2017 IEEE International Conference on Robotics and Automation (ICRA)
Singapore, May 29 - June 3, 2017

Acoustic-Inertial Underwater Navigation

Yulin Yang and Guoquan Huang

Abstract— **In this paper, we introduce a novel acoustic-inertial navigation system (AINS) for Autonomous Underwater Vehicles (AUVs). We are aiming to reduce the cost and latency of current underwater navigation systems that typically employ high-accuracy and thus high-cost inertial sensors. In particular, the proposed approach efficiently fuses the acoustic observations from a 2D imaging sonar and the inertial measurements from a MEMS inertial measurement unit (IMU) within a tightly-coupled EKF framework, while having *no* need to keep the acoustic features in the state vector. As a result, the computational complexity of the proposed AINS is independent from the scale of the operating environment. Moreover, we develop an acoustic feature linear triangulation to provide accurate initial estimates for iterative solvers, and perform an in-depth observability analysis to investigate the effects of sensor motion on the triangulation. Additionally, since it is challenging to perform *a priori* sensor extrinsic calibration underwater, we advocate to calibrate IMU-sonar online. The proposed AINS has been validated extensively in Monte-Carlo simulations.**

I. Introduction

Over the years, there has been increasingly growing demands of Autonomous Underwater Vehicle (AUVs) for a wide range of applications, such as seabed mapping, deep ocean exploring, routine harbor monitoring and oil pipeline maintenance. To successfully accomplish these tasks, an efficient and accurate localization solution is required for AUVs. However, this is challenging for underwater navigation, in part because GPS signal cannot be received underwater, and acoustic beacons require tedious and costly installation before applications. Although high accuracy inertial sensors, such as Doppler Velocity Loggers (DVLs) and fiber optic gyroscopes (FOGs), may provide good localization, the high cost limits their widespread deployments.

Due to the water turbidity and weak illumination in underwater environments, optical cameras only have limited applications and the relatively less expensive two dimensional (2D) forward-looking sonar (FLS) is preferable, which has larger field of view (FOV) and faster operating frequency [1][2], and is often used for short-range underwater detection and imaging [3]. For this reason, substantial research efforts have been taken on sonar-based underwater navigation. Walter et al. [4] used Exactly Sparse Extended Information Filter (ESEIF) with manually extracted sonar features to estimated the trajectory of AUV. Johannsson et al. [5] and Hover et al. [6] both adopted incremental smoothing and mapping (iSAM [7]) to estimate the vehicle motion and produce the environment maps for harbor surveillance and ship hull inspection. In their work automatic feature extraction and Normal Distribution Transformation (NDT) based image registration were introduced. This is different from Hurtos et al. [8], who registered images with Fourier-based methods. Aykin et al. [9] improved [5] by using Gaussian Distribution Transform for image registration instead of NDT. Based on that, Negahdaripour [10] integrated visual cues from acoustic shadows of stationary objects and devised 3D sonar motion estimation solution. Mallios et al. [11] utilized two extended Kalman filters (EKFs) together with a mechanical scanning imaging sonar (MSIS) to solve the full simultaneous localization and mapping (SLAM) problem. Assalih [12] tried a similar idea of stereo vision. Instead of optical cameras, the author imaged with two sonars and estimated the sonar motion by matching corresponding acoustic features between image pairs. Similarly, Negahdaripour et al. [13] proposed an opti-acoustic stereo system, which combined both a DIDSON sonar and an optical camera. But this system is not applicable when there exists strong water turbidity. Based on bundle adjustment (BA) [14], Huang et al. [15] introduced acoustic structure from motion (ASFM), which uses multiple sonar viewpoints to reconstruct 3D structure as well as the motion of sonar.

In contrast to the aforementioned work, in this paper, rather than solely relying on acoustic/optical images, we propose to employ an acoustic sonar and IMU to develop a low-cost acoustic-inertial navigation system (AINS). The proposed AINS can efficiently fuse acoustic measurements from a 2D imaging sonar and inertial measurements from a MEMS IMU within a tightly-coupled EKF framework. In particular, the main theoretical contributions of this work are as follows:

- We develop a novel acoustic-inertial odometry algorithm to fuse acoustic and inertial information without keeping the sonar features in the state vector. Thus, the computational complexity of the proposed approach is independent of the number of features observed.
- We propose an acoustic feature triangulation method to provide accurate initial estimates for the iterative algorithms for solving the corresponding least-squares problem. Moreover, an in-depth observability analysis is conducted to examine the effects of sensor motion on acoustic feature triangulation.
- We perform online extrinsic calibration between the sonar and the IMU, because it is often challenging in practice to pre-calibrate these sensors operating in

This work was partially supported by the University of Delaware College of Engineering, UD Cybersecurity Initiative, the Delaware NASA/EPSCoR Seed Grant, the NSF (IIS-1566129), and the DTRA (HDTRA1-16-1-0039).

The authors are with the Department of Mechanical Engineering, University of Delaware, Newark, DE 19716, USA. Email: {yuyang, ghuang}@udel.edu

underwater environments.

II. Problem Statement

In this work, we consider a low-cost AUV navigating underwater equipped with a 2D forward looking sonar (FLS) and a MEMS IMU and aim to efficiently localize the vehicle only using these onboard sensor measurements. In what follows, we briefly describe the IMU kinematic model and the acoustic sonar measurement model within the EKF framework, which will serve as the basis for our proposed AINS.

