Hey! Be Positive! Nonnegative Matrix Factorization

An gentle tour

Andersen Ang

Department of Combinatorics and Optimization,
University of Waterloo, Waterloo, Canada

ms$\mathbf{x}$ang@uwaterloo.ca, where $\mathbf{x} = \lfloor \pi \rfloor$
Homepage: angms.science

First draft: May 10, 2021
Last update: May 12, 2021

Where we are now

Nonnegative Matrix Factorization (NMF)

Given :

- A matrix $\mathbf{X} \in \mathbb{R}_{+}^{m \times n}$.
- A positive integer $r \in \mathbb{N}$.

Find :

Nonnegative Matrix Factorization (NMF)

Given :

- A matrix $\mathbf{X} \in \mathbb{R}_{+}^{m \times n}$.
- A positive integer $r \in \mathbb{N}$.

Find :

- Matrices $\mathbf{W} \in \mathbb{R}_{+}^{m \times r}, \mathbf{H} \in \mathbb{R}_{+}^{r \times n}$ s.t. $\mathbf{X} = \mathbf{WH}$.

Nonnegative Matrix Factorization (NMF)

Given :

- A matrix $\mathbf{X} \in \mathbb{R}_{+}^{m \times n}$.
- A positive integer $r \in \mathbb{N}$.

Find :

- Matrices $\mathbf{W} \in \mathbb{R}_{+}^{m \times r}, \mathbf{H} \in \mathbb{R}_{+}^{r \times n}$ s.t. $\mathbf{X} = \mathbf{WH}$.
- Constraint Satisfaction Problem

Nonnegative Matrix Factorization (NMF)

Given :

- A matrix $\mathbf{X} \in \mathbb{R}_{+}^{m \times n}$.
- A positive integer $r \in \mathbb{N}$.

Find :

- Matrices $\mathbf{W} \in \mathbb{R}_{+}^{m \times r}, \mathbf{H} \in \mathbb{R}_{+}^{r \times n}$ s.t. $\mathbf{X} = \mathbf{WH}$.
- Constraint Satisfaction Problem
- Important: everything is nonnegative.

Other notations: $\mathbf{WH}^\top, \mathbf{UV}, \mathbf{UV}^\top$.

Examples

Given: $M(i,j) = (i-j)^2$, Euclidean Distance Matrix.

- Trivial solution that is not interesting

$$\mathbf{M} = \begin{bmatrix} 0 & 1 & 4 & 9 & 16 & 25 \\ 1 & 0 & 1 & 4 & 9 & 16 \\ 4 & 1 & 0 & 1 & 4 & 9 \\ 9 & 4 & 1 & 0 & 1 & 4 \\ 16 & 9 & 4 & 1 & 0 & 1 \\ 25 & 16 & 9 & 4 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 4 & 9 & 16 & 25 \\ 1 & 0 & 1 & 4 & 9 & 16 \\ 4 & 1 & 0 & 1 & 4 & 9 \\ 9 & 4 & 1 & 0 & 1 & 4 \\ 16 & 9 & 4 & 1 & 0 & 1 \\ 25 & 16 & 9 & 4 & 1 & 0 \end{bmatrix}$$

Examples

Given: $M(i,j) = (i-j)^2$, Euclidean Distance Matrix.

- Trivial solution that is not interesting

$$\mathbf{M} = \begin{bmatrix} 0 & 1 & 4 & 9 & 16 & 25 \\ 1 & 0 & 1 & 4 & 9 & 16 \\ 4 & 1 & 0 & 1 & 4 & 9 \\ 9 & 4 & 1 & 0 & 1 & 4 \\ 16 & 9 & 4 & 1 & 0 & 1 \\ 25 & 16 & 9 & 4 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 4 & 9 & 16 & 25 \\ 1 & 0 & 1 & 4 & 9 & 16 \\ 4 & 1 & 0 & 1 & 4 & 9 \\ 9 & 4 & 1 & 0 & 1 & 4 \\ 16 & 9 & 4 & 1 & 0 & 1 \\ 25 & 16 & 9 & 4 & 1 & 0 \end{bmatrix}$$

- Non-trivial solution that is more interesting

$$\begin{bmatrix} 0 & 1 & 4 & 9 & 16 & 25 \\ 1 & 0 & 1 & 4 & 9 & 16 \\ 4 & 1 & 0 & 1 & 4 & 9 \\ 9 & 4 & 1 & 0 & 1 & 4 \\ 16 & 9 & 4 & 1 & 0 & 1 \\ 25 & 16 & 9 & 4 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 4 & 1 & 0 \\ 3 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 4 \\ 0 & 1 & 0 & 1 & 4 \\ 0 & 3 & 1 & 0 & 1 \\ 0 & 5 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 & 3 & 5 \\ 5 & 3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- Even more non-trivial fact: $\operatorname{rank}(\mathbf{M}) = 3 < 5 = \operatorname{rank}_+(\mathbf{M})$.

A bunch of problemsssssss

$$\begin{array}{rl} \text{Given} & \mathbf{M} \in \mathbb{R}_+^{m \times n},\ r \in \mathbb{N} \\ \text{Find} & \mathbf{W} \in \mathbb{R}_+^{m \times r},\ \mathbf{H} \in \mathbb{R}_+^{r \times n} \ \text{ s.t. } \ \mathbf{WH} = \mathbf{M} \end{array}$$

- Is such problem even solvable?
- How to tell if a pair $(\mathbf{M}, r)$ has a NMF?

A bunch of problemssssss

$$\begin{array}{rl} \text{Given} & \mathbf{M} \in \mathbb{R}_+^{m\times n},\ r \in \mathbb{N} \\ \text{Find} & \mathbf{W} \in \mathbb{R}_+^{m\times r},\ \mathbf{H} \in \mathbb{R}_+^{r\times n}\ \text{ s.t. } \mathbf{WH} = \mathbf{M} \end{array}$$

- Is such problem even solvable?
- How to tell if a pair $(\mathbf{M}, r)$ has a NMF?
- If the problem is solvable
  - Is the sol. unique?
  - When will the sol. be unique?

