Nonnegative Matrix Factorization

Book · December 2020

DOI: 10.1137/1.9781611976410

1 author:

Nicolas Gillis
Université de Mons

Data Science Book Series

Nonnegative matrix factorization (NMF) in its modern form has become a standard tool in the analysis of high-dimensional data sets. This book provides a comprehensive and up-to-date account of the most important aspects of the NMF problem and is the first to detail its theoretical aspects, including geometric interpretation, nonnegative rank, complexity, and uniqueness. It explains why understanding these theoretical insights is key to using this computational tool effectively and meaningfully.

Nonnegative Matrix Factorization

- is accessible to a wide audience and is ideal for anyone interested in the workings of NMF,
- discusses some new results on the nonnegative rank and the identifiability of NMF, and
- makes available MATLAB codes for readers to run the numerical examples presented in the book.

Graduate students starting to work on NMF and researchers interested in better understanding the NMF problem and how they can use it will find this book useful. It can be used in advanced undergraduate and graduate-level courses on numerical linear algebra and on advanced topics in numerical linear algebra and requires only a basic knowledge of linear algebra and optimization.

Nicolas Gillis is an associate professor in the department of Mathematics and Operational Research at the University of Mons in Belgium. He is a recipient of the Householder Award and an ERC Starting Grant. His research interests include optimization, numerical linear algebra, machine learning, signal processing, and data mining. A member of SIAM and IEEE, he serves as an associate editor of *SIAM Journal on Matrix Analysis and Applications* and *IEEE Transactions on Signal Processing*.

siam.
Society for Industrial and Applied Mathematics
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
+1-215-382-9800 • Fax: +1-215-386-7999
siam@siam.org • *siam.org*

DI02

ISBN: 978-1-611976-40-3
90000
9781611976403

SIAM Textbooks

Nonnegative Matrix Factorization

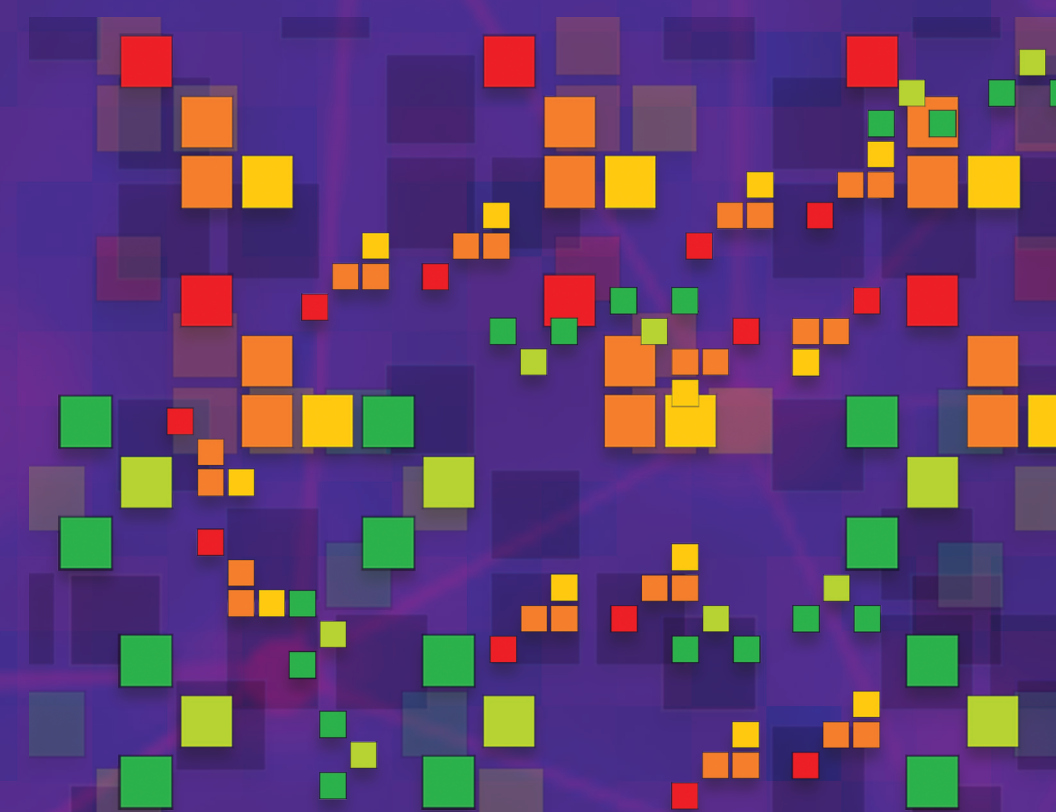
Nicolas Gillis

siam.

DI02

Data Science Book Series

Nonnegative Matrix Factorization



Nicolas Gillis

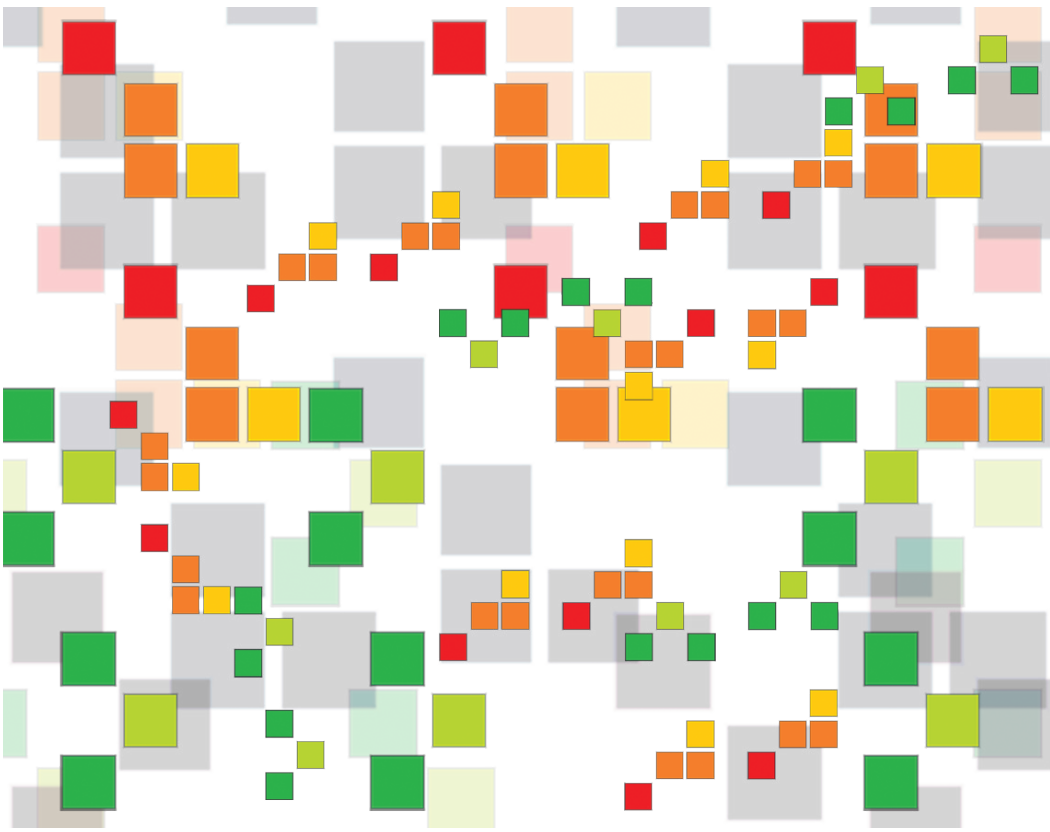
siam.

Nonnegative Matrix Factorization

Data Science Book Series

Daniela Calvetti and Erkki Somersalo, *Mathematics of Data Science: A Computational Approach to Clustering and Classification*
Nicolas Gillis, *Nonnegative Matrix Factorization*

Nonnegative Matrix Factorization



Nicolas Gillis

University of Mons
Mons, Belgium

siam

Society for Industrial and Applied Mathematics
Philadelphia

10 9 8 7 6 5 4 3 2 1

Publications Director	Kivmars H. Bowling
Executive Editor	Elizabeth Greenspan
Developmental Editor	Mellisa Pascale
Managing Editor	Kelly Thomas
Production Editor	Lisa Briggeman
Copy Editor	Susan Fleshman
Production Manager	Donna Witzleben
Production Coordinator	Cally A. Shrader
Compositor	Cheryl Hufnagle
Graphic Designer	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://lccn.loc.gov/2020042037*
LC ebook record available at *https://lccn.loc.gov/2020042038*

Pour Aline, Elinor, et Rose

Contents

Preface

Identifying the underlying structure of a data set and extracting meaningful information is a key problem in data analysis. Simple and powerful methods to achieve this goal are *linear dimensionality reduction* (LDR) techniques, which are equivalent to low-rank matrix approximations (LRMA). Examples of LDR techniques are principal component analysis (PCA), independent component analysis, sparse PCA, robust PCA, low-rank matrix completion, and sparse component analysis. The reason for the success of this type of methods is that, although simple, they are applicable in a wide range of applications such as recommender systems, model-order reduction and system identification, clustering, image analysis, and blind source separation, to cite a few.

Among LRMA techniques, nonnegative matrix factorization (NMF) requires the factors of the low-rank approximation to be componentwise nonnegative. This makes it possible to interpret them meaningfully, for example when they correspond to nonnegative physical quantities. Applications of NMF include extracting parts of faces (such as eyes, noses, and lips) in a set of facial images, identifying topics in a set of documents, learning hidden Markov models, extracting materials and their abundances in hyperspectral images, separating audio sources from their mixture, detecting communities in large networks, analyzing medical images, and decomposing gene expression microarrays.

Aim of the book The aim of this book is to provide a comprehensive account of the most important aspects of the NMF problem:

- Theoretical aspects: the nonnegative rank, the nonuniqueness/identifiability of NMF solutions, the geometric interpretation of NMF, and computational complexity issues.

- Models: choice of the objective function and regularizations, link with well-known techniques such as k-means, and use of additional constraints such as orthogonality or symmetry.

- Algorithms: heuristic algorithms using standard nonlinear optimization schemes such as block coordinate descent methods, and provably correct algorithms under appropriate assumptions.

- Applications: they include image analysis, document classification, hyperspectral unmixing, audio source separation, topic modeling, and community detection.

This book is accessible to a wide audience. In particular it is intended for people who want to know about the workings of NMF. It also aims to give more insights to practitioners so that they can use NMF meaningfully. To read this book, basic knowledge of linear algebra and optimization is needed.

Why is this book important? Although NMF has been studied extensively for the last 20 years, there is currently only one book on the topic, by Cichocki et al. [98] (2009) which is already more than 10 years old. It focuses on iterative algorithms and applications, and many aspects of NMF are not covered in that book—especially since many important results have been obtained in the last 10 years.[1]

The aim of this book is to fill in this gap by providing more insights into the theoretical aspects of NMF. These are key to be able to use NMF effectively and meaningfully in practice. This will allow the reader to make better use of NMF as a computational tool. This book is aimed at researchers who want to understand the NMF problem; for example,

- You do not know (much) NMF and want to discover this problem, why and how it works, and what it can be used for. This book would be ideal for example for a master's or Ph.D. student starting to work on NMF.

- You are using NMF for applications but you would like to understand better its subtly difficult aspects such as its computational complexity, its geometric interpretation, or its nonuniqueness/identifiability issues. Also, you would like to know about the state-of-the-art algorithms.

- You are already rather familiar with NMF but have not yet studied all of its aspects (for example you would like to know more about the nonnegative rank, or the nonuniqueness of NMF solutions). This book will allow you to delve into different aspects of the NMF problem and will give you useful references.

Moreover, this book contains a few new results not present in the literature (as far as I know): bounds on the nonnegative rank under rank-one perturbations (Theorem 3.3), the study of the generic value of the nonnegative rank (Section 3.3.2), the identifiability of orthogonal NMF (Section 4.3.2), and a necessary condition for the sufficiently scattered condition, a crucial notion when studying the uniqueness of NMF solutions (see Theorem 4.28).

MATLAB code All algorithms and numerical experiments presented in this book are available from `bookstore.siam.org/di02/bonus`. When we discuss an algorithm, or display results from a numerical experiment, the corresponding MATLAB file will be indicated using

```
[Matlab file: Name of file]
```

It can be found in the folder of the corresponding chapter. Hence the interested reader can easily find the corresponding MATLAB file. To provide a better view of all the NMF algorithms available with this book, there is an *exception* for NMF algorithms: they can all be found in the folder [algorithms]. For example, the separable NMF algorithms presented in Chapter 7 can be found in the folder [algorithms/separable NMF], although the numerical experiments presented in Chapter 7 can be found in the folder [Chapter 7 - Separable NMF].

All tests in this book are performed using MATLAB R2019b on a laptop Intel Core i7-7500U CPU @2.9GHz, 24GB RAM.

How to use this book The book is organized so that it is possible to read only subsets of the chapters depending on the reader's interests. The book was written so that each chapter is

[1] Such as the identifiability results based on the sufficiently scattered condition (Chapter 4), and the polynomial-time algorithms for separable NMF (Chapter 7).

as self-contained as possible. Each chapter can be thought of as a survey on the corresponding topic. Moreover, each chapter ends with take-home messages that summarize and highlight the important results covered in the chapter. The book is organized as follows.

Chapter 1 serves as an introduction and the problem definition. It contains the description of four key applications of NMF that will be used throughout the book as illustrations. It also contain a historical account of the problem explaining when, why, and how the NMF problem came about. Then the book is divided into two parts.

Part I. *Exact factorizations* The first part considers the Exact NMF problem: given the nonnegative input matrix X and a factorization rank r, find nonnegative factors W with r columns and H with r rows such that $X = WH$. Chapter 2 discusses theoretical aspects of Exact NMF and in particular its geometric interpretation; this is crucial to design algorithms but also to use NMF meaningfully in practice. Section 2.1 on the geometric interpretation of NMF is useful to understand Chapters 4 and 7. The second part of this chapter discusses a more constrained NMF problem where the first factor W is required to have the same rank as the input matrix, which is referred to as restricted Exact NMF, and its link with a geometric problem, namely, the nested polytope problem. The third part of the chapter discusses the computational complexity of these problems. Chapter 3 digs into theoretical aspects of the nonnegative rank (the smallest r such that an Exact NMF exists): its properties, its lower and upper bounds, and its link with extended formulations of polytopes and with communication complexity. Chapter 4 discusses the identifiability issues when using NMF in practice. In fact, NMF decompositions are in general not unique, while most applications are looking for the unique ground truth underlying factors. This chapter explains how to recover the true factors, and under which conditions it is possible. Section 4.2 focuses on uniqueness conditions for the plain NMF model, while Section 4.3 discusses regularized NMF models, namely, orthogonal, separable, minimum-volume, and sparse NMF, that lead to unique decompositions under milder conditions.

Part II. *Approximate factorizations* The second part of the book considers approximate NMF decompositions, where $WH \approx X$, which we refer to as NMF for short as it is the standard in the literature. Chapter 5 discusses several important variants of the NMF model that use additional constraints, regularizations, and different objective functions; for example symmetric NMF which requires $W = H^\top$. As discussed in Chapter 4, considering such variants in practice is key to obtain unique solutions. This chapter also discusses different models that are closely related to NMF such as k-means and probabilistic latent semantic analysis/indexing (PLSA/PLSI). Chapter 6 discusses the computational complexity of NMF, which is NP-hard[2] in general. Chapter 7 considers NMF under the separability assumption, referred to as separable NMF, where the columns of the first factor W can be found among the columns of X. Although it is a strong assumption, it makes sense in several applications. Moreover, it allows us to provably solve NMF efficiently, that is, in polynomial time, and renders the solution unique, hence resolving two key issues of NMF (NP-hardness and identifiability). This chapter presents the main algorithms for separable NMF, discusses their robustness to noise, and compares them on several synthetic data sets. Chapter 8 focuses on iterative heuristic algorithms to compute NMF solutions. The state-of-the-art algorithms for NMF are presented; they are based on standard optimization techniques. We also discuss convergence guarantees and provide some numerical comparisons. Chapter 9 presents three more applications of NMF, as well as pointers to more applications.

[2]NP-hardness of a problem implies that unless P=NP, there exists no algorithm running in a number of operations polynomial in the size of the input that solves the problem.

Disclaimer The book presents the NMF problem from my own perspective and is clearly biased toward my own work and research interests. I apologize for not discussing or referring to many relevant works (that I am either unfamiliar with or unaware of). This book is a summary of my current knowledge about NMF and is by no means comprehensive. Any feedback on the book is more than welcome and is highly encouraged.

Tensors The NMF model can be directly generalized to tensors, in which case it is referred to as nonnegative tensor factorization (NTF) or nonnegative canonical polyadic decomposition (nonnegative CPD). It is out of the scope of this book to discuss this important extension, and we stick to the matrix case. However, one should keep this connection in mind as results from the matrix case can be used in the tensor case. For example, NP-hardness results for NMF (see Chapter 6) directly apply to NTF since NTF is a generalization of NMF, and NMF algorithms based on the block coordinate descent framework described in Chapter 8 directly extend to NTF. We refer the interested reader to [283, 98, 418] and the references therein for more details on tensor factorizations.

Acknowledgments Writing this book would not have been possible without the many fruitful collaborations I have been lucky to have over the years. In particular, I am grateful to my two mentors, François Glineur, who introduced me to NMF (the topic of my master's thesis in 2007), and Steve Vavasis, who welcomed me as a postdoc at the University of Waterloo. I have also been lucky that many enthusiastic researchers joined me in Mons: Arnaud, Punit, Andersen, Jérémy, Valentin, Hien, Junjun, François, Nicolas^2, Tim, Christophe, Pierre, and Maryam.

I am thankful to the persons who gave me feedback on early drafts of this book: Tim Marrinan, Arnaud Vandaele, Ken Ma, Jérémy Cohen, Christophe Kervazo, Pierre De Hanschutter, Le Hien, Andersen Ang, Nicolas Nadisic, Hamza Fawzi, Yaroslav Shitov, Anthony Degleris, Marc Pirlot, Junjun Pan, Fabian Lecron, and the anonymous reviewers. I also thank Kaie Kubjas, Xiao Fu, and Kejun Huang for insightful discussions.

Writing a book on NMF has been in the back of my mind for several years, but I thank Andersen Ang and Jennifer Pestana, who motivated me and pushed me to finally undertake this endeavor.

I would like to thank the SIAM editorial team, who have done a wonderful job improving the book through their reviewing and copyediting. It was a pleasure collaborating with them on this project.

I am grateful for the support from the European Research Council (ERC starting grant 679515, COLORAMAP project), the Fonds de la Recherche Scientifique-FNRS through several grants (Incentive Grant for Scientific Research F.4501.16, and Excellence of Science grant O005318F-RG47 which is also supported by the Fonds Wetenschappelijk Onderzoek-Vlaanderen), and the Franqui Foundation (Francqui research professor).

I thank my family and my friends.

Notation

Sets of scalars, vectors, matrices

$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$	set of nonnegative real numbers
$\mathbb{R}_{++}$	set of positive real numbers
$\mathbb{R}^n$	set of real column vectors of dimension n
$\mathbb{R}^{m\times n}$	set of real m-by-n matrices
$\mathbb{R}^n_+$	set of nonnegative real column vectors of dimension n
$\mathbb{R}^{m\times n}_+$	set of m-by-n nonnegative real matrices
$\mathbb{S}^n$	set of n-by-n symmetric matrices
$\mathbb{S}^n_+$	set of n-by-n positive semidefinite matrices
$\mathbb{S}^n_{++}$	set of n-by-n positive definite matrices
$\mathfrak{C}^n$	set of n-by-n copositive matrices
$\mathfrak{C}^n_+$	set of n-by-n completely positive matrices
$\mathcal{C}$	second-order cone $\left\{x \in \mathbb{R}^r \mid e^\top x \geq \sqrt{r-1}\|x\|_2\right\}$
$\mathcal{C}^*$	second-order cone $\left\{x \in \mathbb{R}^r \mid e^\top x \geq \|x\|_2\right\}$, the dual of $\mathcal{C}$ (p. 113)
Δ^n	unit simplex of dimension n, that is, $\Delta^n = \{x \in \mathbb{R}^n \mid x \geq 0, \sum_{i=1}^n x_i = 1\}$
$\mathcal{S}^n$	convex hull of the unit simplex and the origin, that is, $\mathcal{S}^n = \{x \in \mathbb{R}^n \mid x \geq 0, \sum_{i=1}^n x_i \leq 1\}$

Submatrices, transpose and inverse

x_i of $x(i)$	ith entry of the vector x
$A_{i:}$ or $A(i,:)$	ith row of A
$A_{:j}$ or $A(:,j)$	jth column of A
A_{ij} or $A(i,j)$	entry at position (i,j) of A
$A(I,J)$	submatrix of A with row (resp. column) indices in I (resp. J)
$[A\ B; C\ D]$	We use Matlab notation: $[A\ B; C\ D] = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$
$A^\top$	transpose of the matrix A, $(A^\top)_{ij} = A_{ji}$
A^{-1}	inverse of the square matrix A, $A^{-1}A = AA^{-1} = I$
$A^{-\top}$	inverse of the transpose of the square matrix A, that is, $A^{-\top}A^\top = A^\top A^{-\top} = I$

Special vectors and matrices

0	matrix of zeros of appropriate dimension
$0_{m\times n}$	m-by-n matrix of zeros
I_n	identity matrix of dimension n
I	identity matrix of appropriate dimension
e	vector of all ones of appropriate dimension
e_k	kth unit vector with $e_k(k) = 1$ and $e_k(i) = 0$ for all $i \neq k$, that is, $e_k = I(:, k)$

Norms

$\\|.\\|_1$	ℓ_1 norm, $\\|x\\|_1 = \sum_{i=1}^n \|x_i\|$, $x \in \mathbb{R}^n$ componentwise matrix ℓ_1 norm, $\\|A\\|_1 = \sum_{i,j} \|A_{i,j}\|$, $A \in \mathbb{R}^{m\times n}$
$\\|.\\|_2$	vector ℓ_2 norm, $\\|x\\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$, $x \in \mathbb{R}^n$ matrix ℓ_2 norm, $\\|A\\|_2 = \max_{x\in\mathbb{R}^n, \\|x\\|_2=1} \\|Ax\\|_2$, $A \in \mathbb{R}^{m\times n}$
$\\|.\\|_F$	Frobenius norm, $\\|A\\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}$, $A \in \mathbb{R}^{m\times n}$
$\\|.\\|_\infty$	vector ℓ_∞ norm, $\\|x\\|_\infty = \max_{1\leq i\leq n} \|x_i\|$, $x \in \mathbb{R}^n$ componentwise matrix ℓ_∞ norm, $\\|A\\|_\infty = \max_{i,j} \|A_{i,j}\|$, $A \in \mathbb{R}^{m\times n}$
$\\|.\\|_0$	ℓ_0 "norm," $\\|x\\|_0 = \big\|\{i \| x_i \neq 0\}\big\|$, $x \in \mathbb{R}^m$
$\\|.\\|_{1,q}$	matrix $\ell_{1,q}$ norm, $\\|A\\|_{1,q} = \sum_{i=1}^m \\|A(i,:)\\|_q$, $A \in \mathbb{R}^{m\times n}$

Inequalities

$A \geq 0$	A is a nonnegative matrix, that is, $A(i,j) \geq 0$ for all i, j
$A \geq B$	This means $A - B \geq 0$
$A \succeq 0$	A is a PSD matrix
$A \succeq B$	$A - B$ is a PSD matrix

