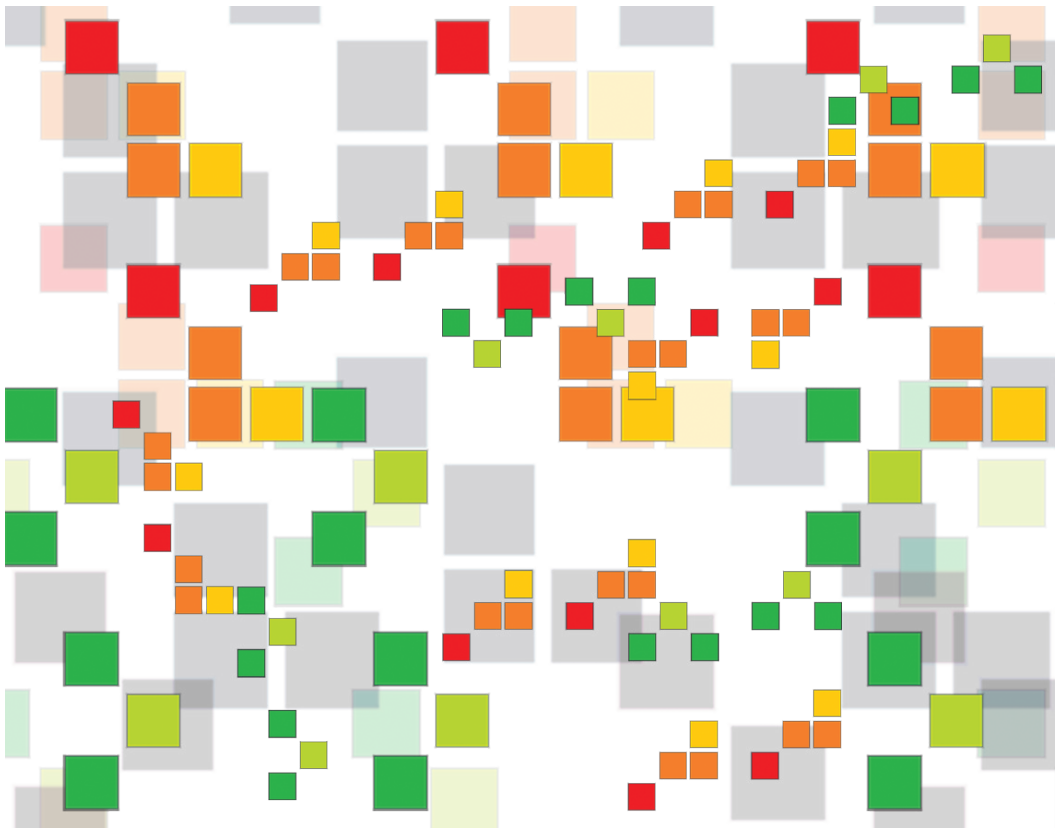

Nonnegative Matrix Factorization



Nicolas Gillis

University of Mons
Mons, Belgium

siam[®]

Society for Industrial and Applied Mathematics
Philadelphia

Copyright © 2021 by the Society for Industrial and Applied Mathematics

10 9 8 7 6 5 4 3 2 1

All rights reserved. Printed in the United States of America. No part of this book may be reproduced, stored, or transmitted in any manner without the written permission of the publisher. For information, write to the Society for Industrial and Applied Mathematics, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688 USA.

No warranties, express or implied, are made by the publisher, authors, and their employers that the programs contained in this volume are free of error. They should not be relied on as the sole basis to solve a problem whose incorrect solution could result in injury to person or property. If the programs are employed in such a manner, it is at the user's own risk and the publisher, authors, and their employers disclaim all liability for such misuse.

Trademarked names may be used in this book without the inclusion of a trademark symbol. These names are used in an editorial context only; no infringement of trademark is intended.

MATLAB is a registered trademark of The MathWorks, Inc. For MATLAB product information, please contact The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098 USA, 508-647-7000, Fax: 508-647-7001, info@mathworks.com, www.mathworks.com.

| | |
|-------------------------------|---------------------|
| <i>Publications Director</i> | Kivmars H. Bowling |
| <i>Executive Editor</i> | Elizabeth Greenspan |
| <i>Developmental Editor</i> | Mellisa Pascale |
| <i>Managing Editor</i> | Kelly Thomas |
| <i>Production Editor</i> | Lisa Briggeman |
| <i>Copy Editor</i> | Susan Fleshman |
| <i>Production Manager</i> | Donna Witzleben |
| <i>Production Coordinator</i> | Cally A. Shrader |
| <i>Compositor</i> | Cheryl Hufnagle |
| <i>Graphic Designer</i> | Doug Smock |

Library of Congress Cataloging-in-Publication Data

Names: Gillis, Nicolas, author.

Title: Nonnegative matrix factorization / Nicolas Gillis, University of Mons, Mons, Belgium.

