

## 0.1 Invariant EKF

**Definition 0.1** — Let  $G$  be a Lie group and  $f_u : G \rightarrow TG$  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

$$g_u(\exp(\xi)) = (A^i \xi)^\vee + O(\|\xi\|^2).$$

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:

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

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:

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

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

$$g(t + dt) = g(t) \exp(h(g(t), t)dt + H(g(t + kdt), t + kdt)dW).$$

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

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

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

We shall use the simplified notation

$$\begin{aligned} (g^{-1}dg)^\vee &= hdt + HdW \\ (g^{-1}dg)^\vee &= h^s dt + H^s \mathbb{S} dW. \end{aligned}$$

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

$$q(t + dt) = q(t) + (J_r^{-1} \tilde{h})(q, t) dt + (J_r^{-1} \tilde{H})(q(t + kdt), t + kdt) dW.$$

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.

$$\begin{aligned} \tilde{h}(x, t) &= h(x, t) + \left( \frac{1}{2} J_r \frac{\partial J_r^{-1}}{\partial x_k} H H^T J_r^{-T} \right) \Big|_{g=\mu \exp((x))} e_k, \\ \tilde{H}(x, t) &= H(\mu \exp(x), t). \end{aligned}$$

Note that if  $H$  does not depend on the state, then Ito's and Stratonovich 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  $\tilde{h}, \tilde{H}$  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

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

$$\int_G \log^\vee(\mu^{-1}g) p(g) dg = 0.$$

The covariance of  $p$  is defined by

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

#### 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

$$\begin{aligned} m' &= \langle J_\ell^{-1} \rangle^{-1} m \\ \mu_m &= \mu \exp(m' + O(|m'|^2)) \\ \Sigma_m &= \Sigma - \text{sym}(\langle J_{\ell^{-1}m'x^T} \rangle) + O(|m'|^2). \end{aligned}$$

#### 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:

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

Note the resemblance of this with its Euclidean counterpart.

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

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

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

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

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

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  $\text{ad}_i S e_i = 0$ .
- (3) The expansion of moments to approximate expectation (with hard-to-bound error):

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

- (4) As long as  $(I - A)^n \rightarrow 0$  converges ( $A$  is close to  $I$ ),  $A^{-1} = \sum_{k=0}^{\infty} (I - A)^k$ . For  $\langle J_l^{-1} \rangle^{-1}$ , since the 1st order term vanish under expectation, setting  $k = 1$  gives the 2nd order approximation.
- (5) Since  $\text{ad}$  is linear in each argument,  $\text{ad}_X = \text{ad}_{(x_i E_i)} = x_i \text{ad}_i$ .

We compute and discard terms with order 3 or higher:

$$\begin{aligned} (\mu^{-1} \dot{\mu})^\vee &\approx \underbrace{\left\langle I - \frac{1}{2}[\text{ad}_X] + \frac{1}{12}[\text{ad}_X]^2 \right\rangle}_{(1)}^{-1} \left\langle \frac{1}{2} \left( \frac{1}{2} \text{ad}_k + \frac{1}{12} (\text{ad}_k [\text{ad}_X] + [\text{ad}_X] \text{ad}_k) \right) H H^T \dots \right. \\ &\quad \left. \left( I + \frac{1}{2}[\text{ad}_X] + \frac{1}{12}[\text{ad}_X]^2 \right)^T e_k + \left( I + \frac{1}{2}[\text{ad}_X] + \frac{1}{12}[\text{ad}_X]^2 \right) \left( h + \frac{\partial h^c}{\partial x_i} x_i + \frac{1}{2} \frac{\partial^2 h^c}{\partial x_i \partial x_j} x_i x_j \right) \right\rangle \\ &\approx \underbrace{\left( I - \frac{1}{12} \langle [\text{ad}_X]^2 \rangle \right)}_{(4)} \left\langle \underbrace{\frac{1}{4} \text{ad}_k H H^T e_k}_{(2)} + \underbrace{\frac{1}{8} \text{ad}_k H H^T [\text{ad}_X]^T e_k + \frac{1}{48} \text{ad}_k H H^T ([\text{ad}_X]^2)^T e_k}_{(1)} \right. \\ &\quad \left. + \underbrace{\frac{1}{24} (\text{ad}_k [\text{ad}_X] + [\text{ad}_X] \text{ad}_k) H H^T e_k}_{(1)} + \underbrace{\frac{1}{48} \text{ad}_k [\text{ad}_X] H H^T [\text{ad}_X]^T e_k + \frac{1}{48} [\text{ad}_X] \text{ad}_k H H^T [\text{ad}_X]^T e_k}_{(2)} \right\rangle \end{aligned}$$

