

VDI

REIHE 10  
INFORMATIK/  
KOMMUNIKATION



# Fortschritt- Berichte VDI

Petrissa Zell, M. Sc.,  
Hannover

NR. 877

## Learning-Based Inverse Dynamics for Human Motion Analysis

BAND  
1|1

VOLUME  
1|1



Institut für Informationsverarbeitung  
[www.tnt.uni-hannover.de](http://www.tnt.uni-hannover.de)



LEARNING-BASED INVERSE DYNAMICS  
FOR HUMAN MOTION ANALYSIS

Von der Fakultät für Elektrotechnik und Informatik  
der Gottfried Wilhelm Leibniz Universität Hannover  
zur Erlangung des akademischen Grades

**Doktor-Ingenieur**

(abgekürzt: Dr.-Ing.)

genehmigte

**Dissertation**

von Frau

**Petrissa Zell, M. Sc.**

geboren am 20. März 1989 in Langenhagen

2022

Hauptreferent: Prof. Dr.-Ing. B. Rosenhahn  
Korreferent: Prof. Dr.-Ing. G. Pons-Moll  
Vorsitzender: Prof. Dr.-Ing. J. Ostermann  
Tag der Promotion: 15. Dezember 2021

VDI

**REIHE 10**  
INFORMATIK/  
KOMMUNIKATION

# Fortschritt- Berichte VDI



Zell, Petrisa, M. Sc.,  
Hannover

**NR. 877**

## Learning-Based Inverse Dynamics for Human Motion Analysis

BAND  
1|1

VOLUME  
1|1



**Institut für Informationsverarbeitung**  
[www.tnt.uni-hannover.de](http://www.tnt.uni-hannover.de)

Zell, Petrisa

## **Learning-Based Inverse Dynamics for Human Motion Analysis**

Fortschritt-Berichte VDI, Reihe 10, Nr. 877. Düsseldorf: VDI Verlag 2022.

160 Seiten, 35 Bilder, 15 Tabellen.

ISBN 987-3-18-387710-2, E-ISBN 978-3-18-687710-9, ISSN 0178-9627,

57,00 EUR/VDI-Mitgliederpreis: 51,30

**Für die Dokumentation:** inverse Dynamik – maschinelles Lernen – menschliche Bewegung – Gelenkmomente – Ganganalyse – künstliche neuronale Netze – selbstüberwachtes Lernen

**Keywords:** inverse dynamics – machine learning – human motion – joint moments – gait analysis – artificial neural networks – self-supervised learning

This dissertation deals with machine learning techniques for inverse dynamics of human motion. Inverse dynamics refers to the derivation of acting forces and moments from the motion of a kinematic model. More precisely, the objective is to estimate joint torques, ground reaction forces and ground reaction moments at both feet based on the three-dimensional input motion of a skeletal model. The problem is solved using a data-driven machine learning approach, proposing several regression models that are particularly suitable with respect to limited data availability. The goal is to exploit the inherent strengths of machine learning, such as fast and noise-resistant data analysis. The described methods are able to predict underlying joint torques and exterior forces with high precision (on gait sequences: relative root mean squared errors of 7.0 %, 16.1 % and 11.9 % for reaction forces, reaction moments and joint moments which correspond to Pearson's correlation coefficients of 0.91, 0.83 and 0.82), while reducing computation times by two orders of magnitude compared to traditional optimization.

### **Bibliographische Information der Deutschen Bibliothek**

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter [www.dnb.de](http://www.dnb.de) abrufbar.

### **Bibliographic information published by the Deutsche Bibliothek (German National Library)**

The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at [www.dnb.de](http://www.dnb.de).

© VDI Verlag GmbH | Düsseldorf 2022

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe (Fotokopie, Mikrokopie), der Speicherung in Datenverarbeitungsanlagen, im Internet und das der Übersetzung, vorbehalten. Als Manuskript gedruckt. Printed in Germany.

ISBN 987-3-18-387710-2, E-ISBN 978-3-18-687710-9, ISSN 0178-9627

## ACKNOWLEDGEMENTS

---

This thesis was written during my time at the Institut für Informationsverarbeitung (TNT) of the Gottfried Wilhelm Leibniz Universität Hannover.

