Second-Order-Optimal Minimum-Energy Filters on Lie groups

Alessandro Saccon, Jochen Trumpf, Robert Mahony, A. Pedro Aguiar

***Abstract*—Systems on Lie groups naturally appear as models for physical systems with full symmetry. We consider the state estimation problem for such systems where both input and output measurements are corrupted by unknown disturbances. We provide an explicit formula for the second-order-optimal nonlinear filter on a general Lie group where optimality is with respect to a deterministic cost measuring the cumulative energy in the unknown system disturbances (minimum-energy filtering). The resulting filter depends on the choice of affine connection which encodes the nonlinear geometry of the state space. As an example, we look at attitude estimation, where we are given a second order mechanical system on the tangent bundle of the special orthogonal group SO(3), namely the rigid body kinematics together with the Euler equation. When we choose the symmetric Cartan-Schouten (0)-connection, the resulting filter has the familiar form of a gradient observer combined with a perturbed matrix Riccati differential equation that updates the filter gain. This example demonstrates how to construct a matrix representation of the abstract general filter formula.**

I. Introduction

Arguably the most prominent approach to state estimation from disturbed measurements of inputs and outputs is via stochastic system models, where the disturbances are modelled as stochastic processes and optimal or sub-optimal solutions are sought that minimize some measure of expected error. The resulting algorithms range from the famous Kalman filter and its various nonlinear generalizations such as the Extended Kalman Filter (EKF) or the Unscented Filter (UF), to particle filters (PF) and other more specialized approximation schemes. An alternative approach to state estimation treats the disturbances as unknown deterministic signals and seeks to optimize some measure of size or "badness" of these signals. The most prominent techniques in the latter domain are H^∞-filtering and minimum-energy filtering, the topic of this paper.

Minimum-energy filtering was first proposed by Mortensen [1] and further developed by Hijab [2]. It is known that the minimum energy-filter for linear systems with quadratic cost coincides with the Kalman-Bucy filter [3]. Krener [4] proved exponential convergence of minimum-energy estimators for uniformly observable systems in $\mathbb{R}^n$. Ongoing research in the area is aimed at generalizing minimum-energy filters to systems whose state evolves on a differentiable manifold such as a Lie group. Aguiar and Hespanha [5] provided a minimum-energy estimator for systems with perspective outputs that can be used for pose estimation, a problem with state space SE(3), the special Euclidean group. Their approach uses an embedding of SE(3) in a linear matrix space and is hence not intrinsic with respect to the geometry of the state space. This means that filter estimates need to be projected back onto SE(3), potentially resulting in suboptimal performance of the filter. Coote et al. [6] derived a near-optimal minimum-energy filter for a system on the unit circle and provided an estimate for the distance to optimality, a result generalized by Zamani et al. to systems on the special orthogonal group SO(3) with full state measurements [7] or vectorial measurements [8]. In the latter case, the resulting filter can be interpreted as a geometric correction to the Multiplicative Extended Kalman Filter (MEKF) and shows asymptotically optimal performance in simulation [9], [10].

In this paper we provide an explicit formula for a second-order optimal minimum-energy filter for systems on general Lie groups with vectorial outputs. It is an established fact that the exact minimum-energy filter is typically an infinite dimensional filter and therefore a truncation of the filter is necessary for practical implementability (the linear dynamics case with quadratic cost being the exception). The proposed second-order-optimal minimum-enery filter is a truncation of the exact minimum-energy filter when third order terms of the associated value function are neglected: in this sense, it is *second order optimal*. Details of this truncation are provided in the derivation of the filter formula. The resulting filter takes the form of a gradient observer coupled with a perturbed operator Riccati differential equation that updates the filter gain. The filter explicitly depends on the choice of affine connection on the state space which encodes its nonlinear geometry. In the GAME filter presented in [8], a special case of the filter derived in this work, the connection is implicitly selected by the choice of the Riemannian metric employed on SO(3). In this work, we show that there is no need to have a Riemannian metric on the group to derive the filter, but it suffices to have an affine connection. As a by-product, the resulting (modified) Riccati differential equation lives now in the cotangent space (better, the dual of the Lie algebra) of the group, rather than in the tangent space as it was in [8].

We provide a worked example, applying the developed theory to the case of attitude estimation given a second order (dynamic) system model on the tangent bundle of the special orthogonal group SO(3) and vectorial measurements. The gain equation specializes to a perturbed matrix Riccati differential equation in this case. We choose the usual symmetric Cartan-Schouten (0)-connection and the Cartan-Schouten (-)-connection on SO(3) for illustration, but different choices would be possible, resulting in different gain equations. To the best of our knowledge, this is the first such filter published for a (second-order) mechanical system, except for the conference version of this paper [11].

Besides the numerical results presented in [8] and [10], another successful application of the Lie group minimum enegy filter presented in this work is reported in [12]. There, the authors are interested in obtaining accurate camera motion estimation to be used as input data for computer vision algorithms and it is concluded that the second-order optimal minimum energy filter on Lie groups improves rotational velocity estimation and otherwise is on par with the state-of-the-art.

This paper is divided in six sections, including this introduction and the conclusion section. Mathematical preliminaries are given in Section II. In Section III, we formulate the problem of minimum-energy filtering for systems on Lie groups. The explicit expression for the second-order-optimal filter, the filter that agrees up to second order terms with the optimal minimum-energy filter, and its derivation are detailed in Section IV. A worked example is discussed in Section V.

A. Saccon is with the Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands. a.saccon@tue.nl

J. Trumpf and R. Mahony are with the Research School of Engineering, Australian National University, Canberra, Australia. (Jochen.Trumpf|Robert.Mahony)@anu.edu.au

A. Pedro Aguiar is with the Faculty of Engineering, University of Porto (FEUP), Portugal. pedro.aguiar@fe.up.pt

Research partially supported by Australian Research Council grant DP120100316 "Geometric observer theory for mechanical control systems" and by projects CONAV/FCT-PTDC/EEACRO/113820/2009 and MORPH [EU FP7 ICT 288704].

II. Notation and Mathematical Preliminaries

We begin by establishing the notation used throughout this paper. The basic notation and methodology is fairly standard within the differential geometry literature and we have attempted to use traditional symbols and definitions wherever feasible. We refer the reader to the books [13], [14] and [15] for a review of differentiable manifolds and covariant differentiation and to [16], [17] and [18] for a review of the theory of Lie groups and Lie algebras. Many of these topics are also covered in the system theory literature, specifically [19], [20] and [21]. The following symbols will be used frequently:

G	a connected Lie group
n	the dimension of the group G
g, h	elements of G
$\mathfrak{g}$	the Lie algebra associated with G
X, Y	elements of the Lie algebra $\mathfrak{g}$
$[\cdot,\cdot]$	the Lie bracket of $\mathfrak{g}$
$\mathfrak{g}^*$	the dual of the Lie algebra $\mathfrak{g}$
μ	an element of $\mathfrak{g}^*$
$L_g \colon G \to G$	left translation $L_g h = gh$
$T_h L_g$	the tangent map of L_g at $h \in G$
gX	shorthand for $T_e L_g(X) \in T_g G$
$\langle \cdot, \cdot \rangle$	duality paring $\langle \mu, X \rangle = \mu(X)$
V	finite dimensional vector space
$f \colon G \to V$	differentiable map
$\mathbf{d}\, f(g)$	differential of f at g, $\mathbf{d}\, f(g) \colon T_g G \to V$ identifying $T_{f(g)} V$ with V
$\mathbf{d}_1, \mathbf{d}_2, \ldots$	differentials with respect to individual arguments of a multiple argument map
$\nabla_{\mathbf{X}} \mathbf{Y}$	covariant derivative ($\mathbf{X}$ and $\mathbf{Y}$ are vector fields on G)
$\omega \colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$	connection function associated with ∇
$\omega_X \colon \mathfrak{g} \to \mathfrak{g}$	$\omega_X(Y) = \omega(X, Y)$
$\omega_Y^{\leftrightharpoons} \colon \mathfrak{g} \to \mathfrak{g}$	$\omega_Y^{\leftrightharpoons}(X) = \omega_X(Y)$
$\omega_X^* \colon \mathfrak{g}^* \to \mathfrak{g}^*$	$\langle \omega_X^*(\mu), Y \rangle = \langle \mu, \omega_X\, Y \rangle$
$\omega_\mu^{*\leftrightharpoons} \colon \mathfrak{g} \to \mathfrak{g}^*$	$\langle \omega_\mu^{*\leftrightharpoons}(X), Y \rangle = \langle \omega_X^*(\mu), Y \rangle = \langle \mu, \omega_X\, Y \rangle$
$\omega_Y^{\leftrightharpoons *} \colon \mathfrak{g}^* \to \mathfrak{g}^*$	$\langle \omega_Y^{\leftrightharpoons *}(\mu), X \rangle = \langle \mu, \omega_Y^{\leftrightharpoons}\, X \rangle = \langle \mu, \omega_X\, Y \rangle$
$T(X, Y) \in \mathfrak{g}$	torsion function associated with ω
$T_X \colon \mathfrak{g} \to \mathfrak{g}$	partial torsion function $T_X Y = T(X, Y)$
$\operatorname{Hess} f(g)$	Hessian operator of a twice differentiable function $f \colon G \to \mathbb{R}$ (or a map $f \colon G \to V$)
$(\phi)^W \colon \mathfrak{L}(W, U) \to \mathfrak{L}(W, V)$	Exponential functor $(\cdot)^W$ applied to a linear map $\phi \colon U \to V$

Dual and symmetric maps. We will use the canonical identification of the Lie algebra $\mathfrak{g}$ with its bidual $\mathfrak{g}^{**}$ allowing us to treat the dual $\phi^* \colon \mathfrak{g}^{**} \to \mathfrak{g}^*$ of a linear map $\phi \colon \mathfrak{g} \to \mathfrak{g}^*$ again as a map $\phi^* \colon \mathfrak{g} \to \mathfrak{g}^*$. We can hence call ϕ *symmetric* (with respect to the duality pairing) if $\phi = \phi^*$. This idea extends to arbitrary linear maps between a (finite-dimensional) vector space and its dual, for example the Hessian operator defined below.

Connection function. A left-invariant affine connection ∇ on G is fully characterized by its bilinear connection function $\omega \colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ through the identity $\nabla_{gX}(gY) = g\,\omega(X, Y)$ [22, Theorem 8.1]. For further reading on invariant connections on Lie groups, we refer the interested reader to the notation and definitions subsection in [23, Section II.A] and references therein.

Swap operator. The connection function ω allows us to introduce a convenient operator calculus that we will use extensively in the derivation of our filter. Other than the partial connection functions ω_X, $\omega_Y^{\leftrightharpoons}$, ω_X^*, $\omega_\mu^{*\leftrightharpoons}$, and $\omega_Y^{\leftrightharpoons *}$ defined in the notation table above, thinking of the 'swap' $\leftrightharpoons$ and 'dual' $*$ operations as formal operations, we can close off the calculus with the two additional operators $\omega^{*\leftrightharpoons *}$ and $\omega^{\leftrightharpoons * \leftrightharpoons}$, defined in the obvious way, which turn out to be equal. This yields identities like $\omega^{*\leftrightharpoons * \leftrightharpoons} = \omega^{\leftrightharpoons *}$. We will use this latter identity at one point in the filter derivation.

Hessian operator. Given a twice differentiable function $f \colon G \to \mathbb{R}$ we can define the Hessian operator $\operatorname{Hess} f(g) \colon T_g G \to T_g^* G$ at a point $g \in G$ by $\operatorname{Hess} f(g)(gX)(gY) = \mathbf{d}\big(\mathbf{d}\, f(g)(gY)\big)(gX) - \mathbf{d}\, f(g)\big(\nabla_{gX}(gY)\big)$ for all $gX, gY \in T_g G$ [24]. Here, $\mathbf{d}\big(\mathbf{d}\, f(g)(gY)\big) \colon T_g G \to \mathbb{R}$ is shorthand for the differential of the function $g \mapsto \mathbf{d}\, f(g)(gY)$ at the point $g \in G$.