Functions and sets on matrices

$\langle .,. \rangle$	Euclidean scalar product, that is, for $A, B \in \mathbb{R}^{m\times n}$, $\langle A, B\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ij}$,
$\sigma_i(A)$	ith singular values of matrix A, in nondecreasing order
$\sigma_{\max}(A)$	largest singular value of A, that is, $\sigma_1(A)$
$\sigma_{\min}(A)$	smallest singular value of $A \in \mathbb{R}^{m\times n}$, that is, $\sigma_{\min(m,n)}(A)$
$\kappa(A)$	condition number of A, $\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$
$\det(A)$	determinant of A
$\mathrm{tr}(A)$	trace of A, that is, sum of its diagonal entries
$\mathrm{diag}(.)$	For $a \in \mathbb{R}^n$, $A = \mathrm{diag}(a) \in \mathbb{R}^{n\times n}$ is a diagonal matrix such that $A_{ii} = a_i$ for all i For $A \in \mathbb{R}^{n\times n}$, $a = \mathrm{diag}(A) \in \mathbb{R}^n$ is the vector containing the diagonal entries of A
$\overline{A}$	$\overline{A}$ is the matrix A without the last row
$\mathrm{rank}(A)$	rank of A
$\mathrm{rank}_+(A)$	nonnegative rank of A (p. 55)
$\mathrm{rank}_+^*(A)$	restricted nonnegative rank of A (p. 35)
cp-rank(A)	completely positive rank of A (p. 79)
k-rank(A)	Kruskal rank of A (p. 151)
$\mathrm{cone}(A)$	conical hull of the columns of A (p. 20)
$\mathrm{conv}(A)$	convex hull of the columns of A (p. 21)
$\mathrm{col}(A)$	column space of A (p. 21)
$\mathrm{aff}(A)$	affine hull of the columns of A (p. 21)
$A \circ B$	componentwise multiplication between A and B, that is, $(A \circ B)_{ij} = A_{ij}B_{ij}$
$\frac{[A]}{[B]}$	componentwise division between A and B, $\left(\frac{[A]}{[B]}\right)_{ij} = \frac{A_{ij}}{B_{ij}}$
$A^{\circ a}$	componentwise exponent of matrix A by the scalar a, that is, $A^{\circ a}(i,j) = A(i,j)^a$ for all i,j
$\mathrm{supp}(A)$	support (index set of nonzero entries) of matrix A, that is, $\mathrm{supp}(A) = \{(i,j) \mid A(i,j) \neq 0\}$
$\omega(A)$	fooling set bound for A (p. 64)
$\mathrm{bin}(A)$	$\mathrm{bin}(A)$ is the binarization of A, that is, the nonzero entries of A are set to 1 (p. 70)
$\mathrm{rank}_{01}(A)$	this is the Boolean rank of $\mathrm{bin}(A)$ (p. 70)
$\mathrm{rc}(A)$	rectangle covering bound for A (p. 72)
$\mathrm{rrc}(A)$	refined rectangle covering bound for A (p. 73)

Miscellaneous

$a{:}b$	set $\{a, a+1, \ldots, b-1, b\}$ (for a and b integers with $a \leq b$)
$[a, b]$	closed interval for reals $a \leq b$
(a, b)	open interval for reals $a \leq b$
∇f	gradient of the function f
$\nabla^2 f$	Hessian of the function f
$\lceil . \rceil$	$\lceil x \rceil$ is the smallest integer greater or equal to $x \in \mathbb{R}$
$\lfloor . \rfloor$	$\lfloor x \rfloor$ is the largest integer smaller or equal to $x \in \mathbb{R}$
$\backslash$	subtraction of two sets, that is, $R \backslash S$ is the set of elements that are in R but not in S
$\lvert . \rvert$	cardinality of a set, $\lvert S \rvert$ is the number of elements in S
k-sparse	the vector x is k-sparse if it has k nonzero entries, that is, $\lvert \operatorname{supp}(x) \rvert = k$
$\mathbb{P}(x)$	probability of the event $X = x$
$\mathbb{E}(X)$	expected value of a random variable X
$\mathcal{N}(\mu, \sigma)$	normal distribution of mean μ and standard deviation σ
$\mathcal{U}(a, b)$	uniform distribution in the interval $[a, b]$
$f(x) = \mathcal{O}(g(x))$	Big $\mathcal{O}$ notation: there exists K and x_0 such that $f(x) \leq K g(x)$ for all $x \geq x_0$
$f(x) = o(g(x))$	small o notation: $\lim_{x \to +\infty} \frac{f(x)}{g(x)} = 0$
$f(x) = \Omega(g(x))$	Big Omega notation, equivalent to $g(x) = \mathcal{O}(f(x))$
$f(x) = \Theta(g(x))$	Big Theta notation, equivalent to $f(x) = \mathcal{O}(g(x))$ and $f(x) = \Omega(g(x))$
$\min_{x \in \mathcal{X}} f(x)$	minimum value of $f(x)$ over the feasible set $\mathcal{X}$
$\operatorname{argmin}_{x \in \mathcal{X}} f(x)$	set of minimizers of $f(x)$ over the feasible set $\mathcal{X}$

Abbreviations

w.l.o.g.	without loss of generality
i.i.d.	independently and identically distributed
w.r.t.	with respect to

Acronyms

Page number indicates where the acronym is first defined.

2-BCD	two-block coordinate descent (p. 261)
ADMM	alternating direction method of multipliers (p. 289)
ANLS	alternating nonnegative least squares (p. 281)
BCD	block coordinate descent (p. 266)
BSUM	block successive upper-bound minimization (p. 268)
EDM	Euclidean distance matrix (p. 64)
Exact NMF	exact nonnegative matrix factorization (p. 19)
FAW	fast anchor words (p. 230)
FPGM	fast projected gradient method (p. 288)
HALS	hierarchical alternating least squares (p. 283)

Acronyms (continued)

HU	hyperspectral unmixing (p. 7)
IS	Itakura–Saito (p. 162)
KKT	Karush–Kuhn–Tucker (p. 264)
KL	Kullback–Leibler (p. 161)
k-sparse MF	k-sparse matrix factorization (p. 151)
LDR	linear dimensionality reduction (p. 1)
LP	linear programming (p. 249)
LRMA	low-rank matrix approximation (p. 1)
min-vol	minimum-volume (p. 138)
MLP	multiple linear programs (p. 249)
MM	majorization-minimization (p. 265)
MVE	minimum-volume ellipsoid (p. 237)
MU	multiplicative updates (p. 270)
NMF	nonnegative matrix factorization (p. 4)
NMU	nonnegative matrix underapproximation (p. 176)
NNLS	nonnegative least squares (p. 280)
NPP	nested polytope problem (p. 36)
ONMF	orthogonal NMF (p. 136)
PCA	principal component analysis (p. 2)
PGM	projected gradient method (p. 287)
PLSA	probabilistic latent semantic analysis (p. 189)
PLSI	probabilistic latent semantic indexing (p. 189)
PMF	positive matrix factorization (p. 4)
PPI	pure-pixel index (p. 247)
PSD	positive semidefinite (p. 79)
RE-NMF	restricted Exact NMF (p. 35)
SD-LP	self-dictionary via linear programming (p. 254)
SMCR	self-modeling curve resolution (p. 12)
SNPA	successive nonnegative projection algorithm (p. 232)
SPA	successive projection algorithm (p. 223)
SSC	sufficiently scattered condition (p. 104)
SSM	stochastic sequential machine (p. 13)
SVD	singular value decomposition (pp. 2, 196)
symNMF	symmetric NMF (p. 178)
tri-NMF	nonnegative matrix trifactorization (p. 181)
tri-symNMF	symmetric nonnegative matrix trifactorization (p. 181)
TSP	traveling salesman problem (p. 15)
VCA	vertex component analysis (p. 232)
WLRA	weighted low-rank matrix approximation (p. 3)

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we elaborate on the main reason why nonnegative matrix factorization (NMF) became a popular and standard tool in data analysis, namely because NMF is an easily interpretable linear dimensionality reduction technique for nonnegative data (Section 1.1). After providing a formal definition of NMF, the notation, and the terminology (Section 1.2), we illustrate the ability of NMF to extract meaningful components in nonnegative data sets with four applications: feature extraction in images, hyperspectral unmixing, text mining, and audio source separation[3] (Section 1.3). The last part of the chapter provides a historical overview of how NMF came about (Section 1.4). In each chapter of this book, we conclude with some take-home messages (Section 1.5).

1.1 ▪ Linear dimensionality reduction techniques for data analysis

Most works on NMF are motivated by its applicability in data analysis, more precisely, by the capability of NMF to automatically extract meaningful information in a data set. Extracting the underlying structure within data sets is one of the central problems in data science, and numerous techniques exist to perform this task. One of the oldest approaches is linear dimensionality reduction (LDR). LDR represents each data point as a linear combination of a small number of basis elements. Mathematically, given a data set of n data points $x_j \in \mathbb{R}^m$ $(1 \leq j \leq n)$, LDR looks for a small number r of basis vectors $w_k \in \mathbb{R}^m$ $(1 \leq k \leq r)$ such that each data point is well-approximated by a linear combination of these basis vectors, that is,

$$x_j \approx \sum_{k=1}^{r} w_k h_{kj} \quad \text{for all } j,$$

where the h_{kj}'s are scalars; see Figure 1.1 for an illustration in three dimensions with $m = 3$, $r = 2$, and $n = 50$.

LDR is equivalent to low-rank matrix approximation (LRMA):

$$\underbrace{[x_1, x_2, \dots, x_n]}_{X \in \mathbb{R}^{m \times n}} \approx \underbrace{[w_1, w_2, \dots, w_r]}_{W \in \mathbb{R}^{m \times r}} \underbrace{[h_1, h_2, \dots, h_n]}_{H \in \mathbb{R}^{r \times n}},$$

[3] Sections 1.1 and 1.3 follow closely the introductions from [189, 190, 191].

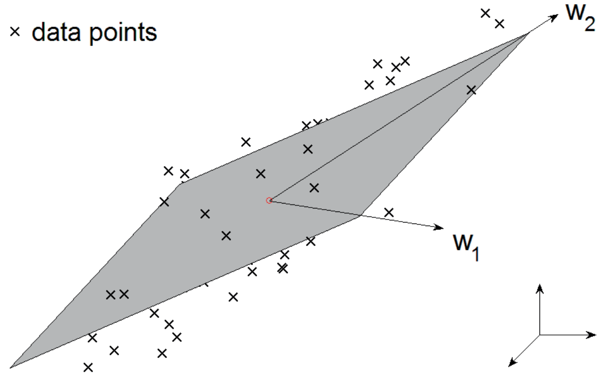


Figure 1.1. *Illustration of LDR: approximation of three-dimensional data points with a two-dimensional subspace generated by* w_1 *and* w_2.

where

- each column of the matrix $X \in \mathbb{R}^{m\times n}$ is a data point, that is, $X(:,j) = x_j$ for $1 \leq j \leq n$;
- each column of the matrix $W \in \mathbb{R}^{m\times r}$ is a basis element, that is, $W(:,k) = w_k$ for $1 \leq k \leq r$; and
- each column of $H \in \mathbb{R}^{r\times n}$ contains the coordinates of a data point $X(:,j)$ in the basis W, that is, $H(:,j) = h_j$ for $1 \leq j \leq n$.

Hence LDR provides a rank-r approximation WH of X, and each data point is mapped into the basis W using the corresponding column of H:

$$x_j \approx W h_j \quad \text{for all } j.$$

Typically, the number of basis vectors is much smaller than the ambient dimension m and the number of data points n, that is, $r \ll \min(m,n)$. This allows LDR and LRMA to compress the data, which is achieved for $r < \frac{mn}{m+n}$ since X contains mn entries, while W and H contain only $mr + nr$ entries.[4] Note that if the input matrix is sparse, with $\mathrm{nnz}(X)$ nonzero entries, compression requires $r < \frac{\mathrm{nnz}(X)}{m+n}$.

In order to compute W and H given X and r, one needs to define an error measure. For example, when the solution (W,H) minimizes the sum of the squares of the entries of the residual $X-WH$, that is, the squared Frobenius norm of the residual $\|X-WH\|_F^2 = \sum_{i,j}(X-WH)_{ij}^2$, LRMA is equivalent to principal component analysis (PCA) [266] which can be computed via the singular value decomposition (SVD) [216]. LRMA is a workhorse in numerical linear algebra, with the SVD being a central technique, and is closely related to the eigenvalue decomposition and factorizations such as Cholesky, QR, and LU, to cite a few. LRMA is at the heart of many fields in applied mathematics and computer science, for example in

- statistics, data analysis, and machine learning to perform regression, prediction, clustering, classification, and noise filtering [266, 144];
- numerical linear algebra to solve linear systems of equations [216];

[4]The factorization WH only has $(m+n-1)r$ degrees of freedom due to the scaling degree of freedom of the columns of W and rows of H, that is, $W(:,k)H(k,:) = (\alpha W(:,k))(H(k,:)/\alpha)$ for any $\alpha \neq 0$ and any k; see Chapter 4.

- signal and image processing to perform blind source separation [104, 87];
- graph theory to cluster the vertices of a graph [94];
- optimization to gain computational efficiency [68]; and
- systems theory and control to perform model-order reduction and system identification [341].

Although PCA has been around for more than 100 years, LRMA models have gained momentum in the last 20 years. The reason is mostly twofold: (i) data analysis has become more and more important due to the recent deluge of data in many different areas (the big data era), and (ii) despite a simple model, LRMA is very powerful since many high-dimensional data sets are well-approximated by low-rank matrices [458]. LRMA models are used to compress the data, filter the noise, reduce the computational effort for further manipulation of the data, or to directly identify hidden structure in a data set; see for example the survey [457]. Many variants of LRMA have been developed over the last few years [422]. They differ in two key aspects:

1. The error measure can vary and should be chosen depending on the noise statistic assumed on the data. For example, PCA uses least squares, that is, it minimizes $\|X - WH\|_F^2$, which implicitly assumes independently and identically distributed (i.i.d.) Gaussian noise.

   If data is missing or if weights are assigned to the entries of X (for example because the noise is not identically distributed over the entries of X), the problem can be cast as a weighted low-rank matrix approximation (WLRA) problem that minimizes $\sum_{i,j} P_{i,j}(X - WH)_{i,j}^2$ for some nonnegative weight matrix P, where $P_{i,j} = 0$ when the entry at position (i, j) is missing [434]. Note that if P contains entries only in $\{0, 1\}$, then the problem is also referred to as PCA with missing data or low-rank matrix completion with noise. WLRA is widely used in recommender systems [32, 287] for predicting the preferences of users for a given product based on the product's attributes and user preferences, such as in the Netflix prize competition; see [35] and Section 9.5.

   If the sum of the absolute values of the entries of the error $\sum_{i,j} |X - WH|_{i,j}$ is used as the error measure, we obtain yet another variant more tolerant to outliers; this is closely related to robust PCA [73, 212]. It can be used, for example, for background subtraction in video sequences where the noise (the moving objects) is assumed to be sparse while the background has low rank (the background does not change much between consecutive images in video sequences).

2. Different constraints can be imposed on the factors W and H. These constraints depend on the application at hand and allow, for example, for meaningful interpretation of the factors.

   For example, the k-means problem, which is the problem of finding a set of r centroids w_k ($1 \leq k \leq r$) such that the sum of the distances between each data point and the closest centroid is minimized, is equivalent to LRMA where the factor H is required to have a single nonzero entry in each column that is equal to one, so that the columns of W are the cluster centroids.

   If instead one wants to explain each data point using as few basis vectors as possible, each column of the matrix H should contain as many zero entries as possible. This LRMA variant is referred to as sparse component analysis [221] and is closely related to dictionary learning [5, 455] and sparse PCA [115]. It yields a more compact and easily interpretable decomposition.

NMF, an LDR technique for nonnegative data Among LRMA models, nonnegative matrix factorization (NMF) requires the factor matrices W and H to be componentwise nonnegative, which we denote $W \geq 0$ and $H \geq 0$. These nonnegativity constraints play an instrumental role in various applications as they allow one to extract meaningful and interpretable components in nonnegative data sets. For example, in some applications, the entries in W and H can be interpreted as physical quantities. Before presenting four such applications in Section 1.3, let us first define the NMF problem rigorously.

1.2 ▪ Problem definition

Let us formally define the NMF problem and discuss some of its aspects.

Problem 1.1 (Nonnegative matrix factorization). *Given a nonnegative matrix $X \in \mathbb{R}^{m \times n}_+$, a factorization rank r, and a distance measure $D(.,.)$ between two matrices, compute two nonnegative matrices $W \in \mathbb{R}^{m \times r}_+$ and $H \in \mathbb{R}^{r \times n}_+$ such that $D(X, WH)$ is minimized, that is, solve*

$$\min_{W \in \mathbb{R}^{m \times r}_+, H \in \mathbb{R}^{r \times n}_+} D(X, WH). \tag{1.1}$$

Terminology NMF in its modern form was first referred to as positive matrix factorization (PMF) by Paatero and Tapper [371] in 1994, but this name has not been used much, most likely because "positive" means strictly larger than zero, while NMF usually generates a factor with many zeros entries; see the discussion below in the paragraph Sparsity. The name NMF became standard after the paper of Lee and Seung published in 1999 [303]; see Section 1.4 for the historical overview of NMF. In most data analysis applications (see Section 1.3 and Chapter 9), the solution (W, H) is only an approximation of the data matrix X; hence $X \neq WH$ as with most applications using LRMA mentioned in Section 1.1. This is due to the presence of noise and the linear model being, in most cases, only an approximate model ("All models are wrong but some are useful" as mentioned by George Box[5] in 1976). Therefore, the use of the term "factorization" might be misleading since it usually refers to an exact decomposition $X = WH$. Hence some authors have argued that it would, for example, make more sense to refer to (1.1) as nonnegative matrix approximation [433]. However, in this book we subscribe to the widely accepted standard that the name NMF refers to the associated approximation problem, and we will further specify when we are considering the exact factorization, $WH = X$, by calling it Exact NMF. Exact NMF is important in linear algebra as it allows us to compute the nonnegative rank of a matrix X, denoted $\text{rank}_+(X)$, which is the smallest r such that an Exact NMF of X exists; see Chapters 2 and 3.

Objective function The objective function of the NMF problem is defined as

$$D\colon \mathbb{R}^{m \times n}_+ \times \mathbb{R}^{m \times n}_+ \mapsto \mathbb{R}_+ \text{ given by } (A, B) \mapsto D(A, B)$$

and is also referred to as the error measure. The choice of this function is crucial when designing LRMA models and often depends on the assumptions made about the noise statistics. It may greatly influence the solution (W, H) and leads to rather different optimization problems; see

[5] `https://en.wikipedia.org/wiki/All_models_are_wrong` (consulted May 27, 2020).

Chapter 8. Most error measures give the same importance to each entry of the data matrix X and hence have the form

$$D(A, B) = \sum_{i,j} d(A_{ij}, B_{ij})$$

for some function $d\colon \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that $d(x, y) = 0$ if and only if $x = y$. A standard choice is $d(x, y) = (x - y)^2$ which leads to $D(X, WH) = \|X - WH\|_F^2$. There are two main reasons for this choice: (1) it corresponds to the assumption of i.i.d. Gaussian noise which is reasonable for many data sets, and (2) it leads to a smooth optimization, which is easier to handle (see Chapter 8 for a discussion). We refer the reader to Section 5.1 for a discussion on the choice of D in the context of NMF.

Choice of the symbols In the linear algebra community, authors consistently use the symbols $A = U\Sigma V^\top$ to represent the SVD of matrix A. However, in the NMF literature, there is no consensus on the choice of the symbols used for the data matrix X and the factor matrices (W, H), and many combinations of symbols exist; examples include $V \approx WH$ [303], $X \approx CS^\top$ [263], $A \approx BC$ [433], $X \approx UV$ [478], $Y \approx AX$ [98], $X \approx WH$ [189], or $X \approx WH^\top$ [170]. In this book we choose the notation $X \approx WH$.

Transpose: $\boldsymbol{WH}$ vs. $\boldsymbol{WH}^\top$ In the numerical linear algebra community, most authors would likely prefer the use of $X \approx WH^\top$, similarly as for the SVD that uses $U\Sigma V^\top$ (see Section 6.1.1), as it preserves the symmetry by transposition, that is, $X \approx WH^\top$ if and only if $X^\top \approx HW^\top$. However, we choose $X \approx WH$ for the following reason. When interpreting NMF as an LDR technique, which is the main motivation behind NMF, the columns of H play the role of the coefficients of the columns of X in the subspace spanned by the columns of W since $X(:, j) \approx WH(:, j)$ for all j. In other words, there is a one-to-one correspondence between the high-dimensional columns of X and their low-dimensional representations as the columns of H. We believe this is the reason why it is the most common choice in the NMF literature.

NMF variants The problem (1.1) is the formulation of the standard NMF problem. However, it is important to keep in mind that there exist many variants of this problem. Moreover, as we will argue in Chapter 4, it is in general crucial to consider such variants in practice to obtain unique decompositions and be able to identify the true underlying factors that generated the data. Some variants use regularization in order to obtain solutions with some structure, such as sparsity; see Sections 4.3.4 and 5.3. Other variants use additional constraints; for example symmetric NMF (symNMF) requires $H = W^\top$, and orthogonal NMF (ONMF) requires H to have orthogonal rows (see Section 5.4).

Sparsity Because of the nonnegativity constraints, NMF solutions (W, H) are expected to contain zero entries and hence to naturally have some degree of sparsity; see for example Figures 1.2, 1.4, 1.6, and 1.8 in the next section. Mathematically, this is explained by the first-order optimality conditions of a smooth optimization problem with nonnegativity constraints

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{such that} \quad x \geq 0,$$

which are given by

$$x \geq 0, \quad \nabla f(x) \geq 0, \quad \text{and} \quad x_i(\nabla f(x))_i = 0 \text{ for all } i.$$

This enforces $x_i = 0$ whenever $(\nabla f(x))_i > 0$; see Section 8.1.2 for more details. In some applications, one might need to obtain even sparser solutions, which requires the use of additional constraints or regularization; see Chapters 4 and 5.

Applications In the next section, we describe four important applications of NMF in data analysis (see Chapter 9 for other NMF applications). However, it is important to stress that *NMF is not motivated only by its use as an LDR technique for data analysis*; see in particular the historical overview in Section 1.4. Another important motivation is exact factorizations. In particular, the nonnegative rank of a nonnegative matrix X, denoted $\mathrm{rank}_+(X)$, is the smallest r such that there exists an Exact NMF $X = WH$ where W has r columns and H has r rows; see Chapters 2 and 3 for more details. An application where the nonnegative rank has had tremendous impact is in the study of the extension complexity of polytopes; see Section 3.6.

1.3 ▪ Four applications of NMF in data analysis

In this section, we describe four important applications of NMF that will be used throughout the book as illustrative examples. They show that NMF is a particularly meaningful LRMA model as it leads to interpretable decompositions. In Chapter 9, we review other applications of NMF.

