Fully Efficient Nonnegative Matrix Factorization

Mohammad Hossein Rohban

ISS Group, ECE Dept. Boston University

14 Dec. 2012

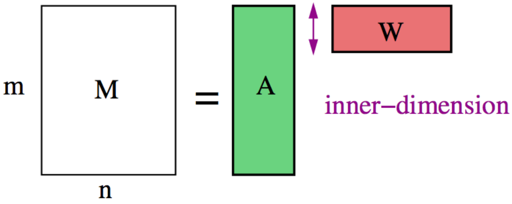
This presentation is some rephrasing of the 5^{th} section of the following paper :

- Arora, S., Ge, R., Kannan, R., and Moitra, A., "*Computing a nonnegative matrix factorization, provably*," In ACM Symposium on Theory of Computing (STOC), 2012.

Table of Contents

Problem Definition

- Noise Free Case : Given a non-negative $m \times n$ matrix $\mathbf{M}$, find $m \times r$ matrix $\mathbf{A}$ and $r \times n$ matrix $\mathbf{W}$ such that $\mathbf{M} = \mathbf{AW}$.
- Small Perturbation Case : Suppose $\mathbf{M} = \mathbf{AW}$, but instead of $\mathbf{M}$, only a noisy version $\mathbf{M}'$ is given.
  - Assumption : $\|\mathbf{M}_i - \mathbf{M}'_i\|_1 \le \epsilon$.
  - Goal : Find $\mathbf{A}'$ and $\mathbf{W}'$ such that $\|\mathbf{M}'_j - (\mathbf{A}'\mathbf{W}')_j\|_1 \le \mathcal{O}(\epsilon)$
- WLoG : Assume that $\mathbf{M}$ and $\mathbf{W}$ are row normalized.



Picture Courtesy : A. Moitra, "Computing a Nonnegative Matrix Factorization – Provably," Dec. 2011.

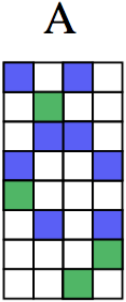
Challenges

- The problem can be solved efficiently using SVD, if **A** and **W** are not required to be nonnegative.
- It has been recently shown by Vavasis that the problem is $\mathcal{NP}$-hard.
- Under the mild separability assumption of **A**, Arora has shown that a very efficient algorithm exists to solve the problem.

Separability Assumption

Definition

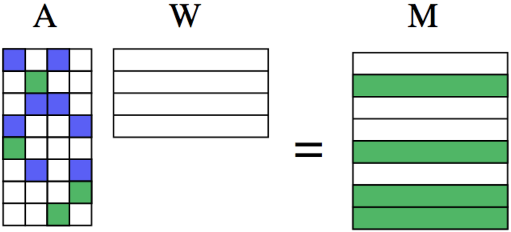
Matrix **A** is a separable matrix, if for each column i, there exists some row $f(i)$ for which the i^{th} element is the only non-zero element in that row. These row indices are called *anchor rows*.



Picture Courtesy : A. Moitra, "Computing a Nonnegative Matrix Factorization – Provably," Dec. 2011.

Why Separability is important?

- After a proper permutation of its rows, **A** contains an identity matrix in itself.
- For each i, $f(i)^{\text{th}}$ row of **M** is just a copy of $f(i)^{\text{th}}$ row of **W**.
- If anchor rows are identified, **W** can be found exactly.
  - Hence we focus of anchor row detection in this presentation.
- Other rows of **A** can be identified easily using a linear program (under some assumption).



Picture Courtesy : A. Moitra, "Computing a Nonnegative Matrix Factorization – Provably," Dec. 2011.

Geometric Picture of the Problem

Definition

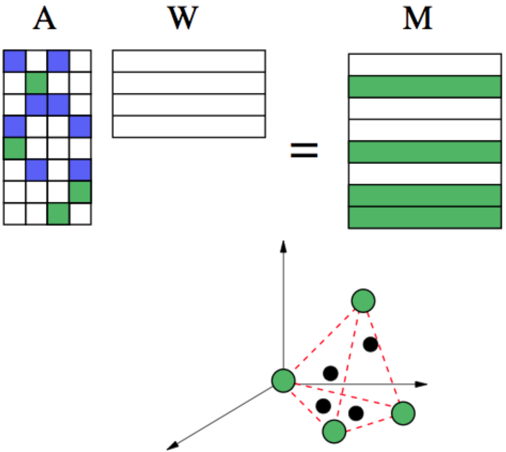
A nonnegative matrix $\mathbf{W}$ is simplicial if no row in $\mathbf{W}$ can be expressed as a linear convex combination of the remaining rows.

