



**HAL**  
open science

## Whole-Body Contact Force Sensing From Motion Capture

Tu-Hoa Pham, Adrien Bufort, Stéphane Caron, Abderrahmane Kheddar

► **To cite this version:**

Tu-Hoa Pham, Adrien Bufort, Stéphane Caron, Abderrahmane Kheddar. Whole-Body Contact Force Sensing From Motion Capture. SII: Symposium on System Integration, Dec 2016, Sapporo, Japan. pp.58-63, <10.1109/SII.2016.7843975>. <hal-01372531v2>

**HAL Id: hal-01372531**

**<https://hal.science/hal-01372531v2>**

Submitted on 19 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Whole-Body Contact Force Sensing From Motion Capture

Tu-Hoa Pham<sup>1,2</sup>, Adrien Bufort<sup>2</sup>, Stéphane Caron<sup>2</sup> and Abderrahmane Kheddar<sup>1,2</sup>

**Abstract**—Using motion capture only, we challenge in this paper the estimation of the contact forces –backed with ground-truth sensing, in human whole-body interaction with the environment. Our method is novel and makes it possible to get rid of cumbersome force sensors in monitoring multi-contact motion and its interaction force data. This problem is very difficult. Indeed, while a given force distribution uniquely determines the resulting kinematics, the converse is generally not true in multi-contact. In such scenarios, physics-based optimization alone may only capture force distributions that are physically compatible with a given motion rather than the actual forces being applied. We address this indeterminacy by collecting a large-scale dataset on whole-body motion and contact forces humans apply in multi-contact scenarios. We then train recurrent neural networks on real human force distribution patterns and complement them with a second-order cone program ensuring the physical validity of the predictions. Extensive validation on challenging dynamic and multi-contact scenarios shows that the method we propose can outperform physical force sensing both in terms of accuracy and usability.

## I. INTRODUCTION

Humans purposefully interact with their environment through physical contact to manipulate and move themselves or objects. The contact forces that are applied during a given task are informative on both the resulting motion and the underlying intent. Thus, force sensing has direct applications in research fields such as robot learning from demonstration and control [1], [2], physics-based animation [3], [4] and visual tracking [5], [6]. Contact forces are typically measured using force transducers that are costly, cumbersome and of limited, varying accuracy under repeated stress [7]. In this work, we propose a method to infer human whole-body contact forces from motion capture alone. If combined with markerless visual tracking technologies [8], this would enable the non-intrusive monitoring of contact forces in daily activities. However, the problem is very challenging.

By means of the equations of motion for articulated rigid bodies, the knowledge of external and internal forces uniquely determines the resulting kinematics. In contrast, the reverse problem is generally indeterminate in multi-contact with the environment and the knowledge of a given motion may not suffice to fully characterize the underlying force distribution. For instance, one can stand still while applying foot forces of varying magnitude in opposite, lateral directions.

This work was partially supported by the H2020 RIA COMANOID project [www.comanoid.eu](http://www.comanoid.eu) and the Japan Society for the Promotion of Science (JSPS): Kakenhi B No. 25280096.

<sup>1</sup> with the CNRS-AIST Joint Robotics Laboratory, UMI3218/RL, Tsukuba, Japan.

<sup>2</sup> with the Interactive Digital Human group of CNRS-UM LIRMM, UMR5506, Montpellier, France.



(a) Motion capture suit, contact force (red) and torque (yellow) visualization.



(b) Shoes equipped with inertial measurement units and force-torque sensors.

Fig. 1. Acquisition system for whole-body kinematics and contact forces.

The force distribution problem is an active research topic in multiple fields (Section II). We address it as follows:

- We construct a dataset on whole-body kinodynamics in the form of 100 min of motion and force measurements for diverse contact configurations (Section III).
- We propose a force estimation framework relying jointly on a recurrent neural network that learns how humans instinctively distribute contact forces in multi-contact and a second-order cone program that guarantees the physical plausibility of the resulting distributions (Section IV).
- We consistently validate our approach with ground-truth measurements throughout our work and demonstrate its accuracy on challenging scenarios (Section V).

Finally, we discuss the current limitations of our work as well as possible applications and extensions (Section VI). To accelerate the research on this new topic and encourage alternative implementations, we make our datasets and algorithms publicly available<sup>1</sup>.

## II. RELATED WORK

The identification of the objective function optimized by the central nervous system in daily activities (e.g., locomotion, manipulation) is a long-standing problem in kinesiology

<sup>1</sup>The dataset and algorithms will be released at <https://github.com/jrl-umi3218/WholeBodyKinodynamics>.

research [9]. A major difficulty lies in the high dimensionality of the human body and the limited observability of its parameters without invasive surgery [10]. Notably, [11] and [12] showed that inverse dynamics solutions during gait and other motions could vary significantly by using different existing models of body segment inertial parameters (BSIP) [13], [14]. Towards this issue, [15] introduced an optimization framework for the online estimation of robot and human BSIPs from force-torque measurements.

