alpha}(t+1|t)$  and  $y \doteq y(t+1)$ ,

$$h[\hat{\alpha}, y] = -C(\hat{\alpha}, y)\delta\alpha(t+1) - n(t+1) \quad (58)$$

where we have defined  $n \doteq D(\hat{\alpha}, y)w(t+1)$ . This, together with the equation (55) defines a linear and *explicit* variational model, for which we can finally write the update equation based on the traditional linear Kalman filter:

$$\delta\hat{\alpha}(t+1|t+1) = \delta\hat{\alpha}(t+1|t) + L(t+1)[h[\hat{\alpha}, y] + C\delta\hat{\alpha}(t+1|t)] \quad (59)$$

where

$$\begin{aligned} L(t+1) &= -P(t+1|t)C^T\Lambda^{-1}(t+1) \\ \Lambda(t+1) &= CP(t|t)C^T + R_n(t+1) \\ P(t+1|t+1) &= \Gamma(t+1)P(t+1|t)\Gamma^T(t+1) + LR_n(t+1)L^T \\ \Gamma &= (I - LC). \end{aligned}$$

Since  $\delta\hat{\alpha}(t+1|t) = 0$  and  $\delta\hat{\alpha}(t+1|t+1) = \hat{\alpha}(t+1|t+1) - \hat{\alpha}(t+1|t)$ , we may write the update equation for the original model:

$$\hat{\alpha}(t+1|t+1) = \hat{\alpha}(t+1|t) + L(t+1)h[\hat{\alpha}(t+1|t), y(t+1)]. \quad (60)$$

In this formulation the quantity  $h[\hat{\alpha}(t+1|t), y(t+1)]$  plays the role of the pseudo-innovation. The noise  $n$  defined in (58) has a variance which is calculated from its definition:

$$R_n(t) = D(\hat{\alpha}, y)R_w(t)D^T(\hat{\alpha}, y).$$

The implicit Kalman filter was used by other researchers such as Darmon [12], Faugeras [38, 69] and Heel [27].

## B Projection onto the essential space

We have defined the projection operator onto the essential manifold without proving that the result is in fact an element of the essential manifold. In fact the following theorem, which was first stated by Faugeras in 1990 [23, 41], shows that a characterizing property of the essential manifold is that its elements have two non-zero equal singular values and a zero singular value.

### Theorem B.1 .

Let  $Q = U\Sigma V^T$  be the SVD of an element of  $GL(3)$ . Then

$$Q \in E \Leftrightarrow \Sigma = \Sigma_0 = \text{diag}\{\lambda \ \lambda \ 0\} \mid \lambda \in \mathbb{R}^+.$$

#### Proof:

( $\Rightarrow$ ) let  $Q = RS \mid R \in SO(3), S \in so(3)$ ;  $\sigma(Q)$ , the set of singular values of  $Q$ , is such that  $\sigma(Q) = \sqrt{\sigma(QQ^T)}$ . Next observe that  $QQ^T = RSS^TR^T = SS^T = -S^2$ . Also  $\forall S \in so(3) \exists ! T \mid S = (T^\wedge)$ , and the singular values of  $S^2$  are  $\{0, \|T\|^2, \|T\|^2\}$ . Hence if  $Q \in E$ , it has two equal singular values and a zero singular value.

( $\Leftarrow$ ) let  $Q = U\Sigma_0 V^T$  for some orthonormal  $U, V$  and for some  $\lambda$ . Let furthermore  $R_Z(\frac{\pi}{2})$  be a rotation of  $\frac{\pi}{2}$  about the  $Z$  axis, then

$$Q = U\Sigma_0 V^T = UR_Z(\frac{\pi}{2})^T V^T V R_Z(\frac{\pi}{2}) \Sigma_0 V^T.$$

Now call  $R \doteq UR_Z(\frac{\pi}{2})^T V^T$  and  $S \doteq V R_Z(\frac{\pi}{2}) \Sigma_0 V^T$ ; it is immediate to see that  $RR^T = R^T R = I_3$  and  $S^T = -S$ , hence the claim. **Q.E.D.**

