## 0.1 Invariant EKF

> **Definition 0.1** — Let $G$ be a Lie group and $f_u : G \to TG$ be a input-dependent vector field. Then $f_u$ is **group-affine** if it satisfies
>
> $$f_u(ab) = f_u(a)b + af_u(b) - af_u(Id)b \in T_{ab}G.$$

Group-affine is the weakest assumption one needs on the dynamics to have trajectory-independent error with log-linear error dynamics.

Given two state trajectories $X, \widehat{X}$ that follows the same deterministic dynamics, let $\eta^L := X^{-1}\widehat{X}$ and $\eta^R := \widehat{X}X^{-1}$ be their left and right errors. Then the first theorem says that $f_u$ is group affine iff $\eta^L$ is state trajectory-independent iff $\eta^R$ is state trajectory-independent.

> **Theorem 0.2**
>
> Let $i = L, R$. Let $\xi_0^i = (\log(\eta_0^i))^\vee$. Define $A^i$ as
>
> $$g_u(\exp(\xi)) = \left(A^i\xi\right)^\vee + O(\|\xi\|^2).$$
>
> If $\xi^i$ is defined $\forall\, t > 0$ by the linear diffeq in $\mathbb{R}^n$ :
>
> $$\frac{d}{dt}\xi^i = A^i\xi^i.$$
>
> Then for the true nonlinear error $\eta^i$, we have $\eta^i = \exp(\xi^i) \,\forall\, t \geq 0$.

Let $\eta = X^{-1}\widehat{X}$ be the left-invariant error. In EKF, $\widehat{X}$ represents estimated state and $X$ is the true state, where they share the same deterministic dynamics and only differ in initial state. In the context of uncertainty propagation, $X$ is modeled by a stochastic process governed by an SDE, and $\widehat{X}$ is the propagated mean.

When the dynamics $f_u$ is group-affine, the evolution of the error $\eta$ satisfies

$$\frac{d}{dt}\eta = g_u(\eta)$$

where $g_u$ satisfies $g_u(\eta) = f_u(\eta) - f_u(Id)\eta$. It turns out that $g_u$ can be converted to a linear dynamic under the exponential coordinates. Let $(A\xi)^\wedge$ be the first-order approximation of

$g_u(\exp(\xi))$. Then we have $\eta = \exp(\xi)$ for all time where

$$\frac{d}{dt}\xi = A\xi.$$

This means that the error $\eta$ is trajectory-independent and can be linearized in the Lie algebra without approximation.

Note that if $f_u$ is affine (*i.e.* $f_u(x) = f_0 + Ax + Bu$), we know that the error has the linear dynamics $\dot{e} = Ae$, independent of the state-trajectory. This is a generalization of that.

So in a stochastic setting, if the drift of SDE is group-affine, then no matter how state deviates from the mean, the drift error is trajectory-independent and can be propagated using the same ODE in the Lie algebra. If the drift is not group-affine, then using the drift equation (Ricatti?) along the mean trajectory should give additional error.

Euler-Poincare Equation:

$$\mathbb{I}\dot{\xi} = \mathrm{ad}^*_\xi(\mathbb{I}\xi) + u$$

This can be linearized and then used for trajectory tracking. This requires already having a nominal trajectory, but in covariance steering we have to produce the nominal.

## 0.2 Uncertainty Propagation on Lie Groups

In the Euclidean case, an SDE $dx = h(x,t)dt + H(t)dW$ has the following evolution of the first two moments:

$$\begin{cases} \dot{\mu} = \langle h \rangle, \\ \dot{\Sigma} = \left\langle h(x-\mu)^T + (x-\mu)h^T \right\rangle + HH^T. \end{cases}$$

We can define SDE on $G$ in three different ways. The first is the McKean-Gangolli injection:

$$g(t + \mathrm{d}t) = g(t)\exp(h(g(t),t)\mathrm{d}t + H(g(t+k\mathrm{d}t), t + k\mathrm{d}t)\mathrm{d}W).$$

If $k = 0$, then the SDE is Ito's. If $k = \frac{1}{2}$, then it is Stratonovich's, which we distinguish by using ⓢ.

Stratonovich is natural and obeys the geometry and chain rule. However, it is extremely difficult to evaluate the expectation TODO. Thus, we rely on Ito's to compute expectations and then convert to Stratonovich's if needed. The conversion is

$$h = h^s + \frac{1}{2} E_i^r \left( H_{kj}^s \right) H_{ij}^s e_k, \qquad H = H^s.$$

We shall use the simplified notation

$$(g^{-1}\mathrm{d}g)^\vee = h\mathrm{d}t + H\mathrm{d}W$$

$$(g^{-1}\mathrm{d}g)^\vee = h^s\mathrm{d}t + H^s \text{\small S}\mathrm{d}W.$$

The second way is to parametrize group element $g = g(q)$ and write

$$q(t + \mathrm{d}t) = q(t) + (J_r^{-1}\widetilde{h})(q, t)\mathrm{d}t + (J_r^{-1}\widetilde{H})(q(t + k\mathrm{d}t), t + k\mathrm{d}t)\mathrm{d}W.$$

Using the exponential coordinates $g(x) = \mu(t)\exp(x)$, under Stratonovich interpretation, to describe the same stochastic process, the drift and diffusion of both ways equal each other. Under Ito's interpretation, there is an additional drift term coming from nonlinearity of exponential map.

$$\widetilde{h}(x, t) = h(x, t) + \left( \frac{1}{2} J_r \frac{\partial J_r^{-1}}{\partial x_k} H H^T J_r^{-T} \right) \bigg|_{g=\mu\exp((x))} e_k,$$

$$\widetilde{H}(x, t) = H(\mu\exp(x), t).$$

Note that if $H$ does not depend on the state, then Ito's and Stranotovich coincide for $H$. However, $J_r$ could still be different under the two interpretations.

In reality, we are given $h^s$ and $H^s$ TODO, which describe an SDE in the tangent space at each $g$. We would have to compute the $h, H$ or $\widetilde{h}, \widetilde{H}$ that would yield the same stochastic process.

If $g$ is not near the identity, then the exponential coordinates might not be convenient to use because of high nonlinearity of exponential map away from the identity and non-bijectiveness of the exponential map. In this case, a moving coordinate frame by re-centering solves the

3

issue. However, a moving frame makes it harder to do covariance steering. We might have to do covariance steering on the error space.

The goal is to derive the error mean and covariance dynamics using the exponential coordinates. We can achieve this via computing the Fokker-Planck equation of

**Definition 0.3** — Let $p$ be a probability density function on an unimodular Lie group $G$. The **group-theoretic mean** $\mu$ is defined by

$$\int_G \log^\vee \left(\mu^{-1}g\right) p(g)\mathrm{d}g = 0.$$

The covariance of $p$ is defined by

$$\Sigma = \int_G \left[\log^\vee(\mu^{-1}g)\right] \left[\log^\vee(\mu^{-1}g)\right]^T p(g)\mathrm{d}g.$$

**Theorem 0.4**

Let $x \in \mathbb{R}^N$ be a random variable with concentrated probability distribution around the origin of $D$. Let its mean, covariance, and probability density function be $m, \Sigma, \tilde{p}$. Then the random variable defined by $g = \mu \exp(x)$ obeys a distribution whose group-theoretic mean $\mu_m$ and covariance $\Sigma_m$ are estimated by

$$m' = \left\langle J_\ell^{-1}\right\rangle^{-1} m$$

$$\mu_m = \mu \exp(m' + O(|m'|^2))$$

$$\Sigma_m = \Sigma - \text{sym}\left(\langle J_{\ell^{-1}m'x^T}\rangle\right) + O(|m'|^2).$$

**Theorem 0.5**

The group-theoretic mean $\mu(t)$ and covariance $\Sigma(t)$ of a stochastic process $g(t)$ described by the McKean-Gangolli injection Ito's SDE obey the following equations:

$$\left(\mu^{-1}\dot{\mu}\right)^\vee \approx \left\langle J_\ell^{-1}\right\rangle^{-1} \left\langle \frac{1}{2}\frac{\partial J_r^{-1}}{\partial x_k}\left(HH^T J_r^{-T}e_k\right) + J_r^{-1}h^c\right\rangle,$$

$$\dot{\Sigma} \approx \left\langle \text{sym}\left[\left(\frac{1}{2}\frac{\partial J_r^{-1}}{\partial x_k}\left(HH^T J_r^{-T}\right)e_k - J_\ell^{-1}(\mu^{-1}\dot{\mu})^\vee + J_r^{-1}h^c\right)x^T\right] + J_r^{-1}HH^T J_r^{-T}\right\rangle$$

Note the resemblance of this with its Euclidean counterpart.

$$J_l(q) = \left[\left(\frac{\partial g}{\partial q_1}g^{-1}\right)^{\vee}, \cdots, \left(\frac{\partial g}{\partial q_n}g^{-1}\right)^{\vee}\right],$$

$$J_r(q) = \left[\left(g^{-1}\frac{\partial g}{\partial q_1}\right)^{\vee}, \cdots, \left(g^{-1}\frac{\partial g}{\partial q_n}\right)^{\vee}\right]$$

We shall use the Taylor second-order approximation of inverse Jacobians:

$$J_\ell^{-1}(x) = I - \frac{1}{2}[\mathrm{ad}_X] + \frac{1}{12}[\mathrm{ad}_X]^2 - \ldots$$

$$J_r^{-1}(x) = I + \frac{1}{2}[\mathrm{ad}_X] + \frac{1}{12}[\mathrm{ad}_X]^2 - \ldots$$

We use the following tricks for the derivations:

(1) Any linear term of $x$ inside the expectation vanishes because we recenter the mean to 0 at all times.

(2) For any symmetric matrix $S$, we have $\mathrm{ad}_i Se_i = 0$.