In the context of simulation, physically plausible forces are commonly computed using constrained optimization [16]. However, in multi-contact, such methods may fail to capture workless forces that humans apply in addition to the ‘minimal’ distributions supporting a given stance or task. Penalty or deformation-based approaches depend on the accurate physical identification of the contacting materials, and their relative pose can be difficult to accurately estimate in real experiments (e.g., from vision). Such methods were used to estimate manipulation forces on deformable objects [17] or by considering the human body elastic [18]. Our work was also inspired by [19], which computed ground reaction forces and joint torques from vision using a damped spring model.

The recent successes of deep learning for whole-body control [20] showed that data-driven approaches could successfully account for model or perception uncertainties. Artificial neural networks were used in [21] to capture recurring patterns in human manipulation from ground-truth measurements and infer realistic hand-object interaction forces from vision. Our current work expands this idea to whole-body interactions with the environment.

### III. WHOLE-BODY KINODYNAMICS DATASET

#### A. Experimental Setup

We collect kinodynamic measurements (motion and forces) on human activities using two types of sensors in parallel. First, the human whole-body motion is tracked using a motion capture system (Xsens MVN Awinda) consisting of 17 wireless inertial measurement units (IMU) and batteries strapped at specified body landmarks. The choice of this motion capture technology is motivated by our intention to collect human kinodynamic measurements in confined and eventually outdoor environments. Vision-based systems (e.g., Vicon) are limited by strong occlusions occurring in whole-body interactions with the environment, and difficult to apply in uncontrolled environments on the fly (e.g., outdoor).

The motion of the subject’s body, modeled as a 23-segment skeleton, is recorded at 100 Hz. For each sample, the system provides the 6-DoF pose of each segment as well as the corresponding linear and rotational velocity and acceleration. To facilitate the dynamics analysis, we transform the motion capture output into a kinematic tree comprised of 23 segments linked by 22 spherical joints and rooted at the subject’s pelvis. The BSIPs are computed from the subject’s weight and measurements using anthropomorphic tables [22]. Contact forces at the subject’s feet are monitored with instrumented shoes (Xsens ForceShoe), equipped with 6-DoF force-torque sensors at the heel and toes and IMUs recording

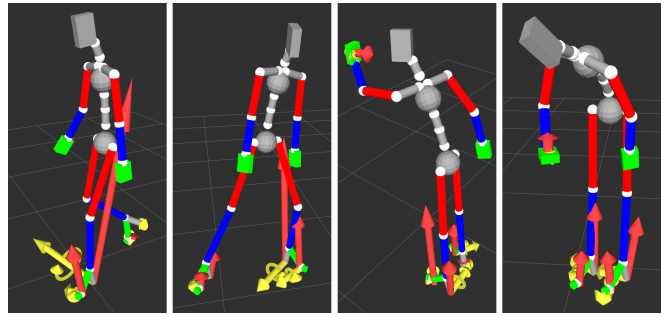


Fig. 2. Dataset examples (from left to right): walking, upper-body balancing with static contacts, taking support on a wall, leaning on a table.

the sensor orientations. We measure other interaction forces with the environment using an additional 6-DoF force-torque sensor (ATI Mini-45) held in the subject’s hand. All force-torque measurements are recorded at 100 Hz. We depict our acquisition setup in Fig. 1.

Being based on inertial measurements, the motion capture system is prone to drift compared to marker-based tracking methods (e.g., Vicon). We are working on a solution to attenuate this problem. Similarly, wearable force sensors can be of lower accuracy than force plates due to repeated pressure and deformations. Still, a major benefit of our lightweight setup is the efficient and continuous acquisition of kinematics and contact forces on highly-dynamic motions through time, which is generally not possible with static force plates. Additionally, the simultaneous monitoring of the whole-body motion and forces allows their correction as follows. First, the measurements are subject to noise, either from the sensors themselves or due to interferences in the wireless transmission. We attenuate it by smoothing all signals with a Gaussian filter of kernel  $\sigma = 0.05$  s. Second, a slow-varying bias can appear in the force-torque measurements with repeated stress and battery drain. We compute the bias through time by averaging the signals that persist when a sensor is not in contact with the environment, which should only be caused by the sensor moving parts.

