

# Legged Robot State-Estimation Through Combined Forward Kinematic and Preintegrated Contact Factors

Ross Hartley, Josh Mangelson, Lu Gan, Maani Ghaffari Jadidi, Jeffrey M. Walls,  
Ryan M. Eustice, and Jessy W. Grizzle

**Abstract**—State-of-the-art robotic perception systems have achieved sufficiently good performance using Inertial Measurement Units (IMUs), cameras, and nonlinear optimization techniques, that they are now being deployed as technologies. However, many of these methods rely significantly on vision and fail when visual tracking is lost because of lighting or scarcity of features. In addition, such systems are formulated such that the system is independent of the operating platform. This paper presents state-estimation technique for legged robots that takes into account the robot’s kinematic model as well as its contact with the environment. This is achieved by introducing forward kinematic factors and preintegrated contact factors into a factor graph framework that can incrementally solved in real-time. The forward kinematics factor relates the robot’s base pose to a contact frame through noisy encoder measurements. The preintegrated contact factor provides odometry measurements of this contact frame, while accounting for possible foot slippage. Taken together, the two factors constrain the graph optimization problem allowing the robot’s trajectory to be estimated. The paper evaluates the method using simulated and real sensory IMU and kinematic data from experiments with a Cassie-series robot designed by Agility Robotics. These experiments show that using the proposed method in addition to IMU significantly decreases drift and improves localization accuracy, suggesting that its use can enable successful recovery from loss of visual tracking.

## I. INTRODUCTION AND RELATED WORK

Legged locomotion enables robots to adaptively operate in an unstructured and unknown environment with possibly rough and discontinuous ground; and arguably is one of the most advanced forms of locomotion in nature [1]. The state-of-the-art control algorithms for dynamic biped locomotion are capable of providing stabilizing feedback for a biped robot blindly walking through sinusoidally varying [2] or discrete [3] terrain; however, without perceiving the environment, the application of legged robots remains extremely limited. Accurate estimates of the robot’s state and environment are essential prerequisites for stable control and motion planning [4], [5], [6]. In addition, real-time performance of the state estimation and perception system are required to enable online decision making [7], [8].

Legged robots, unlike ground, flying, and underwater platforms, are in direct and switching contacts with the environment. Leg odometry involves estimating relative transformations using kinematic and contact information, which can be noisy due to the encoder noise and foot slip [9].

The authors are with the College of Engineering, University of Michigan, Ann Arbor, MI 48109 USA {rosshart, mangelson, ganlu, maanigj, eustice, grizzle}@umich.edu, with the exception of Jeff Walls who is with Toyota Research Institute.



Fig. 1: Experiments were conducted on a Cassie-series robot designed by Agility Robotics. The biped robot has 20 degrees of freedom, 10 actuators, joint encoders, an IMU, and a Multisense S7 stereo camera.

Thus, without a sound sensor fusion framework, the resulting localization or odometry trajectory is inaccurate as a consequence of significant drift over the traversed distance. State-of-the-art visual-inertial odometry systems [10], and generally Simultaneous Localization and Mapping (SLAM) [11], [12], use graphical models (factor graphs) and nonlinear optimization techniques to achieve a probabilistically sound sensor fusion framework and real-time performance by exploiting the sparse structure of the SLAM problem [13], [14]. The high-dynamic motion and noise characterization is captured using preintegration of high-frequency sensors such as Inertial Measurement Units (IMUs) [15]. Therefore, the factor graphs framework and the preintegration theory are mathematically suitable approaches for legged robot state estimation.

In this paper, we develop two novel factors that integrate the use of multi-link Forward Kinematics (FK) and the notion of contact between the robotic system and the environment into the factor graph smoothing framework. Multi-body platforms such as biped robots exhibit complex movement to achieve dynamic stability. The FK factor can model such movement that relates the contact frame to a sensor frame such as camera or IMU. On the other hand, the novel contact factor preintegrates high-frequency feet contact measurements that constrain the robot net movement despite its complicated kinematic configurations and change

of the supporting leg. By augmenting the state to include the contact frame of the feet at each time step we enable the estimator to provide valuable information to the gait controller regarding the changes of the terrain. In particular, this work has the following contributions:

- i. A FK factor that incorporates joint encoder noise to estimate the end-effector pose at any time-step;
- ii. Rigid and point preintegrated contact factors that constrain robot poses between successive time-steps while accommodating noise from foot slip;
- iii. Integration of leg odometry into the factor graph smoothing framework;
- iv. Real-time implementation of the proposed FK and preintegrated contact measurement models.

Section II provides the required preliminaries including the notation and mathematical prerequisites. We formulate the problem and our factor graph approach in Section III. Section IV explains forward kinematic modeling. The forward kinematic factor is developed in Section V. Rigid and point contact factors are derived in Sections VI and VII, respectively. Simulation and experimental evaluations of the proposed methods on a 3D biped robot (Fig. 1) are presented in Section VIII. Finally, Section IX concludes and provides future work suggestions.

## II. PRELIMINARIES

In this section we present preliminary materials necessary for the developments in the following sections. We first establish the mathematical notation where we assume readers are already familiar with basics of Lie groups, Lie Algebra, and optimization on matrix Lie groups [16], [17], [18]. In addition, as this work is partly motivated by IMU preintegration on manifold, we adopt the notation from [10], where possible, so that readers can connect the two papers more easily.

### A. Mathematical Notation and Background

Matrices are capitalized in bold, such as in  $\mathbf{X}$ , and vectors are in lower case bold type, such as in  $\mathbf{x}$ . Vectors are column-wise and  $[n]$  means the set of integers from 1 to  $n$ , i.e.  $\{1:n\}$ . The Euclidean norm is shown by  $\|\cdot\|$ .  $\|\mathbf{e}\|_{\Sigma}^2 \triangleq \mathbf{e}^T \Sigma^{-1} \mathbf{e}$ . The  $n$ -by- $n$  identity and zero matrices are denoted by  $\mathbf{I}_n$  and  $\mathbf{O}_n$ , respectively, and where it is evident simply by  $\mathbf{I}$  and  $\mathbf{O}$ .  $\mathbf{0}$  and  $\mathbf{1}$  denote the vector of zeros and ones with appropriate dimensions, respectively.  $\text{vec}(x_1, \dots, x_n)$  denotes a vector such as  $\mathbf{x}$  constructed by stacking  $x_i$ ,  $\forall i \in [n]$ .  $\text{Cov}[\cdot]$  denotes the covariance of a random vector. Finally, we denote the base frame of the robot by  $\mathbf{B}$ , the world frame by  $\mathbf{W}$ , and contact frames by  $\mathbf{C}$ . With the rotation matrix representing the orientation of the base in the world frame by  $\mathbf{R}_{\mathbf{WB}}$  and the position of the base in the world frame by  $\mathbf{w}\mathbf{p}_{\mathbf{B}}$ .

The *general linear group* of degree  $n$ , denoted by  $\text{GL}_n(\mathbb{R})$ , is the set of  $n \times n$  invertible matrices, where the group binary operation is the ordinary matrix multiplication. The *special orthogonal group*, denoted by  $\text{SO}(3) = \{\mathbf{R} \in \text{GL}_3(\mathbb{R}) | \mathbf{R}\mathbf{R}^T = \mathbf{I}, \det \mathbf{R} = 1\}$ ,

contains valid three-dimensional (3D) rotation matrices. The *special Euclidean group*, denoted by  $\text{SE}(3) = \{\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{p} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \in \text{GL}_4(\mathbb{R}) | \mathbf{R} \in \text{SO}(3), \mathbf{p} \in \mathbb{R}^3\}$ , is the 3D rigid body motion group. Let  $\boldsymbol{\omega} \triangleq \text{vec}(\omega_1, \omega_2, \omega_3)$ . The *hat* operator is defined as  $\boldsymbol{\omega}^\wedge \triangleq \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}$ , where the right hand side is known as the *skew-symmetric matrix*.