(3) The expansion of moments to approximate expectation (with hard-to-bound error):

$$\langle M(x)\rangle = \left\langle A_0 + A_1^i x_i + A_2^{ij} x_i x_j + \ldots\right\rangle \approx A_0 + A_2^{ij}\langle x_i x_j\rangle.$$

(4) As long as $(I-A)^n \to 0$ converges ($A$ is close to $I$), $A^{-1} = \sum_{k=0}^{\infty}(I-A)^k$. For $\left\langle J_l^{-1}\right\rangle^{-1}$, since the 1st order term vanish under expectation, setting $k=1$ gives the 2nd order approximation.

(5) Since ad is linear in each argument, $\mathrm{ad}_X = \mathrm{ad}_{(x_i E_i)} = x_i\,\mathrm{ad}_i$.

We compute and discard terms with order 3 or higher:

$$\left(\mu^{-1}\dot\mu\right)^{\vee} \approx \left\langle \underbrace{I - \frac{1}{2}[\mathrm{ad}_X]}_{(1)} + \frac{1}{12}[\mathrm{ad}_X]^2\right\rangle^{-1}\left\langle \frac{1}{2}\left(\frac{1}{2}\,\mathrm{ad}_k + \frac{1}{12}\left(\mathrm{ad}_k[\mathrm{ad}_X] + [\mathrm{ad}_X]\,\mathrm{ad}_k\right)\right)HH^T\ldots\right.$$

$$\left(I + \frac{1}{2}[\mathrm{ad}_X] + \frac{1}{12}[\mathrm{ad}_X]^2\right)^T e_k + \left(I + \frac{1}{2}[\mathrm{ad}_X] + \frac{1}{12}[\mathrm{ad}_X]^2\right)\left(h + \frac{\partial h^c}{\partial x_i}x_i + \frac{1}{2}\frac{\partial^2 h^c}{\partial x_i \partial x_j}x_i x_j\right)\right\rangle$$

$$\approx \underbrace{\left(I - \frac{1}{12}\left\langle[\mathrm{ad}_X]^2\right\rangle\right)}_{(4)}\left\langle \underbrace{\frac{1}{4}\,\mathrm{ad}_k HH^T e_k}_{(2)} + \underbrace{\frac{1}{8}\,\mathrm{ad}_k HH^T[\mathrm{ad}_X]^T e_k + \frac{1}{48}\,\mathrm{ad}_k HH^T([\mathrm{ad}_X]^2)^T e_k}_{(1)}\right.$$

$$+ \underbrace{\frac{1}{24}(\mathrm{ad}_k[\mathrm{ad}_X] + [\mathrm{ad}_X]\,\mathrm{ad}_k)HH^T e_k}_{(1)} + \underbrace{\frac{1}{48}\,\mathrm{ad}_k[\mathrm{ad}_X]HH^T[\mathrm{ad}_X]^T e_k + \frac{1}{48}[\mathrm{ad}_X]\,\mathrm{ad}_k HH^T[\mathrm{ad}_X]^T e_k}_{(2)}$$

5

$$+ h + \underbrace{\frac{\partial h^c}{\partial x_i} x_i}_{(1)} + \frac{1}{2}\frac{\partial^2 h^c}{\partial x_i \partial x_j} x_i x_j + \underbrace{\frac{1}{2}[\mathrm{ad}_X]h}_{(1)} + \frac{1}{2}x_j[\mathrm{ad}_X]\frac{\partial h^c}{\partial x_j} + \frac{1}{12}[\mathrm{ad}_X]^2 h \Big\rangle$$

$$\approx h + \Big\langle \frac{1}{48}\,\mathrm{ad}_k\, HH^T ([\mathrm{ad}_X]^2)^T e_k + \frac{1}{48}[\mathrm{ad}_X]\,\mathrm{ad}_k\, HH^T[\mathrm{ad}_X]^T e_k + \frac{1}{2}\frac{\partial^2 h^c}{\partial x_i \partial x_j}x_i x_j$$

$$+ \frac{1}{2}x_j[\mathrm{ad}_X]\frac{\partial h^c}{\partial x_j} + \underbrace{\frac{1}{12}[\mathrm{ad}_X]^2 h - \frac{1}{12}\langle[\mathrm{ad}_X]^2\rangle h}_{\text{cancel after next step}} \Big\rangle$$

$$\approx h + \Big( \frac{1}{48}\,\mathrm{ad}_k\, HH^T\, \mathrm{ad}_j^T\, \mathrm{ad}_i^T\, e_k + \frac{1}{48}\,\mathrm{ad}_i\, \mathrm{ad}_k\, HH^T\, \mathrm{ad}_j^T\, e_k + \frac{1}{2}\frac{\partial^2 h^c}{\partial x_i \partial x_j}$$

$$+ \frac{1}{2}\,\mathrm{ad}_i\, \frac{\partial h^c}{\partial x_j} \Big) \langle x_i x_j\rangle \qquad\qquad \text{use (5) for all and then use (3)}$$

$$=: h + M_{ij}^\mu \Sigma_{ij}.$$

For the covariance, the same simplifications apply. Keeping only up to second-order terms, we have

$$\dot\Sigma \approx \Big\langle \mathrm{sym}\Big\{ \Big[\frac{1}{2}\Big(\frac{1}{2}\mathrm{ad}_k + \frac{1}{12}(\mathrm{ad}_k[\mathrm{ad}_X] + \underbrace{[\mathrm{ad}_X]\,\mathrm{ad}_k}_{(2)})\Big) HH^T \Big(I + \frac{1}{2}[\mathrm{ad}_X]\Big)^T e_k$$

$$- \Big(I - \frac{1}{2}[\mathrm{ad}_X]\Big)\Big(\mu^{-1}\dot\mu\Big)^\vee + \Big(I + \frac{1}{2}[\mathrm{ad}_X]\Big)\Big(h + \frac{\partial h^c}{\partial x_i}x_i\Big)\Big]x^T\Big\}$$

$$+ \Big(I + \frac{1}{2}[\mathrm{ad}_X] + \frac{1}{12}[\mathrm{ad}_X]^2\Big) HH^T \Big(I + \frac{1}{2}[\mathrm{ad}_X] + \frac{1}{12}[\mathrm{ad}_X]^2\Big)\Big\rangle$$

$$\approx HH^T + \Big\langle \mathrm{sym}\Big\{ \Big[\frac{1}{8}\mathrm{ad}_k\, HH^T[\mathrm{ad}_X]^T e_k + \frac{1}{24}\mathrm{ad}_k[\mathrm{ad}_X]HH^T e_k + \frac{1}{2}[\mathrm{ad}_X]\Big(h + \big(\mu^{-1}\dot\mu\big)^\vee\Big) + \frac{\partial h^c}{\partial x_i}x_i\Big]x^T$$

$$+ \frac{1}{12}[\mathrm{ad}_X]^2 HH^T\Big\} + \frac{1}{4}[\mathrm{ad}_X]HH^T[\mathrm{ad}_X]^T \Big\rangle$$

$$\approx HH^T + \Big( \mathrm{sym}\Big\{ \Big[\frac{1}{8}\mathrm{ad}_k\, HH^T\, \mathrm{ad}_i^T\, e_k + \frac{1}{24}\mathrm{ad}_k\, \mathrm{ad}_i\, HH^T e_k + \frac{1}{2}\mathrm{ad}_i\Big(h + \big(\mu^{-1}\dot\mu\big)^\vee\Big) + \frac{\partial h^c}{\partial x_i}\Big]e_j^T$$

$$+ \frac{1}{12}\mathrm{ad}_i\, \mathrm{ad}_j\, HH^T\Big\} + \frac{1}{4}\mathrm{ad}_i\, HH^T\, \mathrm{ad}_j^T \Big) \langle x_i x_j\rangle$$

$$=: HH^T + M_{ij}^\Sigma \Sigma_{ij}.$$

$$M_{ij}^\mu = \frac{1}{48}\mathrm{ad}_i\, \mathrm{ad}_k\, HH^T\, \mathrm{ad}_j^T\, e_k + \frac{1}{48}\mathrm{ad}_k\, HH^T\, \mathrm{ad}_j^T\, \mathrm{ad}_i^T\, e_k + \frac{1}{2}\mathrm{ad}_i\, \frac{\partial h^c}{\partial x_j} + \frac{1}{2}\frac{\partial^2 h^c}{\partial x_i \partial x_j}$$

$$M_{ij}^{\Sigma} = \text{sym} \left\{ \left[ \frac{1}{8} \operatorname{ad}_k HH^T \operatorname{ad}_i^T e_k + \frac{1}{24} \operatorname{ad}_k \operatorname{ad}_i HH^T e_k + \frac{1}{2} \operatorname{ad}_i \left( h + \left( \mu^{-1}\dot{\mu} \right)^\vee \right) + \frac{\partial h^c}{\partial x_i} \right] e_j^T \right.$$
$$\left. + \frac{1}{12} \operatorname{ad}_i \operatorname{ad}_j HH^T \right\} + \frac{1}{4} \operatorname{ad}_i HH^T \operatorname{ad}_j^T .$$

We note that the only difference between equations from right FPE and from left FPE from the paper is that right FPE equations has all positive signs.

Let us now consider the Euler-Poincaré attitude dynamics with linear feedback control. The Lie group of interest is $G = \mathrm{SO}(3) \ltimes \mathbb{R}^3$, where the space of angular momenta $\mathbb{R}^3$, an additive Lie group, is identified with $\mathfrak{so}(3)$ via the skew-symmetric operations $(\cdot)^\vee$ and $(\cdot)^\wedge$. Let $R$ be the attitude in matrix form, $\ell$ be the body-frame angular momentum, and $\omega$ be the body-frame angular velocity, so $g = (R, \ell)$ with $\ell = \mathbb{I}\omega$. Next, we present the exponential map $\exp : \mathfrak{g} \supset U \to V \subset G$ of $G$. We also denote the exponential map of $\mathrm{SO}(3)$ as $\exp(\xi) := \operatorname{expm}(\xi^\wedge)$. The input should resolve any ambiguity. We claim that TODO

$$\exp \begin{pmatrix} \xi \\ x \end{pmatrix} = \begin{pmatrix} \exp(\xi) \\ J_l(\xi)x \end{pmatrix},$$

where

$$J_l(x) := I + \frac{1 - \cos \|x\|}{\|x\|^2} x^\wedge + \frac{\|x\| - \sin \|x\|}{\|x\|^3} (x^\wedge)^2$$

is the left Jacobian of the $\mathrm{SO}(3)$ exponential map. Note that $J_l(0) = I$.