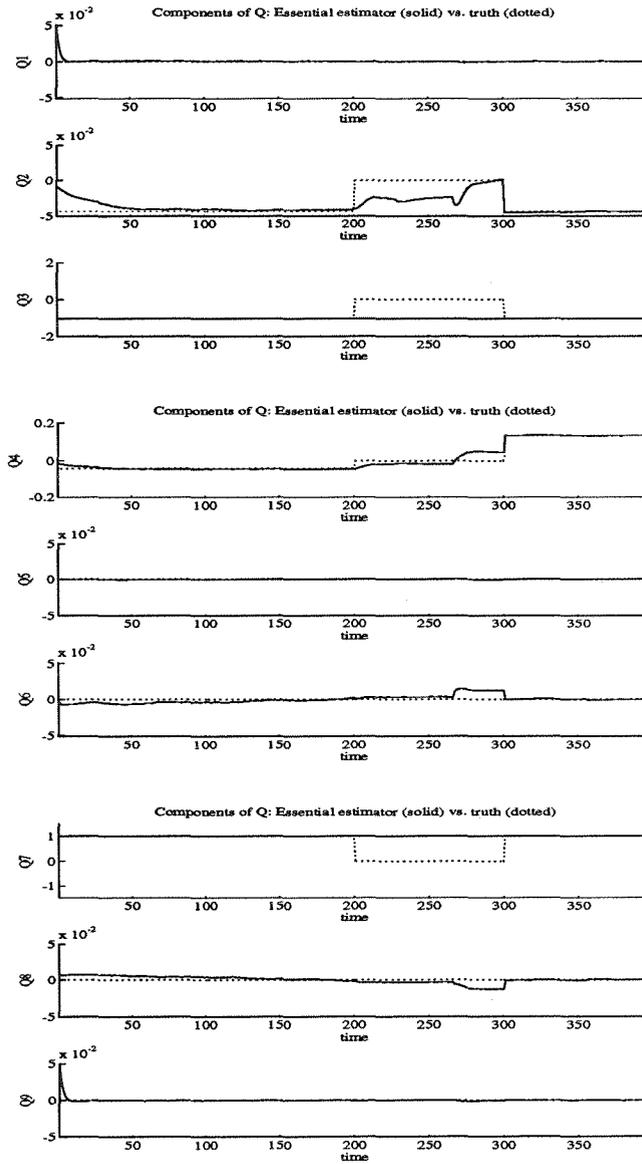


Figure 8: Components of the essential matrix as estimated by the essential estimator. Note that there are no spikes and the estimate is smooth. Note that the estimates between time 200 and 300 are not significant, as the ground truth (dotted line) is scaled to zero.

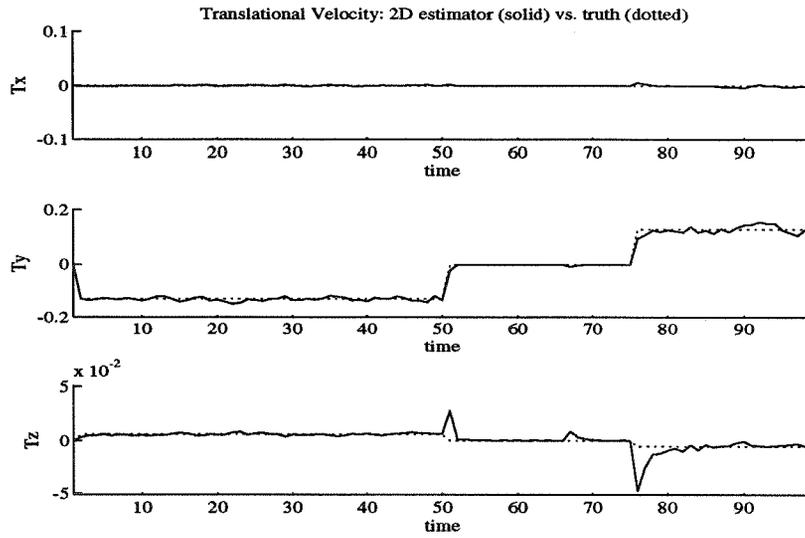


Figure 9: Components of translational velocity as estimated by the double iteration estimator.