#### B. The Dataset

In a preliminary study, four male volunteers took part as subjects in our experiments. Their weights (between 69.6 kg and 79.8 kg, plus the 3.5 kg acquisition system), heights (between 1.79 m and 1.94 m), and individual body segment lengths were measured to initialize the motion capture skeletal tracking model and BSIPs following [22]. All sensors (motion and force-torque) were calibrated and reset between experiments following the manufacturers’ recommended acquisition procedure to reduce the effects of measurement drift and hysteresis. The subjects were instructed to perform the following tasks:

- walking at different speeds (slow, normal, fast) and following different trajectories (circular, back and forth);
- moving the upper body while maintaining the feet static;
- taking support against a wall with the left or right hand;
- leaning on a table with the left or right hand.

We illustrate these experiments in Fig. 2. The goal of the first task is to allow neural networks to capture the centroidal dynamics relationship between motion and forces in bipedal contact. The second task follows the same principle and also provides examples of static indeterminacy, i.e., how humans apply forces that cancel each other out and do not affect their state of equilibrium. The third and fourth tasks go further and are typical scenarios where the straightforward minimization of multi-contact forces leads to distributions that are physically plausible but not representative of those humans really apply, as discussed in Section IV. Overall, we construct a dataset of total duration 100 min comprising synchronized motion and force-torque measurements on 51 experiments.

#### IV. FORCE SENSING FROM WHOLE-BODY MOTION

##### A. Whole-Body Force Optimization

For an articulated system of rigid bodies subject to  $n_\tau$  joint torques  $\boldsymbol{\tau}$  and  $n_c$  external wrenches  $\mathbf{F}_k = (\boldsymbol{\omega}_k, \mathbf{f}_k)$ , with  $\boldsymbol{\omega}_k$  and  $\mathbf{f}_k$  the respective external torque and force at contact  $k$ , the equations of motion can be expressed as:

$$\mathbf{H}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}) = \boldsymbol{\tau} + \sum_{k=1}^{n_c} \mathbf{J}_k^T \mathbf{F}_k, \quad (1)$$

with  $\mathbf{q}$  the vector of the position and orientation of the base link and joint angles,  $\mathbf{H}$  and  $\mathbf{C}$  the mass and bias matrices, respectively, and  $\mathbf{J}_k$  the Jacobian matrices at contact  $k$ .

With  $\mathbf{z}_k$  the normal at contact  $k$  and  $\mathbf{x}_k$  and  $\mathbf{y}_k$  two orthogonal vectors in the tangential plane, we express the wrench  $\mathbf{F}_k$  in the contact space  $\mathcal{C}_k = (\mathbf{x}_k, \mathbf{y}_k, \mathbf{z}_k)$ :

$${}^{\mathcal{C}_k} \mathbf{F}_k = (\omega_k^x, \omega_k^y, \omega_k^z, f_k^x, f_k^y, f_k^z)^T. \quad (2)$$

With the convention that  $\mathbf{z}$  is oriented from the environment to the contacting segment, the normal force is such that:

$$f_k^z \geq 0. \quad (3)$$

The Coulomb model with friction coefficient  $\mu_k$  implies that:

$$\|f_k^x \mathbf{x} + f_k^y \mathbf{y}\| \leq \mu_k f_k^z. \quad (4)$$

Eq. (1), (3) and (4) can be respectively integrated as equality, linear inequality and cone constraints of a second-order cone program (SOCP) of optimization parameters:

$$\begin{aligned} \mathbf{x} &= (\boldsymbol{\tau}, ({}^{\mathcal{C}_k} \mathbf{F}_k)_{k=1, n_c}) \\ &= (\tau_1, \dots, \tau_{n_\tau}, \omega_1^x, \omega_1^y, \omega_1^z, \dots, f_{n_c}^x, f_{n_c}^y, f_{n_c}^z)^T. \end{aligned} \quad (5)$$

From this formulation, physically plausible force distributions can be constructed by minimizing a cost function depending only on the optimization parameters, e.g., the (squared)  $L^2$  norm of the internal and external wrenches:

$$\mathcal{C}_{L^2}(\mathbf{x}) = \|\mathbf{x}\|^2 = \sum_{k=1}^{n_\tau} \tau_k^2 + \sum_{k=1}^{n_c} \|\mathbf{F}_k\|^2. \quad (6)$$

The resulting forces, by construction, are necessarily compatible with the observation motion. However, in multi-contact, when there exists more than a single distribution for a given task, there is no guarantee that the  $L^2$ -optimal