Assume control has the form $u = K\ell$ and shares the same channel as the noise with scaling matrix $B$. Then the stochastic control process is modeled by the following Stratonovich SDE (Ito's is non-physical):

$$g^{-1}\mathrm{d}g = h^s(g)\mathrm{d}t + H\circledS\mathrm{d}W,$$

where

$$h^s \begin{pmatrix} R \\ \ell \end{pmatrix} = \begin{pmatrix} \mathbb{I}^{-1}\ell \\ -\ell^\wedge \times \mathbb{I}^{-1}\ell + BK\ell \end{pmatrix}, \qquad H = \begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}.$$

To obtain the 2nd-order approximation propagation, we need to compute the Jacobian and Hessian of following function:

$$h^c \begin{pmatrix} \xi \\ x \end{pmatrix} := h^s \left( g \exp \begin{pmatrix} \xi \\ x \end{pmatrix} \right)$$

$$= h^s \begin{pmatrix} \overline{R}\exp(\xi) \\ \overline{\ell} + J_l(\xi)x \end{pmatrix}$$

$$=$$

$$D(x^\wedge I^{-1}x)e_j = \left(-\left(I^{-1}x\right)^\wedge + x^\wedge I^{-1}\right)e_j$$

Then

$$\partial x_i \left(-\left(I^{-1}x\right)^\wedge + x^\wedge I^{-1}\right)e_j = \partial x_i \left((-I^{-1}x)^\wedge e_j\right) + E_i^{\mathrm{SO}(3)} I^{-1} e_j$$
$$= \partial x_i \left(e_j^\wedge I^{-1}x\right) + E_i^{\mathrm{SO}(3)} I^{-1} e_j$$
$$= E_j^{\mathrm{SO}(3)} I^{-1} e_i + E_i^{\mathrm{SO}(3)} I^{-1} e_j.$$

Since the term is symmetric, we can combine them into one when summed and scaled by entries of a symmetric matrix.

# 1 Tangent Bundle of A Lie Group

## 1.1 Group Theory

We first review two key concepts from group theory: group actions and semidirect products.

> **Definition 1.1** — Let $G$ be a group and let $A$ be a set. A **group action** $\rho : G \to \mathrm{Aut}(A), g \mapsto \rho_g$ is a group homomorphism. That is, each element $g \in G$ corresponds to an automorphism of $A$, which means $\rho_g$ sends any element $a$ to another element $\rho_g(a)$ in $A$ in an invertible way. There are two types of group actions, namely left and right actions. The former is denoted $g.a = \rho_g(a)$ and satisfies $(gh).a = g.(h.a)$, and the latter is denoted $a.g = \rho_g(a)$ and satisfies $a.(gh) = (a.g).h$.

The idea of group action is that since group elements have inverses, compose associatively, and contain an identity element, we can use them to encode automorphisms of a set, which

themselves forms a group and thus share the same behaviors. Thus, it is easy to check that $(\rho_g)^{-1} = \rho_{g^{-1}}$. Since the action is only required to be a group homomorphism but not isomorphism, it is possible that multiple group elements encode the same automorphism.

If $A = G$, then there are three natural actions of $G$ on itself: given any $g$, we have left multiplication $g.h = gh$, right multiplication $h.g = hg$, and conjugation $g.h = ghg^{-1}$.

In the context of a Lie group $G$, left and right multiplications are smooth automorphisms of $G$ that induces derivatives on the tangent spaces. If $v_h \in T_h G$, then the induced derivative of left multiplication is $T_h L_g(v_h)$ and that of right multiplication is $T_h R_g(v_h)$. Notice that we need to keep careful track of the base point. Together, they induce an action on the Lie algebra $\mathfrak{g} := T_e G$ called the **adjoint action**. Given any $\xi \in \mathfrak{g}$, the adjoint action of $g$, $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ is defined as

$$\mathrm{Ad}_g(\xi) := T_{g^{-1}} L_g \circ T_e R_{g^{-1}}(\xi),$$
$$:= T_g R_{g^{-1}} \circ T_e L_g(\xi).$$

In a sense, the adjoint action is conjugation by the derivatives. Notice that the derivatives of left and right multiplications commute with each other, just as left and right multiplications themselves. We simply need to make sure the base point is correct.

We shall see later that the definition of Ad can indeed be derived from differentiating group conjugation. For now, let us review semidirect products.

**Definition 1.2** — Let $(K, *_K), (H, *_H)$ be groups and suppose there exists a group action $\rho : K \to \text{Aut}(H)$. Then the **semidirect product** $K \ltimes_\rho H$ is the set $K \times H$ with the identity element $(e_K, e_H)$ and is endowed with the following the group operation $*$ induced by the group action $\rho$:

$$(k_1, h_1) * (k_2, h_2) = (k_1 *_K k_2, h_1 *_H \rho_{k_1}(h_2))$$

Thus, $(k, h)^{-1} = (k^{-1}, \rho_{k^{-1}}(h^{-1}))$ since $(k, h) * (k^{-1}, \rho_{k^{-1}}(h^{-1})) = (e_K, h *_H [\rho_k \circ \rho_{k^{-1}}(h^{-1})]) = (e_K, e_H)$, where we use the fact that $\rho_{k^{-1}} = \rho_k^{-1}$.

Alternatively, we can define a different group operation $*$ such that

$$(k_1, h_1) * (k_2, h_2) = (k_1 *_K k_2, \rho_{k_2^{-1}}(h_1) *_H h_2).$$

This is an equally valid definition yet is rarely mentioned in the literature. We will see that this definition is vital for our purpose so we haphazardly name it the **cosemidirect product** and denote it as $K \ltimes_{\rho^{-1}}^* H$. Then, $(k, h)^{-1} = (k^{-1}, \rho_k(h^{-1}))$ since

$$(k, h) * (k^{-1}, \rho_k(h^{-1})) = (e_K, \rho_k(h) *_H \rho_k(h^{-1})) = (e_K, \rho_k(h *_H h^{-1})) = (e_K, e_H).$$

**Example 1.3**

Consider the Lie groups $(\text{SO}(3), \text{matmul})$ and $(\mathbb{R}^3, +)$. The 3D special Euclidean group is defined as $\text{SE}(3) := \text{SO}(3) \ltimes_{\text{Ad}} \mathbb{R}^3$, where $\text{Ad}_R(v) = Rv$. Then its group operation $*$ is

$$(R_1, v_1) * (R_2, v_2) = (R_1 R_2, v_1 + \text{Ad}_{R_1}(v_2)) = (R_1 R_2, v_1 + R_1 v_2).$$

Observe that we can embed $\text{SE}(3)$ into $\text{SL}(4, \mathbb{R})$ and convert $*$ to simply matrix multiplication:

$$(R_1, v_1) * (R_2, v_2) \mapsto \begin{pmatrix} R_1 & v_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R_2 & v_2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} R_1 R_2 & v_1 + R_1 v_2 \\ 0 & 1 \end{pmatrix}.$$

**Example 1.4**

We define the 3D left special Euclidean group as $\text{LSE}(3) := \text{SO}(3) \ltimes_{\text{Ad}^{-1}}^* \mathbb{R}^3$. Since

$R^{-1} = R^T$ in SO(3), the group operation $*$ is

$$(R_1, v_1) * (R_2, v_2) = (R_1 R_2, \mathrm{Ad}_{R_2^T}(v_1) + v_2) = (R_1 R_2, R_2^T v_1 + v_2).$$

Since the group operation is different from that of SE(3), its matrix embedding must be different as well. Consider the embedding $(R, v) \mapsto \begin{pmatrix} R & 0 \\ v^T & 1 \end{pmatrix}$, then $*$ also becomes matrix multiplication:

$$(R_1, v_1) * (R_2, v_2) \mapsto \begin{pmatrix} R_1 & 0 \\ v_1^T & 1 \end{pmatrix} \begin{pmatrix} R_2 & 0 \\ v_2^T & 1 \end{pmatrix} = \begin{pmatrix} R_1 R_2 & 0 \\ v_1^T R_2 + v_2^T & 1 \end{pmatrix} \mapsto (R_1 R_2, R_2^T v_1 + v_2).$$

## 1.2   Lie Group Structure of Tangent Bundle

The right trivialization results from this section can be found at [?] Chapter 4.1 (note the right action semidirect product mentioned there is not the cosemidirect product we have here). TODO Engo 2001 mentions some results but did not cite nor elaborate.

Let $G$ be a Lie group of dimension $n$. The goal of this section is to show that the tangent bundle $TG$ has two group-structure-preserving trivializations (meaning we can parametrize the group by another group whose underlying manifold is the trivial vector bundle $G \times \mathbb{R}^n$), which we call **left and right trivializations**. First, notice that $G$ acts on itself by left or right multiplication. The action induces an action on the fibers of $TG$: $g.(h, v) = (gh, T_h L_g(v))$ or $(h, v).g = (hg, T_h R_g(v))$. This induced action allows us to move the base point of all tangent vectors of $G$ to the identity, so that all tangent vectors in $TG$ are translated to the Lie algebra $\mathfrak{g}$. This way, we can define left and right trivialization maps as

$$LT : TG \to G \ltimes_{\mathrm{Ad}^{-1}}^* \mathfrak{g}, (g, v) \mapsto (g, T_g L_{g^{-1}}(v)),$$
$$RT : TG \to G \ltimes_{\mathrm{Ad}} \mathfrak{g}, (g, v) \mapsto (g, T_g R_{g^{-1}}(v)).$$

We claim that the two maps are Lie group isomorphisms. That is, it is a diffeomorphism to the direct product manifold and respects the (co)semidrect product structure. The inverse is $(g, u)^{-1} = (g^{-1}, -\mathrm{Ad}_g(u))$. To prove the claim, we shall use the inverses of left and right trivialization maps instead, for convenience.