$$\begin{aligned}
& + h + \underbrace{\frac{\partial h^c}{\partial x_i} x_i}_{(1)} + \frac{1}{2} \frac{\partial^2 h^c}{\partial x_i \partial x_j} x_i x_j + \underbrace{\frac{1}{2} [\text{ad}_X] h + \frac{1}{2} x_j [\text{ad}_X] \frac{\partial h^c}{\partial x_j}}_{(1)} + \frac{1}{12} [\text{ad}_X]^2 h \Big\rangle \\
& \approx h + \left\langle \frac{1}{48} \text{ad}_k H H^T ([\text{ad}_X]^2)^T e_k + \frac{1}{48} [\text{ad}_X] \text{ad}_k H H^T [\text{ad}_X]^T e_k + \frac{1}{2} \frac{\partial^2 h^c}{\partial x_i \partial x_j} x_i x_j \right. \\
& \quad \left. + \frac{1}{2} x_j [\text{ad}_X] \frac{\partial h^c}{\partial x_j} + \underbrace{\frac{1}{12} [\text{ad}_X]^2 h - \frac{1}{12} \langle [\text{ad}_X]^2 \rangle h}_{\text{cancel after next step}} \right\rangle \\
& \approx h + \left( \frac{1}{48} \text{ad}_k H H^T \text{ad}_j^T \text{ad}_i^T e_k + \frac{1}{48} \text{ad}_i \text{ad}_k H H^T \text{ad}_j^T e_k + \frac{1}{2} \frac{\partial^2 h^c}{\partial x_i \partial x_j} \right. \\
& \quad \left. + \frac{1}{2} \text{ad}_i \frac{\partial h^c}{\partial x_j} \right) \langle x_i x_j \rangle \quad \text{use (5) for all and then use (3)} \\
& =: h + M_{ij}^\mu \Sigma_{ij}.
\end{aligned}$$

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

$$\begin{aligned}
\dot{\Sigma} & \approx \left\langle \text{sym} \left\{ \left[ \frac{1}{2} \left( \frac{1}{2} \text{ad}_k + \frac{1}{12} (\text{ad}_k [\text{ad}_X] + \underbrace{[\text{ad}_X] \text{ad}_k}_{(2)}) \right) H H^T \left( I + \frac{1}{2} [\text{ad}_X] \right)^T e_k \right. \right. \right. \\
& \quad \left. \left. - \left( I - \frac{1}{2} [\text{ad}_X] \right) (\mu^{-1} \dot{\mu})^\vee + \left( I + \frac{1}{2} [\text{ad}_X] \right) \left( h + \frac{\partial h^c}{\partial x_i} x_i \right) \right] x^T \right\} \right. \\
& \quad \left. + \left( I + \frac{1}{2} [\text{ad}_X] + \frac{1}{12} [\text{ad}_X]^2 \right) H H^T \left( I + \frac{1}{2} [\text{ad}_X] + \frac{1}{12} [\text{ad}_X]^2 \right) \right\rangle \\
& \approx H H^T + \left\langle \text{sym} \left\{ \left[ \frac{1}{8} \text{ad}_k H H^T [\text{ad}_X]^T e_k + \frac{1}{24} \text{ad}_k [\text{ad}_X] H H^T e_k + \frac{1}{2} [\text{ad}_X] \left( h + (\mu^{-1} \dot{\mu})^\vee \right) + \frac{\partial h^c}{\partial x_i} x_i \right] x^T \right. \right. \\
& \quad \left. \left. + \frac{1}{12} [\text{ad}_X]^2 H H^T \right\} + \frac{1}{4} [\text{ad}_X] H H^T [\text{ad}_X]^T \right\rangle \\
& \approx H H^T + \left( \text{sym} \left\{ \left[ \frac{1}{8} \text{ad}_k H H^T \text{ad}_i^T e_k + \frac{1}{24} \text{ad}_k \text{ad}_i H H^T e_k + \frac{1}{2} \text{ad}_i \left( h + (\mu^{-1} \dot{\mu})^\vee \right) + \frac{\partial h^c}{\partial x_i} \right] e_j^T \right. \right. \\
& \quad \left. \left. + \frac{1}{12} \text{ad}_i \text{ad}_j H H^T \right\} + \frac{1}{4} \text{ad}_i H H^T \text{ad}_j^T \right) \langle x_i x_j \rangle \\
& =: H H^T + M_{ij}^\Sigma \Sigma_{ij}.
\end{aligned}$$