First and foremost, I would like to thank my thesis advisor, Prof. Dr.-Ing. Bodo Rosenhahn, for giving me the opportunity to do my doctoral studies under his supervision and for his scientific guidance and support during the entire time. I am very grateful for his motivational and open-minded style of mentoring and of course for the evaluation of my thesis as first examiner. I would also like to thank Prof. Dr.-Ing. Gerard Pons-Moll for the examination of my thesis and for bringing the research field of motion capture to my attention in the first place. I also cordially thank Prof. Dr.-Ing. Jörn Ostermann, who chaired the examination committee and whose assessment I learned to value very much.

I would like to express my sincere thanks to all of my colleagues for numerous constructive discussions and the productive working atmosphere. They made my time at TNT really memorable to me. I especially thank Bastian Wandt, who always supported me with goal-oriented discussions and advice and with whom I enjoyed sharing an office. I would also like to say a special thank you to Thorsten Laude, Felix Kuhnke, the TNT Alpine team and the entire eNIFE for the great time we spent during and outside of work. Moreover, I thank all participants in my data set recordings for their active contribution despite the physically demanding exercises and Matthias Schuh for setting up the measurement equipment. I also like to thank Martin Pahl, Marco Munderloh and Thomas Wehberg for their constant administrative support and Doris Jaspers-Göhring, Melanie Huch and Ellen Sylla for their help in organizational tasks.

I sincerely thank Carolin Scheler for the joint working meetings during the COVID-19 pandemic, which made the lengthy writing process much more pleasant, and for proofreading parts of the thesis. Finally, my deepest gratitude goes to my family and my partner Hendrik Hachmann for their tireless encouragement and motivation and for always having an open ear for my thoughts and concerns.

Dedicated to the loving memory of Jürgen Zell.  
1944 – 2017

# CONTENTS

---

1	INTRODUCTION	1
1.1	Applications and Challenges of Inverse Dynamics	1
1.2	Learning Inverse Dynamics	4
1.3	Contributions	6
1.4	Structure of the Thesis	8
1.5	Publications	11
2	RELATED WORK	15
2.1	Inverse Dynamics by Physical Simulation	15
2.1.1	Inverse Approaches	16
2.1.2	Forward Approaches	17
2.1.3	Implicit Approaches	18
2.2	Learning-Based Inverse Dynamics	18
2.3	Decreasing Supervision	20
3	FUNDAMENTALS	23
3.1	Rigid Body Motion	23
3.1.1	Representation of Position	23
3.1.2	Representation of Orientation	24
3.1.3	Homogeneous Transformations	26
3.2	Kinematics of a Rigid Body System	27
3.2.1	Kinematic Trees	27
3.2.2	The Denavit-Hartenberg Convention	28
3.2.3	Velocity and Acceleration Kinematics	30
3.3	Dynamics of a Rigid Body System	41
3.3.1	TMT-Method	42
3.4	Machine Learning	44
3.4.1	Terminology and General Concepts	45
3.4.2	Support Vector Machines	46
3.4.3	Ridge Regression	48
3.4.4	Random Forests	49
3.4.5	Neural Networks	52
3.4.6	Generalization	58
3.4.7	Transfer Learning	59

4	HUMAN MOTION DATASET	62
4.1	Motion Capture and Kinematic Optimization	62
4.2	Force Plate Measurements	64
4.3	Estimation of Inertial Properties	65
4.4	Optimization of Joint Torques	68
4.5	Data Specification	71
4.6	Generation of Training Data Points	73
5	SUPERVISED LEARNING OF INVERSE DYNAMICS	76
5.1	Methodology	77
5.1.1	End-to-End Regression	79
5.1.2	Multi-Stage Regression	79
5.2	Experimental Evaluation	81
5.2.1	Predictive Dynamics Dataset	82
5.2.2	Public Dataset	89
5.2.3	Application to Reconstructed Motions	90
5.3	Discussion	91
6	SELF-SUPERVISION BY DYNAMICS-BASED LAYERS	95
6.1	Datasets	97
6.2	Dynamics Network	98
6.2.1	Forward Layer	100
6.2.2	Inverse Layer	102
6.2.3	Contact Loss	103
6.2.4	Training Modes	103
6.3	Experimental Evaluation	104
6.3.1	Comparison in the Supervised Setting	105
6.3.2	Semi-Supervision with Small Labeled Datasets	106
6.3.3	Domain Adaptation	109
6.3.4	Ablation of Input Structure	115
6.3.5	Effect of Noise	116
6.4	Discussion	117
7	CONCLUSIONS	123
A	APPENDIX	127
A.1	Evaluation Based on Additional Metrics	127
A.2	Data-Driven Inverse Dynamics Optimization	127
	BIBLIOGRAPHY	130