> **Definition 1.5** — The left translation and right translation maps on $TG$ are respectively defined as
>
> $$TL : G \ltimes^*_{\mathrm{Ad}^{-1}} \mathfrak{g} \to TG, \ (g, \xi) \mapsto (g, T_e L_g(\xi)),$$
>
> $$TR : G \ltimes_{\mathrm{Ad}} \mathfrak{g} \to TG, (g, \xi) \mapsto (g, T_e R_g(\xi)).$$

> **Proposition 1.6**
>
> Let $G$ be a Lie group and $\mathfrak{g}$ be its Lie algebra. Then the left and right translation maps $TL$ and $TR$ are Lie group isomorphisms.

*Proof.* Recall that multiplication on $TG$ is the differential of the multiplication on $G$. That is, if $\mu : G \times G \to G, (g, h) \mapsto gh$, then product rule gives

$$d\mu : TG \times TG \to TG, \ ((g, \xi), (h, \eta)) \mapsto (gh, T_h L_g(\eta) + T_g R_h(\xi)).$$

The left translation map $TL : G \ltimes_{\mathrm{Ad}^{-1}} \mathfrak{g} \to TG, \ (g, \xi) \mapsto (g, T_e L_g \xi)$ is smooth since $d\mu$ is smooth, the map $(g, \xi) \mapsto ((e, \xi), (g, 0))$ is smooth, and $d\mu((e, \xi), (g, 0)) = (g, T_e L_g(\xi))$ is a composition of smooth maps. It is a diffeomorphism since it has inverse $LT : (g, v) \mapsto (g, T_g L_{g^{-1}}(v))$. It remains to show that it is also a Lie group homomorphism.

Since $G$ is a multiplicative group whereas $\mathfrak{g}$ as a vector space is an additive group, we have

$$
\begin{aligned}
TL((g, \xi) * (h, \eta)) &= TL((gh, \ \mathrm{Ad}_{h^{-1}}(\xi) + \eta)) \\
&= (gh, T_e L_{gh}(\mathrm{Ad}_{h^{-1}}(\xi) + \eta)) \\
&= (gh, T_e L_{gh}(T_h L_{h^{-1}} \circ T_e R_h(\xi) + \eta)) \\
&= (gh, T_h L_g \circ T_e R_h(\xi) + T_h L_g \circ T_e L_h(\eta)) \\
&= (gh, T_h L_g \circ T_e L_h(\eta) + T_g R_h \circ T_e L_g(\xi)) \qquad \text{LR commute} \\
&= d\mu[(g, T_e L_g(\xi)), \ (h, T_e L_h(\eta))] \\
&= d\mu[TL(g, \xi), \ TL(h, \eta)].
\end{aligned}
$$

Thus, $TL$ is a Lie group homomorphism and therefore an isomorphism. The results for $TR$ exactly mirror those of $TL$. For completeness, we show

$$TR((g, \xi) * (h, \eta)) = TR((gh, \ \xi + \mathrm{Ad}_g(\eta)))$$

$$= (gh, T_e R_{gh}(\xi + \mathrm{Ad}_g(\eta)))$$

$$= (gh, T_e R_{gh}(\xi + T_{g^{-1}} L_g \circ T_e R_{g^{-1}}(\eta)))$$

$$= (gh, T_g R_h \circ T_e R_g(\xi) + T_g R_h \circ T_e L_g(\eta))$$

$$= (gh, T_g R_h(T_e R_g(\xi)) + T_h L_g(T_e R_h(\eta))) \qquad \text{LR commute}$$

$$= d\mu[(g, T_e R_g(\xi)), \ (h, T_e R_h(\eta))]$$

$$= d\mu[TR(g, \xi), \ TR(h, \eta)].$$

$$\square$$

It turns out that both left and right trivializations yield the same Lie algebra! TODO: why? intuition? Thus we denote the Lie algebra by $\mathfrak{g} \ltimes \mathfrak{g}$, which we claim is endowed with the following Lie bracket:

$$[(\xi, u), (\eta, v)] := ([\xi, \eta], [\xi, v] - [\eta, u]).$$

Its **adjoint representation** $\mathrm{ad}_{(\xi, u)} := [(\xi, u), -] : \mathfrak{g} \ltimes \mathfrak{g} \to \mathfrak{g} \ltimes \mathfrak{g}$. Note that the notations $\mathrm{Ad}, \mathrm{ad}, [-, -]$ can come from $G$ or the semidirect product, and any ambiguity can be resolved by inspecting their inputs. Now we derive the bracket formula.

Loosely speaking, the adjoint action $\mathrm{Ad}$ is the derivative of group conjugation action at identity, and adjoint representation $\mathrm{ad}$ is the derivative of adjoint action at identity. We make this precise in the context of the tangent bundle of $G$: for any $(g, u) \in G \ltimes^*_{\mathrm{Ad}^{-1}} \mathfrak{g}$, its action on $G \ltimes^*_{\mathrm{Ad}^{-1}} \mathfrak{g}$ by conjugation is

$$\phi_{(g,u)} : G \ltimes^*_{\mathrm{Ad}^{-1}} \mathfrak{g} \to G \ltimes^*_{\mathrm{Ad}^{-1}} \mathfrak{g}, (h, v) \mapsto (g, u) * (h, v) * (g, u)^{-1}.$$

Explicitly, we have

$$\phi_{(g,u)}(h, v) = (g, u) * (h, v) * \left( g^{-1}, -\mathrm{Ad}_g(u) \right)$$

$$= (gh, \mathrm{Ad}_{h^{-1}}(u) + v) * \left( g^{-1}, -\mathrm{Ad}_g(u) \right)$$

$$= \left( ghg^{-1}, \mathrm{Ad}_g(\mathrm{Ad}_{h^{-1}}(u) + v) - \mathrm{Ad}_g(u) \right).$$

Let $(\gamma, c)$ be a smooth curve on $G \times \mathfrak{g}$ such that $\gamma(0) = e$, $\dot{\gamma}(0) = \eta \in \mathfrak{g}$, $c(0) = 0$, $\dot{c}(0) = v$. Then the derivative of $\phi_{(g,u)}$ at $(e, 0)$, the adjoint action $\mathrm{Ad}_{(g,u)} : \mathfrak{g} \ltimes \mathfrak{g} \to \mathfrak{g} \ltimes \mathfrak{g}$ is given by

$$\mathrm{Ad}_{(g,u)}(\eta, v) = \frac{d}{dt}\Big|_{t=0} \phi_g(\gamma(t), c(t))$$

$$= \frac{d}{dt}\Big|_{t=0} \left( g\gamma(t)g^{-1}, \mathrm{Ad}_g(\mathrm{Ad}_{\gamma(t)^{-1}}(u) + c(t)) - \mathrm{Ad}_g(u) \right)$$
$$= \left( \mathrm{Ad}_g(\eta), \mathrm{Ad}_g(-\mathrm{ad}_{\dot{\gamma}(0)}(u) + \dot{c}(0)) \right)$$
$$= \left( \mathrm{Ad}_g(\eta), \mathrm{Ad}_g(-[\eta, u] + v) \right),$$

where we use the fact that $\mathrm{ad}_{\dot{\gamma}(0)} := \frac{d}{dt}\Big|_{t=0} \mathrm{Ad}_{\gamma(t)}$ and the following result:

$$\mathrm{Ad}_{\gamma(t)^{-1}} \circ \mathrm{Ad}_{\gamma(t)} = \mathrm{id}$$
$$\frac{d}{dt}\Big|_{t=0} \left( \mathrm{Ad}_{\gamma(t)^{-1}} \circ \mathrm{Ad}_{\gamma(t)} \right) = 0$$
$$\frac{d}{dt}\Big|_{t=0} \mathrm{Ad}_{\gamma(t)^{-1}} \circ \mathrm{Ad}_{\gamma(0)} + \mathrm{Ad}_{\gamma(0)^{-1}} \circ \mathrm{ad}_{\dot{\gamma}(0)} = 0 \qquad \text{chain rule}$$
$$\frac{d}{dt}\Big|_{t=0} \mathrm{Ad}_{\gamma(t)^{-1}} \circ \mathrm{Ad}_e + \mathrm{Ad}_e \circ \mathrm{ad}_{\dot{\gamma}(0)} = 0$$
$$\frac{d}{dt}\Big|_{t=0} \mathrm{Ad}_{\gamma(t)^{-1}} = -\mathrm{ad}_{\dot{\gamma}(0)} \qquad \mathrm{Ad}_e = \mathrm{id}.$$

Finally, let $(\gamma, c)$ be another smooth curve on $G \times \mathfrak{g}$ such that $\gamma(0) = e$, $\dot{\gamma}(0) = \xi$, $c(0) = 0$, $\dot{c}(0) = u$. Then the derivative of $\mathrm{Ad}_{(\gamma(t),c(t))}$ at $t = 0$ is the adjoint representation $\mathrm{ad}_{(\xi,u)} : \mathfrak{g} \ltimes^* \mathfrak{g} \to \mathfrak{g} \ltimes^* \mathfrak{g}$:

$$\mathrm{ad}_{(\xi,u)}(\eta, v) = \frac{d}{dt}\Big|_{t=0} \mathrm{Ad}_{(\gamma(t),c(t))}(\eta, v)$$
$$= \frac{d}{dt}\Big|_{t=0} \left( \mathrm{Ad}_{\gamma(t)}(\eta), \mathrm{Ad}_{\gamma(t)}(-[\eta, c(t)] + v) \right)$$
$$= \left( \mathrm{ad}_\xi(\eta), \mathrm{ad}_\xi(-[\eta, c(0)]) + \mathrm{Ad}_{\gamma(0)}(-[\eta, \dot{c}(0)]) + \mathrm{ad}_\xi(v) \right)$$
$$= ([\xi, \eta], [\xi, v] - [\eta, u]).$$

