

A Framework for Active Estimation: Application to Structure from Motion

Riccardo Spica, Paolo Robuffo Giordano

▶ **To cite this version:**

Riccardo Spica, Paolo Robuffo Giordano. A Framework for Active Estimation: Application to Structure from Motion. IEEE Conference on Decision and Control, CDC'13, Dec 2013, Firenze, Italy. hal-00865321

HAL Id: hal-00865321
https://inria.hal.science/hal-00865321

Submitted on 24 Sep 2013

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

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

A Framework for Active Estimation: Application to Structure from Motion

Riccardo Spica and Paolo Robuffo Giordano

***Abstract*— State estimation is a fundamental and challenging problem in many applications involving planning and control, in particular when dealing with systems exhibiting nonlinear dynamics. While the design of nonlinear observers is an active research field, the issue of *optimizing* over time the transient response of the estimation error has not received, to the best of our knowledge, a comparable attention. In this paper, an *active strategy* for tuning the transient response of a particular class of nonlinear observers is discussed. This is achieved by suitably acting on the estimation gains and on the inputs applied to the system under observation. The theory is validated by simulation results applied to two visual estimation tasks (Structure from Motion — SfM).**

I. INTRODUCTION

The possibility of estimating the state of a dynamical system by observing its input/output behavior is a very classical topic in control theory. Indeed, state observation is central to many applications: as an example, within the scope of robotics, onboard sensors are typically only able to provide partial information about the surrounding environment. In these cases, one usually tries to recover the 'world state' by means of observation techniques such as the well-known Extended Kalman Filter (EKF) for subsequent planning and/or control purposes.

While state observation is a solved problem for linear time-invariant systems, a substantial literature has instead been produced for the (significantly) more challenging case of *nonlinear* state observation, mainly because of the lack of a general methodology w.r.t. the presence of case-by-case or class-by-class solutions, as typical in nonlinear settings. For instance, still within the context of robotics, a large number of EKF-based solutions have been proposed to deal with specific nonlinear state observation problems in the presence of noise. Although successful in practice, EKF-based approaches usually imply some level of linearization/approximation of the system dynamics, thus preventing a rigorous analysis of the filter transient response.

Alternatively, the exploitation of nonlinear observation tools from classical control theory has also yielded interesting results, although typically in a deterministic context as opposed to the EKF and its variants. For instance, one could mention [1], [2] which address the problem of recovering the structure of the scene (depth for feature points or normal vector for planar scenes) from a moving camera with known linear/angular velocity. The proposed method exploits the linearity of the state dynamics w.r.t. an invertible mapping of the quantities to be estimated, and determines the observability conditions under which the estimation task can be solved. Other approaches sharing the same 'theoretical background' have also been presented in [3], [4], [5] under different conditions and/or attacking different SfM problems. Successful attempts to solve state estimation problems on *non-Euclidean* spaces can instead be found in [6], [7], while [8] discusses an observation algorithm able to recover in *closed-form* the unmeasurable states of interest.

While all these works, as well as many others not referenced here, address the general problem of nonlinear state estimation, to the best of our knowledge much less attention has been devoted to the issue of *actively optimizing* the transient response of a nonlinear observation scheme. We note that this problem is 'well-posed' since, for nonlinear systems, the observability conditions are not (in general) time-invariant but may both depend on the current state and on the *current applied inputs*. This is for instance the case of *nonuniformly observable systems* characterized by 'singular inputs' that prevent state observation [9]. It is then meaningful to study if and how one can suitably act on the estimation gains and/or system inputs in order to (i) not only guarantee the fulfilment of any specific observability criterium (as typically done), but to also (ii) 'optimize' the observation process by, e.g., shaping the error transient response.

In this paper we show how these goals can be achieved for a particular class of nonlinear observers requiring (partial) linearity of the system dynamics w.r.t. an *invertible function* of the unmeasurable states. We note that the considered class of observers, although not spanning all the possibilities, is general enough to also encompass examples of *nonuniformly observable systems* such as all SfM estimation problems: indeed, in the SfM case, (a suitable function of) the missing scene structure can be treated as unmeasurable state and the camera linear velocity (a system input) directly affects the estimation convergence.

The paper is then organized as follows: Sect. II introduces the nonlinear observation scheme considered in this work. Subsequently, Sect. III provides a detailed characterization of the associated error transient response and of the (active) strategies needed to shape its behavior over time. Then, in Sect. IV, the proposed general theory is applied to two SfM problems taken as case studies meant to concretely illustrate the various claims of the paper. Finally, Sect. V reports some simulation results validating the presented theory, and Sect. VI concludes the paper with some future directions.

R. Spica is with the University of Rennes 1 at Irisa and Inria Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France riccardo.spica@irisa.fr

P. Robuffo Giordano is with the CNRS at Irisa and Inria Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France prg@irisa.fr.

II. A NONLINEAR OBSERVATION SCHEME

Let $\boldsymbol{x} = [\boldsymbol{x}_m^T \ \boldsymbol{x}_u^T]^T \in \mathbb{R}^{m+p}$ be the state of a dynamical system in the form

$$\left\{ \begin{array}{rcl} \dot{\boldsymbol{x}}_m & = & \boldsymbol{f}_m(\boldsymbol{x}_m, \boldsymbol{u}, t) + \boldsymbol{\Omega}^T(t)\boldsymbol{x}_u \\ \dot{\boldsymbol{x}}_u & = & \boldsymbol{f}_u(\boldsymbol{x}_m, \boldsymbol{x}_u, \boldsymbol{u}, t) \end{array} \right. \tag{1}$$

with vector $\boldsymbol{x}_m$ representing a *measurable* component of the state, vector $\boldsymbol{x}_u$ an *unmeasurable* component, $\boldsymbol{u} \in \mathbb{R}^v$ the system input vector, $\boldsymbol{\Omega}(t) \in \mathbb{R}^{p\times m}$ a generic but *known* time-varying quantity, and $\boldsymbol{f}_m(\cdot)$, $\boldsymbol{f}_u(\cdot)$ sufficiently smooth functions w.r.t. their arguments[1]. Note that $\boldsymbol{x}_u$ is assumed to appear *linearly* in the dynamics of $\boldsymbol{x}_m$.

When dealing with this class of nonlinear systems, a possible estimation scheme for retrieving the (unmeasurable) value of $\boldsymbol{x}_u$ can be devised as follows: let $\hat{\boldsymbol{x}} \in \mathbb{R}^{m+p}$ be the estimated state, $\boldsymbol{\xi} = \boldsymbol{x}_m - \hat{\boldsymbol{x}}_m$, $\boldsymbol{z} = \boldsymbol{x}_u - \hat{\boldsymbol{x}}_u$, $\boldsymbol{e} = [\boldsymbol{\xi}^T \ \boldsymbol{z}^T]^T$ the total error vector, and consider the following observer

$$\left\{ \begin{array}{rcl} \dot{\hat{\boldsymbol{x}}}_m & = & \boldsymbol{f}_m(\boldsymbol{x}_m, \boldsymbol{u}, t) + \boldsymbol{\Omega}^T(t)\hat{\boldsymbol{x}}_u + \boldsymbol{H}\boldsymbol{\xi} \\ \dot{\hat{\boldsymbol{x}}}_u & = & \boldsymbol{f}_u(\boldsymbol{x}_m, \hat{\boldsymbol{x}}_u, \boldsymbol{u}, t) + \boldsymbol{\Lambda}\boldsymbol{\Omega}(t)\boldsymbol{P}\boldsymbol{\xi} \end{array} \right. \tag{2}$$

with $\boldsymbol{H} > 0$, $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^T > 0$, and $\boldsymbol{P} = \boldsymbol{P}^T > 0$. The corresponding error dynamics takes the expression

$$\left\{ \begin{array}{l} \dot{\boldsymbol{\xi}} = -\boldsymbol{H}\boldsymbol{\xi} + \boldsymbol{\Omega}^T(t)\,\boldsymbol{z} \\ \dot{\boldsymbol{z}} = -\boldsymbol{\Lambda}\boldsymbol{\Omega}(t)\,\boldsymbol{P}\boldsymbol{\xi} + (\boldsymbol{f}_u(\boldsymbol{x}_m, \boldsymbol{x}_u\,\boldsymbol{u}, t) - \boldsymbol{f}_u(\boldsymbol{x}_m, \hat{\boldsymbol{x}}_u\,\boldsymbol{u}, t)) \\ \quad = -\boldsymbol{\Lambda}\boldsymbol{\Omega}(t)\,\boldsymbol{P}\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{e}, t) \end{array} \right. \tag{3}$$

with $\boldsymbol{g}(\boldsymbol{e}, t)$ being a *vanishing* term w.r.t. the error vector $\boldsymbol{e}$, i.e., such that $\boldsymbol{g}(\boldsymbol{0}, t) = \boldsymbol{0}$, $\forall t$.