1.3.1 ▪ Feature extraction in a set of images

Given a set of gray-scale images of the same dimensions, let us construct the matrix X such that each column of X corresponds to a vectorized gray-level image. Vectorization means that the two-dimensional images are transformed into a long one-dimensional vector, for example, by stacking the columns of the image on top of each other.[6] This means that each row of X corresponds to the same pixel location among the images. The entry of X at position (i, j), that is, $X(i, j)$, is equal to the intensity of the ith pixel within the jth image, which is nonnegative. As explained in Section 1.1, factorizing X with NMF as $X \approx WH$ where $W \geq 0$ and $H \geq 0$ provides an LDR where the columns of W form a basis for the columns of X. Because W is nonnegative, its columns can be interpreted as basis images: the columns of W are vectors of pixel intensities whose linear combinations allow us to approximate each input image. Moreover, because of the nonnegativity constraints on the weight matrix H, no cancellation is possible between the basis images to reconstruct all the input images. Hence these basis images typically correspond to localized features that are shared among the input images, and the entries of H indicate which input image contains which feature. For example, if the columns of X are facial images, the columns of W correspond to facial features such as eyes, noses, mustaches, and lips. Figure 1.2 illustrates such a decomposition and shows that NMF is able to extract a part-based representation of a set of facial images. Note that to obtain such decompositions, the images need to be well-registered/aligned (for example, pixels corresponding to noses should be located at the same position in the input images).

Another example is the synthetic swimmer data set [138], where the columns of X are vectorized images of a swimmer whose four limbs take four different positions for a total of $n = 256$ images; see Figure 1.3. For an NMF of rank 17, the columns of W correspond to the body and the limbs in the 16 different positions; see Figure 1.4.

Remark 1.1 (Uniqueness, minimality, and sparsity). *The decomposition shown in Figure 1.4 is not minimal, that is, there exists an Exact NMF with fewer basis elements. In particular, there exist several Exact NMFs with factorization rank 16: for example the body can be put together with the limbs in all positions with an intensity of 1/4 (since each image in the data set contains four limbs), or the body can be put together with one limb in its four positions with an intensity of 1 (since each image in the data set contains each limb in one position). Hence the Exact NMF of rank 16 of this data set is not unique; see Chapter 4 for a discussion*

[6] In MATLAB, this is achieved via the function `vec`.

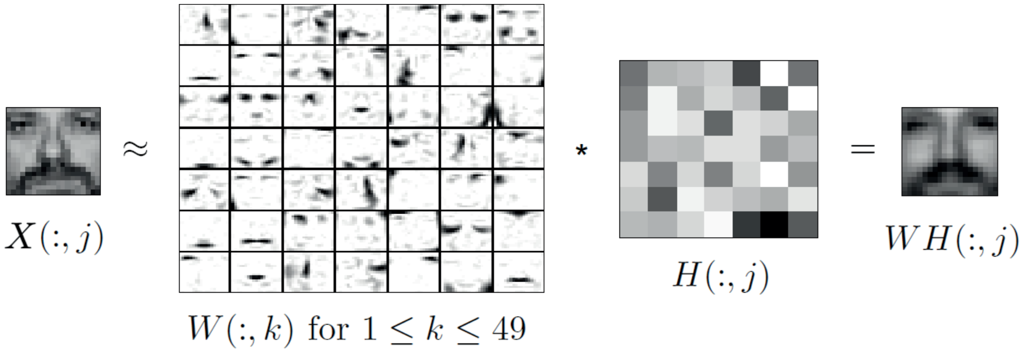


Figure 1.2. *NMF applied on the CBCL face data set with $r = 49$ (2429 images with 19×19 pixels each), as in the seminal paper of Lee and Seung [303]. On the left is a column of X reshaped as an image. In the middle are the 49 columns of the basis W reshaped as images and displayed in a 7×7 grid and the reshaped column of H in the same 7×7 grid that shows which features are present in that particular face. On the right is the reshaped approximation $WH(:,j)$ of $X(:,j)$ as an image. [`Matlab file: CBCL.m`].*

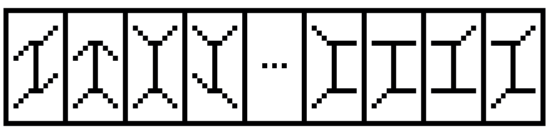


Figure 1.3. *Sample images of the swimmer data set.*

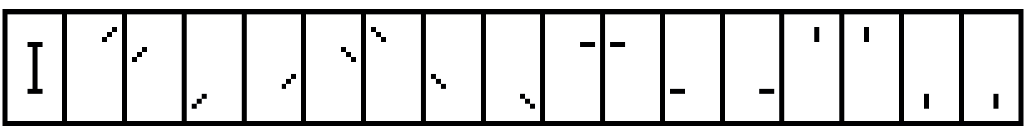


Figure 1.4. *NMF basis images for the swimmer data set. [`Matlab file: Swimmer.m`].*

on this important issue. However, additional constraints on the decomposition may make it unique. For example, the decomposition of rank 17 from Figure 1.4 can be obtained as the unique decomposition using sparse NMF; see Section 4.3.4. It can also be obtained with separable NMF (Chapter 7), minimum-volume NMF (Section 4.3.3), ONMF (Section 4.3.2), or nonnegative matrix underapproximation (NMU; Section 5.4.5).

1.3.2 ▪ Blind hyperspectral unmixing

A hyperspectral image measures the intensity of the light within a scene for many different wavelengths, typically between 100 and 200 wavelengths; see for example [427] for a gentle introduction to hyperspectral imaging. Hence, for each pixel, a vector of intensities is recorded that is equal to the fraction of light reflected by that pixel depending on the wavelength; this is referred to as the spectral signature of the pixel. Given a hyperspectral image, the goal of blind hyperspectral unmixing (blind HU) is to recover the materials present in an image, referred to as the endmembers, and their proportions in each pixel, referred to as abundances. Under the

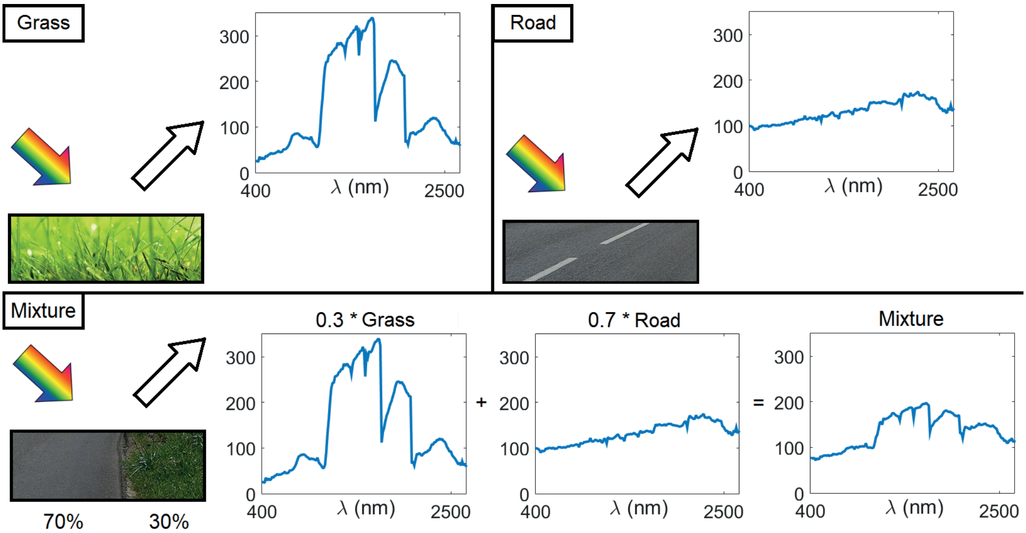


Figure 1.5. *Linear mixing model for hyperspectral imaging.*

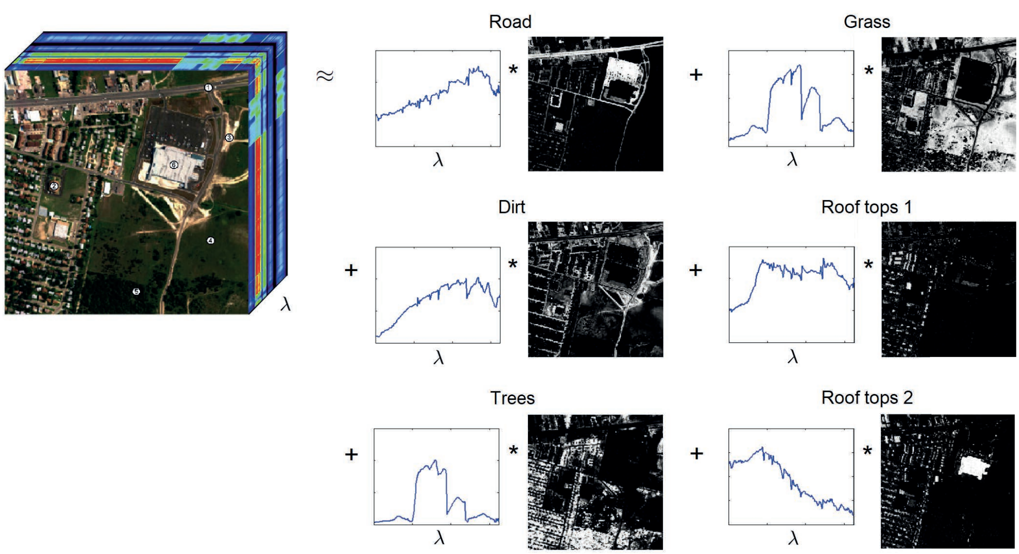


Figure 1.6. *Blind HU of an urban image taken above the Walmart in Copperas Cove, Texas, using NMF with $r = 6$ (162 spectral bands, 307 × 307 pixels). This is the so-called Urban data set. Each factor corresponds to the spectral signature of an endmember (a column of W) with its abundance map (a row of H, reshaped as an image on the figure), where light tones represent high abundances. The Urban hyperspectral image is mostly made up of six materials, namely road, grass, trees, dirt, and roof tops 1 and 2. In particular, the rank-6 NMF above explains more than 95% of the data, that is, $\|X - WH\|_F \leq 0.05\|X\|_F$. Image adapted from [191].* `[Matlab file: Urban.m]`.

linear mixing model, the spectral signature of a pixel is equal to the linear combination of the spectral signatures of the materials it contains. For example, if a pixel contains 30% grass and 70% road surface, its spectral signature is equal to 0.3 times the spectral signature of the grass plus 0.7 times the spectral signature of the road surface; see Figure 1.5 for an illustration. Hence letting each column of the matrix X be the spectral signature of a pixel, blind HU boils down to the NMF of matrix X. In fact, NMF decomposes X as

$$X(:,j) \approx \sum_{k=1}^{r} W(:,k)H(k,j),$$

where the columns of W are the spectral signatures of the endmembers, while the entries of H give the abundance of each endmember in each pixel; see Figure 1.6 for an illustration of NMF on the widely used Urban data set.

1.3.3 ▪ Text mining: topic recovery and document classification

Let each column of the matrix X correspond to a document, that is, a nonnegative vector of word counts. For example, the entry of X at position (i,j) can be the number of times word i appears in document j. The matrix X can also be constructed in different, more sophisticated ways, for example, with the term frequency-inverse document frequency (tf-idf) [389]. This is the so-called bag of words model where the positions of the words in a document are not taken into account. The NMF of X provides the model

$$X(:,j) \approx \sum_{k=1}^{r} W(:,k)H(k,j),$$

where the nonnegative columns of W can also be interpreted as bags of words, that is, as vectors of word count. Since the number of columns of W is much smaller than the number of columns of X ($r \ll n$), each column of W must be used to reconstruct many documents. Moreover, because of the nonnegativity of H, no cancellation is possible, and hence each column of W must contain words that appear simultaneously in these documents. In practice, it has been observed that the columns of W correspond to different topics; see Figure 1.7 for an illustration. Moreover, the columns of the factor H indicate the importance of the topics discussed in the corresponding documents.

Remark 1.2. *We will provide more details on topic modeling and its link with NMF later in the book. In Section 5.5.4, we will show the equivalence between NMF and probabilistic latent semantic analysis/indexing (PLSA/PLSI) which is a simple probabilistic topic model. In Section 5.4.9.1, we will discuss the limitations of NMF for topic modeling and describe a better suited NMF model for this task.*

1.3.4 ▪ Audio source separation

Given an audio signal recorded from a single microphone, its magnitude spectrogram can be constructed as follows. The signal is split into small time frames with some overlap (usually 50%), and each frame is multiplied by some window function (such as the Hamming window) to avoid artifacts due to the truncation of the signal. Each column of X is obtained by taking the short-time Fourier transform of each time frame. More precisely, the entry of X at position (i,j) is the magnitude of the Fourier coefficient for the jth time frame at the ith frequency. Given such

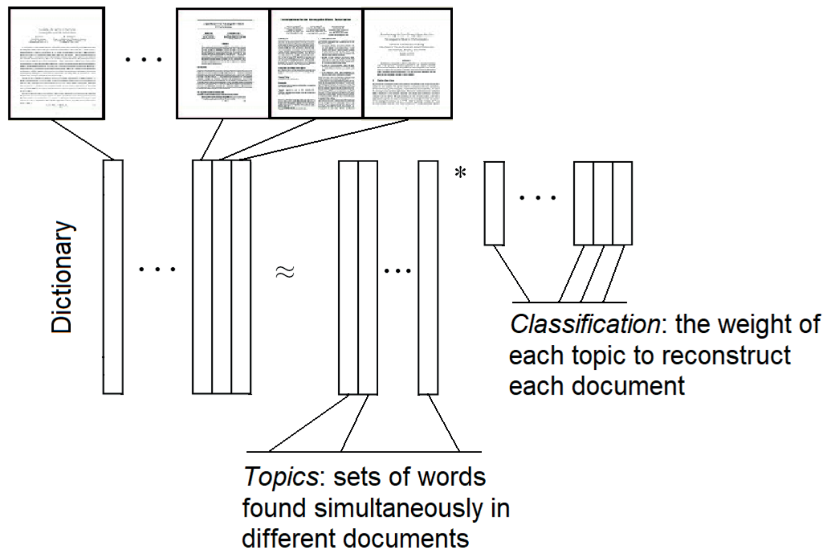


Figure 1.7. *Illustration of NMF for text mining: extraction of topics, and classification of each document with respect to these topics. The columns of X are sets of words, with $X(i, j)$ being the importance of word i in document j. Because H is nonnegative, these sets are approximated as the union of a smaller number of sets defined by the columns of W that correspond to topics.*

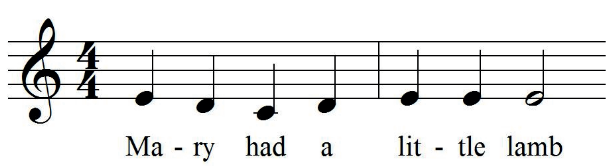
a signal, the goal is to blindly separate the sources that compose the signal, for example, separate the voice and the instruments in a song. Under the two following assumptions, this separation problem is yet another NMF problem:

1. The spectrogram of the mixture is a nonnegative linear combination of the spectrogram of the sources. Nonnegativity means that sound cancellation is neglected. Linearity is a natural model; nonlinear effects, such as the saturation of the microphone or reverberations of the sound, are not taken into account.

2. The spectrograms of the sources have low rank. This has been validated on many experiments and makes sense physically. For example, the spectrogram of an instrument is composed of rank-one factors made of the signature of each note in the frequency domain along with its activation over time.

We refer the interested reader to [158] and the references therein for more information on this model.

Let us use a simple monophonic signal for illustrative purposes, namely a piano recording of "Mary Had a Little Lamb," whose musical score is shown below.



The sequence is composed of three notes, C_4, D_4, and E_4, that activate as follows: E_4, D_4, C_4, D_4, E_4, E_4, E_4. For this data set, the three main sources are the piano notes whose spectrograms

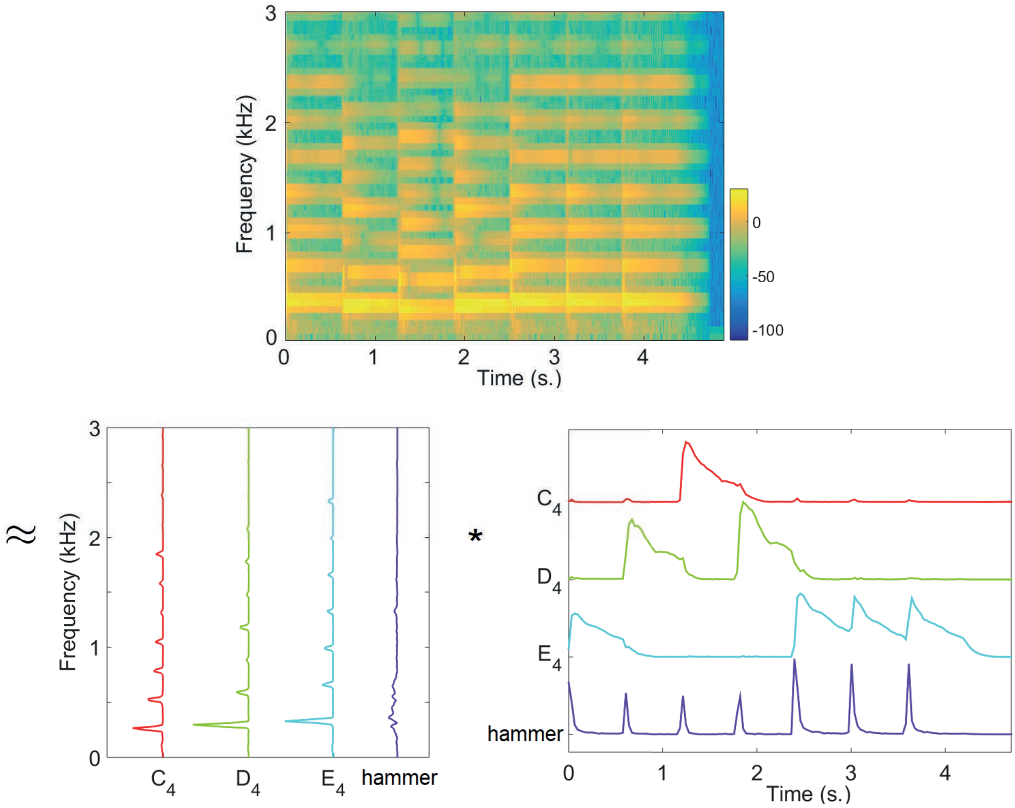


Figure 1.8. *Decomposition of the piano recording "Mary Had a Little Lamb" using NMF: (top) amplitude spectrogram X in dB; (bottom left) basis matrix W corresponding to the three notes C_4, D_4, and E_4, and the hammer noise; (bottom right) activation matrix H that indicates when each note is active. Image adapted from [191].* `[Matlab file: Mary_piano.m]`.

have rank one: each column of W is the signature of each note in the frequency domain, while the entries of H indicate when a note is active. However, there is a fourth source in this data set: it captures the common mechanism that triggers a note; in particular the hammer within the piano. Figure 1.8 shows the NMF decomposition of the magnitude spectrogram using $r = 4$. As expected, the three notes and the hammer noise are extracted as the columns of W, while the rows of H provide the activation of each note in the time domain. NMF is able to blindly separate the different sources and identify which source is active at which moment in time. NMF has been used, for example, for automatic music transcription; see the survey [34].

NMF shows its full potential for polyphonic music analysis when several notes, and even several instruments, are played at once. However, for such complicated scenarios, refined NMF models should be used, such as sparse NMF [471] or convolutive NMF [424]; see Chapter 5. Moreover sources usually have rank higher than one (for example, the voice of a person), and a postprocessing step is necessary to assign each rank-one factor to its corresponding source. Note that the phase information is missing when reconstructing the sound signal from the spectrogram of the sources. This is an important issue, and in practice one often uses the phase of the input signal; see [337, 468, 338] for more information on this topic.

1.4 ▪ History

The first use of the NMF model can be traced back to the fields of analytical chemistry and of geoscience and remote sensing. In both these fields, NMF corresponds to a meaningful physical model. In both cases,

- the columns of the matrix X correspond to nonnegative spectra of samples (also known as the spectral signature),
- the columns of W correspond to the pure component spectra, and
- the columns of matrix H provide the concentration of the pure components within each sample.

This linear mixing model makes sense physically as the spectrum of a sample is, in ideal conditions, proportional to the spectra of the pure components it contains; this is the so-called Lambert–Beer law [297, 30]; see Figure 1.5 (page 8). This model, which is equivalent to NMF, was discovered and used independently in the early 1960s in these two fields.

1.4.1 ▪ Analytical chemistry

The columns of $X \in \mathbb{R}^{m\times n}_+$ are the spectra of a chemical reaction measured over time, so that m is the number of measured spectral bands and n is the number of time steps for which the reaction is measured. These spectra can be obtained in different ways, for example using Raman spectroscopy; see Section 9.3 for a numerical example. Given these nonnegative spectra, the goal is to recover the pure spectra of the chemical compounds present in the reaction (the columns of W) along with their proportions in each sample (the rows of H). This is an NMF problem which is referred to as self-modeling curve resolution (SMCR); see Section 9.3 for more details.

To the best of our knowledge, Wallace in 1960 [473] was the first to describe the model. He then discussed a way to estimate the number of sources (that is, r). Later in 1971, Lawton and Sylvestre [301] focused on the case $r = 2$ which can be solved easily; see Sections 2.1.3 and 4.1. In 1985, Borgen and Kowalski focused on the case $r = 3$, and, in 1986, Borgen et al. studied the general case [53]. The literature on the topic has grown rapidly since then, and we refer the reader to [263, 390, 366] and the references therein for more information on SMCR. Note that the problem of multivariate curve resolution (MCR) is more general than SMCR as it does not necessarily assume nonnegativity [119, 403].

Interestingly, in the SMCR literature, most works have analyzed the noiseless case, that is, $X = WH$, with the additional assumption that $\mathrm{rank}(X) = \mathrm{rank}(W) = r$. Under these assumptions, Exact NMF is equivalent to finding a transformation of an unconstrained factorization (such as the SVD) to make it nonnegative; see Section 5.5.1. A unique feature within the SMCR literature is that, in most works, the goal is to recover all possible factorizations, that is, all possible nonnegative (W, H) such that $X = WH$, referred to as the set of feasible solutions. The task of choosing the “right” factorization is left to the expert analyzing such chemical reactions. This is a particularly challenging problem, which may explain why an important part of the literature has focused on the exact case, as pointed out in [366].

1.4.2 ▪ Geoscience and remote sensing

The columns of X are the spectra of pixels within a hyperspectral image. The spectrum of a pixel records the fraction of light reflected by that pixel for many wavelengths. The goal is to recover the pure materials present in the image (the columns of W) and the abundance of the materials in each pixel (the columns of H). This is referred to as blind HU; see Section 1.3.2. This model was first used in 1964 by Imbrie and Van Andel for analyzing the mixture of heavy mineral data [254]. Early findings in this field include the works by Craig in the early 1990s [108, 109], by Boardman [48] (1993), and by Winter [482] (1999). The main contribution of these authors is they have shed light on the geometric interpretation of NMF (Section 2.1) and devised algorithms based on this intuition; see Chapters 4 and 7.

The literature on blind HU has grown considerably since the 1990s, motivated by the development of hyperspectral cameras that are becoming higher performing and more affordable as the years go by. We refer the reader to [276, 377, 45, 334] and the references therein for more information on blind HU.

1.4.3 ▪ Stochastic sequential machines

Another very early account of the NMF model appears in the study of stochastic finite state systems, also known as stochastic sequential machines (SSMs). An important problem in this context is to obtain a minimal state representation of the given system. The SSM can be represented with a nonnegative matrix that contains the transition probabilities between the different states. To the best of our knowledge, Ott [369] (1966) was the first to show that computing an Exact NMF of this matrix is equivalent to finding the minimal state representation of the SSM. Moreover, Ott discovered the connection between Exact NMF and the nested polytope problem (NPP), an important problem in computational geometry. Recall that a polytope is a bounded set defined via affine inequalities, that is, a bounded polyhedron. In two dimensions, polytopes are convex polygons. The NPP requires finding a polytope with the minimum number of vertices nested between two given nested polytopes. As we will discuss in length in Chapter 2, the Exact NMF problem with $r = \mathrm{rank}(X)$ is equivalent to the NPP where the sought solution is required to have r vertices. Follow-up works investigating this connection and providing new insights include the papers of Paz [378] (1968) and Bancilhon [22] (1974); see also the first chapter of the book of Paz [379] (1971). The problem of minimal state representation of an SSM is equivalent to the minimal covering of a labeled Markov chain. We refer the reader to [88] for more details and for the proof of the equivalence between Exact NMF and minimal covering of a labeled Markov chain.