The dual Hessian operator is also a map $\big(\operatorname{Hess} f(g)\big)^* \colon T_g G \to T_g^* G$ since we identify the bidual $T_g^{**} G$ with $T_g G$. Note that the Hessian operator is not always symmetric (in the sense defined above). It is, however, symmetric at any critical point of the function f since $\mathbf{d}\, f(g) = 0$ causes the second, potentially non-symmetric term in the definition of the Hessian operator to vanish. This term, and hence the Hessian operator, is always symmetric if the connection ∇ is symmetric [24].

The concept of a Hessian operator naturally extends to vector-valued twice differentiable maps $f \colon G \to V$. The Hessian operator at a point $g \in G$ is then a map $\operatorname{Hess} f(g) \colon T_g G \to \mathfrak{L}(T_g G, V)$, where $\mathfrak{L}(T_g G, V)$ denotes the set of linear maps from $T_g G$ to V. The Hessian operator is defined component-wise with respect to a basis in V [24]. It is easy to check that the resulting operator is independent of the choice of basis.

Exponential functor. Given a linear map $\phi \colon U \to V$ and a third vector space W, the exponential functor $(\cdot)^W$ lifts the map ϕ to the linear map $\phi^W \colon \mathfrak{L}(W, U) \to \mathfrak{L}(W, V)$ defined by $\phi^W(\xi) = \phi \circ \xi$.

III. Problem Formulation

The problem of minimum-energy state estimation for systems on Lie groups is as follows. Consider the deterministic system on a Lie group G defined by

$$\dot{g}(t) = g(t)\big(\lambda(g(t), u(t), t) + B\delta(t)\big), \qquad g(t_0) = g_0 \tag{1}$$

with state $g(t) \in G$, input $u(t) \in \mathbb{R}^m$ a known exogenous signal, nominal (left-trivialized) dynamics $\lambda \colon G \times \mathbb{R}^m \times \mathbb{R} \to \mathfrak{g}$, and unknown *model error* $\delta(t) \in \mathbb{R}^d$. The known map $B : \mathbb{R}^d \to \mathfrak{g}$ is linear and $g_0 \in G$, the initial condition at the initial time $t_0 \in \mathbb{R}$, is unknown. After a choice of basis for $\mathfrak{g}$, the model error space $\mathbb{R}^d$ can be taken to be the Lie algebra $\mathfrak{g}$ or, alternatively, a vector space of smaller dimension. This latter case will be illustrated in Section V.

The two main applications that motivate our work are the kinematics and dynamics of mechanical systems. In the case where just the kinematics of a mechanical system are considered, the left-trivialized dynamics are simply the system velocity leading to the classical nominal left-invariant kinematics $\dot{g} = g(u + B\delta)$ [25], [26]. In this case the system 'input' u is the measured velocity and the model error δ is best thought of as measurement error associated with inexact measurement of the physical velocity $u + B\delta$.

In the case of a (dynamic) mechanical system then $G = TC$ is the tangent bundle of a smaller Lie-group C that is a representation of the configuration space of the mechanical system [20], [21]. In this case the model error δ is an additive term that includes unmodeled dynamics as well as acceleration measurement error and only applies to the dynamics that model the evolution of the velocity of the system and not to the kinematics. This property can be incorporated into (1) by suitable choice of the linear operator B and the dimension d of the model error space $\mathbb{R}^d$. Section V provides an example of the second case while the first case has been considered in a number of prior works including [8], [10].

The known measurement output, denoted by $y \in \mathbb{R}^p$, is related to the state g through the nominal output map $h\colon G \times \mathbb{R} \to \mathbb{R}^p$ as

$$y(t) = h(g(t), t) + D\varepsilon(t) \quad (2)$$

where $\varepsilon \in \mathbb{R}^p$ is the unknown *measurement error* and $D\colon \mathbb{R}^p \to \mathbb{R}^p$ is an invertible linear map.

In the minimum energy filtering approach, both the 'error' signals, δ and ε, are modeled as unknown deterministic functions of time. Along with the unknown initial condition g_0 these three signals are the unknowns in the filtering problem. Given measurements $y(\tau)$ and inputs $u(\tau)$ taken over a period $\tau \in [t_0, t]$ then there are only certain possible unknown signals $(\delta(\tau), \varepsilon(\tau), g_0)$ for $\tau \in [t_0, t]$ that are compatible with (1) and (2). Each triple of compatible unknown signals corresponds to a separate state trajectory $g(\tau)$. The principle of minimum energy filtering is that the 'best' estimate of the state is the trajectory induced by the set of unknown signals $(\delta, \varepsilon, g_0)$ that are 'smallest' in a specific sense. To quantify the concept of small it is necessary to introduce a cost functional, typically a measure of energy in the unknown error signals δ and ε, along with some form of initial cost (initial "energy") in g_0, leading to the terminology of minimum energy filtering.

Define two quadratic forms

$$\mathcal{R}\colon \mathbb{R}^d \to \mathbb{R}, \quad \mathcal{Q}\colon \mathbb{R}^p \to \mathbb{R} \quad (3)$$

that measure instantaneous energy $\mathcal{R}(\delta(\tau))$ and $\mathcal{Q}(\varepsilon(\tau))$ of the error signals. Let $\alpha \geq 0$ be a non-negative scalar and define an incremental cost $l\colon \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by

$$l(\delta, \varepsilon, t, \tau) := 1/2\, e^{-\alpha(t-\tau)}\big(\mathcal{R}(\delta) + \mathcal{Q}(\varepsilon)\big). \quad (4)$$

The constant α is the *discount rate*, the rate at which old information in the incremental cost is discounted and forgotten. In addition, we introduce a cost $m\colon G \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ on the initial condition g_0.

$$m(g_0, t, t_0) := 1/2\, e^{-\alpha(t-t_0)} m_0(g_0), \quad (5)$$

where $m_0\colon G \to \mathbb{R}$ is a bounded smooth function with a unique global minimum on G. The initial cost m_0 can be thought of as encoding the *a-priori* information about the state at time t_0. It provides a boundary condition for the Hamilton-Jacobi-Bellman equation used in the derivation of the minimum energy filter in Section IV.

The cost functional that we consider is

$$J(\delta, \varepsilon, g_0; t, t_0) := m(g_0, t, t_0) + \int_{t_0}^{t} l(\delta(\tau), \varepsilon(\tau), t, \tau)\mathrm{d}\tau. \quad (6)$$

Note that J depends on the values of the signals δ and ε on the whole time interval $[t_0, t]$. In order that the cost functional is well defined we will assume that all error signals considered are square integrable.

Consider a time-interval $[t_0, t]$ and let $\widehat{g}(t)$ denote the filter estimate, at the terminal time t, for the minimum energy filter. That is $\widehat{g}(t) := g^*_{[t_0,t]}(t)$ is the final value of the state trajectory $g^*_{[t_0,t]}$ that is associated with the signals $(\delta^*, \varepsilon^*, g_0^*)$ that minimize the cost functional (6) on $[t_0, t]$ and are compatible with (1) and (2) for given measurements $y(\tau)$ and inputs $u(\tau)$, $\tau \in [t_0, t]$. Note that this correspondence of $\widehat{g}(t)$ and $g^*_{[t_0,t]}(t)$ will only necessarily hold at the terminal condition, and indeed, in general $g^*_{[t_0,t]}(\tau) \neq \widehat{g}(\tau)$ for $\tau \neq t$. The minimum energy filter can only be posed on the whole interval $[t_0, t]$ since this is the domain of definition of the cost functional. Nevertheless, it is not necessary to resolve the whole optimization problem for each new time t since the Hamilton-Jacobi-Bellman (HJB) equation provides a model for the evolution of the solution of the filter equation, in terms of the value function associated to the cost functional, with changing terminal condition. Finding a suitable solution to the HJB equation is known to be difficult, and indeed expected to yield an infinite dimensional evolution equation for the value function. In the remainder of the paper we go on to show how a second order approximation to the filter equation can be derived by using Mortensen's approach to approximating the Taylor expansion of the value function at the terminal condition of the filter [1]. *Taking a second order approximation of the value function yields what we term the second-order-optimal minimum-energy filter equation.*

In order to write down the filter equation it is necessary to associate gain operators with the quadratic forms $\mathcal{R}$ and $\mathcal{Q}$ that appear in the incremental cost (4). Let $\bar{R}\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ resp. $\bar{Q}\colon \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ be the (unique) symmetric positive definite bilinear forms associated with $\mathcal{R}$ resp. $\mathcal{Q}$, i.e. $\bar{R}(\delta, \delta) = \mathcal{R}(\delta)$ for all $\delta \in \mathbb{R}^d$ and $\bar{Q}(\varepsilon, \varepsilon) = \mathcal{Q}(\varepsilon)$ for all $\varepsilon \in \mathbb{R}^p$. The duality pairing $\langle \cdot, \cdot \rangle$ can then be used to uniquely define symmetric positive definite linear maps $R\colon \mathbb{R}^d \to (\mathbb{R}^d)^*$ and $Q\colon \mathbb{R}^p \to (\mathbb{R}^p)^*$ from the bilinear forms $\bar{R}$ and $\bar{Q}$ by

$$\langle R(X_1), X_2 \rangle = \bar{R}(X_1, X_2), \ \langle Q(y_1), y_2 \rangle = \bar{Q}(y_1, y_2) \quad (7)$$

for all $X_1, X_2 \in \mathbb{R}^d$ and $y_1, y_2 \in \mathbb{R}^p$, respectively.

IV. The filter and its derivation

This section presents the second-order-optimal filter and it details how to obtain it, revisiting the optimal estimation problem, its corresponding optimal Hamiltonian and the associated Hamilton-Jacobi-Bellman (HJB) equation. We then show how to compute a second-order Taylor expansion of the value function that provides the filter equation.

A. The second-order-optimal filter equation

Assume that the error terms δ and ε are square integrable deterministic functions of time and adopt the shorthand notation $h_t(g)$ and $\lambda_t(g, u)$ for $h(g(t), t)$ and $\lambda(g(t), u(t), t)$, respectively. For ease of presentation, we drop the explicit dependence on time of the input, output, state, and error signals from our notation. The following theorem is the main result of this paper.

Theorem 4.1: Consider the system defined by (1) and (2) along with the energy cost functional (6) with incremental cost (4) and initial cost (5). The second-order-optimal minimum-energy filter in the sense descibed in Section III is given by

$$\widehat{g}^{-1}\dot{\widehat{g}} = \lambda_t(\widehat{g}, u) + K(t)\, r_t(\widehat{g}), \qquad \widehat{g}(t_0) = \widehat{g}_0, \quad (8)$$

where $K(t)\colon \mathfrak{g}^* \to \mathfrak{g}$ is a time-varying linear map satisfying the perturbed operator Riccati equation (11) given below,

$$\widehat{g}_0 = \operatorname*{arg\,min}_{g \in G} m_0(g), \quad (9)$$

and the residual $r_t(\widehat{g}) \in \mathfrak{g}^*$ is given by

$$r_t(\widehat{g}) = T_e L^*_{\widehat{g}}\Big[\Big(\big(D^{-1}\big)^* \circ Q \circ D^{-1}\big(y - h_t(\widehat{g})\big)\Big) \circ \mathbf{d}\, h_t(\widehat{g})\Big]. \quad (10)$$

The second-order-optimal symmetric gain operator $K(t)\colon \mathfrak{g}^* \to \mathfrak{g}$ satisfies the perturbed operator Riccati equation

$$\begin{aligned}\dot{K} = & -\alpha \cdot K + A \circ K + K \circ A^* - K \circ E \circ K + B \circ R^{-1} \circ B^* \\ & - \omega_{Kr} \circ K - K \circ \omega^*_{Kr}, \end{aligned} \quad (11)$$

with initial condition $K(t_0) = X_0^{-1}$ where the operators $X_0\colon \mathfrak{g} \to \mathfrak{g}^*$, $A(t)\colon \mathfrak{g} \to \mathfrak{g}$, and $E(t)\colon \mathfrak{g} \to \mathfrak{g}^*$ are given by

$$X_0 = T_e L^*_{\hat{g}_0} \circ \operatorname{Hess}\ m_0(\hat{g}_0) \circ T_e L_{\hat{g}_0}, \quad (12)$$

$$A(t) = \mathbf{d}_1\, \lambda_t(\hat{g}, u) \circ T_e L_{\hat{g}} - \operatorname{ad}_{\lambda_t(\hat{g},u)} - T_{\lambda_t(\hat{g},u)}, \quad (13)$$

and

$$E(t) = -T_e L_{\hat g}^* \circ \Big[\Big(\big(D^{-1}\big)^* \circ Q \circ D^{-1}\big(y - h_t(\hat g)\big) \Big)^{T_{\hat g} G} \circ \tag{14}$$
$$\operatorname{Hess} h_t(\hat g) - \big(\mathbf{d}\, h_t(\hat g)\big)^* \circ \big(D^{-1}\big)^* \circ Q \circ D^{-1} \circ \mathbf{d}\, h_t(\hat g)\big) \Big] \circ T_e L_{\hat g},$$

ω is the connection function and Kr is shorthand notation for $K(t)\, r_t(\widehat{g})$. □

Compared to the standard Riccati equation, the gain update equation (11) contains additional quadratic terms (namely, the two terms $\omega_{Kr} \circ K$ and $K \circ \omega^*_{Kr}$) that depend on the choice of affine connection and can generally not be absorbed into the standard quadratic term $K \circ E \circ K$. Moreover, the coefficient $A(t)$ of the linear term contains an additional dependence on the torsion. A coordinate version of equation (11) can be obtained by choosing bases for $\mathfrak{g}$, $\mathbb{R}^d$, and $\mathbb{R}^p$. We provide a worked example in Section V where we obtain a perturbed matrix Riccati differential equation for the filter gain. In the remaining part of this section, we prove Theorem 4.1.