We collect the above results into a proposition:

<div style="border:1px solid; border-radius:10px; padding:10px;">

**Proposition 1.7**

The left trivialization $G \ltimes^*_{\mathrm{Ad}^{-1}} \mathfrak{g}$ of $TG$ has

$$\mathrm{Ad}_{(g,u)}(\eta, v) = \left( \mathrm{Ad}_g(\eta), \mathrm{Ad}_g(v - [\eta, u]) \right),$$
$$\mathrm{ad}_{(\xi,u)}(\eta, v) = ([\xi, \eta], [\xi, v] - [\eta, u]).$$

</div>

**Remark 1.8** The right trivialization $G \ltimes_{\mathrm{Ad}} \mathfrak{g}$ instead has adjoint action

$$\mathrm{Ad}_{(g,u)}(\eta, v) = \left(\mathrm{Ad}_g(\eta), \mathrm{Ad}_g(v) - [\mathrm{Ad}_g(\eta), u]\right).$$

This is different from that of the left trivialization. However, the adjoint representation is exactly the same. Moreover, SE(3) manifests this structure for TSO(3) through matrix multiplication. However, body-frame dynamics demand left trivialization, so we are stuck with the cosemidirect product LSE(3).

## 1.3   Demystifying LSE(3)

Recall that any finite-dimensional Lie algebra is a finite-dimensional vector space and thus is isomorphic to $\mathbb{R}^n$ after choosing a basis, and we convert back and forth using $x^\wedge = X \in \mathfrak{g}$ and $X^\vee = x \in \mathbb{R}^n$.

The Lie algebra elements of $\mathfrak{lse}(3) = \mathfrak{so}(3) \ltimes \mathfrak{so}(3) \cong \mathbb{R}^3 \ltimes \mathbb{R}^3$ have the following matrix embedding:

$$(\xi, v) \mapsto \begin{pmatrix} \xi^\wedge & 0 \\ v^T & 0 \end{pmatrix},$$

where $\xi^\wedge$ is a skew-symmetric matrix in $\mathfrak{so}(3)$.

Now we have a matrix representation, we can easily verify the Ad and ad maps for LSE(3) using matrix multiplication.

$$\mathrm{Ad}_{(R,u)}(\eta, v) = (R, u) * (\eta, v) * (R, u)^{-1}$$

$$\mapsto \begin{pmatrix} R & 0 \\ u^T & 1 \end{pmatrix} \begin{pmatrix} \eta^\wedge & 0 \\ v^T & 0 \end{pmatrix} \begin{pmatrix} R^T & 0 \\ -Ru^TR^T & 1 \end{pmatrix}$$

$$= \begin{pmatrix} R\eta^\wedge & 0 \\ u^T\eta^\wedge + v^T & 0 \end{pmatrix} \begin{pmatrix} R^T & 0 \\ -Ru^TR^T & 1 \end{pmatrix}$$

$$= \begin{pmatrix} R\eta^\wedge R^T & 0 \\ (u^T\eta^\wedge + v^T)R^T & 0 \end{pmatrix}$$

$$\mapsto (R\eta^\wedge R^{-1}, R(v - \eta^\wedge u)) \qquad\qquad (\eta^\wedge)^T = -\eta^\wedge$$

$$\mapsto (\mathrm{Ad}_R(\eta), \mathrm{Ad}_R(v - [\eta, u])).$$

In the last step, we use the following facts: 1. adjoint action of matrix Lie group is literally conjugation, $\text{Ad}_R(\Omega) = R\Omega R^{-1}$; 2. adjoint action of matrix Lie group on $\mathfrak{g}$ identified as $\mathbb{R}^n$ is simply matrix-vector multiplication, $\text{Ad}_R(\omega) = R\omega = R\omega^\wedge R^{-1}$.

The adjoint representation of matrix Lie group is just the commutator $\text{ad}_X(Y) := [X,Y] = XY - YX$:

$$\text{ad}_{(\xi,u)}(\eta,v) \mapsto \begin{pmatrix} \xi^\wedge & 0 \\ u^T & 0 \end{pmatrix} \begin{pmatrix} \eta^\wedge & 0 \\ v^T & 0 \end{pmatrix} - \begin{pmatrix} \eta^\wedge & 0 \\ v^T & 0 \end{pmatrix} \begin{pmatrix} \xi^\wedge & 0 \\ u^T & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \xi^\wedge \eta^\wedge & 0 \\ u^T \eta^\wedge & 0 \end{pmatrix} - \begin{pmatrix} \eta^\wedge \xi^\wedge & 0 \\ v^T \xi^\wedge & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \xi^\wedge \eta^\wedge - \eta^\wedge \xi^\wedge & 0 \\ \left( (\eta^\wedge)^T u - (\xi^\wedge)^T v \right)^T & 0 \end{pmatrix}$$

$$\mapsto ([\xi,\eta], \xi^\wedge v - \eta^\wedge u) \qquad\qquad \xi,\eta \text{ skew-symmetric}$$

$$= ([\xi,\eta], [\xi,v] - [\eta,u]) \,.$$

The results match those from the general case!

## 1.4  Preparation for Propagation

Using this matrix representation of LSE(3), we can apply the propagation formula.

A natural basis for $\mathfrak{lse}(3)$ is

$$E_i = \begin{cases} \begin{pmatrix} E_i^{\mathfrak{so}(3)} & 0 \\ 0 & 0 \end{pmatrix} & i = 1,2,3, \\ \\ \begin{pmatrix} 0 & 0 \\ e_{i-3}^T & 0 \end{pmatrix} & i = 4,5,6. \end{cases}$$

Under this basis, we can express the linear action $\text{Ad}_{(R,u)} : \mathfrak{so}(3) \ltimes \mathfrak{so}(3) \to \mathfrak{so}(3) \ltimes \mathfrak{so}(3)$ as a $6 \times 6$ matrix, whose $j$th column can be computed by $[(R,u)^\wedge E_j (R,u)^{-\wedge}]^\vee$, yielding

$$[\text{Ad}_{(R,u)}] = \begin{pmatrix} R & 0 \\ Ru^\wedge & R \end{pmatrix} \,.$$

Similarly, since $\text{ad}_{(\xi,u)} : \mathfrak{so}(3) \ltimes \mathfrak{so}(3) \to \mathfrak{so}(3) \ltimes \mathfrak{so}(3)$ is linear, let $[\text{ad}_{(\xi,u)}]$ denote the $6 \times 6$ matrix representation under this basis. Thus, its $j$th column can be computed by

$[(\xi, u)^\wedge, E_j]^\vee$, yielding

$$[\mathrm{ad}_{(\xi,u)}] = \begin{pmatrix} \xi^\wedge & 0 \\ u^\wedge & \xi^\wedge \end{pmatrix}.$$

We can readily verify that

$$[\mathrm{ad}_{(\xi,u)}] \begin{pmatrix} \eta \\ v \end{pmatrix} = \begin{pmatrix} \xi^\wedge & 0 \\ u^\wedge & \xi^\wedge \end{pmatrix} \begin{pmatrix} \eta \\ v \end{pmatrix} = \begin{pmatrix} \xi^\wedge \eta \\ \xi^\wedge v + u^\wedge \eta \end{pmatrix} = \begin{pmatrix} [\xi, \eta] \\ [\xi, v] - [\eta, u] \end{pmatrix} = \mathrm{ad}_{(\xi,u)}(\eta, v).$$

From this, we obtain a key expression in the propagation formula.

$$\mathrm{ad}_i := [\mathrm{ad}_{E_i}] = \begin{cases} \begin{pmatrix} E_i^{\mathfrak{so}(3)} & 0 \\ 0 & E_i^{\mathfrak{so}(3)} \end{pmatrix} & i = 1, 2, 3 \\ \begin{pmatrix} 0 & 0 \\ e_{i-3}^T & 0 \end{pmatrix} & i = 4, 5, 6. \end{cases}$$

Let $x = \begin{pmatrix} \eta \\ v \end{pmatrix}$. The exponential map $\exp : \mathfrak{g} \ltimes \mathfrak{g} \to G \ltimes_{\mathrm{Ad}^{-1}}^* \mathfrak{g}$ is

$$\exp(x^\wedge) = \begin{pmatrix} \exp(\eta^\wedge) \\ J_\ell(\eta)^T v \end{pmatrix},$$

because

$$\begin{pmatrix} \eta^\wedge & 0 \\ v^T & 0 \end{pmatrix}^k = \begin{pmatrix} (\eta^\wedge)^k & 0 \\ v^T (\eta^\wedge)^{k-1} & 0 \end{pmatrix},$$

$$\begin{aligned} \exp \begin{pmatrix} \eta^\wedge & 0 \\ v^T & 0 \end{pmatrix} &= \sum_{k=0}^\infty \frac{1}{k!} \begin{pmatrix} \eta^\wedge & 0 \\ v^T & 0 \end{pmatrix}^k \\ &= \begin{pmatrix} \sum_{k=0}^\infty \frac{1}{k!} \eta^\wedge & 0 \\ v^T \sum_{k=0}^\infty \frac{1}{k!} (\eta^\wedge)^{k-1} & 0 \end{pmatrix} \\ &= \begin{pmatrix} \exp(\eta^\wedge) & 0 \\ \left( J_\ell(\eta)^T v \right)^T & 0 \end{pmatrix} \\ &= \begin{pmatrix} \exp(\eta^\wedge) & 0 \\ (J_r(\eta)v)^T & 0 \end{pmatrix} \qquad \text{only in } \mathfrak{se}(3). \end{aligned}$$

See [?]. Due to skew-symmetry, it is easy to see via Rodriguez's formula that $J_r = J_\ell^T$ in SO(3).