The stability of the error system (3) can be characterized by exploiting the following classical result from the adaptive control literature known as the *persistecy of excitation* Lemma, see, e.g., [10]:

Lemma 1 (Persistency of Excitation): Consider the system:

$$\left\{ \begin{array}{l} \dot{\boldsymbol{\xi}} = -\boldsymbol{H}\boldsymbol{\xi} + \boldsymbol{\Omega}^T(t)\,\boldsymbol{z} \\ \dot{\boldsymbol{z}} = -\boldsymbol{\Lambda}\boldsymbol{\Omega}(t)\,\boldsymbol{P}\boldsymbol{\xi} \end{array} \right. \tag{4}$$

where $\boldsymbol{\xi} \in \mathbb{R}^m$, $\boldsymbol{z} \in \mathbb{R}^p$, $\boldsymbol{H} > 0$, $\boldsymbol{P} = \boldsymbol{P}^T > 0$ such that

$$\boldsymbol{H}^T\boldsymbol{P} + \boldsymbol{P}^T\boldsymbol{H} = \boldsymbol{Q}, \quad \text{with } \boldsymbol{Q} > 0, \tag{5}$$

and $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^T > 0$. If $\|\boldsymbol{\Omega}(t)\|$ and $\left\|\dot{\boldsymbol{\Omega}}(t)\right\|$ are uniformly bounded and the *persistency of excitation* condition is satisfied, that is, there exists a $T > 0$ and $\gamma > 0$ such that

$$\int_t^{t+T} \boldsymbol{\Omega}(\tau)\,\boldsymbol{\Omega}^T(\tau)\,d\tau \geq \gamma\boldsymbol{I} > 0, \quad \forall t \geq t_0, \tag{6}$$

then $(\boldsymbol{\xi}, \boldsymbol{z}) = (\boldsymbol{0}, \boldsymbol{0})$ is a globally exponentially stable equilibrium point.

We note that the only difference among the actual error dynamics (3) and the 'unperturbed nominal system' (4) is the term $\boldsymbol{g}(\boldsymbol{e}, t)$ which, as explained, acts as a *vanishing* disturbance. It is then typically possible to conclude *local* exponential stability of (3) owing to the properties of $\boldsymbol{g}(\boldsymbol{e}, t)$ and to the global exponential conrvergence of the unperturbed system (4) (see [1], [2] for some examples in this sense exploiting bounds on $\|\boldsymbol{g}\|$ and on $\|\boldsymbol{e}(t_0)\|$).

The PE condition (6) plays the role of an observability criterium: convergence of the estimation error $\boldsymbol{e}(t) \to \boldsymbol{0}$ is possible iff the square matrix $\boldsymbol{\Omega}(t)\,\boldsymbol{\Omega}^T(t) \in \mathbb{R}^{p\times p}$ keeps being full rank in the integral sense of (6). We note that if $m \geq p$, that is, if more measurements $\boldsymbol{x}_m$ are available than the number of estimated quantities $\boldsymbol{x}_u$, it is in principle possible to instantaneously satisfy (6) by enforcing

$$\boldsymbol{\Omega}(t)\,\boldsymbol{\Omega}^T(t) \geq \frac{\gamma}{T}\boldsymbol{I}, \quad \forall t. \tag{7}$$

On the other hand, if $m < p$ then $\det(\boldsymbol{\Omega}(t)\,\boldsymbol{\Omega}^T(t)) \equiv 0$ by construction. Nevertheless, in this case it could still be possible to satisfy (6) in an integral sense if the $r-$dimensional range space of $\boldsymbol{\Omega}(t)\,\boldsymbol{\Omega}^T(t)$ ($r \leq m$) can span $\mathbb{R}^p$ during the period T. In this work, however, we will only consider the first situation $m \geq p$ and thus aim at fulfilling the (more restrictive) condition (7). Finally, we note that in the special situation $\dot{\boldsymbol{x}}_u = \boldsymbol{0}$ one obviously has $\boldsymbol{g}(\boldsymbol{e}, t) \equiv \boldsymbol{0}$, thus resulting in a perfect match among (3) and (4): in case of an unknown but *constant* $\boldsymbol{x}_u$, the estimation convergence becomes global[2].

We now perform some manipulations of system (3) in order to slightly simplify its structure and highlight some important features exploited in the next Sections. Being symmetric and positive-definite, we can let $\boldsymbol{P} = \boldsymbol{P}^{\frac{1}{2}}\boldsymbol{P}^{\frac{1}{2}}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}$, with $\boldsymbol{P}^{\frac{1}{2}} = \boldsymbol{P}^{\frac{1}{2}T} > 0$ and $\boldsymbol{\Lambda}^{\frac{1}{2}} = \boldsymbol{\Lambda}^{\frac{1}{2}T} > 0$, and analogously for $\boldsymbol{P}^{-1}$ and $\boldsymbol{\Lambda}^{-1}$. Consider then the following invertible change of coordinates

$$\left\{ \begin{array}{l} \tilde{\boldsymbol{\xi}} = \boldsymbol{P}^{\frac{1}{2}}\boldsymbol{\xi} \\ \tilde{\boldsymbol{z}} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{z} \end{array} \right. . \tag{8}$$

In the new coordinates, system (3) takes the form

$$\begin{pmatrix} \dot{\tilde{\boldsymbol{\xi}}} \\ \dot{\tilde{\boldsymbol{z}}} \end{pmatrix} = \left[\begin{pmatrix} \boldsymbol{0} & \tilde{\boldsymbol{\Omega}}^T(t) \\ -\tilde{\boldsymbol{\Omega}}(t) & \boldsymbol{0} \end{pmatrix} - \begin{pmatrix} \tilde{\boldsymbol{H}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \right] \begin{pmatrix} \tilde{\boldsymbol{\xi}} \\ \tilde{\boldsymbol{z}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} \\ \tilde{\boldsymbol{g}} \end{pmatrix}, \tag{9}$$

with $\tilde{\boldsymbol{H}} = \boldsymbol{P}^{\frac{1}{2}}\boldsymbol{H}\boldsymbol{P}^{-\frac{1}{2}}$, $\tilde{\boldsymbol{\Omega}}(t) = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Omega}(t)\,\boldsymbol{P}^{\frac{1}{2}}$, and $\tilde{\boldsymbol{g}} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{g}$. We can then note the following facts: system (9) has an evident port-Hamiltonian structure which is, again, perfectly recovered in the unperturbed case ($\tilde{\boldsymbol{g}} \equiv \boldsymbol{0}$). The Hamiltonian (storage function) for (9) is the lower-bounded scalar function

$$\mathcal{H}(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{z}}) = \frac{1}{2}\tilde{\boldsymbol{\xi}}^T\tilde{\boldsymbol{\xi}} + \frac{1}{2}\tilde{\boldsymbol{z}}^T\tilde{\boldsymbol{z}} = \frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{P}\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{z}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{z} \geq 0.$$

Furthermore, the gain matrices $\boldsymbol{P}$ and $\boldsymbol{\Lambda}$, free design parameters, can be suitably exploited to fulfill two independent objectives. First, since

$$\dot{\mathcal{H}} = -\tilde{\boldsymbol{\xi}}^T\tilde{\boldsymbol{H}}_s\tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{z}}^T\tilde{\boldsymbol{g}} = -\tilde{\boldsymbol{\xi}}^T\tilde{\boldsymbol{H}}_s\tilde{\boldsymbol{\xi}} + \boldsymbol{z}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{g},$$

one can conclude that, for a bounded disturbance $\|\boldsymbol{g}\| \leq M$, it is always possible to attenuate at will its (possibly

[1]Despite their explicit time dependency, we also assume $\boldsymbol{f}_m(\cdot)$ to be a function of only *known* quantities and similarly for $\boldsymbol{f}_u(\cdot)$ apart, in this case, from the unknown value of $\boldsymbol{x}_u$.