A bunch of problemssssss

$$\begin{array}{rl} \text{Given} & \mathbf{M} \in \mathbb{R}_+^{m \times n},\ r \in \mathbb{N} \\ \text{Find} & \mathbf{W} \in \mathbb{R}_+^{m \times r},\ \mathbf{H} \in \mathbb{R}_+^{r \times n} \ \text{ s.t. } \ \mathbf{WH} = \mathbf{M} \end{array}$$

- Is such problem even solvable?
- How to tell if a pair $(\mathbf{M}, r)$ has a NMF?
- If the problem is solvable
  - Is the sol. unique?
  - When will the sol. be unique?
- How to solve NMF?

A bunch of problemssssss

$$\begin{array}{rl} \text{Given} & \mathbf{M} \in \mathbb{R}_+^{m\times n},\ r \in \mathbb{N} \\ \text{Find} & \mathbf{W} \in \mathbb{R}_+^{m\times r},\ \mathbf{H} \in \mathbb{R}_+^{r\times n}\ \text{s.t.}\ \mathbf{WH} = \mathbf{M} \end{array}$$

- Is such problem even solvable?
- How to tell if a pair $(\mathbf{M}, r)$ has a NMF?
- If the problem is solvable
  - Is the sol. unique?
  - When will the sol. be unique?
- How to solve NMF?
- What if r is unknown? Given a $\mathbf{M} \in \mathbb{R}_+^{m\times n}$, how to find r such that $(\mathbf{M}, r)$ has a NMF?
  - How difficult it is to find r?

A bunch of problemssssss

$$\begin{array}{rl} \text{Given} & \mathbf{M} \in \mathbb{R}_+^{m \times n},\ r \in \mathbb{N} \\ \text{Find} & \mathbf{W} \in \mathbb{R}_+^{m \times r},\ \mathbf{H} \in \mathbb{R}_+^{r \times n}\ \text{s.t.}\ \mathbf{WH} = \mathbf{M} \end{array}$$

- Is such problem even solvable?
- How to tell if a pair $(\mathbf{M}, r)$ has a NMF?
- If the problem is solvable
  - Is the sol. unique?
  - When will the sol. be unique?
- How to solve NMF?
- What if r is unknown? Given a $\mathbf{M} \in \mathbb{R}_+^{m \times n}$, how to find r such that $(\mathbf{M}, r)$ has a NMF?
  - How difficult it is to find r?
- If $(\mathbf{M}, r)$ has no NMF, what's the "smallest perturbation" Δ such that $(\mathbf{M} + \Delta, r)$ has a NMF?
  - What's the "smallest perturbation" Δ such that $(\mathbf{M} + \Delta, r)$ with $\mathbf{M} + \Delta \geq \mathbf{0}$ has a NMF?

Exact NMF and Approximate NMF

Given $(\mathbf{M} \in \mathbb{R}_+^{m \times n},\ r \in \mathbb{N})$, find $\mathbf{W} \in \mathbb{R}_+^{m \times r},\ \mathbf{H} \in \mathbb{R}_+^{r \times n}$ such that

[1] Vavasis, "On the complexity of nonnegative matrix factorization", SIAM J. Optim. 2010

Exact NMF and Approximate NMF

Given $(\mathbf{M} \in \mathbb{R}_+^{m\times n},\ r \in \mathbb{N})$, find $\mathbf{W} \in \mathbb{R}_+^{m\times r},\ \mathbf{H} \in \mathbb{R}_+^{r\times n}$ such that

- **Exact NMF**: $\mathbf{WH} = \mathbf{M}$.
  - Vavasis[1] (From UWaterloo!):
    - it is equivalent to a problem in polyhedral combinatorics
    - it is NP-hard
    - a polynomial-time local search heuristic exists.

[1] Vavasis, "On the complexity of nonnegative matrix factorization", SIAM J. Optim. 2010

Exact NMF and Approximate NMF

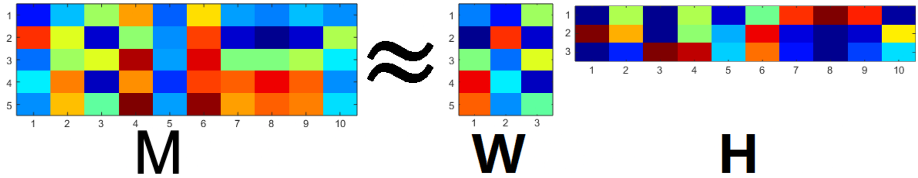
Given $(\mathbf{M} \in \mathbb{R}_+^{m\times n},\ r \in \mathbb{N})$, find $\mathbf{W} \in \mathbb{R}_+^{m\times r},\ \mathbf{H} \in \mathbb{R}_+^{r\times n}$ such that

- **Exact NMF**: $\mathbf{WH} = \mathbf{M}$.
  - Vavasis[1] (From UWaterloo!):
    - it is equivalent to a problem in polyhedral combinatorics
    - it is NP-hard
    - a polynomial-time local search heuristic exists.
- **Approximate NMF**: $\mathcal{D}_\xi\big(\mathbf{WH}, \mathbf{M}\big) \leq \epsilon$ under some "distance" ξ
  - i.e. $\mathbf{M} \approxeq \mathbf{WH}$
  - Example of $\mathcal{D}$
    - Frobenius norm $\|\mathbf{M} - \mathbf{WH}\|_F^2$
    - Kullback – Leibler divergence $[M]_{ij} \log \dfrac{[M]_{ij}}{[WH]_{ij}} - [M]_{ij} + [WH]_{ij}$
    - Itakura – Saito divergence $\dfrac{[M]_{ij}}{[WH]_{ij}} - \log \dfrac{[M]_{ij}}{[WH]_{ij}} - 1$

