On the Exact Linearization of Structure From Motion*

Stefano Soatto and **Pietro Perona**
California Institute of Technology 116-81
Pasadena – CA 91125
soatto@systems.caltech.edu

keywords: Structure From Motion, nonlinear estimation, observer linearization

Abstract

The estimation of structure from motion has been a central task of computational vision over the last decade. As it is very well known, the problem is *nonlinear* due to the perspective nature of the measurements.

One may ask at this point: *does there exist a clever choice of coordinates which simplifies the estimation task*? In particular, since "linearity" is a coordinate-dependent notion, *is there a choice of coordinates such that the problem of estimating structure from motion becomes linear*? In this paper we prove that the answer to the above question is **no**.

An immediate consequence is that all choices of coordinates representations are *structurally equivalent*, in the sense that, at the current state of understanding of nonlinear estimation, none of them has an advantage based on geometric properties; instead, the difference between them is based purely on *computational* (numerical) ground.

A further consequence of our result is the legitimation of the use of local linearization-based techniques (such as the Extended Kalman Filter) for estimating structure from known motion.

1 Introduction

Estimating "Structure From Motion" (SFM) consists of reconstructing the structure of a moving object from its projection onto a camera. A number of schemes have been proposed for estimating motion from known structure, structure for known motion and both structure and motion recursively from an image sequence (see [3, 21] for a review of the existing methods).

In this paper we restrict our attention to the *recursive* estimation of *point-based* structure for *known motion*. It has been known for a while [12] that SFM can be formulated as the estimation of the state of a nonlinear dynamical system. Such estimation task has been traditionally addressed using Extended Kalman Filters (EKF) [2, 6, 8], as for example in [12, 14, 16, 19].

The EKF is a general purpose technique for estimating the state of a nonlinear dynamical system and is based upon a linear update of the original nonlinear model with a gain computed on the *local-linearization* of the model about the best current estimate of the trajectory. The estimation error can be described as the state of a nonlinear dynamical system as well. In the case of a linear system, the Kalman Filter has the property of minimizing the variance of the state estimation

*Research funded by the California Institute of Technology, a scholarship from the University of Padova and a fellowship from the "A. Gini" Foundation. This work is registered as Technical Report **CIT-CDS** 94-011, California Institute of Technology, May 1994

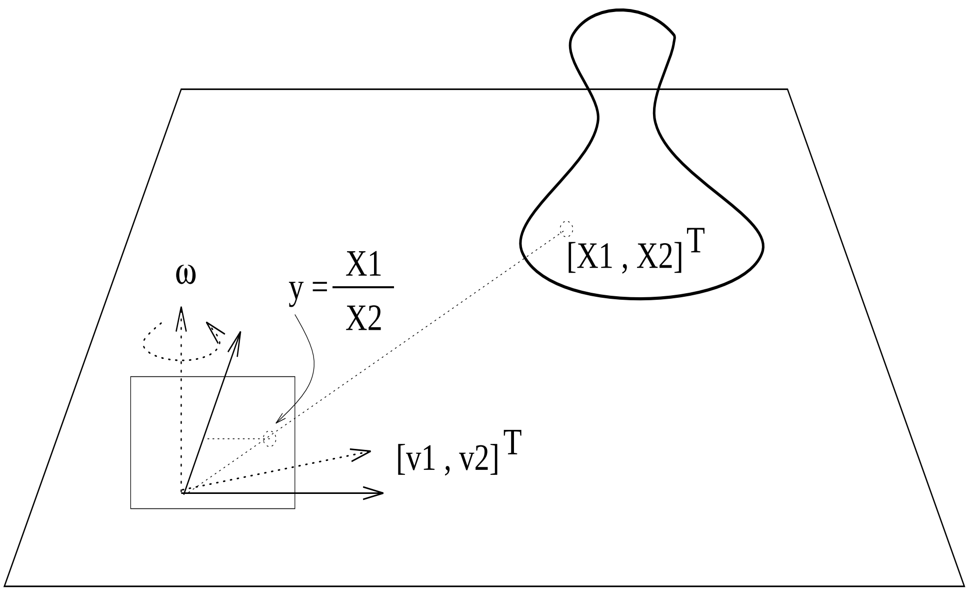

Figure 1: *Planar Structure From Motion*

error, and a number of results is available on the asymptotic behavior of the filter, its convergence properties, the error dynamics etc. .

We consider in this paper the most general class of dynamic state estimators, also called *"observers"*, of which the Kalman Filter is an instance. In particular, since *"linearity" is a coordinates-dependent notion*, we want to see if *there exists a change of coordinates such that the estimation error of the observer has a linear dynamic.* In such case we may be able to *assign* its modes and achieve arbitrarily fast error decays. This problem has been known for a decade in the Control community as the "observer linearization problem"(see [5] for a review).