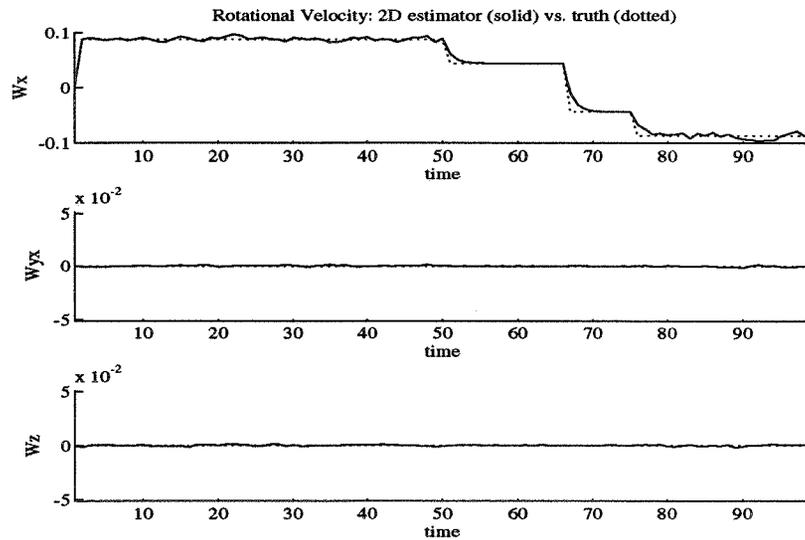


Figure 10: Components of rotational velocity as estimated by the double iteration estimator.

