# Second-Order-Optimal Filters on Lie groups

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

*Abstract*— **We provide an explicit formula for the second-order-optimal nonlinear filter for state estimation of systems on general Lie groups with disturbed measurements of inputs and outputs. Optimality is with respect to a deterministic cost measuring the cumulative energy in the unknown system disturbances (minimum-energy filtering). We show that the resulting filter will depend on the choice of affine connection, thus encoding the nonlinear geometry of the state space. For the case of attitude estimation, where we are given a second order (dynamic) mechanical system on the tangent bundle of the special orthogonal group SO$(3)$, and where we choose the symmetric Cartan-Schouten (0)-connection, the resulting filter has the familiar form of a gradient observer combined with a matrix Riccati differential equation that updates the filter gain.**

## I. Introduction

The problem of state estimation from disturbed measurements of inputs and outputs is one of the basic problems in system theory and has been studied in many different modelling frameworks. Arguably the most prominent approach 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 (Extended Kalman Filter (EKF), 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 $H^\infty$-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 coincides with the Kalman filter [3]. Krener [4] recently 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 matrix vector 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. [7] to systems on the special orthogonal group SO$(3)$ with full state measurements. In a paper published last year, Zamani [8] provided a second-order-optimal minimum-energy filter for attitude estimation (state space SO$(3)$) from vectorial measurements, resulting in a filter that can be interpreted as a geometric correction to the Multiplicative Extended Kalman Filter (MEKF) [9].

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. The resulting filter takes the form of a gradient observer coupled with an operator Riccati differential equation that updates the filter gain. The filter explicitly depends on the choice of affine connection on the state space, thus encoding its nonlinear geometry. 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 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.

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.

## 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 [10], [11], [12] for a review on differentiable manifolds and covariant differentiation and to [13], [14], [15] for a review of the theory of Lie groups and Lie algebras. Many of these topics

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]

are also covered in [16] and [17]. 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 $\mathfrak{g}$ |
| $\mu$ | 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 $V$ |
| $\mathbf{d}_1$, $\mathbf{d}_2$, ... | 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 $\nabla$ |
| $\omega_X\colon \mathfrak{g}\to\mathfrak{g}$ | $\omega_X(Y) = \omega(X,Y)$ |
| $\omega_Y^{\leftrightharpoons}\colon \mathfrak{g}\to\mathfrak{g}$ | $\omega_Y^{\leftrightharpoons}(X) = \omega_X(Y)$ |
| $\omega_X^*\colon \mathfrak{g}^*\to\mathfrak{g}^*$ | $\langle\omega_X^*(\mu), Y\rangle = \langle \mu, \omega_X Y\rangle$ |
| $\omega_\mu^{*\leftrightharpoons}\colon \mathfrak{g}\to\mathfrak{g}^*$ | $\langle\omega_\mu^{*\leftrightharpoons}(X), Y\rangle = \langle\omega_X^*(\mu), Y\rangle = \langle \mu, \omega_X Y\rangle$ |
| $\omega_Y^{\leftrightharpoons *}\colon \mathfrak{g}^*\to\mathfrak{g}^*$ | $\langle\omega_Y^{\leftrightharpoons *}(\mu), X\rangle = \langle\mu, \omega_Y^{\leftrightharpoons} X\rangle = \langle \mu, \omega_X Y\rangle$ |
| $T(X,Y)\in\mathfrak{g}$ | torsion function associated with $\omega$ |
| $T_X\colon \mathfrak{g}\to\mathfrak{g}$ | 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 $\phi$ *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 $\nabla$ on $G$ is fully characterized by its bilinear connection function $\omega\colon \mathfrak{g}\times\mathfrak{g}\to\mathfrak{g}$ through the identity $\nabla_{gX}(gY) = g\,\omega(X,Y)$ [18].

**Swap operator.** The connection function $\omega$ 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 $\omega_X$, $\omega_Y^{\leftrightharpoons}$, $\omega_X^*$, $\omega_\mu^{*\leftrightharpoons}$, 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 $gX, gY \in T_g G$ [19]. 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 $\nabla$ is symmetric [19].

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 $V$ [19]. 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 $\phi$ to the linear map $\phi^W\colon \mathfrak{L}(W,U)\to\mathfrak{L}(W,V)$ defined by $\phi^W(\eta) = \phi\circ\eta$.