It is conceivable that the success of an observer as a state estimator depends on the structure of the system to be observed. In particular, since the observer tries to reconstruct the state of a system by measuring its output, if two states produce *the same output*, the observer will not be able to *distinguish* those states apart. The condition under which there are no *indistinguishable* states is called *observability* of the model [5, 7], and will be discussed later.

1.1 Structure from motion using observers

Let us first simplify the problem by assuming that the motion of the object is *rigid, constrained on a plane*, and has *constant velocity.* It can be shown [17] that the planar motion case is structurally equivalent to the full 3D motion, as far as observability is concerned.

The "structure" of the scene is represented by a number of point-features whose coordinates in the ambient plane are $\mathbf{x} \doteq [\mathbf{x}_1 \ \mathbf{x}_2]^T$; $v = [v_1 \ v_2]^T$ indicates the relative translational velocity between the object and the viewer and ω is the rotational velocity about an axis orthogonal to the plane and to the optical axis (see figure 1). If we measure the "horizontal" coordinate of the *projection* of the point onto an image plane, $y \doteq \mathbf{x}_1/\mathbf{x}_2$, then we can write a nonlinear dynamical model having the position of the point in the ambient plane as the state, and the projection as output/measurement equation:

$$
\begin{cases} \frac{d}{dt}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ y = \frac{\mathbf{x}_1}{\mathbf{x}_2} \qquad \mathbf{x}_2 \neq 0 \end{cases} \tag{1}
$$

which is in the form

$$
\begin{cases} \frac{d}{dt}\mathbf{x} = f(\mathbf{x}) \qquad \mathbf{x}(t_0) = \mathbf{x_0} \in \mathbb{R}^n \\ y = h(\mathbf{x}) \end{cases} \tag{2}
$$

where

$$
\begin{cases} n = 2 \\ f(\mathbf{x}) \doteq F\mathbf{x} + G \\ F \doteq \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix} \\ G \doteq \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ h(\mathbf{x}) \doteq \frac{\mathbf{x}_1}{\mathbf{x}_2} \qquad \mathbf{x}_2 \neq 0. \end{cases} \tag{3}
$$

We call the above model the *standard model for SFM.* One may argue that the choice of the refecence frame (the viewer-reference in this case) and of the model of projection (an ideal pinhole camera with unit focal length) are arbitrary. We fully agree. There are other possible reference frames (object-centered, world-centered etc.) and models for the perspective projection (with the center of projection displaced in the ambient plane). More than that, there are other possible *nonlinear* changes of coordinates (not simply changes of the reference frame) that one may consider.

In this paper we are interested in studying whether any of these changes of coordinates *simplifies* the structure of the estimation problem.

2 The linear observer

Let us pretend for the moment that the model of SFM is linear:

$$
\begin{cases} \frac{d}{dt}\mathbf{x} = A\mathbf{x} \qquad \mathbf{x}(0) = \mathbf{x_0} \in \mathbb{R}^n \\ y = C\mathbf{x} \end{cases} \tag{4}
$$

for some matrices A, C of the appropriate dimensions. Then we may apply standard results from linear systems theory [7] and write another linear dynamical system with state $\hat{\mathbf{x}}$ starting from an <u>*arbitrary* initial condition $\hat{\mathbf{x}}_0$</u> and satisfying

$$
\frac{d}{dt}\hat{\mathbf{x}} = (A + LC)\hat{\mathbf{x}} - Ly \tag{5}
$$

for some "gain" matrix L. Then the *estimation error*, defined as $\mathbf{e} \doteq \mathbf{x} - \hat{\mathbf{x}}$, satisfies the linear differential equation

$$
\frac{d}{dt}\mathbf{e} = (A + LC)\mathbf{e}. \tag{6}
$$

Suppose now that the pair of constant matrices (C, A) is such that, for each choice of n (pairwise conjugate) complex numbers, we can find a gain matrix L such that $(A + LC)$ has exactly these numbers as eigenvalues. In such case, the pair (C, A) is said to be *completely (linearly) observable* $(C - O)$:

$$
C - O \;\Leftrightarrow\; \forall\, \{\lambda_1, \ldots, \lambda_n\}\; \exists\, L \mid \sigma\{A + LC\} = \{\lambda_1, \ldots, \lambda_n\}
$$

where σ denotes the spectrum (set of eigenvalues). For a detailed treatment of these concepts in the linear case, see for instance [7, 20].

