A Differentiable Formulation for Uncertain Pose Estimation during Contact

Jeongmin Lee*, Minji Lee*, and Dongjun Lee

Abstract— **For many robotic manipulation and contact tasks, it is crucial to accurately estimate uncertain object poses, for which certain geometry and sensor information are fused in some optimal fashion. Previous results for this problem primarily adopt sampling-based or end-to-end learning methods, which yet often suffer from the issues of efficiency and generalizability. In this paper, we propose a novel framework for modeling and solving this uncertain pose estimation problem in differentiable form. To this end, we first devise a new type of geometric definition which is versatile and can provide differentiable contact features. In conjunction with this, we develop an efficient bi-level algorithm to solve the problem. Several scenarios are implemented to demonstrate how the proposed framework can improve existing methods.**

I. Introduction

In this paper, we present a novel differentiable framework which estimates the uncertain pose in contact tasks from sensor measurements. Our framework has a wide range of applications, from simple external impact localization to interactive manipulation such as peg-in-hole assembly. The main contribution of this paper is two-fold: 1) a new geometry representation based on a prescribed support function with differentiable contact features and their efficient computation algorithm; and 2) an efficient bi-level solution scheme based on differentiable optimization for uncertain pose estimation problem. The proposed methods are validated against both in simulation and experiment, demonstrating the efficacy of our differentiable framework for contact tasks.

Multiple studies have explored the identification of uncertainty in interaction, using a range of sensors such as vision, tactile, and force/torque (FT). These studies typically utilize learning-based frameworks to encode the relevant information. For example, [1] combines vision and FT sensor information using self-supervised learning. In [2], a certain action is performed to acquire FT measurements when contact occurs, and the plotted results are passed through neural network to estimate of the peg pose. For tactile sensor, the work [3] estimates the pose of grasped object using neural network and [4] perform tracking of extrinsic contact between object and environment based on neural contact fields. These methods are still lacking in their exploitation of the dynamic/kinematic structures of the problems, and data must be collected and learned again as the use cases expands.

Model-based methods that address geometry and sensor information together have been primarily relied on sampling strategies. For instance, contact particle filter (CPF) [5], [6] presents the way for external contact localization using proprioceptive sensors or force sensors. Object grasp pose estimation method is also conducted in [7] on the extension of CPF. Similarly, [8] presents the Bayesian framework for multi-modal fusion. These sampling-based methods may struggle with estimating poses in the case involving multiple contacts or steps. Recently, [9] and [10] develop the optimization based extrinsic contact sensing frameworks using various structured constraints. In comparison to above works, we aim for a differentiable formulation that can be applied to more general geometric types.

II. Problem Formulation

The main purpose of this paper is to develop the differentiable and general-purposed framework for uncertain pose estimation in interaction. We define the basic structure of the problem as follows:

Problem 1 (Uncertain Pose Estimation in Contact): Given the measurement $\gamma \in \mathbb{R}^{n_\gamma}$, estimate uncertain pose parameter $\xi \in \mathbb{R}^{n_\xi}$ through following optimization problem:

$$\begin{aligned}\min_{\xi, f\in\mathcal{C}} \ & \frac{1}{2}\left\| \gamma - \sum_{k=1}^{m(\xi)} P_k(\xi) f_k \right\|_{\Sigma^{-1}}^2 \\ \text{s.t.} \ & g_k(\xi) \geq 0, \quad (g_k(\xi))^+ f_k = 0 \quad \forall k\end{aligned} \tag{1}$$

where m is the number of collision, $g_k \in \mathbb{R}$, $f_k \in \mathbb{R}^3$, $P_k \in \mathbb{R}^{n_\gamma \times 3}$ are the gap, contact force, and contact mapping matrix (to the measurement) for the k-th contact. Note that P_k is a function of the contact witness points and normal. Also, $\|\cdot\|_{\Sigma^{-1}}^2$ is the Mahalanobis distance defined under the covariance matrix Σ, $(\cdot)^+ = \max(\cdot, 0)$, and $\mathcal{C}$ denotes the friction cone set:

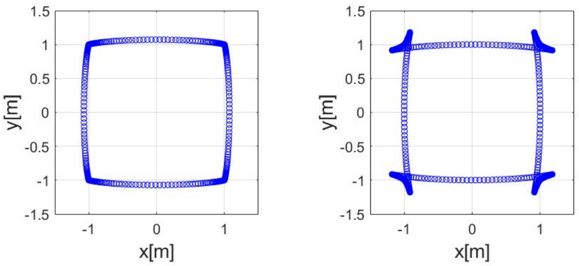
$$\begin{aligned}\mathcal{C} &= \mathcal{C}_1 \times \cdots \times \mathcal{C}_m \\ \mathcal{C}_k &= \{f_k \mid \mu_k f_{k,n} \geq \|f_{k,t}\|\}\end{aligned} \tag{2}$$

with μ, n, t being the friction coefficient[1], subscripts for the normal and tangential direction.

Here, the measurement γ is typically the FT or joint torque sensor value. It can also be a stack of measurements rather than a single measurement. Problem 1 can be interpreted as finding the most likely pose and contact force for the given sensor measurement. It has wide-ranging applications in robotics including grasp pose identification, object tracking, and external contact localization and is easily extensible. However, there are several challenges to solving a problem:

The authors are with the Department of Mechanical Engineering, IAMD and IOER, Seoul National University, Seoul, Republic of Korea. {ljmlgh,mingg8,djlee}@snu.ac.kr. *equal contribution

[1] In practice, it is difficult to accurately know the friction coefficient value, so the rough upper value is mainly used.



(a) Proposed

(b) Softmax

Fig. 1: Comparison of geometry obtained by the proposed support function and the naive softmax support function based on exponential. Vertex set is defined as $\{[1,1],[1,-1],[-1,1],[-1,-1]\}$.

1) the problem is non-linear with multiple complementarity constraints, and 2) the differentiability of m, g, P is ambiguous, making it difficult to find a proper gradient direction to optimize.

III. Differentiable Contact Features via Prescribed Support Function

A. *Prescribing Support Function*

For given set and direction, the support function [11] describes a distance to the supporting hyperplane. We model the geometry by *prescribing* the support function, based on the following theorem:

Theorem 1 ([11]): If $h:\mathbb{R}^3\rightarrow\mathbb{R}$ is a sublinear function i.e., a function that satisfies:

$$\begin{aligned}\text{Positive homogeneity:}\quad & h(\lambda x)=\lambda h(x)\ \forall\lambda\geq 0, x\in\mathbb{R}^3\\ \text{Subadditivity:}\quad & h(x+y)\leq h(x)+h(y)\ \forall x,y\in\mathbb{R}^3\end{aligned}$$

then there is a unique convex body with this support function. This theorem implies the one-to-one relationship between a sublinear function and correspoding convex body.

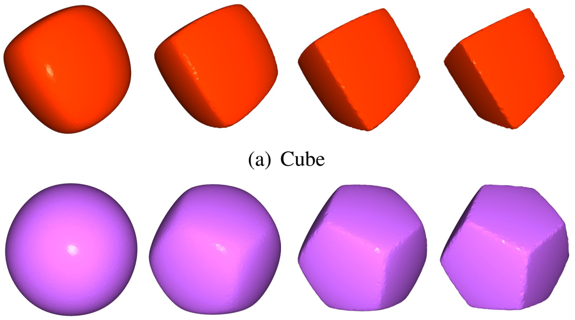
The question remained is then how to define the prescribed form of the support function. We first consider the set of vertices i.e., $v_1,\cdots,v_n\in\mathbb{R}^3$. This vertex set can be determined by the user or obtained from data such as mesh or point cloud. As it will be generalized under SE(3) transformation in Sec. III-B), here we assume that the origin is inside the convex hull of the vertices. Then we can easily find that the support function of the geometry defined as a convex hull is written as

$$h(x)=\max\left(v_1^Tx,\cdots,v_n^Tx\right) \tag{3}$$

which is discontinuous. Instead of using the max operator, we consider using a smoothed version of (3) which can be used for differentiable contact feature computation. The proposed function form is as follows:

$$h(x)=\left(\sum_{i=1}^{n}\left\{\max(v_i^Tx,0)\right\}^p\right)^{\frac{1}{p}} \tag{4}$$

where $p>2$. Equation (4) is similar to the p-norm function, but the abs$(\cdot)$ is replaced by $\max(\cdot,0)$, which naturally culls negative elements. Then Theorem 2 summarizes an important property of (4).