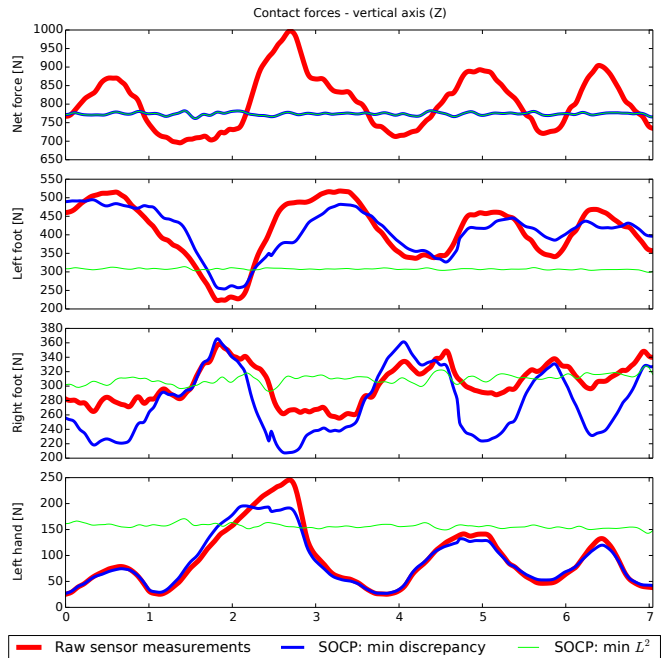


Fig. 3. In this sequence, the subject stays still while applying varying forces in triple contact with the environment. The equations of motion dictate that the net contact force should be constant (top row), which is not apparent on the force sensor measurements (red line) due to sensing uncertainties. Forces compatible with the observed kinematics can be computed using an SOCP (green and blue lines). The minimization of the  $L^2$  norm alone yields forces that are physically plausible but differ significantly from the measurements. Instead, minimizing the discrepancy to the uncertain measurements yields forces that are realistic both physically and compared to actual distributions.

distribution coincides with the actual forces being applied. As discussed in Section II, the identification of the actual cost function optimized by the human central nervous system is still an active research topic. Recall that due to measurement uncertainties, the preliminary correction process described in Section III-A does not fully ensure that the force-torque measurements are physically consistent with the motion. We do so by incorporating the force-torque sensor measurements  $\tilde{\mathbf{F}}_k$  into the SOCP cost function [23]:

$$\mathcal{C}_{\text{disc}}(\mathbf{x}) = \sum_{k=1}^{n_c} \left\| \mathbf{F}_k - \tilde{\mathbf{F}}_k \right\|^2. \quad (7)$$

Thus, we extract the force distribution that is the closest to the uncertain sensor measurements, while also being physically compatible with the observed motion, see Fig. 3. In the following, we denote by ground truth the SOCP-corrected sensor measurements (relative to the dynamic model).

##### B. Learning Features

Our goal is to construct a mapping  $\mathcal{F}$  between a set of input features  $\mathbf{K}$  representing the whole-body kinematics and contact configuration, and output features  $\mathbf{D}$  representing the underlying dynamics, i.e., external wrenches:

$$\mathbf{D} = \mathcal{F}(\mathbf{K}). \quad (8)$$

For the sake of generality, we aim at modelling human force distribution patterns based on an optimal selection of high-

level features rather than a large set of hand-engineered parameters. Alternatively to the whole-body equations of motion described in Eq. (1), we consider the corresponding centroidal dynamics [24]:

$$\dot{\mathbf{h}} = \sum_{k=1}^{n_c} \mathbf{G} \mathbf{F}_k, \quad (9)$$

with  $\mathbf{h}$  the  $6 \times 1$  centroidal momentum vector and  $\mathbf{G} \mathbf{F}_k$  the external wrenches applied to the center of mass  $\mathbf{G}$ . Denoting by  $\mathbf{P}_k$  the location of contact  $k$ , Eq. (9) can be rewritten as:

$$\dot{\mathbf{h}} = \sum_{k=1}^{n_c} \begin{bmatrix} \boldsymbol{\omega}_k + \overline{\mathbf{G} \mathbf{P}_k} \times \mathbf{f}_k \\ \mathbf{f}_k \end{bmatrix}. \quad (10)$$

This compact representation makes it a good candidate for the selection of optimal learning features extracting the gist of locomotory dynamics. Recall that the whole-body motion is expressed in the world frame. To account for translational and rotational invariances, we consider the quantities of Eq. (10) in a reference frame  $\mathcal{G}$  of origin  $\mathbf{G}$  and parallel to a chosen body frame (e.g., that of the pelvis). Thus, walking straight to the North is perceptually equivalent to walking straight to the East. The kinematics-based features are:

- The centroidal momentum time derivative  ${}^{\mathcal{G}}\dot{\mathbf{h}}$
- The position of each contact  ${}^{\mathcal{G}}\mathbf{P}_k$