Description: Philadelphia : Society for Industrial and Applied Mathematics, [2021] | Series: Data science ; 2 | Includes bibliographical references and index. | Summary: "This book provides a comprehensive and up-to-date account of the NMF problem and its most significant features"-- Provided by publisher.

Identifiers: LCCN 2020042037 (print) | LCCN 2020042038 (ebook) | ISBN 9781611976403 (paperback) | ISBN 9781611976410 (ebook)

Subjects: LCSH: Non-negative matrices. | Factorization (Mathematics) | Computer algorithms. | Data mining.

Classification: LCC QA188 .G566 2021 (print) | LCC QA188 (ebook) | DDC 512.9/434--dc23

LC record available at <https://lcn.loc.gov/2020042037>

LC ebook record available at <https://lcn.loc.gov/2020042038>

Contents

| | |
|--|------------|
| Preface | xi |
| Notation | xv |
| List of Figures | xxi |
| List of Tables | xxv |
| 1 Introduction | 1 |
| 1.1 Linear dimensionality reduction techniques for data analysis | 1 |
| 1.2 Problem definition | 4 |
| 1.3 Four applications of NMF in data analysis | 6 |
| 1.4 History | 12 |
| 1.5 Take-home messages | 16 |
| I Exact factorizations | 17 |
| 2 Exact NMF | 19 |
| 2.1 Geometric interpretation | 19 |
| 2.2 Restricted Exact NMF | 35 |
| 2.3 Computational complexity of RE-NMF and Exact NMF | 43 |
| 2.4 Take-home messages | 54 |
| 3 Nonnegative rank | 55 |
| 3.1 Some properties of the nonnegative rank | 55 |
| 3.2 The nonnegative rank under perturbations | 57 |
| 3.3 Generic values of the nonnegative rank | 61 |
| 3.4 Lower bounds on the nonnegative rank | 62 |
| 3.5 Upper bounds for the nonnegative rank | 82 |
| 3.6 Lower bounds on extended formulations via the nonnegative rank | 83 |
| 3.7 Link with communication complexity | 94 |
| 3.8 Other applications of the nonnegative rank | 97 |
| 3.9 Take-home messages | 97 |
| 4 Identifiability | 99 |
| 4.1 Case $\text{rank}(X) \leq 2$ | 101 |
| 4.2 Exact NMF with $r = \text{rank}(X)$ | 103 |
| 4.3 Regularized Exact NMF | 133 |
| 4.4 Take-home messages | 156 |

| | | |
|-----------|--|------------|
| II | Approximate factorizations | 157 |
| 5 | NMF models | 159 |
| 5.1 | Error measures | 160 |
| 5.2 | Model-order selection | 167 |
| 5.3 | Regularizations | 168 |
| 5.4 | NMF variants | 170 |
| 5.5 | Models related to NMF | 185 |
| 5.6 | Take-home messages | 193 |
| 6 | Computational complexity of NMF | 195 |
| 6.1 | Frobenius norm | 195 |
| 6.2 | Kullback–Leibler divergence | 200 |
| 6.3 | Infinity norm | 202 |
| 6.4 | Weighted Frobenius norm | 203 |
| 6.5 | Componentwise ℓ_1 norm | 203 |
| 6.6 | Other NMF models | 204 |
| 6.7 | Take-home messages | 205 |
| 7 | Near-separable NMF | 207 |
| 7.1 | Context and applications | 208 |
| 7.2 | Preliminaries | 211 |
| 7.3 | Idealized algorithm | 216 |
| 7.4 | Greedy/sequential algorithms | 223 |
| 7.5 | Heuristic algorithms | 247 |
| 7.6 | Convex-optimization-based algorithms | 248 |
| 7.7 | Summary of provably robust near-separable NMF algorithms | 256 |
| 7.8 | Separable tri-symNMF | 257 |
| 7.9 | Further readings | 259 |
| 7.10 | Take-home messages | 260 |
| 8 | Iterative algorithms for NMF | 261 |
| 8.1 | Preliminaries | 263 |
| 8.2 | The multiplicative updates | 270 |
| 8.3 | Algorithms for the Frobenius norm | 280 |
| 8.4 | Number of inner iterations and acceleration | 291 |
| 8.5 | Stopping criteria | 297 |
| 8.6 | Initialization | 299 |
| 8.7 | Alternative algorithmic approaches | 301 |
| 8.8 | Further readings | 304 |
| 8.9 | Online resources | 305 |
| 8.10 | Take-home messages | 305 |
| 9 | Applications | 307 |
| 9.1 | Beware of scaling ambiguity | 307 |
| 9.2 | Should your data set be approximately of low rank? | 308 |
| 9.3 | Self-modeling curve resolution | 308 |
| 9.4 | Gene expression analysis | 309 |
| 9.5 | Recommender systems and collaborative filtering | 311 |
| 9.6 | Other applications | 313 |
| 9.7 | Take-home messages | 314 |

| | |
|---------------------|------------|
| Contents | ix |
| Bibliography | 315 |
| Index | 343 |