(a) Cube

(b) Dodecahedron

Fig. 2: Visualization of geometries represented by the prescribed support function (4). From left to right, $p=5,10,20,40$ are used.

Theorem 2: Given vertex set $v_1,\cdots,v_n$, the function (4) is sublinear and twice-differentiable on $\mathbb{R}^3\setminus\mathbf{0}$.

Proof: Positive homegeneity is trivial. Subadditivity can be shown as

$$\begin{aligned}h(x)+h(y) &= \left(\sum_{i=1}^{n}\left\{(v_i^Tx)^+\right\}^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}\left\{(v_i^Ty)^+\right\}^p\right)^{\frac{1}{p}}\\ &\geq\left(\sum_{i=1}^{n}\left\{(v_i^Tx)^++(v_i^Ty)^+\right\}^p\right)^{\frac{1}{p}}\\ &\geq\left(\sum_{i=1}^{n}\left\{(v_i^T(x+y))^+\right\}^p\right)^{\frac{1}{p}}\\ &= h(x+y)\end{aligned}$$

using the Minkowski inequality, where $\max(\cdot,0)$ is simplified as $(\cdot)^+$. Therefore, the function is sublinear. Twice-differentiablity can be easily verified by using the fact that

$$\sum_{i=1}^{n}\left\{(v_i^Tx)^+\right\}^p>0$$

for $x\in\mathbb{R}^3\setminus\mathbf{0}$ as the origin is inside the vertex set. ■

The properties in Theorem 2 is crucial, as it ensures that any (4) always corresponds to some convex geometry - note from Fig. 1 that other classes of support function are not necessarily able to do so. Fig. 2 depicts various smoothed geometries generated by the support function (4). We can find that smoothness of the geometry can be easily adjusted using p while retaining convexity and differentiability.

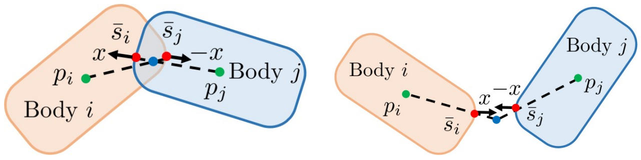
B. *Support Point and* SE(3) *Transformation*

Support point $s(x)$ can be derived as follows:

$$s(x)=s(x)+x^T\frac{ds}{dx}=\frac{dh}{dx} \tag{5}$$

since $x^T\frac{ds}{dx}=0$ holds from the homogeneity. By computing support points (5) for various x direction, we can visualize the corresponding shape of geometry.

The aforementioned support function and point can be generalized for SE(3) transformation. Given h and configuration vector $q\in\mathbb{R}^7$ (i.e., position and quaternion), the



(a) Active contact (b) Inactive contact

Fig. 3: Visualization of the condition in (7). Support points (red points) on both bodies extended by the growth factor should meet exactly (blue point).

support function $\bar{h}$ and support point $\bar{s}$ for q and x can be derived as follows:

$$\begin{aligned}\bar{h}(q,x) &= h(R(q)^T x) + p(q)^T x \\ \bar{s}(q,x) &= R(q)s(R(q)^T x) + p(q)\end{aligned} \tag{6}$$

where $p(q) \in \mathbb{R}^3$ and $R(q) \in \mathrm{SO}(3)$ are the translation and rotation by q.

C. *Contact Feature Computation*

We compute the contact features based on the growth distance (GD) model [12]. Combined with our geometry definition described above, we present an efficient unconstrained formulation.

1) Unconstrained nonlinear equation: Our unconstrained formulation employs the solution variables as x (i.e., normal vector) and growth factor $\sigma \in \mathbb{R}$, resulting in 4 dimensions. Then the conditions that the solution must satisfy are: 1) the two support points of each body corresponding to x coincide exactly when extended to σ; and 2) the normal vector x has unit norm. These can be written as:

$$F(x,\sigma,q) = \begin{bmatrix} \sigma(\bar{s}_i - \bar{s}_j) + (1-\sigma)(p_i - p_j) \\ \|x\|^2 - 1 \end{bmatrix} = 0 \tag{7}$$

for given bodies i and j where $\bar{s}_i = \bar{s}_i(x, q_i), \bar{s}_j = \bar{s}_j(-x, q_j)$ and $p = p(q)$. See Fig. 3 for visualization. The contact detection process is then reduced to solve (7) with respect to x, σ given the configuration q_i and q_j. Note that the formulation is of fixed dimension (i.e., 4) regardless of the number of vertices used in the geometries.

