Efficient Moment Methods and Mixture Models

Joe Kileel

University of Texas at Austin, Department of Mathematics

Oden Institute for Computational Engineering and Sciences

January 11, 2024

Symmetric Moment Tensor

Given data $\mathbf{x}_1, \ldots, \mathbf{x}_p \in \mathbb{R}^n$. It is often useful to form the moment

$$\mathbf{M}_d = \frac{1}{p}\sum_{i=1}^{p} \mathbf{x}_i^{\otimes d} \in S^d(\mathbb{R}^n)$$

where $(\mathbf{x}^{\otimes d})_{i_1,\ldots,i_d} = \mathbf{x}_{i_1} \ldots \mathbf{x}_{i_d}$ for each $(i_1, \ldots, i_d) \in [n]^d$.

Symmetric Moment Tensor

Given data $\mathbf{x}_1, \ldots, \mathbf{x}_p \in \mathbb{R}^n$. It is often useful to form the moment

$$\mathbf{M}_d = \frac{1}{p} \sum_{i=1}^{p} \mathbf{x}_i^{\otimes d} \in S^d(\mathbb{R}^n)$$

where $(\mathbf{x}^{\otimes d})_{i_1,\ldots,i_d} = \mathbf{x}_{i_1} \ldots \mathbf{x}_{i_d}$ for each $(i_1, \ldots, i_d) \in [n]^d$.

- $d = 1 \rightsquigarrow$ sample average
- $d = 2 \rightsquigarrow$ sample covariance matrix (uncentered)
- $d = 3 \rightsquigarrow n \times n \times n$ real symmetric tensor (sample third moment), etc.

Moment Tensor Decompositions: A Mix of Algebra and Computation

When the data $\mathbf{x}_1, \ldots, \mathbf{x}_p$ follows a nice model, typically the moment tensors $\mathbf{M}_d$ admit a nice algebraic decomposition. We'll see computing it helps to e.g. estimate model parameters.

Moment Tensor Decompositions: A Mix of Algebra and Computation

When the data $\mathbf{x}_1, \ldots, \mathbf{x}_p$ follows a nice model, typically the moment tensors $\mathbf{M}_d$ admit a nice algebraic decomposition. We'll see computing it helps to e.g. estimate model parameters.

I will focus on two examples of (*model*, *decomposition*) pairs:

I. Gaussian Mixture Models & CP Tensor Decompositions

II. Mixture of Products & Coupled Incomplete CP Tensor Decompositions

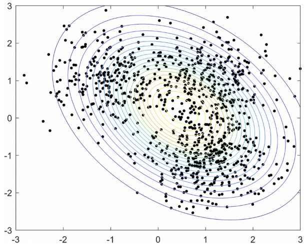
PART I: *Gaussian Mixture Models & CP Tensor Decompositions*

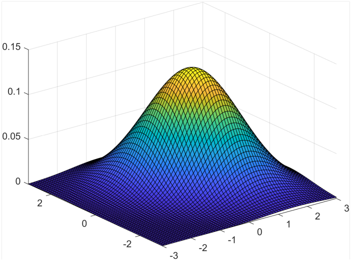
Gaussian Distribution

Gaussian vector: $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$

probability density function: $\frac{\exp(-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu))}{\sqrt{(2\pi)^n \det(\Sigma)}}$

parameters: $\mu = \mathbb{E}[\mathbf{x}] \in \mathbb{R}^n, \; \Sigma = \mathbb{E}[(\mathbf{x}-\mu)^{\otimes 2}] \in S^2(\mathbb{R}^n)$



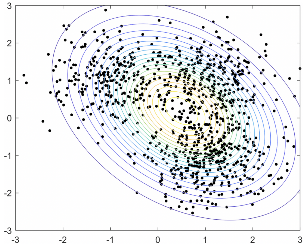


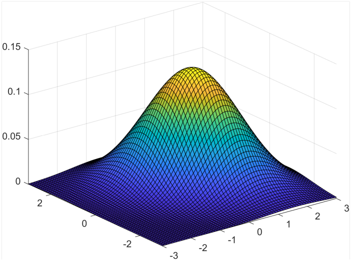
Gaussian Distribution

Gaussian vector: $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$

probability density function: $\frac{\exp(-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu))}{\sqrt{(2\pi)^n \det(\Sigma)}}$

parameters: $\mu = \mathbb{E}[\mathbf{x}] \in \mathbb{R}^n, \; \Sigma = \mathbb{E}[(\mathbf{x}-\mu)^{\otimes 2}] \in S^2(\mathbb{R}^n)$





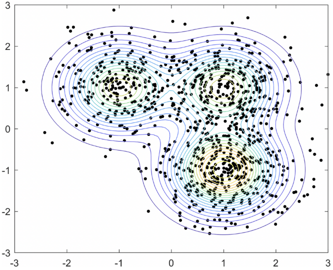
- Limiting average of any (suff. integrable) i.i.d. random vectors
- Marginals are themselves lower-dimensional Gaussians