## III. Problem formulation

In this section, we introduce the problem of minimum-energy state estimation for systems on Lie groups. The section concludes with a statement of the main result of the paper.

Consider the deterministic system on a Lie group $G$ defined by

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

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 $B\colon \mathbb{R}^d\to\mathfrak{g}$ 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 $\lambda(g,u,t) = u$ leading to the classical nominal left-invariant kinematics $\dot g = gu$ [20], [21]. In this case the system 'input' is the measured velocity and the model error $\delta$ is best thought of as measurement error associated with inexact measurement of the velocity.

In the case of a dynamical 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 [16], [17]. In this case the model error $\delta$ 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], [9].

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

$$y(t) = h(g(t), t) + D\varepsilon(t) \tag{2}$$

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, $\delta$ and $\varepsilon$, 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 $\delta$ and $\varepsilon$, along with some form of initial cost (initial "energy") in $g_0$, leading to the terminology of minimum energy filtering.

Define two quadratic forms

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

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

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

The constant $\alpha$ 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 $g_0$.

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

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 is a necessary part of the development and cannot be ignored since 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

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

Note that $J$ depends on the values of the signals $\delta$ and $\varepsilon$ 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 $\tau \neq t$. That is 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 equations.*

In order to write down the filter equations 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

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

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 details how to obtain it, revisiting the optimal estimation problem, its corresponding optimal Hamiltonian and the associated Hamilton-Jacobi-Bellman (HJB) equation.

### A. The second-order-optimal filter equations

Assume that the error terms $\delta$ and $\varepsilon$ 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

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

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

$$\widehat{g}_0 = \arg\min_{g\in G}\, m_0(g), \tag{9}$$

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

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

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

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

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

$$X_0 = T_eL^*_{\hat{g}_0}\circ \operatorname{Hess}\ m_0(\hat{g}_0)\circ T_eL_{\hat{g}_0}, \tag{12}$$

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

and

$$\begin{aligned}E(t) = -T_eL^*_{\hat{g}}\circ\Big[\Big(\big(D^{-1}\big)^*\circ Q\circ D^{-1}\big(y - h_t(\hat{g})\big)\Big)\overset{T_{\hat{g}}G}{\circ} \\ \operatorname{Hess} h_t(\hat{g}) - \big(\mathbf{d}\,h_t(\hat{g})\big)^*\!\circ\big(D^{-1}\big)^*\!\circ Q\circ D^{-1}\circ\mathbf{d}\,h_t(\hat{g})\big)\Big]\circ T_eL_{\hat{g}},\end{aligned} \tag{14}$$

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

A coordinate version of the operator statement of the Riccati equation can be obtained by choosing bases for $\mathfrak{g}$, $\mathbb{R}^d$, and $\mathbb{R}^p$. We provide a worked example in Section V. 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

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

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 $\tau\in[t_0,t]$.

Note how the control input in the above optimal estimation problem is the model error $\delta$ 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 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 truncation of 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) [9], see also [22].

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 [23] 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 (time-varying) Hamiltonian $\tilde{H}\colon T^*G\times\mathbb{R}^d\times\mathbb{R}\to\mathbb{R}$ is given by

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

where $g$ is the base point of $p\in T^*_gG\subset T^*G$. 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 typical for optimal control problems defined on Lie groups [23], the cotangent vector $p\in T^*_gG$ can be identified via left translation with the element $\mu\in\mathfrak{g}^*$, defined as $\mu = T_eL^*_g(p)$. 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

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

We are now ready to compute the left-trivialized *optimal* Hamiltonian that characterizes the optimal estimation 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 estimation problem (15)-(16) is given by

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

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

### 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

$$\frac{\partial}{\partial t}V(g,t) - H^-\left(g, T_eL_g^*\big(\mathbf{d}_1 V(g,t)\big),t\right) = 0. \tag{23}$$