A. IMU Kinematic Model

The IMU navigation state $\mathbf{x}_{\mathbf{IMU}}$ is given by [16]:

$$\mathbf{x}_{\mathbf{IMU}} = \begin{bmatrix} {}^I_G\bar{q}^T & \mathbf{b}_g^T & {}^G\mathbf{v}_I^T & \mathbf{b}_a^T & {}^G\mathbf{p}_I^T \end{bmatrix}^T \quad (1)$$

where ${}^I_G\bar{q}$ is the unit quaternion representing the rotation from the global frame $\{G\}$ to the current IMU frame $\{I\}$ [16]. b_g and b_a are gyroscope and accelerometer biases for IMU measurements, respectively. ${}^G\mathbf{v}_I$ and ${}^G\mathbf{p}_I$ are the IMU velocity and position in the global frame $\{G\}$.

The time evolution of the IMU is described as [16], [17]:

$$\begin{aligned} {}^I_G\dot{\bar{q}}(t) &= \frac{1}{2}\mathbf{\Omega}(\omega_m(t) - \mathbf{b}_g(t) - \mathbf{n}_g(t)){}^I_G\bar{q}(t) && (2)\\ \dot{\mathbf{b}}_g(t) &= \mathbf{n}_{\omega g}(t) && (3)\\ {}^G\dot{\mathbf{v}}_I(t) &= \mathbf{R}^T({}^I_G\bar{q})({}^I\mathbf{a}_m - \mathbf{b}_a(t) - \mathbf{n}_a(t)) + {}^G\mathbf{g} && (4)\\ \dot{\mathbf{b}}_a(t) &= \mathbf{n}_{\omega a}(t) && (5)\\ {}^G\dot{\mathbf{p}}_I(t) &= {}^G\dot{\mathbf{v}}_I(t) && (6) \end{aligned}$$

where $\mathbf{\Omega}(\omega) = \begin{bmatrix} -\lfloor\omega\times\rfloor & \omega \\ -\omega^T & 0 \end{bmatrix}$, and $\mathbf{R}({}^I_G\bar{q})$ represents the rotation matrix corresponding to ${}^I_G\bar{q}$. ω_m and ${}^I\mathbf{a}_m$ are the direct measurements of angular velocity and linear acceleration from IMU, while $\mathbf{n}_g$ and $\mathbf{n}_a$ denote the white Gaussian noises that corrupt the corresponding measurements. $\mathbf{n}_{\omega g}$ and $\mathbf{n}_{\omega a}$ represents white Gaussian noise vectors driving the IMU biases $\mathbf{b}_g$ and $\mathbf{b}_a$. Note that in real underwater system, the gravity term ${}^G\mathbf{g}$ needs to be re-calibrated due to the buoyancy of AUV.

B. Sonar Measurement Model

An imaging sonar (e.g., FLS) provides ranges and azimuth angles to features in the underwater surroundings. The acoustic measurement model is depicted in Figure 1. We assume a single feature f_j has been observed and tracked in a set of n sonar frames, where the set is denoted as M_j. We denote f_j in the i-th frame of M_j as ${}^{S_i}\mathbf{p}_{f_j}$, and in the global frame as ${}^G\mathbf{p}_{f_j}$, then:

$${}^{S_i}\mathbf{p}_{f_j} = \begin{bmatrix} {}^{S_i}x_j \\ {}^{S_i}y_j \\ {}^{S_i}z_j \end{bmatrix} = \begin{bmatrix} r^{(j)}_{S_i}\cos\phi^{(j)}_{S_i}\cos\theta^{(j)}_{S_i} \\ r^{(j)}_{S_i}\sin\phi^{(j)}_{S_i}\cos\theta^{(j)}_{S_i} \\ r^{(j)}_{S_i}\sin\theta^{(j)}_{S_i} \end{bmatrix} \quad (7)$$

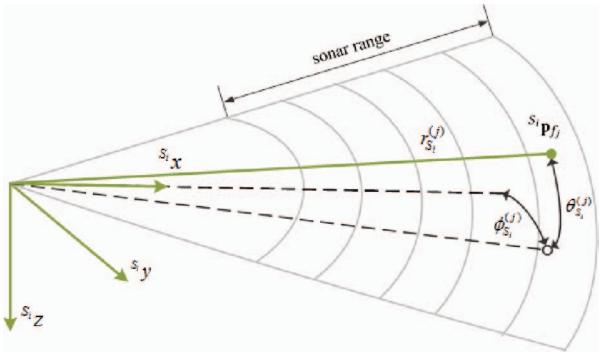

Fig. 1. Illustration of the sonar measurement model: The feature f_j in the sonar frame $\{S_i\}$, ${}^{S_i}\mathbf{p}_{f_j}$, can be represented in a spherical coordinate form: $(r^{(j)}_{S_i}, \phi^{(j)}_{S_i}, \theta^{(j)}_{S_i})$. Note that the range $r^{(j)}_{S_i}$ and the azimuth angle $\phi^{(j)}_{S_i}$ of feature f_j can be derived from this sonar measurement, while the elevation angle $\theta^{(j)}_{S_i}$ is lost in the 2D sonar image.

From the sonar measurement, we can get the range $r^{(j)}_{S_i}$ and the azimuth angle $\phi^{(j)}_{S_i}$ measurements. Thus, the measurement model can be described as:

$$\mathbf{z}^{(j)}_{S_i} = \begin{bmatrix} r^{(j)}_{S_i} \\ \phi^{(j)}_{S_i} \end{bmatrix} + \mathbf{n}^{(j)}_{S_i} = \begin{bmatrix} \sqrt{{}^{S_i}x_j^2 + {}^{S_i}y_j^2 + {}^{S_i}z_j^2} \\ \arctan\left(\frac{{}^{S_i}y_j}{{}^{S_i}x_j}\right) \end{bmatrix} + \mathbf{n}^{(j)}_{S_i} \quad (8)$$

where $\mathbf{n}^{(j)}_i$ is white Gaussian noise vectors with covariance matrix $\mathbf{R}^{(j)}_{S_i}$.

