

MIT Open Access Articles

Object Manipulation Through Contact Configuration Regulation: Multiple and Intermittent Contacts

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Taylor, Orion, Doshi, Neel and Rodriguez, Alberto. 2023. "Object Manipulation Through Contact Configuration Regulation: Multiple and Intermittent Contacts."

As Published: 10.1109/iros55552.2023.10341362

Publisher: IEEE|2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

Persistent URL: <https://hdl.handle.net/1721.1/155755>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-ShareAlike



Massachusetts Institute of Technology

Object manipulation through contact configuration regulation: multiple and intermittent contacts

Orion Taylor¹ Neel Doshi² Alberto Rodriguez

Massachusetts Institute of Technology

Abstract—In this work, we build on our method for manipulating unknown objects via *contact configuration regulation*: the estimation and control of the location, geometry, and mode of all contacts between the robot, object, and environment. We further develop our estimator and controller to enable manipulation through more complex contact interactions, including intermittent contact between the robot/object, and multiple contacts between the object/environment. In addition, we support a larger set of contact geometries at each interface. This is accomplished through a factor graph based estimation framework that reasons about the complementary kinematic and wrench constraints of contact to predict the current contact configuration. We are aided by the incorporation of a limited amount of visual feedback; which when combined with the available F/T sensing and robot proprioception, allows us to differentiate contact modes that were previously indistinguishable. We implement this revamped framework on our manipulation platform, and demonstrate that it allows the robot to perform a wider set of manipulation tasks. This includes, using a wall as a support to re-orient an object, or regulating the contact geometry between the object and the ground. Finally, we conduct ablation studies to understand the contributions from visual and tactile feedback in our manipulation framework. Our code can be found at: <https://github.com/mcubelab/pbal>.

I. INTRODUCTION

Robotic manipulation is driven by contact: robots manipulate objects by imparting wrenches and motions through the contact interfaces of the system. The geometric and frictional properties of these contacts determine which wrenches and motions are feasible. As such, regulating an object’s *contact configuration* – the location, mode and geometry of all contacts between the object, robot, and environment – is a fundamental aspect of manipulation. For example, when folding origami or tying one’s shoes, regulating sticking contact between the hand and the paper or shoe-lace ensures that the latter moves as intended.

In our previous work [1], we developed a joint estimation and control framework for manipulating 2D objects in the gravity plane via contact configuration regulation. The key estimation challenge was inferring the properties of the object/environment contact(s) from measurements taken at the object/robot contact(s), without prior knowledge of the object’s geometry. Previously, we restricted the set of possible

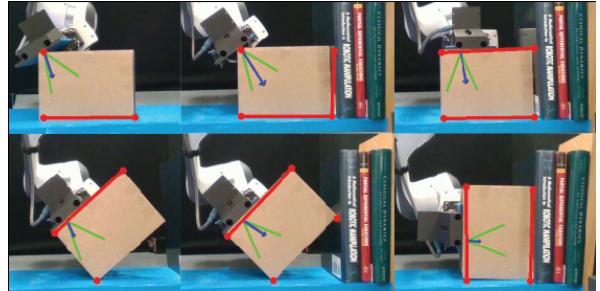


Fig. 1. Manipulating a box by first pushing it against a horizontal wall, and then using the wall as a support to reorient the box. Executing this involves several different combinations of contact geometries between the object and the environment/hand. In our framework, the measured robot wrench (blue) and the robot pose are used to estimate the friction constraints (green) and the contact locations and geometries (red).

contact geometries that were considered to prevent ambiguity. While effective, this assumption constrained the capabilities of the robot. This work pushes beyond these limitations by expanding our framework to include a more diverse set of contact interactions between the robot/object/environment.

Consider, for example, the common warehouse task of loading (or unloading) boxes from a storage shelf (Fig. 1). To accomplish this, it is natural to reason about the different possible combinations of contact formations between the box and the robot/environment. With the above in mind, the key contribution of our research is the development of a manipulation system that utilizes the mechanics of contact (Section III) to generate dexterous behaviors with unknown objects (Section VII) through a variety of contact formations. In particular, we:

- Present an improved estimation framework that uses factor graphs and limited visual feedback in order to localize, characterize, and differentiate a more diverse set of contact configurations (Section IV).
- Use these improved estimates to develop new manipulation primitives for our controller to regulate the system through these additional contact types, including manipulating an object against a corner (Section V).

We conclude that, for our purposes, the unique strength of vision is in revealing the global geometry/topology of the system, as opposed to the dense local information provided by F/T sensing and proprioception. Even coarse, low-frequency vision measurements can significantly magnify the inferential power of the available high-quality tactile sensing, allowing us to disambiguate contact configurations

¹ This work was supported by an ARA/Sponsored research award from Amazon.

² Neel Doshi (ndd@amazon.com) is with Amazon Robotics R&D. This research was conducted prior to Neel joining Amazon.

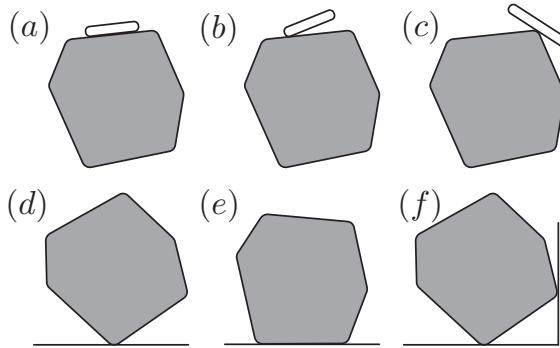


Fig. 2. Different contact geometries between the robot hand/object (top) and object/environment (bottom) supported by our system.

that would be indistinguishable when relying on feedback from a single source (Section VI). We demonstrate that this allows us to perform more complex manipulation tasks while still being agnostic to object geometry (Section VII).