[1] Vavasis, "On the complexity of nonnegative matrix factorization", SIAM J. Optim. 2010

Low-rank approximate NMF

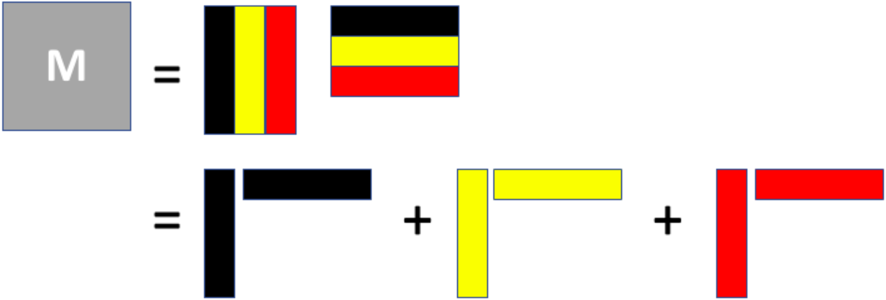
$$\mathbf{M} \approx \mathbf{W}\mathbf{H}, \; 1 \le r \le \min\{m, n\}.$$



- Usually seen in applications.
- r: rank of factorization.

What is r?

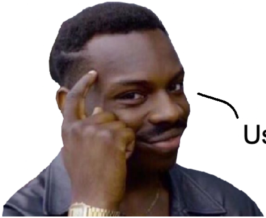
$$\mathbf{M} = \mathbf{W}\mathbf{H} = \sum_{i=1}^{r} \mathbf{w}_i \mathbf{h}^i.$$



- $\operatorname{rank}(\mathbf{M}) = r$: $\mathbf{M}$ is sum of r rank-1 matrices
- $\operatorname{rank}_+(\mathbf{M}) = r$: $\mathbf{M}$ is sum of r rank-1 nonnegative matrices
- $\operatorname{rank}_{\mathsf{cp}}(\mathbf{M}) = r$: $\mathbf{M}$ admits a rank-r NMF with $\mathbf{W} = \mathbf{H}^\top$
- Technique of lower bounding rank_+ has had a tremendous impact on the study of extended formulations.

Where we are now

Why NMF ???



Useful ! Fun !

The "fun": is NMF related to

- Linear Algebra
  - Nonnegative rank rank_+, cp-rank $\operatorname{rank}_{\mathsf{cp}}$
- Multilinear algebra (tensors)
- Discrete Optimization
  - Combinatorics: Extended formulations
  - Graph theory
- Continuous Optimization
  - Semidefinite Optimization
  - Structural nonsmooth nonconvex optimization
- Convex analysis
- Computational geometry
- Probability
- Communication complexity, Information Theory

So much fun!

The “Usefulness” of NMF

- Spectral unmixing in analytical chemistry (one of the earliest work)
- Representation learning on human face (the work that popularizes NMF)
- Topic modeling in text mining
- Geoscience and remote sensing: hypersepctral imaging
- Probability distribution application on identification of Hidden Markov Model
- Bioinformatics, gene expression
- Time-frequency matrix decompositions for neuroinformatics
- Audio blind source separation
- Stochastic sequential machines
- Data compression
- Data clustering
- Sparse coding, dictionary learning
- Community detection, social network
- Legal document analysis
- Identification of low-dimensional features within large-scale neural record ings
- Speech denoising, noise reduction
- Natural language processing
- Recommender system and collaborative filtering
- Face recognition
- Video summarization
- Medical imaging – image processing on small object
- Mid-infrared astronomy – image processing on large object
- Prediction of epileptic seizures using electroencephalographic
- Non-intrusive appliance load monitoring in energy disaggregation
- Forensics
- Art work conservation (identify true color used in painting)
- Computation of the temporal psycho-visual modulation
- Discovering of the signatures characterizing geothermal resources
- Spatio-temporal atmospheric chemistry
- Food processing: tells whether a banana or a fish is healthy
- Gas chromatography on Belgian beer
- Image annotation

list is too long

Analysis of Legal Documents via Non-negative Matrix Factorization Methods

Ryan Budahazy[1], Lu Cheng[2], Yihuan Huang[2], Andrew Johnson[3], Pengyu Li[2], Joshua Vendrow[2], Zhoutong Wu[2], Denali Molitor[2], Elizaveta Rebrova[2], Deanna Needell[2]

April 30, 2021 Not long ago!

Abstract

The California Innocence Project (CIP), a clinical law school program aiming to free wrongfully convicted prisoners, evaluates thousands of mails containing new requests for assistance and corresponding case files. Processing and interpreting this large amount of information presents a significant challenge for CIP officials, which can be successfully aided by topic modeling techniques. In this paper, we apply Non-negative Matrix Factorization (NMF) method and implement various offshoots of it to the important and previously unstudied data set compiled by CIP. We identify underlying topics of existing case files and classify request files by crime type and case status (decision type). The results uncover the semantic

better than lawyer ... ?

For non-NMF people : why NMF?

- Interpretability

  NMF beats similar tools (PCA, SVD, ICA) due to the interpretability on nonnegative data.

- Model correctness

  You can find the underlying hidden structure of the data (under some conditions).

- Mathematical curiosity ♡♡♡

  You find it fun.