## ACRONYMS

---

EOM	Equation Of Motion
GRF	Ground Reaction Force
GRM	Ground Reaction Moment
GRF/M	Ground Reaction Force and Moment
COP	Center Of Pressure
IMU	Inertial Measurement Unit
DOF	Degree Of Freedom
SVM	Support Vector Machine
RF	Random Forest
CART	Classification And Regression Trees
NN	artificial Neural Network
SGD	Stochastic Gradient Descent
ReLU	Rectified Linear Unit
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
rRMSE	relative Root Mean Squared Error
MAE	Mean Absolute Error
CE	Cross Entropy
PDO	Predictive Dynamics Optimization
PD-set	Predictive Dynamics Set
F-net	neural network trained with forward loss
cFI-net	neural network trained with contact, forward and inverse loss
CMU	Carnegie Mellon University

## NOTATIONS

---

$\mathbf{p}$	joint positions
$\mathbf{q}$	generalized coordinates
$\boldsymbol{\xi}$	linear and angular body velocities
$\mathbf{v}$	linear body velocity
$\boldsymbol{\omega}$	angular body velocity
$\mathbf{x}$	kinematic state
$\hat{\mathbf{x}}$	kinematic state without global coordinates
$\mathbf{x}_s$	segment center of mass positions and velocities
$\mathbf{x}_j$	joint positions and velocities
$\mathbf{M}$	inertia matrix
$\mathcal{M}$	reduced inertia matrix
$\mathcal{F}$	reduced force
$m_i$	mass of segment $i$
$\mathbf{I}_i$	tensor of inertia of segment $i$
$\mathbf{f}_c$	concatenation of contact forces and moments
$\boldsymbol{\tau}$	joint torques
$\mathbf{u}$	controls (concatenation of $\mathbf{f}_c$ and $\boldsymbol{\tau}$ )
$\mathbf{T}$	Jacobian of kinematic coordinate transformations
$\mathbf{T}_{v_i}$	rows of Jacobian transforming to linear velocity of segment $i$
$\mathbf{T}_{\omega_i}$	rows of Jacobian transforming to rotational velocity of segment $i$
$\mathbf{z}_{j-1}$	z-axis of link $j$
$\mathbf{t}_{j-1}$	origin of link $j$
$\kappa(j)$	support set of link $j$
$\mu(j)$	subtree set starting at link $j$
$\nu(i, j)$	subchain set between links $i$ and $j$
$\mathbf{R}_n^0$	rotation matrix from frame $n$ to the global frame
$\zeta$	convective acceleration
$\mathbf{g}$	gravitational acceleration

$\mathbf{f}_r$	ground reaction force
$\mathbf{m}_r$	ground reaction moment
$\mathbf{r}_{(\cdot)}$	position vector of point $(\cdot)$
$l$	segment lengths
$T$	window size (number of frames)
$\alpha_{(\cdot)}$	polynomial coefficients of variable $(\cdot)$
$\mathbf{v}_c$	contact point velocity vector
$\boldsymbol{\theta}$	feature vector
$l_c$	class label of gait phase
$c_i$	contact state of foot $i$
$d$	damping in forward layer
$f(), \mathbf{f}()$	functions implemented by neural networks
$L$	loss function
$\epsilon_{(\cdot)}$	relative root mean squared error of variable $(\cdot)$
$\rho_{(\cdot)}$	Pearson's correlation coefficient related to variable $(\cdot)$
$e_{\text{EOM}}$	equation of motion error

## ABSTRACT

---

This dissertation deals with machine learning techniques for inverse dynamics of human motion. Inverse dynamics refers to the derivation of acting forces and moments from the motion of a kinematic model. More precisely, the objective is to estimate joint torques, ground reaction forces and ground reaction moments at both feet based on the three-dimensional input motion of a skeletal model. Of particular interest are the joint torques, also specified as net joint moments, since they correspond to the total effect of all forces on the joints. In the context of biomechanical investigations, they represent a common measure of the load on joints.

Traditional approaches formulate the problem as an optimization that incorporates the equation of motion (EOM) of a physical model of the human body. The EOM is either used in a forward or an inverse sense which implies either integration or differentiation of kinematics. Both processes are prone to error propagation and complicate the convergence of the optimization algorithms built on the formulation. Furthermore, the EOM belonging to a multi-body system, such as the modeled human body, gives rise to a highly non-linear and non-convex objective function which is notoriously hard to optimize. Last but not least, conventional methods generally rely on measured external reaction forces and moments, which severely limits the motions that can be analyzed due to the laboratory environment required.