2) Newton solver: Theorem 2 ensures that h is twice-differentiable everywhere. Therefore we can always compute the Jacobian of F in (7) as follows:

$$\begin{aligned} J &= \left[\frac{\partial F}{\partial x}, \frac{\partial F}{\partial \sigma}\right] = \begin{bmatrix} \sigma\left(\frac{d\bar{s}_i}{dx} + \frac{d\bar{s}_j}{dx}\right) & y \\ 2x^T & 0 \end{bmatrix} \\ y &= R_i s_i(R_i^T x) - R_j s_j(-R_j^T x) \end{aligned} \tag{8}$$

and (8) can be applied to Newton-type algorithm to solve nonlinear equation in (7). Specifically, we utilize the trust-region-dogleg method [13] to achieve stable convergence property. Due to the simple structure of (4), $\frac{d\bar{s}}{dx}$ is also very easy to compute, much like s.

D. *Feature Differentiation*

After obtaining the contact features, the differential values can be computed and used to obtain the gradients for P and g. The conciseness of our GD model solver also makes the process of obtaining contact feature differentiation very efficient. Applying implicit differentiation to the nonlinear equation (7), we get

$$\frac{\partial F^*}{\partial q} + J^* \left[\frac{dx^*}{dq}; \frac{d\sigma^*}{dq}\right] = 0 \tag{9}$$

where the superscript $*$ denotes the value at the solution. As J^* is only a 4×4 matrix (and its factorization have already been computed in the solver step), we can obtain differentiation of contact normal and growth factor (and consequently, witness points) very efficiently.

IV. Bi-level Solver

A. *Predefined Number of Contact*

Despite this, we have differentiable formulation for $P(\xi)$ and $g(\xi)$, still the number of contacts $m(\xi)$ can change discretely. We address this issue by decomposing two interacting objects into m_1 and m_2 convex geometries, each of which is represented using the method described in Sec. III. Accordingly, we can predefine the collision number as constant, i.e., $m(\xi) = m = m_1 m_2$. Note that we can suppress contact forces for inactive contact by imposing the constraint $(g_k(\xi))^+ f_k = 0$.

B. *Bi-level Formulation*

1) Low-level problem: For the fixed ξ, problem (1) reduces to find the optimal contact force f^* as

$$\min_{f \in \mathcal{C}} \frac{1}{2} \|\gamma - P(\xi) f\|^2_{\Sigma^{-1}} \quad \text{s.t.}\ (g_k(\xi))^+ f_k = 0 \tag{10}$$

which is a second-order cone programming (SOCP). Since (10) is a convex problem, it is possible to obtain the global optimum efficiently. Additionally, the problem is independent for each touch step, which allows for parallel computation.

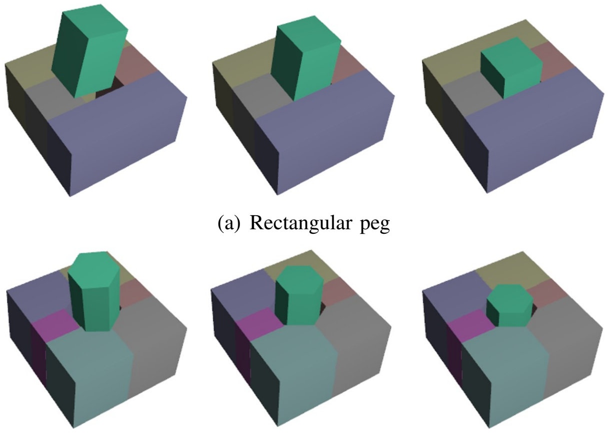
2) Differentiation: In terms of differentiable optimization [], the derivative of the solution to (10) with respect to the target parameter ξ can be obtained. For better smoothness, we put constraint $(g_k(\xi))^+ f_k = 0$ into a quadratic penalty term, and utilize the smoothed friction cone.