- ~~Your boss tell you to do it.~~

The scope of this talk: a gentle tour of NMF

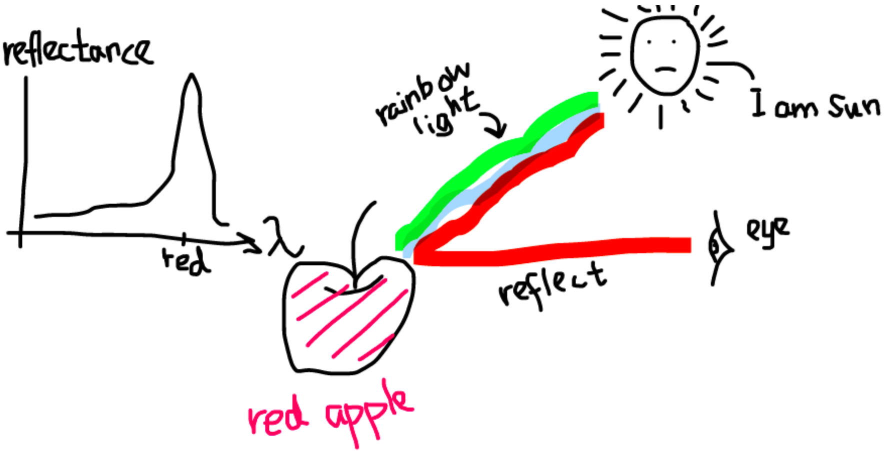
- Applications
- Geometry $\rightarrow$ Separable NMF $\rightarrow$ Minimum volume NMF
- Structural nonsmooth nonconvex optimization

Ideas only, no crazy maths in this talk.

Where we are now

Where we are now

High school optics: how you see



Remote sensing : satellite taking pictures of the Earth

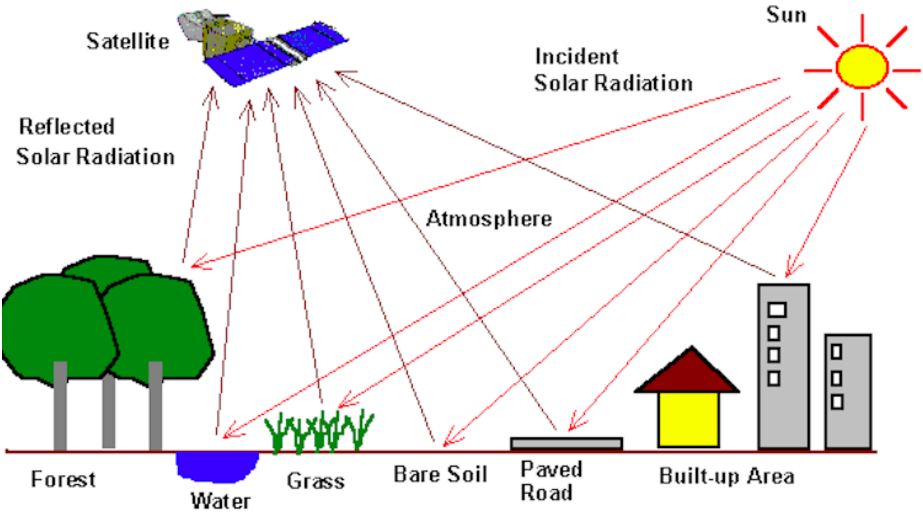


Figure: Image source: internet.

Hyperspectral imaging

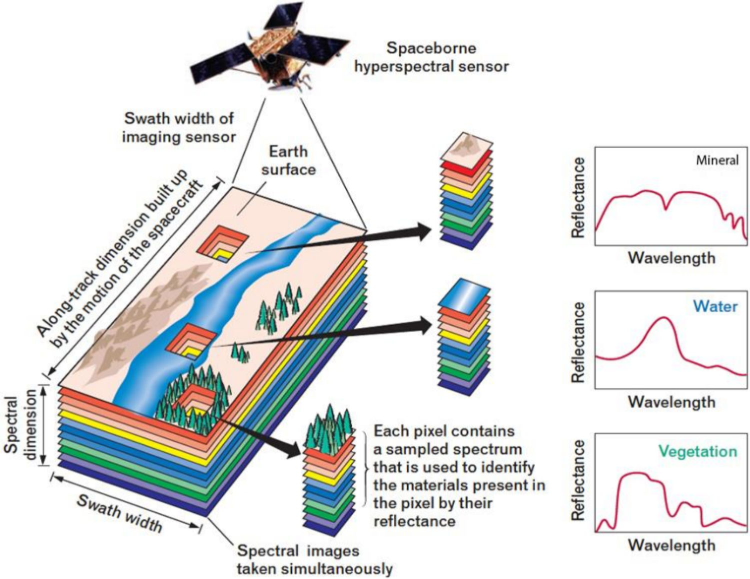


Figure: Image source: internet.

Decomposition of Hyperspectral image (1/2)

$$\underbrace{X(:,j)}_{\substack{\text{spectral signature}\\ \text{of } j\text{th pixel}}} \approx \sum_{k=1}^{r} \underbrace{W(:,k)}_{\substack{\text{spectral signature}\\ \text{of } k\text{th endmember}}} \quad \underbrace{H(k,j)}_{\substack{\text{abundance of } k\text{th endmember}\\ \text{in } j\text{th pixel}}} \quad .$$

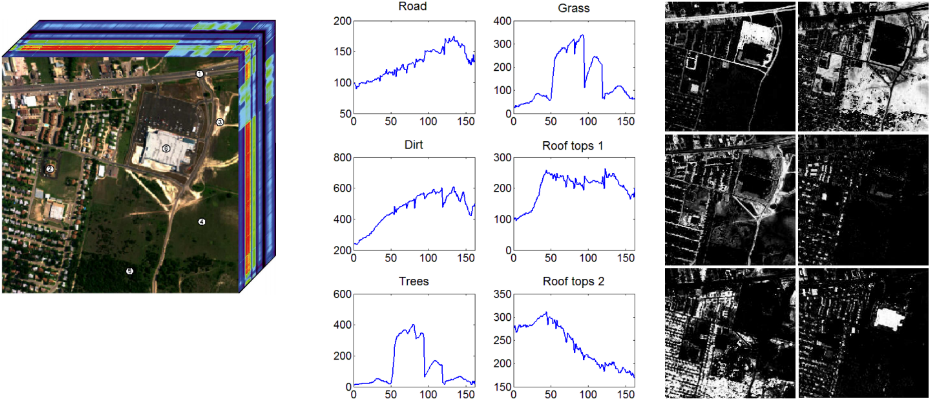


Figure: Hyper-spectral image decomposition. Figure shamelessly copied from (Gillis,2014).