II. RELATED WORK

Prior work on contact configuration regulation has often focused on tasks such as polishing [2] and deburring [3], as well as opening a variety of drawers [4] or doors [5]. The ability to directly observe all contacts facilitates estimation and control. Less attention has been given to manipulation tasks where not all contacts are directly observable (e.g., during non-prehensile manipulation). In this area, prior research usually either focuses on control assuming a known model of the world [6], [7], [8], or estimation assuming stable interactions [9], [10]. While there is more recent work on joint estimation and control, individual papers often make simplifying assumptions (e.g., frictionless models of contact) [11], [12], or learn task-specific policies from data (e.g., for cable manipulation [13], part insertion [14], or manipulating rigid objects on a shelf [15]). In contrast, our contribution is an object-agnostic joint estimation and control framework that reasons about all frictional interactions between the robot, object, and environment, leveraging limited visual feedback to execute more complex manipulations than in our previous work [1].

Estimation The primary focus of this work is the fusion of vision and tactile information for improved estimation of the location, geometry, and mode of all contacts between the robot, object, and environment. There is a large body of work in the area of localizing contacts/objects and estimating contact configurations using vision and/or tactile feedback. Many of these works present solutions to specific aspects of the problem we tackle here. Prior information about the end-effector geometry can be exploited to localize contacts [16], [17], [18], [19]. Similarly, knowledge of the object geometry can be used estimate its pose using only tactile sensing [20], or a fusion of both visual and tactile feedback [21], [22].

Particle filters are a popular formulation for nonlinear estimation problems, and have been often used in the context of tactile localization. Koval et al. [23] develop an approach to particle filters that takes into account how contact restricts

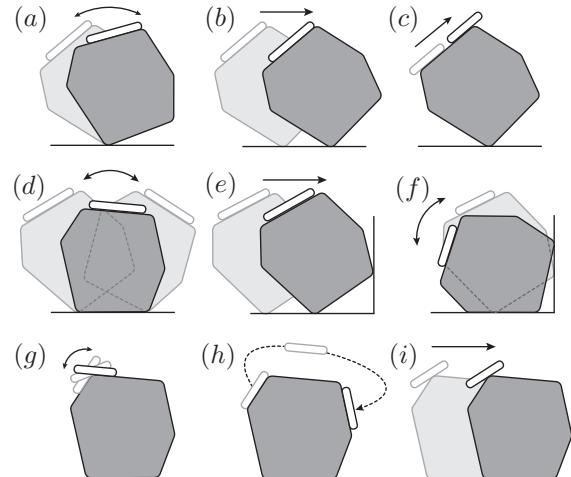


Fig. 3. Contact configuration transitions that can occur in our system.

an object's degrees-of-freedom. Li et al. [24] explicitly reason about the complementarity constraints of frictional contact, while Meeussen et al. [25] apply particle filters to fuse visual and tactile data. These works partially overlap with the problem that we are trying to solve.

A core part of our estimation framework is also inferring the contact wrench constraints (i.e., the generalized friction cone [26]). Prior work in this field has focused on the problem of planar pushing [27], [28], [29]. Of particular relevance is the work of Zhou et al. [29], who estimate the set of wrenches that can be transmitted through planar frictional contact (i.e., the limit-surface [30]).

Finally, several approaches have been presented to the problem of joint shape and pose estimation for 2D objects [31], [32], [33]. Suresh et al. [32] use factor graphs [34], [35] to encode the kinematic constraints of contact, a method that we also employ. However, our work places a significant emphasis on estimating *extrinsic contact formations* [36] (i.e., between the object and environment) that are not directly measured.

III. SYSTEM OVERVIEW

We focus on quasi-static manipulation of objects on top of a horizontal work surface. The system consists of: the **robot hand**, which we represent as a line of length $2l_h$, the **object**, which we treat as planar convex polygon, and the **environment**, which features the ground (a fixed horizontal line), and up to two vertical walls.

In our prior work, we assumed that the object was always in flush contact with the hand (Fig. 2a), the object was always in single point contact with the ground (Fig. 2d), and that the object's respective contact face and point at these two interfaces were fixed. These restrictions greatly simplified the system model, while still allowing for a contact-rich set of behaviors, primarily: pivoting the object about its ground contact, (Fig. 3a), translating the object (Fig. 3b), and sliding the hand relative to the object (Fig. 3c).

Here we build on this framework, relaxing some of these assumptions to regulate more complex geometric interactions

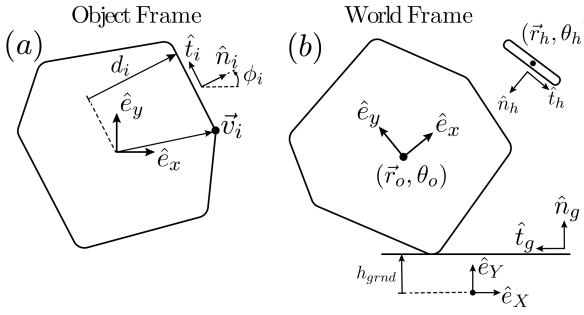


Fig. 4. System parameterization.

between the object and the hand/environment. We expand the set of admissible contact interactions to include:

- Object-line/hand-point contact (Fig. 2b), and object point/hand-line contact (Fig. 2c).
- Flush ground contact (Fig. 2e), and other external contacts, e.g., a wall (Fig. 2f).
- Transitions between contact geometries (Fig. 3d-i). This could be simple, like the object making and breaking contact with a wall (Fig. 3e); or more complex, like the hand transitioning from flush contact with one object face to the next by pivoting across a vertex (Fig. 3g).

A. System Parameterization

The system's state consists of the planar poses of the hand $\vec{x}_h = (\vec{r}_h, \theta_h)$ and object $\vec{x}_o = (\vec{r}_o, \theta_o)$ in the world-frame (Fig. 4b). The object, which is treated as a convex polygon, is parameterized by its (object-frame) vertex positions $\vec{v}_i = (x_i, y_i)$, outward facing surface normals $\hat{n}_i = (\cos \phi_i, \sin \phi_i)$, and contact face offsets $d_i = \hat{n}_i \cdot \vec{v}_i$ (Fig. 4a). We denote the (world-frame) contact normals and contact tangents of the hand and ground as (\hat{n}_h, \hat{t}_h) and (\hat{n}_g, \hat{t}_g) respectively. The constant h_{grnd} represents the height of the ground. During flush contact with the hand, we also consider the relative tangential displacement between the hand and the object, s .