B. The optimal estimation problem

The minimum-energy estimation problem stated in Section III is to find the state-control trajectory pair $(g^*_{[t_0,t]}(\tau), \delta^*_{[t_0,t]}(\tau))$, $\tau \in [t_0, t]$, that solves

$$\min_{(g(\cdot),\delta(\cdot))} m(g(t_0), t, t_0) + \int_{t_0}^{t} l\left(\delta(\tau), D^{-1}\big(y(\tau) - h(g(\tau), \tau)\big), t, \tau\right) d\tau \tag{15}$$

subject to the dynamic constraint

$$\dot g(t) = g(t)\big(\lambda(g(t), u(t), t) + B\delta(t)\big) \tag{16}$$

with free initial and final conditions. Here, $u(\tau)$ and $y(\tau)$ are *known* for $\tau \in [t_0, t]$, and we have substituted (2) into (6) to obtain (15).

Note how the 'control' input in the above optimal control problem is the model error δ while the applied input u is simply a known function of time. As we show in the following, the above rewriting of the minimum-energy estimation problem allows one to easily compute the associated optimal Hamiltonian, which is then used to obtain the explicit expression for the Hamilton-Jacobi-Bellman (HJB) equation. A suitable approximation of the solution to the HJB equation will then lead to the second-order-optimal filter presented at the beginning of this section.

We denote by $V(g, t)$ the minimum energy value among all trajectories of (16) within the interval $[t_0, t]$ that reach the state $g \in G$ at time t. The optimal estimate $\widehat{g}(t)$ is therefore equal to

$$\widehat{g}(t) = g^*_{[t_0,t]}(t) = \arg\min_{g \in G} V(g, t),$$

for $t \in [t_0, \infty)$ and $V(g, t_0) = m(g, t_0, t_0)$. The key observation in [1] is that if we assume $V(g, t)$ to be differentiable in a neighborhood of the optimal estimate $\widehat{g}(t)$ then, as $V(g, t)$ attains its minimum at $\widehat{g}(t)$, we must have

$$\mathbf{d}_1\, V(\widehat{g}(t), t) \equiv 0 \tag{17}$$

for $t \geq t_0$. Assuming that $V(g, t)$ is smooth, the above expression can be further differentiated with respect to time obtaining a set of necessary conditions (actually, a set of differential equations) that fully characterize the optimal filter. Unfortunately, such a program has the drawback that we obtain an *infinite* number of conditions and therefore, for practical application, the optimal filter has to be truncated after a certain order, obtaining a suboptimal filter. However, such filters have shown promising performance for systems on SO(3), outperforming established nonlinear filters such as the Multiplicative Extended Kalman Filter (MEKF) [8], see also [9], [10].

The simplest optimal filter is that obtained by truncating the series expansion at the second order. This requires only two differentiations of (17). It is worth recalling that for linear dynamics and quadratic cost, the minimum-energy filter obtained in this way is actually optimal and its equations are equivalent to the Kalman-Bucy filter [3].

C. The optimal Hamiltonian

Aiming for the Hamilton-Jacobi-Bellman (HJB) equation associated with the optimal estimation problem (15)-(16), in this subsection we derive the optimal Hamiltonian. Special care has to be taken in obtaining such a function because the system dynamics evolves on a smooth manifold and not, as would be more common, on the vector space $\mathbb{R}^n$. We refer to [19] for a review of optimal control theory on smooth manifolds (and, in particular, on Lie groups).

Given the estimator vector field (16) and incremental cost (4), the associated (time-varying) Hamiltonian $\tilde H \colon T^*G \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ is given by

$$\tilde H(p, \delta, t) := \frac{1}{2} e^{-\alpha(t-t_0)} \left(\mathcal{R}(\delta) + \mathcal{Q}(D^{-1}(y(t) - h(g, t)))\right) + \left\langle p, -g\big(\lambda(g, u(t), t) + B\delta\big)\right\rangle, \tag{18}$$

where g is the base point of $p \in T^*_g G \subset T^*G$. The optimal filtering problem (15)-(16) can be thought of as a standard optimal control problem which is solved *backward in time*. This justifies the presence of the minus sign in the pairing between the state dynamics and the Lagrange multiplier p on the right hand side of (18). In this way, one interprets the function m in (15) as the *terminal* cost and the minimum energy $V(g, t)$ as the cost-to-go.

As is typical for optimal control problems defined on Lie groups [19], the cotangent vector $p \in T^*_g G$ can be identified via left translation with the element $\mu \in \mathfrak{g}^*$, defined as $\mu = T_e L^*_g(p)$. Using $(g, \mu) \in G \times \mathfrak{g}^*$ in place of $p \in T^*G$ in (18), one obtains the left-trivialized Hamiltonian $\tilde H^- \colon G \times \mathfrak{g}^* \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ defined as

$$\tilde H^-(g, \mu, \delta, t) = \frac{1}{2} e^{-\alpha(t-t_0)} \big(\mathcal{R}(\delta) + \mathcal{Q}(D^{-1}(y(t) - h(g, t)))\big) - \big\langle \mu, \lambda(g, u(t), t) + B\delta \big\rangle. \tag{19}$$

We are now ready to compute the left-trivialized *optimal* Hamiltonian that characterizes the optimal control problem (15)-(16).

Proposition 4.2: The left-trivialized optimal Hamiltonian $H^- \colon G \times \mathfrak{g}^* \times \mathbb{R} \to \mathbb{R}$ associated with the optimal control problem (15)-(16) is given by

$$H^-(g, \mu, t) = -\frac{1}{2} e^{\alpha(t-t_0)} \big\langle \mu, B \circ R^{-1} \circ B^*(\mu) \big\rangle + \frac{1}{2} e^{-\alpha(t-t_0)} \mathcal{Q}\left(D^{-1}(y(t) - h(g, t))\right) - \big\langle \mu, \lambda(g, u(t), t) \big\rangle. \tag{20}$$

Proof: The vector field $g(\lambda(g, u(t), t) + B\delta)$ given in (16) is linear in δ, while the incremental cost $l(\delta, \varepsilon, t, \tau)$ given in (4) is quadratic in δ. It is straightforward to see that the unique minimum $\delta^{opt}(g, \mu, t)$ of the left-trivialized Hamiltonian (19) with respect to δ is attained at

$$\arg\min_{\delta} \tilde H^-(g, \mu, \delta, t) = e^{\alpha(t-t_0)} \cdot R^{-1} \circ B^*(\mu). \tag{21}$$

Substituting δ^{opt} into the left-trivialized Hamiltonian (19), the result follows. ■

The reason to study left-trivialized versions of the Hamiltonian and subsequently the Hamilton-Jacobi-Bellman equation will become clear in the next two sections where we use it to derive the second-order time evolution of the filter.

D. The left-trivialized HJB equation and the structure of the optimal filter

The Hamilton-Jacobi-Bellman equation associated with the optimal control problem (15)-(16) is given by

$$\frac{\partial}{\partial t}V(g,t) - H(\mathbf{d}_1 V(g,t),t) = 0 \tag{22}$$

with initial condition $V(g,t_0) = m(g,t_0,t_0)$. Here, $H\colon T^*G\times\mathbb{R}\to\mathbb{R}$ is the optimal Hamiltonian.

The presence of the minus sign in (22) is justified, as mentioned in the previous subsection, by the fact that the energy $V(g,t)$ should be thought of as the cost-to-go associated with the minimization of the cost functional in the interval $[t_0,t]$ while evolving the dynamics backwards in time, starting with g as final condition.

Equation (22) can be written in terms of the left-trivialized Hamiltonian as

$$\frac{\partial}{\partial t}V(g,t) - H^-\left(g,\mu(g,t),t\right) = 0, \tag{23}$$

where $\mu : G\times\mathbb{R}\to\mathfrak{g}^*$ is defined as

$$\mu(g,t) := T_eL_g^*\big(\mathbf{d}_1 V(g,t)\big). \tag{24}$$

The minimum energy estimator defines the estimate of the state at time t as the element $g\in G$ that minimizes the value function $V(g,t)$, that is

$$\widehat{g}(t) := \arg\min_{g\in G} V(g,t). \tag{25}$$

Assuming differentiability of the value function in a neighborhood of the minimum value, we obtain the necessary condition

$$\mathbf{d}_1 V(\widehat{g}(t),t) = 0 \quad \text{or, equivalently,} \quad \mu(\widehat{g}(t),t) = 0 \tag{26}$$

for all $t\geq 0$ [1]. Differentiating with respect to time it follows that

$$\mathrm{Hess}_1\, V(\hat{g}(t),t)\left(\dot{\hat{g}}(t)\right) + \mathbf{d}_1\left(\frac{\partial}{\partial t}V\right)(\hat{g}(t),t) = 0 \tag{27}$$

for $t \geq 0$, cf. Lemma A.1 in the appendix. Here, $\mathrm{Hess}_1\, V(\hat{g}(t),t)\colon T_{\hat{g}(t)}G \to T^*_{\hat{g}(t)}G$ is the Hessian operator, see Section II. Since $\frac{\partial}{\partial t}V\colon G\times\mathbb{R}\to\mathbb{R}$ satisfies the HJB equation (23), by straightforward application of the chain rule we obtain

$$\begin{aligned}\mathbf{d}_1\left(\frac{\partial}{\partial t}V\right)(g,t) &= \mathbf{d}_1\, H^-\big(g,\mu(g,t),t\big)\\ &\quad + \mathbf{d}_2\, H^-\big(g,\mu(g,t),t\big)\circ \mathbf{d}_1\,\mu(g,t).\end{aligned} \tag{28}$$

Using (26), when substituting in the equation above $g=\hat{g}(t)$, yields

$$\begin{aligned}\mathbf{d}_1\left(\frac{\partial}{\partial t}V\right)(\hat{g}(t),t) &= \mathbf{d}_1\, H^-(\hat{g}(t),0,t)\\ &\quad + \mathbf{d}_2\, H^-(\hat{g}(t),0,t)\circ \mathbf{d}_1\,\mu(\hat{g}(t),t).\end{aligned} \tag{29}$$

By Lemma A.2 in the appendix, recalling (24), we get

$$\begin{aligned}\mathbf{d}_1\,\mu(g,t) &= T_eL_g^*\circ \mathrm{Hess}_1\, V(g,t)\\ &\quad + \omega^{*\leftrightharpoons}_{T_eL_g^*(\mathbf{d}_1\, V(g,t))}\circ T_gL_{g^{-1}}.\end{aligned} \tag{30}$$

See Section II for the meaning of the second term. But then

$$\mathbf{d}_1\,\mu(\hat{g}(t),t) = T_eL_{\hat{g}(t)}^*\circ \mathrm{Hess}_1\, V(\hat{g}(t),t) \tag{31}$$

since the second term in (30) vanishes due to (26).