N. Gillis, "The why and how of nonnegative matrix factorization", 2014

Decomposition of Hyperspectral image (2/2)

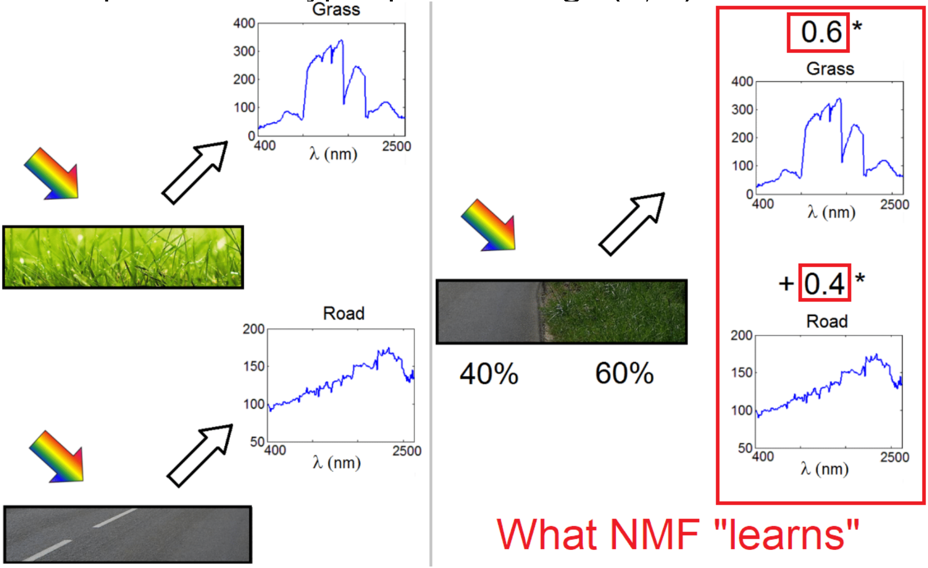


Figure: Hyper-spectral imaging. Figure modified from N. Gillis.

Art work conservation (1/2)

Figure: Amateur restoration ruined a 19th century fresco painting. Image source: internet.

Art work conservation (2/2)

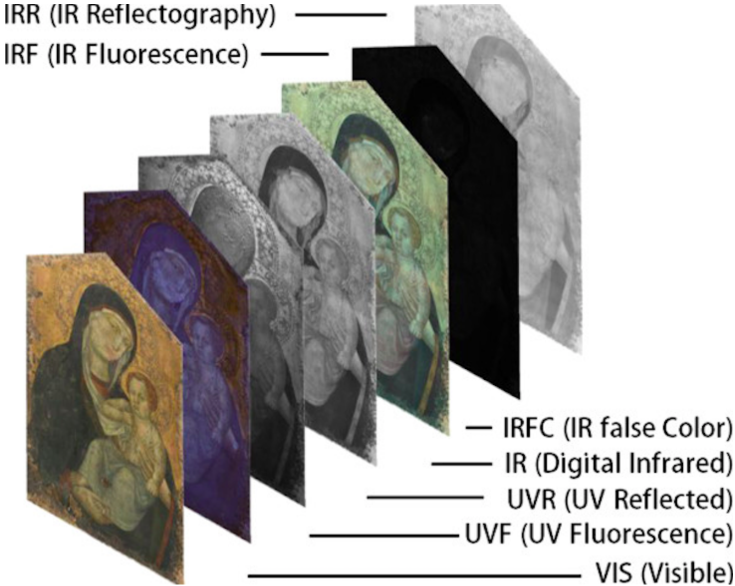


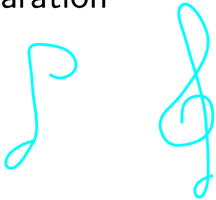
Figure: Image source: internet.

Where we are now

Demo

Where we are now

Audio source separation



https://www.youtube.com/watch?v=1BrpxvpghKQ

Where we are now

Conic combination

- Linear combination of two vectors

$$\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_1, \quad \theta_1, \theta_2 \text{ can be any number}$$

- Conic combination = nonnegative linear combination

$$\theta_1\mathbf{x} + \theta_1\mathbf{y}, \quad \theta_1, \theta_2 \geq 0$$

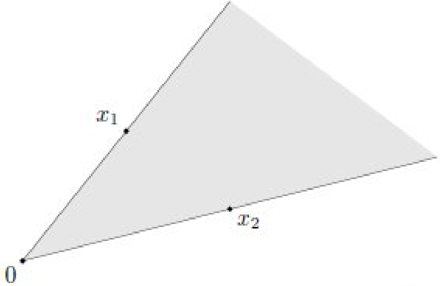
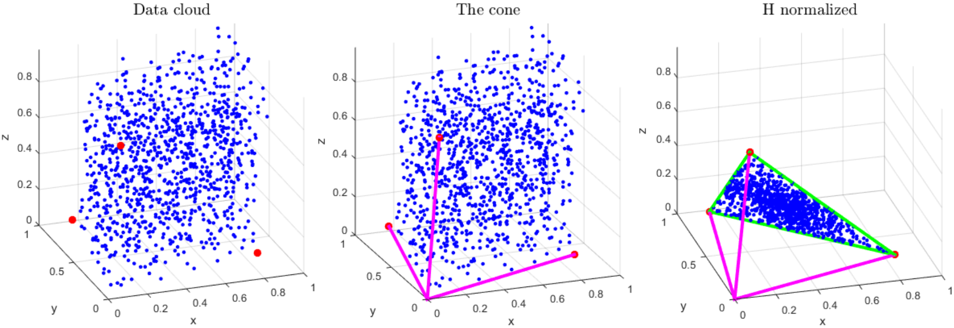