The Lie algebra of  $\text{SO}(3)$  is its tangent space at the identity together with Lie bracket and canonically determined as  $\mathfrak{so}(3) = \{\boldsymbol{\omega}^\wedge \in \text{GL}_3(\mathbb{R}) | \exp(t\boldsymbol{\omega}^\wedge) \in \text{SO}(3) \text{ and } t \in \mathbb{R}\}$ . Conversely, the *vee* operator maps a skew symmetric matrix to a vector in  $\mathbb{R}^3$ , that is  $(\boldsymbol{\omega}^\wedge)^\vee = \boldsymbol{\omega}$ . In addition,  $\forall \mathbf{a}, \mathbf{b} \in \mathbb{R}^3$  we have  $\mathbf{a}^\wedge \mathbf{b} = -\mathbf{b}^\wedge \mathbf{a}$ . The *exponential map* associates an element of the Lie Algebra around a neighborhood of zero to a rotation matrix in  $\text{SO}(3)$  around a neighborhood of the identity and can be derived as  $\exp(\phi^\wedge) = \mathbf{I} + \frac{\sin(\|\phi\|)}{\|\phi\|} \phi^\wedge + \frac{1 - \cos(\|\phi\|)}{\|\phi\|^2} (\phi^\wedge)^2$ ; and using a first-order approximation reduces to  $\exp(\phi^\wedge) \approx \mathbf{I} + \phi^\wedge$ . For any  $\|\phi\| < \pi$ , the *logarithm map* uniquely associates a rotation matrix in  $\text{SO}(3)$  to an element of its Lie algebra  $\mathfrak{so}(3)$ ;  $\log(\mathbf{R}) = \frac{\varphi(\mathbf{R} - \mathbf{R}^T)}{2 \sin(\varphi)}$  with  $\varphi = \cos^{-1}\left(\frac{\text{tr}(\mathbf{R}) - 1}{2}\right)$ . We use the following simplified notations from [19]:

$$\begin{aligned} \text{Exp} : \mathbb{R}^3 &\ni \boldsymbol{\phi} \rightarrow \exp(\boldsymbol{\phi}^\wedge) \in \text{SO}(3) \\ \text{Log} : \text{SO}(3) &\ni \mathbf{R} \rightarrow \log(\mathbf{R})^\vee \in \mathbb{R}^3 \end{aligned}$$

The following first order approximations of above maps allow for relating increments in the Lie Algebra to increments in the Lie group and vice versa [19]

$$\text{Exp}(\boldsymbol{\phi} + \delta\boldsymbol{\phi}) \approx \text{Exp}(\boldsymbol{\phi})\text{Exp}(\mathbf{J}_r(\boldsymbol{\phi})\delta\boldsymbol{\phi}) \quad (1)$$

$$\text{Log}(\text{Exp}(\boldsymbol{\phi})\text{Exp}(\delta\boldsymbol{\phi})) \approx \boldsymbol{\phi} + \mathbf{J}_r^{-1}(\boldsymbol{\phi})\delta\boldsymbol{\phi} \quad (2)$$

where  $\mathbf{J}_r(\boldsymbol{\phi}) = \mathbf{I} - \frac{1 - \cos(\|\boldsymbol{\phi}\|)}{\|\boldsymbol{\phi}\|^2} \boldsymbol{\phi}^\wedge + \frac{\|\boldsymbol{\phi}\| - \sin(\|\boldsymbol{\phi}\|)}{\|\boldsymbol{\phi}\|^3} (\boldsymbol{\phi}^\wedge)^2$  is the right Jacobian of  $\text{SO}(3)$  [17]. Finally, from the *adjoint representation* of  $\text{SO}(3)$  we have

$$\text{Exp}(\boldsymbol{\phi})\mathbf{R} = \mathbf{R}\text{Exp}(\mathbf{R}^T \boldsymbol{\phi}) \quad (3)$$

### B. Modeling Noise and Optimization on Matrix Lie Groups

We model the uncertainty in  $\text{SO}(3)$  by defining a noise distribution in the tangent space and then mapping it to  $\text{SO}(3)$  via the exponential map [20], [19];  $\tilde{\mathbf{R}} = \mathbf{R}\text{Exp}(\boldsymbol{\epsilon})$  and  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ , where  $\mathbf{R}$  is a given noise-free rotation (the mean) and  $\boldsymbol{\epsilon}$  is a small normally distributed perturbation with zero mean and covariance  $\Sigma$ . Through approximating the normalization factor as constant, the negative log-likelihood of a rotation  $\mathbf{R}$  given a measurement  $\tilde{\mathbf{R}}$  distributed according to the defined perturbation is  $\mathcal{L}(\mathbf{R}) \propto \frac{1}{2} \|\text{Log}(\mathbf{R}^{-1} \tilde{\mathbf{R}})\|_{\Sigma}^2 = \frac{1}{2} \|\text{Log}(\tilde{\mathbf{R}}^{-1} \mathbf{R})\|_{\Sigma}^2$  which is a *bi-invariant Riemannian metric*. For the translation part of  $\text{SE}(3)$ , noise can be characterized using the usual additive white Gaussian noise assumptions.

Solving an optimization problem where the cost function is defined on a manifold is slightly different than the usual problems in  $\mathbb{R}^n$ . The tangent vectors lie in the tangent space.

As such, a *retraction* that maps a vector in the tangent space to an element in the manifold is required [18]. Given a retraction mapping and the associated manifold, we can optimize over the manifold by iteratively *lifting* the cost function of our optimization problem to the tangent space, solving the reparameterized problem, and then mapping the updated solution back to the manifold using the retraction. For  $SO(3)$ , the exponential map is this retraction. For  $SE(3)$  we adopt the retraction used in [19]:

$$\mathcal{R}_T(\delta\phi, \delta\mathbf{p}) = (\mathbf{R}\text{Exp}(\delta\phi), \mathbf{p} + \mathbf{R}\delta\mathbf{p}), \quad \text{vec}(\delta\phi, \delta\mathbf{p}) \in \mathbb{R}^6 \quad (4)$$

### III. PROBLEM STATEMENT AND FORMULATION

In this section, we formulate the state estimation problem of the legged robot. The biped robot is equipped with a stereo camera, an IMU, joint encoders, and contact sensors in the feet. Without loss of generality we assume the IMU and camera are collocated with the base frame of the robot. In current state-of-the-art visual-inertial navigation systems, the state includes the camera pose and velocity, and system calibration parameters such as IMU bias [10]. In the factor graph framework, independent measurements from additional sensors can be incorporated by introducing additional factors based on the associated measurement models. Foot slip is the major source of drift in leg odometry; as such, to isolate the noise at the contact point we augment the state at time-step  $i$  to include the contact frame pose of both feet  $\mathbf{C}_i \triangleq \{\mathbf{R}_{\text{WC},l}(i)\}_{l=1}^2$  and  $\mathbf{d}_i \triangleq \{\mathbf{w}\mathbf{p}_{\text{C},l}(i)\}_{l=1}^2$ . (However, without loss of generality, all following derivations are for a single contact frame.) Thus, the state at any time-step  $i$  is represented as  $\mathcal{S}_i \triangleq \{\mathbf{R}_i, \mathbf{p}_i, \mathbf{v}_i, \mathbf{C}_i, \mathbf{d}_i, \mathbf{b}_i\}$ , where  $\mathbf{R}_i \triangleq \mathbf{R}_{\text{WB}}(i)$  and  $\mathbf{p}_i \triangleq \mathbf{w}\mathbf{p}_{\text{B}}(i)$  represent the base pose,  $\mathbf{v}_i \triangleq \mathbf{w}\mathbf{v}_{\text{B}}(i)$  is the base velocity, and  $\mathbf{b}_i \triangleq \mathbf{b}(i)$  is the IMU bias. In addition  $\mathcal{X}_k \triangleq \bigcup_{i=1}^k \mathcal{S}_i$  denotes the state up to time-step  $k$ .