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

$$\widehat{g}(t) := \arg\min_{g\in G} V(g,t). \tag{24}$$

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

$$\mathbf{d}_1 V(\widehat{g}(t),t) = 0 \tag{25}$$

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

$$\operatorname{Hess}_1 V(\hat{g}(t),t)\left(\dot{\hat{g}}(t)\right) + \mathbf{d}_1(\frac{\partial}{\partial t}V)(\hat{g}(t),t) = 0 \tag{26}$$

for $t\geq 0$. Here, $\operatorname{Hess}_1 V(\hat{g}(t),t)\colon T_{\hat{g}(t)}G \to T^*_{\hat{g}(t)}G$ 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), we have

$$\begin{aligned}\mathbf{d}_1(\frac{\partial}{\partial t}V)(g,t) =& \,\mathbf{d}_1 H^-\big(g,\mu(g,t),t\big) + \\ & \mathbf{d}_2 H^-\big(g,\mu(g,t),t\big)\circ \mathbf{d}_1\mu(g,t),\end{aligned} \tag{27}$$

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

$$\mu(g,t) := T_eL_g^*\big(\mathbf{d}_1 V(g,t)\big). \tag{28}$$

Using (25) this yields

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

Now,

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

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

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

since the second term vanishes due to (25). Define $Z(g,t)\colon\mathfrak{g}\to\mathfrak{g}^*$ as

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

then we can rewrite

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

and similarly

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

Moreover,

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

Using (29), (33), (34) and (35), equation (26) can be rewritten as

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

for $t\geq 0$. By equation (25), the point $\hat{g}(t)\in G$ is a critical point of the value function, and hence the Hessian $\operatorname{Hess}_1 V(\hat{g}(t),t)$ 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 $\operatorname{Hess}_1 V(\hat{g}(t),t)$

and hence $Z(\hat{g}(t),t)$ is invertible, then equation (36) is equivalent to

$$\hat{g}(t)^{-1}\dot{\hat{g}}(t) = -\,\mathbf{d}_2\,H^-(\hat{g}(t),0,t) - Z(\hat{g}(t),t)^{-1}\circ T_eL^*_{\hat{g}(t)}\big(\mathbf{d}_1\,H^-(\hat{g}(t),0,t)\big). \tag{37}$$

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

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

We obtain

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

and

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

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 $r_t(\widehat{g}) = r(\widehat{g}(t),t)\in\mathfrak{g}^*$ by

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

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). Unfortunally, 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

$$\dot{X} = e^{-\alpha(t-t_0)}\cdot S - A^*\circ X - X\circ A - e^{\alpha(t-t_0)}\cdot X\circ B\circ R^{-1}\circ B^*\circ X, \qquad X(t_0)=X_0 \tag{44}$$

with

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

$$A(t) = -\,\omega_{\hat{g}^{-1}\dot{\hat{g}}} + \omega_{\overleftarrow{\lambda_t(\hat{g},u)}} + \mathbf{d}_1\,\lambda_t(\hat{g},u)\circ T_eL_{\hat{g}}, \tag{46}$$

$$S(t) = -T_eL^*_{\hat{g}}\circ \Bigg(\Big(\big(D^{-1}\big)^*\circ Q\circ D^{-1}\big(y-h_t(\hat{g})\big)\Big)^{T_{\hat{g}}G}\circ\text{Hess}\,h_t(\hat{g}) + -\big(\mathbf{d}\,h_t(\hat{g})\big)^*\circ\big(D^{-1}\big)^*\circ Q\circ D^{-1}\circ\mathbf{d}\,h_t(\hat{g})\Bigg)\circ T_eL_{\hat{g}}, \tag{47}$$

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

*Proof:* From (32), setting $g=\hat{g}(t)$, we get

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

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 $\omega$ and $\omega^*$. The partial time derivative commutes with covariant differentiation on $G$, so we can use equation (27) 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 (27) into the following form,

$$\mathbf{d}_1(\frac{\partial}{\partial t}V)(g,t) = \mathbf{d}_1\,H^-\big(g,\mu(g,t),t\big) + \big(\mathbf{d}_1\,\mu(g,t)\big)^*\big(\mathbf{d}_2\,H^-\big(g,\mu(g,t),t\big)\big). \tag{49}$$

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

$$\big(\mathbf{d}_1\,\mu(g,t)\big)^*(W) = \big(\text{Hess}_1\,V(g,t)\big)^*\circ T_eL_g(W) + T_gL^*_{g^{-1}}\circ\omega^*_{\overleftarrow{W}}\big(T_eL^*_g(\mathbf{d}_1\,V(g,t))\big),$$