Figure 2.4 The pie slice shows all points of the form $\theta_1 x_1 + \theta_2 x_2$, where $\theta_1,\ \theta_2 \geq 0$. The apex of the slice (which corresponds to $\theta_1 = \theta_2 = 0$) is at 0; its edges (which correspond to $\theta_1 = 0$ or $\theta_2 = 0$) pass through the points x_1 and x_2.

Figure: Image source: internet.

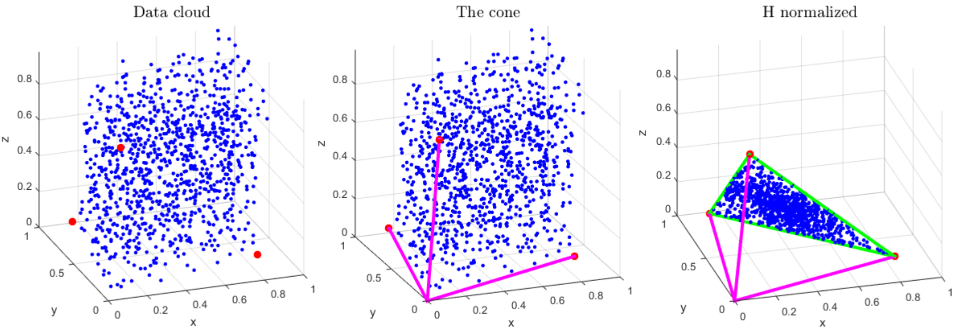
NMF tells a picture of a cone

- Given $\mathbf{M}$, the NMF $\mathbf{M} = \mathbf{W}\mathbf{H}$ tells a picture of a (nonnegative polyhedral simplicial† convex) cone.



NMF tells a picture of a cone

- Given $\mathbf{M}$, the NMF $\mathbf{M} = \mathbf{W}\mathbf{H}$ tells a picture of a (nonnegative polyhedral simplicial† convex) cone.



- If the columns of $\mathbf{H}$ are normalized (sum-to-1), the cone becomes (compressed into) a convex hull.

†Assumes $\mathbf{W}$ is full rank.

How NMF is telling a picture of a cone

$$\underbrace{[\mathbf{m}_1, \mathbf{m}_2, \dots]}_{\mathbf{M}} = \underbrace{\begin{bmatrix} | & | & \\ \mathbf{w}_1 & \mathbf{w}_2 & \dots \\ | & | & \end{bmatrix}}_{\mathbf{W}} \underbrace{\begin{bmatrix} \underbrace{\begin{bmatrix} a \\ b \\ \vdots \end{bmatrix}}_{\mathbf{h}_1}, \underbrace{\begin{bmatrix} * \\ * \\ \vdots \end{bmatrix}}_{\mathbf{h}_2}, \dots \end{bmatrix}}_{\mathbf{H}}$$

$$\mathbf{m}_1 = a\mathbf{w}_1 + b\mathbf{w}_2 + \cdots, \quad a, b, \cdots \geq 0$$

- $\mathbf{m}_i$: a point in the data cloud
- $\mathbf{w}_i$: a vertex / extreme point of the cone
- $\mathbf{h}_i$: how much each columns of $\mathbf{W}$ contribute (conically) to $\mathbf{m}_i$
  - If $\mathbf{h}_i \geq \mathbf{0} \implies$ conical combination
  - If $\mathbf{h}_i \geq \mathbf{0}$ AND $\|\mathbf{h}_i\|_1 = 1 \implies$ convex combination
- If $\|\mathbf{m}_i\|_1 = \|\mathbf{w}_i\|_1 = 1 \implies$ points on unit simplex

NMF tells a picture of a conical hull and a convex hull

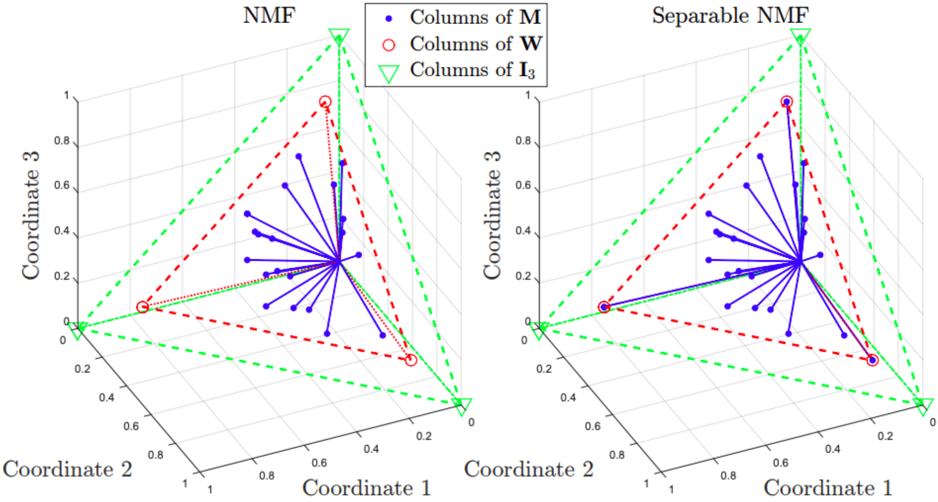


Fig. 1.3. Example of NMF and SNMF. Here $r = m = 3$, $n = 20$. The blue rays are data points $\mathbf{M}$, the red rays are the columns of $\mathbf{W}$ and the green rays are the standard basis vectors in $\mathbb{R}^3$. In both cases, blue cone $\subseteq$ red cone $\subseteq$ green cone.

