Global Identifiability of Overcomplete Dictionary Learning via L1 and Volume Minimization

Yuchen Sun
yuchen.sun@ufl.edu
University of Florida

Kejun Huang
kejun.huang@ufl.edu
University of Florida

The Thirteenth International Conference
on Learning Representations (ICLR 2025)

Dictionary Learning

- Given data matrix $\boldsymbol{X}$, find dictionary matrix $\boldsymbol{A}_{m\times k}$ and sparse coefficients $\boldsymbol{S}_{k\times n}$

$$\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{s}_i, \qquad i = 1, \ldots, n \qquad \Rightarrow \qquad \boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$$

- Completeness of the dictionary $\boldsymbol{A}_{m\times k}$: usually assume $m \leq k$:
  - complete DL: $m = k$
  - overcomplete DL: $m < k$ (in this paper)
- Assume there exists a groundtruth generative model $\boldsymbol{X} = \boldsymbol{A}_\natural \boldsymbol{S}_\natural$,

  (unconstrained) matrix factorization not identifiable without additional assumption

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} = (\boldsymbol{A}\boldsymbol{Q})(\boldsymbol{Q}^{-1}\boldsymbol{S}) = \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{S}}$$

Previous Work

★ Model is **identifiable** if for any $(\boldsymbol{A}_\star, \boldsymbol{S}_\star)$ there exist permutation $\boldsymbol{\Pi}$ and diagonal $\boldsymbol{D}$ that

$$\boldsymbol{A}_\natural = \boldsymbol{A}_\star \boldsymbol{\Pi} \boldsymbol{D}, \quad \boldsymbol{S}_\natural = \boldsymbol{D}^{-1} \boldsymbol{\Pi}^\top \boldsymbol{S}_\star$$

- Identifiability of hard sparsity-constrained DL: $\|\boldsymbol{s}_i\|_0 \leq s$ for all i
  - sample size n is $O((k+1)\binom{k}{s})$ [Aharon et al., 2006],[Hillar and Sommer, 2015]
  - sample size $O(k^3/(k-s)^2)$ [Cohen and Gillis, 2019]

Previous Work

- **Local** identifiability with ℓ_1 regularization:

$$\underset{\boldsymbol{A},\boldsymbol{S}}{\text{minimize}} \quad \|\boldsymbol{S}\|_1 \qquad \text{subject to} \quad \boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}, \|\boldsymbol{A}_{:,c}\| \leq 1, c = 1, \ldots, k$$

[Gribonval and Schnass, 2010],[Wu and Yu, 2017],[Wang et al., 2020]

  - Sample size requirement has been down to $n = O(k \log(k))$.
  - But the results are dominantly local.

Previous Work

- **Global** identifiability is achieved by Hu and Huang [2023], Sun and Huang [2024]
  - using a matrix volume criterion $|\det \boldsymbol{A}|$ while constraining the ℓ_1 norms of the rows of $\boldsymbol{S}$ with same sample complexity $n = O(k \log(k))$
  - although as the criterion suggests it **only** applies to **complete dictionaries**

Identifiability

Definition: identifiability

Consider the generative model $\boldsymbol{X} = \boldsymbol{A}_\natural \boldsymbol{S}_\natural$, where $\boldsymbol{A}_\natural$ and $\boldsymbol{S}_\natural$ are the groundtruth latent factors. Let $(\boldsymbol{A}_\star, \boldsymbol{S}_\star)$ be optimal for an identification criterion q

$$(\boldsymbol{A}_\star, \boldsymbol{S}_\star) = \arg\min_{\boldsymbol{X}=\boldsymbol{A}\boldsymbol{S}} q(\boldsymbol{A}, \boldsymbol{S}).$$

If $\boldsymbol{A}_\natural$ and/or $\boldsymbol{S}_\natural$ satisfy some condition such that for any $(\boldsymbol{A}_\star, \boldsymbol{S}_\star)$, there exist a permutation matrix $\boldsymbol{\Pi}$ and a diagonal matrix $\boldsymbol{D}$ such that $\boldsymbol{A}_\natural = \boldsymbol{A}_\star \boldsymbol{D} \boldsymbol{\Pi}$, then we say $\boldsymbol{A}_\natural$ is essentially identifiable, up to permutation and scaling, under that condition; if we further have that $\boldsymbol{S}_\natural = \boldsymbol{\Pi}^\top \boldsymbol{D}^{-1} \boldsymbol{S}_\star$, then we say that the matrix factorization model is essentially identifiable, up to permutation and scaling, under that condition.