Given these limitations, data-driven machine learning techniques open up tremendous opportunities by enabling fast and noise-resistant data analysis. This thesis investigates the applicability of such methods to inverse dynamics of human motion and addresses the design of suitable regression models. The proposed methods are able to predict underlying joint torques and exterior forces with high precision (on gait sequences: relative root mean squared errors of 7.0 %, 16.1 % and 11.9 % for reaction forces, reaction moments and joint moments which correspond to Pearson's correlation coefficients of 0.91, 0.83 and 0.82), while reducing computation times by two orders of magnitude compared to traditional optimization.

A general feature of human motion data is the discontinuity at contact phase transitions, e. g. at the moment the foot touches the ground. By changing the number of contact points of the human model to its environment, the set of dynamic equations is fundamentally altered to the extent that external influences are allowed or forbidden at the corresponding points. Motivated by this property, a multi-stage regression approach is presented. The method initially identifies the current gait phase and limits the inference of joint torques

as well as contact forces to the resulting sub-space. This way, the regression of unrealistic non-zero forces during swing phases is significantly reduced compared to a model that estimates the forces without knowledge of the contact state.

Current problems of machine learning methods for solving inverse dynamics are a lack of suitable datasets and that the compliance with the EOM is not guaranteed for the predictions. Both issues are addressed by a self-supervised learning method presented in this thesis. The approach allows cycle consistent training of an artificial neural network with pure motion data, i. e. without any ground truth forces and moments. Instead of minimizing a direct loss on the target forces, the model solves an initial value problem based on predicted forces and minimizes the distance between the resulting simulation and the input motion. This is realized by implementing a differentiable forward dynamics loss layer that allows backward flow of gradients and can be integrated into the training of the neural network. In addition, the model includes a corresponding inverse dynamics layer that evaluates the estimated contact forces decoupled from predicted joint torques. Thus, the model not only allows training on readily available motion data, but also constrains the predicted variables using both dynamic directions for optimal satisfaction of the EOM. The neural network maintains stable performance even with small labeled datasets consisting of dynamics data of only two or three subjects by learning generalization capability on larger unlabeled motion sets. Furthermore, the method enables self-supervised transfer learning to different motion types, movement speeds and skeleton characteristics.

The presented learning-based inverse dynamics approaches are evaluated using a self-recorded dataset of walking and running sequences performed by 22 subjects as well as a public dynamics dataset [39] and gait sequences from the well-known CMU database [18]. The self-recorded dataset is available to the research community.

## KURZFASSUNG

---

Diese Dissertation beschäftigt sich mit maschinellen Lernverfahren für die inverse Dynamik der menschlichen Bewegung. Unter inverser Dynamik versteht man die Ableitung von wirkenden Kräften und Momenten aus der Bewegung eines kinematischen Modells. Genauer gesagt geht es um die Abschätzung von Gelenkmomenten, Bodenreaktionskräften und Bodenreaktionsmomenten an beiden Füßen basierend auf der dreidimensionalen Eingangsbewegung eines Skelettmodells. Von besonderem Interesse sind die Gelenkmomente, die auch als Netto-Gelenkmomente bezeichnet werden, da sie der Gesamtwirkung aller Kräfte an den Gelenken entsprechen. Im Rahmen biomechanischer Untersuchungen stellen sie ein gängiges Maß für die Beanspruchung von Gelenken dar.

Traditionelle Ansätze formulieren das Problem als eine Optimierung, die die Bewegungsgleichung (Equation of Motion, EOM) eines physikalischen Modells des menschlichen Körpers einbezieht. Die EOM wird entweder in einem vorwärts gerichteten oder einem inversen Sinn verwendet, was entweder eine Integration oder eine Differenzierung der Kinematik impliziert. Beide Verfahren sind anfällig für Fehlerfortpflanzung und erschweren die Konvergenz des Optimierungsalgorithmus. Darüber hinaus führt die zu einem Mehrkörpersystem, wie dem modellierten menschlichen Körper, gehörende EOM, zu einer hochgradig nichtlinearen und nichtkonvexen Zielfunktion, die schwer zu optimieren ist. Zudem stützen sich konventionelle Methoden in der Regel auf gemessene externe Reaktionskräfte und -momente, was die zu analysierenden Bewegungen aufgrund der erforderlichen Laborumgebung stark einschränkt.