C. EKF with Stochastic Cloning

Determining the AUV's poses is often performed as SLAM problems and the EKF (or its variants) is frequently used for solutions (e.g., see [4]). To better address the nonlinear partially-observable sonar measurement [see (8)], we propose to employ stochastic cloning [18] in the EKF framework. Also, we advocate to perform online IMU-Sonar extrinsic calibration, because it is often difficult (if not possible) to pre-calibrate these sensors in an underwater workspace.

Specifically, the state vector at time-step k contains the current IMU state $\mathbf{x}_{\mathbf{IMU}_k}$, the extrinsic calibration (i.e., 6 DoF rigid transformation) between IMU and sonar $\mathbf{x}_{calib}$, the cloned N latest IMU poses $\mathbf{x}_{I_i}, i = 1\dots N$ and all the detected features $\mathbf{x}_f$:

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}^T_{\mathbf{IMU}_k} & \mathbf{x}^T_{calib} & \mathbf{x}^T_{I_1} & \dots & \mathbf{x}^T_{I_N} & \mathbf{x}^T_f \end{bmatrix}^T \quad (9)$$

where $\mathbf{x}_{calib} = \begin{bmatrix} {}^S_I\bar{q}^T & {}^I\mathbf{p}^T_S \end{bmatrix}^T$ is the extrinsic IMU-Sonar calibration; $\mathbf{x}_{I_i} = \begin{bmatrix} {}^{I_i}_G\bar{q}^T & {}^G\mathbf{p}^T_{I_i} \end{bmatrix}^T$ is the i-th cloned IMU pose, and $\mathbf{x}_f = \begin{bmatrix} {}^G\mathbf{p}^T_{f_1} & \cdots & {}^G\mathbf{p}^T_{f_{N_L}} \end{bmatrix}^T$ contains all N_L features detected thus far.

The standard EKF is then employed to propagate and update the state estimates and covariance [19]. In particular, as a new sonar image is acquired and processed, the current IMU pose estimate (${}^I_G\hat{\bar{q}}$ and ${}^G\hat{\mathbf{p}}_I$) and newly detected features corresponding to this sonar image will be appended

to the state vector and the covariance matrix is augmented accordingly [18]:

$$\mathbf{P}_{k|k} \leftarrow \begin{bmatrix} \mathbf{I}_{6(N+N_L)+21} \\ \mathbf{J} \end{bmatrix} \mathbf{P}_{k|k} \begin{bmatrix} \mathbf{I}_{6(N+N_L)+21} \\ \mathbf{J} \end{bmatrix}^T \quad (10)$$

where the Jacobian $\mathbf{J}$ is given by (see [18]):

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times9} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times6(N+N_L+1)} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times9} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times6(N+N_L+1)} \end{bmatrix}$$

Note that when cloning a new IMU pose, the oldest pose will be removed if the total number of cloned states exceeds the pre-defined threshold.

III. Acoustic-Inertial Odometry

It is clear from the preceding section that the system in the SLAM formulation may suffer from ever-increasing computational/storage complexity, as new features are being included in the state vector. In particular, this easily occurs when operating in large-scale environments. To address this issue, inspired by visual-inertial odometry [20], we introduce acoustic-inertial odometry for low-cost underwater navigation. In particular, we linearly marginalize out the acoustic features to keep the state vector a constant size, while still utilizing the information from sonar measurements of these features to update the state estimates.

Specifically, based on the sonar measurement model (8), the measurement residual for feature f_j is given by:

$$\mathbf{r}_{S_i}^{(j)} = \mathbf{z}_{S_i}^{(j)} - \hat{\mathbf{z}}_{S_i}^{(j)} \quad (11)$$

Linearizing the above equation around the current state estimates and feature estimates, the measurement residual can be computed as:

$$\mathbf{r}_{S_i}^{(j)} \simeq \mathbf{H}_{\mathbf{x}_{S_i}}^{(j)} \widetilde{\mathbf{x}}^* + \mathbf{H}_{f_{S_i}}^{(j)\,G} \widetilde{\mathbf{p}}_{f_j} + \mathbf{n}_{S_i}^{(j)} \quad (12)$$

where $\mathbf{H}_{\mathbf{x}_{S_i}}^{(j)}$ and $\mathbf{H}_{f_{S_i}}^{(j)}$ are the Jacobians corresponding to the state vector $\mathbf{x}^* = \begin{bmatrix} \mathbf{x}_{\mathbf{IMU}_k}^T & \mathbf{x}_{calib}^T & \mathbf{x}_{I_1}^T & \cdots & \mathbf{x}_{I_N}^T \end{bmatrix}^T$ and the sonar feature f_j respectively. By stacking all the measurement residuals corresponding to the same feature f_j within the set of M_j, we have:

$$\mathbf{r}^{(j)} \simeq \mathbf{H}_{\mathbf{x}}^{(j)} \widetilde{\mathbf{x}}^* + \mathbf{H}_{f}^{(j)\,G} \widetilde{\mathbf{p}}_{f_j} + \mathbf{n}^{(j)} \quad (13)$$

Note that since features are not in the state vector, we cannot perform EKF update based on this residual (13). To overcome this issue, similar to [20], we multiply the left nullspace $\mathbf{U}$ of the Jacobian matrix $\mathbf{H}_f^{(j)}$ to both sides (13), and arrive at:

$$\mathbf{r}_o^{(j)} = \mathbf{U}^T \mathbf{r}^{(j)} \simeq \mathbf{U}^T \mathbf{H}_{\mathbf{x}}^{(j)} \widetilde{\mathbf{x}}^* + \mathbf{U}^T \mathbf{n}^{(j)} \quad (14)$$

$$= \mathbf{H}_o^{(j)} \widetilde{\mathbf{x}}^* + \mathbf{n}_o^{(j)} \quad (15)$$

Note that here we essentially have *linearly* marginalized the features from the linearized measurement model. Thus, eliminating the need to keep features in the state vector.