Define the left-trivialized Hessian operator $Z(g,t)\colon\mathfrak{g}\to\mathfrak{g}^*$ as

$$Z(g,t) := T_eL_g^*\circ \mathrm{Hess}_1\, V(g,t)\circ T_eL_g. \tag{32}$$

The key idea behind left-trivialization is that it converts the bundle map

$$\mathrm{Hess}_1\, V(\cdot,t)\colon G\to\mathfrak{L}(T_{(\cdot)}G, T^*_{(\cdot)}G)$$

into a map

$$Z(\cdot,t)\colon G\to\mathfrak{L}(\mathfrak{g},\mathfrak{g}^*)$$

with a (fixed) vector space codomain. This allows one to construct a global matrix representation for Z by choosing bases in $\mathfrak{g}$ and $\mathfrak{g}^*$. When taking the time derivative of $\mathrm{Hess}_1\, V(g(t),t)$ along a curve $g(t)$, left-trivialization plus choice of bases turns this into simple differentiation of a time-varying matrix $Z(t) = Z(g(t),t)$.

Continuing with the rewriting of (29), using (31) and (32)

$$\begin{aligned}&\mathbf{d}_2\, H^-(\hat{g}(t),0,t)\circ \mathbf{d}_1\,\mu(\hat{g}(t),t)\circ T_eL_{\hat{g}(t)}\\ &= \mathbf{d}_2\, H^-(\hat{g}(t),0,t)\circ Z(\hat{g}(t),t)\\ &= Z(\hat{g}(t),t)^*(\mathbf{d}_2\, H^-(\hat{g}(t),0,t))\end{aligned} \tag{33}$$

and similarly

$$\mathbf{d}_1\, H^-(\hat{g}(t),0,t)\circ T_eL_{\hat{g}(t)} = T_eL_{\hat{g}(t)}^*\big(\mathbf{d}_1\, H^-(\hat{g}(t),0,t)\big). \tag{34}$$

If we now turn the attention to the first term of (27), then

$$\begin{aligned}&\mathrm{Hess}_1\, V(\hat{g}(t),t)\left(\dot{\hat{g}}(t)\right)\circ T_eL_{\hat{g}(t)} = \\ &T_eL_{\hat{g}(t)}^*\circ \mathrm{Hess}_1\, V(\hat{g}(t),t)\left(\dot{\hat{g}}(t)\right) = Z(\hat{g}(t),t)\left(\hat{g}(t)^{-1}\dot{\hat{g}}(t)\right).\end{aligned} \tag{35}$$

Finally, using (29), (33), (34) and (35), we rewrite (27) as

$$\begin{aligned}Z(\hat{g}(t),t)\left(\hat{g}(t)^{-1}\dot{\hat{g}}(t)\right) &= -T_eL_{\hat{g}(t)}^*\big(\mathbf{d}_1\, H^-(\hat{g}(t),0,t)\big)\\ &\quad - Z(\hat{g}(t),t)^*(\mathbf{d}_2\, H^-(\hat{g}(t),0,t))\end{aligned} \tag{36}$$

for $t\geq 0$. By equation (26), the point $\hat{g}(t)\in G$ is a critical point of the value function, and hence the Hessian $\mathrm{Hess}_1\, V(\hat{g}(t),t)$ is symmetric, see Section II. By equation (32) then $Z(\hat{g}(t),t)$ is symmetric, i.e., $Z(\hat{g}(t),t) = Z(\hat{g}(t),t)^*$. Here we have used the identification of the bidual $\mathfrak{g}^{**}$ with $\mathfrak{g}$. Assuming further that $\mathrm{Hess}_1\, V(\hat{g}(t),t)$ and hence $Z(\hat{g}(t),t)$ is invertible, then equation (36) is equivalent to

$$\begin{aligned}\hat{g}(t)^{-1}\dot{\hat{g}}(t) &= -\,\mathbf{d}_2\, H^-(\hat{g}(t),0,t)\\ &\quad - Z(\hat{g}(t),t)^{-1}\circ T_eL_{\hat{g}(t)}^*\big(\mathbf{d}_1\, H^-(\hat{g}(t),0,t)\big).\end{aligned} \tag{37}$$

We can now develop (37) further, exploiting the specific expression for H^- given in (20). In the following, we adopt the shorthand notation $h_t(g)$ and $\lambda_t(g,u)$ for $h(g(t),t)$ and $\lambda(g(t),u(t),t)$, respectively, and drop the explicit dependence on time of signals from our notation where convenient. From (20),

$$\mathbf{d}_2\, H^-(g,\mu,t) = -\,e^{\alpha(t-t_0)}\cdot B\circ R^{-1}\circ B^*(\mu) - \lambda_t(g,u) \tag{38}$$

and, using Lemma A.5 in the appendix,

$$\begin{aligned}\mathbf{d}_1\, H^-(g,\mu,t) &= -e^{-\alpha(t-t_0)}\cdot\Big(\big(D^{-1}\big)^*\circ Q\circ D^{-1}\\ &\quad \big(y-h_t(g)\big)\Big)\circ \mathbf{d}\, h_t(g) - \mu\circ \mathbf{d}_1\,\lambda_t(g,u).\end{aligned} \tag{39}$$

Substituting $g=\hat{g}$, the two expressions above become

$$\mathbf{d}_2\, H^-(\hat{g},0,t) = -\lambda_t(\hat{g},u) \tag{40}$$

and

$$\begin{aligned}\mathbf{d}_1\, H^-(\hat{g},0,t) &= -e^{-\alpha(t-t_0)}\cdot\Big(\big(D^{-1}\big)^*\circ Q\circ D^{-1}\\ &\quad \big(y-h_t(\widehat{g})\big)\Big)\circ \mathbf{d}\, h_t(\widehat{g}).\end{aligned} \tag{41}$$

Here we have again used the identification of $\mathfrak{g}$ with its bidual $\mathfrak{g}^{**}$, allowing us to interpret the differential $\mathbf{d}_2\, H^-(\hat{g}(t),\mu,t)\colon\mathfrak{g}^*\to\mathbb{R}$ as an element of $\mathfrak{g}$. Defining $r_t(\widehat{g})\in\mathfrak{g}^*$ by

$$r_t(\widehat{g}) := T_eL_{\hat{g}}^*\Big[\Big(\big(D^{-1}\big)^*\circ Q\circ D^{-1}\big(y-h_t(\widehat{g})\big)\Big)\circ \mathbf{d}\, h_t(\widehat{g})\Big] \tag{42}$$

we can then write (37) as

$$\widehat{g}^{-1}\dot{\widehat{g}} = \lambda_t(\widehat{g}, u) + e^{-\alpha(t-t_0)} \cdot Z(\widehat{g}, t)^{-1} r_t(\widehat{g}). \tag{43}$$

Compare this to equations (8) and (10) in Theorem 4.1, noting that $e^{-\alpha(t-t_0)} \cdot Z(\hat{g}(t), t)^{-1}$ maps $\mathfrak{g}^*$ to $\mathfrak{g}$. Since the integral part of the cost (6) vanishes at the initial time $t = t_0$, the initial condition for the optimal filter is as in (9).

E. Approximate time evolution of Z

Ideally, one would like to compute a differential equation for $Z(\hat{g}(t), t)$ so that coupling it with (43) one obtains the optimal filter for (1)-(2). Unfortunately, it is well known – in the flat case – that such an approach is going to fail as Z satisfies an infinite dimensional differential equation, the linear dynamics with quadratic cost being one of the most important exceptions [1]. For this reason, in the following we compute an approximation of the time evolution of $Z(g, t)$ along the optimal solution $\hat{g}(t)$ by neglecting the third covariant derivative of the value function V. Such an approximation is denoted by $X(g, t)$. In the case of linear dynamics with quadratic cost the value function is itself quadratic, meaning that its third derivative is zero and $X(g, t) = Z(g, t)$ in that case. In the general Lie group case, we have the following result.

Proposition 4.3: $X(t) := X(\hat{g}(t), t) \in \mathfrak{L}(\mathfrak{g}, \mathfrak{g}^*)$ fulfills the operator Riccati equation

$$\begin{aligned} \dot{X} &= e^{-\alpha(t-t_0)} \cdot S - F^* \circ X - X \circ F \\ &\quad - e^{\alpha(t-t_0)} \cdot X \circ B \circ R^{-1} \circ B^* \circ X, \qquad X(t_0) = X_0 \end{aligned} \tag{44}$$

with

$$X_0 = T_e L_{\hat{g}_0}^* \circ \text{Hess } m_0(\hat{g}_0) \circ T_e L_{\hat{g}_0}, \tag{45}$$

$$F(t) = -\omega_{\hat{g}^{-1}\dot{\hat{g}}} + \omega_{\lambda_t(\hat{g},u)}^{\leftrightarrows} + \mathbf{d}_1 \lambda_t(\hat{g}, u) \circ T_e L_{\hat{g}}, \tag{46}$$

$$\begin{aligned} S(t) &= -T_e L_{\hat{g}}^* \circ \\ &\Big(\Big(\big(D^{-1}\big)^* \circ Q \circ D^{-1}\big(y - h_t(\hat{g})\big) \Big)^{T_{\hat{g}} G} \circ \text{Hess}\, h_t(\hat{g}) + \\ &\quad - \big(\mathbf{d}\, h_t(\hat{g})\big)^* \circ \big(D^{-1}\big)^* \circ Q \circ D^{-1} \circ \mathbf{d}\, h_t(\hat{g}) \Big) \circ T_e L_{\hat{g}}, \end{aligned} \tag{47}$$

and $\hat{g}_0$ as in (9).

Remark. Note that, as (46) depends on $\hat{g}^{-1}\dot{\hat{g}}$ given in (43), the terms $F^* \circ X$ and $X \circ F$ in (44) are quadratic and not just linear in X, if the connection function ω is not trivial. This fact is ultimately responsible for the appearance, in the perturbed Riccati equation (11), of the nonstandard quadratic terms $\omega_{K_r} \circ K$ and $K \circ \omega_{K_r}^*$ and is one of the key discoveries of this work. □

Proof: From (32), setting $g = \hat{g}(t)$, and using Lemma A.3 in the appendix we get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} Z(\hat{g}(t), t) &= \frac{\mathrm{d}}{\mathrm{d}t}\big(T_e L_{\hat{g}(t)}^* \circ \text{Hess}_1 V(\hat{g}(t), t) \circ T_e L_{\hat{g}(t)}\big) \\ &= \omega_{\hat{g}^{-1}\dot{\hat{g}}}^* \circ Z(\hat{g}(t), t) + Z(\hat{g}(t), t) \circ \omega_{\hat{g}^{-1}\dot{\hat{g}}} \\ &\quad + T_e L_{\hat{g}(t)}^* \circ \frac{\partial}{\partial t}(\text{Hess}_1 V)(\hat{g}(t), t) \circ T_e L_{\hat{g}(t)} \\ &\quad + h.o.t., \end{aligned} \tag{48}$$

where the higher order terms ($h.o.t.$) will be neglected to obtain a finite dimensional approximation to the (infinite dimensional) optimal filter. See Section II for the meaning of the operators ω_X and ω_X^*. The partial time derivative commutes with covariant differentiation on G, so we can use equation (28) to compute $\frac{\partial}{\partial t}(\text{Hess}_1 V)(\hat{g}(t), t) = \text{Hess}_1(\frac{\partial}{\partial t} V)(\hat{g}(t), t)$. We start by rewriting equation (28) into the following form,

$$\begin{aligned} \mathbf{d}_1 \left(\frac{\partial}{\partial t} V\right)(g, t) &= \mathbf{d}_1 H^-\big(g, \mu(g, t), t\big) \\ &\quad + \big(\mathbf{d}_1 \mu(g, t)\big)^* \big(\mathbf{d}_2 H^-\big(g, \mu(g, t), t\big)\big). \end{aligned} \tag{49}$$