$$M^\mu = \frac{1}{48} \text{ad}_i \text{ad}_k H H^T \text{ad}_j^T e_k + \frac{1}{48} \text{ad}_k H H^T \text{ad}_j^T \text{ad}_i^T e_k + \frac{1}{2} \text{ad}_i \frac{\partial h^c}{\partial x_j} + \frac{1}{2} \frac{\partial^2 h^c}{\partial x_i \partial x_j}$$

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

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  $G = \text{SO}(3) \ltimes \mathbb{R}^3$ , 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  $\exp : \mathfrak{g} \supset U \rightarrow V \subset G$  of  $G$ . We also denote the exponential map of  $\text{SO}(3)$  as  $\exp(\xi) := \text{expm}(\xi^\wedge)$ . 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  $\text{SO}(3)$  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):

$$g^{-1}dg = h^s(g)dt + H \odot dW,$$

where

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

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

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

$$\begin{aligned}
&= h^s \left( \begin{array}{c} \overline{R} \exp(\xi) \\ \bar{\ell} + J_l(\xi)x \end{array} \right) \\
&=
\end{aligned}$$

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

Then

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

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

Let  $G$  be a Lie group. The tangent bundle  $TG$  has exactly two trivializations, 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  $TG$ :  $g.(h, v) = (gh, T_h L_g(v))$  or  $g.(h, v) = (hg, T_h R_g(v))$ . This induced action allows us to move the base point of all tangent vectors of  $G$  to the identity, so that all tangent vectors are translated to the Lie algebra  $\mathfrak{g}$ . Explicitly, left trivialization is the map

$$\phi : TG \rightarrow G \ltimes \mathfrak{g}, (g, v) \mapsto (g, T_g L_{g^{-1}}(v)).$$

where the semidirect product  $G \ltimes \mathfrak{g}$  has the following unconventional product structure:

$$(g, u) * (h, v) = (gh, \text{Ad}_{h^{-1}}(u) + v).$$

We claim that this map is a Lie group isomorphism. That is, it is a diffeomorphism to the direct product manifold and respects the semidirect product structure. The inverse is  $(g, u)^{-1} = (g^{-1}, -\text{Ad}_g(u))$ .



**Definition 1.1** — Let  $(K, *_K), (H, *_H)$  be groups and suppose there exists a group action  $\rho : K \rightarrow \text{Aut}(H)$ . Then the **semidirect product**  $K \ltimes_\rho H$  is the set  $K \times H$  endowed with the following the group operation  $*$  induced by the group action  $\rho$ :

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

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  $K \ltimes_{\rho^{-1}}^* H$ .

Note that  $\rho_{k^{-1}} = \rho_k^{-1}$  which justifies the notation.

**Proposition 1.2**

Let  $G$  be a Lie group and  $\mathfrak{g}$  be its Lie algebra. Then  $TG \cong G \ltimes_{\text{Ad}} \mathfrak{g}$  as a Lie group under .