With the measurement residual formulation (14) for the j-th single sonar feature, we can stack all the available feature measurements, and thus, the stacked residual vector for all features is written as:

$$\mathbf{r} = \mathbf{H}\widetilde{\mathbf{x}}^* + \mathbf{n} \quad (16)$$

With (16), the standard EKF can be used to update the state estimate and covariance [19].

IV. Determining Acoustic Feature Positions

In order to perform the linear marginalization for the acoustic features as explained in the previous section, the 3D position estimate of the acoustic feature f_j is needed [see (13)]. Inspired by [21], we present in detail our method of localizing acoustic features.

A. Linear Triangulation

We first formulate a linear triangulation to obtain the feature position estimates by transforming the nonlinear sonar measurements (8) into linear equations. These triangulation results will be used as the initial estimates for the iterative nonlinear least-squares solver (see Section IV-B).

Specifically, let us first consider the bearing constraint. With (7), it is not difficult to see that the bearing of feature f_j in $\{S_i\}$ can be written as:

$$\mathbf{b}_{S_i}^{(j)} = \begin{bmatrix} \cos\phi_{S_i}^{(j)} \cos\theta_{S_i}^{(j)} \\ \sin\phi_{S_i}^{(j)} \cos\theta_{S_i}^{(j)} \\ \sin\theta_{S_i}^{(j)} \end{bmatrix} \quad (17)$$

and its perpendicular vector can be computed as:

$$\mathbf{b}_{S_i}^{(j)\perp} = \begin{bmatrix} -\sin\phi_{S_i}^{(j)} & \cos\phi_{S_i}^{(j)} & 0 \end{bmatrix}^T \quad (18)$$

Assuming that $\{S_N\}$ is the first sonar frame of M_j and using (17) and (18), we have:

$$\begin{aligned} \mathbf{p}_{f_j} &= \mathbf{p}_{S_i} + \left[\mathbf{R}(^{S_i}_{S_N}\bar{q})\right]^T {}^{S_i}\mathbf{p}_{f_j} \Rightarrow \\ \mathbf{R}(^{S_i}_{S_N}\bar{q})\mathbf{p}_{f_j} &= \mathbf{b}_{S_i}^{(j)} r_{S_i}^{(j)} + \mathbf{R}(^{S_i}_{S_N}\bar{q})\mathbf{p}_{S_i} \Rightarrow \\ \left[\mathbf{b}_{S_i}^{(j)\perp}\right]^T \mathbf{R}(^{S_i}_{S_N}\bar{q})\mathbf{p}_{f_j} &= \left[\mathbf{b}_{S_i}^{(j)\perp}\right]^T \mathbf{R}(^{S_i}_{S_N}\bar{q})\mathbf{p}_{S_i} \end{aligned} \quad (19)$$

where we have denoted $\mathbf{p}_{f_j}$ and $\mathbf{p}_{S_i}$ as ${}^{S_N}\mathbf{p}_{f_j}$ and ${}^{S_N}\mathbf{p}_{S_i}$ for simplicity. We also employed the identities that ${}^{S_i}\mathbf{p}_{f_j} = \mathbf{b}_{S_i}^{(j)} r_{S_i}^{(j)}$ and $\left[\mathbf{b}_{S_i}^{(j)\perp}\right]^T \mathbf{b}_{S_i}^{(j)} = 0$. Now consider the range constraint. The geometry of range constraint is shown in Figure 2. Based on the law of cosine, we have:

$$(r_{S_N}^{(j)})^2 + \|\mathbf{p}_{S_i}\|^2 - (r_{S_i}^{(j)})^2 = 2 r_{S_N}^{(j)} \|\mathbf{p}_{S_i}\| \left[\frac{\mathbf{p}_{S_i}}{\|\mathbf{p}_{S_i}\|}\right]^T \mathbf{b}_{S_N}^{(j)}$$

$$\Rightarrow \; \mathbf{p}_{S_i}^T \mathbf{p}_{f_j} = \frac{1}{2}\left[(r_{S_N}^{(j)})^2 + \|\mathbf{p}_{S_i}\|^2 - (r_{S_i}^{(j)})^2\right] \quad (20)$$

Since the feature f_j has been observed and tracked in the set of M_j frames, which contain n sonar images, we

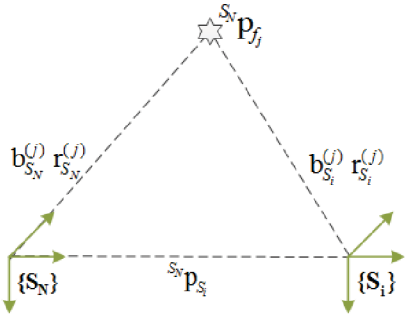

Fig. 2. Illustration of the geometry of the acoustic feature triangulation.