Let  $\mathcal{L}_{ij} \in SE(3)$  be a perceptual loop closure measurement relating poses at time-steps  $i$  and  $j$  ( $j > i$ ) computed from the stereo camera raw data, e.g. using a point cloud matching algorithm. Let  $\alpha_i$  be the joint angles measurements at time-step  $i$ . The IMU and contact sensors provide measurements at higher frequencies. Between any two time-steps  $i$  and  $j$ , we denote the set of all IMU and contact-angles measurements by  $\mathcal{I}_{ij}$  and  $\mathcal{F}_{ij}$ , respectively.  $\mathcal{F}_{ij}$  includes measurements from the contact sensor as well as joint angles measurements,  $\alpha_{i-j}$ , between  $i$  and  $j$  excluding  $\alpha_i$  and  $\alpha_j$ . Let  $\mathcal{K}_k$  be the index set of time-steps (or key-frames) up to time-step  $k$ . We denote the set of all measurements up to time-step  $k$  by  $\mathcal{Z}_k \triangleq \{\mathcal{L}_{ij}, \alpha_i, \mathcal{I}_{ij}, \mathcal{F}_{ij}\}_{i,j \in \mathcal{K}_k}$ .

The posterior probability of the *full SLAM* problem by assuming the measurements are conditionally independent and are corrupted by additive zero mean white Gaussian noise can be written as  $p(\mathcal{X}_k | \mathcal{Z}_k) \propto p(\mathcal{X}_0) p(\mathcal{Z}_k | \mathcal{X}_k)$ . Furthermore,

$$p(\mathcal{Z}_k | \mathcal{X}_k) = \prod_{i,j \in \mathcal{K}_k} p(\mathcal{L}_{ij} | \mathcal{X}_j) p(\mathcal{I}_{ij} | \mathcal{X}_j) p(\alpha_i | \mathcal{X}_i) p(\mathcal{F}_{ij} | \mathcal{X}_j) \quad (5)$$

The *Maximum-A-Posteriori* (MAP) estimate of  $\mathcal{X}_k$  can be computed by solving the following optimization problem:

$$\underset{\mathcal{X}_k}{\text{minimize}} \quad -\log p(\mathcal{X}_k | \mathcal{Z}_k) \quad (6)$$

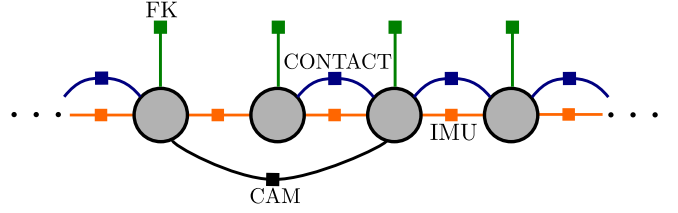


Fig. 2: An example factor graph for the proposed system. Forward kinematic factors are added at each node and constrain the pose of the contact frames on the feet of the robot with respect to the robot base. Contact factors are added into the graph over timesteps where the given contact frame remained in contact with environment. This enables the system to handle failures of the visual tracking or loop closure system.

in which due to the noise assumption mentioned earlier is equivalent to the following nonlinear least-squares problem:

$$\begin{aligned} \underset{\mathcal{X}_k}{\text{minimize}} \quad & \|\mathbf{r}_0\|_{\Sigma_0}^2 + \sum_{i,j \in \mathcal{K}_k} \|\mathbf{r}_{\mathcal{L}_{ij}}\|_{\Sigma_{\mathcal{L}_{ij}}}^2 + \sum_{i,j \in \mathcal{K}_k} \|\mathbf{r}_{\mathcal{I}_{ij}}\|_{\Sigma_{\mathcal{I}_{ij}}}^2 \\ & + \sum_{i \in \mathcal{K}_k} \|\mathbf{r}_{\alpha_i}\|_{\Sigma_{\alpha_i}}^2 + \sum_{i,j \in \mathcal{K}_k} \|\mathbf{r}_{\mathcal{F}_{ij}}\|_{\Sigma_{\mathcal{F}_{ij}}}^2 \end{aligned} \quad (7)$$

where  $p(\mathcal{X}_0)$  is the prior over the initial state and serves to anchor the graph,  $\mathbf{r}_0$ ,  $\mathbf{r}_{\mathcal{L}_{ij}}$ ,  $\mathbf{r}_{\mathcal{I}_{ij}}$ ,  $\mathbf{r}_{\alpha_i}$ ,  $\mathbf{r}_{\mathcal{F}_{ij}}$  are the residual terms associated with the loop closure, IMU, forward kinematic, and contact measurements respectively, i.e. the error between the measured and predicted values given the state, and  $\Sigma_0$ ,  $\Sigma_{\mathcal{L}_{ij}}$ ,  $\Sigma_{\mathcal{I}_{ij}}$ ,  $\Sigma_{\alpha_i}$ ,  $\Sigma_{\mathcal{F}_{ij}}$  are the corresponding covariance matrices.

### IV. FORWARD KINEMATICS

Forward kinematics refers to the process of computing the position and orientation of an end-effector frame using measurements of the robot's joint positions. This section derives a representation of the forward kinematic functions needed to compute the pose of a contact frame relative to base frame of the robot.

Let the *contact frame* be a coordinate frame on the robot located at a point of contact with the environment. This contact frame is separated from the robot's *base frame*, by  $N$  links. These  $N$  links are assumed to be connected by  $N - 1$  revolute joints, each equipped with an encoder to measure the joint angle. The vector of encoder angles is denoted by  $\alpha \in \mathbb{R}^{N-1}$ . In total, this amounts to  $N + 1$  frames, where frames 1 and  $N + 1$  are the base and contact frame respectively, as shown in Fig. 3. The forward kinematics can be computed through a product of homogeneous transforms:

$$\mathbf{H}(\alpha) = \prod_{n=1}^N \mathbf{H}_{n,n+1}(\alpha_n) \triangleq \begin{bmatrix} \mathbf{f}_R(\alpha) & \mathbf{f}_p(\alpha) \\ \mathbf{0}_3^T & 1 \end{bmatrix} \in SE(3) \quad (8)$$

The relative transformation between link frames  $n \in [N - 1]$  and  $n + 1$  is given by:

$$\mathbf{H}_{n,n+1}(\alpha_n) \triangleq \begin{bmatrix} \mathbf{A}_n \text{Exp}(\alpha_n^\dagger) & \mathbf{t}_n \\ \mathbf{0}_3^T & 1 \end{bmatrix} \quad (9)$$

where  $\mathbf{A}_n \in SO(3)$  and  $\mathbf{t}_n \in \mathbb{R}^3$  are a constant rotation and translation defined by the kinematic model of the robot. Each joint is assumed to be revolute with an angle  $\alpha_n \in \mathbb{R}$ . The *dagger* operator maps a scalar to a vector in  $\mathbb{R}^3$  based on