By (30) and using the operator calculus from Section II,

$$\begin{aligned} \big(\mathbf{d}_1 \mu(g, t)\big)^*(W) &= \big(\text{Hess}_1 V(g, t)\big)^* \circ T_e L_g(W) + \\ &\quad T_g L_{g^{-1}}^* \circ \omega_W^{\leftrightarrows *}\big(T_e L_g^*(\mathbf{d}_1 V(g, t))\big), \end{aligned}$$

for $W \in \mathfrak{g}^{**} \simeq \mathfrak{g}$. Combining this with equation (49) we arrive at

$$\begin{aligned} \mathbf{d}_1 \left(\frac{\partial}{\partial t} V\right)(g, t) &= \mathbf{d}_1 H^-\big(g, \mu(g, t), t\big) \\ &\quad + \big(\text{Hess}_1 V(g, t)\big)^* \circ T_e L_g\big(\mathbf{d}_2 H^-\big(g, \mu(g, t), t\big)\big) \\ &\quad + T_g L_{g^{-1}}^* \circ \omega_{\mathbf{d}_2 H^-(g,\mu(g,t),t)}^{\leftrightarrows *}\big(T_e L_g^*(\mathbf{d}_1 V(g, t))\big). \end{aligned}$$

Then, using the chain rule and Lemma A.4 in the appendix,

$$\begin{aligned} &\frac{\partial}{\partial t}(\text{Hess}_1 V)(\hat{g}(t), t) = \text{Hess}_1 \left(\frac{\partial}{\partial t} V\right)(\hat{g}(t), t) = \\ &\quad \text{Hess}_1 H^-(\hat{g}, 0, t) + \mathbf{d}_2(\mathbf{d}_1 H^-)(\hat{g}, 0, t) \circ \mathbf{d}_1 \mu(\hat{g}, t) + \\ &\quad \text{Hess}_1 V(\hat{g}, t) \circ T_e L_{\hat{g}} \circ \omega_{\mathbf{d}_2 H^-(\hat{g},0,t)}^{\leftrightarrows} \circ T_{\hat{g}} L_{\hat{g}^{-1}} + \\ &\quad \text{Hess}_1 V(\hat{g}, t) \circ T_e L_{\hat{g}} \circ \mathbf{d}_1(\mathbf{d}_2 H^-)(\hat{g}, 0, t) + \\ &\quad \text{Hess}_1 V(\hat{g}, t) \circ T_e L_{\hat{g}} \circ \text{Hess}_2 H^-(\hat{g}, 0, t) \circ \mathbf{d}_1 \mu(\hat{g}, t) + \\ &\quad T_{\hat{g}} L_{\hat{g}^{-1}}^* \circ \omega_{\mathbf{d}_2 H^-(\hat{g},0,t)}^{\leftrightarrows *} \circ T_e L_{\hat{g}}^* \circ \text{Hess}_1 V(\hat{g}, t) + \\ &\quad h.o.t. \end{aligned} \tag{50}$$

Here we have used equation (26) and the fact that the Hessian operator at a critical point is symmetric. Combining equations (48), (50) and (31) and neglecting higher order terms, we arrive at

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} Z(\hat{g}(t), t) \approx \\ &\quad \omega_{\hat{g}^{-1}\dot{\hat{g}}}^* \circ Z(\hat{g}, t) + Z(\hat{g}, t) \circ \omega_{\hat{g}^{-1}\dot{\hat{g}}} + \\ &\quad T_e L_{\hat{g}}^* \circ \text{Hess}_1 H^-(\hat{g}, 0, t) \circ T_e L_{\hat{g}} + \\ &\quad T_e L_{\hat{g}}^* \circ \mathbf{d}_2(\mathbf{d}_1 H^-)(\hat{g}, 0, t) \circ Z(\hat{g}, t) + \\ &\quad \omega_{\mathbf{d}_2 H^-(\hat{g},0,t)}^{\leftrightarrows *} \circ Z(\hat{g}, t) + Z(\hat{g}, t) \circ \omega_{\mathbf{d}_2 H^-(\hat{g},0,t)}^{\leftrightarrows} + \\ &\quad Z(\hat{g}, t) \circ \mathbf{d}_1(\mathbf{d}_2 H^-)(\hat{g}, 0, t) \circ T_e L_{\hat{g}} + \\ &\quad Z(\hat{g}, t) \circ \text{Hess}_2 H^-(\hat{g}, 0, t) \circ Z(\hat{g}, t). \end{aligned} \tag{51}$$

Differentiating equation (39) and using Lemma A.5 in the appendix we obtain

$$\begin{aligned} &\text{Hess}_1 H^-(\hat{g}(t), 0, t) = -e^{-\alpha(t-t_0)} \\ &\quad \cdot \Big(\big(D^{-1}\big)^* \circ Q \circ D^{-1}\big(y - h_t(\hat{g})\big) \Big)^{T_{\hat{g}} G} \circ \text{Hess}\, h_t(\hat{g}) \\ &\quad + e^{-\alpha(t-t_0)} \cdot \big(\mathbf{d}\, h_t(\hat{g})\big)^* \circ \big(D^{-1}\big)^* \circ Q \circ D^{-1} \circ \mathbf{d}\, h_t(\hat{g}), \end{aligned} \tag{52}$$

where $(\cdot)^{T_{\hat{g}} G}$ is the exponential functor, see Section II. Also, from (39) and (38),

$$\begin{aligned} \mathbf{d}_2(\mathbf{d}_1 H^-)(\hat{g}(t), 0, t) &= \big(\mathbf{d}_1(\mathbf{d}_2 H^-)(\hat{g}(t), 0, t)\big)^* \\ &= -\big(\mathbf{d}_1 \lambda(\hat{g}(t), u(t), t)\big)^*, \end{aligned} \tag{53}$$

and differentiating (38) yields

$$\text{Hess}_2 H^-(\hat{g}(t), 0, t) = -e^{\alpha(t-t_0)} \cdot B \circ R^{-1} \circ B^*. \tag{54}$$

Using (51)-(54) and (40) it is straightforward to show that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} Z(\hat{g}(t), t) &\approx e^{-\alpha(t-t_0)} \cdot S - F^* \circ Z(\hat{g}, t) - Z(\hat{g}, t) \circ F \\ &\quad - e^{\alpha(t-t_0)} \cdot Z(\hat{g}, t) \circ B \circ R^{-1} \circ B^* \circ Z(\hat{g}, t), \end{aligned}$$

with $S(t)$ and $F(t)$ as in (47) and (46), respectively. As we neglected the third order covariant derivative of V in (48) and (50), the above equation is only an approximation of $\frac{\mathrm{d}}{\mathrm{d}t}Z(\hat{g}(t),t)$. To highlight this fact, in (44), we write X instead of Z. The formula (45) for the initial condition follows immediately from the initial condition for the HJB. This completes the proof. ∎

It remains to prove Theorem 4.1 stated in Section III. To this end, note that in (43) the inverse of the matrix X is required. It is however unnecessary to compute the inverse of X. Indeed, defining $K(t) := e^{-\alpha(t-t_0)}X^{-1}(t)$, with $X(t)$ satisfying (44), the second-order-optimal filter can be computed from Proposition 4.3, equation (43) and a straightforward application of the well known formula for the derivative of the inverse of an operator.

V. A WORKED EXAMPLE

In this section, we detail two different second-order-optimal filters for the rotational dynamics of a rigid body subject to external torques, assuming some directional measurements are available. The two filters correspond to different choices of affine connection. The section concludes with a numerical simulation illustrating the performance of the two second-order-optimal filters.

A. Derivation of the filter equation

We represent the orientation of a rigid body in space by the rotation matrix $\mathbf{R} \in \mathrm{SO}(3)$ that encodes the coordinates of a body-fixed frame $\{B\}$ with respect to the coordinates of an inertial frame $\{A\}$. We denote by $\mathbb{I}$ the inertia tensor, by $\mathbf{\Omega}$ the angular velocity, and by τ the applied external torque, all of them expressed in the body-fixed frame $\{B\}$. The rotational dynamics of a rigid body evolves on $T\mathrm{SO}(3)$, the tangent bundle of the special orthogonal group SO(3). It is standard practice to identify $T\mathrm{SO}(3)$ with $\mathrm{SO}(3)\times\mathbb{R}^3$ via left translation [17] and write the rotation dynamics as

$$\mathbf{R}^T\dot{\mathbf{R}} = \mathbf{\Omega}^\times, \tag{55}$$

$$\dot{\mathbf{\Omega}} = \mathbb{I}^{-1}\left((\mathbb{I}\mathbf{\Omega})^\times\mathbf{\Omega} + \tau\right), \tag{56}$$

with $(\mathbf{R},\mathbf{\Omega}) \in \mathrm{SO}(3)\times\mathbb{R}^3$.

We consider the nonlinear filtering problem of reconstructing the attitude matrix $\mathbf{R}$ and the angular velocity of the rigid body assuming that we have *two* (possibly time-varying) reference direction measurements $\mathring{a}_1$ and $\mathring{a}_2 \in \mathbb{S}^2$ corrupted by measurement noise. Here, $\mathbb{S}^2$ denotes the 2-sphere of unit norm in the inertial reference frame $\{A\}$. Typical examples of reference directions are the magnetic or gravitational fields at the location in which the system is operating.

As measurement output model, we employ

$$\mathbb{R}^6 \ni y(t) = \begin{bmatrix} a_1(t) \\ a_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{R}^T(t)\mathring{a}_1(t) \\ \mathbf{R}^T(t)\mathring{a}_2(t) \end{bmatrix} + D\varepsilon(t), \tag{57}$$

where ε represents the unknown measurement error. Equation (57) has the structure of (2). where, for ease of presentation, we assume that D a block diagonal structure, namely

$$D = \begin{bmatrix} d_1 I_{3\times3} & 0 \\ 0 & d_2 I_{3\times3} \end{bmatrix}. \tag{58}$$

As error model for the dynamics, we choose

$$\mathbf{R}^T\dot{\mathbf{R}} = (\mathbf{\Omega})^\times, \tag{59}$$

$$\dot{\mathbf{\Omega}} = \mathbb{I}^{-1}\left((\mathbb{I}\mathbf{\Omega})^\times\mathbf{\Omega} + \tau\right) + B_2\delta. \tag{60}$$

Equations (59)-(60) are a particular case of (1) where $\delta \in \mathbb{R}^3$ and $B : \mathbb{R}^3 \to \mathbb{R}^3_\times \times \mathbb{R}^3$, $\delta \mapsto (0, B_2\delta)$, with $B_2 \in \mathbb{R}^{3\times3}$.

We assume that the optimal filtering problem is posed in terms of the minimization of the cost functional (6) with the quadratic incremental cost (4). Without loss of generality and for ease of presentation, in (4) we assume that the quadratic form $\mathcal{Q}$ is *block diagonal*, while no additional conditions are imposed to the quadratic form $\mathcal{R}$ other than being strictly positive definite. As done in (7), the quadratic forms $\mathcal{Q}$ and $\mathcal{R}$ are specified in terms of two symmetric positive definite linear maps Q and R. As mentioned, for the linear map Q we will assume a block diagonal structure. Namely, we take the matrix representation

$$Q = \begin{bmatrix} q_1 I_{3\times3} & 0 \\ 0 & q_2 I_{3\times3} \end{bmatrix}, \tag{61}$$

with respect to the standard bases in $\mathbb{R}^p$ and $(\mathbb{R}^p)^*$, $p = 6$, where q_1 and q_2 are strictly positive constants.

By choosing different group operations, we can assign different Lie group structures to the tangent bundle $T\mathrm{SO}(3) \approx \mathrm{SO}(3)\times\mathbb{R}^3$. Here, we follow the approach detailed in, e.g., [27], [23] and select the product group structure, that is, for $(\mathbf{R},\mathbf{X})$ and $(\mathbf{S},\mathbf{Y}) \in \mathrm{SO}(3)\times\mathbb{R}^3$, we define the group product as

$$(\mathbf{R},\mathbf{X})\cdot(\mathbf{S},\mathbf{Y}) = (\mathbf{RS},\mathbf{X}+\mathbf{Y}). \tag{62}$$

The Lie algebra of the product group $\mathrm{SO}(3)\times\mathbb{R}^3$ is the product algebra $\mathfrak{so}(3)\times\mathbb{R}^3 \approx \mathbb{R}^3_\times\times\mathbb{R}^3$. Here, $\mathbb{R}^3_\times$ is the Lie algebra of $\mathbb{R}^3$ with the cross product Lie bracket. Given (η^R,η^Ω) and $(\xi^R,\xi^\Omega) \in \mathbb{R}^3_\times\times\mathbb{R}^3$, the adjoint representation of the Lie algebra $\mathbb{R}^3_\times\times\mathbb{R}^3$ onto itself is simply given by

$$\mathrm{ad}_{(\eta^R,\eta^\Omega)}(\xi^R,\xi^\Omega) = (\eta^R\times\xi^R, 0). \tag{63}$$

In matrix form, $\mathrm{ad}_{(\eta^R,\eta^\Omega)}$ is represented by the 6×6 matrix

$$\mathrm{ad}_{(\eta^R,\eta^\Omega)} = \begin{bmatrix} (\eta^R)^\times & 0 \\ 0 & 0 \end{bmatrix}, \tag{64}$$

where we have chosen the standard basis for the Lie algebra $\mathbb{R}^3_\times\times\mathbb{R}^3$.