Under the conditions above, it is possible to assign arbitrarily the spectrum of the estimation error, in particular it is possible for $\hat{\mathbf{x}}$ to converge to $\mathbf{x}$ arbitrarily fast regardless the initial condition $\hat{\mathbf{x}}_0$. The idea behind the Kalman Filter is to compute L so that the estimation error has least two-norm.

Of course the model (2) is *nonlinear*, and the observer is defined as a nonlinear dynamical system of the form

$$\frac{d}{dt}\hat{\mathbf{x}} = g(\hat{\mathbf{x}}, y) \qquad \hat{\mathbf{x}}(0) = \hat{\mathbf{x}}_0$$

which has the measurements y as inputs and produces the estimates of the state of the original model. The error $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ also satisfies a set of *nonlinear* differential equations. In such a case, unlike in the linear context, it is not easy in general to design observers such that the estimation error has prescribed dynamical properties.
However, suppose that there exists a coordinates transformation of $\mathbb{R}^n$

$$\begin{aligned} \Phi : \mathbb{R}^n &\rightarrow \mathbb{R}^n && (7) \\ \mathbf{x} &\mapsto \mathbf{z} = \Phi(\mathbf{x}) && (8) \end{aligned}$$

such that the original model is transformed into

$$\begin{cases} \frac{d}{dt}\mathbf{z} \doteq \left[\frac{\partial\Phi}{\partial\mathbf{x}} f(\mathbf{x})\right]_{\mathbf{x}=\Phi^{-1}(\mathbf{z})} = A\mathbf{z} + k(C\mathbf{z}) \\ y = h(\Phi^{-1}(\mathbf{z})) = C\mathbf{z} \end{cases} \tag{9}$$

for some A, C and k such that the pair (C, A) is completely observable. Then an observer of the form

$$\frac{d}{dt}\hat{\mathbf{z}} = (A + LC)\hat{\mathbf{z}} - Ly + k(y) \tag{10}$$

yields an estimation error $\mathbf{e} \doteq \mathbf{z} - \hat{\mathbf{z}}$ satisfying the differential equation

$$\frac{d}{dt}\mathbf{e} = (A + LC)\mathbf{e} \tag{11}$$

that is *linear and spectrally assignable* under the assumption that (C, A) is observable. In such case we can resort to the linear case and achieve arbitrarily fast error decays. This technique was proposed and studied in the last ten years [4, 9, 10, 11, 13, 15].

3 The observer linearization problem

As a result of the above discussion, we may give a precise definition of what we mean by the *solution of the "observer linearization problem"*.

> We say that the "observer linearization problem" (OLP) is *solvable* for the model (2) if and only if we can find U_0, $\mathbf{x_0} \in U_0$, $\Phi : U_0 \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$ as above, and $k : h(U_0) \rightarrow \mathbb{R}^2$ such that
>
> $$\begin{aligned} \frac{d}{dt}\mathbf{z} &\doteq \left[\frac{\partial\Phi}{\partial\mathbf{x}} f(\mathbf{x})\right]_{\mathbf{x}=\Phi^{-1}(\mathbf{z})} = A\mathbf{z} + k(C\mathbf{z}) \qquad \forall \mathbf{z} \in \Phi(U_0) && (12) \\ y &= h(\Phi^{-1}(\mathbf{z})) = C\mathbf{z} && (13) \\ & (C, A) \text{ is observable.} && (14) \end{aligned}$$

Theorem 3.1 *(Isidori [5])*
OLP is solvable only if $\dim(\text{span}\{dh\ ,\ dL_f h\}_{|\mathbf{x}}) = 2$, *where* $L_f h_{|\mathbf{x}} \doteq \frac{\partial h}{\partial \mathbf{x}} f(\mathbf{x})$ *denotes the Lie derivative of* h *along* f

Definition 3.1 *The* $\text{span}\{dh\ ,\ dL_f h\}$ *is called the observability Lie algebra. When the observability Lie algebra has full normal rank, the model is said to be* locally (weakly) observable.

Given the above result, we may define τ as the *unique* vector field on $U_0 \mid \mathbf{x}_0 \in U_0$ that satisfies

$$\begin{cases} L_\tau h(\mathbf{x}) = 0 \\ L_\tau L_f h(\mathbf{x}) = 1 \end{cases} \qquad \forall\ \mathbf{x} \in U_0 \tag{15}$$

which is equivalent to

$$\begin{bmatrix} dh \\ dL_f h \end{bmatrix} \tau = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{16}$$

Suppose now that we can find a diffeomorphism $F : \mathbb{R}^2 \to \mathbb{R}^2$ mapping $\mathbf{x}$ into $\mathbf{z}$ such that

$$\frac{\partial F}{\partial \mathbf{z}} = [\tau\ \ L_f \tau]\,. \tag{17}$$

Then it is easy to check that $\Phi \doteq F^{-1}$ and $k(\mathbf{z}) = \left[\frac{\partial \Phi}{\partial \mathbf{x}} f(\mathbf{x})\right]_{\Phi^{-1}(\mathbf{z})} = \begin{bmatrix} 0 & \mathbf{z}_1 \end{bmatrix}$ solve the observer linearization problem (see Isidori [5]).