When the object is in single point contact with the ground, we model the effect of gravity using the parameters (α_i, β_i) , which encapsulate the object's weight (mg), the length (l_i) of the gravitational moment arm (the vector from the ground contact point to the object's COM), and the angle $(\theta_o + \psi_i)$ of the gravitational moment arm:

$$mgl_i \sin(\theta_o + \psi_i) = \alpha_i \cos(\theta_o) + \beta_i \sin(\theta_o) \quad (1)$$

From the quasi-static motion assumption, it follows that:

$$\sum \vec{w}_{net} = \vec{w}_h + \vec{w}_e + \vec{w}_{grav} = 0, \quad (2)$$

where $\vec{w}_h = (\vec{F}_h, \tau_h)$, $\vec{w}_e = (\vec{F}_e, \tau_e)$, and $\vec{w}_{grav} = (-mg\hat{n}_g, 0)$ are the wrenches exerted on the object by the hand, environment, and gravity, respectively. Because the wrench exerted by gravity is constant, it follows that $\vec{w}_e = -\vec{w}_h + const$. This relation allows us to use the wrenches measured at the hand contact to infer the external contact behavior.

Measured vs. estimated quantities: We are able to directly measure the hand pose (\vec{r}_h, θ_h) and wrench \vec{w}_e via robot proprioception and a force-torque sensor in the wrist of the robot. We have added limited visual feedback, which

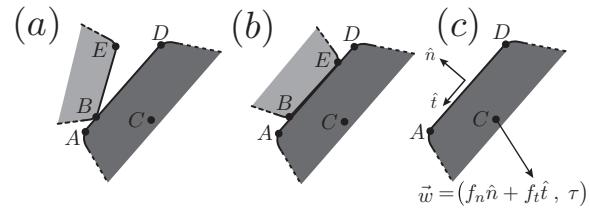


Fig. 5. a) Point/line contact. b) Flush contact. c) Normal/tangent vectors.

estimates the world-frame positions of the object vertices $\vec{v}_{i,world}$. Compared to the tactile feedback, vision is relatively noisy and operates at a lower frequency. We primarily use it to seed the estimator with an initial prior, though the estimator is also capable of making use of live visual feedback. Besides the hand pose and wrench, every other quantity, including the object vertices, is estimated.

B. The Constraints of Contact

Our framework is built on the idea that the constraints of contact can be exploited to regulate the system. Under the quasi-static assumption, these constraints describe the relationship between the contact kinematics and wrenches.

In our previous work, we reasoned about the friction cone, to detect and regulate sticking/sliding at the hand/environment contact interfaces; and the torque cone, to enforce flush contact between the object and end-effector. These constraints (the generalized friction cone [26]) are still sufficient to describe the expanded set of interactions, however we must express them in terms of the updated parameterization. We focus on the torque cone, to help reason about more diverse contact geometries.

We consider two contacting bodies, depicted in Fig. 5. Here, A and D are material points fixed to the first body frame, while B and E are material points fixed to the second body frame. The unit vectors (\hat{n}, \hat{t}) are the normal and tangent of the first body contact face. The wrench $\vec{w} = (f_n \hat{n} + f_t \hat{t}, \tau)$ is the net contact wrench that the second body exerts on the first body, measured with respect to reference point C , which is fixed in the first body frame. We consider point/line contact (Fig. 5a), flush contact (Fig. 5b), and no contact.

Making and breaking contact: The contact wrench can only be nonzero during contact. During periods of no contact, the contact wrench is zero. For the interaction in Fig. 5:

$$f_n > 0 \rightarrow (\vec{r}_B - \vec{r}_A) \cdot \hat{n} = 0 \text{ and/or } (\vec{r}_E - \vec{r}_A) \cdot \hat{n} = 0 \quad (3)$$

Sticking, sliding, and the friction cone: We use the standard linear complementarity constraints to reason about friction and sticking/slipping. Given friction coefficient μ , and assuming either of the contact geometries shown in Fig. 5:

$$v_{slide} = \frac{d}{dt} ((\vec{r}_b - \vec{r}_a) \cdot \hat{t}) = v_{slide}^+ - v_{slide}^- \quad (4)$$

$$0 \leq -\mu f_n - f_t \perp v_{slide}^+ \geq 0 \quad (5)$$

$$0 \leq -\mu f_n + f_t \perp v_{slide}^- \geq 0 \quad (6)$$

The \perp symbol is a shorthand, where $0 \leq u \perp w \geq 0$ implies $u \geq 0$, $w \geq 0$, and $uw = 0$. In our implementation, we do

not directly estimate the friction coefficient μ , but instead estimate the wrench constraints themselves (see [1]).

Contact geometry and torque constraints: We define the resultant contact torque, τ_P , about some reference point P :

$$\tau_P = \tau + (\vec{r}_C - \vec{r}_P) \times (f_n \hat{n} + f_t \hat{t}) \quad (7)$$

For the contact geometries shown in Fig. 5, we see that:

$$\tau_A \leq 0, \tau_B = 0, \tau_E \geq 0 \text{ (point/line contact at } B\text{)} \quad (8)$$

$$\tau_A \leq 0, \tau_B \leq 0, \tau_D \geq 0, \tau_E \geq 0 \text{ (flush contact)} \quad (9)$$

These constraints allow us to disambiguate contact geometries, estimate contact locations, and regulate the contact geometry. Here, it is convenient to consider the center of pressure (COP) of a line contact defined by points G and H e.g., the point on the contact patch for which the resultant contact torque is zero:

$$\vec{r}_Q = COP(\vec{r}_G, \vec{r}_H) \iff \tau_Q = 0, \vec{r}_Q = \gamma \vec{r}_G + (1 - \gamma) \vec{r}_H \quad (10)$$

For the point contact illustrated in Fig. 5a, B is the COP of AD . During estimation, we frequently solve (10) to compute the COP of a potential contact patch, given its endpoints and the measured contact wrench.