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

$$\begin{aligned}\mathbf{d}_1(\frac{\partial}{\partial t}V)(g,t) &= \mathbf{d}_1\,H^-\big(g,\mu(g,t),t\big) +\\ &\quad\big(\text{Hess}_1\,V(g,t)\big)^*\circ T_eL_g\big(\mathbf{d}_2\,H^-\big(g,\mu(g,t),t\big)\big) +\\ &\quad T_gL^*_{g^{-1}}\circ\omega^*_{\overleftarrow{\mathbf{d}_2\,H^-(g,\mu(g,t),t)}}\big(T_eL^*_g(\mathbf{d}_1\,V(g,t))\big).\end{aligned}$$

Then

$$\begin{aligned}\frac{\partial}{\partial t}(\text{Hess}_1\,V)(\hat{g}(t),t) &= \text{Hess}_1(\frac{\partial}{\partial t}V)(\hat{g}(t),t) =\\ &\text{Hess}_1\,H^-(\hat{g},0,t) + \mathbf{d}_2(\mathbf{d}_1\,H^-)(\hat{g},0,t)\circ\mathbf{d}_1\,\mu(\hat{g},t) +\\ &\text{Hess}_1\,V(\hat{g},t)\circ T_eL_{\hat{g}}\circ\omega_{\overleftarrow{\mathbf{d}_2\,H^-(\hat{g},0,t)}}\circ T_{\hat{g}}L_{\hat{g}^{-1}} +\\ &\text{Hess}_1\,V(\hat{g},t)\circ T_eL_{\hat{g}}\circ\mathbf{d}_1(\mathbf{d}_2\,H^-)(\hat{g},0,t) +\\ &\text{Hess}_1\,V(\hat{g},t)\circ T_eL_{\hat{g}}\circ\text{Hess}_2\,H^-(\hat{g},0,t)\circ\mathbf{d}_1\,\mu(\hat{g},t) +\\ &T_{\hat{g}}L^*_{\hat{g}^{-1}}\circ\omega^*_{\overleftarrow{\mathbf{d}_2\,H^-(\hat{g},0,t)}}\circ T_eL^*_{\hat{g}}\circ\text{Hess}_1\,V(\hat{g},t) +\\ &h.o.t.\end{aligned} \tag{50}$$

Here we have used equation (25) 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

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

Differentiating equation (39) we obtain

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

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

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

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

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

with $S(t)$ and $A(t)$ as in (47) and (46), respectively. As we neglect the third order derivative of $V$, 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 near 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 a second-order-optimal filter for the rotational dynamics of a rigid body subject to external torques, assuming some directional measurements are available.

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 $\tau$ 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 a standard practice to identify $T\mathrm{SO}(3)$ with $\mathrm{SO}(3)\times\mathbb{R}^3$ via left translation [14] 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 to have *two* (possibly time-varying) reference direction measurements $\mathring{a}_1$ and $\mathring{a}_2 \in \mathbb{S}^2$ corrupted by measurement noise ($\mathbb{S}^2$ 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

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

where $\varepsilon$ represents the unknown measurement error. Equation (57) has the structure of (2), where, for ease of presentation, we assume that $D$ equals the identity.

As error model for the dynamics, we choose

$$\mathbf{R}^T\dot{\mathbf{R}} = (\mathbf{\Omega})^\times, \tag{58}$$

$$\dot{\mathbf{\Omega}} = \mathbb{I}^{-1}\left((\mathbb{I}\mathbf{\Omega})^\times\mathbf{\Omega} + \tau\right) + \delta. \tag{59}$$

Equations (58)-(59) are a particular case of (1) where $\delta \in \mathbb{R}^3$ and $B : \mathbb{R}^3 \to \mathbb{R}^3_\times \times \mathbb{R}^3$, $\delta \mapsto (0,\delta)$.

We assume that the optimal filtering problem is posed in terms of the minimization of 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

$$Q = \begin{bmatrix} q_1 I_{3\times 3} & 0 \\ 0 & q_2 I_{3\times 3}\end{bmatrix}, \tag{60}$$

with $q_1$ and $q_2$ strictly positive constants.