Therefore the solution to the OLP boils down to the solution of the partial differential equation (PDE) of eq. (17). The first question, however, is whether the OLP is solvable at all. In order to discover that, we do not need to try to solve explicitly the PDE, for there is an equivalent condition expressed only in terms of the vector field τ:

Theorem 3.2 *(Isidori [5])*
The OLP is solvable if and only if

$$\begin{cases} 1)\ \dim(\text{span}\{dh\ ,\ dL_f h\}_{\mathbf{x}}) = 2 \\ 2)\ \tau \text{ is such that } [L_f^i \tau\ ,\ L_f^j \tau] = 0\ \forall\ i,j\ = 0,1 \end{cases} \tag{18}$$

where $[\ ,\ ]$ *denotes the Lie bracket of two vector fields:* $[f(\mathbf{x}), g(\mathbf{x})] \doteq \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} f(\mathbf{x}) - \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} g(\mathbf{x})$.

Proof:

See Isidori [5]. Note that it follows from the properties of the Lie bracket [1] that $[\tau, \tau] = [L_f \tau, L_f \tau] = 0$ and $[L_f \tau, \tau] = -[\tau, L_f \tau]$. Therefore we only need to check $[\tau, L_f \tau] = 0$.

4 Structure from motion and the observer linearization problem

Claim 4.1 *The "Observer Linearization Problem" is not solvable in the case of "Structure From Motion".*

Proof:

We start by studying the local observability of structure from motion: after some simple algebra we get

$$\left[\begin{array}{c} dh \\ dL_f h \end{array}\right] = \left[\begin{array}{cc} \mathbf{x}_2^2 & -\mathbf{x}_1\mathbf{x}_2 \\ -v_2\mathbf{x}_1 - 2\omega\mathbf{x}_1\mathbf{x}_2 & -2v_2\mathbf{x}_1 + v_1\mathbf{x}_2 - 2\omega\mathbf{x}_1^2 \end{array}\right]. \tag{19}$$

Since the normal rank of the rightmost matrix, which is defined for $\mathbf{x}_2 \neq 0$, is 2, we conclude that SFM is locally nonlinearly observable anywhere away from the center of projection and the necessary conditions of theorem 3.1 are met, and so for condition 1) of theorem 3.2.

However, condition 2) of theorem 3.2 does not hold. In fact, by solving equation 16 we get

$$\tau_1 = \frac{\mathbf{x}_1}{\mathbf{x}_2}\tau_2 \tag{20}$$

$$\tau_2 = \frac{\mathbf{x}_2^3}{v_1\mathbf{x}_2 - \mathbf{x}_1 v_2} \tag{21}$$

and, after some tedious algebra,

$$[\tau\ ,\ L_f\tau] = [\frac{{\mathbf{x}_2}^3\ (-4\,{v_2}^2\,{\mathbf{x}_1}^2 - 2\,v_2\,w\,{\mathbf{x}_1}^3 + 4\,v_1\,v_2\,\mathbf{x}_1\,\mathbf{x}_2 + v_1\,w\,{\mathbf{x}_1}^2\,\mathbf{x}_2 - 2\,v_2\,w\,\mathbf{x}_1\,{\mathbf{x}_2}^2 + v_1\,w\,{\mathbf{x}_2}^3)}{(v_2\,\mathbf{x}_1 - v_1\,\mathbf{x}_2)^3} \quad \frac{v_2\,{\mathbf{x}_2}^4\ (4\,v_2\,\mathbf{x}_1 + w\,{\mathbf{x}_1}^2 - 4\,v_1\,\mathbf{x}_2 + w\,{\mathbf{x}_2}^2)}{(-v_2\,\mathbf{x}_1 + v_1\,\mathbf{x}_2)^3}] \neq [0\ 0] \tag{22}$$

therefore we conclude from theorem 3.2 that *the observer linearization problem is* **not** *solvable in the case of structure from motion.*

The previous claim tells us that SFM is a "structurally nonlinear" estimation problem, in the sense that there exists no set of coordinates that makes the estimation error *linear*. However, the state of the model that defines the structure from motion problem is not only locally weakly observable, but also its local linearization is completely (linearly) observable. This is a very favorable situation for using local linearization-based observers, as for example the Extended Kalman Filter (EKF) [2, 6, 8]. Experimental results confirm the EKF as an appropriate tool for estimating structure from motion (see for example [21] for a review).