Angesichts dieser Einschränkungen eröffnen datengesteuerte maschinelle Lernverfahren enorme Möglichkeiten, da sie generell eine schnelle und rauschresistente Datenanalyse erlauben. Diese Arbeit untersucht die Anwendbarkeit solcher Methoden auf die inverse Dynamik der menschlichen Bewegung und beschäftigt sich mit dem Entwurf geeigneter Regressionsmodelle. Die vorgeschlagenen Methoden sind in der Lage, die zugrundeliegenden Gelenkmomente und äußeren Kräfte mit hoher Genauigkeit (bei Gangsequenzen: relative mittlere quadratische Fehler von 7,0 %, 16,1 % und 11,9 % für Reaktionskräfte, Reaktionsmomente und Gelenkmomente, was Pearson's Korrelationskoeffizienten von 0,91, 0,83 und 0,82 entspricht) vorherzusagen und gleichzeitig die Berechnungszeiten um zwei Größenordnungen im Vergleich zur traditionellen Optimierung zu reduzieren.

Ein allgemeines Merkmal menschlicher Bewegungsdaten ist die Diskontinuität an Kontaktphasenübergängen, z.B. im Moment der Bodenberührung des Fußes. Durch Veränderung der Anzahl der Kontaktpunkte des menschlichen Modells zu seiner Umgebung wird der

Satz der dynamischen Gleichungen grundlegend verändert, und zwar in dem Sinn, dass äußere Einflüsse an den entsprechenden Punkten erlaubt oder verboten werden. Motiviert durch diese Eigenschaft, wird ein mehrstufiger Regressionsansatz vorgestellt. Das Verfahren identifiziert zunächst die aktuelle Gangphase und beschränkt die Inferenz von Gelenkmomenten und Kontaktkräften auf den resultierenden Unterraum. Auf diese Weise wird die Regression unrealistischer endlicher Kräfte während der Schwungphasen im Vergleich zu einem Modell, das die Kräfte ohne Kenntnis des Kontaktzustandes schätzt, deutlich reduziert.

Aktuelle Probleme von maschinellen Lernmethoden zur Lösung der inversen Dynamik sind ein Mangel an geeigneten Datensätzen und dass die Einhaltung der EOM durch die vorhergesagten Größen nicht garantiert ist. Beide Probleme werden durch ein in dieser Arbeit vorgestelltes selbst-überwachtes Lernverfahren adressiert. Der Ansatz erlaubt ein zyklus konsistentes Training eines künstlichen neuronalen Netzes mit reinen Bewegungsdaten, d.h. ohne jegliche Ground-Truth-Kräfte und -Momente. Anstatt einen direkten Verlust auf die Zielkräfte zu minimieren, löst das Modell ein Anfangswertproblem basierend auf den vorhergesagten Kräften und minimiert den Abstand zwischen der resultierenden Simulation und der Eingangsbewegung. Dies wird durch die Implementierung einer differenzierbaren vorwärtsdynamischen Verlustschicht realisiert, die einen Rückwärtsfluss von Gradienten erlaubt und in das Training des neuronalen Netzes integriert werden kann. Zusätzlich enthält das Modell eine entsprechende Schicht für die inverse Dynamik, die die geschätzten Kontaktkräfte entkoppelt von den vorhergesagten Gelenkmomenten auswertet. Somit ermöglicht das Modell nicht nur das Training auf leicht verfügbaren Bewegungsdaten, sondern beschränkt auch die vorhergesagten Variablen unter Verwendung beider dynamischer Richtungen zur optimalen Erfüllung der EOM. Das neuronale Netzwerk behält seine stabile Leistung auch bei kleinen gelabelten Datensätzen, die aus Dynamikdaten von nur 2 bis 3 Probanden bestehen, indem es die Fähigkeit zu generalisieren auf größeren nicht gelabelten Bewegungsdatensätzen lernt. Darüber hinaus ermöglicht die Methode selbst-überwachtes Transferlernen unbekannter Bewegungstypen, Bewegungsgeschwindigkeiten und abweichender Skelettmerkmale.

Die vorgestellten lernbasierten inversen Dynamikansätze werden anhand eines selbst aufgezeichneten Datensatzes von Geh- und Laufsequenzen, die von 22 Probanden ausgeführt wurden, sowie eines öffentlichen Dynamikdatensatzes [39] und Gangsequenzen aus der bekannten CMU-Datenbank [18] evaluiert. Der selbst aufgezeichnete Datensatz steht der Forschungsgemeinschaft zur Verfügung.