In practice, the mapping  $\mathcal{F}$  of Eq. (8) may require a fixed-size input vector. Instead of only considering the location of the contacts  ${}^{\mathcal{G}}\mathbf{P}_{k,i}$  existing at timestamp  $i$ , we monitor a set of  $N_c$  possible contact locations through time and encode contact configurations using parameters  $\delta_{k,i}$  such that:

$$\delta_{k,i} = \begin{cases} 1 & \text{if contact } k \text{ is active at timestamp } i \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

In our experiments, we continuously monitored the position of the subject's feet and hands. Since our whole-body tracking system relies on inertial measurements, simple contact identification criteria based on contact staticity or distance to the environment could not be reliably applied. Therefore, we extracted the parameters  $\delta_k$  by applying thresholding to force sensor measurements. Note that this material limitation does not affect the generality of our approach and can be fully circumvented with additional visual observations. Overall, the complete input features at timestamp  $i$  are:

$$\mathbf{K}_i = \left( {}^{\mathcal{G}}\dot{\mathbf{h}}_i, \left( {}^{\mathcal{G}}\mathbf{P}_{k,i}, \delta_{k,i} \right)_{k=1, N_c} \right)^T \quad (12)$$

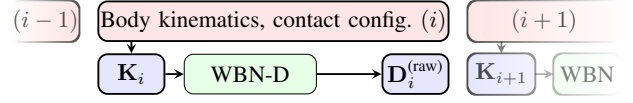
Similarly, the output features are the target wrenches in  $\mathcal{G}$ :

$$\mathbf{D}_i = \left( \left( {}^{\mathcal{G}}\mathbf{F}_{k,i} \right)_{k=1, N_c} \right)^T \quad (13)$$

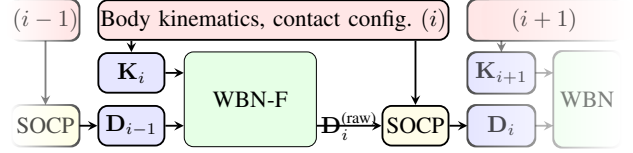
### C. Neural Network Model

In Eq. (8), the mapping  $\mathcal{F}$  does not account for temporal continuity. As such, consecutive force distributions are independent of each other. Instead, we introduce a dependency on both the current motion and the past trajectory using the following formulation:

$$\mathbf{D}_i = \mathcal{F} \left( \left( \mathbf{K}_j \right)_{j=1, i} \right). \quad (14)$$



(a) Forces are directly computed from the kinematics and contact configuration.



(b) Force predictions are corrected between consecutive time steps.

Fig. 4. Direct and feedback whole-body network architectures.

We model this time series structure using recurrent neural networks (RNN) [25] with long short-term memory neurons (LSTM) [26]. A simple network architecture, which we denote by WBN-D (whole-body network, direct), thus consists in a simple RNN directly mapping  $\mathbf{K}_i$  to  $\mathbf{D}_i$  while keeping track of long-term dependencies to the past:

$$\mathbf{D}_i = \text{WBN-D}(\mathbf{K}_i). \quad (15)$$

A typical iteration at timestamp  $i$  is as follows:

- 1) from the whole-body motion, compute the kinematics-based input features  $\mathbf{K}_i$
- 2) feed  $\mathbf{K}_i$  into WBN-D, get raw predictions  $\mathbf{D}_i^{(\text{raw})}$
- 3) extract predicted contact wrenches  $\tilde{\mathbf{F}}_{k,i}$

We illustrate the WBN-D architecture in Fig. 4a. Although the RNN is expected to implicitly capture the relationship between kinematics and forces, the raw predicted forces are not guaranteed to fully comply with the whole-body equations of motion and friction constraints. Therefore, we compute physically plausible solutions  $\mathbf{F}_k$  in the vicinity of the raw wrench predictions  $\tilde{\mathbf{F}}_k$  using the SOCP of section IV-A with the discrepancy cost function of Eq. (7). This step can be done offline, after the prediction of the complete raw wrench sequence from kinematics alone.

Alternatively, we propose an architecture that implements a feedback loop, WBN-F, allowing the correction of raw wrenches between consecutive predictions:

$$\mathbf{D}_i = \text{WBN-F}(\mathbf{K}_i, \mathbf{D}_{i-1}). \quad (16)$$

For prediction, the force distribution sequence is initialized with the distribution  $\mathbf{D}_0$  of minimal  $L^2$  norm, as described in Eq. (6). Such a distribution is computed from the kinematics alone. Subsequent iterations  $i$  are then as follows:

- 1) fetch the previous force features  $\mathbf{D}_{i-1}$
- 2) from the current whole-body motion, compute  $\mathbf{K}_i$
- 3) feed  $\mathbf{K}_i, \mathbf{D}_{i-1}$  into WBN-F, get raw predictions  $\mathbf{D}_i^{(\text{raw})}$
- 4) extract predicted contact wrenches  $\tilde{\mathbf{F}}_{k,i}$
- 5) find correct forces  $\mathbf{F}_{k,i}$  in the vicinity of  $\tilde{\mathbf{F}}_{k,i}$
- 6) extract corrected features  $\mathbf{D}_i$  for iteration  $i + 1$

We depict the WBN-F architecture in Fig. 4b.