Problem 2.3 (Nested Polytope Problem - NPP). *Let $\mathcal{A} \subseteq \mathcal{B} \subset \mathbb{R}^d$ be two full-dimensional nested polytopes, that is, the dimension of $\mathcal{A}$ and $\mathcal{B}$ is equal to d. The polytope $\mathcal{A}$, referred to as the inner polytope, is given via the convex hull of n points*

$$\mathcal{A} = \operatorname{conv}\left(\{v_1, v_2, \ldots, v_n\}\right), \; v_j \in \mathbb{R}^d \text{ for } j = 1, 2, \ldots, n,$$

and the polytope $\mathcal{B}$, referred to as the outer polytope, via m inequalities

$$\mathcal{B} = \{x \in \mathbb{R}^d \mid Fx + g \geq 0\},$$

where $F \in \mathbb{R}^{m \times d}$ and $g \in \mathbb{R}^m$. Given k, find, if possible, a polytope $\mathcal{E}$ with k vertices, referred to as the nested polytope, such that

$$\mathcal{A} \subseteq \mathcal{E} \subseteq \mathcal{B}.$$

Figure: Image source: (Nicolas Gillis 2020)

Theorem (Restricted) NMF = NPP

Proof: 5 pages.

If points have unit ℓ_1-norm

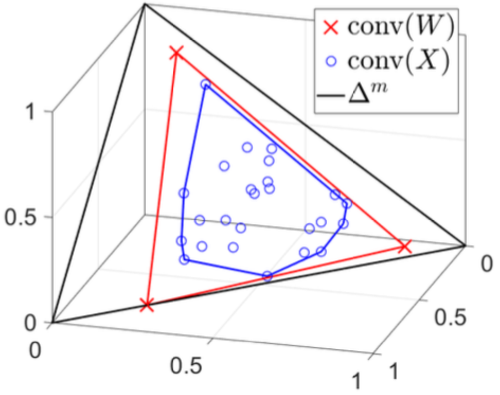


Figure: Image source: (Nicolas Gillis 2020)

blue convex hull $\subseteq$ red convex hull $\subseteq$ unit simplex

Solving the Minimum-volume NMF

https://angms.science/eg_SNPA_ini.gif

Where we are now

Structural Nonconvex Nonsmooth Optimization

- Given $(\mathbf{M}, r)$, find $(\mathbf{W}, \mathbf{H})$ by solving

$$\operatorname*{argmin}_{\mathbf{W},\mathbf{H}} \frac{1}{2}\|\mathbf{M} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda g(\mathbf{W}, \mathbf{H})$$

  - $\frac{1}{2}\|\mathbf{M} - \mathbf{W}\mathbf{H}\|_F^2$: data fitting term in Frobenius norm
  - g: constraint / regularizer

Structural Nonconvex Nonsmooth Optimization

- Given $(\mathbf{M}, r)$, find $(\mathbf{W}, \mathbf{H})$ by solving

$$\underset{\mathbf{W},\mathbf{H}}{\operatorname{argmin}} \ \frac{1}{2}\|\mathbf{M} - \mathbf{WH}\|_F^2 + \lambda g(\mathbf{W}, \mathbf{H})$$

  - $\frac{1}{2}\|\mathbf{M} - \mathbf{WH}\|_F^2$: data fitting term in Frobenius norm
  - g: constraint / regularizer

- Example: Minimum-volume NMF

$$\underset{\mathbf{W},\mathbf{H}}{\operatorname{argmin}} \ \frac{1}{2}\|\mathbf{M} - \mathbf{WH}\|_F^2 + i_+(\mathbf{W}) + i_+(\mathbf{H}) + \lambda \log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I})$$

- i_+: indicator function

$$i_+(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ +\infty & \text{else} \end{cases}$$

i_+ is not differentiable!

Where we are now

Algorithms

- How to solve the problem in the following form?

$$\min_{\mathbf{x},\mathbf{y}} \; f(\mathbf{x}) + g(\mathbf{y}) + H(\mathbf{x},\mathbf{y})$$

  - $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is an extended value function, possibly nonsmooth and nonconvex
  - $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is an extended value function, possibly nonsmooth and nonconvex
  - $H : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is partially smooth, possibly nonconvex

- A bunch of algorithmssssss
  - BCD: **B**lock **C**oordinate **D**escent
  - BSUM: **B**lock **S**uccessive **U**pper-Bound **M**inimization
  - PALM: **P**roximal **A**lternating **L**inearized **M**inimization
  - IBPG: **I**nertial **B**lock **P**roximal **G**radient
  - TITAN: Iner**T**ial block major**I**zation minimization framework for non-smooth non-convex op**T**imiz**A**tio**N**
  - BBPG: **B**lock **B**regman **P**roximal **G**radient
  - HER: **H**eursitic **E**xtrapolation with **R**estart

Generic structure of these algorithm

$$\min_{\mathbf{x},\mathbf{y}} \; f(\mathbf{x}) + g(\mathbf{y}) + H(\mathbf{x},\mathbf{y})$$

Algorithm 1: Generic structure of an algorithm

Result: Sol. to the optimization problem

Initialization of $\mathbf{x}_0, \mathbf{y}_0$;

for $\underline{k = 1, 2, \dots}$ **do**

   $\mathbf{x}_{k+1}$ =Update$(\mathbf{x}_k, \mathbf{y}_k)$;

   $\mathbf{y}_{k+1}$ =Update$(\mathbf{x}_{k+1}, \mathbf{y}_k)$;

end

Update = do something on the variable.