A (left-invariant) connection has to be chosen to derive the second order optimal filter on a Lie group. First, as done in [27], [23], we make use of the symmetric Cartan-Schouten (0)-connection, characterized by the connection function

$$\omega^{(0)} = \frac{1}{2}\,\mathrm{ad}. \tag{65}$$

Proposition 5.1: Consider the system (55)-(56) with dynamics error model (59)-(60), measurement output model (57), and incremental cost (4) with block diagonal structure given by (61). The second-order optimal filter with respect to the Cartan-Schouten (0)-connection is given by

$$\widehat{\mathbf{R}}^T\dot{\widehat{\mathbf{R}}} = \left(\widehat{\mathbf{\Omega}} + K_{11}\,r^R + K_{12}\,r^\Omega\right)^\times \tag{66}$$

$$\dot{\widehat{\mathbf{\Omega}}} = \mathbb{I}^{-1}\left((\mathbb{I}\widehat{\mathbf{\Omega}})^\times\widehat{\mathbf{\Omega}} + \tau\right) + K_{21}\,r^R + K_{22}\,r^\Omega, \tag{67}$$

where the residual $r_t = (r^R; r^\Omega)$ and the second-order optimal gain $K = (K_{11}, K_{12}; K_{21}, K_{22})$ are given below. Let

$$\hat{a}_1 = \widehat{\mathbf{R}}^T\mathring{a}_1, \quad \text{and} \quad \hat{a}_2 = \widehat{\mathbf{R}}^T\mathring{a}_2.$$

then the residual r_t is given by

$$r_t = \begin{bmatrix} r^R \\ r^\Omega \end{bmatrix} = \begin{bmatrix} -(q_1/d_1^2)(\hat{a}_1\times a_1) - (q_2/d_2^2)(\hat{a}_2\times a_2) \\ 0 \end{bmatrix}, \tag{68}$$

while the gain K is the solution of the perturbed matrix Riccati differential equation

$$\begin{aligned}\dot{K} = &-\alpha K + AK + KA^\top - KEK + BR^{-1}B^\top \\ &- W(K,r_t)\,K - K\,W(K,r_t)^\top,\end{aligned} \tag{69}$$

where

$$A = \begin{bmatrix} -\widehat{\mathbf{\Omega}}^\times & I \\ 0 & \mathbb{I}^{-1}\left[(\mathbb{I}\widehat{\mathbf{\Omega}})^\times - \widehat{\mathbf{\Omega}}^\times \mathbb{I}\right] \end{bmatrix}, \tag{70}$$

$$E = \begin{bmatrix} \sum_{i=1}^{2} -(q_i/d_i^2)\left(\hat{a}_i^\times a_i^\times + a_i^\times \hat{a}_i^\times\right)/2 & 0 \\ 0 & 0_{3\times 3} \end{bmatrix}, \tag{71}$$

$$BR^{-1}B^\top = \begin{bmatrix} 0_{3\times 3} & 0 \\ 0 & B_2 R^{-1} B_2^\top \end{bmatrix}, \quad \text{and} \tag{72}$$

$$W(K, r_t) = \begin{bmatrix} 1/2(K_{11} r^R + K_{12} r^\Omega)^\times & 0 \\ 0 & 0_{3\times 3} \end{bmatrix}. \tag{73}$$

The reported filter equations result from a straightforward application of the theory presented in this paper.

To illustrate the dependence of the filter on the choice of affine connection, we repeat the filter construction with the alternative choice of the Cartan-Schouten (-)-connection, characterized by the connection function

$$\omega^{(-)} = 0. \tag{74}$$

Proposition 5.2: Consider the system (55)-(56) with dynamics error model (59)-(60), measurement output model (57), and incremental cost (4) with block diagonal structure given by (61). The second order optimal filter with respect to the Cartan-Schouten (-)-connection is given by (66)-(67) where the second-order optimal gain $K = (K_{11}, K_{12}; K_{21}, K_{22})$ is the solution of the perturbed matrix Riccati differential equation (69) where

$$A = \begin{bmatrix} 0 & I \\ 0 & \mathbb{I}^{-1}\left[(\mathbb{I}\widehat{\mathbf{\Omega}})^\times - \widehat{\mathbf{\Omega}}^\times \mathbb{I}\right] \end{bmatrix}, \tag{75}$$

$$E = \begin{bmatrix} \sum_{i=1}^{2} -(q_i/d_i^2)\; a_i^\times \hat{a}_i^\times & 0 \\ 0 & 0_{3\times 3} \end{bmatrix}, \tag{76}$$

$$BR^{-1}B^\top = \begin{bmatrix} 0_{3\times 3} & 0 \\ 0 & B_2 R^{-1} B_2^\top \end{bmatrix}, \quad \text{and} \tag{77}$$

$$W(K, r_t) = 0. \tag{78}$$

Note that the Riccati equation in Proposition 5.2 is a standard Riccati differential equation, however, it is not symmetric. In contrast, the Riccati equation in Proposition 5.1 is perturbed but symmetric.

The performance of the filters introduced in Propositions 5.1 and 5.2 is illustrated by means of numerical simulations in the next subsection.

B. Numerical simulations

Setting up a rigorous analysis of the performance of the proposed filter, perhaps including a comparison with the state of the art, goes beyond the scope of this work. Here, we limit ourself to present numerical results to demonstrate the practical implementability of the second-order-optimal minimum-energy filter on Lie groups in the example presented in the previous subsection. To this end, we consider a rigid body with inertia tensor given by

$$\mathbb{I} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad \text{Kg m}^2 \tag{79}$$

subject to the following nominal input torque

$$\tau(t) = \left(\sin\left(\frac{2\pi}{3}t\right); \cos\left(\frac{2\pi}{1}t\right); \sin\left(\frac{2\pi}{5}t\right)\right) \quad \text{Nm}, \tag{80}$$

where ; denotes row concatenation. The total simulation time is chosen to be $T = 50$ s.

The model error $B_2\delta$ appearing in (60) is generated as a Gaussian white noise with zero mean and standard deviation $\sigma_\delta = 1$ rad/s^2. Assuming for simplicity of exposition and without loss of generality, that the only source of error in the dynamics is the incorrect measurement of the input torque, the error $B_2\delta$ can be interpreted as an additive measurement error in the input of the form $\mathbb{I}B_2\delta$. This is clearly evident from rewriting the error model (60) as

$$\dot{\mathbf{\Omega}} = \mathbb{I}^{-1}\left((\mathbb{I}\mathbf{\Omega})^\times \mathbf{\Omega} + \tau + \mathbb{I}B_2\delta\right). \tag{81}$$

The nominal and measured inputs are depicted in Figure 1, where the scaling effect of the inertia tensor $\mathbb{I}$ in the term $\mathbb{I}B_2\delta$ can be directly appreciated.

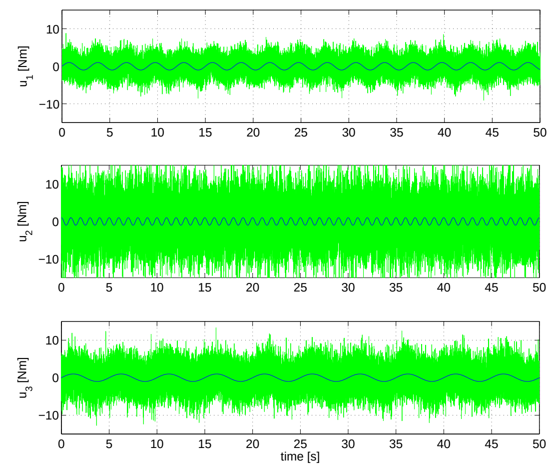

Fig. 1. Input signal. The nominal input torque (blue) is depicted on top of measured input torque (green).

The nominal input (80) is used to generate a nominal trajectory whose initial rotation $R(0)$ is selected to be equal to the identity matrix and whose initial angular velocity has been chosen to be

$$\omega(0) = (0.5; 0.6; 0.4) \text{ rad/s}. \tag{82}$$

The two reference direction measurements used in (57) are selected to be

$$\mathring{a}_1 = (0; 0; 1), \qquad \mathring{a}_2 = (0; 1; 0). \tag{83}$$

The nominal output (57) resulting from the integration of the dynamics (55)-(56) with nominal input (80) and initial condition (82) is shown in Figure 2. The integration of the nominal trajectory has been conducted with a constant time step of $h = 1 \cdot 10^{-3}$ s using a geometric forward Euler method, i.e., the attitude update is obtained as $\mathbf{R}(t+h) = \mathbf{R}(t)\exp(h\mathbf{\Omega}(t))$ to ensure $\mathbf{R}(t+h) \in \mathrm{SO}(3)$.

The measured output is obtained by adding a Gaussian white noise with standard deviation $\sigma_\varepsilon = 0.5/3$ m, corresponding to assume that with probability 99.7% (3 standard deviations) the output samples have a distance of less than 0.5 m from the nominal output samples. Note that a sample error of 0.5 m corresponds to an error of about 30 deg in the measurement of the reference directions (83). The measured output is depicted, together with the nominal output, in Figure 2.

The measurement errors selected for the input and output signals suggest to employ in (58) and (60)

$$d_1 = d_2 = \sigma_\varepsilon, \qquad B_2 = \sigma_\delta I_{3\times 3}, \tag{84}$$

respectively. The matrices Q in (61) and R are then chosen as

$$q_1 = q_2 = \sigma_\varepsilon^2, \qquad R = \sigma_\delta^2 I_{3\times 3}, \tag{85}$$

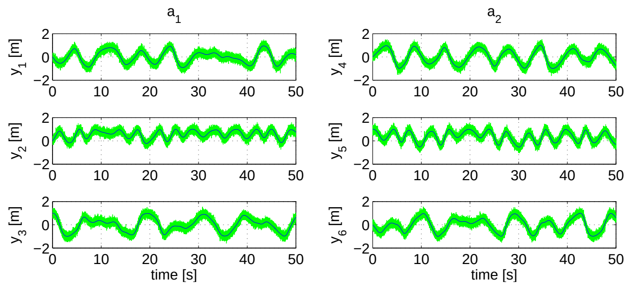

Fig. 2. Output signal. The nominal output (blue), corresponding to the measurement directions a_1 and a_2 in body coordinates, is depicted together with the measured output (green).

so that the filter cost is the sum of the energies of the errors $D\varepsilon$ and $B\delta$ since, by construction, $Q = D^\top D$ and $R = B^\top B$. The selection of the Q and R matrices, drawing a parallel to what happens in the (perhaps more familiar to the reader) setting of stochastic Kalman filtering, would correspond to choose as filter's cost the sum of model and output noise variances assuming δ and ε zero-mean unit-variance Gaussian processes.

The resulting estimation error for the second-order-optimal minimum-energy filters is presented in Figure 3. The figure reports, for the (0)- and (-)-connections, the attitude error angle $e^{\mathbf{R}} \in [0, \pi]$ and angular velocity error $e^{\mathbf{\Omega}}(t) \in \mathbb{R}^3$. These errors are defined as

$$e^{\mathbf{R}}(t) := \operatorname{acos}\left(1 - \frac{\operatorname{tr}(I - \mathbf{R}^\top(t)\widehat{\mathbf{R}}(t))}{2}\right), \tag{86}$$

$$e^{\mathbf{\Omega}}(t) := \widehat{\mathbf{\Omega}}(t) - \mathbf{\Omega}(t). \tag{87}$$

In the simulation, the filter was initialized with attitude error angle of about 60 deg, the forgetting factor $\alpha = 0.01$ and the initial condition for the Riccati gain $K(0) = I_{6\times 6}$. Both filters show similar performances, with the (0)-connection based filter performing best. This trend seems to be general but requires further investigation to be confirmed and analyzed.