1.4.4 ▪ Computational geometry

As explained in the previous section, Exact NMF is equivalent to the NPP. In two dimensions, the NPP requires finding a convex polygon with a minimum number of vertices nested between two given convex polygons. Motivated by its applicability for stochastic sequential machines, Silio [420] (1979) solved this problem by proposing a practical polynomial-time algorithm. However, his paper seems to have been overlooked in the computational geometry literature. The problem was later solved independently in a very similar way by Aggarwal et al. [4] (1989). However, the algorithm of Aggarwal et al. has a lower computational cost; see Section 2.3.1.1. In dimension higher than two, the NPP was shown to be NP-hard by Das and Joseph [114] (1990).

1.4.5 ▪ Linear algebra

The Exact NMF problem, with $X = WH$, is closely related to the notion of nonnegative rank. It started to draw attention after a question of Berman and Plemmons [38] in *SIAM Review* in 1973 (Problem 73-14):

> It is well known that an $m \times n$ matrix A of rank r can be factored (in a variety of ways) in the form $A = BG$ where B is of order $m \times r$ and G is of order $r \times n$. Show by counterexample, that when A is nonnegative there need not exist such a "rank factorization" with both B and G nonnegative. If possible, find a simple characterization of the class of nonnegative matrices A for which a nonnegative rank factorization exists.

In other words, this question asks us to characterize when the rank and the nonnegative rank of a matrix coincide, that is, when $\operatorname{rank}(X) = \operatorname{rank}_+(X)$.

Thomas [450] answered the first part of the question in *SIAM Review* in 1974: he showed that $\operatorname{rank}(X) = \operatorname{rank}_+(X)$ always holds when $\operatorname{rank}(X) \leq 2$ and gave a counterexample to show that this is not necessarily true when $\operatorname{rank}(X) \geq 3$:

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \tag{1.2}$$

for which $\operatorname{rank}_+(X) = 4$ while $\operatorname{rank}(X) = 3$; see Section 2.1 for the proofs and more details. It turns out that it is NP-hard to check whether $\operatorname{rank}(X) = \operatorname{rank}_+(X)$ [465]; see Section 2.3, which discusses the computational complexity of Exact NMF.

Early works on the nonnegative rank include the papers of Wall [472] (1979), Jeter and Pye [261] (1981), Campbell and Poole [72] (1981), and Chen [84] (1984). The first comprehensive account of the properties of the nonnegative rank was written by Cohen and Rothblum [100] in 1993; see Section 3.1 for more details.

1.4.6 ▪ Probability

NMF can be used to unravel a particular probabilistic model. Let $Y^{(k)} \in \{1, \dots, m\}$ and $Z^{(k)} \in \{1, \dots, n\}$ be two independent random variables for $1 \leq k \leq r$, and let $P^{(k)}$ be the joint distribution with

$$P_{ij}^{(k)} \;=\; \mathbb{P}\left(Y^{(k)} = i, Z^{(k)} = j\right) \;=\; \mathbb{P}\left(Y^{(k)} = i\right)\mathbb{P}\left(Z^{(k)} = j\right)$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. Each distribution $P^{(k)}$ corresponds to a nonnegative rank-one matrix. Let us define the joint distribution of two random variables Y and Z as follows:

- Choose the distribution $P^{(k)}$ with probability α_k, where $\sum_{k=1}^r \alpha_k = 1$.
- Draw Y and Z from the distribution $P^{(k)}$.

Equivalently, (Y, Z) has the following probability distribution: for $1 \leq i \leq m$ and $1 \leq j \leq n$,

$$\mathbb{P}(Y = i, Z = j) \;=\; \sum_{k=1}^{r} \alpha_k P_{i,j}^{(k)} \;=\; P_{i,j},$$

where the matrix P is the sum of r rank-one nonnegative matrices. In other words, P is the mixture of r independent distributions. In practice, only P is observed and is referred to as a contingency table, and computing its nonnegative rank and a corresponding factorization amounts to explaining the distribution P with as few independent variables as possible. Early work using this connection includes De Leeuw and Van der Heijden [120] (1991) and Ritov and Gilula [398] (1993); see Cohen and Rothblum [100] (1993) and Kubjas, Robeva, and Sturmfels [292] (2015) for more details. However, the first to discuss the above decomposition were Suppes and Zanotti [443] (1981), although without linking it explicitly with the nonnegative rank. In their terms, the nonnegative rank of P is the smallest support of a hidden random variable, which explains the correlation of the two-valued random variable whose joint distribution is represented by P.

Motivated by this application, Mond, Smith, and Van Straten [353] (2003) also discovered the link between Exact NMF and the NPP (which they call the sandwiched simplices problem) and studied the set of feasible solutions of Exact NMF.

1.4.7 ▪ Extended formulations

A standard approach in combinatorial optimization is to model the convex hull of the set of feasible solutions using affine inequalities, and then solve the problem using linear optimization. These linear formulations are referred to as extended formulations, and their size is defined as the number of inequalities; see Section 3.6 for more details. In the 1980s, some researchers were trying to prove P=NP by constructing such formulations of polynomial size for NP-complete combinatorial optimization problems, such as the traveling salesman problem (TSP).[7] If such formulations have polynomial size, then linear optimization can be used to solve them in polynomial time, which would imply P=NP.

In 1988, Yannakakis proved that this was not possible for the matching and TSP polytopes using extended formulations that are symmetric (for the TSP, this means that they are invariant under permutations of the cities in the input) [492, 493]; see also the discussion by Yannakakis in [494]. To do this, Yannakakis unraveled a key result: the minimum-size extended formulation of a polytope, referred to as its extension complexity, is equal to the nonnegative rank of its slack matrix. Given a polytope

$$\mathcal{P} \;=\; \left\{x \in \mathbb{R}^d \mid a_i^\top x \le b_i \text{ for } i = 1, 2, \dots, m\right\},$$

whose vertices are v_j for $j = 1, 2, \dots, n$, its slack matrix $S \in \mathbb{R}^{m\times n}_+$ is an inequality-by-vertex matrix where

$$S(i, j) = b_i - a_i^\top v_j \ge 0$$

is the slack of the jth vertex for the ith inequality.

Many results were obtained 20 years later (starting around 2010) to provide bounds on the extended formulations of combinatorial problems using bounds for the nonnegative rank. Several long-standing open questions were addressed via the nonnegative rank. For example, Fiorini et al. [163, 164] proved that the extension complexity of the TSP polytope is exponential in the number of cities (the difference with Yannakakis' result is that here asymmetric extended formulations are allowed, and asymmetry may reduce the size of extended formulations [270]). This is not surprising if you believe that $P \neq NP$. Rothvoss proved that the perfect matching polytope[8] has exponential extension complexity so that any extended formulation for the perfect matching polytope must have exponential size [400, 401]. This is somewhat surprising since optimizing a linear function over the perfect matching polytope can be performed in polynomial time [142]. Moreover, Braun and Pokutta [60] later established that for all fixed $0 < \epsilon < 1$, even every linear program approximating the matching polytope by a factor $(1 + \epsilon/n)$ must have exponential size, where n is the number of nodes in the graph. We refer the reader to Section 3.6 for more details and examples.

1.4.8 ▪ The first appearance of NMF in its modern form

As far as we know, the first time the NMF problem was explicitly stated as in (1.1) is in the paper by Paatero and Tapper in 1994 [371]. As discussed above, the models previously studied in the literature either considered exact factorizations, assumed $\operatorname{rank}(X) = \operatorname{rank}(W) = r$, or were only based on geometric representations. In their paper, Paatero and Tapper referred to this problem as positive matrix factorization (PMF), proposed an algorithm based on alternatively

[7]Given a set of cities, the TSP requires finding the shortest possible route that visits each city and returns to the origin city.

[8]The perfect matching polytope is the convex hull of the set of all perfect matchings of a complete graph (a perfect matching decomposes the set of vertices into pairs of vertices).

updating W and H (see Section 8.3.1), and explained how it can be used to analyze environmental data, for example, for air emission control. However, until 1999, PMF remained a specialized physical/chemical model confined within the field of chemometrics.

1.4.9 ▪ The "big bang" of NMF: data analysis and machine learning

NMF gained momentum with the seminal paper "Learning the Parts of Objects by Non-negative Matrix Factorization" by Lee and Seung published in *Nature* in 1999 [303]. It explained why NMF is a powerful tool for the analysis of nonnegative data sets. They illustrate their findings on two examples: the extraction of facial features in a set of facial images (see Figure 1.2, page 7) and the identification of topics within a set of documents (see Figure 1.7, page 10). Note that Lee and Seung also proposed algorithms based on multiplicative updates (MU) that became the workhorse in the NMF literature [304]; see Section 8.2.

Let us point out the difference in spirit of Lee and Seung compared to the previous literature by quoting Pentti Paatero:[9]

> The original concept of NMF, as presented in the *Nature* paper, was essentially different from our PMF: they did not search for the (hopefully unique) model of a data set that describes a physical/chemical situation. Instead, their goal could be described as "data compression." Such compressed version of data is normally not unique, it is not "THE solution."

In fact, prior to the paper of Lee and Seung, most works using NMF for data analysis focused on the physical model behind NMF and were hoping that NMF would produce the true underlying sources (such as pure spectra; see Sections 1.4.1 and 1.4.2). In their paper, Lee and Seung did not discuss this issue, and their goal was rather to compute one possible NMF decomposition that compresses the data and can be meaningfully interpreted. We will discuss this key identifiability question for NMF in Chapter 4.

1.5 ▪ Take-home messages

NMF is a linear dimensionality reduction technique with nonnegativity constraints on the factors and is able to extract meaningful components in various applications. From our account of the early literature on the NMF model, we observe that, before the paper of Lee and Seung, the study of NMF was motivated by either specific applications or theoretical questions. Lee and Seung were able to popularize NMF for the analysis of nonnegative data sets via the extraction of sparse and part-based components. They showed that NMF is very versatile and can be used in many different settings. This has been confirmed since then as NMF has been used successfully in many applications; see Chapter 9. In summary, the paper of Lee and Seung set a spark on NMF that has ignited a fire that has been growing steadily since then.

[9] Private communication.

Bibliography

[1] Abbe, E.: Community detection and stochastic block models: recent developments. Journal of Machine Learning Research **18**(177), 1–86 (2018) (Cited on p. 184)

[2] Abdolali, M., Gillis, N.: Simplex-structured matrix factorization: sparsity-based identifiability and provably correct algorithm. arXiv preprint arXiv:2007.11446 (2020) (Cited on pp. 155, 213)

[3] Abraham, B., Naomi, S.M.: Completely Positive Matrices. World Scientific (2003) (Cited on pp. 79, 179)

[4] Aggarwal, A., Booth, H., O'Rourke, J., Suri, S.: Finding minimal convex nested polygons. Information and Computation **83**(1), 98–110 (1989) (Cited on pp. xxi, 13, 44, 45, 46, 49)

[5] Aharon, M., Elad, M., Bruckstein, A., et al.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing **54**(11), 4311 (2006) (Cited on pp. 3, 150)

[6] Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P.: Mixed membership stochastic blockmodels. Journal of Machine Learning Research **9**(Sep), 1981–2014 (2008) (Cited on p. 184)

[7] Alexe, G., Alexe, S., Crama, Y., Foldes, S., Hammer, P.L., Simeone, B.: Consensus algorithms for the generation of all maximal bicliques. Discrete Applied Mathematics **145**(1), 11–21 (2004) (Cited on p. 72)

[8] Ambikapathi, A., Chan, T.H., Chi, C.Y., Keizer, K.: Hyperspectral data geometry-based estimation of number of endmembers using p-norm-based pure pixel identification algorithm. IEEE Transactions on Geoscience and Remote Sensing **51**(5), 2753–2769 (2012) (Cited on p. 229)

[9] Ambikapathi, A., Chan, T.H., Ma, W.K., Chi, C.Y.: Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing **49**(11), 4194–4209 (2011) (Cited on p. 138)

[10] Ang, A.M.S.: Nonnegative matrix and tensor factorizations: Models, algorithms and applications. Ph.D. thesis, University of Mons (2020), `https://angms.science/doc/PhDThesis.pdf` (Cited on p. 169)

[11] Ang, A.M.S., Cohen, J.E., Gillis, N., Hien, L.T.K.: Accelerating block coordinate descent for nonnegative tensor factorization. arXiv preprint arXiv:2001.04321 (2020) (Cited on pp. 293, 294, 296)

[12] Ang, A.M.S., Gillis, N.: Accelerating nonnegative matrix factorization algorithms using extrapolation. Neural Computation **31**(2), 417–439 (2019) (Cited on p. 293)

[13] Ang, A.M.S., Gillis, N.: Algorithms and comparisons of nonnegative matrix factorizations with volume regularization for hyperspectral unmixing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **12**(12), 4843–4853 (2019) (Cited on pp. 144, 148)

[14] Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems **57**(2), 65–73 (2001) (Cited on pp. 224, 229)

[15] Arora, S., Ge, R., Halpern, Y., Mimno, D., Moitra, A., Sontag, D., Wu, Y., Zhu, M.: A practical algorithm for topic modeling with provable guarantees. In: Proceedings of the 30th International Conference on Machine Learning, pp. 280–288 (2013) (Cited on pp. 183, 193, 208, 210, 226, 230, 231, 257, 258, 259, 308)

[16] Arora, S., Ge, R., Kannan, R., Moitra, A.: Computing a nonnegative matrix factorization–provably. In: Proceedings of STOC '12, pp. 145–162 (2012) (Cited on pp. 52, 53, 200, 208, 232, 249, 257)

[17] Arora, S., Ge, R., Kannan, R., Moitra, A.: Computing a nonnegative matrix factorization—provably. SIAM Journal on Computing **45**(4), 1582–1611 (2016) (Cited on pp. 52, 53, 200)

[18] Asteris, M., Papailiopoulos, D., Dimakis, A.G.: Orthogonal NMF through subspace exploration. In: Advances in Neural Information Processing Systems (NIPS), pp. 343–351 (2015) (Cited on p. 136)

[19] Atif, S.M., Qazi, S., Gillis, N.: Improved SVD-based initialization for nonnegative matrix factorization using low-rank correction. Pattern Recognition Letters **122**, 53–59 (2019) (Cited on p. 301)

[20] Ban, F., Bhattiprolu, V., Bringmann, K., Kolev, P., Lee, E., Woodruff, D.P.: A PTAS for ℓ_p-low rank approximation. In: Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 747–766. SIAM (2019) (Cited on pp. 204, 205)

[21] Ban, F., Woodruff, D., Zhang, R.: Regularized weighted low rank approximation. In: Advances in Neural Information Processing Systems (NIPS), pp. 4061–4071 (2019) (Cited on p. 203)

[22] Bancilhon, F.: A geometric model for stochastic automata. IEEE Transactions on Computers **100**(12), 1290–1299 (1974) (Cited on p. 13)

[23] Baranzini, S.E., Mousavi, P., Rio, J., Caillier, S.J., Stillman, A., Villoslada, P., Wyatt, M.M., Comabella, M., Greller, L.D., Somogyi, R., et al.: Transcription-based prediction of response to IFNβ using supervised computational methods. PLoS Biology **3**(1), e2 (2004) (Cited on p. 310)

[24] Barefoot, C., Hefner, K.A., Jones, K.F., Lundgren, J.R.: Biclique covers of the complements of cycles and paths in a digraph. Congressus Numerantium **53**, 133–146 (1986) (Cited on p. 72)

[25] Barioli, F., Berman, A.: The maximal cp-rank of rank k completely positive matrices. Linear Algebra and Its Applications **363**, 17–33 (2003) (Cited on p. 79)

[26] Bärmann, A.: Polyhedral Approximation of Second-Order Cone Robust Counterparts, pp. 163–177. Springer Fachmedien Wiesbaden (2016) (Cited on p. 90)

[27] Basu, S., Pollack, R., Roy, M.F.: On the combinatorial and algebraic complexity of quantifier elimination. Journal of the ACM **43**(6), 1002–1045 (1996) (Cited on p. 52)

[28] Beasley, L.B., Laffey, T.J.: Real rank versus nonnegative rank. Linear Algebra and Its Applications **431**(12), 2330–2335 (2009) (Cited on pp. 65, 70)

[29] Beck, A.: First-Order Methods in Optimization. MOS-SIAM Series on Optimization vol. 25. SIAM (2017) (Cited on pp. 268, 270, 290)

[30] Beer, A.: Bestimmung der Absorption des rothen Lichts in farbigen Flussigkeiten. Annalen der Physik **162**, 78–88 (1852) (Cited on p. 12)

[31] Belachew, M.T.: Efficient algorithm for sparse symmetric nonnegative matrix factorization. Pattern Recognition Letters **125**, 735–741 (2019) (Cited on pp. 185, 287)

[32] Bell, R.M., Koren, Y.: Lessons from the Netflix prize challenge. SiGKDD Explorations **9**(2), 75–79 (2007) (Cited on p. 3)

[33] Ben-Tal, A., Nemirovski, A.: On polyhedral approximations of the second-order cone. Mathematics of Operations Research **26(2)**, 193–205 (2001) (Cited on p. 90)

[34] Benetos, E., Dixon, S., Duan, Z., Ewert, S.: Automatic music transcription: An overview. IEEE Signal Processing Magazine **36**(1), 20–30 (2019) (Cited on p. 11)

[35] Bennett, J., Lanning, S., et al.: The Netflix prize. In: Proceedings of KDD Cup and Workshop, p. 35. New York (2007) (Cited on p. 3)

[36] Benson, A.R., Lee, J.D., Rajwa, B., Gleich, D.F.: Scalable methods for nonnegative matrix factorizations of near-separable tall-and-skinny matrices. In: Advances in Neural Information Processing Systems (NIPS), pp. 945–953 (2014) (Cited on p. 216)

[37] Berman, A., Dur, M., Shaked-Monderer, N.: Open problems in the theory of completely positive and copositive matrices. Electronic Journal of Linear Algebra **29**(1), 46–58 (2015) (Cited on p. 179)

[38] Berman, A., Plemmons, R.J.: Rank factorization of nonnegative matrices. SIAM Review **15**(3), 655 (1973) (Cited on p. 13)

[39] Berman, A., Plemmons, R.J.: Nonnegative Matrices in the Mathematical Sciences, Classics in Applied Mathematics vol. 9. SIAM (1994) (Cited on pp. 105, 174, 197)

[40] Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., Plemmons, R.J.: Algorithms and applications for approximate nonnegative matrix factorization. Computational Statistics & Data Analysis **52**(1), 155–173 (2007) (Cited on p. 314)

[41] Bertsekas, D.: Corrections for the Book Nonlinear Programming: Second Edition (1999). `http://www.athenasc.com/nlperrata.pdf` (Cited on p. 268)

[42] Bertsekas, D.: Nonlinear Programming: Second Edition. Athena Scientific (1999) (Cited on p. 268)

[43] Bhadury, J., Chandrasekaran, R.: Finding the set of all minimal nested convex polygons. In: Proceedings of the 8th Canadian Conference on Computational Geometry, pp. 26–31 (1996) (Cited on p. 44)

[44] Biggs, M., Ghodsi, A., Vavasis, S.: Nonnegative matrix factorization via rank-one downdate. In: Proceedings of the 25th International Conference on Machine Learning, pp. 64–71 (2008) (Cited on p. 302)

[45] Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., Chanussot, J.: Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **5**(2), 354–379 (2012) (Cited on pp. 13, 209, 247, 248)

[46] Blei, D.M.: Probabilistic topic models. Communicatoins of the ACM **55**(4), 77–84 (2012) (Cited on p. 192)

[47] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. Journal of Machine Learning Research **3**(Jan), 993–1022 (2003) (Cited on pp. 183, 192, 193)

[48] Boardman, J.W.: Automating spectral unmixing of AVIRIS data using convex geometry concepts. In: Summaries of the 4th Annual JPL Airborne Geoscience Workshop, vol. 1, pp. 11–14 (1993) (Cited on p. 13)

[49] Boardman, J.W., Kruse, F.A., Green, R.O.: Mapping target signatures via partial unmixing of AVIRIS data. In: Summaries of the JPL Airborne Earth Science Workshop, Pasadena, CA, pp. 23–26 (1995) (Cited on pp. 209, 247)

[50] Bobin, J., Starck, J.L., Fadili, J., Moudden, Y.: Sparsity and morphological diversity in blind source separation. IEEE Transactions on Image Processing **16**(11), 2662–2674 (2007) (Cited on p. 170)

[51] Bocci, C., Carlini, E., Rapallo, F.: Perturbation of matrices and nonnegative rank with a view toward statistical models. SIAM Journal on Matrix Analysis and Applications **32**(4), 1500–1512 (2011) (Cited on pp. 47, 57)

[52] Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Mathematical Programming **146**(1–2), 459–494 (2014) (Cited on p. 270)

[53] Borgen, O.S., Davidsen, N., Mingyang, Z., Øyen, Ø.: The multivariate n-component resolution problem with minimum assumptions. Microchimica Acta **89**(1–6), 63–73 (1986) (Cited on p. 12)

[54] Boutsidis, C., Gallopoulos, E.: SVD based initialization: A head start for nonnegative matrix factorization. Pattern Recognition **41**(4), 1350–1362 (2008) (Cited on p. 300)

[55] Boutsidis, C., Mahoney, M.W., Drineas, P.: An improved approximation algorithm for the column subset selection problem. In: Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 968–977. SIAM (2009) (Cited on p. 211)

[56] Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2004) (Cited on p. 238)

[57] Braun, G., Fiorini, S., Pokutta, S., Steurer, D.: Approximation limits of linear programs (beyond hierarchies). In: Proceedings of the Annual IEEE Symposium on Foundations of Computer Science, pp. 480–489 (2012) (Cited on pp. 76, 92)

[58] Braun, G., Fiorini, S., Pokutta, S., Steurer, D.: Approximation limits of linear programs (beyond hierarchies). Mathematics of Operations Research **40**(3), 756–772 (2015) (Cited on pp. 76, 92)

[59] Braun, G., Pokutta, S.: Common information and unique disjointness. In: 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, pp. 688–697. IEEE (2013) (Cited on p. 92)

[60] Braun, G., Pokutta, S.: The matching polytope does not admit fully-polynomial size relaxation schemes. In: Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 837–846. SIAM (2015) (Cited on pp. 15, 92)

[61] Braun, G., Pokutta, S.: Common information and unique disjointness. Algorithmica **76**(3), 597–629 (2016) (Cited on pp. 82, 83, 92, 95)

[62] Braverman, M., Moitra, A.: An information complexity approach to extended formulations. In: Proceedings of the 45th Annual ACM Symposium on Theory of Computing, pp. 161–170. ACM (2013) (Cited on p. 92)

[63] Bro, R.: Multi-way analysis in the food industry: Models, algorithms, and applications. Ph.D. thesis, University of Copenhagen (1998). `http://curis.ku.dk/ws/files/13035961/Rasmus_Bro.pdf` (Cited on p. 286)

[64] Bro, R., Acar, E., Kolda, T.G.: Resolving the sign ambiguity in the singular value decomposition. Journal of Chemometrics **22**(2), 135–140 (2008) (Cited on p. 197)