IV. ESTIMATION FRAMEWORK

The estimator consists of two separate modules: the friction estimator and the kinematic estimator (Fig. 6). The friction estimator, described in [1], builds a model for the friction cones at the hand and ground contact interfaces, which is used to detect and regulate sticking/sliding at both contacts. Our primary focus is the kinematic estimator (Fig. 6 bottom), which detects the location and geometry of each contact. This process has two stages. First, we use a set of heuristics to guess the contact geometry of each interface. Second, these estimates (and the stick/slip estimates from the friction module) are mapped to a set of kinematic/wrench constraints, which are then added as constraint factors in a factor graph. The graph is then solved to find an estimate of the system state and parameters. We warm start this loop using a prior provided by vision.

A. Contact Geometry Estimation

We rely on a sequence of heuristics to identify both the active vertices and contact geometry using tactile feedback and the previous state/parameter estimate. These heuristics are built from the kinematic/wrench constraints laid out in Section III-B and some geometric reasoning.

The object/hand interface: There are four possible contact geometries that we will consider at the object/hand contact interface: no contact, object-line/hand-point, object-point/hand-line and flush contact. We omit object-point/hand-point contact, because it is difficult to maintain and regulate.

Contact detection: The robot and object are assumed to be in contact if the magnitude of the measured force felt at the object/hand contact interface is above a set threshold (around 2–3N). Below this threshold, we assume no contact.

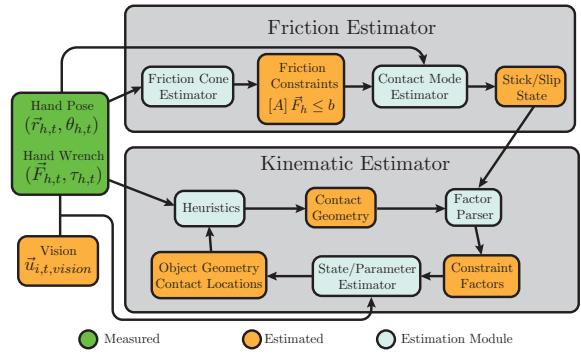


Fig. 6. The friction and kinematic estimators.

Object-line/hand-point: To test for this geometry (Fig. 2b), we use the measured wrench to compute the COP of the hand contact patch. If this COP is sufficiently close to one of the end-effector boundary points, then object-line/hand-point contact is detected.

Object-point/hand-line: For this case (Fig. 2c), we once again examine the COP of the hand. As per (8), during object-point/hand-line contact, the hand COP is coincident with the contact vertex. If the estimated positions of the hand COP and one of the object vertices are close, then object-point/hand-line contact is likely. In this case, we then test the kinematic feasibility of this geometry, taking into account the other active contact constraints. A similar feasibility test is performed for flush contact, and the results are compared to determine which geometry is more likely. We supplement this with a heuristic that leverages intuition of how the hand COP changes as a function of the end-effector pose. During flush contact, the COP often fluctuates while the hand remains stationary; whereas during object-point/hand-line contact, the COP motion is constrained by (8). We use this to compare the likelihood of these two geometries.

Flush contact: If contact has been detected, but the previous two geometries have been ruled out as possibilities, then we assume that the system is in flush contact (Fig. 2a). We then identify the current contact face by comparing the surface normal of the end-effector to the outward facing surface normals of the object in the most recent state estimate.

The object/environment contact(s): Here, we assume that the object is always in contact with the ground. Ground contact is limited to either object-point/ground-line or flush contact. We also check to see if wall contact is occurring.

Object-point/ground-line vs. flush contact: Given the most recent kinematic estimate and tactile feedback, we perform a sequence of tests. First, if one of the object vertices is sufficiently below the rest, then we assume single point contact with the ground at that vertex (Fig. 2d). Otherwise, if the two lowest object vertices have similar altitude e.g., the bottom edge of the object polygon is near horizontal, we compute the COP of that edge (assuming a weightless object). If the COP is in the interior of this edge by some margin, then we assume that the edge is in flush contact with the ground (Fig. 2e). Otherwise, the closest vertex to the COP is in single point contact with the ground.

Wall contact: To determine whether or not the object is

in contact with a wall (Fig. 2f), we compare the measured force at the hand contact with the friction constraints that we estimate for the ground contact [1]. These estimated constraints are computed during periods when there are no external contacts besides the ground; after which they are frozen in place (during periods when wall contact is allowed). Wall contact is detected when the measured contact force significantly violates one of the friction constraints for ground contact, since it implies the existence of a new contact that is generating the wrench causing the constraint violation.

B. Object and Contact Localization

We now synthesize the contact mode/geometry estimates with tactile feedback to estimate the contact locations.

Factor graphs and GTSAM: We use factor graphs to formulate the problem of localizing the object and contacts (Tactile SLAM [32]). Factor graphs are bipartite graphs which encode variables and constraints as vertices, and their input-output relationships as edges. For our purposes, factor graphs provide a straightforward way to leverage our knowledge of the contact constraints in order to localize the object and contacts (Fig. 7). Here, the contact geometry factors (depicted in blue) constrain the instantaneous contact pose and wrench, while the contact mode factors (purple) constrain how the contact pose evolves over time. Each factor is expressed as a function that maps the estimated variables to a constraint violation error. Each constraint has an associated variance that reflects our confidence in its accuracy. There are various algorithms available for computing the optimal values of the estimated variables. We rely on the factor graph software GTSAM [35], and its implementation of the incremental smoothing and mapping algorithm [34].