All of this is true as long as motion is *known*. If motion is to be inserted in the estimation process, then the model of structure from motion is no longer locally observable. Therefore alternative models have to be considered. This issue is addressed in [17].

From a geometric point of view, there is no change of coordinates which structurally modifies the observer task. However, from a *computational* (numerical) point of view, the choice of the reference frame may make a difference, depending on the application. In each specific case (broad field of view, small apertures etc.) the user has to evaluate what is the best reference frame in terms of conditioning with respect to error in the location of the projection of the feature points in the image plane as well as in the components of motion.

5 Conclusions

In this paper we have recalled the "observer linearization problem" as the problem of building a nonlinear observer for a nonlinear dynamical system, having an error which evolves according to a linear and spectrally assignable dynamic model.

We have applied results from the theory of nonlinear control and estimation theory for proving that, in the case of "structure from motion", there does not exist a change of coordinates that solves the observer linearization problem. In particular, *linear* change of coordinates, such as the transformation to object-centered or to world-centered, or alternative models of the perspective projection, cannot yield to a structural advantage in the estimation process. The only difference is based on computational (numeric) ground.

Our result also legitimates the use of local-linearization based techniques (such as the EKF) for solving structure from *known motion*. However, when motion has to be estimated as well, the geometry of the problem changes and the standard model is no longer locally observable, so that *global* techniques have to be used [18].

Acknowledgements

We wish to thank Prof. Ruggero Frezza and Prof. Giorgio Picci for their constant support and advice, Prof. Richard Murray and Prof. Shankar Sastry for their observations and useful suggestions. Also discussions with Michiel van Nieuwstadt and Andrea Mennucci were helpful.

References

[1] W. Boothby. *Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press, 1986.

[2] R.S. Bucy. Non-linear filtering theory. *IEEE Trans. A.C. AC-10, 198*, 1965.

[3] O. Faugeras. *Three dimensional vision, a geometric viewpoint.* MIT Press, 1993.

[4] R. Hermann and A. J. Krener. Nonlinear controllability and observability. *IEEE Trans. Aut. Contr. AC-22 pp. 728-740*, 1977.

[5] A. Isidori. *Nonlinear Control Systems.* Springer Verlag, 1989.

[6] A.H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, 1970.

[7] T. Kailath. *Linear Systems.* Prentice Hall, 1980.

[8] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. of the ASME-Journal of basic engineering.*, 35-45, 1960.

[9] A. J. Krener and A. Isidori. Linearization by output injection and nonlinear observers. *Systems and Control Letters vol. 3*, 1983.

[10] A. J. Krener and W. Respondek. Nonlinear observers with linearizable error dynamics. *SIAM J. Control and Optimization*, 1985.

[11] W. Lee and K. Nam. Observer design for autonomous discrete-time nonlinear systems. *Systems and Control Letters vol. 17*, 1991.

[12] L. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. of computer vision*, 1989.

[13] H. Nijmeijer. Observability of autonomous discrete time nonlinear systems. *Int. J. Control vol. 36 (5)*, 1982.

[14] J. Oliensis and J. Inigo-Thomas. Recursive multi-frame structure from motion incorporating motion error. *Proc. DARPA Image Understanding Workshop*, 1992.

[15] A. J. Van Der Shaft. Observability and controllability for smooth nonlinear systems. *SIAM J. Control and Optim. vol. 20 (3)*, 1982.

[16] C. Shekhar and R. Chellappa. Passive ranging using a moving camera. *J. of Robotics S. vol. 9*, 1992.

[17] S. Soatto. Observability/identifiability of rigid motion under perspective. *Technical Report CIT-CDS 94-001, California Institute of Technology. Reduced version submitted to the invited session "Dynamic Vision" at the 33rd Conf. on Decision and Control. Submitted to Automatica*, 1994.

[18] S. Soatto, R. Frezza, and P. Perona. Motion estimation on the essential manifold. *Computer Vision ECCV 94 – In "Lecture Notes in Computer Sciences vol. 801", Springer Verlag*, May 1994.

[19] 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.

[20] E. Sontag. *Mathematical Control Theory*. Springer Verlag, 1992.

[21] Z. Zhang and O. Faugeras. *3D dynamic scene analysis*, volume 27 of *Information Sciences*. Springer-Verlag, 1992.