Our Work

★ Novel formulation for overcomplete dictionary

$$\underset{\boldsymbol{A},\boldsymbol{S}}{\text{minimize}} \quad \frac{1}{2}\log\det \boldsymbol{A}\boldsymbol{A}^\top + \max_{\|\boldsymbol{d}\|_2^2=m} \sum_{c=1}^{k} d_c \|\boldsymbol{e}_c^\top \boldsymbol{S}\|_1 \qquad \text{subject to} \quad \boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} \tag{1}$$

- (1) is our identification criterion
- What are the conditions for the identifiability of $\boldsymbol{A}_\natural$ and $\boldsymbol{S}_\natural$?

Identifiability of $A_\natural$

★ Novel formulation for overcomplete dictionary

$$\underset{A,S}{\text{minimize}} \quad \frac{1}{2}\log\det AA^\top + \max_{\|d\|_2^2=m} \sum_{c=1}^{k} d_c \|e_c^\top S\|_1 \qquad \text{subject to} \quad X = AS \tag{1}$$

► Optimal scaling: $d_{\star c}$ are the optimal weights that reach the maximum of $\sum_c d_c \|e_c^\top S_\star\|_1$.

$$\|e_c^\top S_\star\|_1 = \sqrt{\left[A_\star^\top \left(A_\star A_\star^\top\right)^{-1} A_\star\right]_{cc}} = d_{\star c}, \qquad c = 1, \ldots, k. \tag{2}$$

Identifiability of $A_\natural$

Assumption 1:

The columns of $\boldsymbol{A}_\natural$ and rows of $\boldsymbol{S}_\natural$ are scaled to satisfy (2).

Assumption 2:

Rows of $\boldsymbol{A}_\natural$ and $\boldsymbol{S}_\natural$ are both linearly independent. Matrix $\boldsymbol{A}_\natural$ does not contain zero columns.

Assumption 3: m-strongly scattered in the k-hypercube

Identifiability of $A_\natural$

Assumption 3: m-strongly scattered in the k-hypercube

Let $\mathcal{C}_k$ denote the k-hypercube $\mathcal{C}_k = \{\boldsymbol{x} \in \mathbb{R}^k \mid \|\boldsymbol{x}\|_\infty \leq 1\}$. Define $\mathcal{B}_m$ as the following set

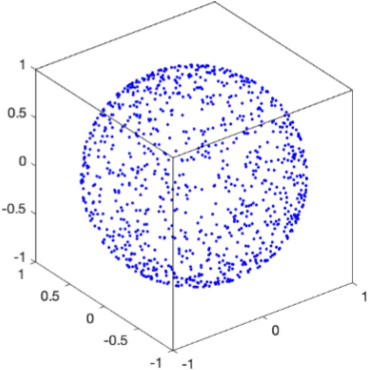
$$\mathcal{B}_m = \{ \operatorname{Diag}(\|\boldsymbol{q}_1\|_2, \ldots, \|\boldsymbol{q}_k\|_2)^\dagger \boldsymbol{Q}\boldsymbol{p} \mid \forall\, \boldsymbol{Q} \in \mathbb{R}^{k\times m} : \boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I},\ \boldsymbol{p} \in \mathbb{R}^m : \|\boldsymbol{p}\|_2 = 1\},$$

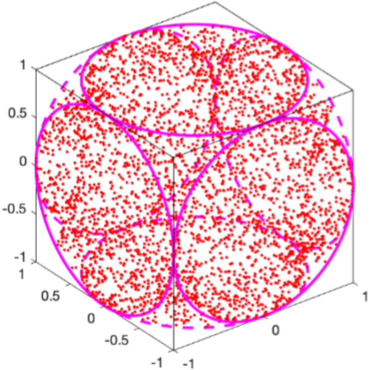
where $\boldsymbol{q}_c$ denotes the cth row of $\boldsymbol{Q}$. A set $\mathcal{S}$ is m-strongly scattered in the k-hypercube if:

1. $\mathcal{B}_m \subseteq \mathcal{S} \subseteq \mathcal{C}_k$;
2. $\partial\mathcal{B}_m \cap \partial\mathcal{S} = \{\operatorname{Diag}(\|\boldsymbol{q}_1\|_2, \ldots, \|\boldsymbol{q}_k\|_2)^\dagger \boldsymbol{Q}\boldsymbol{q}/\|\boldsymbol{q}\|_2\}$, where $\boldsymbol{q}$ are rows of $\boldsymbol{Q}$ with $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}$ and ∂ denotes the boundary of the set.

Assumption 3: m-strongly scattered in the k-hypercube

- An illustration of sufficiently scattered and 2-strongly scattered in 3-hypercube:





Assumption 3: m-strongly scattered in the k-hypercube

★ Strong generalization of Sufficiently Scattered condition in Hu and Huang [2023].

★ A **significant breakthrough** not only in Dictionary Learning but also in general matrix factorization models when the ambient dimension is smaller than the latent dimension.

Theorom 1: Identifiability of $\boldsymbol{A}_\natural$

An overcomplete dictionary $\boldsymbol{A}_\natural$ is identifiable if the groundtruth $\boldsymbol{A}_\natural$ and $\boldsymbol{S}_\natural$ satisfies Assumption 1, 2 and $\mathrm{cell}(\widetilde{\boldsymbol{S}}_\natural)$ satisfies Assumption 3.

- It will be imposed on the cellular hull of $\boldsymbol{S}_\natural$, which is defined as follows:

Definition (Cellular hull)

The cellular hull of a finite set of vectors $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$, stacked as the columns of the matrix $\boldsymbol{S}$, is

$$\mathrm{cell}(\boldsymbol{S}) = \left\{ \boldsymbol{S}\boldsymbol{\theta} \;\middle|\; \|\boldsymbol{\theta}\|_\infty \leq 1 \right\}.$$

Identifiability of $S_\natural$

- With the knowledge of the dictionary, the identifiability of $\boldsymbol{S}_\natural$ has been studied extensively. A general condition for $\boldsymbol{S}_\natural$ is identifiable:

Assumption 4:

Every column of $\boldsymbol{S}_\natural$ contains at most s nonzeros. In addition, $\boldsymbol{A}_\natural$ is a dictionary such that for every $\boldsymbol{s}_0$ with no more than s nonzeros, $\boldsymbol{s}_0$ is the unique solution to the following optimization problem

$$\underset{\boldsymbol{s}}{\text{minimize}} \quad \|\boldsymbol{s}\|_1 \qquad \text{subject to} \quad \boldsymbol{A}_\natural \boldsymbol{D}_\natural^{-1} \boldsymbol{s} = \boldsymbol{A}_\natural \boldsymbol{D}_\natural^{-1} \boldsymbol{s}_0,$$

where $\boldsymbol{D}_\natural$ is a diagonal matrix with

$$[\boldsymbol{D}_\natural]_{cc} = \sqrt{\left[\boldsymbol{A}_\natural^\top \left(\boldsymbol{A}_\natural \boldsymbol{A}_\natural^\top\right)^{-1} \boldsymbol{A}_\natural\right]_{cc}}, \qquad c = 1, \ldots, k.$$

Identifiability Analysis

Theorom 1: Identifiability of $A_\natural$

An overcomplete dictionary $\boldsymbol{A}_\natural$ is identifiable if the groundtruth $\boldsymbol{A}_\natural$ and $\boldsymbol{S}_\natural$ satisfies Assumption 1, 2 and $\mathrm{cell}(\widetilde{\boldsymbol{S}}_\natural)$ satisfies Assumption 3.