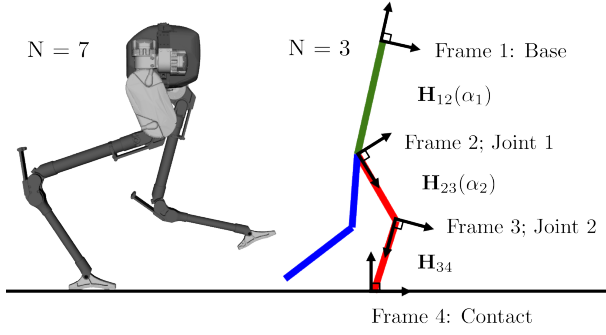


Fig. 3: The contact frame is separated from the robot's base frame by  $N$  links. Cassie (left) has 7 links between the base and the bottom of the toe. A simpler, planar biped (right) is shown to demonstrate the forward kinematics.

the joint's axis of rotation:

$$\alpha_n^\dagger \triangleq \text{vec}(\alpha_n, 0, 0), \text{vec}(0, \alpha_n, 0), \text{ or } \text{vec}(0, 0, \alpha_n) \quad (10)$$

The final transformation between link frames  $N$  and  $N+1$  is constant and denoted by  $\mathbf{H}_{N,N+1} \triangleq \begin{bmatrix} \mathbf{A}_N & \mathbf{t}_N \\ \mathbf{0}_3^\top & 1 \end{bmatrix}$ . When (8) is multiplied out, the orientation and position of the contact with respect to the base frame are:

$$\mathbf{f}_R(\alpha) = \mathbf{A}_{1,N+1}(\alpha) \quad (11)$$

$$\mathbf{f}_p(\alpha) = \sum_{n=1}^N \mathbf{A}_{1n}(\alpha) \mathbf{t}_n \quad (12)$$

where the relative rotation between links frames  $i$  and  $j$  is given by:

$$\mathbf{A}_{ij}(\alpha) \triangleq \begin{cases} \mathbf{I} & \text{for } (i, j) \in [N], i = j \\ \prod_{k=i}^{j-1} \mathbf{A}_k \text{Exp}(\alpha_k^\dagger) & \text{for } (i, j) \in [N], j > i \\ \mathbf{A}_{iN} \mathbf{A}_N & \text{for } i \in [N-1], j = N+1 \end{cases}$$

Changes in joint angles affect the orientation of all link frames further down the kinematic tree. To deal with encoder noise, it is important to understand how these angle offsets propagate through the forward kinematics functions.

**Lemma 1** (Relative rotation between two frames [21]). *Let  $\beta \in \mathbb{R}^N$  be a vector of joint angle offsets. The relative rotation between frames  $i$  and  $j$  can be factored into the nominal rotation and an offset rotation:*

$$\mathbf{A}_{ij}(\alpha + \beta) = \mathbf{A}_{ij}(\alpha) \prod_{k=i}^{j-1} \text{Exp}(\mathbf{A}_{k+1,j}^\top(\alpha) \beta_k^\dagger) \quad (13)$$

## V. FORWARD KINEMATICS FACTOR

The goal of this section is to formulate a general *forward kinematics factor* that accounts for uncertainty in joint encoder readings. This factor can be included in a factor graph to allow estimation of end-effector poses. Here, the end-effector coincides with the contact frame. We note that this factor only constrains the relative transformation between the base and contact frames; therefore, the full utility of this factor depends on a separate measurement of the contact frame described in Section VI.

### A. Contact Pose through Encoder Measurements

The encoder measurements are assumed to be affected by additive Gaussian noise,  $\eta^\alpha \sim \mathcal{N}(0, \Sigma^\alpha)$ .

$$\tilde{\alpha}(t) = \alpha(t) + \eta^\alpha(t) \quad (14)$$

The orientation and position of the contact frame in the world frame are given by:

$$\begin{aligned} \mathbf{R}_{WC}(t) &= \mathbf{R}_{WB}(t) \mathbf{R}_{BC}(t) \\ \mathbf{w}\mathbf{p}_C(t) &= \mathbf{w}\mathbf{p}_B(t) + \mathbf{R}_{WB}(t) \mathbf{B}\mathbf{p}_C(t) \end{aligned} \quad (15)$$

Rewriting (15) in terms of the state vector and encoder measurements yields:

$$\begin{aligned} \mathbf{C}(t) &= \mathbf{R}(t) \mathbf{f}_R(\tilde{\alpha}(t) - \eta^\alpha(t)) \\ \mathbf{d}(t) &= \mathbf{p}(t) + \mathbf{R}(t) \mathbf{f}_p(\tilde{\alpha}(t) - \eta^\alpha(t)) \end{aligned} \quad (16)$$

The isolation of the noise terms in orientation and position of the contact frame are derived in the following Lemmas. The dependence on time is assumed, so  $t$  is omitted for readability.

**Lemma 2** (FK factor orientation noise isolation [21]). *Using (11), (13), and (3), the rotation term and the noise quantity  $\delta \mathbf{f}_R$  can be derived as*

$$\begin{aligned} \mathbf{R}^\top \mathbf{C} &= \mathbf{f}_R(\tilde{\alpha} - \eta^\alpha) \\ &= \mathbf{f}_R(\tilde{\alpha}) \prod_{k=1}^{N-1} \text{Exp}(-\mathbf{A}_{k+1,N+1}^\top(\tilde{\alpha}) \eta_k^{\alpha^\dagger}) \triangleq \mathbf{f}_R(\tilde{\alpha}) \text{Exp}(-\delta \mathbf{f}_R) \end{aligned} \quad (17)$$

$$\begin{aligned} \delta \mathbf{f}_R &= -\text{Log} \left( \prod_{k=1}^{N-1} \text{Exp}(-\mathbf{A}_{k+1,N+1}^\top(\tilde{\alpha}) \eta_k^{\alpha^\dagger}) \right) \\ &\approx \sum_{k=1}^{N-1} \mathbf{A}_{k+1,N+1}^\top(\tilde{\alpha}) \eta_k^{\alpha^\dagger} \end{aligned} \quad (18)$$

Through repeated first order approximation, the noise quantity  $\delta \mathbf{f}_R$  is approximately zero mean and Gaussian.

**Lemma 3** (FK factor position noise isolation [21]). *Using (12), (13), (1), and anticommutativity of skew-symmetric matrices, the position term can be approximated as:*

$$\begin{aligned} \mathbf{R}^\top (\mathbf{d} - \mathbf{p}) &= \mathbf{f}_p(\tilde{\alpha} - \eta^\alpha) \\ &\approx \mathbf{f}_p(\tilde{\alpha}) + \sum_{k=1}^{N-1} \sum_{n=k}^{N-1} \mathbf{A}_{1,n+1}(\tilde{\alpha}) \mathbf{t}_{n+1}^\wedge \mathbf{A}_{k+1,n+1}^\top(\tilde{\alpha}) \eta_k^{\alpha^\dagger} \\ &\triangleq \mathbf{f}_p(\tilde{\alpha}) - \delta \mathbf{f}_p \end{aligned} \quad (19)$$

The noise quantity  $\delta \mathbf{f}_p$  is a linear combination of zero mean Gaussians, and is therefore also zero mean and Gaussian.