Lemma

If a nonnegative matrix $\mathbf{M}$ *has a separable factorization* $\mathbf{AW}$*, then there is one in which* $\mathbf{W}$ *is simplicial.*

- This implies that anchor rows correspond to the extreme points of the convex hull containing other rows.

Geometric Picture of the Problem (cont.)



Picture Courtesy : A. Moitra, "Computing a Nonnegative Matrix Factorization – Provably," Dec. 2011.

Anchor Detection Algorithm (Noise Free Case)

Definition

The row $\mathbf{M}_j$ is a *loner* if (ignoring other rows that are its copies), it is not in the convex hull of the remaining rows.

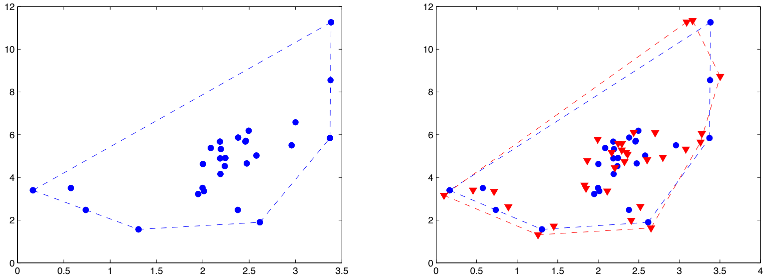
- Using Linear Programming, it is easy check whether a point is a loner.

Lemma

A row is a loner iff it is an anchor row (equal to some row $\mathbf{W}_j$).

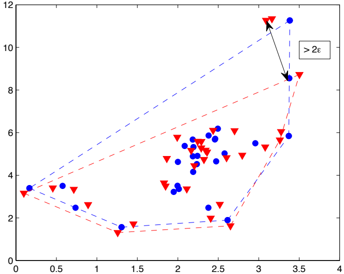
Anchor Detection Algorithm (Small Perturbation Case)

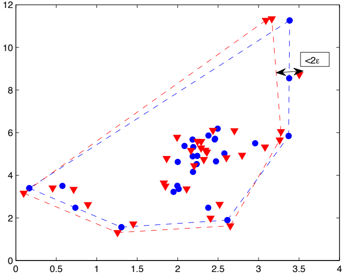
- Why can't we just apply the previous algorithm?
  - There may be multiple copies of an extreme point.
  - Their distance to the convex hull the remaining rows may be indistinguishable with these distances for the ones on/near some face of the simplex.



Small Perturbation Case (cont.)

- Main idea : remove all points in an small ℓ_1 neighborhood of the point which is going to be tested, then check whether it is loner.
  - What should be the size of that neighborhood in order to guarantee a robust anchor row detection?
  - What is the right threshold for the distance of a loner to the convex hull of remaining rows?
  - How close would be the robust loners to the rows of $\mathbf{W}$?





Small Perturbation Case (cont.)

Definition

W is called *α-robust simplicial*, if ℓ_1 distance of any row in **W** to the convex hull of the remaining rows is larger than α.

- This condition holds (with high probability in terms of n) for many reasonable prior distributions on **W**.

Definition

A point $\mathbf{M}'_j$ is called a *robust loner*, if ignoring rows that are within ℓ_1 distances of $d = 10\epsilon/\alpha + 2\epsilon$, its ℓ_1 distance to convex hull of the remaining rows be more than 2ϵ.

- Robust loners can be found using Linear Programming.

Main Facts about Robust Loners

Lemma

For all i, $\mathbf{M}'_{f(i)}$ will be a robust loner.

- Remind that $\|\mathbf{W}_i - \mathbf{M}'_{f(i)}\|_1 \leq \epsilon$.
- Hence for this set of robust loners, the retrieved row is within ϵ ℓ_1 distance of the ground truth.

Lemma

If $\mathbf{M}'_j$ has ℓ_1 distance more than $d + \epsilon = 10\epsilon/\alpha + 3\epsilon$ to all $\mathbf{W}_i$s, then it can't be a robust loner.

- These two lemmas imply that
  - For all $\mathbf{W}_i$s, we retrieve at least one row of $\mathbf{M}'$ which is ϵ close to it.
  - The other retrieved rows have distance less than $\mathcal{O}(\epsilon/\alpha)$ to some $\mathbf{W}_j$.
  - ϵ should be small with respect to α to have reasonable estimation of $\mathbf{W}$.

Concluding Remarks

- Using row normalization + assuming simplicial second factor + Separability assumption implies the possibility of totally efficient NMF in the noise free case.
- Using same components as the previous one and assuming robust simplicial factor implies possibility of totally efficient approximate NMF in the small perturbation case.