Corollary:

Consider the overcomplete DL model $\boldsymbol{X} = \boldsymbol{A}_\natural \boldsymbol{S}_\natural$, where $\boldsymbol{A}_\natural \in \mathbb{R}^{m \times k}$ is the groundtruth mixing matrix and $\boldsymbol{S}_\natural \in \mathbb{R}^{k \times n}$ is the groundtruth sparse coefficient matrix. Suppose $\boldsymbol{A}_\natural$ and $\boldsymbol{S}_\natural$ satisfies Assumptions 1–4. Then for any solution of (1), denoted as $(\boldsymbol{A}_\star, \boldsymbol{S}_\star)$, there exist a permutation matrix $\boldsymbol{\Pi}$ and a diagonal matrix $\boldsymbol{D}$ such that $\boldsymbol{A}_\natural = \boldsymbol{A}_\star \boldsymbol{D} \boldsymbol{\Pi}$ and $\boldsymbol{S}_\natural = \boldsymbol{\Pi}^\top \boldsymbol{D}^{-1} \boldsymbol{S}_\star$.

Sample Complexity Analysis

- Sparse-Gaussian model: $\boldsymbol{S} \sim \mathcal{SG}(s)$ with parameter $s < k$, if every column of $\boldsymbol{S}$ is independently and identically distributed from the following process
  - a subset $\mathcal{I}$ of size s is uniformly drawn from all size-s subsets of $\{1, \ldots, k\}$
  - let $\boldsymbol{s} \in \mathbb{R}^k$ be such that $s_i = 0$ if $i \in \mathcal{I}$ and $s_i \sim \mathcal{N}(0, 1)$ if $i \notin \mathcal{I}$, where $\mathcal{N}(0, 1)$ stands for a standard normal distribution

Sample Complexity Analysis

- Sparse-Gaussian model: $\boldsymbol{S} \sim \mathcal{SG}(s)$ with parameter $s < k$.

Theorem 2:

Suppose $\boldsymbol{S} \in \mathbb{R}^{k \times n}$ is generated from the sparse-Gaussian model $\mathcal{SG}(s)$, where $s < m$, and $\widetilde{\boldsymbol{S}}$ is obtained by scaling its rows to have unit ℓ_1 norm. Then

$$\Pr\left[\sup_{\substack{\|\boldsymbol{w}^\top \tilde{\boldsymbol{S}}\|_1 \leq 1 \\ \boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}}} \|\boldsymbol{Q}^\top \boldsymbol{D}^\dagger \boldsymbol{w}\| > 1\right] \leq 4 \exp\left(\frac{k}{2} \log \frac{k^2}{m} - n \frac{s^2 m}{k^3}\right). \quad (3)$$

The probability goes to zero exponentially fast as

$$n = O\left(\frac{k^2}{m} \log \frac{k^2}{m}\right).$$

References I

Michal Aharon, Michael Elad, and Alfred M Bruckstein. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. **Linear algebra and its applications**, 416(1):48–67, 2006.

Jeremy E Cohen and Nicolas Gillis. Identifiability of complete dictionary learning. **SIAM Journal on Mathematics of Data Science**, 1(3):518–536, 2019.

Rémi Gribonval and Karin Schnass. Dictionary identification——sparse matrix-factorization via ℓ_1-minimization. **IEEE Transactions on Information Theory**, 56(7):3523–3539, 2010.

Christopher J Hillar and Friedrich T Sommer. When can dictionary learning uniquely recover sparse data from subsamples? **IEEE Transactions on Information Theory**, 61(11):6290–6297, 2015.

Jingzhou Hu and Kejun Huang. Global identifiability of ℓ_1-based dictionary learning via matrix volume optimization. In **Advances in Neural Information Processing Systems (NeurIPS)**, volume 36, 2023.

Yuchen Sun and Kejun Huang. Improved identifiability and sample complexity analysis of complete dictionary learning. In **2024 IEEE 13rd Sensor Array and Multichannel Signal Processing Workshop (SAM)**, pages 1–5. IEEE, 2024.

Yu Wang, Siqi Wu, and Bin Yu. Unique sharp local minimum in l1-minimization complete dictionary learning. **Journal of Machine Learning Research**, 21(63):1–52, 2020.

Siqi Wu and Bin Yu. Local identifiability of l1-minimization dictionary learning: a sufficient and almost necessary condition. **The Journal of Machine Learning Research**, 18 (1):6121–6176, 2017.