We can now write out the *forward kinematics measurement model*:

$$\begin{aligned} \mathbf{f}_R(\tilde{\alpha}) &= \mathbf{R}^\top \mathbf{C} \text{Exp}(\delta \mathbf{f}_R) \\ \mathbf{f}_p(\tilde{\alpha}) &= \mathbf{R}^\top (\mathbf{d} - \mathbf{p}) + \delta \mathbf{f}_p \end{aligned} \quad (20)$$

where the forward kinematics noise characterized by  $\text{vec}(\delta \mathbf{f}_R, \delta \mathbf{f}_p) \sim \mathcal{N}(0, \Sigma_f)$ .

### B. Unary Forward Kinematic Factor

The forward kinematics factor is a unary factor that relates the robot's base frame to an end-effector frame. Using (20),



we can write the residual errors,  $\mathbf{r}_{\alpha_i} \triangleq \text{vec}(\mathbf{r}_{f_{R_i}}, \mathbf{r}_{f_{p_i}})$  from (7), at time  $t_i$  as follows.

$$\begin{aligned}\mathbf{r}_{f_{R_i}} &= \text{Log} \left( \mathbf{f}_R(\tilde{\alpha}_i)^\top \mathbf{R}_i^\top \mathbf{C}_i \right) \\ \mathbf{r}_{f_{p_i}} &= \mathbf{R}_i^\top (\mathbf{d}_i - \mathbf{p}_i) - \mathbf{f}_p(\tilde{\alpha}_i)\end{aligned}\quad (21)$$

The forward kinematics noise can be rewritten as a linear system:

$$\begin{bmatrix} \delta \mathbf{f}_R \\ \delta \mathbf{f}_p \end{bmatrix} = \begin{bmatrix} \mathbf{Q} \\ \mathbf{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_1^{\alpha^\dagger} \\ \boldsymbol{\eta}_2^{\alpha^\dagger} \\ \vdots \\ \boldsymbol{\eta}_{N-1}^{\alpha^\dagger} \end{bmatrix}\quad (22)$$

where columns of the  $3 \times 3(N-1)$  matrices  $\mathbf{Q}$  and  $\mathbf{S}$  are given by:

$$\begin{aligned}\mathbf{Q}_i &= \mathbf{A}_{i+1, N+1}^\top (\tilde{\alpha}) \\ \mathbf{S}_i &= - \sum_{n=i}^{N-1} \mathbf{A}_{1, n+1}(\tilde{\alpha}) \mathbf{t}_{n+1}^\wedge \mathbf{A}_{k+1, n+1}^\top (\tilde{\alpha})\end{aligned}\quad (23)$$

The covariance can be computed using the linear noise model and the sensor covariance matrix  $\Sigma_{\eta^{\alpha^\dagger}}$  that describes the encoder noise  $\eta^{\alpha^\dagger}$ , i.e.  $\Sigma_{\eta^{\alpha^\dagger}} = \text{Cov}[\eta^{\alpha^\dagger}]$ :

$$\Sigma_f = \begin{bmatrix} \mathbf{Q} \\ \mathbf{S} \end{bmatrix} \Sigma_{\eta^{\alpha^\dagger}} \begin{bmatrix} \mathbf{Q}^\top & \mathbf{S}^\top \end{bmatrix}\quad (24)$$

The Jacobians for the forward kinematics factor are given in the supplementary material [21]. We note that  $\Sigma_f = \Sigma_{\alpha_i}$  as in (7).

## VI. RIGID CONTACT FACTOR

This section formulates a *contact factor* based on the assumption that the contact frame remains fixed with respect to the world frame over time. Slip is accommodated by incorporating noise on the contact frames' velocities. When combined with the forward kinematics factor introduced in Section V, an additional odometry measurement of the robot's base frame is obtained, which can improve the MAP estimate.

### A. Rigid Contact Model

In addition to the encoders, it is assumed that a separate sensor can measure when the robot is in contact with the static world. If the contact is rigid (6-DOF constraint), then both the angular and linear velocity of the contact frame are zero; i.e.  $c\omega_{\text{WC}}(t) = w\mathbf{v}_C(t) = 0$ . Therefore,

$$\begin{aligned}\dot{\mathbf{R}}_{\text{WC}}(t) &= \mathbf{R}_{\text{WC}}(t)(c\omega_{\text{WC}}(t) + \boldsymbol{\eta}^\omega(t))^\wedge = \mathbf{R}_{\text{WC}}(t)\boldsymbol{\eta}^{\omega^\wedge}(t) \\ w\dot{\mathbf{p}}_C(t) &= w\mathbf{v}_C(t) + \mathbf{R}_{\text{WC}}(t)\boldsymbol{\eta}^v(t) = \mathbf{R}_{\text{WC}}(t)\boldsymbol{\eta}^v(t)\end{aligned}\quad (25)$$

where  $\boldsymbol{\eta}^\omega \sim \mathcal{N}(0, \Sigma^\omega)$  and  $\boldsymbol{\eta}^v \sim \mathcal{N}(0, \Sigma^v)$  are additive Gaussian noise terms that capture contact slip. Both noise terms are represented in the contact frame and are rotated to align with the world frame. Rewriting (25) in terms of the state vector yields:

$$\begin{aligned}\dot{\mathbf{C}}(t) &= \mathbf{C}(t)\boldsymbol{\eta}^{\omega^\wedge}(t) \\ \dot{\mathbf{d}}(t) &= \mathbf{C}(t)\boldsymbol{\eta}^v(t)\end{aligned}\quad (26)$$

If the robot maintains rigid contact with the world from  $t$  to  $t + \Delta t$ , Euler integration can be applied to obtain the pose of the contact frame at time  $t + \Delta t$ .

$$\begin{aligned}\mathbf{C}(t + \Delta t) &= \mathbf{C}(t)\text{Exp}(\boldsymbol{\eta}^{\omega^\wedge}(t)\Delta t) \\ \mathbf{d}(t + \Delta t) &= \mathbf{d}(t) + \mathbf{C}(t)\boldsymbol{\eta}^{vd}(t)\Delta t\end{aligned}\quad (27)$$

Integrating from the initial time of contact,  $t_i$ , to the final time of contact,  $t_j$ , yields:

$$\begin{aligned}\mathbf{C}_j &= \mathbf{C}_i \prod_{k=i}^{j-1} \text{Exp}(\boldsymbol{\eta}_k^{\omega^\wedge} \Delta t) \\ \mathbf{d}_j &= \mathbf{d}_i + \sum_{k=i}^{j-1} \mathbf{C}_k \boldsymbol{\eta}_k^{vd} \Delta t\end{aligned}\quad (28)$$

where  $\boldsymbol{\eta}^{\omega^\wedge}$  and  $\boldsymbol{\eta}^{vd}$  are discrete time noise terms computed using the sampling time;  $\boldsymbol{\eta}^d = \frac{1}{\Delta t} \text{Cov}(\boldsymbol{\eta}(t))$ .