collect all the measurements from this set and formulate the following linear least-squares equations:

$$\underbrace{\begin{bmatrix} \left[\mathbf{b}_{S_1}^{(j)\perp}\right]^T \mathbf{R}({}^{S_1}_{S_N}\bar{q}) \\ \mathbf{p}_{S_1}^T \\ \vdots \\ \left[\mathbf{b}_{S_i}^{(j)\perp}\right]^T \mathbf{R}({}^{S_i}_{S_N}\bar{q}) \\ \mathbf{p}_{S_i}^T \\ \vdots \\ \left[\mathbf{b}_{S_n}^{(j)\perp}\right]^T \mathbf{R}({}^{S_n}_{S_N}\bar{q}) \\ \mathbf{p}_{S_n}^T \end{bmatrix}}_{\mathbf{B}_{2n\times 3}} {}^{S_N}\mathbf{p}_{f_j} = \underbrace{\begin{bmatrix} \left[\mathbf{b}_{S_1}^{(j)\perp}\right]^T \mathbf{R}({}^{S_1}_{S_N}\bar{q})\mathbf{p}_{S_1} \\ \frac{1}{2}\left[(r_{S_N}^{(j)})^2 + \|\mathbf{p}_{S_1}\|^2 - (r_{S_1}^{(j)})^2\right] \\ \vdots \\ \left[\mathbf{b}_{S_i}^{(j)\perp}\right]^T \mathbf{R}({}^{S_i}_{S_N}\bar{q})\mathbf{p}_{S_i} \\ \frac{1}{2}\left[(r_{S_N}^{(j)})^2 + \|\mathbf{p}_{S_i}\|^2 - (r_{S_i}^{(j)})^2\right] \\ \vdots \\ \left[\mathbf{b}_{n}^{(j)\perp}\right]^T \mathbf{R}({}^{S_n}_{S_N}\bar{q})\mathbf{p}_{S_n} \\ \frac{1}{2}\left[(r_{S_N}^{(j)})^2 + \|\mathbf{p}_{S_n}\|^2 - (r_{S_n}^{(j)})^2\right] \end{bmatrix}}_{\mathbf{b}_{2n\times 1}} \tag{21}$$

It is clear from (21) that each sonar measurement can provide 2 constraint equations. Therefore, if there are n $(n \geq 2)$ measurements, we are able to determine the feature's 3D position in the local sonar frame. Thus, the solution of (21) is given by the normal equation:

$${}^{S_N}\mathbf{p}_{f_j} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{b} \tag{22}$$

B. Nonlinear Least-Squares

Since the above linear triangulation does not take into account the measurement uncertainty, the result would not be optimal in the maximum likelihood sense. In order to find the maximum likelihood estimate of the feature, we formulate the equivalent (under mild assumptions) nonlinear least-squares optimization to refine the triangulation result:

$$\min_{{}^{S_N}\mathbf{p}_{f_j}} \sum_{i=1}^{n} \left|\left|\mathbf{z}_{S_i}^{(j)} - h({}^{S_N}x_j, {}^{S_N}y_j, {}^{S_N}z_j)\right|\right|^2_{\mathbf{R}_{S_i}^{(j)}} \tag{23}$$

The Gauss-Newton iterative algorithm can be employed to solve this problem by using the triangulation solution (22) as the initial guess.

C. Observability Analysis

A close inspection of matrix $\mathbf{B}$ in (22) reveals that it comprises three components: the bearing perpendicular vector $\mathbf{b}_{S_i}^{(j)\perp}$, the sensor rotation $\mathbf{R}({}^{S_i}_{S_N}\bar{q})$, and the sensor translation ${}^{S_N}\mathbf{p}_{S_i}$. Therefore, it would be important to examine how the sensor motion impacts the feature triangulation, and we need to find out what are the necessary conditions for feasible feature triangulation. Compared to [15] which has only qualitative analysis of 3 degenerate cases for ASFM, we will provide more thorough analysis with a complete mathematical proof for sonar feature triangulation.

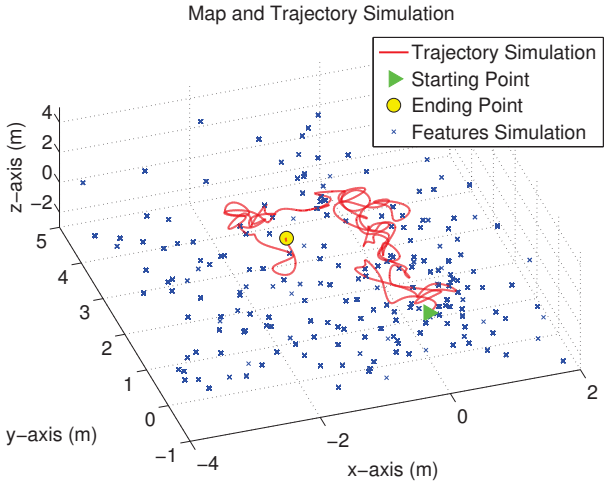

Fig. 3. Simulated AUV's trajectory and the feature map.

Lemma 1: The effects of sensor motion on feature triangulation are summarized in Table I.

- If the sonar rotates solely around z axis, moves along only x axis or y axis, or undergoes any combination of these three basic motion primitives, the feature cannot be triangulated.
- If the sonar motion contains one of the other three basic motion primitives (i.e., x rotation, y rotation and z translation), the feature can be triangulated with at least 2 or 3 measurements.

TABLE I
EFFECTS OF SONAR MOTION ON FEATURE TRIANGULATION

Sensor Motion	Conditions for Feature Triangulation
pure x rotation	$\phi_{S_i}^{(j)} \neq 0$ for at least 3 measurements
pure y rotation	$\phi_{S_i}^{(j)} \neq 0$ for at least 3 measurements
pure z rotation	Cannot triangulate
pure x translation	Cannot triangulate
pure y translation	Cannot triangulate
pure z translation	$\phi_{S_i}^{(j)} \neq 0$ for at least 2 measurements

Proof: See Appendix. ■

V. SIMULATION RESULTS

To validate our proposed AINS algorithm, we perform 50 Monte-Carlo (MC) simulations under various conditions. For the results presented in this section, we consider an AUV that randomly moves in the environments where point features are also randomly populated. In particular, the AUV is performing a general motion to avoid unobservable motions and all these point features are assumed to be static. Since features are randomly generated, some of them might be clustered together or spread out. Figure 3 shows the vehicle's