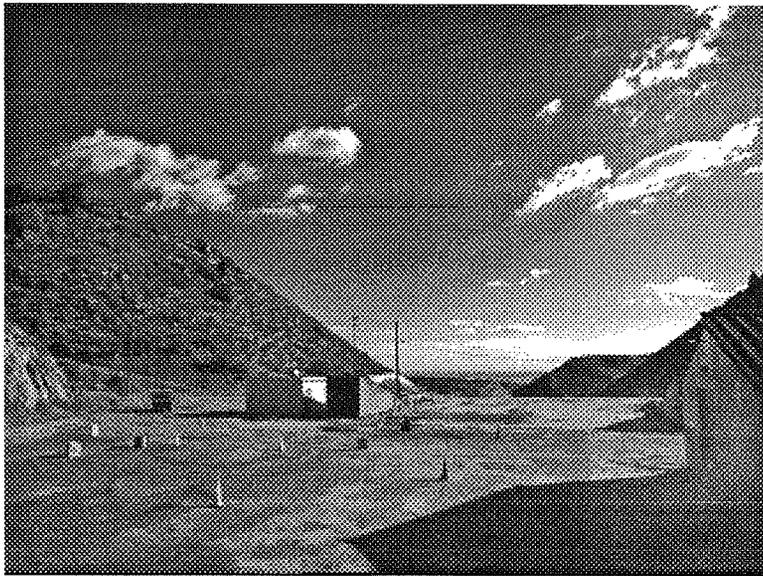


Figure 11: *One image of the rocket scene.*

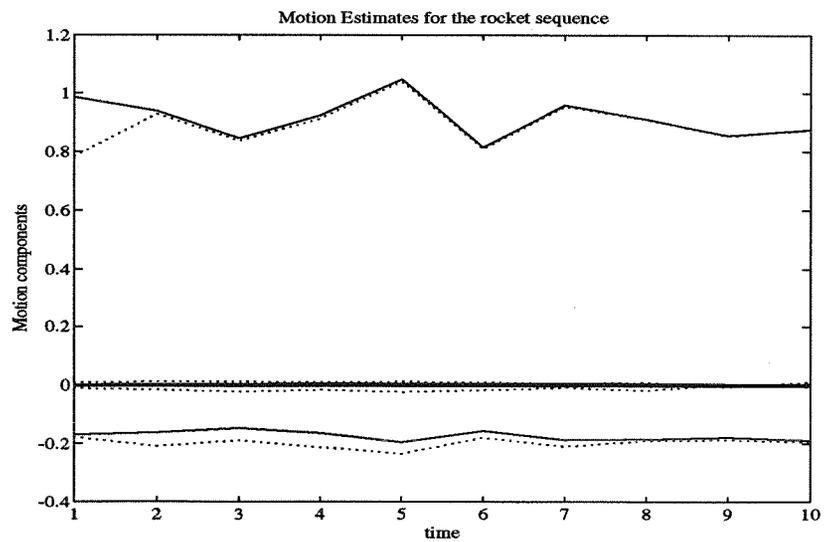


Figure 12: Motion estimates for the rocket sequence: The six components of motion as estimated by the local coordinates estimator are showed in solid lines. The corresponding ground truth is in dotted lines.

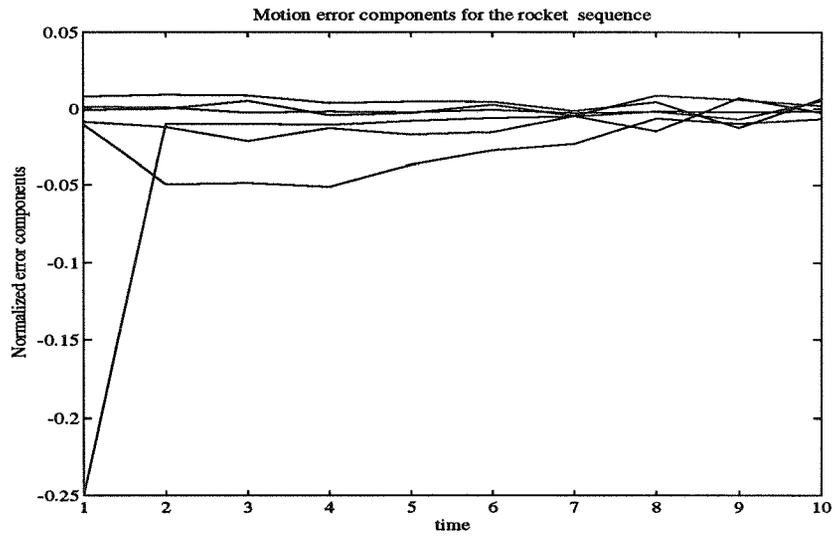


Figure 13: Error in the motion estimates for the rocket sequence. All components are within 5% of the true motion.

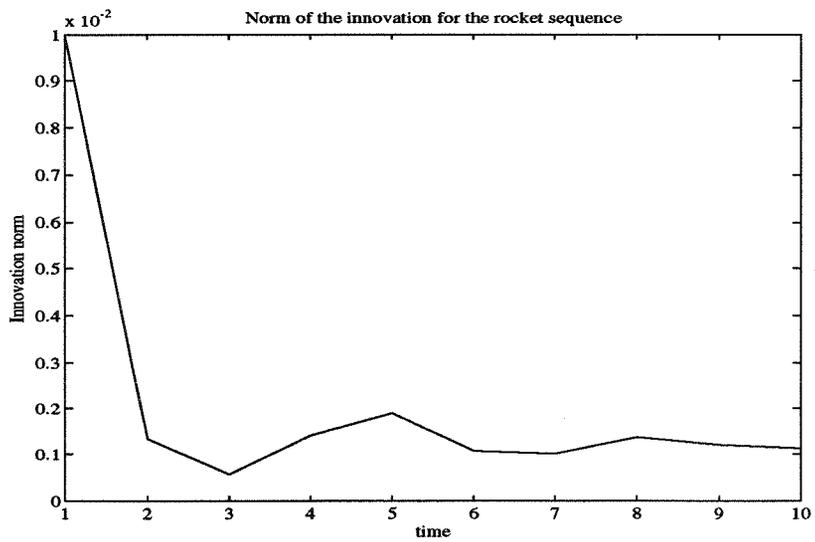


Figure 14: Norm of the pseudo-innovation process of the local estimator for the rocket scene. Convergence is reached in less than 5 steps.