### B. Rigid Contact preintegration

We can now rearrange (28) to create relative increments that are independent of the state at times  $t_i$  and  $t_j$ .

$$\begin{aligned}\Delta \mathbf{C}_{ij} &= \mathbf{C}_i^\top \mathbf{C}_j = \prod_{k=i}^{j-1} \text{Exp}(\boldsymbol{\eta}_k^{\omega^\wedge} \Delta t) \\ \Delta \mathbf{d}_{ij} &= \mathbf{C}_i^\top (\mathbf{d}_j - \mathbf{d}_i) = \sum_{k=i}^{j-1} \Delta \mathbf{C}_{ik} \boldsymbol{\eta}_k^{vd} \Delta t\end{aligned}\quad (29)$$

Next, we wish to isolate the noise terms. First, we will deal with the rotation of the contact frame. The product of multiple incremental rotations can be expressed as one larger rotation. Therefore,

$$\Delta \mathbf{C}_{ij} \triangleq \Delta \tilde{\mathbf{C}}_{ij} \text{Exp}(-\delta \boldsymbol{\theta}_{ij})\quad (30)$$

where  $\Delta \tilde{\mathbf{C}}_{ij} = \mathbf{I}$  due to the rigid contact assumption. Furthermore, through repeated first order approximation,  $\delta \boldsymbol{\theta}_{ij}$  is approximately zero mean and Gaussian.

$$\delta \boldsymbol{\theta}_{ij} = -\text{Log} \left( \prod_{k=i}^{j-1} \text{Exp}(\boldsymbol{\eta}_k^{\omega^\wedge} \Delta t) \right) \approx \sum_{k=i}^{j-1} \boldsymbol{\eta}_k^{\omega^\wedge} \Delta t\quad (31)$$

Now, we can isolate the noise in the position of the contact frame by substituting (30) into  $\Delta \mathbf{d}_{ij}$  and dropping the higher order noise terms.

$$\begin{aligned}\Delta \mathbf{d}_{ij}^{\text{eq.(1)}} &\approx \sum_{k=i}^{j-1} \Delta \tilde{\mathbf{C}}_{ik} (\mathbf{I} - \delta \boldsymbol{\theta}_{ik}^\wedge) \boldsymbol{\eta}_k^{vd} \Delta t \approx - \sum_{k=i}^{j-1} \boldsymbol{\eta}_k^{vd} \Delta t \\ &\triangleq \Delta \tilde{\mathbf{d}}_{ij} - \delta \mathbf{d}_{ij}\end{aligned}\quad (32)$$

where  $\Delta \tilde{\mathbf{d}}_{ij} = 0$ . The noise term,  $\delta \mathbf{d}_{ij} = \sum_{k=i}^{j-1} \boldsymbol{\eta}_k^{vd} \Delta t$ , is zero mean and Gaussian.

Finally, we arrive at the *preintegrated contact measurement model*:

$$\begin{aligned}\Delta \tilde{\mathbf{C}}_{ij} &= \mathbf{C}_i^\top \mathbf{C}_j \text{Exp}(\delta \boldsymbol{\theta}_{ij}) = \mathbf{I} \\ \Delta \tilde{\mathbf{d}}_{ij} &= \mathbf{C}_i^\top (\mathbf{d}_j - \mathbf{d}_i) + \delta \mathbf{d}_{ij} = \mathbf{0}\end{aligned}\quad (33)$$

where the rigid contact noise characterized by  $\text{vec}(\delta \boldsymbol{\theta}_{ij}, \delta \mathbf{d}_{ij}) \sim \mathcal{N}(0, \Sigma_{ij})$ . More detailed derivations are provided in the supplementary material [21].

### C. Preintegrated Rigid Contact Factor

Once the noise terms are separated out, we can write down the residual errors,  $\mathbf{r}_{\mathcal{F}_{ij}} = \text{vec}(\mathbf{r}_{\Delta \mathbf{C}_{ij}}, \mathbf{r}_{\Delta \mathbf{d}_{ij}})$  from (7), as follows.

$$\begin{aligned}\mathbf{r}_{\Delta \mathbf{C}_{ij}} &= \text{Log}(\mathbf{C}_i^\top \mathbf{C}_j) \\ \mathbf{r}_{\Delta \mathbf{d}_{ij}} &= \mathbf{C}_i^\top (\mathbf{d}_j - \mathbf{d}_i)\end{aligned}\quad (34)$$

Furthermore, since both noise terms are simply additive Gaussians, the covariance can easily be computed. If the contact noise is constant, the covariance is simply:

$$\Sigma_{ij} = \begin{bmatrix} \Sigma_{\eta^\omega} & \mathbf{O}_3 \\ \mathbf{O}_3 & \Sigma_{\eta^v} \end{bmatrix} \Delta t_{ij} \quad (35)$$

where  $\Sigma_{\eta^\omega}$  and  $\Sigma_{\eta^v}$  are the continuous covariance matrices of the contact frame's angular and linear velocities,  $\eta^\omega$  and  $\eta^v$ , and  $\Delta t_{ij} = \sum_{k=i}^j \Delta t$ . We note that  $\Sigma_{ij} = \Sigma_{\mathcal{F}_{ij}}$  as in (7).

If the contact noise is time-varying, the covariance can be computed iteratively. This would be particularly useful if the noise was modeled to depend on contact pressure. The iterative noise propagation and the Jacobians for the contact factor are fully derived in the supplementary material [21].

## VII. POINT CONTACT FACTOR

The rigid contact factor can be modified to support additional contact types. For a point contact, the contact frame position remains fixed in the world frame; however, the orientation can change over time. Therefore,  $\Delta \tilde{\mathbf{C}}_{ij} \neq \mathbf{I}$ , and subsequently becomes unobservable using only encoder and contact measurements. To track the contact noise appropriately, a gyroscope needs to be used.

In the following section, the preintegrated contact factor is formulated as an extension to preintegrated IMU factor described in [19]. This allows support for point contacts in our factor graph formulation.

### A. Point Contact Model

A point contact is defined as a 3-DOF constraint on the position of the contact frame. All rotational degree of freedoms are unconstrained. Since  $\Delta \tilde{\mathbf{C}}_{ij}$  is unobservable without the use of a gyroscope, we shall remove this term in the preintegrated contact factor by utilizing the preintegrated IMU measurements. Replacing  $\mathbf{C}_k$  in (28) with its definition in (15), yields

$$\mathbf{d}_j = \mathbf{d}_i + \sum_{k=i}^{j-1} \mathbf{R}_k \mathbf{f}_R(\tilde{\alpha}_k - \eta_k^\alpha) \eta_k^{vd} \Delta t \quad (36)$$

We can rewrite this equation to be independent of the state at time  $t_i$  and  $t_j$ .

$$\mathbf{R}_i^\top (\mathbf{d}_j - \mathbf{d}_i) = \sum_{k=i}^{j-1} \Delta \mathbf{R}_{ik} \mathbf{f}_R(\tilde{\alpha}_k - \eta_k^\alpha) \eta_k^{vd} \Delta t \quad (37)$$

where  $\Delta \mathbf{R}_{ij} = \mathbf{R}_i^\top \mathbf{R}_j = \prod_{i=1}^{j-1} \text{Exp}((\tilde{\omega}_k - \mathbf{b}_k^g - \eta_k^{gd}) \Delta t)$  is the relative rotation increment from the IMU preintegration model [19].