Let $\mu = (\overline{R}, \overline{\omega})$. The remaining terms in the propagation equations involve 1st-order and 2nd-order derivatives at zero of

$$
\begin{aligned}
h^c(x) := h^s(\mu \exp(x^\wedge)) &= h^s \begin{pmatrix} \overline{R}\exp(\eta^\wedge) \\ [\mathrm{Ad}_{\exp(-\eta^\wedge)}]\overline{\omega} + J_r(\eta)v \end{pmatrix} \\
&= h^s \begin{pmatrix} \overline{R}\exp(\eta^\wedge) \\ \exp(-[\mathrm{ad}_{\eta^\wedge}])\overline{\omega} + J_r(\eta)v \end{pmatrix} \qquad \mathrm{Ad}_{\exp(X)} = \exp(\mathrm{ad}_X) \\
&= h^s \begin{pmatrix} \overline{R}\exp(\eta^\wedge) \\ \exp(-\eta^\wedge)\overline{\omega} + J_r(\eta)v \end{pmatrix} \qquad [\mathrm{ad}_X] = X \in \mathfrak{so}(3) \\
&= \begin{pmatrix} \widetilde{\omega} \\ \mathbb{I}^{-1}\left(BK\widetilde{\omega} - \widetilde{\omega}^\wedge \mathbb{I}\widetilde{\omega}\right) \end{pmatrix},
\end{aligned}
$$

where $\widetilde{\omega} := \exp(-\eta^\wedge)\overline{\omega} + J_r(\eta)v$ and $h^s(R,\omega) = \begin{pmatrix} \omega \\ \mathbb{I}^{-1}\left(BK\omega - \omega^\wedge\mathbb{I}\omega\right) \end{pmatrix}$ is just the attitude dynamics with linear feedback control $u(\omega) = K\omega$ from the Stratonovich SDE $g^{-1}\mathrm{d}g = h^s\mathrm{d}t + H^s \textcircled{s}\mathrm{d}W$, where $H^s = \begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}$.

It is easier to find the 1st and 2nd order derivatives of $\widetilde{\omega}$ first:

$$
a_j := \left.\frac{\partial \omega}{\partial x_j}\right|_{(0,0)} = \begin{cases} -\exp(-\eta^\wedge)\big|_0 E_j^{\mathfrak{so}(3)}\overline{\omega} + \partial_j Jr(\eta)v\big|_{(0,0)} = -E_j^{\mathfrak{so}(3)}\overline{\omega} & j = 1,2,3, \\[2mm] \partial_j J_r(\eta)v\big|_{(0,0)} + J_r(\eta)\big|_0 e_{j-3} = e_{j-3} & j = 4,5,6. \end{cases}
$$

$$
c_{ij} := \left.\frac{\partial^2 \omega}{\partial x_i \partial x_j}\right|_{(0,0)} = \begin{cases} \exp(-\eta^\wedge)\big|_0 E_i^{\mathfrak{so}(3)} E_j^{\mathfrak{so}(3)}\overline{\omega} = E_i^{\mathfrak{so}(3)} E_j^{\mathfrak{so}(3)}\overline{\omega} & i,j = 1,2,3, \\[2mm] \partial_j J_r(\eta)\big|_0 e_{i-3} = -\frac{1}{2}E_j^{\mathfrak{so}(3)} e_{i-3} & i = 4,5,6, \quad j = 1,2,3, \\[2mm] c_{ji} & i = 1,2,3, \quad j = 4,5,6 \\[2mm] 0 & i,j = 4,5,6. \end{cases}
$$

Note that we use the series expansion of $J_r(\eta) = I - \frac{1}{2}\eta^\wedge + \cdots$ to obtain $\partial_j J_r(\eta) = -\frac{1}{2}E_j^{\mathfrak{so}(3)}$ for $j = 1,2,3$.

Denote $h^c = (h_1^c, h_2^c)$, we can compute the composite derivatives blockwise:

$$
\left.\frac{\partial h_1^c}{\partial x_j}\right|_{(0,0)} = a_j, \qquad \left.\frac{\partial^2 h_1^c}{\partial x_i \partial x_j}\right|_{(0,0)} = c_{ij},
$$

and

$$\left.\frac{\partial h_2^c}{\partial x_j}\right|_{(0,0)} = \mathbb{I}^{-1}\left(BKa_j - \left.\left(\partial_j(\widetilde{\omega})^\wedge \mathbb{I}\widetilde{\omega} + \widetilde{\omega}^\wedge \mathbb{I}\partial_j\widetilde{\omega}\right)\right|_{(0,0)}\right)$$

$$= \mathbb{I}^{-1}\left(BKa_j - \left(a_j^\wedge \mathbb{I}\overline{\omega} + \overline{\omega}^\wedge \mathbb{I}a_j\right)\right),$$

$$\left.\frac{\partial^2 h_2^c}{\partial x_i \partial x_j}\right|_{(0,0)} = \mathbb{I}^{-1}\left(BKc_{ij} - \left.\left(\partial_{ij}^2(\widetilde{\omega})^\wedge \mathbb{I}\widetilde{\omega} + \partial_j(\widetilde{\omega})^\wedge \mathbb{I}\partial_i\widetilde{\omega} + \partial_i(\widetilde{\omega})^\wedge \mathbb{I}\partial_j\widetilde{\omega} + \widetilde{\omega}^\wedge \mathbb{I}\partial_{ij}^2\widetilde{\omega}\right)\right|_{(0,0)}\right)$$

$$= \mathbb{I}^{-1}\left(BKc_{ij} - \left(c_{ij}^\wedge \mathbb{I}\overline{\omega} + a_j^\wedge \mathbb{I}a_i + a_i^\wedge \mathbb{I}a_j + \overline{\omega}^\wedge \mathbb{I}c_{ij}\right)\right),$$

where we use the fact that $\widetilde{w}|_{(0,0)} = \overline{\omega}$. Due to symmetry, further simplifications can be achieved when their sums are weighted by a symmetric matrix like the covariance. But we will leave that for later.

We now have all the ingredients for the 2nd order mean and covariance propagation on LSE(3).

## 2 Jacobian

Since $\exp: \mathfrak{g} \to G$, its derivative at $X = x^\wedge \in \mathfrak{g}$ is $T_X \exp: \mathfrak{g} \to T_{\exp(X)}G$. Let $Y = y^\wedge \in \mathfrak{g}$. We can compute the derivative by

$$\begin{aligned}
T_X \exp(Y) &= \left.\frac{d}{dt}\right|_{t=0} \exp(X + tY) \\
&= \left.\frac{d}{dt}\right|_{t=0} \lim_{n\to\infty}\left(I + \frac{X + tY}{n}\right)^n \\
&= \lim_{n\to\infty}\left.\frac{d}{dt}\right|_{t=0}\left(I + \frac{X + tY}{n}\right)^n && \text{uniform convergence of derivatives} \\
&= \lim_{n\to\infty}\sum_{k=0}^{n}\left(I + \frac{X}{n}\right)^k \frac{Y}{n}\left(I + \frac{X}{n}\right)^{n-k-1} && \text{generalized Leibniz rule} \\
&= \int_0^1 \exp(tX)Y\exp((1-t)X)\,dt && \text{known identity} \\
&= \int_0^1 \exp(tX)Y\exp(-tX)\,dt\,\exp(X) \\
&= \int_0^1 \mathrm{Ad}_{\exp(tX)}(Y)\,dt\,\exp(X).
\end{aligned}$$

But since we not interested in any tangent space other than the Lie algebra, we should right or left translate this derivative back to the Lie algebra, which results in left and right

19

Jacobian, respectively. That is,

$$
\begin{aligned}
J_\ell(x)[y] &:= \left( T_{\exp(X)} R_{\exp(X)^{-1}} \circ T_X \exp(y^\wedge) \right)^\vee \\
&= \left( T_X \exp(y^\wedge) \exp(X)^{-1} \right)^\vee && \text{matrix Lie group} \\
&= \left( \int_0^1 \mathrm{Ad}_{\exp(tX)}(y^\wedge) \ \mathrm{d}t \exp(X) \exp(X)^{-1} \right)^\vee \\
&= \left( \int_0^1 [\mathrm{Ad}_{\exp(tX)}] \ \mathrm{d}t \right) y \\
&= \int_0^1 \exp([\mathrm{ad}_{tX}]) \ \mathrm{d}t \ y.
\end{aligned}
$$

Thus, we see that

$$
\begin{aligned}
J_\ell(x) &= \int_0^1 \exp([t \, \mathrm{ad}_X]) \ \mathrm{d}t \\
&= \int_0^1 \sum_{k=0}^\infty \frac{t^k}{k!} [\mathrm{ad}_X]^k \mathrm{d}t \\
&= \sum_{k=0}^\infty \int_0^1 \frac{t^k}{k!} [\mathrm{ad}_X]^k \mathrm{d}t \\
&= \sum_{k=0}^\infty \frac{t^{k+1}}{(k+1)!} [\mathrm{ad}_X]^k \bigg|_0^1 \\
&= \sum_{k=0}^\infty \frac{[\mathrm{ad}_X]^k}{(k+1)!}.
\end{aligned}
$$

Similarly, by a change of variable $s = -t$ and left translation, we obtain that

$$
J_r(x) = \int_0^1 \exp([-t \, \mathrm{ad}_X]) \ \mathrm{d}t \ = \sum_{k=0}^\infty \frac{[-\mathrm{ad}_X]^k}{(k+1)!}.
$$