Variable factors and measured quantities: The following are variables in our factor graph:

- The object-frame vertex positions, $\vec{v}_i = (x_i, y_i)$.
- The parameters describing the object-frame surface normals ϕ_i and face offsets d_i .
- The object pose in the world-frame, $(\vec{r}_{o,t}, \theta_{o,t})$.
- During flush contact, the tangential displacement between the hand and the object, s_t .
- The height of the ground h_{grnd} .
- The gravitational torque parameters (α_i, β_i) .

We use F/T sensing and robot proprioception to measure:

- The hand contact wrench $\vec{w}_{h,t} = (\vec{F}_{h,t}, \tau_{h,t})$.
- The hand pose $(\vec{r}_{h,t}, \theta_{h,t})$.

We also assume that the ground is horizontal, meaning that \hat{n}_g and \hat{l}_g are respectively vertical and horizontal.

When describing the constraint factors in this paper, we let $\vec{u}_{i,t}$ denote the current world-frame positions of the object vertices. These are computed as a function of the aforementioned estimation variables, specifically the object pose and the object-frame vertex positions:

$$\vec{u}_{i,t} = R(\theta_{o,t})\vec{v}_i + \vec{r}_{o,t} \quad (11)$$

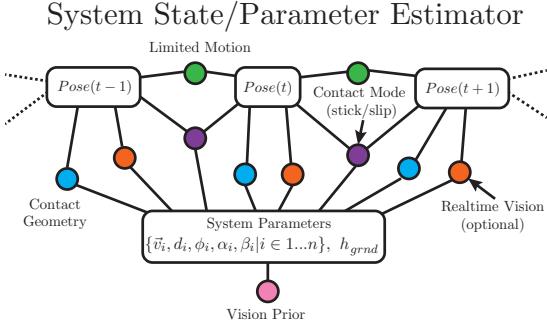


Fig. 7. Factor graph used for the kinematic estimator. Different sets of contact mode and geometry factors are used at each time-step, with the composition depending upon the stick/slippage and contact geometry estimates.

where $R(\theta)$ is the rotation matrix for angle θ . When the hand is in flush contact with object face c , we use:

$$\vec{u}_{i,t} = R(\theta_{h,t} - \phi_c + \pi)\vec{v}_i + d_c \hat{n}_{h,t} + s_t \hat{l}_{h,t} \quad (12)$$

The estimator may be provided with noisy live visual feedback of the world-frame vertex positions, $\vec{u}_{i,t,vision}$.

Constraint factors: We convert the estimated contact mode and geometry into a set of constraint factors that are added to the factor graph at each time-step.

Contact: The end-effector intersects any object point(s) it is in contact with. The same holds true for the ground:

$$(\vec{u}_{i,t} - \vec{r}_{h,t}) \cdot \hat{n}_{h,t} = 0 \mid i \in \{\text{hand contact vertices}\} \quad (13)$$

$$\vec{u}_{i,t} \cdot \hat{n}_g - h_{grnd} = 0 \mid i \in \{\text{ground contact vertices}\} \quad (14)$$

During flush contact with the hand, the hand contact constraint is implicit to (12), so we do not include a factor.

Sticking contact: If the object is in sticking contact with the hand, then the tangential motion (relative to the hand) of the contact point(s) is zero:

$$(\vec{u}_{i,t+1} - \vec{r}_{h,t+1}) \cdot \hat{l}_{h,t+1} - (\vec{u}_{i,t} - \vec{r}_{h,t}) \cdot \hat{l}_{h,t} = 0 \quad (15)$$

$$s_{t+1} - s_t = 0 \quad (\text{flush contact}) \quad (16)$$

Similarly, if the object is in sticking contact with the ground, the horizontal motion of the ground contact(s) is zero:

$$(\vec{u}_{i,t+1} - \vec{u}_{i,t}) \cdot \hat{l}_g = 0 \quad (17)$$

Torque Balance: When the object is in single point contact with ground, but not in wall contact, then by the quasi-static assumption, the net torque about the ground contact is zero.

$$(\vec{r}_{h,t} - \vec{u}_{i,t}) \times \vec{F}_{h,t} + \tau_{h,t} + \alpha_i \cos \theta_{o,t} + \beta_i \sin \theta_{o,t} = 0 \quad (18)$$

Here, the quantity $\alpha_i \cos \theta_{o,t} + \beta_i \sin \theta_{o,t}$ represents the gravitational torque about the ground contact vertex.

When the object is in single point contact with the hand, then the object contact vertex should be coincident with the COP of the hand contact patch, which we can compute by applying (10) to the hand pose and measured contact wrench:

$$\vec{u}_{i,t} - \vec{r}_{COP,h,t} = 0, \quad i = \text{hand contact vertex} \quad (19)$$

Vision: Any vision estimates are added as extra constraints:

$$\vec{u}_{i,t} - \vec{u}_{i,t,vision} = 0 \quad (20)$$

Geometric consistency: These constraints ensure that object surface normal ϕ_i and face offset d_i parameters are geometrically consistent with the object vertices \vec{v}_i :

$$d_i = \hat{n}_i \cdot \vec{v}_i \rightarrow x_i \cos \phi_i + y_i \sin \phi_i - d_i = 0 \quad (21)$$

$$d_i = \hat{n}_i \cdot \vec{v}_{i+1} \rightarrow x_{i+1} \cos \phi_i + y_{i+1} \sin \phi_i - d_i = 0 \quad (22)$$

Regularization constraints: We include low-weight constraint factors to limit the object motion between time-steps.

$$\vec{r}_{o,t+1} - \vec{r}_o = 0, \quad \theta_{o,t+1} - \theta_{o,t} = 0 \quad (23)$$

$$s_{t+1} - s_t = 0 \quad (\text{flush contact}) \quad (24)$$

Optimization weights: Each constraint factor has an associated variance that reflects our confidence in its accuracy. This is essentially an optimization weight, with lower variances increasing the significance of their corresponding factor in the optimization. We tuned these variances manually.