TABLE I  
FORCE ESTIMATION ERRORS ON TESTING SET (16 min)

	Raw	Corrected
Force sensors	-4.58 (46.1)	ground truth
Min. $L^2$	N/A	2.19 (46.0)
WBN-D	0.75 (38.4)	0.89 (29.4)
WBN-F	1.26 (48.1)	0.77 (47.3)

## V. EXPERIMENTS

### A. Results on Complete Dataset

For the purpose of training, validation and testing, we construct a random partitioning of the whole-body kinodynamics dataset into three subsets of respective size 70 %, 15 % and 15 %. We implement the WBN-D and WBD-F neural network architectures within the Torch7 framework [27]. Both architectures take the kinematics features  $\mathbf{K}_i$  as input (as well as  $\mathbf{D}_{i-1}$  for WBD-F), pass them into two LSTM hidden layers of size 256, and compose the results with a linear layer returning the dynamics features  $\mathbf{D}_i$ . We train the networks using mini-batch stochastic gradient descent with a standard regression criterion (mean square error) and dropout to avoid overfitting [28]. The SOCP correction is implemented using the CVXOPT library for convex optimization [29].

In Table I, we summarize the average error and standard deviation (between parentheses) between ground truth and the following force data:

- raw force sensor measurements
- forces obtained from direct  $L^2$  norm minimization
- WBN-D outputs: raw and corrected offline
- WBN-F outputs: raw and corrected after each iteration

We observe the following:

- Force-torque sensors are rather unreliable without physics-based correction (Table I, first row, first column), in terms of bias (average error) and repeatability (standard deviation).
- On the other hand, forces computed with a direct  $L^2$  criterion also greatly differ from actual measurements (see Fig. 3). Thus, physics-based optimization, by itself, is not enough to capture the real forces being applied.
- Finally, the accuracy of all methods relying on learning and optimization is at least comparable (WBN-F) or significantly better (WBN-D) than that of the force-torque sensors.

Our main outcome is thus that, provided a rich dataset on human kinodynamics, the method we propose can outperform physical force sensing both in terms of accuracy and usability.

### B. Results on Restricted Training

During walking, most of the time is spent with only one foot on the ground. In single contact, the equations of centroidal dynamics, see Eq. (9), dictate that the contact wrench can be uniquely identified from the body kinematics. Therefore, it may not be necessary to extensively train neural networks on such examples. Instead, the prediction accuracy

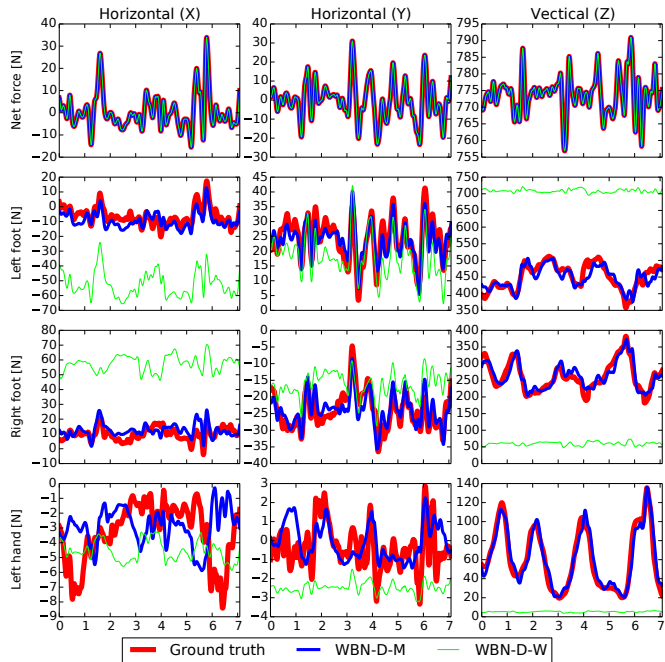


Fig. 5. Triple contact example. Trained on similar examples, WBN-D-M successfully estimates the actual forces being applied. In contrast, WBN-D-W predicts physically valid but significantly different force distributions.

may suffer if multi-contact examples (where the difficulty resides) represent a minority of the dataset. We assess this effect by training the previous neural network architectures not on the whole dataset, but on two sets containing only either walking or multi-contact examples. Both are again randomly partitioned into training (70 %), validation (15 %) and testing (15 %) subsets. We denote by WBN-D-W and WBN-F-W the respective direct and feedback architectures trained on walking examples only, and by WBN-D-M and WBN-F-M the networks trained on multi-contact examples.