[2]We stress, however, that in this work we are not interested in the identification of unknown *constant* plant parameters, but in a *state observation problem* in which the unknown $\boldsymbol{x}_u$ obeys the (non-negligible) dynamics (1).

destabilizing) contribution by letting the gain $\boldsymbol{\Lambda} \to \infty$, basically obtaining a 'semi-global' vs. local stability condition. Furthermore, being

$$\tilde{\boldsymbol{\Omega}}(t)\tilde{\boldsymbol{\Omega}}^T(t) = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Omega}(t)\boldsymbol{P}\boldsymbol{\Omega}^T(t)\boldsymbol{\Lambda}^{\frac{1}{2}}, \tag{10}$$

it is also possible to directly affect the norm of $\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{\Omega}}^T$ by acting on the gains $\boldsymbol{P}$ and $\boldsymbol{\Lambda}$. For instance, if $\boldsymbol{\Lambda}$ is chosen for coping with the disturbance $\boldsymbol{g}$, one can still exploit the other matrix gain $\boldsymbol{P}$ to compensate for a possibly large value of $\boldsymbol{\Lambda}$ so as to keep a limited norm of $\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{\Omega}}^T$. Having an explicit control over the norm of $\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{\Omega}}^T$, or equivalently over its (real) eigenvalues, will be pivotal for the next developments.

Finally, it is straightforward to show that the asymptotic stability of system (9) is preserved when adopting a sufficiently smooth *time-varying* gain matrix $h_1\boldsymbol{I} \leq \tilde{\boldsymbol{H}}(t) \leq h_2\boldsymbol{I}$, $0 < h_1 \leq h_2 < \infty$, with bounded $\left\|\dot{\tilde{\boldsymbol{H}}}(t)\right\|$: this important feature allows to suitably shape the gain matrix $\tilde{\boldsymbol{H}}$ over time in order to fulfil additional objectives of interest, as it will be the case in the next developments.

III. A STRATEGY FOR ACTIVE ESTIMATION

The goal of this Section is to present a strategy for actively tuning the transient response of the estimation error vector $\tilde{\boldsymbol{z}}(t) = \boldsymbol{\Lambda}^{-\frac{1}{2}}(\boldsymbol{x}_u(t) - \hat{\boldsymbol{x}}_u(t))$ in (9). To this end, we assume that $\tilde{\boldsymbol{\Omega}}(t) = \tilde{\boldsymbol{\Omega}}(\boldsymbol{x}_m, \boldsymbol{u}(t))$: presence of this structure allows to (actively) exploit the input vector $\boldsymbol{u}(t)$ for affecting matrix $\tilde{\boldsymbol{\Omega}}$ and, as a consequence, the transient response of (9) as sought[3]. The following analysis will also neglect presence of the disturbance term $\tilde{\boldsymbol{g}}(t)$, since, as explained before, its distorting effects can be typically made arbitrarily small by a proper choice of the gain $\boldsymbol{\Lambda}$.

A. Characterization of the system transient response

Following the port-Hamiltonian interpretation of system (9), in particular with in mind a standard mechanical system, one can identify vector $\tilde{\boldsymbol{z}}$ as playing the role of a 'position'-like quantity, and vector $\tilde{\boldsymbol{\xi}}$ as that of a 'velocity'-like quantity upon which a dissipative action is present. Therefore, analogously to a mechanical system, we focus the analysis on the dynamics of vector $\ddot{\tilde{\boldsymbol{z}}}$.

Being $\dot{\tilde{\boldsymbol{z}}} = -\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{\xi}}$, it is

$$\begin{aligned}\ddot{\tilde{\boldsymbol{z}}} =& -\dot{\tilde{\boldsymbol{\Omega}}}\tilde{\boldsymbol{\xi}} - \tilde{\boldsymbol{\Omega}}\dot{\tilde{\boldsymbol{\xi}}} = -\dot{\tilde{\boldsymbol{\Omega}}}\tilde{\boldsymbol{\xi}} - \tilde{\boldsymbol{\Omega}}(-\tilde{\boldsymbol{H}}\tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\Omega}}^T\tilde{\boldsymbol{z}}) = \\ =&(\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{H}} - \dot{\tilde{\boldsymbol{\Omega}}})\tilde{\boldsymbol{\xi}} - \tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{\Omega}}^T\tilde{\boldsymbol{z}} = (\dot{\tilde{\boldsymbol{\Omega}}}\tilde{\boldsymbol{\Omega}}^\dagger - \tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{\Omega}}^\dagger)\dot{\tilde{\boldsymbol{z}}} - \tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{\Omega}}^T\tilde{\boldsymbol{z}}\end{aligned} \tag{11}$$

with $\tilde{\boldsymbol{\Omega}}^\dagger \in \mathbb{R}^{m\times p}$ denoting the pseudo-inverse of matrix $\tilde{\boldsymbol{\Omega}}$. Let $\tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^T = \tilde{\boldsymbol{\Omega}}$ be the singular value decomposition of matrix $\tilde{\boldsymbol{\Omega}}$, where $\tilde{\boldsymbol{\Sigma}} = [\tilde{\boldsymbol{S}} \quad \boldsymbol{0}]$, $\tilde{\boldsymbol{S}} = \mathrm{diag}(\tilde{\sigma}_i) \in \mathbb{R}^{p\times p}$, and $0 \leq \tilde{\sigma}_1 \leq \ldots \leq \tilde{\sigma}_p$, from which it directly follows $\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{\Omega}}^T = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{S}}^2\tilde{\boldsymbol{U}}^T$ and $\tilde{\boldsymbol{\Omega}}^\dagger = \tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{\Sigma}}^\dagger\tilde{\boldsymbol{U}}$ where, as usual, $\tilde{\boldsymbol{\Sigma}}^\dagger = [\tilde{\boldsymbol{S}}^{-1} \quad \boldsymbol{0}]^T$ with $\tilde{\boldsymbol{S}}^{-1} = \mathrm{diag}(\tilde{\varsigma}_i)$, $\tilde{\varsigma}_i = 1/\tilde{\sigma}_i$ if $\tilde{\sigma}_i > 0$ and $\tilde{\varsigma}_i = 0$ otherwise.

[3]We note that this assumption is not too restrictive in our context, since it is typically encountered when facing 'non-trivial' state estimation problems, i.e., problems for which a linear time-invariant approximation of the original systems is unobsevable. Furthermore, this assumption is certainly met for SfM problems, see Sect. IV and [1], [2].

As for $\dot{\tilde{\boldsymbol{\Omega}}}$ it is $\dot{\tilde{\boldsymbol{\Omega}}} = \dot{\tilde{\boldsymbol{U}}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^T + \tilde{\boldsymbol{U}}\dot{\tilde{\boldsymbol{\Sigma}}}\tilde{\boldsymbol{V}}^T + \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\dot{\tilde{\boldsymbol{V}}}^T$. Exploiting the orthonormality of $\tilde{\boldsymbol{U}}$, we have $\tilde{\boldsymbol{U}}^T\tilde{\boldsymbol{U}} = \boldsymbol{I} \implies \dot{\tilde{\boldsymbol{U}}}^T\tilde{\boldsymbol{U}} + \tilde{\boldsymbol{U}}^T\dot{\tilde{\boldsymbol{U}}} = \boldsymbol{0}$. Denoting the skew-symmetric matrix $\tilde{\boldsymbol{U}}^T\dot{\tilde{\boldsymbol{U}}} = \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{U}}$, it is $\dot{\tilde{\boldsymbol{U}}} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{U}}$ and, following the same arguments, one has $\dot{\tilde{\boldsymbol{V}}}^T\tilde{\boldsymbol{V}} = \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{V}} = -\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{V}}^T$ and $\dot{\tilde{\boldsymbol{V}}}^T = \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$. Therefore,

$$\dot{\tilde{\boldsymbol{\Omega}}} = \tilde{\boldsymbol{U}}(\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}} + \dot{\tilde{\boldsymbol{\Sigma}}} + \tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{V}})\tilde{\boldsymbol{V}}^T. \tag{12}$$

