
Reinforcement Learning Aided Performance Optimization of Feedback Control Systems

Changsheng Hua

Reinforcement Learning Aided Performance Optimization of Feedback Control Systems

 Springer Vieweg

Changsheng Hua
Duisburg, Germany

Von der Fakultät für Ingenieurwissenschaften, Abteilung Elektrotechnik und Informationstechnik der Universität Duisburg-Essen zur Erlangung des akademischen Grades Doktor der Ingenieurwissenschaften (Dr.-Ing.) genehmigte Dissertation von Changsheng Hua aus Jiangsu, V.R. China.

1. Gutachter: Prof. Dr.-Ing. Steven X. Ding
 2. Gutachter: Prof. Dr. Yuri A.W. Shardt
- Tag der mündlichen Prüfung: 23.01.2020

ISBN 978-3-658-33033-0 ISBN 978-3-658-33034-7 (eBook)
<https://doi.org/10.1007/978-3-658-33034-7>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Fachmedien Wiesbaden GmbH, part of Springer Nature 2021

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Responsible Editor: Stefanie Eggert

This Springer Vieweg imprint is published by the registered company Springer Fachmedien Wiesbaden GmbH part of Springer Nature.

The registered company address is: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

*To my parents, my brother, and my wife
Xiaodi*

Acknowledgments

First and foremost, I am deeply indebted to my advisor, Prof. Dr.-Ing. Steven X. Ding, for his guidance, encouragement and insightful discussions during my Ph.D. studies. He has continually pointed me in the right direction and provided me inspiration to do the best work I could. I would also like to express my heartfelt appreciation to Prof. Dr. Yuri A.W. Shardt, who has sparked my interest in performance optimization of control systems and mentored me a lot. He has shared with me his rich experience in academic research and scientific writing. I feel very lucky to have him as one of my major collaborators.

I would particularly like to give thanks to Dr.-Ing. Linlin Li and Dr.-Ing. Hao Luo for many insightful discussions and constructive comments during my studies. From both of them, I have learned a lot on robust and optimal control. I would also like to thank Dr.-Ing. Birgit Köppen-Seliger, Dr.-Ing. Chris Louen, Dr.-Ing. Minjia Krüger and Dr.-Ing. Tim Könings for giving me support and valuable advice in supervision of exercises and research projects.

I owe a great debt of gratitude to Dr.-Ing. Zhiwen Chen, Dr.-Ing. Kai Zhang, Dr.-Ing. Yunsong Xu, Dr.-Ing. Lu Qian, who offered enormous help and support during the early days of my life in Duisburg. I am particularly thankful to M.Sc. Micha Obergfell, M.Sc. Yuhong Na, M.Sc. Frederick Hesselmann, M.Sc. Hogir Rafiq, M.Sc. Ting Xue, M.Sc. Deyu Zhang, M.Sc. Reimann Christopher, M.Sc. Caroline Zhu, M.Sc. Yannian Liu, M.Sc. Tieqiang Wang, M.Sc. Jiarui Zhang for making my life in AKS much enjoyable, and for giving me valuable suggestions, generous support and encouragement. I would like to extend my thanks to numerous visiting scholars for all the advice, support, help and the great times. I am also very grateful for the administrative and technical assistance given by Mrs. Sabine Bay, Dipl.-Ing. Klaus Göbel and Mr. Ulrich Janzen.

Lastly, I would like to dedicate this work to my family, to my parents for their unconditional love and care, to my brother for igniting my passion for engineering and all the years of care and support, and especially to my dear wife Xiaodi for being with me all these years with patience and faithful support.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Scope of the Work	4
1.3	Objective of the Work	5
1.4	Outline of the Thesis	6
2	The Basics of Feedback Control Systems	9
2.1	Design of Feedback Controllers	9
2.1.1	Description of a Nominal System	9
2.1.2	A Coprime Factorization Design Tool	10
2.1.3	Well-posedness and Internal Stability	12
2.1.4	Parameterization of Stabilizing Controllers	15
2.2	Model Uncertainty and Robustness	18
2.2.1	Small Gain Theorem	19
2.2.2	Coprime Factor Representation of Model Uncertainty	20
2.2.3	Dual YK Representation of Model Uncertainty	22
2.3	Concluding Remarks	25
3	Reinforcement Learning and Feedback Control	27
3.1	An Overview of RL Methods	28
3.2	Infinite Horizon Linear Quadratic Regulator	30
3.2.1	An Overview of the Infinite Horizon LQR Problem and DP	30
3.2.2	Policy Iteration and Value Iteration	32
3.2.3	Q -learning	35
3.2.4	SARSA	38