C. Estimator Initialization and the Role of Vision

The contact geometry heuristics (Section IV-A) and the factor graph estimator (Section IV-B) are mutually dependent: the contact geometry estimator needs a somewhat accurate state/parameter estimate to disambiguate contact geometries through tactile feedback, and the factor graph based estimator relies on the contact geometry estimator to determine which constraints are imposed at each time-step. This presents a chicken and egg problem: if the initial estimation error is too large, it will never correctly converge.

We resolve this by using vision to generate a prior of the object’s geometry and starting pose to warm start the kinematic estimator. Once it has been initialized, the kinematic estimator is capable of operating on tactile feedback alone. Live visual feedback is still useful when available, and acts to regularize the object pose. This safeguards against errors from accumulating when the heuristics incorrectly guess the contact geometries.

The vision heuristic we use directly estimates the positions of the object vertices in the world-frame, under the assumption that the object is a convex polygon. We start with a few seed pixels that belong to the object’s interior, and assume that the object is a somewhat uniform color that is distinct from its surroundings. With these seed pixels, we perform a breadth-first-search on a down-sampled version of the image to sparsely flood fill the portion of the object in the image, giving us a blob that approximately covers the object. Going back to the original high-resolution image, we then randomly sample pixels near the boundary of this blob to check if they should belong to the object’s interior. This gives us a more accurate picture of the object’s boundaries. Finally, we extract the object vertices using a heuristic to approximate the 2D convex hull for a noisy data-set (the same method we use to compute the friction constraints from the measured contact wrenches in [1]). This heuristic was sufficiently reliable for developing the kinematic estimator, but should be replaced with a standard perception method during real-world implementation.

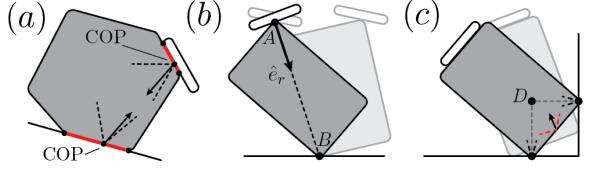


Fig. 8. The controller relies on the kinematic estimates to regulate the system: a) The controller enforces flush contact at an interface by keeping the COP on the interior of the **estimated** contact patch. b) During corner/corner pivoting, the admissible motions are pure rotations about the respective contact points (A, B), and the constraint direction corresponds to pure translation in the radial (\hat{e}_r) direction. c) During wall pivoting, the admissible motion direction corresponds to pure rotation about the ICR, D , which is a function of the external contact locations. In addition, the external friction cone (red) is expanded due to the wall contact

V. CONTROL FRAMEWORK

We further develop the control framework that we presented in [1]. In this framework, the contact configuration controller is built on top of a lower level impedance controller that regulates the end-effector pose. At each time-step, the controller solves a quadratic program (QP) to determine the incremental change of the impedance target, $\Delta\vec{x}_{tar}$:

$$\min_{\Delta\vec{w}_h, \Delta\vec{x}_{tar}} \alpha_0 \|\Delta\vec{x}_{tar}\|^2 + \sum \alpha_i (\Delta\vec{x}_{tar} \cdot \Delta\vec{q}_i - \beta_i \Delta\varepsilon_i)^2 \quad (25)$$

$$\text{s.t. } \hat{n}_j \cdot (\vec{w}_{h,\text{meas}} + \gamma_j \Delta\vec{w}_h) \leq b_j \quad \forall j \quad (26)$$

$$\Delta\vec{w}_h = K \Delta\vec{x}_{tar} \quad (27)$$

Here, the sum $\vec{w}_{h,\text{meas}} + \Delta\vec{w}_h$ is the predicted wrench that the hand will exert after the impedance target has been incremented by $\Delta\vec{x}_{tar}$. The constraints (26) correspond to the estimated value of the wrench space constraints, which are weighted by γ_j . The equality constraint (27) corresponds to the stiffness law of the low-level impedance controller. Finally, the cost (25) is used to regulate the pose along the admissible motion directions of the system, $\Delta\vec{q}_i$. The composition of cost terms and wrench constraints is specified by the motion primitive that the controller has been commanded to execute. The main change to this control framework is the adjustment/addition of individual constraints and cost terms, as well as the addition of a few new primitives; however, the underlying QP remains the unchanged.

Contact Geometry Regulation: In our previous work, the controller enforced flush contact with the hand via. the inclusion of two torque constraints in the QP:

$$-l_h(\vec{F}_{h,\text{meas}} + \Delta\vec{F}_h) \cdot \hat{n}_h - (\tau_{h,\text{meas}} + \Delta\tau_h) \leq -\varepsilon \quad (28)$$

$$l_h(\vec{F}_{h,\text{meas}} + \Delta\vec{F}_h) \cdot \hat{n}_h + (\tau_{h,\text{meas}} + \Delta\tau_h) \leq -\varepsilon \quad (29)$$

This ensured that the COP of the object/hand contact remained in the interior of the hand by some safety margin ε . These constraints rely on the assumption that the boundaries of the object contact face extends past the hand, and will fail in the cases of overhang (Fig. 8a). Since the improved estimator keeps track of the object vertices at the hand contact, we can adjust the torque constraints as follows:

$$-l_a(\vec{F}_{h,\text{meas}} + \Delta\vec{F}_h) \cdot \hat{n}_h - (\tau_{h,\text{meas}} + \Delta\tau_h) \leq -\varepsilon \quad (30)$$

$$l_b(\vec{F}_{h,\text{meas}} + \Delta\vec{F}_h) \cdot \hat{n}_h + (\tau_{h,\text{meas}} + \Delta\tau_h) \leq -\varepsilon \quad (31)$$

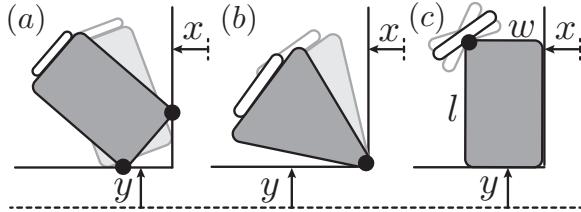


Fig. 9. Experimental design: we manipulate an object while enforcing wall contact, which provides ground truth for measuring the estimation error. a) Pivoting a rectangle against a corner, and computing the estimation error of the y-coordinate of the bottom vertex and the x-coordinate of the wall contact vertex. b) Pivoting a triangle against a corner, and computing the estimation error of the pivot vertex. c) Pivoting about the vertex of a rectangle that is in flush contact with a wall, and computing the estimation error of the pivot vertex.