We highlight that, as shown in [11], matrices $\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{U}}$, $\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{V}}$, $\dot{\tilde{\boldsymbol{\Sigma}}}$ can be computed in *closed-form* from the knowledge of $(\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}}, \tilde{\boldsymbol{\Sigma}})$ and of the closed-form expression of $\tilde{\boldsymbol{\Omega}}$. This is also valid in our context since an explicit expression of $\tilde{\boldsymbol{\Omega}}$ is assumed available, while matrices $\tilde{\boldsymbol{U}}$, $\tilde{\boldsymbol{V}}$, $\tilde{\boldsymbol{\Sigma}}$ can be numerically retrieved from $\tilde{\boldsymbol{\Omega}}$ via any standard SVD routine. Finally, exploiting (12) we have

$$\begin{aligned}\dot{\tilde{\boldsymbol{\Omega}}}\tilde{\boldsymbol{\Omega}}^\dagger =&\tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\Sigma}}^\dagger\tilde{\boldsymbol{U}}^T + \tilde{\boldsymbol{U}}\dot{\tilde{\boldsymbol{\Sigma}}}\tilde{\boldsymbol{\Sigma}}^\dagger\tilde{\boldsymbol{U}}^T + \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{V}}\tilde{\boldsymbol{\Sigma}}^\dagger\tilde{\boldsymbol{U}}^T \\ =&\tilde{\boldsymbol{U}}(\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{U}} + \dot{\tilde{\boldsymbol{S}}}\tilde{\boldsymbol{S}}^{-1} + \tilde{\boldsymbol{S}}\bar{\boldsymbol{\Gamma}}_{\boldsymbol{V}}\tilde{\boldsymbol{S}}^{-1})\tilde{\boldsymbol{U}}^T\end{aligned} \tag{13}$$

where $\bar{\boldsymbol{\Gamma}}_V = -\bar{\boldsymbol{\Gamma}}_V^T$ is the $p \times p$ upper-left block of matrix $\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{V}}$.

At this point, the dissipation matrix is purposely taken as

$$\tilde{\boldsymbol{H}} = \tilde{\boldsymbol{V}}\begin{bmatrix}\boldsymbol{D}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D}_2\end{bmatrix}\tilde{\boldsymbol{V}}^T \tag{14}$$

with $\boldsymbol{D}_1 \in \mathbb{R}^{p\times p} > 0$, $\boldsymbol{D}_2 \in \mathbb{R}^{(m-p)\times(m-p)} > 0$, and, thus, $\tilde{\boldsymbol{H}} > 0$ as well. This choice in fact yields

$$\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{\Omega}}^\dagger = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{S}}\boldsymbol{D}_1\tilde{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{U}}^T. \tag{15}$$

By combining (11) with (13)–(15), and exploiting the diagonal form of matrix $\tilde{\boldsymbol{S}}$, we finally obtain

$$\begin{aligned}\ddot{\tilde{\boldsymbol{z}}} =&\tilde{\boldsymbol{U}}(\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{U}} + \dot{\tilde{\boldsymbol{S}}}\tilde{\boldsymbol{S}}^{-1} + \tilde{\boldsymbol{S}}\bar{\boldsymbol{\Gamma}}_{\boldsymbol{V}}\tilde{\boldsymbol{S}}^{-1} - \tilde{\boldsymbol{S}}\boldsymbol{D}_1\tilde{\boldsymbol{S}}^{-1})\tilde{\boldsymbol{U}}^T\dot{\tilde{\boldsymbol{z}}} - \tilde{\boldsymbol{U}}\tilde{\boldsymbol{S}}^2\tilde{\boldsymbol{U}}^T\tilde{\boldsymbol{z}} \\ =&(\tilde{\boldsymbol{U}}\tilde{\boldsymbol{S}})(\tilde{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{U}}\tilde{\boldsymbol{S}} + \dot{\tilde{\boldsymbol{S}}}\tilde{\boldsymbol{S}}^{-1} + \bar{\boldsymbol{\Gamma}}_{\boldsymbol{V}} - \boldsymbol{D}_1)(\tilde{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{U}}^T)\dot{\tilde{\boldsymbol{z}}} - \tilde{\boldsymbol{U}}\tilde{\boldsymbol{S}}^2\tilde{\boldsymbol{U}}^T\tilde{\boldsymbol{z}} \\ =&(\tilde{\boldsymbol{U}}\tilde{\boldsymbol{S}})(\tilde{\boldsymbol{\Pi}} - \boldsymbol{D}_1)(\tilde{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{U}}^T)\dot{\tilde{\boldsymbol{z}}} - (\tilde{\boldsymbol{U}}\tilde{\boldsymbol{S}})\tilde{\boldsymbol{S}}^2(\tilde{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{U}}^T)\tilde{\boldsymbol{z}}\end{aligned} \tag{16}$$

where

$$\tilde{\boldsymbol{\Pi}} = \tilde{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{U}}\tilde{\boldsymbol{S}} + \dot{\tilde{\boldsymbol{S}}}\tilde{\boldsymbol{S}}^{-1} + \bar{\boldsymbol{\Gamma}}_{\boldsymbol{V}}. \tag{17}$$

The expression obtained in (16) has a clear and neat structure: it indicates presence of a change of coordinates

$$\boldsymbol{\eta} = (\tilde{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{U}}^T)\tilde{\boldsymbol{z}} \tag{18}$$

in which, in the approximation $\tilde{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{U}}^T \approx const$, the system exhibits the simple (and almost diagonal) form

$$\ddot{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\Pi}} - \boldsymbol{D}_1)\dot{\boldsymbol{\eta}} - \tilde{\boldsymbol{S}}^2\boldsymbol{\eta}, \tag{19}$$

that is, a (unit-)mass-spring-damper system with diagonal stiffness matrix $\tilde{\boldsymbol{S}}^2$.

The convergence rate of (19) is then related to the slowest mode of the system, i.e., that associated to the element $\tilde{\sigma}_1^2$ in $\tilde{\boldsymbol{S}}^2$. Therefore, in order to impose a given convergence speed and overall transient response to (19) (resp. to the estimation error dynamics (11)), one can try to '*place the poles*' of (19) by (i) regulating $\tilde{\sigma}_1^2$ to a desired value $\tilde{\sigma}_{1,des}^2$ and, at the same time, (ii) shaping the damping factor $\boldsymbol{D}_1$ in order to prevent the occurrence of oscillatory modes ($\sim$ complex poles).

B. Shaping the Damping Factor

A reasonable choice for matrix $\boldsymbol{D}_1$ could be

$$\boldsymbol{D}_1 = \tilde{\boldsymbol{\Pi}} + \boldsymbol{C} \tag{20}$$

with $\boldsymbol{C}$ any positive definite matrix, w.l.o.g. a diagonal one $\boldsymbol{C} = \mathrm{diag}(c_i)$, $c_i > 0$, so as to obtain a completely decoupled transient behavior for (19)

$$\ddot{\eta}_i + c_i \dot{\eta}_i + \sigma_i^2 \eta_i = 0, \quad i = 1 \dots p. \tag{21}$$

For instance, taking $c_i = c_i^* = 2\sigma_i$ would (conveniently) result in a critically damped state evolution. Matrix $\boldsymbol{D}_1$, however, is bound to remain positive definite over time, a constraint which, clearly, is not necessarily met by (20) for any arbitrary pair $(\boldsymbol{C}, \tilde{\boldsymbol{\Pi}})$.

By suitably bounding $\tilde{\boldsymbol{\Pi}}$ as, e.g., $\|\tilde{\boldsymbol{\Pi}}\| \leq q\boldsymbol{I}$, $q \geq 0$, any $\boldsymbol{C} > q\boldsymbol{I}$ could guarantee $\boldsymbol{D}_1 > 0$. This possibility, however, could result in an *over-damped transient response* for the system since, in the general case, one could have $\boldsymbol{C} > q\boldsymbol{I} > \mathrm{diag}(c_i^*)$. In other words, aiming at obtaining a completely decoupled behavior for the evolution of $\boldsymbol{\eta}(t)$ as in (21) could require a unnecessary degradation of the transient response. Therefore, in the following we will *not* aim for a cancelation of matrix $\tilde{\boldsymbol{\Pi}}$, but we will rather neglect its effects on the transient by just taking $\boldsymbol{D}_1 = \mathrm{diag}(c_i^*) > 0$. This can of course result in a poorer overall behavior (for not compensating for $\tilde{\boldsymbol{\Pi}}$), but avoids the introduction of unnecessary lower bounds on $\boldsymbol{D}_1$. The simulation results reported in Sect. V will also show that not compensating for matrix $\tilde{\boldsymbol{\Pi}}$ has a marginal effect.