## 3  $SO(3)$ **Facts**

The following facts can be found in [?]. For $SO(3)$, by Cayley-Hamilton Theorem, the infinite series of the exponential map (infinite basis with constant coefficients) can be reduced to an order-2 exact formula with trigonometric coefficients, called the Rodriguez formula, as long as we stay within the injectivity radius $\|x\| < \pi$ and set $\exp(0^\wedge) = I$:

$$
\exp(x^\wedge) = I + \frac{\sin \|x\|}{\|x\|} X + \frac{(1 - \cos \|x\|)}{\|x\|^2} X^2.
$$

We can derive the left and right Jacobians accordingly:

$$
J_\ell(x) = I + \frac{1 - \cos \|x\|}{\|x\|^2} X + \frac{\|x\| - \sin \|x\|}{\|x\|^3} X^2,
$$

$$J_r(x) = I - \frac{1 - \cos \|x\|}{\|x\|^2} X + \frac{\|x\| - \sin \|x\|}{\|x\|^3} X^2.$$

Since $X^T = -X$ in $\mathfrak{so}(3)$, we see that $J_\ell = J_r^T$ in SO(3). Then we can also obtain closed-formed expressions of their inverses:

$$J_\ell^{-1}(x) = I - \frac{1}{2} X + \left( \frac{1}{\|x\|^2} - \frac{1 + \cos \|x\|}{2\|x\| \sin \|x\|} \right) X^2,$$

$$J_r^{-1}(x) = I + \frac{1}{2} X + \left( \frac{1}{\|x\|^2} - \frac{1 + \cos \|x\|}{2\|x\| \sin \|x\|} \right) X^2.$$

In this document we use the following notations: for any smooth manifold $M$ and $p \in M$, the tangent space at $p$ is denoted as $T_p M$. If $f : M \to N$ is a smooth map, then the induced map (derivative) of $f$ at $p$ is $T_p f : T_p M \to T_p N$. Gluing induced maps at all points yields a smooth map between the tangent bundles $Tf : TM \to TN$. The reason to use the $T$ notation here is that it is natural to consider $T$ as a functor: it acts on both the manifolds and morphisms between manifolds naturally. Moreover, this is the most common notation in Lie theory textbooks. Let $G$ denote a general Lie group of dimension $n$ and $\mathfrak{g}$ be its Lie algebra. We tend to use $g$ for a group element in $G$ and $e$ for the identity element. We tend to use $\xi, \eta, u, v$ for Lie algebra elements. For matrix Lie groups, we use capital letters such as $X$ for matrices, and lower case letters such as $x$ for vectors. Since the Lie algebra of a matrix Lie group is a vector space of matrices, we can choose a basis $\{E_i\}_{i=1}^n$ and convert between a Lie algebra element $X$ and its coordinate vector $x$ via the hat $x^\wedge = X$ and vee $X^\vee = x$ operations. Likewise we can do this for any finite-dimensional Lie algebra, including the pure quaternions.

A Lie group can be parametrized differently with generalized coordinates. For example, SO(3) can be parametrized by Euler angles at the cost of singularities. A more natural and convenient parametrization is the exponential coordinates as long as the exponential map is surjective. This is the case for compact connected Lie groups like SO($n$) and SU($n$). It is also the case for SE($n$) or the tangent bundle Lie group of any Lie group that has a surjective exponential TODO. Thus, we will use $g(q)$ for generalized coordinates, and $g(x) = \exp(x^\wedge)$ for exponential coordinates. It is beneficial for the analysis to use a moving exponential coordinates instead, so later we shall use $g(x) = \mu \exp(x^\wedge)$ where $\mu$ is the current mean so we continuously reset the origin of the coordinates to current mean.

> **Proposition 3.1**
>
> For any $\xi, \eta \in \mathfrak{g}$, we have the following identity
>
> $$\mathrm{Ad}_{\exp(\xi)}(\eta) = \exp(\mathrm{ad}_\xi(\eta)).$$
>
> This is simply a corollary of the fact that exponential map commutes with differentiation, and ad is the derivative of Ad. This matches the intuition of exponential on the real numbers, where the derivative of the exponential is itself.

# 4 Intuition of Hamiltonian Mechanics

A symplectic manifold has a conanical 2-form $\omega$, which locally (in Darboux coordinates) acts as a symplectic bilinear form $J := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$. Intuitively this "turns the tangent coordinates 90 degrees". That is, horizontal direction becomes vertical and vice versa (not too different from $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ if we think $x$-coordinate as horizontal tangent coordinates and $y$-coordinate as vertical tangent coordinates). When this is applied to gradient vector field of a function $H$, since gradient goes perpendicular to the contour lines of $H$, by turning 90 degrees now the vector field goes along the contour lines (level sets), which means that along the flow of the rotated vector field, $H$ is conserved. This vector field is the Hamiltonian vector field $X_H$ of $H$. Bloch Exercise 3.1.2 shows that in local Darboux coordinates, indeed we have $X_H = Jdf^T = J\nabla f$. However, since there is no metric on the manifold, we cannot actually obtain gradient vector field of a function $H$ nor rotate it. Yet we can still do this rotation to flow along the contour lines thanks to the 2-form. Instead of rotating the gradient vector field of $H$, we rotate the differential $dH$ contravariantly and get a vector field via $\iota_{X_H}\omega = dH$.

Moreover, when the concept is applied to the cotangent bundle of a manifold, we have a canonical 1-form $\theta$ and a canonical 2-form $\omega := -d\theta$ that is symplectic. We can write this 2-form in local coordinates, and then compute the $X_H$ in local coordinates as well for any $H$. If we choose $H$ to be the energy function, then the resulting coordinates are exactly Hamilton's equations, which give us the flows where energy is conserved. Notice that we

do not need to use calculus of variations at all here. Geometry completely determines the equations of motion.

$$\dot{z} = X_H(z)$$
$$\dot{f}(z) = \langle df(z), X_H(z) \rangle \ \forall \ f$$
$$\dot{f} = \{f, H\} \ \forall \ f$$

# 5   Dual Quaternions

Consider $G = \mathbb{H}_u$ with scalar-first convention. Then $\mathfrak{g} = T_1\mathbb{H}_u$ the set of pure quaternions, which we can identify with 3D vectors via $^\vee : \mathfrak{g} \to \mathbb{R}^3, (0, v) \mapsto v$ and $^\wedge : \mathbb{R}^3 \to \mathfrak{g}, v \mapsto (0, v)$. The exponential map can be defined using the infinite series $\exp(v) = \sum_{k=0}^{\infty} \frac{(v^\wedge)^k}{k!}$, which by applying the identity $(v^\wedge)^2 = -\|v\|^2$ simplifies to

$$\exp(v) = \begin{pmatrix} \cos\|v\| \\ \frac{\sin\|v\|}{\|v\|} v \end{pmatrix}.$$

We see that the injectivity radius remains $\|v\| < \pi$, the same as the exponential map of SO(3). Recall that $\mathbb{H}_u$ is a double-cover of SO(3) with isomorphic Lie algebra $\mathfrak{so}(3)$. Since their exponential maps have the same injectivity radius, in order for the exponential map of $\mathbb{H}_u$ to cover the entire $S^3$ that wraps around SO(3) twice within the same injectivity radius, the exponential map of $\mathbb{H}_u$ should go twice as fast as that of SO(3). Therefore, in order to encode the same rotation, axis-angle-wise we need half the angle $\|v\|$ in $T_1\mathbb{H}_u$ than we do in $\mathfrak{so}(3)$. Another way to see this is just deriving everything from SU(2) which is isomorphic to $\mathbb{H}_u$ as a Lie group.

As a result, for $u^\wedge, v^\wedge \in \mathfrak{g}$, define pure quaternion multiplication to be the cross product $u^\wedge v^\wedge := (u \times v)^\wedge$, we have

$$\mathrm{ad}_{u^\wedge}(v^\wedge) = [u^\wedge, v^\wedge] = u^\wedge v^\wedge - v^\wedge u^\wedge = (u \times v - v \times u)^\wedge = (2u \times v)^\wedge = 2u^\wedge v^\wedge.$$

Thus, we can write

$$\mathrm{ad}_{u^\wedge} = 2u^\wedge.$$

Indeed, we see that adjoint operator for pure quaternions goes twice as fast as well.

Note that we omit explicitly writing out two types of multiplication for quaternions. A unit quaternion that encodes a rotation acts on other quaternions by the quaternion multiplication, but a pure quaternion that encodes a tangent vector in the Lie algebra can only act on other pure quaternions by the cross product. Thus, ambiguity can be resolved by inspecting the types of the objects.

Since the adjoint representation for $\mathbb{H}_u$ has the same form and only differ by a factor of 2, the left and right Jacobians and their inverses of $\mathbb{H}_u$ follow from the same infinite series as those from $SO(3)$ modulo a factor of 2. By following Barfoot 25, we can then derive the left and right Jacobians of the cosemidirect product $G \ltimes^*_{\mathrm{Ad}^{-1}} \mathfrak{g}$, which has the form

$$J_\ell \begin{pmatrix} \xi \\ u \end{pmatrix} = \begin{pmatrix} J_\ell(\xi) & 0 \\ Q(\xi, u) & J_\ell(\xi) \end{pmatrix}$$
$$J_r \begin{pmatrix} \xi \\ u \end{pmatrix} = \begin{pmatrix} J_r(\xi) & 0 \\ Q(\xi, u)^T & J_r(\xi) \end{pmatrix}.$$

And then by Schur's complement, we can invert the block matrices and obtain

$$J_r^{-1} \begin{pmatrix} \xi \\ u \end{pmatrix} = \begin{pmatrix} J_r^{-1} & -J_r^{-1} Q^T J_r^{-1} \\ 0 & J_r^{-1} \end{pmatrix}.$$

And likewise for $J_\ell^{-1}$.