TABLE II
SIMULATION SETUP PARAMETERS

Parameter	Value
Sonar Range (m)	[0.1, 7]
Sonar Azimuth FOV (deg)	[-60,60]
Sonar Elevation FOV (deg)	[-10,10]
Sonar Angular Resolution (deg)	1
Sonar Range Resolution (m)	0.01
Calib Orientation Error (deg)	[3 -3 0]
Calib Orientation σ (deg)	4.58
Calib Position Error (m)	[0 0 0.01]
Calib position σ (m)	0.2
IMU rotation σ (rad/s)	1.1220×10^{-4}
IMU rotation bias σ (rad/s)	5.6323×10^{-5}
IMU acc. σ (m/s^2)	5.0119×10^{-4}
IMU acc. bias σ (m/s^2)	3.9811×10^{-5}
Monte-Carlo Trials	50

trajectory and the feature map. All pertinent parameters of the simulation setup are summarized in Table II. It should be pointed out that the sensor parameters used in this test are realistic. The performance metrics used are the root mean squared error (RMSE) and the normalized estimation error squared (NEES) [19]. The former quantifies the estimation accuracy while the latter is the standard criterion for estimation consistency.

In particular, Figure 4 (a) and (d) show the average RMSE of MC simulations for the vehicle's orientation and position. Note that the total distance travelled is about 40 meters, while the average position RMSE is about 1 meter. This indicates that the navigation error of the proposed AINS is about 2.5% of the distance traveled. Figure 4 (c) and (f) depicts the average NEES for the vehicle's orientation and position. Figure 5 shows the estimation errors and the corresponding 3σ bounds that are obtained from one typical trial of the 50 MC simulations. As evident from these results, the proposed AINS achieves reasonably consistent performance.

Figure 4 (b) and (e) show the average RMSE of calibration parameters (rotation and translation). The RMSE for calibration rotation decreases quickly within the first 10 seconds and then stays at a low error value. Similarly, the average RMSE for calibration translation converges to a small value. This implies that the online calibration achieves a better accuracy than the initial estimate which is typically obtained by manual measures in practice. Moreover, Figure 6 shows estimate errors and their 3σ bounds of the online calibration for a typical trial. These results show that the online extrinsic calibration for IMU and sonar reaches steady state quickly provided good initial estimates. Thus in practice, we may stop performing online calibration after a short period of time once its estimate becomes matured in order to save resources.

VI. CONCLUSIONS

We have developed a low-cost acoustic-inertial navigation system (AINS) that efficiently fuses acoustic and inertial measurements within a tightly-coupled, stochastic cloning-based EKF framework. In particular, we linearly marginalize out the acoustic features from the state while still utilizing all the corresponding sonar measurements. As a result, the computational complexity of the proposed approach is independent from the scale of the environment. We have also introduced an acoustic feature linear triangulation to generate initial estimates for the nonlinear least-squares based feature localization. A rigorous, detailed observability analysis has been performed to better understand the impact of the sensor motion on the feature triangulation. Additionally, motivated by the practical pre-calibration challenges, the proposed AINS advocates online sonar-IMU extrinsic calibration.

APPENDIX
PROOF OF LEMMA 1

We first introduce the following notations that will be useful for the ensuing analysis:

$$\mathbf{R}({}^{S_i}_{S_N}\bar{q}) = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{24}$$

$${}^{S_N}\mathbf{p}_{S_i} = \begin{bmatrix} \Delta x_{S_i} & \Delta y_{S_i} & \Delta z_{S_i} \end{bmatrix}^T \tag{25}$$

$${}^{S_N}\mathbf{p}_{f_j} = \begin{bmatrix} x_j & y_j & z_j \end{bmatrix}^T \tag{26}$$

1) Pure rotation around x axis: If the sonar only rotates around the x axis of the local frame, then the rotation matrix ${}^{S_i}_{S_N}\mathbf{R}$ and the translation vector ${}^{S_N}\mathbf{p}_{S_i}$ become:

$$\mathbf{R}({}^{S_i}_{S_N}\bar{q}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r_{22} & r_{23} \\ 0 & r_{32} & r_{33} \end{bmatrix}, {}^{S_N}\mathbf{p}_{S_i} = \begin{bmatrix} \Delta x_{S_i} \\ \Delta y_{S_i} \\ \Delta z_{S_i} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This yields:

$$\begin{aligned} &\left[\mathbf{b}_{S_i}^{(j)\perp}\right]^T \mathbf{R}({}^{S_i}_{S_N}\bar{q}){}^{S_N}\mathbf{p}_{f_j} = \\ &\begin{bmatrix} -\sin\phi_{S_i}^{(j)} & r_{22}\cos\phi_{S_i}^{(j)} & r_{23}\cos\phi_{S_i}^{(j)} \end{bmatrix} {}^{S_N}\mathbf{p}_{f_j} \end{aligned} \tag{27}$$

Thus, from (21), we have:

$$\mathbf{B} = \begin{bmatrix} \vdots & \vdots & \vdots \\ -\sin\phi_{S_i}^{(j)} & r_{22}\cos\phi_{S_i}^{(j)} & r_{23}\cos\phi_{S_i}^{(j)} \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{28}$$

Clearly, in this case, at least 3 measurements of feature f_j with $\phi_{S_i}^{(j)} \neq 0$ are needed for the triangulation when the sonar is purely rotating around its x axis.