C. Tuning the stiffness matrix $\tilde{\boldsymbol{S}}^2$

We recall that matrix $\tilde{\boldsymbol{S}}^2 = \mathrm{diag}(\tilde{\sigma}_i^2)$ contains the p eigenvalues of the square symmetric matrix $\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{\Omega}}^T$ in (10). Let then $\boldsymbol{S}^2 = \mathrm{diag}(\sigma_i^2)$ represent the eigenvalues of matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ in the original coordinates $(\boldsymbol{\xi}, \boldsymbol{z})$. From (10) it follows that, in order to affect $\tilde{\boldsymbol{S}}^2$, one can either act on the gains $\boldsymbol{P}$, $\boldsymbol{\Lambda}$ for a given $\boldsymbol{S}^2$, or actively adjust $\boldsymbol{S}^2$ for a given set of gains $\boldsymbol{P}$, $\boldsymbol{\Lambda}$ (or, of course, any combination of both actions). The effect of gains $\boldsymbol{P}$, $\boldsymbol{\Lambda}$ has already been discussed in Sect. II: in short, one can exploit them to freely amplify/attenuate the eigenvalues of $\tilde{\boldsymbol{S}}^2$ as clear from (10). However, we note that, whatever the gain choice, one still needs to ensure a minimum threshold $\sigma_1^2(t) \geq \sigma_{min}^2 > 0$ for the estimation to converge, i.e., for fulfilling condition (7): this can only be achieved by actively tuning matrix $\boldsymbol{S}^2$ as explained in the following.

We start assuming for simplicity of exposition (but without a significant loss of generality) a 'scalar' form for matrices $\boldsymbol{P}$ and $\boldsymbol{\Lambda}$, i.e., $\boldsymbol{P} = \alpha\boldsymbol{I}$ and $\boldsymbol{\Lambda} = \beta\boldsymbol{I}$, $\alpha > 0$, $\beta > 0$, a choice yielding the simple relationship $\tilde{\boldsymbol{S}}^2 = \alpha\beta\boldsymbol{S}^2 \implies \tilde{\sigma}_i^2 = \alpha\beta\sigma_i^2$. Therefore, seeking a desired value $\tilde{\sigma}_{i,des}^2$ is equivalent to imposing $\sigma_i^2 \to \sigma_{i,des}^2 = \tilde{\sigma}_{i,des}^2/(\alpha\beta)$. We can then focus on the regulation of the eigenvalues σ_i^2.

An explicit expression of the time derivative of the eigenvalues σ_i^2 can be obtained as follows: being $\boldsymbol{\Omega}(t) = \boldsymbol{\Omega}(\boldsymbol{x}_m, \boldsymbol{u}(t))$, it is $\sigma_i^2(t) = \sigma_i^2(\boldsymbol{x}_m, \boldsymbol{u}(t))$ which, exploiting the results of [11], [12], allows to conclude

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_i^2 = \sum_{j=1}^{v} \left(\boldsymbol{v}_i^T \frac{\partial(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)}{\partial u_j} \boldsymbol{v}_i \dot{u}_j \right) + \sum_{j=1}^{n} \left(\boldsymbol{v}_i^T \frac{\partial(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)}{\partial x_{m_j}} \boldsymbol{v}_i \dot{x}_{m_j} \right) \tag{22}$$

where $\boldsymbol{v}_i \in \mathbb{R}^p$ is the *normalized* eigenvector associated to σ_i^2. Letting

$$\boldsymbol{J}_{u,i} = \left[\begin{array}{ccc} \boldsymbol{v}_i^T \dfrac{\partial(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)}{\partial u_1} \boldsymbol{v}_i & \dots & \boldsymbol{v}_i^T \dfrac{\partial(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)}{\partial u_v} \boldsymbol{v}_i \end{array} \right] \in \mathbb{R}^{1 \times v} \tag{23}$$

and

$$\boldsymbol{J}_{x,i} = \left[\begin{array}{ccc} \boldsymbol{v}_i^T \dfrac{\partial(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)}{\partial x_{m_1}} \boldsymbol{v}_i & \dots & \boldsymbol{v}_i^T \dfrac{\partial(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)}{\partial x_{m_n}} \boldsymbol{v}_i \end{array} \right] \in \mathbb{R}^{1 \times n}, \tag{24}$$

eq. (22) can be compactly rewritten as

$$\dot{(\sigma_i^2)} = \boldsymbol{J}_{u,i}\dot{\boldsymbol{u}} + \boldsymbol{J}_{x,i}\dot{\boldsymbol{x}}_m. \tag{25}$$

Note, again, that the Jacobian matrices $\boldsymbol{J}_{u,i}$ and $\boldsymbol{J}_{x,i}$ in (23–24) can be computed in *closed-form* from the knowledge of the eigenvectors $\boldsymbol{v}_i$ and of a closed-form expression for matrix $\boldsymbol{\Omega}$.

At this point, any differential inversion technique can be applied to (25) in order to affect the behavior of the i-th eigenvalue $\sigma_i^2(t)$ by acting upon vector $\dot{\boldsymbol{u}}$: this must then be treated as the 'actual' input vector, with $\boldsymbol{u}$ regarded, instead, as an internal state. The next Section will discuss some examples in this sense.

We finally note that, when inverting (25), it is in general not possible to fully compensate for the term $\boldsymbol{J}_{x,i}\dot{\boldsymbol{x}}_m$: the system dynamics in (1) implies a direct dependence of $\dot{\boldsymbol{x}}_m$ from the unmeasurable $\boldsymbol{x}_u$, so that an exact evaluation of $\dot{\boldsymbol{x}}_m$ is not obtainable in practice. However, when viable a simple possibility is to just enforce $\dot{\boldsymbol{x}}_m \simeq 0$ during the system evolution as shown in the next case studies.

IV. CASE STUDIES

In this Section we present two case studies meant to illustrate how to exploit the active estimation framework introduced so far. We focus our attention on the issue of 3D structure identification with a monocular camera, in particular on the problems of depth estimation for a point feature [1] and identification of the normal and distance to a planar scene using image moments, namely area and barycenter [2]. These two examples have been chosen because of their practical relevance and also because of their complementarity w.r.t. the proposed active estimation framework. In fact, the first case is characterized by a single unknown quantity ($p = 1$) vs. two available measurements ($m = 2$), thus resulting in a scalar $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$. The second case, instead, involves three unknown quantities vs. three available measurements, hence yielding a 'square' problem with $p = m = 3$.

A. Depth estimation for a point feature

Let $\boldsymbol{p} = [x \;\; y \;\; 1]^T = [X/Z \;\; Y/Z \;\; 1]^T \in \mathbb{R}^3$ be the perspective projection of a 3D point (X, Y, Z) onto

the image plane of a calibrated pinhole camera. As well known [13], the interaction matrix of a point feature $\boldsymbol{p}$ is

$$\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} = \begin{bmatrix} -\frac{1}{Z} & 0 & \frac{x}{Z} & xy & -\left(1+x^2\right) & y \\ 0 & -\frac{1}{Z} & \frac{y}{Z} & 1+y^2 & -xy & -x \end{bmatrix} \boldsymbol{u} \tag{26}$$

where $\boldsymbol{u} = \begin{bmatrix}\boldsymbol{v}^T & \boldsymbol{\omega}^T\end{bmatrix}^T \in \mathbb{R}^6$ is the camera linear/angular velocity in the camera frame, and Z the *unmeasurable depth* of the feature point. The dynamics of Z is

$$\dot{Z} = \begin{bmatrix}0 & 0 & -1 & -yZ & xZ & 0\end{bmatrix} \boldsymbol{u}.$$

The expression in (26) is not linear in Z but it is linear in $1/Z$. Therefore, by defining $\boldsymbol{x}_m = [x\ y]^T$ and $x_u = 1/Z$, with then $m = 2$ and $p = 1$, we obtain for (1)–(3)

$$\begin{cases} \boldsymbol{f}_m(\boldsymbol{x}_m, \boldsymbol{u}) = \begin{bmatrix}0 & 0 & 0 & xy & -\left(1+x^2\right) & y \\ 0 & 0 & 0 & 1+y^2 & -xy & -x\end{bmatrix} \boldsymbol{u} \\ \boldsymbol{\Omega}(\boldsymbol{x}_m, \boldsymbol{u}) = \begin{bmatrix} xv_z - v_x & yv_z - v_y \end{bmatrix} \\ f_u(\boldsymbol{x}_m, x_u, \boldsymbol{u}) = v_z x_u^2 + (yw_x - xw_y)\, x_u \\ g(\boldsymbol{e}, t) = v_z\left(x_u^2 - \hat{x}_u^2\right) + (yw_x - xw_y)\, z \end{cases}. \tag{27}$$

We note that $g(\mathbf{0}, t) = 0$ as expected.