The constants l_a and l_b take into account the possibility of overhang; and are a function of the estimated object vertex positions, and the hand's pose and length.

We have also constructed an analogous set of torque constraints that enforce flush contact with the environment. With these constraints, we have created variants of the three original motion primitives (Fig. 3 a-c). These new primitives can slide the object or hand while maintaining flush external contact, or rotate an object to force a transition from single point external contact to flush contact (Fig. 3d).

Corner/corner pivoting: Corner/corner pivoting is a new primitive during which the object is in sticking single point contact with both the ground and the hand (Fig. 8b). Here, the admissible motion directions $\Delta\vec{q}_\phi = (R(90^\circ)(\vec{r}_h - \vec{r}_A), 1)$ and $\Delta\vec{q}_\theta = (R(90^\circ)(\vec{r}_h - \vec{r}_B), 1)$ are rotations about the hand contact and external contact respectively; and the constraint direction $\hat{n}_r = (\hat{e}_r, 0)$ corresponds to pure translation along the line segment connecting the two contacts. We build a new QP using the cost terms and constraints corresponding to $(\Delta\vec{q}_\phi, \Delta\vec{q}_\theta, \hat{n}_r)$, which gives us a controller that regulates the system during this contact configuration.

Wall pivoting: Finally, we have constructed a primitive that regulates the orientation of an object, while maintaining sticking flush contact with the hand and sliding contact at both external contacts (Fig. 8c). This is a variant of our original rotation primitive (Fig. 3a), with two key changes. First, the admissible motion direction for rotation, $\Delta\vec{q}_\theta = (R(90^\circ)(\vec{r}_h - \vec{r}_D), 1)$, takes into account the adjusted instantaneous center of rotation (ICR) location due to wall contact. Second, the external friction constraints (26) take into account the fact that the wall contact expands the environment friction cone, since it can generate an arbitrary amount of force in the direction normal to the wall.

VI. EXPERIMENTS

We perform a set of experiments to validate the estimator. In our setup, we manipulate an object while enforcing contact with a wall (Fig. 9). Under this contact constraint, the known locations of the ground (y), wall (x), and the object dimensions (w, l) act as ground truth for specific coordinates subsets of the contact locations. The composition of those



Fig. 10. Demonstrations: To test our framework, we regulate the system through a sequence of contact configurations. The measured robot wrench (blue) and the robot pose are used to estimate the friction constraints (green) and the contact locations and geometries (red). **Top row:** We execute the corner/corner pivoting primitive to move a rectangle from line contact with the ground to wall contact. **Bottom rows:** We regulate a pentagon through sequence of contact geometry combinations.

TABLE I

| Trial | V only | K + V | K only |
|--------------------|-----------|-----------|-----------|
| (a) Wall Pivot | 3.28 (cm) | 1.68 (cm) | 1.66 (cm) |
| (b) Triangle Pivot | .74 (cm) | .62 (cm) | 1.00 (cm) |
| (c) Object Corner | .38 (cm) | .53 (cm) | .34 (cm) |

subsets is determined the object and the type of wall contact. We compare the root mean square error (RMSE) of these coordinate subsets when computed by our vision heuristic (V only), the kinematic estimator making use of live visual feedback (K+V), and the kinematic estimator using a vision prior, but without live visual feedback (K only). The results (Table I) indicate that F/T sensing and robot proprioception are often sufficient for estimating the geometry and locations of the external contacts, which is especially important in situations where the system does not have constant access to visual feedback (object or wall occlusions).

VII. DEMONSTRATIONS

To demonstrate our framework, we use it to regulate various shapes through a predetermined sequence of contact geometries (Fig. 10). For these tests, the kinematic estimator only relies on tactile feedback (after the initial vision prior). We manually jog the hand to disengage from a face/vertex of the object and then reengage at a different face/vertex.

VIII. CONCLUSION

In conclusion, we show that, with a reasonably accurate prior, it is possible to use tactile feedback to infer complex contact configurations between the object and the hand/environment. This can then be used to regulate the system as it transitions through these configurations.

One of the biggest challenges of this work was developing the heuristic that differentiates object-point/hand-line contact from flush hand contact, using tactile feedback alone. Though it often works, a momentary failure can be disastrous. Since the kinematically feasible object poses for the two geometries can differ significantly, a bad guess can cause the kinematic estimate to permanently diverge from the correct

solution. Though we plan to improve this heuristic, this issue demonstrates how live visual feedback synergizes with tactile estimation: even noisy low-frequency visual estimates can be enough to regularize the object pose, allowing the estimator to recover when a faulty contact geometry guess was made.

In the future, we plan on designing an additional estimator that predicts the feasibility of various motions and discrete transitions of the system using the kinematic and friction estimates. This would then be combined with a high-level planner that would sequence primitives, both to excite the system for the purpose of estimation, and to drive an object to a target state.