[65] Bro, R., Sidiropoulos, N.D.: Least squares algorithms under unimodality and non-negativity constraints. Journal of Chemometrics **12**(4), 223–247 (1998) (Cited on pp. 169, 313)

[66] Brunet, J.P., Tamayo, P., Golub, T.R., Mesirov, J.P.: Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the National Academy of Sciences **101**(12), 4164–4169 (2004) (Cited on p. 309)

[67] Buciu, I., Nikolaidis, N., Pitas, I.: Nonnegative matrix factorization in polynomial feature space. IEEE Transactions on Neural Networks **19**(6), 1090–1100 (2008) (Cited on p. 185)

[68] Burer, S., Monteiro, R.D.: A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Mathematical Programming **95**(2), 329–357 (2003) (Cited on p. 3)

[69] Businger, P., Golub, G.: Linear least squares solutions by Householder transformations. Numerische Mathematik **7**, 269–276 (1965) (Cited on p. 229)

[70] de Caen, D., Gregory, D., Pullman, N.: The Boolean rank of zero-one matrices. In: Proceedings of the 3rd Caribbean Conference on Combinatorics and Computing, pp. 169–173 (1981) (Cited on pp. 69, 70)

[71] Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized nonnegative matrix factorization for data representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(8), 1548–1560 (2011) (Cited on p. 170)

[72] Campbell, S., Poole, G.: Computing nonnegative rank factorizations. Linear Algebra and Its Applications **35**, 175–182 (1981) (Cited on p. 14)

[73] Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM **58**(3), 11 (2011) (Cited on p. 3)

[74] Cao, Y., Eggermont, P.P., Terebey, S.: Cross burg entropy maximization and its application to ringing suppression in image reconstruction. IEEE Transactions on Image Processing **8**(2), 286–292 (1999) (Cited on p. 270)

[75] Casalino, G., Del Buono, N., Mencar, C.: Subtractive clustering for seeding non-negative matrix factorizations. Information Sciences **257**, 369–387 (2014) (Cited on p. 301)

[76] Casalino, G., Gillis, N.: Sequential dimensionality reduction for extracting localized features. Pattern Recognition **63**, 15–29 (2017) (Cited on p. 177)

[77] Çivril, A., Magdon-Ismail, M.: On selecting a maximum volume sub-matrix of a matrix and related problems. Theoretical Computer Science **410**(47–49), 4801–4811 (2009) (Cited on pp. 204, 227, 229)

[78] Çivril, A., Magdon-Ismail, M.: Exponential inapproximability of selecting a maximum volume submatrix. Algorithmica **65**(1), 159–176 (2013) (Cited on pp. 204, 229)

[79] Chan, T.H., Chi, C.Y., Huang, Y.M., Ma, W.K.: A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. IEEE Transactions on Signal Processing **57**(11), 4418–4432 (2009) (Cited on p. 138)

[80] Chan, T.H., Ma, W.K., Ambikapathi, A., Chi, C.Y.: A simplex volume maximization framework for hyperspectral endmember extraction. IEEE Transactions on Geoscience and Remote Sensing **49**(11), 4177–4193 (2011) (Cited on pp. 227, 229)

[81] Chan, T.H., Ma, W.K., Chi, C.Y., Wang, Y.: A convex analysis framework for blind separation of non-negative sources. IEEE Transactions on Signal Processing **56**(10), 5120–5134 (2008) (Cited on p. 209)

[82] Chen, G., Wang, F., Zhang, C.: Collaborative filtering using orthogonal nonnegative matrix tri-factorization. Information Processing & Management **45**(3), 368–379 (2009) (Cited on p. 181)

[83] Chen, J., Huo, X.: Theoretical results on sparse representations of multiple-measurement vectors. IEEE Transactions on Signal processing **54**(12), 4634–4643 (2006) (Cited on p. 252)

[84] Chen, J.C.: The nonnegative rank factorizations of nonnegative matrices. Linear Algebra and Its Applications **62**, 207–217 (1984) (Cited on p. 14)

[85] Chen, R., Varshney, L.R.: Optimal recovery of missing values for non-negative matrix factorization. bioRxiv 647560 (2019) (Cited on p. 311)

[86] Chi, E.C., Kolda, T.G.: On tensors, sparsity, and nonnegative factorizations. SIAM Journal on Matrix Analysis and Applications **33**(4), 1272–1299 (2012) (Cited on pp. 163, 164, 279)

[87] Chi, Y., Lu, Y.M., Chen, Y.: Nonconvex optimization meets low-rank matrix factorization: An overview. IEEE Transactions on Signal Processing **67**(20), 5239–5269 (2019) (Cited on p. 3)

[88] Chistikov, D., Kiefer, S., Marusic, I., Shirmohammadi, M., Worrell, J.: On restricted nonnegative matrix factorization. In: I. Chatzigiannakis, M. Mitzenmacher, Y. Rabani, D. Sangiorgi (eds.) 43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016), Leibniz International Proceedings in Informatics (LIPIcs), vol. 55, pp. 103:1–103:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2016). DOI:10.4230/LIPIcs.ICALP.2016.103 (Cited on pp. 13, 35, 36, 40, 54)

[89] Chistikov, D., Kiefer, S., Marušić, I., Shirmohammadi, M., Worrell, J.: Nonnegative matrix factorization requires irrationality. SIAM Journal on Applied Algebra and Geometry **1**(1), 285–307 (2017) (Cited on p. 54)

[90] Choi, H., Choi, S., Katake, A., Choe, Y.: Learning α-integration with partially-labeled data. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2058–2061. IEEE (2010) (Cited on p. 164)

[91] Choi, S.: Algorithms for orthogonal nonnegative matrix factorization. In: Proceedings of the International Joint Conference on Neural Networks, pp. 1828–1832 (2008) (Cited on p. 136)

[92] Chouh, M., Hanafi, M., Boukhetala, K.: Semi-nonnegative rank for real matrices and its connection to the usual rank. Linear Algebra and Its Applications **466**, 27–37 (2015) (Cited on p. 173)

[93] Chu, M.T., Lin, M.M.: Low-dimensional polytope approximation and its applications to nonnegative matrix factorization. SIAM Journal on Scientific Computing **30**(3), 1131–1155 (2008) (Cited on p. 22)

[94] Chung, F.: Spectral Graph Theory. CBMS Regional Conference Series in Mathematics 92. AMS (1997) (Cited on p. 3)

[95] Cichocki, A., Phan, A.: Fast local algorithms for large scale nonnegative matrix and tensor factorizations. IEICE Transactions on Fundamentals of Electronics **E92-A**(3), 708–721 (2009) (Cited on p. 286)

[96] Cichocki, A., Zdunek, R., Amari, S.i.: New algorithms for non-negative matrix factorization in applications to blind source separation. In: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 5, pp. 621–624. IEEE (2006) (Cited on p. 280)

[97] Cichocki, A., Zdunek, R., Amari, S.i.: Hierarchical ALS Algorithms for Nonnegative Matrix and 3D Tensor Factorization. In: Lecture Notes in Computer Science, vol. 4666, Springer, pp. 169–176 (2007) (Cited on p. 286)

[98] Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.i.: Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation. John Wiley & Sons (2009) (Cited on pp. xii, xiv, 5, 163, 168, 193, 280, 283, 284, 299, 314)

[99] Clarkson, K.: Algorithms for polytope covering and approximation. In: Proceedings of the Third Workshop on Algorithms and Data Structures, pp. 246–252 (1993) (Cited on p. 51)

[100] Cohen, J., Rothblum, U.: Nonnegative ranks, decompositions and factorization of nonnegative matrices. Linear Algebra and Its Applications **190**, 149–168 (1993) (Cited on pp. 14, 23, 53, 55, 63)

[101] Cohen, J.E., Gillis, N.: Dictionary-based tensor canonical polyadic decomposition. IEEE Transactions on Signal Processing **66**(7), 1876–1889 (2017) (Cited on p. 172)

[102] Cohen, J.E., Gillis, N.: Identifiability of complete dictionary learning. SIAM Journal on Mathematics of Data Science **1**(3), 518–536 (2019) (Cited on pp. 152, 153)

[103] Cohen, J.E., Gillis, N.: Nonnegative low-rank sparse component analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8226–8230 (2019) (Cited on p. 153)

[104] Comon, P., Jutten, C.: Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press (2010) (Cited on p. 3)

[105] Conforti, M., Cornuéjols, G., Zambelli, G.: Extended formulations in combinatorial optimization. 4OR: A Quarterly Journal of Operations Research **10**(1), 1–48 (2010) (Cited on p. 84)

[106] Čopar, A., Zupan, B., Zitnik, M.: Fast optimization of non-negative matrix tri-factorization. PloS One **14**(6) (2019) (Cited on p. 287)

[107] Cotter, S.F., Rao, B.D., Engan, K., Kreutz-Delgado, K.: Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Transactions on Signal Processing **53**(7), 2477–2488 (2005) (Cited on p. 252)

[108] Craig, M.D.: Unsupervised unmixing of remotely sensed images. In: Proceedings of the Fifth Australasian Remote Sensing Conference, pp. 324–330 (1990) (Cited on p. 12)

[109] Craig, M.D.: Minimum-volume transforms for remotely sensed data. IEEE Transactions on Geoscience and Remote Sensing **32**(3), 542–552 (1994) (Cited on pp. 12, 138)

[110] Cutler, A., Breiman, L.: Archetypal analysis. Technometrics **36**(4), 338–347 (1994) (Cited on p. 172)

[111] Damle, A., Sun, Y.: A geometric approach to archetypal analysis and nonnegative matrix factorization. Technometrics **59**(3), 361–370 (2017) (Cited on p. 247)

[112] Das, G.: Approximation schemes in computational geometry. Ph.D. thesis, University of Wisconsin–Madison (1990) (Cited on pp. 20, 36, 50, 51)

[113] Das, G., Goodrich, M.: On the complexity of optimization problems for three-dimensional convex polyhedra and decision trees. Computational Geometry: Theory and Applications **8**, 123–137 (1997) (Cited on p. 51)

[114] Das, G., Joseph, D.: The complexity of minimum convex nested polyhedra. In: Proceedings of the 2nd Canadian Conference on Computational Geometry, pp. 296–301 (1990) (Cited on pp. 13, 20, 36, 50, 51)

[115] d'Aspremont, A., El Ghaoui, L., Jordan, M.I., Lanckriet, G.R.G.: A direct formulation for sparse PCA using semidefinite programming. SIAM Review **49**(3), 434–448 (2007) (Cited on pp. 3, 150)

[116] Daube-Witherspoon, M., Muehllehner, G.: An iterative image space reconstruction algorithm suitable for volume ECT. IEEE Transactions on Medical Imaging **5**, 61–66 (1986) (Cited on p. 270)

[117] De Handschutter, P., Gillis, N., Siebert, X.: Deep matrix factorizations. arXiv:2010.00380 (2020) (Cited on p. 184)

[118] De Handschutter, P., Gillis, N., Vandaele, A., Siebert, X.: Near-convex archetypal analysis. IEEE Signal Processing Letters **27**(1), 81–85 (2020) (Cited on pp. 172, 193, 211)

[119] De Juan, A., Tauler, R.: Multivariate curve resolution (MCR) from 2000: progress in concepts and applications. Critical Reviews in Analytical Chemistry **36**(3-4), 163–176 (2006) (Cited on p. 12)

[120] De Leeuw, J., Van der Heijden, P.G.: Reduced rank models for contingency tables. Biometrika **78**(1), 229–232 (1991) (Cited on pp. 14, 313)

[121] De Pierro, A.R.: On the relation between the ISRA and the EM algorithm for positron emission tomography. IEEE Transactions on Medical Imaging **12**(2), 328–333 (1993) (Cited on p. 270)

[122] Debals, O., Van Barel, M., De Lathauwer, L.: Nonnegative matrix factorization using nonnegative polynomial approximations. IEEE Signal Processing Letters **24**(7), 948–952 (2017) (Cited on p. 169)

[123] Degleris, A., Gillis, N.: A provably correct and robust algorithm for convolutive nonnegative matrix factorization. IEEE Transactions on Signal Processing **68** (2020). DOI:10.1109/TSP.2020.2984163 (Cited on pp. 185, 259)

[124] Del Corso, G.M., Romani, F.: Adaptive nonnegative matrix factorization and measure comparisons for recommender systems. Applied Mathematics and Computation **354**, 164–179 (2019) (Cited on p. 311)

[125] Devarajan, K.: Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. PLoS Computational Biology **4**(7), e1000,029 (2008) (Cited on pp. 309, 314)

[126] Deville, Y.: From separability/identifiability properties of bilinear and linear-quadratic mixture matrix factorization to factorization algorithms. Digital Signal Processing **87**, 21–33 (2019) (Cited on p. 185)

[127] Dewez, J., Gillis, N., Glineur, F.: A geometric lower bound on the extension complexity of polytopes based on the f-vector. Discrete Applied Mathematics, to appear (Cited on p. 75)

[128] Dewez, J., Glineur, F.: Lower bounds on the nonnegative rank using a nested polytopes formulation. In: 28th European Symposium on Artificial Neural Networks-Computational Intelligence and Machine Learning (ESANN) (2020) (Cited on p. 47)

[129] Dikmen, O., Yang, Z., Oja, E.: Learning the information divergence. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(7), 1442–1454 (2015) (Cited on p. 164)

[130] Ding, C., He, X., Simon, H.D.: On the equivalence of nonnegative matrix factorization and spectral clustering. In: Proceedings of the 5th SIAM Conference on Data Mining, pp. 606–610 (2005) (Cited on p. 136)

[131] Ding, C., Li, T., Peng, W.: On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. Computational Statistics & Data Analysis **52**(8), 3913–3927 (2008) (Cited on pp. 191, 192)

[132] Ding, C., Li, T., Peng, W., Park, H.: Orthogonal nonnegative matrix t-factorizations for clustering. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 126–135. ACM (2006) (Cited on pp. 136, 181, 280)

[133] Ding, C.H., Li, T., Jordan, M.I.: Convex and semi-nonnegative matrix factorizations. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(1), 45–55 (2008) (Cited on pp. 172, 173)

[134] Doan, X.V., Vavasis, S.: Finding the largest low-rank clusters with Ky Fan 2-k-norm and ℓ_1-norm. SIAM Journal on Optimization **26**(1), 274–312 (2016) (Cited on p. 197)

[135] Dobbins, M.G., Holmsen, A.F., Miltzow, T.: On the complexity of nesting polytopes. In: 35th European Workshop on Computational Geometry (EuroCG 2019) (2019) (Cited on pp. 20, 36, 54)

[136] Dobigeon, N., Tourneret, J.Y., Richard, C., Bermudez, J.C.M., McLaughlin, S., Hero, A.O.: Nonlinear unmixing of hyperspectral images: Models and algorithms. IEEE Signal Processing Magazine **31**(1), 82–94 (2013) (Cited on pp. 185, 209)

[137] Dong, B., Lin, M.M., Chu, M.T.: Nonnegative rank factorization—a heuristic approach via rank reduction. Numerical Algorithms **65**(2), 251–274 (2014) (Cited on pp. 53, 110, 128, 302)

[138] Donoho, D., Stodden, V.: When does non-negative matrix factorization give a correct decomposition into parts? In: Advances in Neural Information Processing Systems (NIPS), pp. 1141–1148 (2004) (Cited on pp. 6, 104, 108, 208)

[139] Du, R., Kuang, D., Drake, B., Park, H.: DC-NMF: nonnegative matrix factorization based on divide-and-conquer for fast clustering and topic modeling. Journal of Global Optimization **68**(4), 777–798 (2017) (Cited on p. 303)

[140] Dür, M.: Copositive programming—a survey. In: Recent Advances in Optimization and Its Applications in Engineering, pp. 3–20. Springer (2010) (Cited on p. 79)

[141] Eckart, C., Young, G.: The approximation of one matrix by another of lower rank. Psychometrika **1**(3), 211–218 (1936) (Cited on p. 196)

[142] Edmonds, J.: Maximum matching and a polyhedron with 0, 1-vertices. Journal of Research of the National Bureau of Standards B **69**(125–130), 55–56 (1965) (Cited on p. 15)

[143] Eftekhari, A., Hauser, R.A.: Principal component analysis by optimization of symmetric functions has no spurious local optima. SIAM Journal on Optimization **30**(1), 439–463 (2020) (Cited on p. 196)

[144] Eldén, L.: Matrix Methods in Data Mining and Pattern Recognition, 2nd ed. SIAM (2019) (Cited on p. 2)

[145] Elhamifar, E., Sapiro, G., Vidal, R.: See all by looking at a few: Sparse modeling for finding representative objects. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1600–1607. IEEE (2012) (Cited on pp. 211, 251, 252)

[146] Erichson, N.B., Mendible, A., Wihlborn, S., Kutz, J.N.: Randomized nonnegative matrix factorization. Pattern Recognition Letters **104**, 1–7 (2018) (Cited on p. 292)

[147] Esposito, F., Gillis, N., Del Buono, N.: Orthogonal joint sparse NMF for microarray data analysis. Journal of Mathematical Biology **79**(1), 223–247 (2019) (Cited on pp. 309, 310, 311)

[148] Esser, E., Moller, M., Osher, S., Sapiro, G., Xin, J.: A convex model for nonnegative matrix factorization and dimensionality reduction on physical space. IEEE Transactions on Image Processing **21**(7), 3239–3252 (2012) (Cited on pp. 211, 251, 252)

[149] Fagot, D., Wendt, H., Févotte, C.: Nonnegative matrix factorization with transform learning. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2431–2435. IEEE (2018) (Cited on p. 308)

[150] Fagot, D., Wendt, H., Févotte, C., Smaragdis, P.: Majorization-minimization algorithms for convolutive NMF with the beta-divergence. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8202–8206. IEEE (2019) (Cited on p. 280)

[151] Fawzi, H.: On representing the positive semidefinite cone using the second-order cone. Mathematical Programming **175**(1), 109–118 (2019) (Cited on p. 93)

[152] Fawzi, H., Gouveia, J., Parrilo, P., Robinson, R., Thomas, R.: Positive semidefinite rank. Mathematical Programming **153**(1), 133–177 (2015) (Cited on p. 93)

[153] Fawzi, H., Gouveia, J., Parrilo, P.A., Saunderson, J., Thomas, R.R.: Lifting for simplicity: concise descriptions of convex sets. arXiv preprint arXiv:2002.09788 (2020) (Cited on pp. 84, 93)

[154] Fawzi, H., Parrilo, P.A.: Lower bounds on nonnegative rank via nonnegative nuclear norms. Mathematical Programming **153**(1), 41–66 (2015) (Cited on pp. 77, 78, 79, 80)

[155] Fawzi, H., Parrilo, P.A.: Self-scaled bounds for atomic cone ranks: applications to nonnegative rank and cp-rank. Mathematical Programming **158**(1-2), 417–465 (2016) (Cited on pp. 47, 80, 81, 82)

[156] Fazel, M., Hindi, H., Boyd, S.P.: Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In: Proceedings of the 2003 American Control Conference, vol. 3, pp. 2156–2162. IEEE (2003) (Cited on p. 148)

[157] Févotte, C.: Majorization-minimization algorithm for smooth Itakura-Saito nonnegative matrix factorization. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1980–1983. IEEE (2011) (Cited on p. 169)

[158] Févotte, C., Bertin, N., Durrieu, J.L.: Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. Neural Computation **21**(3), 793–830 (2009) (Cited on pp. 10, 162, 163, 165, 270)

[159] Févotte, C., Dobigeon, N.: Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization. arXiv preprint arXiv:1401.5649 (2014) (Cited on p. 164)

[160] Févotte, C., Idier, J.: Algorithms for nonnegative matrix factorization with the β-divergence. Neural Computation **23**(9), 2421–2456 (2011) (Cited on pp. 163, 270, 273, 276)

[161] Fiorini, S., Guo, K., Macchia, M., Walter, M.: Lower bound computations for the nonnegative rank. In: 17th Cologne-Twente Workshop on Graphs and Combinatorial Optimization (2019) (Cited on p. 82)

[162] Fiorini, S., Kaibel, V., Pashkovich, K., Theis, D.O.: Combinatorial bounds on nonnegative rank and extended formulations. Discrete Mathematics **313**(1), 67–83 (2013) (Cited on pp. 63, 64, 71, 72, 73, 83, 85)

[163] Fiorini, S., Massar, S., Pokutta, S., Tiwary, H.R., De Wolf, R.: Linear vs. semidefinite extended formulations: exponential separation and strong lower bounds. In: Proceedings of the 44th Annual ACM Symposium on Theory of Computing, pp. 95–106. ACM (2012) (Cited on pp. 15, 83, 85, 86, 95)

[164] Fiorini, S., Massar, S., Pokutta, S., Tiwary, H.R., Wolf, R.D.: Exponential lower bounds for polytopes in combinatorial optimization. Journal of the ACM **62**(2), 17 (2015) (Cited on pp. 15, 86, 95)

[165] Fiorini, S., Rothvoss, T., Tiwary, H.R.: Extended formulations for polygons. Discrete & Computational Geometry **48**(3), 658–668 (2012) (Cited on pp. 82, 88, 90, 92)

[166] Flamant, J., Miron, S., Brie, D.: Quaternion non-negative matrix factorization: Definition, uniqueness, and algorithm. IEEE Transactions on Signal Processing, **68** 1870–1883 (2020) (Cited on p. 185)

[167] Fortunato, S.: Community detection in graphs. Physics Reports **486**(3–5), 75–174 (2010) (Cited on p. 178)

[168] Freund, R.M., Orlin, J.B.: On the complexity of four polyhedral set containment problems. Mathematical Programming **33**(2), 139–145 (1985) (Cited on p. 118)

[169] Fu, X., Huang, K., Sidiropoulos, N.D.: On identifiability of nonnegative matrix factorization. IEEE Signal Processing Letters **25**(3), 328–332 (2018) (Cited on pp. 142, 144, 145)

[170] Fu, X., Huang, K., Sidiropoulos, N.D., Ma, W.K.: Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications. IEEE Signal Processing Magazine **36**(2), 59–80 (2019) (Cited on pp. 5, 108, 122, 124, 125, 148, 149, 184, 259, 314)

[171] Fu, X., Huang, K., Sidiropoulos, N.D., Shi, Q., Hong, M.: Anchor-free correlated topic modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(5), 1056–1071 (2018) (Cited on pp. 139, 193, 259)

[172] Fu, X., Huang, K., Yang, B., Ma, W.K., Sidiropoulos, N.D.: Robust volume minimization-based matrix factorization for remote sensing and document clustering. IEEE Transactions on Signal Processing **64**(23), 6254–6268 (2016) (Cited on pp. 144, 149, 288)

[173] Fu, X., Ibrahim, S., Wai, H., Gao, C., Huang, K.: Block-randomized stochastic proximal gradient for low-rank tensor factorization. IEEE Transactions on Signal Processing **68**, 2170–2185 (2020) (Cited on p. 304)

[174] Fu, X., Ma, W.K.: Robustness analysis of structured matrix factorization via self-dictionary mixed-norm optimization. IEEE Signal Processing Letters **23**(1), 60–64 (2015) (Cited on p. 252)

[175] Fu, X., Ma, W.K., Chan, T.H., Bioucas-Dias, J.M.: Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related. IEEE Journal of Selected Topics in Signal Processing **9**(6), 1128–1141 (2015) (Cited on pp. 229, 252)

[176] Fu, X., Ma, W.K., Huang, K., Sidiropoulos, N.D.: Blind separation of quasi-stationary sources: exploiting convex geometry in covariance domain. IEEE Transactions on Signal Processing **63**(9), 2306–2320 (2015) (Cited on pp. 139, 140, 141)