In this case matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ reduces to its single eigenvalue $\sigma_1^2 = \|\boldsymbol{\Omega}\|^2 = (xv_z - v_x)^2 + (yv_z - v_y)^2$, and it is possible to obtain an explicit expression for the SVD decomposition $\boldsymbol{\Omega} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ as

$$\begin{cases} \boldsymbol{U} = 1 \\ \boldsymbol{\Sigma} = \begin{bmatrix}\sigma_1 & 0\end{bmatrix} \\ \boldsymbol{V} = \frac{1}{\sigma_1^2}\begin{bmatrix}\Omega_1 & -\Omega_2 \\ \Omega_2 & \Omega_1\end{bmatrix} \end{cases}.$$

It is interesting to note that since, in this case, $\boldsymbol{U} = 1$ and $\boldsymbol{\Gamma}_u = 0$, if $\sigma_1(t) \equiv const > 0$ then by construction $\tilde{\boldsymbol{S}}^{-1}\tilde{\boldsymbol{U}}^T = const$ in (18). Therefore, it is always possible to enforce the 'ideal' estimation error dynamics (21) by keeping $\|\boldsymbol{\Omega}\|^2 = \sigma_1^2 = const$.

Finally, we note that $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{x}_m, \boldsymbol{v})$ and that, in this case, the Jacobian $\boldsymbol{J}_{u,1}$ takes the expression

$$\boldsymbol{J}_{u,1} = 2\begin{bmatrix} v_x - xv_z \\ v_y - yv_z \\ (xv_z - v_x)\, x + (yv_z - v_y)\, y \\ 0 \\ 0 \\ 0 \end{bmatrix}^T = \begin{bmatrix}\boldsymbol{J}_{\boldsymbol{v},1} & \mathbf{0}\end{bmatrix}. \tag{28}$$

This allows to conclude the following: first, since $\boldsymbol{\Omega}$ does not depend on $\boldsymbol{\omega}$, and in fact $\boldsymbol{J}_{u,1}$ has a structural $\mathbf{0}$ in its last three components, one can freely exploit the camera angular velocity for fulfilling any desired task without interfering with the regulation of $\sigma_1^2(t)$ (only affected by $\boldsymbol{v}$). For instance, one can use $\boldsymbol{\omega}$ for keeping $\boldsymbol{x}_m \simeq const$ as much as possible in order to render the effect of $\dot{\boldsymbol{x}}_m$ negligible in (25). This possibility will be actually exploited in the next simulation results.

Second, from (28) it follows $\boldsymbol{J}_{\boldsymbol{v},1}\boldsymbol{p} = 0$: the gradient of σ_1^2 is orthogonal to projection ray passing through $\boldsymbol{p}$. Intuitively, when following the gradient $\boldsymbol{J}_{\boldsymbol{v},1}$ for regulating the value of $\sigma_1^2(t)$, no acceleration is imposed along the $\boldsymbol{p}$ direction. If, therefore, the camera initial velocity has a null component along $\boldsymbol{p}$, the resulting motion will lie on a circular path centered at the observed point feature.

B. 3D structure estimation from image moments

Consider a planar object $\mathcal{O}$ with plane equation $\boldsymbol{n}^T\boldsymbol{E} + d = 0$ in the camera frame, where $\boldsymbol{n} \in \mathbb{S}^2$ is the plane unit normal and $d \in \mathbb{R}$ the distance to the plane from the origin of the camera frame. From [14], the depth Z on any 3D point $\boldsymbol{E} \in \mathbb{R}^3$ lying on this plane can be expressed in terms of its normalized image coordinates $\boldsymbol{p} = [x\, y\, 1]^T$ as

$$\frac{1}{Z} = Ax + By + C,$$

where $\boldsymbol{\chi} = [A\ \ B\ \ C]^T = -\boldsymbol{n}/d \in \mathbb{R}^3$ represents an *unmeasurable* 3D scene structure (as with Z for the point feature case).

Consider a generic (i, j)-th order moment m_{ij} evaluated on the image of $\mathcal{O}$: still from [14], the dynamics of m_{ij} can be shown to take the expression

$$\dot{m}_{ij} = \boldsymbol{J}_{m_{ij}}(m_{kl}, \boldsymbol{\chi})\, \boldsymbol{u} = f_{1_{ij}}(m_{kl}, \boldsymbol{u}) + \boldsymbol{f}_{2_{ij}}(m_{kl}, \boldsymbol{u})\boldsymbol{\chi} \tag{29}$$

where m_{kl} stands for a generic (k, l)-th moment of order up to $i + j + 1$. Note that (29) is linear in the unmeasurable $\boldsymbol{\chi}$, while all the other quantities are available to measurement. Let then $\boldsymbol{x}_m = [m_{i_1 j_1} \dots m_{i_m j_m}]^T \in \mathbb{R}^m$ be a collection of m image moments and $\boldsymbol{x}_u = \boldsymbol{\chi}$. From [2] one has $\dot{\boldsymbol{x}}_u = \boldsymbol{x}_u\boldsymbol{x}_u^T\boldsymbol{v} - [\boldsymbol{\omega}]_\times \boldsymbol{x}_u$, so that formulation (1) can be recovered with

$$\begin{cases} \boldsymbol{f}_m(\boldsymbol{x}_m, \boldsymbol{u}) = \begin{bmatrix} f_{1_{i_1 j_1}}(\boldsymbol{x}_m, \boldsymbol{u}) & \dots & f_{1_{i_m j_m}}(\boldsymbol{x}_m, \boldsymbol{u}) \end{bmatrix}^T \\ \boldsymbol{\Omega}(\boldsymbol{x}_m, \boldsymbol{u}) = \begin{bmatrix} \boldsymbol{f}_{2_{i_1 j_1}}^T(\boldsymbol{x}_m, \boldsymbol{u}) & \dots & \boldsymbol{f}_{2_{i_m j_m}}^T(\boldsymbol{x}_m, \boldsymbol{u}) \end{bmatrix}^T \\ \boldsymbol{f}_u(\boldsymbol{x}_m, \boldsymbol{x}_u, \boldsymbol{u}) = \boldsymbol{x}_u\boldsymbol{x}_u^T\boldsymbol{v} - [\boldsymbol{\omega}]_\times \boldsymbol{x}_u \\ \boldsymbol{g}(\boldsymbol{e}, t) = (\boldsymbol{x}_u\boldsymbol{x}_u^T - \hat{\boldsymbol{x}}_u\hat{\boldsymbol{x}}_u^T)\boldsymbol{v} - [\boldsymbol{\omega}]_\times \boldsymbol{z} \end{cases}.$$

where $\boldsymbol{\Omega} \in \mathbb{R}^{p \times m}$, $p = 3$. As before, it is $\boldsymbol{g}(\mathbf{0}, t) = \mathbf{0}$.

Contrarily to the point feature case, choice of which moments to include in $\boldsymbol{x}_m$ is in general not obvious as many possibilities exist (in number and kind). In this work, we opted for the lowest-order moments because of their robustness w.r.t. image noise: $\boldsymbol{x}_m = [a\ \ x_g\ \ y_g]^T$, i.e., the area a and barycenter (x_g, y_g) of the observed object. This choice implies $m = p = 3$, thus yielding a 'square' problem, i.e., a square matrix $\boldsymbol{\Omega}$ with expression

$$\boldsymbol{\Omega} = \begin{bmatrix} 3ax_g v_z - av_x & \left(x_g^2 + 4n_{20}\right) v_z - x_g v_x & \left(x_g y_g + 4n_{11}\right) v_z - x_g v_y \\ 3ay_g v_z - av_y & \left(x_g y_g + 4n_{11}\right) v_z - y_g v_x & \left(y_g^2 + 4n_{02}\right) v_z - y_g v_y \\ 2av_z & x_g v_z - v_x & y_g v_z - v_y \end{bmatrix} \tag{30}$$

where (n_{20}, n_{11}, n_{02}) are the normalized centered moments of order 2.