3.2.5	Simulation Results	41
3.3	Infinite Horizon Linear Quadratic Gaussian	42
3.3.1	An Overview of Infinite Horizon LQG Problem and Stochastic DP	43
3.3.2	On-policy Natural Actor-critic	46
3.3.3	Simulation Results	54
3.4	Concluding Remarks	57
4	<i>Q</i>-learning Aided Performance Optimization of Deterministic Systems	59
4.1	Problem Formulation	59
4.2	Robustness Optimization	60
4.2.1	Existing Robustness Optimization Approaches	61
4.2.2	Input and Output Recovery	63
4.2.3	Robustness Optimization Using <i>Q</i> -learning	64
4.2.4	Simulation Results	69
4.3	Performance Optimization Using a Prescribed Performance Index	72
4.3.1	Performance Optimization Using <i>Q</i> -learning	72
4.3.2	An Extension to Tracking Performance Optimization	75
4.3.3	Simulation Results	78
4.4	Concluding Remarks	82
5	NAC Aided Performance Optimization of Stochastic Systems	85
5.1	Problem Formulation	86
5.2	Robustness Optimization	87
5.2.1	Noise Characteristics	87
5.2.2	Conditions of Closed-loop Internal Stability	90
5.2.3	Robustness Optimization using NAC	92
5.2.4	Simulation Results	95
5.3	Performance Optimization using a Prescribed Performance Index	99
5.3.1	Efficient NAC Learning using Both Nominal Plant Model and Data	99
5.3.2	Data-driven Performance Optimization	100
5.3.3	Simulation Results	103
5.4	Performance Optimization of Plants with General Output Feedback Controllers	106
5.4.1	Problem Formulation	107
5.4.2	Performance Optimization using NAC	108

5.4.3	Experimental Results	111
5.5	Concluding Remarks	117
6	Conclusion and Future Work	119
6.1	Conclusion	119
6.2	Future Work	121
Bibliography	123

Abbreviations and Notation

Abbreviations

Abbreviation	Description
BLDC	brushless direct current
DP	dynamic programming
ECU	electronic control unit
I/O	input/output
IOR	input and output recovery
KL	Kullback-Leibler
LCF	left coprime factorization
LQG	linear quadratic Gaussian
LQR	linear quadratic regulator
LS	least squares
LTI	linear time-invariant
LTR	loop transfer recovery
MIMO	multiple-input multiple-output
NAC	natural actor-critic
PI	proportional-integral
PID	proportional-integral-derivative
RCF	right coprime factorization
RL	reinforcement learning
SARSA	state-action-reward-state-action
SGD	stochastic gradient descent
SISO	single-input single-output
TD	temporal difference

2-DOF	two-degree-of-freedom
YK	Youla-Kučera

Notation

Notation	Description
\forall	for all
\in	belong to
\sim	follow
\approx	approximately equal
\neq	not equal
$:=$	defined as
\Rightarrow	imply
\gg	much greater than
\otimes	Kronecker product
$ \cdot $	determinant of a matrix or absolute value
\mathbb{Z}^+	set of non-negative integers
\mathbb{R}^n	space of n -dimensional column vectors
$\mathbb{R}^{n \times m}$	space of n by m matrices
x	a scalar
$\ln x$	natural logarithm of x
\mathbf{x}	a vector
\mathbf{X}	a matrix
\mathbf{X}^T	transpose of \mathbf{X}
\mathbf{X}^{-1}	inverse of \mathbf{X}
$\text{tr}(\mathbf{X})$	trace of \mathbf{X}
$\mathbf{X} > \mathbf{0}$	\mathbf{X} is a positive definite matrix
$\text{vec}(\mathbf{X})$	vectorization of \mathbf{X} , $\text{vec}(\mathbf{X}) = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{bmatrix} \in \mathbb{R}^{nm}$, for
	$\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_m] \in \mathbb{R}^{n \times m}$, $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, \dots, m$
\mathbf{I}_m	m by m identity matrix
$\mathbf{0}_{m \times n}$	m by n zero matrix
\mathcal{RH}_∞	set of all proper and real rational stable transfer matrices
\mathcal{R}_p	set of all proper and real rational transfer matrices
\mathcal{R}_{sp}	set of all strictly proper and real rational transfer matrices
$\ \mathbf{P}\ _\infty$	\mathcal{H}_∞ norm of a transfer matrix \mathbf{P}