[177] Fu, X., Vervliet, N., De Lathauwer, L., Huang, K., Gillis, N: Computing large-scale matrix and tensor decomposition with structured factors: A unified nonconvex optimization perspective. IEEE Signal Processing Magazine, **37**(5), 78–94 (2020) (Cited on p. 304)

[178] Full, W.E., Ehrlich, R., Klovan, J.: EXTENDED QMODEL–objective definition of external end members in the analysis of mixtures. Journal of the International Association for Mathematical Geology **13**(4), 331–344 (1981) (Cited on p. 138)

[179] Gabriel, K.R., Zamir, S.: Lower rank approximation of matrices by least squares with any choice of weights. Technometrics **21**(4), 489–498 (1979) (Cited on pp. 163, 203)

[180] Garey, M.R., Johnson, D.S.: Computers and Intractability. Freeman (2002) (Cited on pp. 51, 64)

[181] Gaujoux, R., Seoighe, C.: A flexible r package for nonnegative matrix factorization. BMC Bioinformatics **11**(1), 367 (2010) (Cited on p. 305)

[182] Gaussier, E., Goutte, C.: Relation between PLSA and NMF and implications. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 601–602. ACM (2005) (Cited on p. 191)

[183] Ge, R., Zou, J.: Intersecting faces: Non-negative matrix factorization with new guarantees. In: Proceedings of the 32nd International Conference on Machine Learning (ICML), pp. 2295–2303 (2015) (Cited on p. 155)

[184] Gillis, N.: Approximation et sous-approximation de matrices par factorisation positive: algorithmes, complexité et applications. Master's thesis, Université catholique de Louvain (2007) (Cited on p. 302)

[185] Gillis, N.: Nonnegative matrix factorization: Complexity, algorithms and applications. Ph.D. thesis, Université catholique de Louvain (2011) (Cited on pp. 75, 278, 284, 286)

[186] Gillis, N.: Sparse and unique nonnegative matrix factorization through data preprocessing. Journal of Machine Learning Research **13**(Nov), 3349–3386 (2012) (Cited on pp. 127, 129, 169, 176, 287)

[187] Gillis, N.: Robustness analysis of Hottopixx, a linear programming model for factoring nonnegative matrices. SIAM Journal on Matrix Analysis and Applications **34**(3), 1189–1212 (2013) (Cited on pp. 222, 253, 254, 255, 257)

[188] Gillis, N.: Successive nonnegative projection algorithm for robust nonnegative blind source separation. SIAM Journal on Imaging Sciences **7**(2), 1420–1450 (2014) (Cited on pp. 230, 232, 233, 234, 235, 236, 243, 257, 288)

[189] Gillis, N.: The why and how of nonnegative matrix factorization. In: J. Suykens, M. Signoretto, A. Argyriou (eds.) Regularization, Optimization, Kernels, and Support Vector Machines, Machine Learning and Pattern Recognition, chap. 12, pp. 257–291. Chapman & Hall/CRC (2014) (Cited on pp. 1, 5, 298, 314)

[190] Gillis, N.: Introduction to nonnegative matrix factorization. SIAG/OPT Views and News **25**(1), 7–16 (2017) (Cited on pp. 1, 29, 35, 314)

[191] Gillis, N.: Learning with nonnegative matrix factorizations. SIAM News **52**(5), 1–3 (2019), `https://sinews.siam.org/Portals/Sinews2/Issue%20Pdfs/sn_June2019.pdf` (Cited on pp. 1, 8, 11)

[192] Gillis, N.: Separable simplex-structured matrix factorization: Robustness of combinatorial approaches. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5521–5525. IEEE (2019) (Cited on pp. 213, 217, 218, 219, 220, 257)

[193] Gillis, N.: Successive projection algorithm robust to outliers. In: IEEE International Workshop on Computational Advances in Multi-sensor Adaptive Processing (CAMSAP 2019), pp. 331–335. IEEE (2019) (Cited on p. 228)

[194] Gillis, N., Glineur, F.: Nonnegative factorization and the maximum edge biclique problem. arXiv preprint arXiv:0810.4225 (2008) (Cited on pp. 166, 198, 265, 277)

[195] Gillis, N., Glineur, F.: Using underapproximations for sparse nonnegative matrix factorization. Pattern Recognition **43**(4), 1676–1687 (2010) (Cited on pp. 176, 302)

[196] Gillis, N., Glineur, F.: Low-rank matrix approximation with weights or missing data is NP-hard. SIAM Journal on Matrix Analysis and Applications **32**(4), 1149–1165 (2011) (Cited on pp. 203, 312)

[197] Gillis, N., Glineur, F.: Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization. Neural Computation **24**(4), 1085–1105 (2012) (Cited on pp. 278, 279, 284, 287, 292, 294)

[198] Gillis, N., Glineur, F.: A multilevel approach for nonnegative matrix factorization. Journal of Computational and Applied Mathematics **236**(7), 1708–1723 (2012) (Cited on p. 292)

[199] Gillis, N., Glineur, F.: On the geometric interpretation of the nonnegative rank. Linear Algebra and Its Applications **437**(11), 2685–2712 (2012) (Cited on pp. 35, 36, 37, 42, 65, 73, 74, 75, 82, 85, 89, 91)

[200] Gillis, N., Glineur, F.: A continuous characterization of the maximum-edge biclique problem. Journal of Global Optimization **58**(3), 439–464 (2014) (Cited on p. 198)

[201] Gillis, N., Hien, L.T.K., Leplat, V., Tan, V.Y.: Distributionally robust and multi-objective nonnegative matrix factorization. arXiv preprint arXiv:1901.10757 (2019) (Cited on pp. 163, 164, 273)

[202] Gillis, N., Kuang, D., Park, H.: Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization. IEEE Transactions on Geoscience and Remote Sensing **53**(4), 2066–2078 (2014) (Cited on pp. 199, 303)

[203] Gillis, N., Kumar, A.: Exact and heuristic algorithms for semi-nonnegative matrix factorization. SIAM Journal on Matrix Analysis and Applications **36**(4), 1404–1424 (2015) (Cited on pp. 135, 173, 174)

[204] Gillis, N., Luce, R.: Robust near-separable nonnegative matrix factorization using linear optimization. Journal of Machine Learning Research **15**(1), 1249–1280 (2014) (Cited on pp. 243, 254, 255, 256, 257)

[205] Gillis, N., Luce, R.: A fast gradient method for nonnegative sparse regression with self-dictionary. IEEE Transactions on Image Processing **27**(1), 24–37 (2017) (Cited on pp. 243, 255, 256, 257)

[206] Gillis, N., Ma, W.-K.: Enhancing pure-pixel identification performance via preconditioning. SIAM Journal on Imaging Sciences **8**(2), 1161–1186 (2015) (Cited on pp. 237, 241, 257)

[207] Gillis, N., Ohib, R., Plis, S., Potluru, V.: Grouped sparse projection. arXiv preprint arXiv:1912.03896 (2019) (Cited on p. 176)

[208] Gillis, N., Plemmons, R.J.: Dimensionality reduction, classification, and spectral mixture analysis using non-negative underapproximation. Optical Engineering **50**(2), 027,001 (2011) (Cited on p. 302)

[209] Gillis, N., Shitov, Y.: Low-rank matrix approximation in the infinity norm. Linear Algebra and Its Applications **581**, 367–382 (2019) (Cited on pp. 162, 163, 202)

[210] Gillis, N., Vavasis, S.A.: Fast and robust recursive algorithms for separable nonnegative matrix factorization. IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(4), 698–714 (2013) (Cited on pp. 223, 226, 229, 231, 248, 257)

[211] Gillis, N., Vavasis, S.A.: Semidefinite programming based preconditioning for more robust near-separable nonnegative matrix factorization. SIAM Journal on Optimization **25**(1), 677–698 (2015) (Cited on pp. 231, 238, 239, 240, 257)

[212] Gillis, N., Vavasis, S.A.: On the complexity of robust PCA and ℓ_1-norm low-rank matrix approximation. Mathematics of Operations Research **43**(4), 1072–1084 (2018) (Cited on pp. 3, 162, 204)

[213] Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences **99**(12), 7821–7826 (2002) (Cited on p. 179)

[214] Glineur, F.: Computational experiments with a linear approximation of second order cone optimization (2000). Image Technical Report 0001, Service de Mathématique et de Recherche Opérationnelle, Faculté Polytechnique de Mons (Cited on p. 90)

[215] Goemans, M.X.: Smallest compact formulation for the permutahedron. Mathematical Programming **153**(1), 5–11 (2015) (Cited on pp. 65, 66, 84)

[216] Golub, G.H., Van Loan, C.: Matrix Computations, 4th ed. Johns Hopkins University Press (2013) (Cited on pp. 2, 196)

[217] Gouveia, J., Parrilo, P.A., Thomas, R.R.: Lifts of convex sets and cone factorizations. Mathematics of Operations Research **38**(2), 248–264 (2013) (Cited on p. 93)

[218] Gouvert, O., Oberlin, T., Févotte, C.: Ordinal non-negative matrix factorization for recommendation. In: Proceedings of the 37th International Conference on Machine learning (ICML) (2020) (Cited on p. 185)

[219] Gribling, S., de Laat, D., Laurent, M.: Lower bounds on matrix factorization ranks via noncommutative polynomial optimization. Foundations of Computational Mathematics **19**, 1013–1070 (2019) (Cited on p. 82)

[220] Gribonval, R., Jenatton, R., Bach, F.: Sparse and spurious: dictionary learning with noise and outliers. IEEE Transactions on Information Theory **61**(11), 6298–6319 (2015) (Cited on p. 150)

[221] Gribonval, R., Zibulevsky, M.: Sparse component analysis. In: Handbook of Blind Source Separation, pp. 367–420. Elsevier (2010) (Cited on p. 3)

[222] Grippo, L., Sciandrone, M.: On the convergence of the block nonlinear Gauss–Seidel method under convex constraints. Operations Research Letters **26**(3), 127–136 (2000) (Cited on p. 268)

[223] Groemer, H.: On some mean values associated with a randomly selected simplex in a convex set. Pacific Journal of Mathematics **45**(2), 525–533 (1973) (Cited on p. 62)

[224] Grussler, C., Rantzer, A.: On optimal low-rank approximation of non-negative matrices. In: 2015 54th IEEE Conference on Decision and Control (CDC), pp. 5278–5283 (2015) (Cited on p. 100)

[225] Guan, N., Tao, D., Luo, Z., Yuan, B.: NeNMF: An optimal gradient method for nonnegative matrix factorization. IEEE Transactions on Signal Processing **60**(6), 2882–2898 (2012) (Cited on pp. 288, 294)

[226] Guglielmi, N., Scalone, C.: An efficient method for non-negative low-rank completion. Advances in Computational Mathematics **46** (2020). Article number: 31 (Cited on p. 311)

[227] Guillamet, D., Vitrià, J.: Non-negative matrix factorization for face recognition. In: Catalonian Conference on Artificial Intelligence, pp. 336–344. Springer (2002) (Cited on p. 101)

[228] Guo, Z., Zhang, S.: Sparse deep nonnegative matrix factorization. Big Data Mining and Analytics, **3**(1), 13–28 (2019) (Cited on p. 184)

[229] Haeffele, B.D., Vidal, R.: Structured low-rank matrix factorization: Global optimality, algorithms, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(6) (2019) (Cited on p. 197)

[230] Hajinezhad, D., Chang, T.H., Wang, X., Shi, Q., Hong, M.: Nonnegative matrix factorization using ADMM: Algorithm and convergence analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4742–4746. IEEE (2016) (Cited on p. 291)

[231] Hannah, J., Laffey, T.J.: Nonnegative factorization of completely positive matrices. Linear Algebra and Its Applications **55**, 1–9 (1983) (Cited on p. 79)

[232] Hasinoff, S.W.: Photon, Poisson noise. (2014). `http://people.csail.mit.edu/hasinoff/pubs/hasinoff-photon-2011-preprint.pdf` (Cited on p. 161)

[233] Hautecoeur, C., Glineur, F.: Image completion via nonnegative matrix factorization using HALS and B-splines. In: 28th European Symposium on Artificial Neural Networks-Computational Intelligence and Machine Learning (ESANN) (2020) (Cited on p. 313)

[234] Hautecoeur, C., Glineur, F.: Nonnegative matrix factorization over continuous signals using parametrizable functions. Neurocomputing (2020) (Cited on pp. 169, 309)

[235] He, Z., Xie, S., Zdunek, R., Zhou, G., Cichocki, A.: Symmetric nonnegative matrix factorization: Algorithms and applications to probabilistic clustering. IEEE Transactions on Neural Networks **22**(12), 2117–2131 (2011) (Cited on p. 179)

[236] Hendrix, E.M., Garcia, I., Plaza, J., Martin, G., Plaza, A.: A new minimum-volume enclosing algorithm for endmember identification and abundance estimation in hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing **50**(7), 2744–2757 (2011) (Cited on p. 138)

[237] Hien, L.T.K., Gillis, N.: Algorithms for nonnegative matrix factorization with the Kullback-Leibler divergence. arXiv:2010.01935 (2020) (Cited on pp. 270, 271, 305)

[238] Hien, L.T.K., Gillis, N., Patrinos, P.: Inertial block proximal methods for non-convex non-smooth optimization. In: Proceedings of the 37th International Conference on Machine learning (ICML) (2020) (Cited on pp. 266, 293, 304)

[239] Ho, N.D.: Nonnegative matrix factorization algorithms and applications. Ph.D. thesis, Université catholique de Louvain (2008) (Cited on pp. 196, 203, 265, 286, 287, 299)

[240] Ho, N.D., Van Dooren, P.: Non-negative matrix factorization with fixed row and column sums. Linear Algebra and Its Applications **429**(5–6), 1020–1025 (2008) (Cited on pp. 200, 201)

[241] Hong, D., Kolda, T.G., Duersch, J.A.: Generalized canonical polyadic tensor decomposition. SIAM Review **62**(1), 133–163 (2020) (Cited on p. 163)

[242] Hong, M., Razaviyayn, M., Luo, Z.Q., Pang, J.S.: A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing. IEEE Signal Processing Magazine **33**(1), 57–77 (2015) (Cited on pp. 268, 269, 270, 277)

[243] Hoyer, P.O.: Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research **5**(Nov), 1457–1469 (2004) (Cited on pp. 150, 154, 175)

[244] Hrubeš, P.: On the nonnegative rank of distance matrices. Information Processing Letters **112**(11), 457–461 (2012) (Cited on p. 82)

[245] Hsieh, C.J., Dhillon, I.S.: Fast coordinate descent methods with variable selection for non-negative matrix factorization. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1064–1072 (2011) (Cited on p. 285)

[246] Huang, K., Fu, X.: Detecting overlapping and correlated communities without pure nodes: Identifiability and algorithm. In: Proceedings of the 36th International Conference on Machine Learning, pp. 2859–2868 (2019) (Cited on pp. 139, 184, 210, 308, 313)

[247] Huang, K., Fu, X.: Low-complexity proximal Gauss-Newton algorithm for nonnegative matrix factorization. In: 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (2019) (Cited on p. 304)

[248] Huang, K., Sidiropoulos, N.D.: Putting nonnegative matrix factorization to the test: a tutorial derivation of pertinent Cramer–Rao bounds and performance benchmarking. IEEE Signal Processing Magazine **31**(3), 76–86 (2014) (Cited on p. 304)

[249] Huang, K., Sidiropoulos, N.D.: Kullback-Leibler principal component for tensors is not NP-hard. In: 2017 51st Asilomar Conference on Signals, Systems, and Computers, pp. 693–697. IEEE (2017) (Cited on p. 201)

[250] Huang, K., Sidiropoulos, N.D., Liavas, A.P.: A flexible and efficient algorithmic framework for constrained matrix and tensor factorization. IEEE Transactions on Signal Processing **64**(19), 5052–5065 (2016) (Cited on pp. 290, 291, 294, 296)

[251] Huang, K., Sidiropoulos, N.D., Swami, A.: Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. IEEE Transactions on Signal Processing **62**(1), 211–224 (2013) (Cited on pp. 53, 108, 111, 112, 116, 118, 122, 125, 127, 133)

[252] Ibrahim, S., Fu, X.: Recovering joint probability of discrete random variables from pairwise marginals. arXiv preprint arXiv:2006.16912 (2020) (Cited on p. 211)

[253] Ibrahim, S., Fu, X., Kargas, N., Huang, K.: Crowdsourcing via pairwise co-occurrences: Identifiability and algorithms. In: Advances in Neural Information Processing Systems (NIPS), pp. 7845–7855 (2019) (Cited on p. 210)

[254] Imbrie, J., Van Andel, T.H.: Vector analysis of heavy-mineral data. Geological Society of America Bulletin **75**(11), 1131–1156 (1964) (Cited on p. 12)

[255] Iordache, M.D., Bioucas-Dias, J.M., Plaza, A.: Total variation spatial regularization for sparse hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing **50**(11), 4484–4502 (2012) (Cited on p. 170)

[256] Iordache, M.D., Bioucas-Dias, J.M., Plaza, A.: Collaborative sparse regression for hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing **52**(1), 341–354 (2013) (Cited on p. 172)

[257] Jameson, G.J.O.: Summing and Nuclear Norms in Banach Space Theory, vol. 8. Cambridge University Press (1987) (Cited on p. 79)

[258] Janecek, A.G.K., Grotthoff, S.S., Gansterer, W.N.: libNMF—a library for nonnegative matrix factorization. Computing and Informatics **30**(2), 205–224 (2011) (Cited on p. 305)

[259] Jang, B., Hero, A.: Minimum volume topic modeling. In: 22nd International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 3013–3021 (2019) (Cited on p. 139)

[260] Janzamin, M., Ge, R., Kossaifi, J., Anandkumar, A., et al.: Spectral learning on matrices and tensors. Foundations and Trends in Machine Learning **12**(5–6), 393–536 (2019) (Cited on p. 183)

[261] Jeter, M., Pye, W.: A note on nonnegative rank factorizations. Linear Algebra and Its Applications **38**, 171–173 (1981) (Cited on p. 14)

[262] Jia, S., Qian, Y.: Constrained nonnegative matrix factorization for hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing **47**(1), 161–173 (2008) (Cited on p. 169)

[263] Jiang, J.H., Liang, Y., Ozaki, Y.: Principles and methodologies in self-modeling curve resolution. Chemometrics and Intelligent Laboratory Systems **71**(1), 1–12 (2004) (Cited on pp. 5, 12)

[264] Jiang, J.H., Liang, Y.Z., Ozaki, Y.: On simplex-based method for self-modeling curve resolution of two-way data. Chemometrics and Intelligent Laboratory Systems **65**(1), 51–65 (2003) (Cited on p. 229)

[265] John, F.: Extremum problems with inequalities as subsidiary conditions. In: J. Moser (ed.) Fritz John, Collected Papers, pp. 543–560. Birkhaüser (1985). First published in 1948 (Cited on p. 240)

[266] Jolliffe, I.: Principal Component Analysis. Springer (2011) (Cited on p. 2)

[267] Kaibel, V.: Extended formulations in combinatorial optimization. Optima **85**, 2–7 (2011) (Cited on p. 84)

[268] Kaibel, V., Pashkovich, K.: Constructing extended formulations from reflection relations. In: Integer Programming and Combinatorial Optimization, pp. 287–300. Springer (2011) (Cited on p. 90)

[269] Kaibel, V., Pashkovich, K.: Constructing extended formulations from reflection relations. In: M. Jünger, G. Reinelt (eds.) Facets of Combinatorial Optimization, pp. 77–100. Springer (2013) (Cited on p. 90)

[270] Kaibel, V., Pashkovich, K., Theis, D.O.: Symmetry matters for sizes of extended formulations. SIAM Journal on Discrete Mathematics **26**(3), 1361–1382 (2012) (Cited on p. 15)

[271] Kalofolias, V., Gallopoulos, E.: Computing symmetric nonnegative rank factorizations. Linear Algebra and Its Applications **436**(2), 421–435 (2012) (Cited on pp. 179, 258)

[272] Kannan, R., Ishteva, M., Park, H.: Bounded matrix factorization for recommender system. Knowledge and Information Systems **39**(3), 491–511 (2014) (Cited on p. 313)

[273] Ke, Q., Kanade, T.: Robust l_1 norm factorization in the presence of outliers and missing data by alternative convex programming. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 739–746. IEEE (2005) (Cited on pp. 162, 163, 203)

[274] Kervazo, C., Bobin, J., Chenot, C., Sureau, F.: Use of palm for ℓ_1 sparse matrix factorization: Difficulty and rationalization of a two-step approach. Digital Signal Processing, 102611 (2019) (Cited on p. 170)

[275] Kervazo, C., Gillis, N., Dobigeon, N.: Provably robust blind source separation of linear-quadratic near-separable mixtures. In preparation. (Cited on p. 185)

[276] Keshava, N., Mustard, J.F.: Spectral unmixing. IEEE Signal Processing Magazine **19**(1), 44–57 (2002) (Cited on p. 13)

[277] Kim, H., Park, H.: Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics **23**(12), 1495–1502 (2007) (Cited on pp. 150, 154, 168, 281, 282, 309)

[278] Kim, H., Park, H.: Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. SIAM Journal on Matrix Analysis and Applications **30**(2), 713–730 (2008) (Cited on p. 281)

[279] Kim, J., He, Y., Park, H.: Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework. Journal of Global Optimization **58**(2), 285–319 (2014) (Cited on p. 284)

[280] Kim, J., Park, H.: Fast nonnegative matrix factorization: An active-set-like method and comparisons. SIAM Journal on Scientific Computing **33**(6), 3261–3281 (2011) (Cited on pp. 281, 282, 294)

[281] Kingman, J.F.: Random secants of a convex body. Journal of Applied Probability **6**(3), 660–672 (1969) (Cited on p. 62)

[282] Kohjima, M., Matsubayashi, T., Toda, H.: Generalized interval valued nonnegative matrix factorization. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3412–3416. IEEE (2019) (Cited on p. 185)

[283] Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Review **51**(3), 455–500 (2009) (Cited on p. xiv)

[284] Kopriva, I., Hadžija, M., Hadžija, M.P., Korolija, M., Cichocki, A.: Rational variety mapping for contrast-enhanced nonlinear unsupervised segmentation of multispectral images of unstained specimen. American Journal of Pathology **179**(2), 547–554 (2011) (Cited on p. 177)

[285] Kopriva, I., Jerić, I., Brkljačić, L.: Nonlinear mixture-wise expansion approach to underdetermined blind separation of nonnegative dependent sources. Journal of Chemometrics **27**(7–8), 189–197 (2013) (Cited on p. 185)

[286] Kopriva, I., Ju, W., Zhang, B., Shi, F., Xiang, D., Yu, K., Wang, X., Bagci, U., Chen, X.: Single-channel sparse non-negative blind source separation method for automatic 3-D delineation of lung tumor in PET images. IEEE Journal of Biomedical and Health Informatics **21**(6), 1656–1666 (2016) (Cited on p. 177)

[287] Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009) (Cited on pp. 3, 308, 311)

[288] Krone, R., Kubjas, K.: Uniqueness of nonnegative matrix factorizations by rigidity theory. SIAM Journal on Matrix Analysis and Applications. To appear (Cited on pp. 127, 128, 130, 132, 133)

[289] Kruskal, J.B.: Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra and Its Applications **18**(2), 95–138 (1977) (Cited on p. 151)

[290] Kuang, D., Ding, C., Park, H.: Symmetric nonnegative matrix factorization for graph clustering. In: Proceedings of the 12th SIAM International Conference on Data Mining, pp. 106–117. SIAM (2012) (Cited on pp. 179, 185, 287)