REFERENCES

- [1] N. Doshi, O. Taylor, and A. Rodriguez, “Manipulation of unknown objects via contact configuration regulation,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 2693–2699.
- [2] T. Furukawa, D. Rye, M. Dissanayake, and A. Barratt, “Automated polishing of an unknown three-dimensional surface,” *Robotics and Computer-Integrated Manufacturing*, vol. 12, no. 3, pp. 261–270, 1996.
- [3] M. G. Her and H. Kazerooni, “Automated robotic deburring of parts using compliance control,” *Journal of Dynamic Systems, Measurement, and Control*, vol. 113, no. 1, pp. 60–66, 03 1991.
- [4] Y. Karayannidis, C. Smith, F. E. Vina, P. Ögren, and D. Kragic, “Model-free robot manipulation of doors and drawers by means of fixed-grasps,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 4485–4492.
- [5] G. Niemeyer and J.-J. Slotine, “A simple strategy for opening an unknown door,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2, 1997, pp. 1448–1453.
- [6] F. R. Hogan and A. Rodriguez, “Reactive planar non-prehensile manipulation with hybrid model predictive control,” *The International Journal of Robotics Research*, vol. 39, no. 7, pp. 755–773, 2020.
- [7] Y. Hou, Z. Jia, and M. Mason, “Manipulation with Shared Grasping,” in *Proceedings of Robotics: Science and Systems*, Corvallis, Oregon, USA, July 2020.
- [8] F. R. Hogan, J. Ballester, S. Dong, and A. Rodriguez, “Tactile dexterity: Manipulation primitives with tactile feedback,” in *Proceedings of International Conference on Robotics and Automation (ICRA)*, 2020.
- [9] D. Ma, S. Dong, and A. Rodriguez, “Extrinsic contact sensing with relative-motion tracking from distributed tactile measurements,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2021.
- [10] Y. Shirai, D. K. Jha, A. U. Raghunathan, and D. Hong, “Tactile tool manipulation,” in *arXiv preprint*, 2023.
- [11] T. Lefebvre, H. Bruyninckx, and J. D. Schutter, “Polyhedral contact formation modeling and identification for autonomous compliant motion,” *IEEE Transactions on Robotics and Automation*, vol. 19, no. 1, pp. 26–41, 2003.
- [12] J. De Schutter, T. De Laet, J. Rutgeerts, W. Decré, R. Smits, E. Aertbeliën, K. Claes, and H. Bruyninckx, “Constraint-based task specification and estimation for sensor-based robot systems in the presence of geometric uncertainty,” *The International Journal of Robotics Research*, vol. 26, no. 5, pp. 433–455, 2007.
- [13] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, “Cable manipulation with a tactile-reactive gripper,” *The International Journal of Robotics Research*, 2021.
- [14] S. Dong, D. K. Jha, D. Romeres, S. Kim, D. Nikovski, and A. Rodriguez, “Tactile-rl for insertion: Generalization to objects of unknown geometry,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2021.
- [15] J. Liang, X. Cheng, and O. Kroemer, “Learning preconditions of hybrid force-velocity controllers for contact-rich manipulation,” in *arXiv preprint*, 2022.
- [16] A. Bicchi, J. K. Salisbury, and D. L. Brock, “Contact sensing from force measurements,” *The International Journal of Robotics Research*, vol. 12, no. 3, pp. 249–262, 1993.
- [17] L. Manuelli and R. Tedrake, “Localizing external contact using proprioceptive sensors: The contact particle filter,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.
- [18] K.-T. Yu and A. Rodriguez, “Realtime state estimation with tactile and visual sensing for inserting a suction-held object,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.
- [19] S. Wang, A. Bhatia, M. T. Mason, and A. M. Johnson, “Contact localization using velocity constraints,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.
- [20] A. Petrovskaya and O. Khatib, “Global localization of objects via touch,” *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 569–585, 2011.
- [21] J. Bimbo, L. D. Seneviratne, K. Althoefer, and H. Liu, “Combining touch and vision for the estimation of an object’s pose during manipulation,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 4021–4026.
- [22] P. Hebert, N. Hudson, J. Ma, and J. Burdick, “Fusion of stereo vision, force-torque, and joint sensors for estimation of in-hand object location,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011, pp. 5935–5941.
- [23] M. C. Koval, N. S. Pollard, and S. S. Srinivasa, “Pose estimation for planar contact manipulation with manifold particle filters,” *The International Journal of Robotics Research*, vol. 34, no. 7, pp. 922–945, 2015.
- [24] S. Li, S. Lyu, and J. Trinkle, “State estimation for dynamic systems with intermittent contact,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 3709–3715.
- [25] W. Meeussen, J. Rutgeerts, K. Gadeyne, H. Bruyninckx, and J. D. Schutter, “Particle filters for hybrid event sensor fusion with 3d vision and force,” in *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*. IEEE, 2006, pp. 518–523.
- [26] M. Erdmann, “On a representation of friction in configuration space,” *The International Journal of Robotics Research*, vol. 13, no. 3, pp. 240–271, 1994.
- [27] T. Yoshikawa and M. Kurisu, “Identification of the center of friction from pushing an object by a mobile robot,” in *Proceedings of IEEE/RSJ International Workshop on Intelligent Robots and Systems (IROS)*, 1991.
- [28] K. M. Lynch, “Estimating the friction parameters of pushed objects,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1993.
- [29] J. Zhou, M. T. Mason, R. Paolini, and D. Bagnell, “A convex polynomial model for planar sliding mechanics: theory, application, and experimental validation,” *The International Journal of Robotics Research*, vol. 37, no. 2-3, pp. 249–265, 2018.
- [30] S. Goyal, A. Ruina, and J. Papadopoulos, “Planar sliding with dry friction part 1. limit surface and moment function,” *Wear*, vol. 143, no. 2, pp. 307–330, 1991.
- [31] C. Strub, F. Wörgötter, H. Ritter, and Y. Sandamirskaya, “Using haptics to extract object shape from rotational manipulations,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2014, pp. 2179–2186.
- [32] S. Suresh, M. Bauza, K.-T. Yu, J. G. Mangelson, A. Rodriguez, and M. Kaess, “Tactile slam: Real-time inference of shape and pose from planar pushing,” in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 322–11 328.
- [33] K. Yu and A. Rodriguez, “Realtime state estimation with tactile and visual sensing. application to planar manipulation,” in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7778–7785.
- [34] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, “isam2: Incremental smoothing and mapping using the bayes tree,” *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012.
- [35] F. Dellaert and G. Contributors, “borglab/gtsam,” May 2022. [Online]. Available: <https://github.com/borglab/gtsam>
- [36] 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 (ICRA)*. IEEE, 2021, pp. 11 262–11 268.