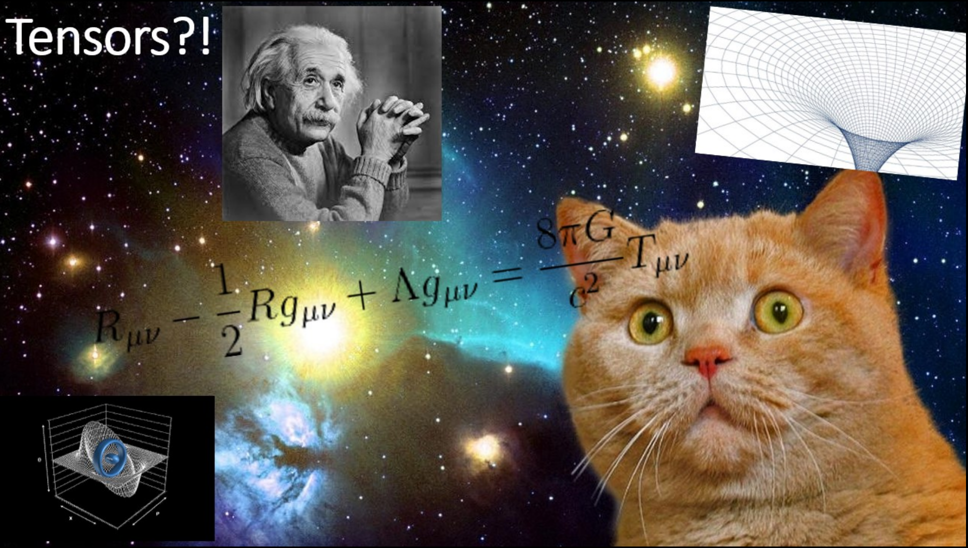
- Projected gradient iteration
- Extrapolated projected gradient iteration (with Nesterov's acceleration)
- Scaled Newton iteration
- Mirror descent or Bregman proximal gradient
- Closed-form sol. (under some specific form)

Another bunch of problemsssss

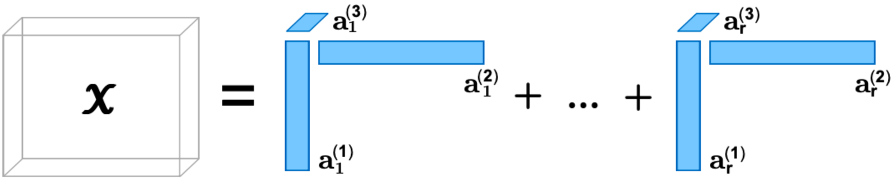
- How to design fast algorithm for solving structural optimization problems?
- Why some methods is slow on certain problem? Why it is slow?
- Why some methods is fast on certain problem? Why it is fast?
- What's the underlying principle behind acceleration phenomenon?
- What's the convergence guarantee for an algorithm? What's the most relaxed condition of convergence guarantee?
- What kind of geometry will affect the convergence of an algorithm? What kind of problems satisfy such geometry?
- What's a good initialization (warm-start) strategy?

Where we are now

Where we are now


Tensors?!
$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^2}T_{\mu\nu}$

Tensors



- Tensor of CPD rank-r mode-N:

$$\mathcal{X} = \sum_{p=1}^{r} \bigotimes_{i=1}^{N} \mathbf{a}_p^{(i)}$$

- Nonnegative Tensor Factorization

$$\min_{\mathbf{a}_p^{(i)}} \frac{1}{2} \left\| \mathcal{X} - \sum_{p=1}^{r} \bigotimes_{i=1}^{N} \mathbf{a}_p^{(i)} \right\|_F^2 + i_+(\mathbf{a}_1^{(1)}) + i_+(\mathbf{a}_1^{(2)}) + \dots$$

Where we are now

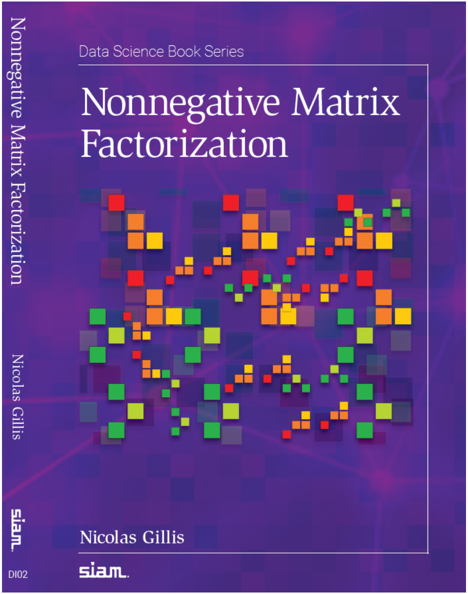
Other NMFs

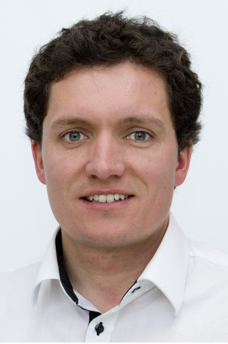
- NMF in other domain
  - Columns of $\mathbf{W}$ are real polynomial / trigonometric polynomial
  - Columns of $\mathbf{W}$ are complex numbers
  - Columns of $\mathbf{W}$ are unimodal / log-cocave vectors
  - $\mathbf{W}, \mathbf{H}$ are 0-1 integers
  - On doubly stochastic matrix
- NMF in other metric
  - Divergence
  - Optimal Transport / Wasserstein distance
  - Hilbert's projective metric
- NMF in other algebra
  - Tropical max-plus semi-ring
  - Clifford algebra
- NMF in other form
  - Deep NMF

Where we are now

Where to learn more?

c'est bon bon !





Nicolas Gillis

https://sites.google.com/site/nicolasgillis/book

Last page - summary

NMF

The end, slide avaliable at angms.science