$\left[\begin{array}{c c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right]$	shorthand for state-space realization $\mathbf{C}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$
$\arg \min_{\mathbf{u}}(f(\mathbf{u}))$	a value of \mathbf{u} at which $f(\mathbf{u})$ takes its minimum value
$\mathcal{N}(\mathbf{a}, \Sigma)$	Gaussian distribution with a mean vector \mathbf{a} and a covariance matrix Σ
$\mathbb{E}(\cdot)$	Mean value/vector
μ/π	deterministic/stochastic policy
μ^*/π^*	(sub)optimal deterministic/stochastic policy
γ	discount factor
$c(\mathbf{x}, \mathbf{u})$	one-step cost
μ^i	i^{th} iteration of policy μ
$\mathbf{R}_j^{\mu^i}$	j^{th} iteration of a parameter matrix \mathbf{R} under the i^{th} iteration of the policy μ
π_{θ}	a stochastic policy corresponding to a parameter vector θ
$\delta(k)$	temporal difference error at the sampling instant k
$V^{\pi}(\mathbf{x})$	value function of policy π
$Q^{\pi}(\mathbf{x}, \mathbf{u})$	Q -function of policy π
$A^{\pi}(\mathbf{x}, \mathbf{u})$	advantage function of policy π
$\nabla_{\theta} f(\theta)$	derivative of $f(\theta)$ with respect to a parameter vector θ

List of Figures

Figure 1.1	Structure for performance optimization	4
Figure 1.2	Regions of performance [9]	5
Figure 2.1	Internal stability analysis diagram	12
Figure 2.2	Representation of a feedforward/feedback controller	14
Figure 2.3	Representation of a feedforward/feedback controller as a feedback controller for an augmented plant	14
Figure 2.4	LCF and RCF of $\mathbf{K}(\mathbf{Q}_r)$	16
Figure 2.5	Observer-based representation of all stabilizing feedback controllers of \mathbf{P}_0	17
Figure 2.6	Observer-based representation of all stabilizing 2-DOF controllers of \mathbf{P}_0	18
Figure 2.7	System representation for robust stability analysis	19
Figure 2.8	Representation of $\mathbf{P}_\Delta(z)$ with left coprime factor uncertainty	21
Figure 2.9	Representation of $\mathbf{P}_\Delta(z)$ with right coprime factor uncertainty	22
Figure 2.10	Equivalent robust stability property	24
Figure 3.1	Interconnection of RL and feedback control	28
Figure 3.2	On-policy and off-policy RL for controller optimization	29
Figure 3.3	Comparison of Q -learning and SARSA	42
Figure 3.4	Parameters Θ and cost J during online learning	55
Figure 3.5	Comparison of system performance with different controllers	56
Figure 4.1	The initial control structure	60

Figure 4.2	Comparison of robustness and performance of different controllers	61
Figure 4.3	The performance optimization structure with Q_r	62
Figure 4.4	Comparison of zero input response of systems before and after perturbations	70
Figure 4.5	Convergence of control parameters of Q_r during online learning	70
Figure 4.6	Comparison of performance of perturbed systems with and without Q_r	71
Figure 4.7	Comparison of robustness of perturbed systems with and without Q_r	71
Figure 4.8	Observer-based I/O data model of the deterministic plant $P(S_f)$	72
Figure 4.9	Tracking control structure	76
Figure 4.10	A schematic sketch of an inverted pendulum system	79
Figure 4.11	Comparison of performance of inverted pendulum systems before and after the perturbation	80
Figure 4.12	Convergence of Θ and F_a during online learning	81
Figure 4.13	Comparison of zero input response of systems consisting of $P(S_f)$ and different controllers	82
Figure 5.1	The system configuration after perturbations	86
Figure 5.2	LCF and RCF of $P(S_f)$	88
Figure 5.3	Comparison of zero input response of stochastic systems before and after perturbations	96
Figure 5.4	Parameters of Q_r during online learning	97
Figure 5.5	Cost J_s during online learning	97
Figure 5.6	Comparison of performance of perturbed stochastic systems with and without Q_r	98
Figure 5.7	Comparison of robustness of perturbed stochastic systems with and without Q_r	98
Figure 5.8	Observer-based I/O data model of the stochastic plant $P(S_f)$	100
Figure 5.9	Data-driven performance optimization of Q_r using NAC	103
Figure 5.10	Comparison of performance of systems before and after the perturbation	104
Figure 5.11	Parameters of Q_r and cost J_u during online learning	105
Figure 5.12	Comparison of performance of systems consisting of $P(S_f)$ and different controllers	106

Figure 5.13	Configuration of the system with a general feedback controller	107
Figure 5.14	Performance optimization with an auxiliary controller K_A	108
Figure 5.15	Data-driven performance optimization of K_A using NAC	112
Figure 5.16	BLDC motor test rig [46]	112
Figure 5.17	Components of the xPC target environment	113
Figure 5.18	Tracking performance of the BLDC motor system with a speed PI controller	113
Figure 5.19	Data-driven optimization of speed tracking performance of the BLDC motor	114
Figure 5.20	Cost J during online learning	115
Figure 5.21	Parameters a_1 , a_2 and a_3 during online learning	116
Figure 5.22	Comparison of performance of systems with and without K_A	117