3) High-level problem: By substituting the obtained low-level solution f^* and handling the gap constraint $g_k(\xi) \geq 0$ as penalty functions, we can formulate the high-level problem as

$$\min_{\xi} \frac{1}{2} \|\gamma - P(\xi) f^*\|^2_{\Sigma^{-1}} + \frac{k_1}{2} \sum_{k=1}^{m} \left((-g_k(\xi))^+\right)^2 \tag{11}$$

where k_1 is the penalty coefficient to penalize penetration between objects. We can find that (11) is a non-linear least squares problem with differentiable error terms. Hence, we can use off-the-shelf algorithms such as the Gauss-Newton method to solve the problem, which also shows good convergence in practice. Note that the framework can be extended by augmenting it with additional cost terms on ξ.

Since the problem (11) is non-convex, there can be multiple local minimum. To enhance the ability of our gradient-based algorithm to discover global minimum, we adopt a strategy of sampling the initial pose parameters and selecting the optimal value from among them after optimization.



(a) Rectangular peg

(b) Hexagonal peg

Fig. 4: Snapshots of simulation results of peg-in-hole manipulation using our uncertain pose estimation framework in online. Different colors are used to represent convex-decomposed shapes.

V. Results and Evaluations

A. Peg-in-Hole

The proposed framework is tested on estimating the uncertain grasp pose (i.e., the pose of the peg with respect to the gripper) in peg-in-hole assembly task. Here, the uncertain parameter $\xi \in \mathbb{R}^3$ is the parameterized grasp pose and the measurement is from the force/torque sensor on gripper.

The experiment employs two distinct peg geometries: a rectangular prism and a hexagonal prism. The rectangular prism has eight vertices, while the hexagonal prism has twelve vertices. As shown in Fig. 4, the hole is decomposed into a total of 4 and 6 convex geometries, and the predefined numbers of collisions m are 4 and 6, respectively.

For the evaluation, we first collect simulation data (FT measurement, ground-truth grasp pose) in a contact situation using the original geometry. Here, the data accumulated over three contacts (i.e., $\gamma \in \mathbb{R}^{18}$) is used. The identification is then performed using the proposed differentiable contact feature, with three initial samples. For the baseline, we implement the particle filter (PF)-based method similar to [7]. The PF solves the high-level problem by using the grasp pose as particles with sampling strategy. For the low-level problem for each particle, we take the same methodology of our framework for better performance. Also, the number of particles is 25 (PF25) and 50 (PF50).

The comparison results are summarized in Table. I. A total of ten datasets and two different amount of noise (standard deviations 0.1 and 0.001) are used. The results clearly demonstrate that the proposed method outperforms the particle filter-based method in terms of accuracy and efficiency. This demonstrates how, our gradient-based method can quickly converge to the solution. Simulation snapshots of peg-in-hole assembly with online estimation are shown in Fig. 4.

B. Real World Experiment: Dish Placing

We deploy our framework in a dish placement task for experimental validation in the real world. The manipulator is built with Franka Emika Panda and a parallel gripper, and ATI Gamma is utilized as the FT sensor. Three different dishes are used, with a narrow-spaced dish rack. Test is conducted as follow: a human makes the gripper to grasp the dish in an arbitrary pose, and the robot identifies the uncertain grasp pose through interaction with the ground. The uncertain grasp parameter is modeled in 3-dimension and the dishes are represented by a smoothed convex hull of mesh with a prescribed support function. Our framework is successfully applied to enable successful performance of dish placement tasks - see Fig. 5 for experiment snapshots.

Noise		Low			High		
Methods		Ours	PF25	PF50	Ours	PF25	PF50
Rect	AT	9.22	44.1	89.1	10.7	43.7	89.3
	APE	4.54	1.91	2.08	3.34	1.99	2.22
	ARE	3.72	1.06	0.94	2.47	0.83	1.16
	AC	5.74	-0.912	-0.66	2.03	-0.70	-0.19
Hexa	AT	18.6	100	197	16.6	99.4	192
	APE	3.71	2.21	2.37	3.87	2.47	2.34
	ARE	2.58	0.87	1.10	2.50	1.06	1.07
	AC	3.28	-0.59	0.36	1.66	0.46	0.19

TABLE I: Evaluation results for the peg-in-hole assembly task. AT: average computation time (ms). APE/ARE/AC: position error (m), rotation error (rad) and cost value are converted using $(-\log(\cdot))$ and averaged, therefore bigger is better.