The encoder measurements at time  $t_i$  is already being used for the forward kinematic factor (Section V). Therefore, to prevent information double counting, the first term in the summation can be replaced with the state estimate at time  $t_i$ .

$$\begin{aligned}\mathbf{R}_i^\top (\mathbf{d}_j - \mathbf{d}_i) &= \\ \mathbf{R}_i^\top \mathbf{C}_i \eta_i^{vd} \Delta t &+ \sum_{k=i+1}^{j-1} \Delta \mathbf{R}_{ik} \mathbf{f}_R(\tilde{\alpha}_k - \eta_k^\alpha) \eta_k^{vd} \Delta t\end{aligned}\quad (38)$$

We can now derive the *preintegrated point contact position measurement*  $\tilde{\Delta \mathbf{d}}_{ij}$  and its noise  $\delta \mathbf{d}_{ij}$  through first-order approximation of (38).

$$\begin{aligned}\mathbf{R}_i^\top (\mathbf{d}_j - \mathbf{d}_i) &\approx \mathbf{R}_i^\top \mathbf{C}_i \eta_i^{vd} \Delta t + \sum_{k=i+1}^{j-1} \Delta \tilde{\mathbf{R}}_{ik} \mathbf{f}_R(\tilde{\alpha}_k) \eta_k^{vd} \Delta t \\ &\triangleq \tilde{\Delta \mathbf{d}}_{ij} - \delta \mathbf{d}_{ij}\end{aligned}\quad (39)$$

where, again,  $\tilde{\Delta \mathbf{d}}_{ij} = \mathbf{0}$ . The full derivation is detailed in the supplementary material [21].

Up to a first order approximation,  $\delta \mathbf{d}_{ij}$  is zero mean, Gaussian, and does not depend on the encoder or IMU noise, i.e. it is decoupled [21]. Since the contact measurement is uncorrelated to the IMU / forward kinematics measurements and all are jointly Gaussian, the measurements are all independent. Therefore, the point contact factor can be written separate to the IMU factor.

### B. Preintegrated Point Contact Factor

Once the noise terms are separated out, we can write down the residual error:

$$\mathbf{r}_{\Delta \mathbf{d}_{ij}} = \mathbf{R}_i^\top (\mathbf{d}_j - \mathbf{d}_i) \quad (40)$$

The noise propagation can be written in an iterative form:

$$\delta \mathbf{d}_{ik+1} = \begin{cases} \delta \mathbf{d}_{ik} - \mathbf{R}_k^\top \mathbf{C}_k \eta_k^{vd} \Delta t & \text{for } k = i \\ \delta \mathbf{d}_{ik} - \Delta \tilde{\mathbf{R}}_{ik} \mathbf{f}_R(\tilde{\alpha}_k) \eta_k^{vd} \Delta t & \text{for } k > i \end{cases} \quad (41)$$

which allows us to write the covariance propagation as a linear system (starting with  $\Sigma_{ii} = \mathbf{O}_3$ ):

$$\Sigma_{ik+1} = \Sigma_{ik} + \mathbf{B} \Sigma_{\eta^v} \mathbf{B}^\top \quad (42)$$

where,

$$\mathbf{B} = \begin{cases} \mathbf{R}_k^\top \mathbf{C}_k \Delta t & \text{for } k = i \\ \Delta \tilde{\mathbf{R}}_{ik} \mathbf{f}_R(\tilde{\alpha}_k) \Delta t & \text{for } k > i \end{cases} \quad (43)$$

The covariance of the discrete contact velocity noise  $\eta_k^{vd}$  is denoted by  $\Sigma_{\eta^v}$ . These equations are derived in complete detail in the supplementary material [21].

## VIII. SIMULATIONS AND EXPERIMENTAL RESULTS

In this section, we evaluate the proposed method and factors. We implemented both factors in GTSAM [22] using iSAM2 as the solver [23]. To handle IMU preintegration we used the implementation built into GTSAM4 [10], [24].

Both simulation and real-world experiments used the Cassie-series robot developed by Agility Robotics shown in Fig. 1.

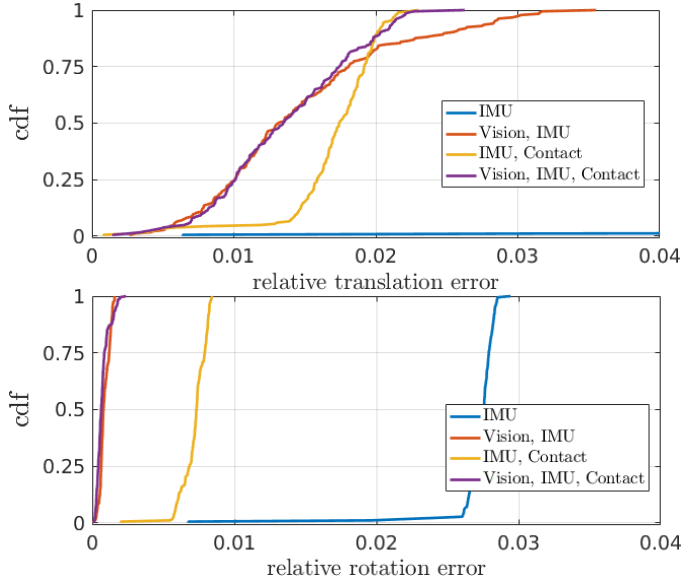


Fig. 4: Cumulative distribution function of the norm translation and rotation error for consecutive poses generated using a model of Cassie in SimMechanics. Measuring error in this way allows us to evaluate the drift of the different combinations of factors. Forward kinematics, contact, and IMU significantly outperform IMU and even outperform camera and IMU in some cases.

TABLE I: Simulation Noise Parameters

Noise	st. dev.
Acceleration	0.0307 m/s <sup>2</sup>
Angular Velocity	0.0014 rad/s
Accelerometer Bias	0.005 m/s <sup>2</sup>
Gyroscope Bias	0.0005 rad/s
Visual Odometry Translation	0.1 m
Visual Odometry Rotation	0.0873
Contact Linear Velocity	0.1 m/s
Joint Encoders	0.00873 rad

#### A. Simulated Evaluation using SimMechanics

For initial evaluation, we used SimMechanics to simulate a full model of Cassie traversing a variety of trajectories. The simulator provided true acceleration, angular velocity, joint angle, and contact values as well as ground-truth trajectory position and velocity.

We generated IMU measurements by adding Gaussian noise and bias according to the models described in [10]. We also simulated stereo vision based loop closure measurements using the ground-truth trajectory, and corrupted them using the methods detailed in Section II-B. Finally, we added white Gaussian noise to the contact and joint angle values to generate forward kinematic and contact measurements. Table I shows the noise parameters used.

In this experiment we compared the following combinations of factors: IMU, Vision and IMU, IMU and Contact/FK, Vision and IMU and Contact/FK. Fig. 4 shows the cumulative distribution function of the norm of the translation and rotation error computed by comparing the relative pose between consecutive time-steps of the trajectory estimated by each combination of methods to the ground-truth. This metric allowing us to evaluate the drift of the

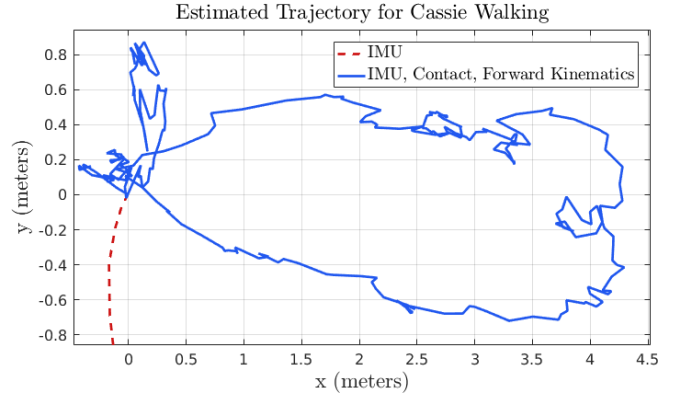


Fig. 5: Estimated trajectory of Cassie experiment data using IMU, forward kinematics, and contact factors. The robot walked in a loop around the lab, starting and ending at approximately the same pose. The video of the experiment is shown at <https://youtu.be/QnFoMR47OBI>. Compared to the IMU only estimation (dashed red line), the addition of contact and forward kinematics factors significantly improves the state estimate. The end-to-end translation error was approximately 1.3 m.

different combinations of factor combinations. The translational error is computed in the usual manner. For rotation we computed the norm of  $\|\text{Log}(\mathbf{R}_{\text{true}}^T \mathbf{R}_{\text{est}})\|$ .