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

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

The “left translation” map  $TL : G \ltimes_{\text{Ad}^{-1}} \mathfrak{g} \rightarrow TG, (g, \xi) \mapsto (g, T_e L_g \xi)$  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  $TL^{-1} : (g, v) \mapsto (g, T_g L_{g^{-1}}(v))$ . We now 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

$$\begin{aligned} \phi((g, \xi)(h, \eta)) &= \phi((gh, \xi + \text{Ad}_g(\eta))) \\ &= (gh, T_e R_{gh}(\xi + \text{Ad}_g(\eta))) \\ &= (gh, T_e R_{gh}(\xi + T_{g^{-1}} L_g \circ T_e R_{g^{-1}}(\eta))) \end{aligned}$$

$$\begin{aligned}
&= (gh, T_g R_h \circ T_e R_g(\xi) + T_g R_h \circ T_e L_g(\eta)) \\
&= (gh, T_g R_h(T_e R_g(\xi)) + T_h L_g(T_e R_h(\eta))) \\
&= d\mu[(g, T_e R_g(\xi)), (h, T_e R_h(\eta))] \\
&= d\mu[\phi(g, \xi), \phi(h, \eta)].
\end{aligned}$$

Thus,  $\phi$  is a Lie group homomorphism and therefore an isomorphism.  $\square$

Its Lie algebra is  $\mathfrak{g} \ltimes \mathfrak{g}$  with the following Lie bracket:

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

Here whether the adjoint comes from  $G$  or the semidirect product is inferred from their inputs. Now we derive this fact.

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

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

Explicitly, we have

$$\begin{aligned}
\phi_{(g,u)}(h, v) &= (g, u) * (h, v) * (g^{-1}, -\text{Ad}_g(u)) \\
&= (gh, \text{Ad}_{h^{-1}}(u) + v) * (g^{-1}, -\text{Ad}_g(u)) \\
&= (ghg^{-1}, \text{Ad}_g(\text{Ad}_{h^{-1}}(u) + v) - \text{Ad}_g(u)).
\end{aligned}$$

Let  $(\gamma, c)$  be a smooth curve on  $G \ltimes \mathfrak{g}$  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  $e$ , the adjoint action  $\text{Ad}_{(g,u)} : \mathfrak{g} \ltimes \mathfrak{g} \rightarrow \mathfrak{g} \ltimes \mathfrak{g}$  is given by

$$\begin{aligned}
\text{Ad}_{(g,u)}(\eta, v) &= \left. \frac{d}{dt} \right|_{t=0} \phi_g(\gamma(t)) \\
&= \left. \frac{d}{dt} \right|_{t=0} (g\gamma(t)g^{-1}, \text{Ad}_g(\text{Ad}_{\gamma(t)^{-1}}(u) + c(t)) - \text{Ad}_g(u)) \\
&= (\text{Ad}_g(\eta), \text{Ad}_g(-\text{ad}_{\dot{\gamma}(0)}(u) + \dot{c}(0))) \\
&= (\text{Ad}_g(\eta), \text{Ad}_g(-[\eta, u] + v)),
\end{aligned}$$

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

$$\begin{aligned}
& \text{Ad}_{\gamma(t)^{-1}} \circ \text{Ad}_{\gamma(t)} = \text{id} \\
& \frac{d}{dt}\bigg|_{t=0} \left( \text{Ad}_{\gamma(t)^{-1}} \circ \text{Ad}_{\gamma(t)} \right) = 0 \\
& \frac{d}{dt}\bigg|_{t=0} \text{Ad}_{\gamma(t)^{-1}} \circ \text{Ad}_{\gamma(0)} + \text{Ad}_{\gamma(0)^{-1}} \circ \text{ad}_{\dot{\gamma}(0)} = 0 \quad \text{chain rule} \\
& \frac{d}{dt}\bigg|_{t=0} \text{Ad}_{\gamma(t)^{-1}} \circ \text{Ad}_e + \text{Ad}_e \circ \text{ad}_{\dot{\gamma}(0)} = 0 \\
& \frac{d}{dt}\bigg|_{t=0} \text{Ad}_{\gamma(t)^{-1}} = -\text{ad}_{\dot{\gamma}(0)} \quad \text{Ad}_e = \text{id}.
\end{aligned}$$