2) Pure rotation around y axis: Similarly, if the sonar only rotates around the y axis, then we have:

$$\mathbf{R}({}^{S_i}_{S_N}\bar{q}) = \begin{bmatrix} r_{11} & 0 & r_{13} \\ 0 & 1 & 0 \\ r_{31} & 0 & r_{33} \end{bmatrix}, {}^{S_N}\mathbf{p}_{S_i} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \vdots & \vdots & \vdots \\ -r_{11}\sin\phi_{S_i}^{(j)} & \cos\phi_{S_i}^{(j)} & -r_{13}\sin\phi_{S_i}^{(j)} \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{29}$$

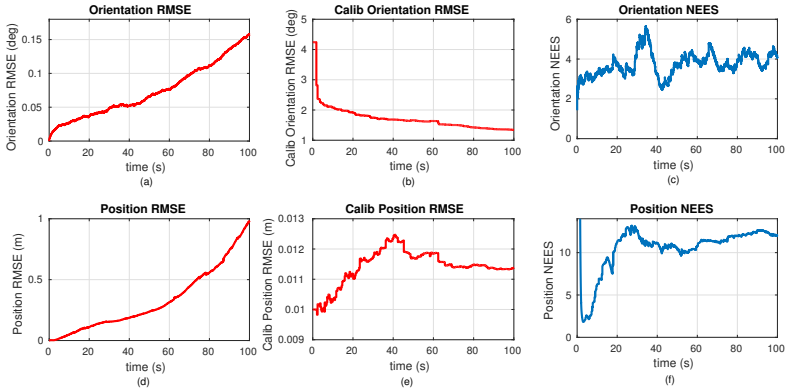

Fig. 4. Average RMSE and NEES of Monte-Carlo simulations for the AUV's position and orientation.

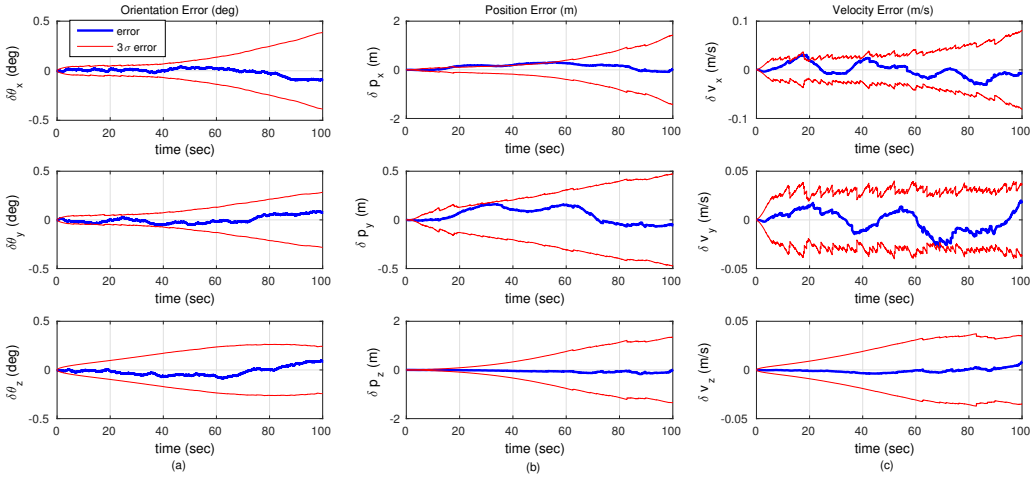

Fig. 5. Estimation errors vs. 3σ bounds. Note that these results are obtained for one typical realization of the 50 Monte-Carlo simulations.

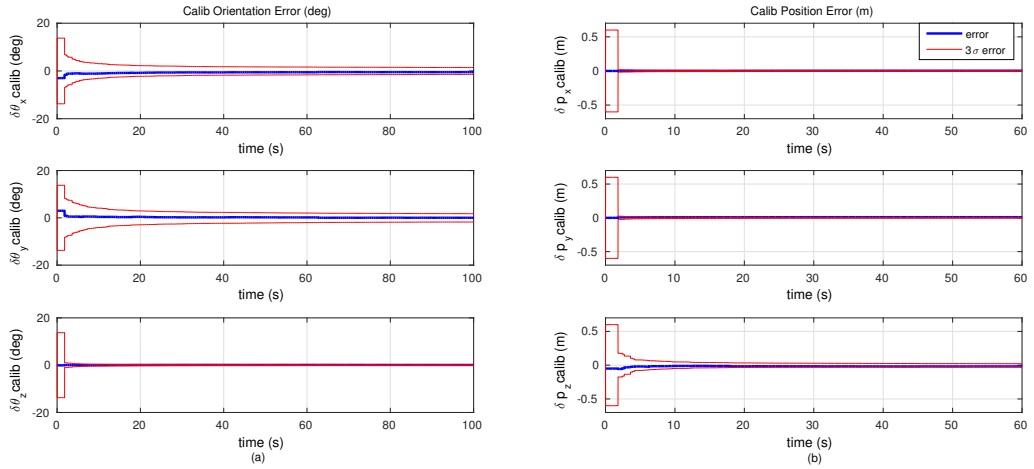

Fig. 6. Online calibration errors vs. 3σ bounds from one typical run of the 50 Monte-Carlo simulations.

In this case at least 3 measurements of feature f_j with $\phi_{S_i}^{(j)} \neq 0$ are needed for the triangulation.

3) Pure rotation around z axis: If the sonar only rotates around the z axis of the local frame, then we have:

$$\mathbf{R}({}^{S_i}_{S_N}\bar{q}) = \begin{bmatrix} r_{11} & r_{12} & 0 \\ r_{21} & r_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix}, {}^{S_N}\mathbf{p}_{S_i} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \vdots & \vdots & \vdots \\ -r_{11}\sin\phi_{S_i}^{(j)} + r_{21}\cos\phi_{S_i}^{(j)} & -r_{12}\sin\phi_{S_i}^{(j)} + r_{22}\cos\phi_{S_i}^{(j)} & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{30}$$

In this case, we notice that no matter how many measurements are acquired for feature f_j, $Rank(\mathbf{B}) \leq 2$. So $\mathbf{B}^T\mathbf{B}$ will always be singular and the feature cannot be triangulated.