We apply each network type on both its own testing set and that of the other type. We illustrate the application of WBN-D-W and WBN-D-M on a triple contact example (leaning on a table) in Fig. 5 and on walking in Fig. 6. In both cases, the raw predictions are corrected with the SOCP to ensure their physical compatibility with the observed motion. As it can be expected, the architecture trained only on walking fails at capturing the actual force distributions applied by humans in multi-contact (see Fig 5). In contrast, the architecture that was not extensively trained on walking accurately reconstructs contact forces even on such scenarios (see Fig. 6). This confirms that physics-based optimization is a valuable complement to recurrent neural networks for the latter to focus on multi-contact indeterminacy.

## VI. DISCUSSION AND FUTURE WORK

Our work introduces a novel method for the inference of contact forces applied by human subjects from their motion only. Our system estimates forces that are both physically plausible and in agreement with ground-truth measurements, even in challenging contact configurations where the force

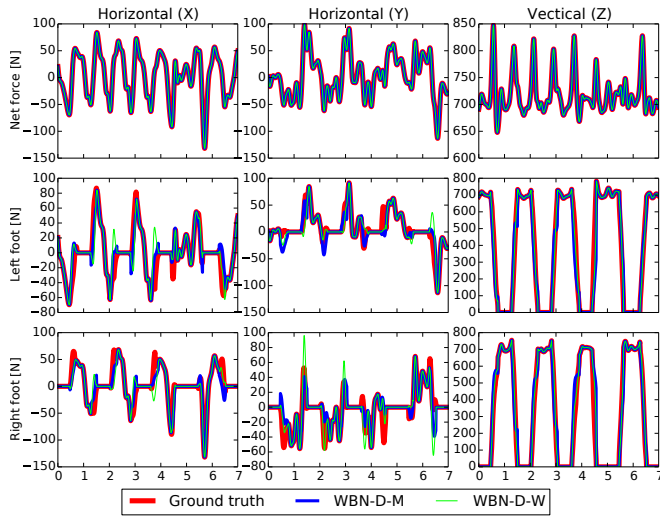


Fig. 6. Walking example. Despite not having been extensively trained on such examples, the performance of WBN-D-M used in conjunction with physics-based optimization is comparable to that of WBN-D-W.

distribution problem is highly indeterminate. Trained on our (public) dataset, the neural network architectures can be applied to any centroidal representation, while the SOCP can be formulated for any articulated body. As such, our approach can be seamlessly generalized to any whole-body tracking system. Applying our method to markerless visual tracking would thus enable fully non-intrusive monitoring of whole-body forces in daily activities. Still, our approach would certainly benefit from the further collection of ground-truth force measurements on even more rich motion and contact configurations. Another possibility could be to integrate convex optimization and learning into a unified computational framework [30]. In the long term, we also plan to apply our framework to force-based robot learning from demonstration, on-line multi-contact motion retargeting and knowledge-based multi-contact planning and control [31].

## REFERENCES

- [1] L. Rozo, P. Jiménez, and T. Carne, “A robot learning from demonstration framework to perform force-based manipulation tasks,” *Intelligent Service Robotics*, vol. 6, no. 1, pp. 33–51, 2013.
- [2] J. Engelsberger, P. Kozłowski, and C. Ott, “Biologically inspired dead-beat controller for bipedal running in 3d,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2015, pp. 989–996.
- [3] S. Ha, Y. Bai, and C. K. Liu, “Human motion reconstruction from force sensors,” in *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2011, pp. 129–138.
- [4] P. Zhang, K. Siu, J. Zhang, C. K. Liu, and J. Chai, “Leveraging depth cameras and wearable pressure sensors for full-body kinematics and dynamics capture,” *ACM Transactions on Graphics (SIGGRAPH Asia)*, vol. 33, 2014.
- [5] N. Kyriazis and A. Argyros, “Physically plausible 3d scene tracking: The single actor hypothesis,” in *CVPR*, 2013.
- [6] T.-H. Pham, A. Kheddar, A. Qammar, and A. A. Argyros, “Capturing and reproducing hand-object interactions through vision-based force sensing,” in *IEEE ICCV Workshop on Object Understanding for Interaction*, December 2015.
- [7] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, “Tactile sensing - from humans to humanoid,” *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 1–20, 2010.