We can note that, as before, $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{x}_m, \boldsymbol{v})$ thus allowing to exploit the camera angular velocity $\boldsymbol{\omega}$ for fulfilling additional tasks of interest. For example, analogously to the point feature case, one could try to keep the barycenter at a constant position $(x_g, y_g) \simeq const$ in order to mitigate the effects of $\dot{\boldsymbol{x}}_m$ in (25). Furthermore, it is interesting to note that matrix $\boldsymbol{\Omega}$ (and, therefore, matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ as well) loses rank whenever $v_z = 0$: in order to meet condition (7), that

is, to keep $\sigma_1^2(t) > 0$, the camera then necessarily needs to translate with a non-zero component along the optical axis[4].

Finally, differently from the point feature case, an explicit expression for the SVD decomposition of $\boldsymbol{\Omega}$ is not easily obtainable, nor special insights can be gained from the inspection of the Jacobian matrices $\boldsymbol{J}_{u,i}$, $i = 1 \dots 3$, apart from the special structure $\boldsymbol{J}_{u,i} = [\boldsymbol{J}_{\boldsymbol{v},i} \;\; \mathbf{0}]$ as in (28).

V. SIMULATION RESULTS

A. Point feature

The reported simulation results are meant to show the benefits of employing the proposed active estimation strategy for generating an 'optimized' camera motion w.r.t. a 'generic' one. Since, as explained, the norm $\|\boldsymbol{\Omega}\boldsymbol{\Omega}^T\|$ (and thus the estimator convergence speed) is related to $\|\boldsymbol{u}\|$ (in particular to $\|\boldsymbol{v}\|$), in order to obtain a fair comparison we consider two simulations in which $\|\boldsymbol{v}\|$ is kept constant over time but, in one case, $\sigma_1^2(t)$ is actively maximized (case I) while in the other case it is left unaffected (case II). The angular velocity input $\boldsymbol{\omega}$ is instead exploited to keep $\boldsymbol{x}_m = (x, y) \simeq \mathbf{0}$ over time (point feature kept at the center of the image plane).

Let then $\kappa = \frac{1}{2}\boldsymbol{v}^T\boldsymbol{v} = \frac{1}{2}\|\boldsymbol{v}\|^2$ and $\kappa_{des} > 0$ represent a desired value for $\kappa(t)$. Being obviously $\dot{\kappa} = \boldsymbol{v}^T\dot{\boldsymbol{v}}$, the following simple law allows to obtain the sought result:

$$\dot{\boldsymbol{v}} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|^2} k_1 \left(\kappa_{des} - \kappa\right) + k_2 \left(\boldsymbol{I} - \frac{\boldsymbol{v}\boldsymbol{v}^T}{\|\boldsymbol{v}\|^2}\right) \boldsymbol{J}_{v,1}^T \qquad (31)$$

with $k_1 > 0$, $k_2 \geq 0$. In fact, the first term in (31) (asymptotically) enforces the constraint $\kappa = \kappa_{des}$ while the second term allows to implement either case I ($k_2 > 0$) or case II ($k_2 = 0$) within the null-space of the first constraint. Note that, when $(x, y) = (0, 0)$, it is $\sigma_1^2 \leq \|\boldsymbol{v}\|^2$ and $\sigma_1^2 = \|\boldsymbol{v}\|^2$ iff $v_z = 0$. Therefore, the maximum achievable value for σ_1^2 is given by the squared norm of the camera linear velocity $\boldsymbol{v}$.

Figures 1(a–d) show the results of the two simulations run with the following conditions: $\alpha\beta = 2000$, $\boldsymbol{\xi}(t_0) = \mathbf{0}$, $z(t_0) = 4$, $\boldsymbol{v}(t_0) = [0.03 \; 0 \; 0.04]^T$, $k_1 = 10$, and $k_2 = 10$ (for case I). In particular, Figs. 1(a–b) report the behavior of $\sigma_1^2(t)$ for the two cases: one can note that in case II (Fig. 1(b)) the obtained value of $\sigma_1^2(t) \simeq 9 \cdot 10^{-4}$ results much smaller than in case I (Fig. 1(a)) where, thanks to the optimization action, the theoretical maximum $\sigma_{1,max}^2 = \|\boldsymbol{v}\|^2 = 2.5 \cdot 10^{-3}$ is (correctly) reached.

Figure 1(c) shows the behavior of $z(t)$ for the two simulations (red line – case I, blue line – case II) together with the 'ideal' response of a critically-damped second-order system in the form (21), i.e., $\ddot{\eta} + 2\sigma_{1,des}\dot{\eta} + \sigma_{1,des}^2\eta = 0$ with $\eta(t_0) = z(t_0)$, $\dot{\eta}(t_0) = 0$ (black dashed line). We can then note the following: while $\kappa(t) = \kappa_d$ for both cases (thus, same control effort), as expected the convergence rate of $z(t)$ for case I results faster than for case II as $\sigma_1^2(t)$ attains a higher value over time ($2.5 \cdot 10^{-3}$ vs. $9 \cdot 10^{-4}$). Furthermore, the system response for case I perfectly matches that of the 'ideal' second order system (21), thus confirming the soundness of the transient analysis performed in Sect. III. Finally, Fig. 1(d) depicts the 3D Cartesian trajectory followed by the camera in both cases, with the 'arrow' representing the camera optical axis: since the angular velocity input $\boldsymbol{\omega}$ was exploited to keep $\boldsymbol{x}_m \simeq \mathbf{0}$, the camera keeps looking at the feature point (red circle) while following a circular path.

[4]A requirement not present in the point-feature case where any non-zero linear velocity not aligned with the projection ray could guarantee fulfillment of (7).

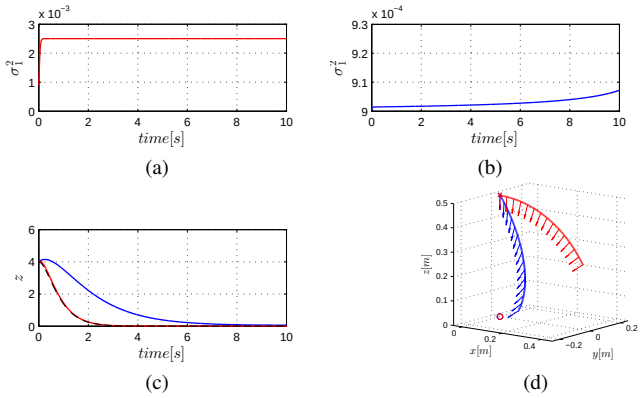


Fig. 1: Simulation results for the point feature case. Figs. (a–b): behavior of $\sigma_1^2(t)$ for case I and case II. Note how in case I (Fig. (a)) $\sigma_1^2(t)$ correctly reaches its maximum value $\sigma_{1,max}^2 = \|\boldsymbol{v}\|^2 = 2.5 \cdot 10^{-3}$. Fig. (c): behavior of $z(t)$ for case I (red line), case II (blue line), and response of an 'ideal' second-order system with the desired poles (dashed black line). Fig. (d): 3D Cartesian trajectory followed by the camera for case I (red line) and case II (blue line)

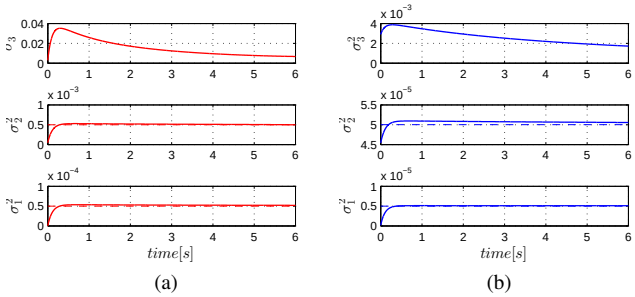


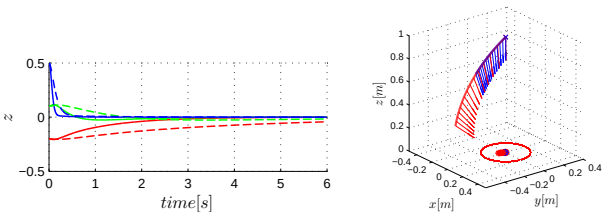
Fig. 2: Simulation results for the moments case. The two pictures depict the behavior of the three eigenvalues of matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ in reaching the desired values $\boldsymbol{\sigma}_{des}$ (dashed lines) for case I (left) and case II (right)