[291] Kuang, D., Park, H.: Fast rank-2 nonnegative matrix factorization for hierarchical document clustering. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining, pp. 739–747 (2013) (Cited on p. 303)

[292] Kubjas, K., Robeva, E., Sturmfels, B.: Fixed points EM algorithm and nonnegative rank boundaries. Annals of Statistics pp. 422–461 (2015) (Cited on pp. 14, 47)

[293] Kumar, A., Sindhwani, V.: Near-separable non-negative matrix factorization with ℓ_1 and Bregman loss functions. In: Proceedings of the 2015 SIAM International Conference on Data Mining, pp. 343–351. SIAM (2015) (Cited on p. 211)

[294] Kumar, A., Sindhwani, V., Kambadur, P.: Fast conical hull algorithms for near-separable nonnegative matrix factorization. In: Proceedings of the 30th International Conference on Machine Learning, pp. 231–239 (2013) (Cited on pp. 142, 189, 236)

[295] Kwan, M., Sauermann, L., Zhao, Y.: Extension complexity of low-dimensional polytopes. arXiv preprint arXiv:2006.08836 (2020) (Cited on p. 60)

[296] Lakshminarayanan, B., Raich, R.: Non-negative matrix factorization for parameter estimation in hidden Markov models. In: 2010 IEEE International Workshop on Machine Learning for Signal Processing, pp. 89–94. IEEE (2010) (Cited on p. 313)

[297] Lambert, J.H.: Photometria sive de mensura et gradibus luminis, colorum et umbrae. Klett (1760) (Cited on p. 12)

[298] Laurberg, H., Christensen, M.G., Plumbley, M.D., Hansen, L.K., Jensen, S.H.: Theorems on positive data: On the uniqueness of NMF. Computational Intelligence and Neuroscience **2008** (2008) (Cited on pp. 100, 104, 105, 107, 108, 111, 126, 127, 208)

[299] Laurberg, H., Hansen, L.K.: On affine non-negative matrix factorization. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 2, pp. II–653. IEEE (2007) (Cited on p. 176)

[300] Lawson, C.L., Hanson, R.J.: Solving Least Squares Problems. Classics in Applied Mathematics, vol. 15. SIAM (1995) (Cited on p. 281)

[301] Lawton, W.H., Sylvestre, E.A.: Self modeling curve resolution. Technometrics **13**(3), 617–633 (1971) (Cited on pp. 12, 103)

[302] Le Roux, J., Weninger, F.J., Hershey, J.R.: Sparse NMF—half-baked or well done? Mitsubishi Electric Research Labs (MERL), Tech. Rep. TR2015-023 **11**, 13–15 (2015) (Cited on p. 280)

[303] Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**, 788–791 (1999) (Cited on pp. 4, 5, 7, 16, 163, 270, 293, 294, 314)

[304] Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing Systems (NIPS), pp. 556–562 (2001) (Cited on pp. 16, 270)

[305] Lee, H., Yoo, J., Choi, S.: Semi-supervised nonnegative matrix factorization. IEEE Signal Processing Letters **17**(1), 4–7 (2009) (Cited on p. 170)

[306] Lee, J.R., Raghavendra, P., Steurer, D.: Lower bounds on the size of semidefinite programming relaxations. In: Proceedings of the 47th Annual ACM Symposium on Theory of Computing, pp. 567–576. ACM (2015) (Cited on p. 93)

[307] Lee, T., Friedman, L.: Nondeterministic communication complexity (2010). Lecture notes, `http://research.cs.rutgers.edu/~troyjlee/lec2.pdf` (Cited on p. 94)

[308] Lee, T., Shraibman, A.: Lower Bounds in Communication Complexity. Now Publishers (2009) (Cited on p. 95)

[309] Leplat, V., Ang, A.M., Gillis, N.: Minimum-volume rank-deficient nonnegative matrix factorizations. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3402–3406. IEEE (2019) (Cited on pp. 148, 149, 288)

[310] Leplat, V., Gillis, N., Ang, M.S.: Blind audio source separation with minimum-volume beta-divergence NMF. IEEE Transactions on Signal Processing **68**, 3400–3410 (2020) (Cited on pp. 143, 145, 280)

[311] Levin, B.: On calculating maximum rank one underapproximations for positive arrays (1985). http://biostats.bepress.com/cgi/viewcontent.cgi?article=1009&context=columbiabiostat (Cited on p. 302)

[312] Li, B., Zhou, G., Cichocki, A.: Two efficient algorithms for approximately orthogonal nonnegative matrix factorization. IEEE Signal Processing Letters **22**(7), 843–846 (2014) (Cited on p. 136)

[313] Li, L., Zhang, Y.J.: FastNMF: highly efficient monotonic fixed-point nonnegative matrix factorization algorithm with good applicability. J. Electron. Imaging **18**, 033004 (2009) (Cited on p. 286)

[314] Li, M.L., Di Mauro, F., Candan, K.S., Sapino, M.L.: Matrix factorization with interval-valued data. IEEE Transactions on Knowledge and Data Engineering (2019) (Cited on p. 185)

[315] Li, W., Li, J., Liu, X., Dong, L.: Two fast vector-wise update algorithms for orthogonal nonnegative matrix factorization with sparsity constraint. Journal of Computational and Applied Mathematics **375**, 112,785 (2020) (Cited on p. 287)

[316] Li, Y., Liang, Y., Risteski, A.: Recovery guarantee of non-negative matrix factorization via alternating updates. In: Advances in Neural Information Processing Systems (NIPS), pp. 4987–4995 (2016) (Cited on p. 156)

[317] Li, Y., Ngom, A.: The non-negative matrix factorization toolbox for biological data mining. Source Code for Biology and Medicine **8**(1), 10 (2013) (Cited on p. 305)

[318] Li, Y., Sima, D.M., Cauter, S.V., Croitor Sava, A.R., Himmelreich, U., Pi, Y., Van Huffel, S.: Hierarchical non-negative matrix factorization (hNMF): a tissue pattern differentiation method for glioblastoma multiforme diagnosis using MRSI. NMR in Biomedicine **26**(3), 307–319 (2013) (Cited on pp. 303, 313)

[319] Lin, C.H., Bioucas-Dias, J.M.: Nonnegative blind source separation for ill-conditioned mixtures via John ellipsoid. IEEE Transactions on Neural Networks and Learning Systems (2020) (Cited on p. 149)

[320] Lin, C.H., Ma, W.K., Li, W.C., Chi, C.Y., Ambikapathi, A.: Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case. IEEE Transactions on Geoscience and Remote Sensing **53**(10), 5530–5546 (2015) (Cited on pp. 111, 118, 119, 139, 141)

[321] Lin, C.-H., Wu, R., Ma, W.-K., Chi, C.-Y., Wang, Y.: Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization. SIAM Journal on Imaging Sciences **11**(2), 1651–1679 (2018) (Cited on pp. 149, 213)

[322] Lin, C.J.: On the convergence of multiplicative update algorithms for nonnegative matrix factorization. IEEE Transactions on Neural Networks **18**(6), 1589–1596 (2007) (Cited on p. 279)

[323] Lin, C.J.: Projected gradient methods for nonnegative matrix factorization. Neural Computation **19**(10), 2756–2779 (2007) (Cited on p. 287)

[324] Liu, J., Liu, J., Wonka, P., Ye, J.: Sparse non-negative tensor factorization using columnwise coordinate descent. Pattern Recognition **45**(1), 649–656 (2012) (Cited on p. 286)

[325] Liu, Z.: Model selection for nonnegative matrix factorization by support union recovery. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3407–3411. IEEE (2019) (Cited on p. 167)

[326] Liu, Z., Tan, V.Y.F.: Rank-one NMF-based initialization for NMF and relative error bounds under a geometric assumption. IEEE Transactions on Signal Processing **65**(18), 4717–4731 (2017) (Cited on p. 259)

[327] Lopes, M.B., Wolff, J.C., Bioucas-Dias, J.M., Figueiredo, M.A.: Near-infrared hyperspectral unmixing based on a minimum volume criterion for fast and accurate chemometric characterization of counterfeit tablets. Analytical Chemistry **82**(4), 1462–1469 (2010) (Cited on p. 139)

[328] López-Serrano, P., Dittmar, C., Özer, Y., Müller, M.: NMF Toolbox: Music processing applications of nonnegative matrix factorization. In: Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx-19), Birmingham, UK (2019) (Cited on p. 305)

[329] Lu, Z., Yang, Z., Oja, E.: Selecting β-divergence for nonnegative matrix factorization by score matching. In: International Conference on Artificial Neural Networks, pp. 419–426. Springer (2012) (Cited on pp. 163, 164)

[330] Lubell, D.: A short proof of Sperner's lemma. Journal of Combinatorial Theory **1**(2), 299 (1966) (Cited on p. 69)

[331] Luce, R., Hildebrandt, P., Kuhlmann, U., Liesen, J.: Using separable nonnegative matrix factorization techniques for the analysis of time-resolved Raman spectra. Applied Spectroscopy **70**(9), 1464–1475 (2016) (Cited on pp. 210, 308, 309, 313)

[332] Lucy, L.B.: An iterative technique for the rectification of observed distributions. Astronomical Journal **79**, 745 (1974) (Cited on p. 270)

[333] Luo, W., Gao, L., Plaza, A., Marinoni, A., Yang, B., Zhong, L., Gamba, P., Zhang, B.: A new algorithm for bilinear spectral unmixing of hyperspectral images using particle swarm optimization. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **9**(12), 5776–5790 (2016) (Cited on p. 247)

[334] Ma, W.K., Bioucas-Dias, J.M., Chan, T.H., Gillis, N., Gader, P., Plaza, A.J., Ambikapathi, A., Chi, C.Y.: A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. IEEE Signal Processing Magazine **31**(1), 67–81 (2014) (Cited on pp. 13, 209, 216, 247, 248)

[335] Machida, K., Takenouchi, T.: Statistical modeling of robust non-negative matrix factorization based on γ-divergence and its applications. Japanese Journal of Statistics and Data Science **2**(2), 441–464 (2019) (Cited on p. 163)

[336] Mackevicius, E.L., Bahle, A.H., Williams, A.H., Gu, S., Denisenko, N.I., Goldman, M.S., Fee, M.S.: Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. eLife **8**, e38,471 (2019) (Cited on pp. 178, 314)

[337] Magron, P., Badeau, R., David, B.: Phase reconstruction of spectrograms with linear unwrapping: application to audio signal restoration. In: 23rd European Signal Processing Conference (EUSIPCO). IEEE (2015) (Cited on p. 11)

[338] Magron, P., Badeau, R., David, B.: Phase recovery in NMF for audio source separation: an insightful benchmark. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 81–85. IEEE (2015) (Cited on p. 11)

[339] Majmudar, J., Vavasis, S.A.: Provable overlapping community detection in weighted graphs. arXiv preprint arXiv:2004.07150 (2020) (Cited on p. 210)

[340] Mao, X., Sarkar, P., Chakrabarti, D.: On mixed memberships and symmetric nonnegative matrix factorizations. In: Proceedings of the 34th International Conference on Machine Learning, pp. 2324–2333 (2017) (Cited on p. 184)

[341] Markovsky, I.: Low Rank Approximation: Algorithms, Implementation, Applications. Springer (2011) (Cited on p. 3)

[342] Megiddo, N.: Towards a genuinely polynomial algorithm for linear programming. SIAM Journal on Computing **12**(2), 347–353 (1983) (Cited on p. 202)

[343] Miao, L., Qi, H.: Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. IEEE Transactions on Geoscience and Remote Sensing **45**(3), 765–777 (2007) (Cited on pp. 138, 139)

[344] Miettinen, P., Mielikäinen, T., Gionis, A., Das, G., Mannila, H.: The discrete basis problem. IEEE Transactions on Knowledge and Data Engineering **20**(10), 1348–1362 (2008) (Cited on p. 185)

[345] Mitchell, J., Suri, S.: Separation and approximation of polyhedral surfaces. Operations Research Letters **11**, 255–259 (1992) (Cited on p. 51)

[346] Miyasawa, A., Fujimoto, Y., Hayashi, Y.: Energy disaggregation based on smart metering data via semi-binary nonnegative matrix factorization. Energy and Buildings **183**, 547–558 (2019) (Cited on p. 313)

[347] Mizutani, T.: Ellipsoidal rounding for nonnegative matrix factorization under noisy separability. Journal of Machine Learning Research **15**(1), 1011–1039 (2014) (Cited on pp. 237, 238, 239)

[348] Mizutani, T.: Robustness analysis of preconditioned successive projection algorithm for general form of separable NMF problem. Linear Algebra and Its Applications **497**, 1–22 (2016) (Cited on p. 239)

[349] Mizutani, T., Tanaka, M.: Efficient preconditioning for noisy separable nonnegative matrix factorization problems by successive projection based low-rank approximations. Machine Learning **107**(4), 643–673 (2018) (Cited on pp. 239, 241)

[350] Mohan, K., Fazel, M.: Iterative reweighted algorithms for matrix rank minimization. Journal of Machine Learning Research **13**(Nov), 3441–3473 (2012) (Cited on p. 148)

[351] Moitra, A.: An almost optimal algorithm for computing nonnegative rank. In: Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'13), pp. 1454–1464 (2013) (Cited on p. 53)

[352] Mollah, M.N.H., Eguchi, S., Minami, M.: Robust prewhitening for ICA by minimizing β-divergence and its application to fastica. Neural Processing Letters **25**(2), 91–110 (2007) (Cited on p. 164)

[353] Mond, D., Smith, J., Van Straten, D.: Stochastic factorizations, sandwiched simplices and the topology of the space of explanations. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **459**(2039), 2821–2845 (2003) (Cited on pp. 14, 29)

[354] Montanari, A., Richard, E.: Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. IEEE Transactions on Information Theory **62**(3), 1458–1484 (2015) (Cited on p. 188)

[355] Moussaoui, S., Brie, D., Idier, J.: Non-negative source separation: range of admissible solutions and conditions for the uniqueness of the solution. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 5, pp. v–289. IEEE (2005) (Cited on pp. 103, 127)

[356] Moutier, F., Vandaele, A., Gillis, N.: Off-diagonal symmetric nonnegative matrix factorization. arXiv preprint arXiv:2003.04775 (2020) (Cited on p. 179)

[357] Mukkamala, M.C., Ochs, P.: Beyond alternating updates for matrix factorization with inertial Bregman proximal gradient algorithms. In: Advances in Neural Information Processing Systems, pp. 4266–4276 (2019) (Cited on p. 262)

[358] Naanaa, W., Nuzillard, J.M.: Blind source separation of positive and partially correlated data. Signal Processing **85**(9), 1711–1722 (2005) (Cited on p. 249)

[359] Nadisic, N., Vandaele, A., Cohen, J.E., Gillis, N.: Sparse separable nonnegative matrix factorization. In: Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) (2020) (Cited on p. 185)

[360] Nascimento, J.M., Bioucas-Dias, J.M.: Hyperspectral unmixing based on mixtures of dirichlet components. IEEE Transactions on Geoscience and Remote Sensing **50**(3), 863–878 (2011) (Cited on pp. 144, 242)

[361] Nascimento, J.M., Dias, J.M.: Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing **43**(4), 898–910 (2005) (Cited on pp. 216, 232)

[362] Nascimento, J.M.P., Bioucas-Dias, J.M.: Nonlinear mixture model for hyperspectral unmixing. In: L. Bruzzone, C. Notarnicola, F. Posa (eds.) Image and Signal Processing for Remote Sensing XV, vol. 7477, pp. 157–164. International Society for Optics and Photonics, SPIE (2009) (Cited on p. 185)

[363] Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM Journal on Computing **24**(2), 227–234 (1995) (Cited on p. 205)

[364] Nesterov, Y.: Introductory Lectures on Convex Optimization: A basic course. Applied Optimization 137, 2nd ed. Springer (2018) (Cited on p. 288)

[365] Newman, D.J.: The double dixie cup problem. American Mathematical Monthly **67**(1), 58–61 (1960) (Cited on p. 122)

[366] Neymeyr, K., Sawall, M.: On the set of solutions of the nonnegative matrix factorization problem. SIAM Journal on Matrix Analysis and Applications **39**(2), 1049–1069 (2018) (Cited on pp. 12, 41, 53, 304, 309)

[367] Nus, L., Miron, S., Brie, D.: An ADMM-based algorithm with minimum dispersion regularization for on-line blind unmixing of hyperspectral images. Chemometrics and Intelligent Laboratory Systems **204**, 104,090 (2020) (Cited on p. 168)

[368] Oelze, M., Vandaele, A., Weltge, S.: Computing the extension complexities of all 4-dimensional 0/1-polytopes. arXiv preprint arXiv:1406.4895 (2014) (Cited on pp. 71, 72, 73)

[369] Ott, G.H.: Reconsider the state minimization problem for stochastic finite state systems. In: 7th Annual Symposium on Switching and Automata Theory (SWAT 1966), pp. 267–273. IEEE (1966) (Cited on p. 13)

[370] Ouedraogo, W.S.B., Souloumiac, A., Jaidane, M., Jutten, C.: Non-negative blind source separation algorithm based on minimum aperture simplicial cone. IEEE Transactions on Signal Processing **62**(2), 376–389 (2013) (Cited on p. 139)

[371] Paatero, P., Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics **5**(2), 111–126 (1994) (Cited on pp. 4, 15, 128, 159, 281, 304, 313)

[372] Packer, A.: NP-hardness of largest contained and smallest containing simplices for V-and H-polytopes. Discrete and Computational Geometry **28**(3), 349–377 (2002) (Cited on p. 148)

[373] Pan, J., Gillis, N.: Generalized separable nonnegative matrix factorization. IEEE Transactions on Pattern Analysis and Machine Intelligence (2019). DOI:10.1109/TPAMI.2019.2956046 (Cited on p. 259)

[374] Panov, M., Slavnov, K., Ushakov, R.: Consistent estimation of mixed memberships with successive projections. In: International Conference on Complex Networks and their Applications, pp. 53–64. Springer (2017) (Cited on pp. 184, 210)

[375] Pashkovich, K.: Extended formulations for combinatorial polytopes. Ph.D. thesis, Otto-von-Guericke-Universität Magdeburg (2012) (Cited on p. 84)

[376] Pashkovich, K.: Tight lower bounds on the sizes of symmetric extensions of permutahedra and similar results. Mathematics of Operations Research **39**(4), 1330–1339 (2014) (Cited on p. 84)

[377] Pauca, V.P., Piper, J., Plemmons, R.J.: Nonnegative matrix factorization for spectral data analysis. Linear Algebra and Its Applications **416**(1), 29–47 (2006) (Cited on p. 13)

[378] Paz, A.: Homomorphisms between stochastic sequential machines and related problems. Mathematical Systems Theory **2**(3), 223–245 (1968) (Cited on p. 13)

[379] Paz, A.: Introduction to Probabilistic Automata. Academic Press (1971) (Cited on p. 13)

[380] Peeters, R.: The maximum edge biclique problem is NP-complete. Discrete Applied Mathematics **131**(3), 651–654 (2003) (Cited on pp. 72, 198, 203)

[381] Peyerimhoff, N.: Areas and intersections in convex domains. American Mathematical Monthly **104**(8), 697–704 (1997) (Cited on p. 62)

[382] Plumbley, M., Cichocki, A., Bro, R.: Non-negative mixtures. In: Handbook of Blind Source Separation, pp. 515–547. Elsevier (2010) (Cited on pp. 193, 313, 314)

[383] Plumbley, M.D.: Algorithms for nonnegative independent component analysis. IEEE Transactions on Neural Networks **14**(3), 534–543 (2003) (Cited on p. 188)

[384] Pompili, F., Gillis, N., Absil, P.A., Glineur, F.: Two algorithms for orthogonal nonnegative matrix factorization with application to clustering. Neurocomputing **141**, 15–25 (2014) (Cited on pp. 137, 189, 282)

[385] Powell, M.J.: On search directions for minimization algorithms. Mathematical Programming **4**(1), 193–201 (1973) (Cited on p. 268)

[386] Psorakis, I., Roberts, S., Ebden, M., Sheldon, B.: Overlapping community detection using Bayesian non-negative matrix factorization. Physical Review E **83**(6), 066,114 (2011) (Cited on p. 313)

[387] Qian, Y., Jia, S., Zhou, J., Robles-Kelly, A.: Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization. IEEE Transactions on Geoscience and Remote Sensing **49**(11), 4282–4297 (2011) (Cited on p. 150)

[388] Qiao, H.: New SVD based initialization strategy for non-negative matrix factorization. Pattern Recognition Letters **63**, 71–77 (2015) (Cited on p. 301)

[389] Rajaraman, A., Ullman, J.D.: Mining of Massive Datasets. Cambridge University Press (2011) (Cited on p. 9)

[390] Rajkó, R., István, K.: Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry. Journal of Chemometrics **19**(8), 448–463 (2005) (Cited on p. 12)

[391] Rapin, J., Bobin, J., Larue, A., Starck, J.L.: Sparse and non-negative BSS for noisy data. IEEE Transactions on Signal Processing **61**(22), 5620–5632 (2013) (Cited on p. 170)

[392] Razaviyayn, M., Hong, M., Luo, Z.-Q.: A unified convergence analysis of block successive minimization methods for nonsmooth optimization. SIAM Journal on Optimization **23**(2), 1126–1153 (2013) (Cited on pp. 268, 269)

[393] Razenshteyn, I., Song, Z., Woodruff, D.P.: Weighted low rank approximations with provable guarantees. In: Proceedings of the 48th Symposium on Theory of Computing (STOC '16), pp. 250–263 (2016) (Cited on p. 203)

[394] Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Review **52**(3), 471–501 (2010) (Cited on pp. 77, 148)

[395] Recht, B., Re, C., Tropp, J., Bittorf, V.: Factoring nonnegative matrices with linear programs. In: Advances in Neural Information Processing Systems (NIPS), pp. 1214–1222 (2012) (Cited on pp. 253, 254, 256, 257)

[396] Ren, H., Chang, C.I.: Automatic spectral target recognition in hyperspectral imagery. IEEE Transactions on Aerospace and Electronic Systems **39**(4), 1232–1249 (2003) (Cited on p. 229)

[397] Richardson, W.H.: Bayesian-based iterative method of image restoration. JoSA **62**(1), 55–59 (1972) (Cited on p. 270)

[398] Ritov, Y., Gilula, Z.: Analysis of contingency tables by correspondence models subject to order constraints. Journal of the American Statistical Association **88**(424), 1380–1387 (1993) (Cited on p. 14)

[399] Rockafellar, R.T.: Convex Analysis. Princeton Landmarks in Mathematics 28. Princeton University Press (1970) (Cited on p. 106)

[400] Rothvoss, T.: The matching polytope has exponential extension complexity. In: Proceedings of the 46th Annual ACM Symposium on Theory of Computing, pp. 263–272. ACM (2014) (Cited on pp. 15, 76, 77, 83, 85, 97)

[401] Rothvoss, T.: The matching polytope has exponential extension complexity. Journal of the ACM **64**(6), 41 (2017) (Cited on pp. 15, 77, 86, 97)

[402] Roughgarden, T., et al.: Communication complexity (for algorithm designers). Foundations and Trends in Theoretical Computer Science **11**(3–4), 217–404 (2016) (Cited on p. 95)

[403] Ruckebusch, C., Blanchet, L.: Multivariate curve resolution: a review of advanced and tailored applications and challenges. Analytica Chimica Acta **765**, 28–36 (2013) (Cited on p. 12)