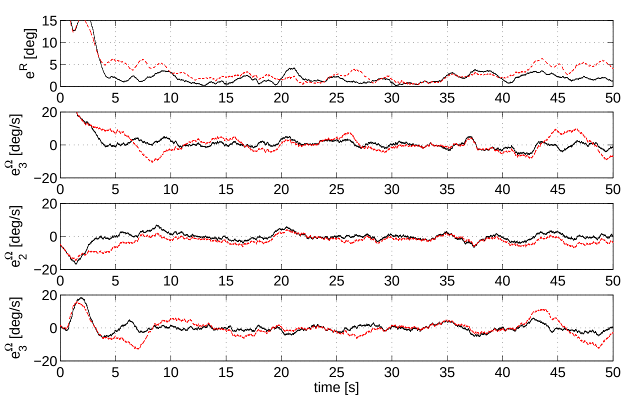

Fig. 3. Filtering error. The plots show the attitude angle error $e^{\mathbf{R}} \in \mathbb{R}$ and the angular velocity error $e^{\mathbf{\Omega}} \in \mathbb{R}^3$ for the minimum-energy filter based on the (0)-connection (black) and the (-)-connection (dashed red).

VI. Conclusions

We provided an explicit formula for the second-order-optimal minimum-energy filter for systems on Lie groups with vectorial measurements. We showed in an example how to use this formula to derive minimum-energy filters for (second-order) mechanical systems. Numerical simulations have been performed to demonstrate the effectiveness of the filter.

The proposed filter does not require a Riemannian metric on the Lie group and, as a by-product, the resulting (modified) Riccati differential equation lives in the dual of the Lie algebra of the group, rather than in the tangent space as in earlier approaches.

Appendix

In this appendix we provide a series of lemmas that are used in the derivation of the filter equation in Section IV. These lemmas should be of independent interest since they provide a second-order left-trivialized calculus on Lie groups. All functions are assumed sufficiently smooth.

Lemma A.1: Given $f\colon G \times \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to G$ then $\mathbf{d}_1 f(g(t), t) = 0$ for all t implies

$$\operatorname{Hess}_1 f(g(t), t)(\dot g(t)) + \mathbf{d}_1\Big(\frac{\partial}{\partial t} f\Big)(g(t), t) = 0.$$

Proof: Let $\mathcal{Y}\colon G \to TG$ be an *arbitrary* vector field on G and introduce the auxiliary maps

$$F\colon G \times \mathbb{R} \to \mathbb{R}, \quad (g, t) \mapsto \mathbf{d}_1 f(g, t)(\mathcal{Y}(g)) \qquad \text{and} \qquad H\colon \mathbb{R} \to G \times \mathbb{R}, \quad t \mapsto (g(t), t).$$

Then, $\mathbf{d}_1 f(g(t), t) \equiv 0$ is equivalent to $F \circ H \equiv 0$ that implies

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}(F \circ H)(t) = \mathbf{d}_1 F(H(t))(\dot g(t)) + \mathbf{d}_2 F(H(t))(1) \tag{88}$$

The first term on the right hand of (88) equals

$$\mathbf{d}\big(g \mapsto \mathbf{d}_1 f(g, t)(\mathcal{Y}(g))\big)\big)(g(t))(\dot g(t))$$

that, as $\mathbf{d}_1 f(g(t), t) = 0$, is equivalent to

$$\mathbf{d}\big(g \mapsto \mathbf{d}_1 f(g, t)(\mathcal{Y}(g))\big)\big)(g(t))(\dot g(t)) \\ - \mathbf{d}_1 f(g(t), t)\big(\nabla_{\dot g(t)}(\mathcal{Y}(g(t))\big)\big).$$

This last expression is $\operatorname{Hess}_1 f(g(t), t)(\dot g(t))(\mathcal{Y}(g(t)))$. By simply swapping the order of differentiation, the second term on the right hand of (88) equals

$$\frac{\partial}{\partial t}(\mathbf{d}_1 f)(g(t), t)(\mathcal{Y}(g(t))) = \mathbf{d}_1\Big(\frac{\partial}{\partial t} f\Big)(g(t), t)(\mathcal{Y}(g(t))).$$

Since $\mathcal{Y}$ was arbitrary, the statement follows.

■

Remark. The conclusion in Lemma A.1 depends on $g(t)$ being a critical path for f. As can be seen from the proof, if we were to compute the time derivative along an arbitrary path, we would incur a third term that depends on the choice of affine connection ∇. This is another expression of the known fact that the Hessian operator at a critical point is independent of the choice of connection. □

Lemma A.2: Given $f\colon G \to \mathbb{R}$ then the derivative of left trivialized differential $T_e L_g^*(\mathbf{d} f(g))$ is given by

$$\mathbf{d}\big(T_e L_g^*(\mathbf{d} f(g))\big) = T_e L_g^* \circ \operatorname{Hess} f(g) + \omega^{*\leftleftarrows}_{T_e L_g^*(\mathbf{d} f(g))} \circ T_g L_{g^{-1}},$$

which equals $T_e L_g^* \circ \operatorname{Hess} f(g)$ wherever $\mathbf{d} f(g) = 0$.

Proof: Let $X, Y \in \mathfrak{g}$ then

$$\begin{aligned} &\mathbf{d}\big(g \mapsto \mathbf{d} f(g)(gY)\big)(g)(gX) \\ &= \mathbf{d}\big(g \mapsto \langle \mathbf{d} f(g), gY\rangle\big)(g)(gX) \\ &= \mathbf{d}\big(g \mapsto \langle T_e L_g^*(\mathbf{d} f(g)), Y\rangle\big)(g)(gX) \\ &= \langle \mathbf{d}\big(g \mapsto T_e L_g^*(\mathbf{d} f(g))\big)(g)(gX), Y\rangle \end{aligned} \tag{89}$$

where we have only used the definition and linearity of the duality pairing $\langle\cdot,\cdot\rangle$. Furthermore,

$$\begin{aligned}
&\mathbf{d}\big(g \mapsto \mathbf{d}\, f(g)(gY)\big)(g)(gX)\\
&= \mathrm{Hess}\, f(g)(gX)(gY) + \mathbf{d}\, f(g)\big(\nabla_{gX}(gY)\big)\\
&= \langle \mathrm{Hess}\, f(g)(gX), gY\rangle + \langle \mathbf{d}\, f(g), g\,\omega(X,Y)\rangle\\
&= \langle T_e L_g^* \circ \mathrm{Hess}\, f(g)(gX), Y\rangle + \langle T_e L_g^*(\mathbf{d}\, f(g)), \omega_X(Y)\rangle\\
&= \langle T_e L_g^* \circ \mathrm{Hess}\, f(g)(gX), Y\rangle + \langle \omega_X^*\big(T_e L_g^*(\mathbf{d}\, f(g))\big), Y\rangle. \quad (90)
\end{aligned}$$

Comparing the two equivalent expressions for the derivative of $\mathbf{d}\, f(g)(gY)$ obtained in (89) and (90) and recalling that Y was arbitrary, we obtain

$$\begin{aligned}
&\mathbf{d}\big(g \mapsto T_e L_g^*(\mathbf{d}\, f(g))\big)(g)(gX)\\
&= T_e L_g^* \circ \mathrm{Hess}\, f(g)(gX) + \omega_X^*\big(T_e L_g^*(\mathbf{d}\, f(g))\big)\\
&= T_e L_g^* \circ \mathrm{Hess}\, f(g)(gX) + \omega^{*\leftrightarrows}_{T_e L_g^*(\mathbf{d}\, f(g))}(X)\\
&= T_e L_g^* \circ \mathrm{Hess}\, f(g)(gX) + \omega^{*\leftrightarrows}_{T_e L_g^*(\mathbf{d}\, f(g))} \circ T_g L_{g^{-1}}(gX).
\end{aligned}$$

Since X was arbitrary, the result follows. ■

Lemma A.3: Given $f\colon G \times \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to G$ then

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\big(T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)}\big)\\
&\quad = \nabla^*_{g(t)^{-1}\dot g(t)} \circ T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)}\\
&\quad + T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)} \circ \nabla_{g(t)^{-1}\dot g(t)}\\
&\quad + T_e L_{g(t)}^* \circ \frac{\partial}{\partial t}(\mathrm{Hess}_1\, f)(g(t),t) \circ T_e L_{g(t)} + h.o.t.
\end{aligned}$$

In the above expression, by higher order terms we mean those terms that depend on the third or higher covariant derivative of f.

Proof: Let $X, Y \in \mathfrak{g}$, then

$$\begin{aligned}
&\big(T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)}\big)(X)(Y)\\
&\qquad = \mathrm{Hess}_1\, f(g(t),t)(g(t)X)(g(t)Y). \quad (91)
\end{aligned}$$

Let $\mathcal{X}, \mathcal{Y}\colon G \to TG$ be arbitrary vector fields on G and introduce the auxiliary maps

$$\begin{aligned}
F\colon \quad & G\times\mathbb{R} \;\to\; \mathbb{R}\\
& (g,t) \;\mapsto\; \mathrm{Hess}_1\, f(g,t)(\mathcal{X}(g))(\mathcal{Y}(g))
\end{aligned}$$

and

$$\begin{aligned}
H\colon \quad & \mathbb{R} \;\to\; G\times\mathbb{R}\\
& t \;\mapsto\; (g(t),t).
\end{aligned}$$

By the product rule for covariant differentiation of tensors

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{Hess}_1\, f(g(t),t)(\mathcal{X}(g(t))(\mathcal{Y}(g(t)))\right)\\
&= \frac{\mathrm{d}}{\mathrm{d}t}(F\circ H)(t) = \mathbf{d}_1\, F(H(t))(\dot g(t)) + \mathbf{d}_2\, F(H(t))(1)\\
&= h.o.t. + \mathrm{Hess}_1\, f(g(t),t)\big(\nabla_{\dot g(t)}(\mathcal{X}(g(t)))\big)(\mathcal{Y}(g(t)))\\
&\quad + \mathrm{Hess}_1\, f(g(t),t)(\mathcal{X}(g(t)))\big(\nabla_{\dot g(t)}(\mathcal{Y}(g(t)))\big)\\
&\quad + \frac{\partial}{\partial t}(\mathrm{Hess}_1\, f)(g(t),t)(\mathcal{X}(g(t)))(\mathcal{Y}(g(t)))
\end{aligned}$$

and hence, recalling (91) and that ∇ is left invariant,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\Big(\big(T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)}\big)\Big)(X)(Y)\\
&= \frac{\mathrm{d}}{\mathrm{d}t}\Big(\big(T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)}\big)(X)(Y)\Big)\\
&= \mathrm{Hess}_1\, f(g(t),t)\big(\nabla_{\dot g(t)}(g(t)X)\big)(g(t)Y)\\
&\quad + \mathrm{Hess}_1\, f(g(t),t)(g(t)X)\big(\nabla_{\dot g(t)}(g(t)Y)\big)\\
&\quad + \frac{\partial}{\partial t}(\mathrm{Hess}_1\, f)(g(t),t)(g(t)X)(g(t)Y) + h.o.t.\\
&= T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)}\big(\nabla_{g(t)^{-1}\dot g(t)}(X)\big)(Y)\\
&\quad + T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)}(X)\big(\nabla_{g(t)^{-1}\dot g(t)}(Y)\big)\\
&\quad + T_e L_{g(t)}^* \circ \frac{\partial}{\partial t}(\mathrm{Hess}_1\, f)(g(t),t) \circ T_e L_{g(t)}(X)(Y) + h.o.t.\\
&= T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)} \circ \omega_{g(t)^{-1}\dot g(t)}(X)(Y)\\
&\quad + \omega^*_{g(t)^{-1}\dot g(t)} \circ T_e L_{g(t)}^* \circ \mathrm{Hess}_1\, f(g(t),t) \circ T_e L_{g(t)}(X)(Y)\\
&\quad + T_e L_{g(t)}^* \circ \frac{\partial}{\partial t}(\mathrm{Hess}_1\, f)(g(t),t) \circ T_e L_{g(t)}(X)(Y) + h.o.t.
\end{aligned}$$