By choosing different group operations, we can assign to tangent bundle $T\mathrm{SO}(3) \approx \mathrm{SO}(3)\times\mathbb{R}^3$ different Lie group structures. Here, we follow the approach detailed in, e.g., [24], [25] 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

$$(\mathbf{R},\mathbf{X})\cdot(\mathbf{S},\mathbf{Y}) = (\mathbf{RS},\mathbf{X}+\mathbf{Y}). \tag{61}$$

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

$(\zeta^R, \zeta^\Omega)$ 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

$$\mathrm{ad}_{(\zeta^R,\zeta^\Omega)}(\xi^R, \xi^\Omega) = (\zeta^R \times \xi^R, 0). \quad (62)$$

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

$$\mathrm{ad}_{(\zeta^R,\zeta^\Omega)} = \begin{bmatrix} (\zeta^R)^\times & 0 \\ 0 & 0 \end{bmatrix}. \quad (63)$$

A (left-invariant) connection has to be chosen to derive the second order optimal filter on a Lie group. Here, as done in [24], [25], we make use of the symmetric (0) Cartan-Schouten connection, characterized by the connection function

$$\omega^{(0)} = \frac{1}{2}\,\mathrm{ad}. \quad (64)$$

*Proposition 5.1:* The second order optimal filter for the system (55)-(55) with dynamics error model (58)-(59), measurement output model (57), and incremental cost (4) with block diagonal structure given by (60) is given by

$$\widehat{\mathbf{R}}^T \dot{\widehat{\mathbf{R}}} = \big(\widehat{\mathbf{\Omega}} + K_{11}\, r^R + K_{12}\, r^\Omega\big)^\times \quad (65)$$

$$\dot{\widehat{\mathbf{\Omega}}}, = \mathbb{I}^{-1}\big( (\mathbb{I}\widehat{\mathbf{\Omega}})^\times \widehat{\mathbf{\Omega}} + \tau\big) + K_{21}\, r^R + K_{22}\, r^\Omega, \quad (66)$$

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

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

The residual $r_t$ is given by

$$r_t = \begin{bmatrix} r^R \\ r^\Omega \end{bmatrix} = \begin{bmatrix} -q_1(\hat{a}_1 \times a_1) - q_2(\hat{a}_2 \times a_2) \\ 0 \end{bmatrix}, \quad (67)$$

while the gain $K$ is the solution of the perturbed Riccati equation (11) where, due to the choice of the (0) connection,

$$A = \begin{bmatrix} -\widehat{\mathbf{\Omega}}^\times & I \\ 0 & \mathbb{I}^{-1}\left[ (\mathbb{I}\widehat{\mathbf{\Omega}})^\times - \widehat{\mathbf{\Omega}}^\times \mathbb{I} \right]. \end{bmatrix}, \quad (68)$$

$$E = \begin{bmatrix} \sum_{i=1}^2 -q_i\big(\hat{a}_i^\times a_i^\times + a_i^\times \hat{a}_i^\times\big)/2 & 0 \\ 0 & 0_{3\times3} \end{bmatrix}, \quad (69)$$

$$BR^{-1}B^* = \begin{bmatrix} 0_{3\times3} & 0 \\ 0 & R^{-1} \end{bmatrix}, \quad \text{and} \quad (70)$$

$$\omega_{Kr} = \begin{bmatrix} 1/2(K_{11} r^R + K_{12} r^\Omega)^\times & 0 \\ 0 & 0_{3\times3} \end{bmatrix}. \quad (71)$$

The reported filter equations result from straightforward application of the theory presented in this paper. Space limitations prevent us to provide the explicit derivation of these formulas, as well as numerical simulations and comparisons with other existing schemes.

## 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.

## 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] ——, "A second order minimum-energy filter on the special orthogonal group," in *Proceedings of the 2012 American Control Conference (ACC)*, 2012, pp. 1895–1900.

[9] ——, "Minimum-energy filtering for attitude estimation," *IEEE Transactions on Automatic Control*, to appear, accepted for publication 24 January 2013.

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

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

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

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

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

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

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

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

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

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

[20] 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.

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

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

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

[24] 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)*, Orlando, FL, USA, 2011, pp. 6973–6978.

[25] 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.