Forward kinematic, contact, and IMU combined, significantly outperform IMU alone. This is partly because the contact and forward kinematic factors constrain the graph enabling the estimator to solve for the IMU bias. Also, note that in translation about 20 % of the time the combination of our proposed factors and the IMU outperforms the combination of the camera and the IMU. Finally, note that the combination of all factors is the most successful.

These results suggest that the proposed method can be used to increase the overall localization accuracy of the system as well as to handle drop-outs in visual tracking.

#### B. Real-world

We evaluated our factor graph implementation using real measurement data collected from a Cassie-series robot that was outfitted with a Multisense S7 camera (developed by Carnegie Robotics). This data included IMU measurements (from the Multisense camera) and joint encoders values. Cassie has two springs located on the each leg that are compressed when the robot is standing on the ground. Contact is computed using measurements of these spring deflections.

Using a controller provided by Agility Robotics, we walked Cassie for about 100 seconds in a loop around a 4.5 meter section of our lab, starting and ending in approximately the same location. Figure 5 compares the estimated trajectory computed using IMU, contact and forward kinematics factors with using only IMU factors. Due to noisy measurements, integration errors, and IMU bias, the IMU only estimate quickly drifts far away from the actual trajectory. After including the contact and kinematic factors, the state estimate is significantly improved resulting in a final translation error of approximately 1.3 m.

## IX. CONCLUSION

We developed two novel factors within the factor graphs framework to provide reliable leg odometry for bipedal locomotion. In particular, the forward kinematic factor computes measurements in the form of relative 3D transformation between the base and contact frames, while the preintegrated contact factor incorporates high-frequency contact measurements as constraints between any two successive poses in the graph. The evaluation showed that in the absence of accurate sensory data such as those from stereo camera, the probabilistic fusion of IMU, FK, and contact factors enable the robot to track its trajectory accurately.

We are currently in the process of developing the open source implementation of the proposed techniques and performing additional evaluation in real experiments. In the future, we plan to use these techniques to build up both local and global maps that can be used for path planning as well as to inform the robot controller about the surrounding terrain.

## ACKNOWLEDGMENT

The authors thank Agility Robotics for both designing the robot and the walking controller used for the reported experiments. Funding for R. Hartley, L. Gan, and M. Ghaffari Jadidi is given by the Toyota Research Institute (TRI), partly under award number N021515, however this article solely reflects the opinions and conclusions of its authors and not TRI or any other Toyota entity. Funding for J. Mangelson is provided by the Office of Naval Research (ONR) under award N00014-16-1-2102. Funding for J. Grizzle was in part provided by TRI and in part by NSF Award No. 1525006.

## REFERENCES

- [1] E. R. Westervelt, J. W. Grizzle, C. Chevallereau, J. H. Choi, and B. Morris, *Feedback control of dynamic bipedal robot locomotion*. CRC press, 2007.
- [2] X. Da, R. Hartley, and J. W. Grizzle, "First steps toward supervised learning for underactuated bipedal robot locomotion, with outdoor experiments on the wave field," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2017, pp. 3476–3483.
- [3] Q. Nguyen, A. Agrawal, X. Da, W. C. Martin, H. Geyer, J. W. Grizzle, and K. Sreenath, "Dynamic walking on randomly-varying discrete terrain with one-step preview," in *Robotics: Science and Systems*, 2017.
- [4] M. Bloesch, M. Hutter, M. A. Hoepflinger, S. Leutenegger, C. Gehring, C. D. Remy, and R. Siegwart, "State estimation for legged robots-consistent fusion of leg kinematics and IMU," in *Robotics: Science and Systems*, Berlin, Germany, June 2013.
- [5] P. Fankhauser, M. Bloesch, C. Gehring, M. Hutter, and R. Siegwart, "Robot-centric elevation mapping with uncertainty estimates," in *International Conference on Climbing and Walking Robots*, 2014, pp. 433–440.
- [6] M. A. Bloesch, "State estimation for legged robots-kinematics, inertial sensing, and computer vision," Ph.D. dissertation, 2017.
- [7] M. F. Fallon, M. Antone, N. Roy, and S. Teller, "Drift-free humanoid state estimation fusing kinematic, inertial and lidar sensing," in *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on*. IEEE, 2014, pp. 112–119.
- [8] S. Nobili, M. Camurri, V. Barasuol, M. Focchi, D. G. Caldwell, C. Semini, and M. Fallon, "Heterogeneous sensor fusion for accurate state estimation of dynamic legged robots," in *Robotics: Science and Systems*, 2017.
- [9] G. P. Roston and E. P. Krotkov, "Dead reckoning navigation for walking robots," Carnegie-Mellon University, Pittsburgh, PA, Robotics Institute, Tech. Rep., 1991.
- [10] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," *IEEE Trans. Robot.*, vol. 33, no. 1, pp. 1–21, 2017.
- [11] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robot. Automat. Mag.*, vol. 13, no. 2, pp. 99–110, 2006.
- [12] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [13] S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, "Simultaneous localization and mapping with sparse extended information filters," *The Int. J. Robot. Res.*, vol. 23, no. 7-8, pp. 693–716, 2004.
- [14] R. M. Eustice, H. Singh, and J. J. Leonard, "Exactly sparse delayed-state filters for view-based slam," *IEEE Trans. Robot.*, vol. 22, no. 6, pp. 1100–1114, 2006.
- [15] T. Lupton and S. Sukkarieh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," *IEEE Trans. Robot.*, vol. 28, no. 1, pp. 61–76, 2012.
- [16] J. M. Lee, *Introduction to smooth manifolds*. Springer, 2015.
- [17] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*. Springer Science & Business Media, 2011.
- [18] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.
- [19] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation," in *Robotics: Science and Systems*, Rome, Italy, July 2015.
- [20] T. D. Barfoot, *State Estimation for Robotics*. Cambridge University Press, 2017.
- [21] R. Hartley, J. Mangelson, L. Gan, M. Ghaffari Jadidi, J. M. Walls, R. M. Eustice, and J. W. Grizzle, "Supplementary material: legged robot state-estimation through combined kinematic and preintegrated contact factors," University of Michigan, Tech. Rep., Sept. 2017. [Online]. Available: <http://web.eecs.umich.edu/faculty/grizzle/papers/ICRASupplement2018MRH.pdf>
- [22] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Georgia Institute of Technology, Tech. Rep., 2012.
- [23] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *The Int. J. Robot. Res.*, vol. 31, no. 2, pp. 216–235, 2012.
- [24] L. Carlone, Z. Kira, C. Beall, V. Indelman, and F. Dellaert, "Eliminating conditionally independent sets in factor graphs: A unifying perspective based on smart factors," in *Proc. IEEE Int. Conf. Robot. Automat.*. IEEE, 2014, pp. 4290–4297.