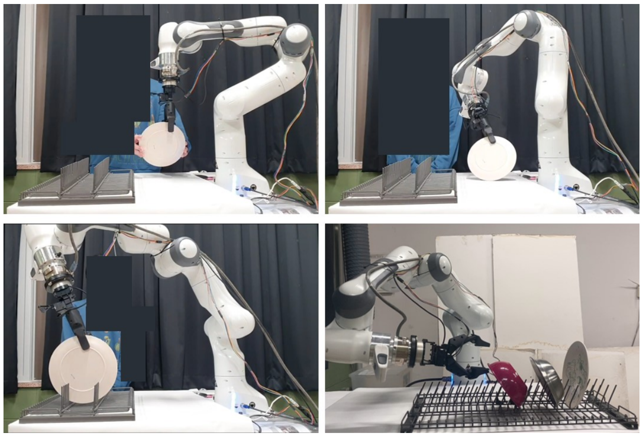


Fig. 5: Experimental demonstration of our framework in dish placing task. Top left: A human gives an arbitrary grasp pose. Top right: The robot estimates the uncertainty through interaction with the ground. Bottom left: Placing succeeded by proper estimation. Bottom right: Three dishes are successfully placed in a row.

VI. Conclusions

In this paper, we propose a differentiable uncertain pose estimation framework for interactive robot tasks. Prescribed support function based geometry definition is first presented to make it possible to express differentiable contact features. This is then combined with differentiable optimization to create an efficient bi-level algorithm for solving the pose estimation problem. Implementation shows how well our method can outperform sampling-based approaches. One of a promising direction for future work will be a combination with more diverse sensors.

References

[1] M. A Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg. Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks. In *IEEE International Conference on Robotics and Automation*, pages 8943–8950, 2019.

[2] S. Jin, X. Zhu, C. Wang, and M. Tomizuka. Contact pose identification for peg-in-hole assembly under uncertainties. In *American Control Conference*, pages 48–53, 2021.

[3] M. B. Villalonga, A. Rodriguez, B. Lim, E. Valls, and T. Sechopoulos. Tactile object pose estimation from the first touch with geometric contact rendering. In *Conference on Robot Learning*, pages 1015–1029, 2021.

[4] C. Higuera, S. Dong, B. Boots, and M. Mukadam. Neural contact fields: Tracking extrinsic contact with tactile sensing. *arXiv preprint arXiv:2210.09297*, 2022.

[5] L. Manuelli and R. Tedrake. Localizing external contact using proprioceptive sensors: The contact particle filter. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016.

[6] S. Wang, A. Bhatia, M. T Mason, and A. M Johnson. Contact localization using velocity constraints. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7351–7358, 2020.

[7] A. Sipos and N. Fazeli. Simultaneous contact location and object pose estimation using proprioception and tactile feedback. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022.

[8] F. von Drigalski, K. Hayashi, Y. Huang, R. Yonetani, M. Hamaya, K. Tanaka, and Y. Ijiri. Precise multi-modal in-hand pose estimation using low-precision sensors for robotic assembly. In *IEEE International Conference on Robotics and Automation*, pages 968–974, 2021.

[9] D. Ma, S. Dong, and A. Rodriguez. Extrinsic contact sensing with relative-motion tracking from distributed tactile measurements. In *IEEE international conference on robotics and automation*, pages 11262–11268, 2021.

[10] S. Kim and A. Rodriguez. Active extrinsic contact sensing: Application to general peg-in-hole insertion. In *IEEE International Conference on Robotics and Automation*, pages 10241–10247, 2022.

[11] R. Schneider. *Convex bodies: the Brunn–Minkowski theory*. Number 151. Cambridge university press, 2014.

[12] C. J. Ong and E. G Gilbert. Growth distances: New measures for object separation and penetration. *IEEE Transactions on Robotics and Automation*, 12(6):888–903, 1996.

[13] D. M Rosen, M. Kaess, and J. J Leonard. An incremental trust-region method for robust online sparse least-squares estimation. In *IEEE International Conference on Robotics and Automation*, pages 1262–1269, 2012.