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

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

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
$\vert\cdot\vert$	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
$\boldsymbol{x}$	a vector
$\boldsymbol{X}$	a matrix
$\boldsymbol{X}^T$	transpose of $\boldsymbol{X}$
$\boldsymbol{X}^{-1}$	inverse of $\boldsymbol{X}$
$\mathrm{tr}(\boldsymbol{X})$	trace of $\boldsymbol{X}$
$\boldsymbol{X} > \boldsymbol{0}$	$\boldsymbol{X}$ is a positive definite matrix
$\mathrm{vec}(\boldsymbol{X})$	vectorization of $\boldsymbol{X}$, $\mathrm{vec}(\boldsymbol{X}) = \begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_m} \end{bmatrix} \in \mathbb{R}^{nm}$, for $\boldsymbol{X} = [\boldsymbol{x_1} \cdots \boldsymbol{x_m}] \in \mathbb{R}^{n\times m}$, $\boldsymbol{x_i} \in \mathbb{R}^n, i = 1, \cdots, m$
$\boldsymbol{I}_m$	m by m identity matrix
$\boldsymbol{0}_{m\times n}$	m by n zero matrix
$\mathcal{RH}_\infty$	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
$\Vert\boldsymbol{P}\Vert_\infty$	$\mathcal{H}_\infty$ norm of a transfer matrix $\boldsymbol{P}$

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

List of Figures