B. Moments

We now show the results of the active estimation framework applied to the case of image moments discussed in Sect. IV-B. As explained, in this case the problem is 'square': 3 available measurements (area and barycenter) for 3 unknowns (vector $\boldsymbol{\chi}$). Let $\boldsymbol{J} = [\boldsymbol{J}_{\boldsymbol{v},1}^T \; \boldsymbol{J}_{\boldsymbol{v},2}^T]^T \in \mathbb{R}^{2\times 3}$ be the Jacobian associated to the first two eigenvalues (σ_1^2, σ_2^2) of matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ from (30). Let also $\boldsymbol{\sigma} = [\sigma_1^2 \; \sigma_2^2]^T \in \mathbb{R}^2$ and $\boldsymbol{\sigma}_{des}$ be a desired value for $\boldsymbol{\sigma}(t)$. In this case, the following law was implemented

$$\dot{\boldsymbol{v}} = k_1 \boldsymbol{J}^{\dagger} \left(\boldsymbol{\sigma}_{des} - \boldsymbol{\sigma}\right) - k_2 (\boldsymbol{I} - \boldsymbol{J}^{\dagger}\boldsymbol{J})\boldsymbol{v}, \quad k_1 > 0, \, k_2 > 0,$$

thus aiming at regulating the first two smallest eigenvalues of $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ while trying to minimize, as much as possible, the norm of $\boldsymbol{v}$. We note that, in principle, one could have



(a) (b)

Fig. 3: Simulation results for the moments case. Fig. (a): behavior of $\boldsymbol{z}(t)$ for case I (solid lines) and case II (dashed lines). Note how the convergence rate of case I is faster than case II, as expected due to the larger values attained by the eigenvalues of matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$. Fig. (b): 3D Cartesian trajectory followed by the camera for case I (red line) and case II (blue line) together with the planar circle used to generate image moments. In case I, the camera travels a longer distance because of the higher value of $\boldsymbol{\sigma}$, that is, larger $\|\boldsymbol{v}\|$

attempted to regulate *all* the three eigenvalues of matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ to some desired value at the same time: however, we found this possibility to be ill-conditioned in practice, with the associated Jacobian matrix being always almost singular, while the choice of only regulating the first two eigenvalues did always result well-conditioned. The implications of these considerations will be addressed in future studies.

For this simulation, we also exploited the camera angular velocity $\boldsymbol{\omega}$ in order to keep the observed barycenter $(x_g,\, y_g)$ stationary during motion. As explained at the end of Sect. III-C, this was meant to (partially) mitigate the effects of a non-zero $\dot{\boldsymbol{x}}_m$ when inverting (25).

Two simulations were run with $\boldsymbol{\sigma}_{des} = [5{\cdot}10^{-5}\ 5{\cdot}10^{-4}]^T$ (case I) and $\boldsymbol{\sigma}_{des} = [5{\cdot}10^{-6}\ 5{\cdot}10^{-5}]^T$ (case II, thus 10 times smaller than case I). These values were employed: $\alpha\beta = 5 \cdot 10^4$, $\boldsymbol{\xi}(t_0) = \boldsymbol{0}$, $\boldsymbol{z}(t_0) = [0.5\ \ -0.2\ \ 0.1]^T$, $\boldsymbol{v}(t_0) = [-0.02\ \ 0.05\ \ 0.01]^T$, $k_1 = 10$, and $k_2 = 10$. The moments were generated from a planar circle of radius $R = 0.2$.

Figures 2(a–b) depict the behavior of the three eigenvalues $(\sigma_1^2(t),\, \sigma_2^2(t),\, \sigma_3^2(t))$ over time for the two cases (solid lines), together with the corresponding desired values $\boldsymbol{\sigma}_{des}$ (dashed lines). We can then note how the first two eigenvalues $(\sigma_1^2(t),\, \sigma_2^2(t))$ approach their desired values with a (small but expected) non-zero steady-state error due to the non-perfect compensation of $\dot{\boldsymbol{x}}_m$. Figure 3(a) reports the behavior of $\boldsymbol{z}(t)$ for the two cases (solid lines – case I, dashed lines – case II). One can then note how (i) the convergence rate for case I is significantly faster than case II due to the larger values attained by $(\sigma_1^2(t),\, \sigma_2^2(t))$, and how (ii) in both cases the transient response does not present significant oscillatory modes thanks to the proper shaping of the damping matrix $\tilde{\boldsymbol{H}}$. This then confirms the validity of our analysis also in the more complex (non-scalar) case of the image moments. Finally, Fig. 3(b) shows the 3D trajectory followed by the camera for case I (red line) and case II (blue line): one can note how (i) the camera keeps looking at the barycenter as expected, and how (ii) over the same amount of time, in case I the camera travels a longer distance compared to case II because of the higher values attained by $\boldsymbol{\sigma}(t)$, that is, larger $\|\boldsymbol{v}\|$.

VI. CONCLUSIONS

In this paper we have addressed the problem of actively tuning the transient response of a nonlinear estimation scheme by suitably shaping the system inputs and internal gains. In particular, we were able to impose an estimation error response close to that of a 'reference second-order linear system' with given poles. Besides its general validity, the proposed theory was also applied to two real-world case studies involving SfM problems: estimation of the depth of a feature point, and estimation of the normal and distance to a planar scene from image moments. The results confirmed the soundness of the various claims in the paper.

We are currently working towards a detailed experimental validation of this framework in the context of visual-based manipulation of unknown objects. For the interested reader, an experimental comparison of the proposed theory against a EKF estimation scheme in the context of flying robots equipped with onboard cameras can also be found in [15].

REFERENCES

[1] A. De Luca, G. Oriolo, and P. Robuffo Giordano, "Feature depth observation for image-based visual servoing: Theory and experiments," *Int. Journal of Robotics Research*, vol. 27, no. 10, pp. 1093–1116, 2008.

[2] P. Robuffo Giordano, A. De Luca, and G. Oriolo, "3D structure identification from image moments," in *2008 IEEE Int. Conf. on Robotics and Automation*, Pasadena, CA, may 2008, pp. 93–100.

[3] F. Morbidi, G. L. Mariottini, and D. Prattichizzo, "Observer design via immersion and invariance for vision-based leader-follower formation control," *Automatica*, vol. 46, no. 1, pp. 148–154, 2010.

[4] M. Sassano, D. Carnevale, and A. Astolfi, "Observer design for range and orientation identification," *Automatica*, vol. 46, no. 8, pp. 1369–1375, 2010.

[5] A. P. Dani, N. R. Fischer, and W. E. Dixon, "Single Camera Structure and Motion," *IEEE Trans. on Automatic Control*, vol. 57, no. 1, pp. 238–243, 2012.

[6] R. Mahony, T. Hamel, and J.-M. Pflimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Trans. on Automatic Control*, vol. 53, no. 5, pp. 1203–1218, 2008.

[7] R. Mahony, T. Hamel, P. Morin, and E. Malis, "Nonlinear complementary filters on the special linear group," *International Journal of Control*, vol. 85, no. 10, pp. 1557–1573, 2012.

[8] A. Martinelli, "Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination," *IEEE Trans. on Robotics*, vol. 28, no. 1, pp. 44–60, 2012.

[9] G. Besançon and H. Hammouri, "On uniform observation of nonuniformity observable systems," *Systems & Control Letters*, vol. 29, no. 1, pp. 9–19, 1996.

[10] R. Marino and P. Tomei, *Nonlinear Control Design: Geometric, Adaptive and Robust*. Prentice Hall, 1995.

[11] T. Papadopoulo and M. I. A. Lourakis, "Estimating the Jacobian of the Singular Value Decomposition: Theory and Applications," in *Proc. European Conf. on Computer Vision, ECCV00*. Springer, 2000, pp. 554–570.

[12] P. Yang, R. A. Freeman, G. J. Gordon, K. M. Lynch, S. S. Srinivasa, and R. Sukthankar, "Decentralized estimation and control of graph connectivity for mobile sensor networks," *Automatica*, vol. 46, no. 2, pp. 390–396, 2010.

[13] F. Chaumette and S. Hutchinson, "Visual servo control, Part I: Basic approaches," *IEEE Robotics and Automation Magazine*, vol. 13, no. 4, pp. 82–90, 2006.

[14] F. Chaumette, "Image moments: a general and useful set of features for visual servoing," *IEEE Trans. on Robotics*, vol. 20, no. 4, pp. 713–723, 2004.

[15] V. Grabe, H. H. Bülthoff, and P. Robuffo Giordano, "A Comparison of Scale Estimation Schemes for a Quadrotor UAV based on Optical Flow and IMU Measurements," in *2013 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2013.