There is no way to convert generic matrix-vector multiplication into quaternion multiplication, since quaternion multiplication can only encodes matrices related to rigid-body motion. Thus, by using $\mathrm{ad}_i^T = -e_i^\wedge$, we obtain

$$M_{ij}^\mu = \frac{2^3}{48} e_i^\wedge e_k^\wedge \left(-BB^T e_j\right)^\wedge e_k^\wedge + \frac{2^3}{48} e_k^\wedge \left(-BB^T e_j\right)^\wedge (-e_i^\wedge) e_k^\wedge + \frac{1}{2} \left(2 e_i^\wedge \left(\frac{\partial h^c}{\partial x_j}\right)^\wedge + \left(\frac{\partial^2 h^c}{\partial x_i \partial x_j}\right)^\wedge\right).$$

# 6   pure dual quaternions

The counterpart of LSE(3) is the unit dual quaternions $\widetilde{\mathbb{H}_u} := \mathbb{H}_{u \; \mathrm{Ad}^{-1}} \ltimes T_1 \mathbb{H}_u = \{q + \varepsilon v^\wedge : q \in \mathbb{H}, v^\wedge \in T_1 \mathbb{H}, \varepsilon^2 = 0\}$. Notice that here we define $\widetilde{\mathbb{H}_u}$ to only take tangent vectors from

the pure quaternions. Multiplication mirrors that of LSE(3):

$$(q_1 + \varepsilon v_1^\wedge) \otimes (q_2 + \varepsilon v_2^\wedge) := q_1 q_2 + \varepsilon(\overline{q_2} v_1^\wedge q_2 + v_2^\wedge),$$

where $\overline{q}$ is the quaternion conjugate, which is also the inverse for unit quaternions.

Its Lie algebra is $T_1 \mathbb{H}_u \ltimes T_1 \mathbb{H}_u$, which we name the **pure dual quaternions**, with the following Lie bracket:

$$\begin{aligned}
[x^\wedge + \varepsilon u^\wedge, y^\wedge + \varepsilon v^\wedge] &= [x^\wedge, y^\wedge] + \varepsilon([x^\wedge, v^\wedge] - [y^\wedge, u^\wedge]) \\
&= 2x^\wedge y^\wedge + \varepsilon(2x^\wedge v^\wedge - 2y^\wedge u^\wedge) \\
&= 2(x^\wedge + \varepsilon u^\wedge) \times (y^\wedge + \varepsilon v^\wedge),
\end{aligned}$$

where $\times$ is the established dual quaternion cross product (see [**?**] Equation 7) simplified for pure dual quaternions. That is, we have

$$(x^\wedge + \varepsilon u^\wedge) \times (y^\wedge + \varepsilon v^\wedge) = x^\wedge y^\wedge + \varepsilon(x^\wedge v^\wedge - y^\wedge u^\wedge).$$

Thus, the adjoint operator of $\widetilde{\mathbb{H}_u}$ is given by

$$\mathrm{ad}_{\widetilde{x}} = 2\widetilde{x},$$

where dual quaternion cross product is implied since it is the only meaningful multiplication for pure dual quaternions.

The hat and vee operations on pure dual quaternions are just inserting or removing 0 for each pure quaternion. That is, for a vector $x = (x_1, x_2) \in \mathbb{R}^6$ where $x_1, x_2 \in \mathbb{R}^3$, we have

$$\widetilde{x} = \begin{pmatrix} 0 \\ x_1 \\ 0 \\ x_2 \end{pmatrix}.$$

Note that in our notation, a lower case letter with no modification, such as $x$, represents a vector, whereas $\widetilde{x}$ represents the pure dual quaternion corresponding to $x$. We reserve the $^\wedge$

and $^\vee$ for quaternions to avoid confusion. We also define the **swap operation** on pure dual quaternions: if $\widetilde{x} = x_1^\wedge + \varepsilon x_2^\wedge$, then

$$\widetilde{x}^s = x_2^\wedge + \varepsilon x_1^\wedge.$$

Let $\{e_i\}_{i=1}^6$ be the standard basis on $\mathbb{R}^6$. A convenient basis of $\widetilde{\mathbb{H}_u}$ is simply $\{\widetilde{e}_i\}_{i=1}^6$. Thus we have

$$\mathrm{ad}_i = \mathrm{ad}_{\widetilde{e}_i} = 2\widetilde{e}_i.$$

Recall that $\mathrm{ad}_i^T$ in LSE(3) acts on $x = (x_1, x_2) \in \mathbb{R}^6$ as

$$\mathrm{ad}_i^T x = \begin{cases} \begin{pmatrix} -E_i^{\mathfrak{so}(3)} & 0 \\ 0 & -E_i^{\mathfrak{so}(3)} \end{pmatrix} x = -\mathrm{ad}_i x & i = 1, 2, 3 \\[2em] \begin{pmatrix} 0 & -E_{i-3}^{\mathfrak{so}(3)} \\ 0 & 0 \end{pmatrix} x = \begin{pmatrix} -E_{i-3}^{\mathfrak{so}(3)} x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} -e_{i-3}^\wedge x_2^\wedge \\ 0 \end{pmatrix} & i = 4, 5, 6. \end{cases}$$

Notice that when converting the above to pure dual quaternions, the second case can be viewed as swapping the quaternion and the dual parts of both $\widetilde{e}_i$ and $\widetilde{x}$:

$$\widetilde{\mathrm{ad}}_i^T \widetilde{x} = \begin{cases} -\widetilde{e}_i \widetilde{x} & i = 1, 2, 3 \\ -\widetilde{e}_i^s \widetilde{x}^s & i = 4, 5, 6. \end{cases}$$

Finally, let $\widetilde{M\widetilde{x}}$ denote the operation where it outputs $\widetilde{Mx}$. As an example, we obtain the propagation equation

$$M_{ij}^\mu = \frac{2^3}{48} \widetilde{e}_i \widetilde{e}_k \widetilde{HH^T} \widetilde{\mathrm{ad}}_j^T \widetilde{e}_k + \frac{2^3}{48} \widetilde{e}_k \widetilde{HH^T} \widetilde{\mathrm{ad}}_j^T \widetilde{\mathrm{ad}}_i^T \widetilde{e}_k + \frac{1}{2} \left( 2\widetilde{e}_i \widetilde{\frac{\partial h^c}{\partial x_j}} + \widetilde{\frac{\partial^2 h^c}{\partial x_i \partial x_j}} \right).$$

In practice, there is no harm forgetting the 0's in dual pure quaternions since they only serve as placeholders and are never involved in any computation.

# 7 PMP

Here is the story: first, we know that the endpoint of optimal trajectory must be on the boundary of the attainable set at final time for the extended system, since the total cost has to

be optimal which is part of the extended system. PMP tells us the necessary conditions for all trajectories that have endpoints on the boundary of attainable set. Since all state trajectories follow the same velocity vector field (that satisfies typical ODE uniqueness assumptions), they are completely determined by their endpoints. Thus we only consider the case when the endpoints are on the boundary of attainable set.

The idea is that instead of imposing velocity constraints on the cost function, we treat the cost function as the velocity of the running cost and obtain an extended system. Then by studying the endpoints, we realized that the Hamiltonian must be maximized along the flow of the Hamiltonian lift of the extended velocity field. This unifies the state dynamics and the cost function dynamics into the same space. The Hamiltonian lift gives us back the trajectory after projection, but along the fiber direction the trajectory simply traces out the level sets of the maximized Hamiltonian since the velocity field is the Hamiltonian vector field. Along the costate trajectory the maximized Hamiltonian is conserved (Remark 12.2 of Agrachev and Sachov). In this case, since Hamiltonian is constant along the trajectory, maximality of Hamiltonian along the trajectory reduces to maximality of Hamiltonian at the end point. This provides another perspective of why we can find a costate trajectory just by knowing that its state trajectory endpoint is on the boundary. TODO

PMP, via its necessary conditions, reduces the infinite-dimensional trajectory optimization problem to the $(n-1)$-dimensional endpoint optimization problem on the boundary (by finding controls that maximize the Hamiltonian on the boundary TODO), which is equivalent to studying the $(n-1)$-dimensional costate equation $\dot{\lambda}_t = X_H(\lambda_t)$ since $X_H$ is homogenous.

How can we find the costate trajectory? It requires finding a supporting hyerplane of the reachable set of state trajectory at (almost) each time point. Pullback flow can give us the supporting hyperplane once we find it at the end point $q_1 = \tilde{q}(t_1)$. If the reachable set near the boundary is locally convex and is differentiable at $q_1$, we are done. But since it might not be locally convex, we must use a local convex approximation. This comes from using all admissible controls at their Lebesgue timepoints (so differentiation of integral gives integrand, which by constructing needlike control variations of $\tilde{u}$ gives all possible directions of controlled velocity at $q_1$) and use the flow to pushforward to $t_1$, then take the convex

cone of them all. This emcompasses all feasible controlled velocities at $q_1$ and ensures local convexity thus a supporting hyperplane. Maximality of the Hamiltonian at $t_1$ is exactly the convexity condition at $q_1$ where $\lambda_{t_1}$ is the supporting hyperplane. Then the adjoint identity on the pullback flow extends this condition to almost all time. We know that the pullback flow is precisely the Hamiltonian vector field corresponding to the Hamiltonian from Chapter 11. And we are done.

# 8 Control-Affine PMP

Consider the optimal control problem with control-affine dynamics and quadratic cost:

$$\ell(\mu, L, u, t) = \frac{1}{2}\|u\|^2$$

$$\begin{pmatrix} \mu^{-1}\dot{\mu} \\ \dot{L} \end{pmatrix} = f_0(\mu, L, t) + \sum_{i=1}^{n} f_i(\mu, L, t)u_i =: f_0(\mu, L, t) + F(\mu, L, t)u.$$