[404] Sadowski, T., Zdunek, R.: Image completion with smooth nonnegative matrix factorization. In: International Conference on Artificial Intelligence and Soft Computing, pp. 62–72. Springer (2018) (Cited on p. 313)

[405] Sandler, R., Lindenbaum, M.: Nonnegative matrix factorization with earth mover's distance metric. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1873–1880. IEEE (2009) (Cited on p. 163)

[406] Sauwen, N., Sima, D.M., Van Cauter, S., Veraart, J., Leemans, A., Maes, F., Himmelreich, U., Van Huffel, S.: Hierarchical non-negative matrix factorization to characterize brain tumor heterogeneity using multi-parametric MRI. NMR in Biomedicine **28**(12), 1599–1624 (2015) (Cited on pp. 303, 313)

[407] Schachtner, R., Pöppel, G., Tomé, A.M., Lang, E.W.: Minimum determinant constraint for nonnegative matrix factorization. In: T. Adali, C. Jutten, J.M.T. Romano, A.K. Barros (eds.) Independent Component Analysis and Signal Separation, pp. 106–113. Springer (2009) (Cited on p. 139)

[408] Scutari, G., Sun, Y.: Parallel and distributed successive convex approximation methods for big-data optimization. In: Multi-agent Optimization, pp. 141–308. Springer (2018) (Cited on p. 265)

[409] Shaked-Monderer, N.: On the number of CP factorizations of a completely positive matrix. arXiv: 2009.12290 (2020) (Cited on p. 133)

[410] Shitov, Y.: Sublinear extensions of polygons. arXiv preprint arXiv:1412.0728v1 (2014) (Cited on p. 82)

[411] Shitov, Y.: An upper bound for nonnegative rank. Journal of Combinatorial Theory, Series A **122**, 126–132 (2014) (Cited on pp. 82, 90)

[412] Shitov, Y.: Nonnegative rank depends on the field II. arXiv:1605.07173 (2016) (Cited on p. 54)

[413] Shitov, Y.: A universality theorem for nonnegative matrix factorizations. arXiv preprint arXiv: 1606.09068 (2016) (Cited on p. 54)

[414] Shitov, Y.: The nonnegative rank of a matrix: Hard problems, easy solutions. SIAM Review **59**(4), 794–800 (2017) (Cited on pp. 52, 54)

[415] Shitov, Y.: Euclidean distance matrices and separations in communication complexity theory. Discrete & Computational Geometry **61**(3), 653–660 (2019) (Cited on p. 93)

[416] Shitov, Y.: Nonnegative rank depends on the field. Mathematical Programming (2019), `doi.org/10.1007/s10107-019-01448-2` (Cited on p. 54)

[417] Shitov, Y.: Sublinear extensions of polygons. arXiv preprint arXiv:1412.0728v2 (2020) (Cited on pp. 82, 92)

[418] Sidiropoulos, N.D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E.E., Faloutsos, C.: Tensor decomposition for signal processing and machine learning. IEEE Transactions on Signal Processing **65**(13), 3551–3582 (2017) (Cited on p. xiv)

[419] Siewert, D.J.: Biclique covers and partitions of bipartite graphs and digraphs and related matrix ranks of $\{0, 1\}$-matrices. Ph.D. thesis, University of Colorado at Denver (2000) (Cited on p. 71)

[420] Silio, C.B.: An efficient simplex coverability algorithm in E2 with application to stochastic sequential machines. IEEE Transactions on Computers **28**(2), 109–120 (1979) (Cited on pp. xxi, 13, 44, 45, 46, 49)

[421] Simsekli, U., Cemgil, A.T., Yilmaz, Y.K.: Learning the beta-divergence in Tweedie compound Poisson matrix factorization models. In: Proceedings of the 30th International Conference on Machine Learning, pp. 1409–1417 (2013) (Cited on p. 163)

[422] Singh, A.P., Gordon, G.J.: A unified view of matrix factorization models. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 358–373. Springer (2008) (Cited on p. 3)

[423] Smaragdis, P.: Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs. In: International Conference on Independent Component Analysis and Signal Separation, pp. 494–499. Springer (2004) (Cited on p. 177)

[424] Smaragdis, P.: Convolutive speech bases and their application to supervised speech separation. IEEE Transactions on Audio, Speech, and Language Processing **15**(1), 1–12 (2007) (Cited on pp. 11, 177)

[425] Smaragdis, P., Févotte, C., Mysore, G.J., Mohammadiha, N., Hoffman, M.: Static and dynamic source separation using nonnegative factorizations: A unified view. IEEE Signal Processing Magazine **31**(3), 66–75 (2014) (Cited on pp. 126, 314)

[426] Smaragdis, P., Venkataramani, S.: A neural network alternative to non-negative audio models. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 86–90. IEEE (2017) (Cited on p. 193)

[427] Smith, R.B.: Introduction to hyperspectral imaging with TNTmips by micro images (1999), `https://www.microimages.com/documentation/Tutorials/hyprspec.pdf` (Cited on p. 7)

[428] Song, Z., Woodruff, D.P., Zhong, P.: Low rank approximation with entrywise ℓ_1-norm error. In: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, pp. 688–701. ACM (2017) (Cited on p. 204)

[429] Sperner, E.: Ein Satz über Untermengen einer endlichen Menge. Mathematische Zeitschrift **27**(1), 544–548 (1928) (Cited on p. 68)

[430] Spielman, D.A., Teng, S.H.: Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. Journal of the ACM **51**(3), 385–463 (2004) (Cited on p. 282)

[431] Spielman, D.A., Wang, H., Wright, J.: Exact recovery of sparsely-used dictionaries. In: Conference on Learning Theory, pp. 37.1–37.18 (2012) (Cited on p. 122)

[432] Squires, S., Prügel-Bennett, A., Niranjan, M.: Rank selection in nonnegative matrix factorization using minimum description length. Neural Computation **29**(8), 2164–2176 (2017) (Cited on p. 167)

[433] Sra, S., Dhillon, I.S.: Nonnegative matrix approximation: algorithms and applications. Computer Science Department, University of Texas at Austin (2006) (Cited on pp. 4, 5, 163, 314)

[434] Srebro, N., Jaakkola, T.: Weighted low-rank approximations. In: Proceedings of the 20th International Conference on Machine learning (ICML), vol. 3, pp. 720–727 (2003) (Cited on pp. 3, 203)

[435] Stoica, P., Selen, Y.: Model-order selection: a review of information criterion rules. IEEE Signal Processing Magazine **21**(4), 36–47 (2004) (Cited on pp. 167, 168)

[436] Stojanović, O., Kuhlmann, L., Pipa, G.: Predicting epileptic seizures using nonnegative matrix factorization. PloS One **15**(2), e0228,025 (2020) (Cited on p. 314)

[437] Strang, G.: Linear Algebra and Its Application, 3rd ed. MIT Press (1988) (Cited on p. 139)

[438] Sun, D.L., Fevotte, C.: Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6201–6205. IEEE (2014) (Cited on p. 291)

[439] Sun, K., Geng, X., Wang, P., Zhao, Y.: A fast endmember extraction algorithm based on gram determinant. IEEE Geoscience and Remote Sensing Letters **11**(6), 1124–1128 (2013) (Cited on pp. 227, 229)

[440] Sun, P., Freund, R.: Computation of minimum-volume covering ellipsoids. Operations Research **52**(5), 690–706 (2004) (Cited on p. 240)

[441] Sun, Y., Babu, P., Palomar, D.P.: Majorization-minimization algorithms in signal processing, communications, and machine learning. IEEE Transactions on Signal Processing **65**(3), 794–816 (2016) (Cited on p. 265)

[442] Sun, Y., Xin, J.: Underdetermined sparse blind source separation of nonnegative and partially overlapped data. SIAM Journal on Scientific Computing **33**(4), 2063–2094 (2011) (Cited on p. 135)

[443] Suppes, P., Zanotti, M.: When are probabilistic explanations possible? Synthese **48**(2), 191–199 (1981) (Cited on p. 14)

[444] Takahashi, N., Hibi, R.: Global convergence of modified multiplicative updates for nonnegative matrix factorization. Computational Optimization and Applications **57**(2), 417–440 (2014) (Cited on p. 277)

[445] Tan, V.Y.F., Févotte, C.: Automatic relevance determination in nonnegative matrix factorization with the β-divergence. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(7), 1592–1605 (2013) (Cited on p. 167)

[446] Teboulle, M., Vaisbourd, Y.: Novel proximal gradient methods for nonnegative matrix factorization with sparsity constraints. SIAM Journal on Imaging Sciences **13**(1), 381–421 (2020) (Cited on p. 304)

[447] Tepper, M., Sapiro, G.: Nonnegative matrix underapproximation for robust multiple model fitting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2059–2067 (2017) (Cited on pp. 177, 302)

[448] Theis, F.J., Stadlthanner, K., Tanaka, T.: First results on uniqueness of sparse non-negative matrix factorization. In: 13th European Signal Processing Conference (EUSIPCO). IEEE (2005) (Cited on p. 150)

[449] Thom, M., Rapp, M., Palm, G.: Efficient dictionary learning with sparseness-enforcing projections. International Journal of Computer Vision **114**(2–3), 168–194 (2015) (Cited on pp. 175, 176)

[450] Thomas, L.B.: Solution to Problem 73-14*: Rank factorization of nonnegative matrices (A. Berman). SIAM Review **16**(3), 393–394 (1974) (Cited on pp. 14, 24, 26, 27, 28, 43, 56, 63, 64, 77, 105)

[451] Thomas, R.R.: Spectrahedral lifts of convex sets. In: Proceedings of the International Congress of Mathematicians, pp. 3819–3842 (2018) (Cited on p. 93)

[452] Thurau, C., Kersting, K., Wahabzada, M., Bauckhage, C.: Descriptive matrix factorization for sustainability adopting the principle of opposites. Data Mining and Knowledge Discovery **24**(2), 325–354 (2012) (Cited on p. 168)

[453] Tikhonov, A.N.: Regularization of incorrectly posed problems. In: Soviet Mathematics Doklady, vol. 4, pp. 1624–1627 (1963) (Cited on p. 168)

[454] Tiwary, H.R., Weltge, S., Zenklusen, R.: Extension complexities of Cartesian products involving a pyramid, Information Processing Letters, **128**, 11–13 (2017). (Cited on p. 94)

[455] Tosic, I., Frossard, P.: Dictionary learning. IEEE Signal Processing Magazine **28**(2), 27–38 (2011) (Cited on p. 3)

[456] Trefethen, L.N., Bau III, D.: Numerical Linear Algebra. SIAM (1997) (Cited on pp. 196, 283)

[457] Udell, M., Horn, C., Zadeh, R., Boyd, S.: Generalized low rank models. Foundations and Trends in Machine Learning **9**(1), 1–118 (2016) (Cited on p. 3)

[458] Udell, M., Townsend, A.: Why are big data matrices approximately low rank? SIAM Journal on Mathematics of Data Science **1**(1), 144–160 (2019) (Cited on p. 3)

[459] Vandaele, A., Gillis, N., Glineur, F.: On the linear extension complexity of regular n-gons. Linear Algebra and Its Applications **521**, 217–239 (2017) (Cited on pp. xxi, 70, 72, 82, 87, 88, 90, 91)

[460] Vandaele, A., Gillis, N., Glineur, F., Tuyttens, D.: Heuristics for exact nonnegative matrix factorization. Journal of Global Optimization **65**(2), 369–400 (2016) (Cited on pp. 53, 88, 96, 132, 299)

[461] Vandaele, A., Gillis, N., Lei, Q., Zhong, K., Dhillon, I.: Efficient and non-convex coordinate descent for symmetric nonnegative matrix factorization. IEEE Transactions on Signal Processing **64**(21), 5571–5584 (2016) (Cited on p. 180)

[462] Vanderbeck, F., Wolsey, L.A.: Reformulation and decomposition of integer programs. In: 50 Years of Integer Programming 1958-2008, pp. 431–502. Springer (2010) (Cited on p. 84)

[463] Varol, E., Nejatbakhsh, A., McGrory, C.: Temporal Wasserstein non-negative matrix factorization for non-rigid motion segmentation and spatiotemporal deconvolution. arXiv preprint arXiv:1912.03463 (2019) (Cited on p. 185)

[464] Vavasis, S.: Nonlinear Optimization: Complexity Issues. Oxford University Press (1991) (Cited on p. 196)

[465] Vavasis, S.A.: On the complexity of nonnegative matrix factorization. SIAM Journal on Optimization **20**(3), 1364–1377 (2009) (Cited on pp. 14, 36, 51, 52, 104, 195, 304)

[466] Velegar, M., Erichson, N.B., Keller, C.A., Kutz, J.N.: Scalable diagnostics for global atmospheric chemistry using Ristretto library (version 1.0). Geoscientific Model Development **12**(4), 1525–1539 (2019) (Cited on p. 313)

[467] Vesselinov, V.V., Mudunuru, M.K., Ahmmed, B., Karra, S., Middleton, R.S.: Discovering signatures of hidden geothermal resources based on unsupervised learning. In: 45th Workshop on Geothermal Reservoir Engineering, Stanford University (2020) (Cited on p. 313)

[468] Vial, P.-H., Magron, P., Oberlin, T., Févotte, C.: Phase retrieval with Bregman divergences and application to audio signal recovery. arXiv:2010.00392 (2020) (Cited on p. 11)

[469] Vincent, E., Bertin, N., Badeau, R.: Adaptive harmonic spectral decomposition for multiple pitch estimation. IEEE Transactions on Audio, Speech, and Language Processing **18**(3), 528–537 (2010) (Cited on p. 164)

[470] Vincent, E., Bertin, N., Gribonval, R., Bimbot, F.: From blind to guided audio source separation: how models and side information can improve the separation of sound. IEEE Signal Processing Magazine **31**(3), 107–115 (2014) (Cited on p. 159)

[471] Virtanen, T.: Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. IEEE Transactions on Audio, Speech, and Language Processing **15**(3), 1066–1074 (2007) (Cited on pp. 11, 150)

[472] Wall, J.: Rank factorizations of positive operators. Linear and Multilinear Algebra **8**(2), 137–144 (1979) (Cited on p. 14)

[473] Wallace, R.M.: Analysis of absorption spectra of multicomponent systems. Journal of Physical Chemistry **64**(7), 899–901 (1960) (Cited on p. 12)

[474] Wang, C.: Finding minimal nested polygons. BIT Numerical Mathematics **31**, 230–236 (1991) (Cited on p. 44)

[475] Wang, H., Huang, H., Ding, C.: Simultaneous clustering of multi-type relational data via symmetric nonnegative matrix tri-factorization. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 279–284. ACM (2011) (Cited on p. 182)

[476] Wang, S., Chang, T.H., Cui, Y., Pang, J.S.: Clustering by orthogonal non-negative matrix factorization: a sequential non-convex penalty approach. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5576–5580. IEEE (2019) (Cited on p. 136)

[477] Wang, Y., Fu, T., Gao, M., Ding, S.: Performance of orthogonal matching pursuit for multiple measurement vectors with noise. In: 2013 IEEE China Summit and International Conference on Signal and Information Processing, pp. 67–71. IEEE (2013) (Cited on p. 252)

[478] Wang, Y.X., Zhang, Y.J.: Nonnegative matrix factorization: A comprehensive review. IEEE Transactions on Knowledge and Data Engineering **25**(6), 1336–1353 (2013) (Cited on pp. 5, 168, 193, 314)

[479] Wedderburn, J.H.M.: Lectures on Matrices. American Mathematical Society Colloquium Publications 17. American Mathematical Society (1934) (Cited on p. 53)

[480] Weiderer, P., Tomé, A.M., Lang, E.W.: Decomposing temperature time series with non-negative matrix factorization. arXiv preprint arXiv:1904.02217 (2019) (Cited on p. 313)

[481] Wild, S., Curry, J., Dougherty, A.: Improving non-negative matrix factorizations through structured initialization. Pattern Recognition **37**(11), 2217–2232 (2004) (Cited on p. 301)

[482] Winter, M.E.: N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In: Proceedings of the SPIE Conference on Imaging Spectrometry V, vol. 3753, pp. 266–276. International Society for Optics and Photonics (1999) (Cited on pp. 13, 248)

[483] Wolsey, L.: Using extended formulations in practice. Optima **85**, 7–9 (2011) (Cited on p. 84)

[484] Wu, X., Zhai, G.: Temporal psychovisual modulation: a new paradigm of information display [exploratory DSP]. IEEE Signal Processing Magazine **30**(1), 136–141 (2012) (Cited on p. 314)

[485] Xu, L., Yu, B., Zhang, Y.: An alternating direction and projection algorithm for structure-enforced matrix factorization. Computational Optimization and Applications **68**(2), 333–362 (2017) (Cited on p. 291)

[486] Xu, Y., Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM Journal on Imaging Sciences **6**(3), 1758–1789 (2013) (Cited on p. 293)

[487] Yang, H., Wen, W., Li, H.: Deephoyer: Learning sparser neural network with differentiable scale-invariant sparsity measures. arXiv preprint arXiv:1908.09979 (2019) (Cited on p. 176)

[488] Yang, J., Leskovec, J.: Overlapping community detection at scale: a nonnegative matrix factorization approach. In: Proceedings of the 6th ACM International Conference on Web Search and Data Mining, pp. 587–596. ACM (2013) (Cited on pp. 308, 313)

[489] Yang, Z., Oja, E.: Linear and nonlinear projective nonnegative matrix factorization. IEEE Transactions on Neural Networks **21**, 734–749 (2010) (Cited on p. 136)

[490] Yang, Z., Oja, E.: Linear and nonlinear projective nonnegative matrix factorization. IEEE Transactions on Neural Networks **21**(5), 734–749 (2010) (Cited on pp. 171, 172, 173, 280)

[491] Yang, Z., Oja, E.: Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization. IEEE Transactions on Neural Networks **22**(12), 1878–1891 (2011) (Cited on p. 276)

[492] Yannakakis, M.: Expressing combinatorial optimization problems by linear programs. In: Proceedings of the 20th Symposium on Theory of Computing (STOC '88), pp. 223–228 (1988) (Cited on pp. 15, 85)

[493] Yannakakis, M.: Expressing combinatorial optimization problems by linear programs. Journal of Computer and System Sciences **43**(3), 441–466 (1991) (Cited on pp. 15, 68, 85, 95)

[494] Yannakakis, M.: On extended LP formulations. Optima **85**, 9–10 (2011) (Cited on pp. 15, 66)

[495] Yoo, J., Choi, S.: Orthogonal nonnegative matrix factorization: multiplicative updates on Stiefel manifolds. In: Proceedings of the 9th International Conference on Intelligent Data Engineering and Automated Learning, pp. 140–147. Springer (2008) (Cited on pp. 136, 280)

[496] Yu, J., Zhou, G., Cichocki, A., Xie, S.: Learning the hierarchical parts of objects by deep non-smooth nonnegative matrix factorization. IEEE Access **6**, 58,096–58,105 (2018) (Cited on p. 184)

[497] Yuan, Z., Oja, E.: Projective nonnegative matrix factorization for image compression and feature extraction. In: Scandinavian Conference on Image Analysis, pp. 333–342. Springer (2005) (Cited on p. 171)

[498] Zachary, W.W.: An information flow model for conflict and fission in small groups. Journal of Anthropological Research **33**(4), 452–473 (1977) (Cited on p. 179)

[499] Zarabie, A.K., Das, S.: An l0-norm constrained non-negative matrix factorization algorithm for the simultaneous disaggregation of fixed and shiftable loads. arXiv preprint arXiv:1908.00142 (2019) (Cited on p. 313)

[500] Zdunek, R.: Approximation of feature vectors in nonnegative matrix factorization with gaussian radial basis functions. In: International Conference on Neural Information Processing, pp. 616–623. Springer (2012) (Cited on p. 172)

[501] Zhang, D., Zhou, Z.H., Chen, S.: Non-negative matrix factorization on kernels. In: Pacific Rim International Conference on Artificial Intelligence, pp. 404–412. Springer (2006) (Cited on p. 185)

[502] Zhang, Y.: An alternating direction algorithm for nonnegative matrix factorization. Rice University, Tech. Rep. (2010) (Cited on p. 291)

[503] Zhang, Y., Yeung, D.Y.: Overlapping community detection via bounded nonnegative matrix tri-factorization. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 606–614. ACM (2012) (Cited on p. 182)

[504] Zhang, Z., Li, T., Ding, C., Zhang, X.: Binary matrix factorization with applications. In: Seventh IEEE International Conference on Data Mining (ICDM 2007), pp. 391–400. IEEE (2007) (Cited on p. 185)

[505] Zhang, Z., Zha, H., Simon, H.: Low-rank approximations with sparse factors I: Basic algorithms and error analysis. SIAM Journal on Matrix Analysis and Applications **23**(3), 706–727 (2002) (Cited on p. 150)

[506] Zhang, Z., Zhang, Y, Zhang, L., Yan, S.: A Survey on concept factorization: From shallow to deep representation learning. arXiv:2007.15840 (2020) (Cited on p. 172)

[507] Zhao, R., Tan, V.Y.F.: Online nonnegative matrix factorization with outliers. IEEE Transactions on Signal Processing **65**(3), 555–570 (2017) (Cited on p. 185)

[508] Zhao, R., Tan, V.Y.F.: A unified convergence analysis of the multiplicative update algorithm for regularized nonnegative matrix factorization. IEEE Transactions on Signal Processing **66**(1), 129–138 (2018) (Cited on p. 277)

[509] Zhong, S., Ghosh, J.: Generative model-based document clustering: a comparative study. Knowledge and Information Systems **8**(3), 374–384 (2005) (Cited on p. 296)

[510] Zhong, Y., Zhao, L., Zhang, L.: An adaptive differential evolution endmember extraction algorithm for hyperspectral remote sensing imagery. IEEE Geoscience and Remote Sensing Letters **11**(6), 1061–1065 (2013) (Cited on p. 247)

[511] Zhou, G., Cichocki, A., Xie, S.: Fast nonnegative matrix/tensor factorization based on low-rank approximation. IEEE Transactions on Signal Processing **60**(6), 2928–2940 (2012) (Cited on p. 292)

[512] Zhou, G., Cichocki, A., Zhao, Q., Xie, S.: Nonnegative matrix and tensor factorizations: An algorithmic perspective. IEEE Signal Processing Magazine **31**(3), 54–65 (2014) (Cited on p. 314)

[513] Zhou, G., Xie, S., Yang, Z., Yang, J.M., He, Z.: Minimum-volume-constrained nonnegative matrix factorization: Enhanced ability of learning parts. IEEE Transactions on Neural Networks **22**(10), 1626–1637 (2011) (Cited on pp. 139, 143, 145)

[514] Zhu, F.: Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey. arXiv preprint arXiv:1708.05125 (2017) (Cited on p. 148)

[515] Zhu, Z., Li, Q., Tang, G., Wakin, M.B.: Global optimality in low-rank matrix optimization. IEEE Transactions on Signal Processing **66**(13), 3614–3628 (2018) (Cited on p. 197)

[516] Zhuang, L., Lin, C.H., Figueiredo, M.A., Bioucas-Dias, J.M.: Regularization parameter selection in minimum volume hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing **57**(12), 9858–9877 (2019) (Cited on pp. 138, 170)

[517] Ziegler, G.: Lectures on Polytopes. Springer (1995) (Cited on pp. 21, 74)

[518] Žitnik, M., Zupan, B.: NIMFA: A python library for nonnegative matrix factorization. Journal of Machine Learning Research **13**(1), 849–853 (2012) (Cited on p. 305)

Index

The letters "f" and "t" following page numbers indicate figures and tables, respectively.