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  $\text{Ad}_{(\gamma(t), c(t))}$  at  $e$  is the adjoint representation  $\text{ad}_{(\xi, u)} : \mathfrak{g} \ltimes \mathfrak{g} \rightarrow \mathfrak{g} \ltimes \mathfrak{g}$ :

$$\begin{aligned}
\text{ad}_{(\xi, u)}(\eta, v) &= \frac{d}{dt}\bigg|_{t=0} \text{Ad}_{(\gamma(t), c(t))}(\eta, v) \\
&= \frac{d}{dt}\bigg|_{t=0} \left( \text{Ad}_{\gamma(t)}(\eta), \text{Ad}_{\gamma(t)}(-[\eta, c(t)] + v) \right) \\
&= \left( \text{ad}_{\xi}(\eta), \text{ad}_{\xi}(-[\eta, c(0)]) + \text{Ad}_{\gamma(0)}(-[\eta, \dot{c}(0)]) + \text{ad}_{\xi}(v) \right) \\
&= ([\xi, \eta], [\xi, v] - [\eta, u]).
\end{aligned}$$

**Remark 1.3** The conventional semi-direct product has

$$(g, u) * (h, v) = (gh, u + \text{Ad}_g(v)).$$

The right trivialization indeed is a Lie group isomorphism to this conventional semidirect product structure. Moreover,  $\text{SE}(3)$  manifests this structure for  $\text{TSO}(3)$  through matrix multiplication. However, body-frame dynamics demand left-trivialization, so we are stuck with the weird one.

Since  $\text{SE}(3)$  corresponds to right trivialization, we need to find another matrix representation of the left trivialization. Consider

$$(g, u) * (h, v) = \begin{pmatrix} h^{-1} & v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} g^{-1} & u \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} (gh)^{-1} & h^{-1}u + v \\ 0 & 1 \end{pmatrix}.$$

We see that when we pass to the matrix, multiplication order is flipped in order to recover  $gh$  and match the semidirect product structure of left trivialization. This is the counterpart to  $SE(3)$  and the correct geometry to use for body-frame attitude dynamics. I do not know if it has a proper name so I will just refer to it as  $LSE(3)$ .

The Lie algebra elements have the form

$$\begin{pmatrix} -\xi^\wedge & v \\ 0 & 0 \end{pmatrix}.$$

This is clear from the fact that  $\exp(-A) = A^{-1}$ .

Now we have a matrix representation, we can easily compute the  $\text{Ad}$  and  $\text{ad}$  maps for this specific case using matrix multiplication.

$$\begin{aligned} \text{Ad}_{(R,u)}(\eta, v) &= (R, u) * (\eta, v) * (R, u)^{-1} \\ &= \begin{pmatrix} R & -RuR^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -\eta^\wedge & v \\ 0 & 0 \end{pmatrix} \begin{pmatrix} R^{-1} & u \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -R\eta^\wedge & Rv \\ 0 & 0 \end{pmatrix} \begin{pmatrix} R^{-1} & u \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -R\eta^\wedge R^{-1} & R(-\eta^\wedge u + v) \\ 0 & 0 \end{pmatrix} \\ &= (R\eta^\wedge R^{-1}, R(v - \eta^\wedge u)). \end{aligned}$$

Let  $\gamma(0) = I, \dot{\gamma}(0) = \xi^\wedge, c(0) = 0, \dot{c}(0) = u$ . Then

$$\begin{aligned} \text{ad}_{(\xi,u)}(\eta, u) &= \frac{d}{dt} \left( \gamma(t)\eta^\wedge\gamma^{-1}(t), \gamma(t)(v - \eta^\wedge c(t)) \right) \\ &= ([\xi, \eta], \dot{\gamma}(0)v - \dot{\gamma}(0)\eta^\wedge c(0) - \gamma(0)\eta^\wedge \dot{c}(0)) \\ &= ([\xi, \eta], \xi^\wedge v - \eta^\wedge u) \\ &= ([\xi, \eta], [\xi, v] - [\eta, u]). \end{aligned}$$

The results match those from the general case!

## 2 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 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, we rotate the differential contravariantly and get a vector field via  $\iota_{X_H}\omega$ .

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.

$$\begin{aligned}\dot{z} &= X_H(z) \\ \dot{f}(z) &= \langle df(z), X_H(z) \rangle \quad \forall f \\ \dot{f} &= \{f, H\} \quad \forall f\end{aligned}$$