- [8] D. Michel, K. Panagiotakis, and A. A. Argyros, “Tracking the articulated motion of the human body with two rgbd cameras,” *Machine Vision Applications*, vol. 26, no. 1, pp. 41–54, 2015.
- [9] V. M. Zatsiorsky, *Kinetics of human motion*. Human Kinetics, 2002.
- [10] B. I. Prilutsky and V. M. Zatsiorsky, “Optimization-based models of muscle coordination,” *Exercise and sport sciences reviews*, vol. 30, no. 1, p. 32, 2002.
- [11] G. Rao, D. Amarantini, E. Berton, and D. Favier, “Influence of body segments parameters estimation models on inverse dynamics solutions during gait,” *Journal of biomechanics*, vol. 39, no. 8, pp. 1531–1536, 2006.
- [12] A. Muller, C. Germain, C. Pontonnier, and G. Dumont, “A comparative study of 3 body segment inertial parameters scaling rules,” *Computer methods in biomechanics and biomedical engineering*, vol. 18, no. sup1, pp. 2010–2011, 2015.
- [13] R. Chandler, C. E. Clauser, J. T. McConville, H. Reynolds, and J. W. Young, “Investigation of inertial properties of the human body,” DTIC Document, Tech. Rep., 1975.
- [14] V. Zatsiorsky, V. Seluyanov, and L. Chugunova, “In vivo body segment inertial parameters determination using a gamma-scanner method,” *Biomechanics of human movement: Applications in rehabilitation, sports and ergonomics*, pp. 186–202, 1990.
- [15] J. Jovic, A. Escande, K. Ayusawa, E. Yoshida, A. Kheddar, and G. Venture, “Humanoid and human inertia parameter identification using hierarchical optimization,” *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 726–735, 2016.
- [16] S. Nakaoka, S. Hattori, F. Kanehiro, S. Kajita, and H. Hirukawa, “Constraint-based dynamics simulator for humanoid robots with shock absorbing mechanisms,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 3641–3647.
- [17] M. Mohammadi, T. L. Baldi, S. Scheggi, and D. Prattichizzo, “Finger-tip force estimation via inertial and magnetic sensors in deformable object manipulation,” in *IEEE Haptics Symposium*, 2016, pp. 284–289.
- [18] Y. Zhu, C. Jiang, Y. Zhao, D. Terzopoulos, and S.-C. Zhu, “Inferring forces and learning human utilities from videos,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3823–3833.
- [19] M. A. Brubaker, L. Sigal, and D. J. Fleet, “Estimating Contact Dynamics,” in *ICCV*, 2009.
- [20] I. Mordatch, K. Lowrey, G. Andrew, Z. Popovic, and E. V. Todorov, “Interactive control of diverse complex characters with neural networks,” in *Advances in Neural Information Processing Systems*, 2015, pp. 3132–3140.
- [21] T.-H. Pham, A. Kheddar, A. Qammar, and A. A. Argyros, “Towards force sensing from vision: Observing hand-object interactions to infer manipulation forces,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [22] R. Dumas, L. Cheze, and J.-P. Verriest, “Adjustments to mcconville et al. and young et al. body segment inertial parameters,” *Journal of biomechanics*, vol. 40, no. 3, pp. 543–553, 2007.
- [23] Y. Nakamura, K. Yamane, Y. Fujita, and I. Suzuki, “Somatosensory computation for man-machine interface from motion-capture data and musculoskeletal human model,” *IEEE Transactions on Robotics*, vol. 21, no. 1, pp. 58–66, 2005.
- [24] D. E. Orin, A. Goswami, and S.-H. Lee, “Centroidal dynamics of a humanoid robot,” *Autonomous Robots*, vol. 35, no. 2-3, pp. 161–176, 2013.
- [25] J. L. Elman, “Finding structure in time,” *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.
- [26] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [27] R. Collobert, K. Kavukcuoglu, and C. Farabet, “Torch7: A matlab-like environment for machine learning,” in *BigLearn, NIPS Workshop*, no. EPFL-CONF-192376, 2011.
- [28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [29] M. Andersen, J. Dahl, and L. Vandenberghe, “Cvxopt: A python package for convex optimization,” [abel.ee.ucla.edu/cvxopt](http://abel.ee.ucla.edu/cvxopt), 2013.
- [30] Z. Yan, J. Fan, and J. Wang, “A collective neurodynamic approach to constrained global optimization,” *IEEE Transactions on Neural Networks and Learning Systems*, 2016.
- [31] K. Bouyarmane and A. Kheddar, “Humanoid robot locomotion and manipulation step planning,” *Advanced Robotics*, vol. 26, no. 10, pp. 1099–1126, 2012.