4) Pure translation along x axis: If the sonar only has translations along the x axis of the local frame, then we have:

$$\mathbf{R}({}^{S_i}_{S_N}\bar{q}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, {}^{S_N}\mathbf{p}_{S_i} = \begin{bmatrix} \Delta x_{S_i} \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \vdots & \vdots & \vdots \\ -\sin\phi_{S_i}^{(j)} & \cos\phi_{S_i}^{(j)} & 0 \\ \Delta x_{S_i} & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{31}$$

In this case, $Rank(\mathbf{B}) \leq 2$. So the feature cannot be triangulated, no matter how many measurements are acquired.

5) Pure translation along y axis: If the sonar only has translations along the y axis of the local frame, then we have:

$$\mathbf{R}({}^{S_i}_{S_N}\bar{q}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, {}^{S_N}\mathbf{p}_{S_i} = \begin{bmatrix} 0 \\ \Delta y_{S_i} \\ 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \vdots & \vdots & \vdots \\ -\sin\phi_{S_i}^{(j)} & \cos\phi_{S_i}^{(j)} & 0 \\ 0 & \Delta y_{S_i} & 0 \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{32}$$

In this case, $Rank(\mathbf{B}) \leq 2$. So the feature cannot be triangulated, no matter how many measurements are acquired.

6) Pure translation along z axis: If the sonar only has translations along the z axis of the local frame, then we have:

$$\mathbf{R}({}^{S_i}_{S_N}\bar{q}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, {}^{S_N}\mathbf{p}_{S_i} = \begin{bmatrix} 0 \\ 0 \\ \Delta z_{S_i} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \vdots & \vdots & \vdots \\ -\sin\phi_{S_i}^{(j)} & \cos\phi_{S_i}^{(j)} & 0 \\ 0 & 0 & \Delta z_{S_i} \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{33}$$

In this case, at least 2 measurements are needed for feature f_j with $\phi_{S_i}^{(j)} \neq 0$ for the triangulation.

REFERENCES

[1] B. Kamgar-Parsi, L. J. Rosenblum, and E. O. Belcher, "Underwater imaging with a moving acoustic lens," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 91–99, 1998.

[2] R. L. Thompson, "Blazed array sonar systems - a new technology for creating low-cost high-resolution imaging sonar system for fisfisher management," *Proc. Puget Sound Georgia Basin Research Conference*, 2005.

[3] S. Negahdaripour, "On 3-d scene interpretation from fs sonar imagery," in *2012 Oceans*. IEEE, 2012, pp. 1–9.

[4] M. Walter, F. Hover, and J. Leonard, "Slam for ship hull inspection using exactly sparse extended information filters," in *IEEE International Conference on Robotics and Automation*. IEEE, 2008, pp. 1463–1470.

[5] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2010, pp. 4396–4403.

[6] F. S. Hover, R. M. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. J. Leonard, "Advanced perception, navigation and planning for autonomous in-water ship hull inspection," *The International Journal of Robotics Research*, vol. 31, no. 12, pp. 1445–1464, 2012.

[7] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Transactions on Robotics*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.

[8] N. Hurtós, Y. Petillot, J. Salvi, *et al.*, "Fourier-based registrations for two-dimensional forward-looking sonar image mosaicing," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 5298–5305.

[9] M. D. Aykin and S. Negahdaripour, "On feature matching and image registration for two-dimensional forward-scan sonar imaging," *Journal of Field Robotics*, vol. 30, no. 4, pp. 602–623, 2013.

[10] S. Negahdaripour, "On 3-d motion estimation from feature tracks in 2-d fs sonar video," *IEEE Transactions on Robotics*, vol. 29, no. 4, pp. 1016–1030, 2013.

[11] A. Mallios, P. Ridao, E. Hernández, D. Ribas, F. Maurelli, and Y. Petillot, "Pose-based slam with probabilistic scan matching algorithm using a mechanical scanned imaging sonar," in *OCEANS 2009-EUROPE*. IEEE, 2009, pp. 1–6.

[12] H. Assalih, "3d reconstruction and motion estimation using forward looking sonar," Ph.D. dissertation, Heriot-Watt University, 2013.

[13] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, "Opti-acoustic stereo imaging: On system calibration and 3-d target reconstruction," *IEEE Transactions on image processing*, vol. 18, no. 6, pp. 1203–1214, 2009.

[14] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment – a modern synthesis," *Lecture Notes in Computer Science*, vol. 1883, pp. 298–375, Jan. 2000.

[15] T. A. Huang and M. Kaess, "Towards acoustic structure from motion for imaging sonar," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 758–765.

[16] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep., Mar. 2005.

[17] A. B. Chatfield, *Fundamentals of High Accuracy Inertial Navigation*. Reston, VA: American Institute of Aeronautics and Astronautics, Inc., 1997.

[18] S. I. Roumeliotis and J. W. Burdick, "Stochastic cloning: A generalized framework for processing relative state measurements," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Washington, DC, May 11-15 2002, pp. 1788–1795.

[19] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York: John Willey and Sons, 2001.

[20] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy, Apr. 10–14, 2007, pp. 3565–3572.

[21] S. Negahdaripour, "On 3-d reconstruction from stereo fs sonar imaging," in *OCEANS 2010 MTS/IEEE SEATTLE*, Sept 2010, pp. 1–6.