Since X and Y were arbitrary, the result follows. ■

Lemma A.4: Let $X \in \mathfrak{g}$ then

$$\mathbf{d}\big(g \mapsto T_e L_g(X)\big) = T_e L_g \circ \omega_X^{\leftrightarrows} \circ T_g L_{g^{-1}}.$$

Proof: Let $Y \in \mathfrak{g}$ then

$$\begin{aligned}
&\mathbf{d}\big(g \mapsto T_e L_g(X)\big)(gY)\\
&\quad = \nabla_{gY}(gX) = g\,\omega(Y,X) = g\,\omega_X^{\leftrightarrows}(Y)\\
&\quad = T_e L_g \circ \omega_X^{\leftrightarrows} \circ T_g L_{g^{-1}}(gY).
\end{aligned}$$

Since Y was arbitrary, the result follows. ■

Lemma A.5: Let V be a vector space, let $A\colon V \to V$ be linear and let $\mathcal{Q}\colon V \to \mathbb{R}$ be a quadratic form with associated symmetric positive definite linear map $Q\colon V \to V^*$ (cf. Section III). Given $f\colon G \to V$ then

$$\mathbf{d}\left(\frac{1}{2}\mathcal{Q}(A(f(g)))\right) = \big(A^* \circ Q \circ A(f(g))\big) \circ \mathbf{d}\, f(g)$$

and

$$\begin{aligned}
\mathrm{Hess}\left(\frac{1}{2}\mathcal{Q}(A(f(g))\right) &= \big(\mathbf{d}\, f(g)\big)^* \circ A^* \circ Q \circ A \circ \mathbf{d}\, f(g)\\
&\quad + \big(A^* \circ Q \circ A(f(g))\big)^{T_g G} \circ \mathrm{Hess}\, f(g),
\end{aligned}$$

where we recall that $(\cdot)^{T_g G}$ denotes the exponential functor.

Proof: Let $X, Y \in \mathfrak{g}$ then

$$\begin{aligned}
&\mathbf{d}\Big(g \mapsto \frac{1}{2}\mathcal{Q}(A(f(g)))\Big)(gX)\\
&= \mathbf{d}\Big(g \mapsto \frac{1}{2}\Big\langle Q(A(f(g))), A(f(g))\Big\rangle\Big)(gX)\\
&= \Big\langle Q(A(f(g))), A\Big(\mathbf{d}\, f(g)(gX)\Big)\Big\rangle\\
&= \big(A^* \circ Q \circ A(f(g))\big) \circ \mathbf{d}\, f(g)(gX),
\end{aligned}$$

where we have used that Q is symmetric in the next to last line. Since X was arbitrary, the first assertion follows.

We now use this result repeatedly, together with the chain rule and the definition of the Hessian operator, to compute

$$\begin{aligned}
&\mathrm{Hess}\Big(g \mapsto \frac{1}{2}\mathcal{Q}(A(f(g)))\Big)(gX)(qY)\\
&= \mathbf{d}\Big(\mathbf{d}\Big(g \mapsto \frac{1}{2}\mathcal{Q}(A(f(g)))\Big)(gY)\Big)(gX)\\
&\quad - \mathbf{d}\Big(g \mapsto \frac{1}{2}\mathcal{Q}(A(f(g)))\Big)\Big(\nabla_{gX}(gY)\Big)\\
&= \mathbf{d}\Big(\big(A^* \circ Q \circ A(f(g))\big) \circ \mathbf{d}\, f(g)(gY)\Big)(gX)\\
&\quad - \big(A^* \circ Q \circ A(f(g))\big) \circ \mathbf{d}\, f(g)\big(\nabla_{gX}(gY)\big)\\
&= \mathbf{d}\Big(A^* \circ Q \circ A(f(g))\Big)(gX) \circ \mathbf{d}\, f(g)(gY)\\
&\quad + \big(A^* \circ Q \circ A(f(g))\big) \circ \mathbf{d}\Big(\mathbf{d}\, f(g)(gY)\Big)(gX)\\
&\quad - \big(A^* \circ Q \circ A(f(g))\big) \circ \mathbf{d}\, f(g)\big(\nabla_{gX}(gY)\big).
\end{aligned}$$

Combining the last two terms yields

$$\big(A^* \circ Q \circ A(f(g))\big) \circ \mathrm{Hess}\, f(g)(gX)(gY),$$

where we need to use the exponential functor (see Section II) in order to drop both arguments. The first term can be rewritten as

$$\begin{aligned}
&\mathbf{d}\Big(A^* \circ Q \circ A(f(g))\Big)(gX) \circ \mathbf{d}\, f(g)(gY)\\
&= \Big(A^* \circ Q \circ A\big(\mathbf{d}\, f(g)(gX)\big)\Big) \circ \mathbf{d}\, f(g)(gY)\\
&= \Big\langle A^* \circ Q \circ A \circ \mathbf{d}\, f(g)(gX), \mathbf{d}\, f(g)(gY)\Big\rangle\\
&= \big(\mathbf{d}\, f(g)\big)^* \circ A^* \circ Q \circ A \circ \mathbf{d}\, f(g)(gX)(gY).
\end{aligned}$$

Since X and Y were arbitrary, the second result follows. This completes the proof. ∎

References

[1] R. E. Mortensen, "Maximum-likelihood recursive nonlinear filtering," *Journal of Optimization Theory and Applications*, vol. 2, no. 6, pp. 386–394, 1968.

[2] O. B. Hijab, "Minimum energy estimation," Ph.D. dissertation, PhD Thesis, University of California, Berkeley, 1980.

[3] A. Jazwinski, *Stochastic processes and filtering theory*. Academic Press, 1970.

[4] A. J. Krener, "The convergence of the minimum energy estimator," in *New Trends in Nonlinear Dynamics and Control and their Applications*, W. Kang, C. Borges, and M. Xiao, Eds. Springer, 2003, pp. 187–208.

[5] A. P. Aguiar and J. P. Hespanha, "Minimum-energy state estimation for systems with perspective outputs," *IEEE Transactions on Automatic Control*, vol. 51, no. 2, pp. 226–241, 2006.

[6] P. Coote, J. Trumpf, R. Mahony, and J. Willems, "Near-optimal deterministic filtering on the unit circle," in *Proceedings of the 48th IEEE Conference on Decision and Control (CDC)*, 2009, pp. 5490–5495.

[7] M. Zamani, J. Trumpf, and R. Mahony, "Near-optimal deterministic filtering on the rotation group," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1411–1414, 2011.

[8] ——, "Minimum-energy filtering for attitude estimation," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2917–2921, 2013.

[9] M. Zamani, "Deterministic attitude and pose filtering, an embedded Lie groups approach," Ph.D. dissertation, Australian National University, 2013.

[10] M. Zamani, J. Trumpf, and R. Mahony, "On the distance to optimality of the geometric approximate minimum-energy attitude filter," in *Proceedings of the American Control Conference (ACC)*, 2014, pp. 4943–4948.

[11] A. Saccon, J. Trumpf, R. Mahony, and A. Aguiar, "Second-order-optimal filters on Lie groups," in *Proceedings of the 52nd IEEE Conference on Decision and Control (CDC)*, 2013, pp. 4434–4441.

[12] J. Berger, A. Neufeld, F. Becker, F. Lenzen, and C. Schnörr, "Second Order Minimum Energy Filtering on SE(3) with Nonlinear Measurement Equations," in *Scale Space and Variational Methods in Computer Vision*. Springer International Publishing, 2015, vol. 9087, pp. 397–409.

[13] W. B. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*. Elsevier, 1986.

[14] R. Abraham, J. E. Marsden, and T. S. Ratiu, *Manifolds, tensor analysis, and applications*. Springer, 1988.

[15] J. M. Lee, *Riemannian manifolds: an introduction to curvature*. Springer, 1997.

[16] V. S. Varadarajan, *Lie groups, Lie algebras, and their representation*. Springer, 1984.

[17] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*. Springer, 1999.

[18] W. Rossmann, *Lie groups: an introduction through linear groups*. Oxford University Press, 2002.

[19] V. Jurdjevic, *Geometric control theory*. Cambridge University Press, 1997.

[20] A. Bloch, *Nonholonomic mechanics and control*. Springer, 2003.

[21] F. Bullo, *Geometric control of mechanical systems*. Springer, 2005.

[22] K. Nomizu, "Invariant affine connections on homogeneous spaces," *American Journal of Mathematics*, vol. 76, no. 1, pp. 33–65, 1954.

[23] A. Saccon, J. Hauser, and A. Aguiar, "Optimal control on Lie groups: the projection operator approach," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2230–2245, 2013.

[24] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[25] R. W. Brockett, "System theory on group manifolds and coset spaces," *SIAM Journal on Control and Optimization*, vol. 10, no. 2, pp. 265–284, 1972.

[26] V. Jurdjevic and H. J. Sussmann, "Control systems on Lie groups," *Journal of Differential Equations*, vol. 12, no. 2, pp. 313–329, 1972.

[27] A. Saccon, A. P. Aguiar, and J. Hauser, "Lie Group Projection Operator Approach: Optimal Control on T SO(3)," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, 2011, pp. 6973–6978.

Alessandro Saccon received the laurea degree cum laude in computer engineering and the Ph.D. degree in control system theory from the University of Padova, Italy, in 2002 and 2006. His thesis received the Claudio Maffezzoni best PhD thesis award by the Politecnico di Milano. After the completion of his PhD studies, he held a research and development position at University of Padova in joint collaboration with Ducati Corse working on control and optimization methods for the exploration of the dynamics of racing motorcycles for virtual prototyping studies until 2009. From 2009 until 2012, he held a post-doctoral research position at the Instituto Superior Tcnico, Lisbon, Portugal, sponsored by the Portuguese Science and Technology Foundation (FCT), working on autonomous marine vehicles modeling, motion planning, and control. Since 2013, he is an Assistant Professor on nonlinear control and robotics at the Department of Mechanical Engineering, Eindhoven University of Technology, the Netherlands. His research interests include modeling, analysis, and control of mechanical systems, nonlinear control theory, and numerical optimal control for exploration of trajectory space of complex and highly manoeuvrable nonlinear systems. Recent work has focused on the development of optimal constrained motion planning strategies for multiple autonomous robotic vehicles and robotics systems with unilateral position constraints.

Jochen Trumpf received the Dipl.-Math. and Dr. rer. nat. degrees in mathematics from the University of Würzburg, Germany, in 1997 and 2002, respectively. He is currently a Fellow in the Research School of Engineering at the Australian National University. His research interests include observer theory and design, linear systems theory and optimisation on manifolds with applications in robotics, computer vision and wireless communication.

Robert Mahony is a Professor in the Research School of Engineering at the Australian National University. He received his BSc in 1989 (applied mathematics and geology) and his PhD in 1995 (systems engineering) both from the Australian National University. He worked firstly as a marine seismic geophysicist and an industrial research scientist before completing a postdoctoral fellowship in France at the Universite de Technologie de Compiegene and a Logan Fellowship at Monash University in Australia. He has held his post at ANU since 2001. His research interests are in non-linear systems theory and optimization with applications in robotics, geometric optimisation techniques and computer vision.

A. Pedro Aguiar received the Licenciatura, M.S. and Ph.D. in electrical and computer engineering from the Instituto Superior Tcnico (IST), Technical University of Lisbon, Portugal in 1994, 1998 and 2002, respectively. Currently, Dr. Aguiar holds an Associate Professor position with the Department of Electrical and Computer Engineering (DEEC), Faculty of Engineering, University of Porto (FEUP). From 2002 to 2005, he was a post-doctoral researcher at the Center for Control, Dynamical-Systems, and Computation at the University of California, Santa Barbara (UCSB). From 2005 to 2012, he was a senior researcher with the Institute for Systems and Robotics at IST (ISR/IST), and an invited assistant professor with the Department of Electrical and Computer Engineering, IST. His research interests include modeling, control, navigation, and guidance of autonomous robotic vehicles, nonlinear control, switched and hybrid systems, tracking, path-following, performance limitations, nonlinear observers, the integration of machine vision with feedback control, networked control, and coordinated/cooperative control of multiple autonomous robotic vehicles.