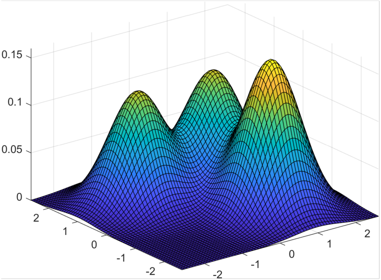
Gaussian Mixture Models

GMM: $\mathbf{x} \sim \sum_{j=1}^{r} \lambda_j \mathcal{N}(\mu_j, \Sigma_j)$

r is the number of components, λ_j are the mixing weights (convex combination)

parameters: $\{(\lambda_j, \mu_j, \Sigma_j) : j = 1, \ldots, r\}$



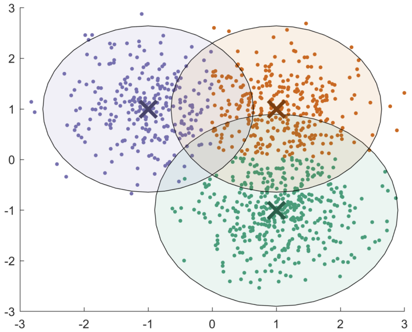


The Many Applications of Gaussian Mixtures

Density Estimation

Clustering

Anomaly Detection

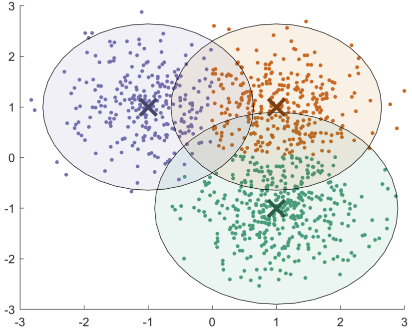


The Many Applications of Gaussian Mixtures

Density Estimation

Clustering

Anomaly Detection



GMMs are one of the most prevalent tools in data analysis!

Neat Formula for Moment Tensors of GMM

Lemma (Wick '50, Pereira-K.-Kolda '22)

Let $\mathbf{x}_1, \ldots, \mathbf{x}_p$ *be i.i.d. realizations of a GMM with parameters* $\{(\lambda_j, \mu_j, \Sigma_j)\}$. *Then*

$$\mathbf{M}_d \longrightarrow \sum_{j=1}^{r} \lambda_j \sum_{k=0}^{\lfloor d/2 \rfloor} \binom{d}{2k} \frac{(2k)!}{k! 2^k} \operatorname{sym}(\mu_j^{\otimes(d-2k)} \otimes \Sigma_j^{\otimes k}) \quad \textit{as } p \to \infty.$$

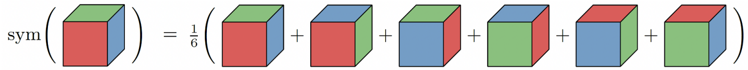
Neat Formula for Moment Tensors of GMM

Lemma (Wick '50, Pereira-K.-Kolda '22)

Let $\mathbf{x}_1, \ldots, \mathbf{x}_p$ *be i.i.d. realizations of a GMM with parameters* $\{(\lambda_j, \mu_j, \Sigma_j)\}$. *Then*

$$\mathbf{M}_d \longrightarrow \sum_{j=1}^{r} \lambda_j \sum_{k=0}^{\lfloor d/2 \rfloor} \binom{d}{2k} \frac{(2k)!}{k! 2^k} \operatorname{sym}(\mu_j^{\otimes(d-2k)} \otimes \Sigma_j^{\otimes k}) \quad \textit{as } p \to \infty.$$

The proof is most easily done using the bijection Φ from symmetric tensors to homogeneous forms, because $\Phi(\operatorname{sym}(S \otimes T)) = \Phi(S)\Phi(T)$.



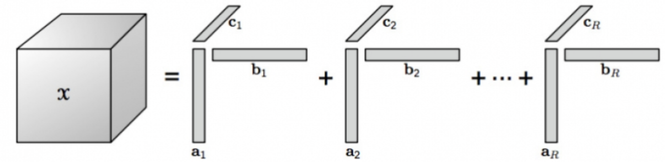
Symmetric CP Tensor Decomposition

Proposition (Pereira-Kileel-Kolda '22)

Let $\mathbf{x}_1, \ldots, \mathbf{x}_p$ *be i.i.d. realizations of a GMM with parameters* $\{(\lambda_j, \mu_j, \Sigma)\}$*, i.e. there is a common covariance. Then as* $p \to \infty$*,*

$$\sum_{k=0}^{\lfloor d/2 \rfloor} (-1)^k \binom{d}{2k} \frac{(2k)!}{k! 2^k} \operatorname{sym}(\mathbf{M}_{d-2k} \otimes \Sigma^{\otimes k}) \longrightarrow \sum_{j=1}^{r} \lambda_j \mu_j^{\otimes d}.$$

The right-hand side is a real symmetric CP tensor decomposition.



Computational Algebraic Geometry: GMM Identifiability

What is the maximal number r of Gaussian components in $\mathbb{R}^n$ that is uniquely determined by the first d moments?

Computational Algebraic Geometry: GMM Identifiability

What is the maximal number r of Gaussian components in $\mathbb{R}^n$ that is uniquely determined by the first d moments?

Symmetric CP tensor decomposition is generically unique when

$$r \dim v_d(\mathbb{P}(\mathbb{R}^n)) + (r-1) < \dim \mathbb{P}(S^d(\mathbb{R}^n)) \Leftrightarrow r < \frac{1}{n}\binom{n+d-1}{d}$$

if $d \geq 3$, with a few classified exceptions for (d, n, r) (see e.g. Chiantini-Ottaviani-Vannieuwenhoven '17 about secants of Veronese varieties). It implies that a GMM with a known common covariance and generic (λ_j, μ_j) is determined by its moments of order $\leq d$ up to such r.

Computational Algebraic Geometry: GMM Identifiability

What is the maximal number r of Gaussian components in $\mathbb{R}^n$ that is uniquely determined by the first d moments?

Symmetric CP tensor decomposition is generically unique when

$$r \dim v_d(\mathbb{P}(\mathbb{R}^n)) + (r-1) < \dim \mathbb{P}(S^d(\mathbb{R}^n)) \Leftrightarrow r < \frac{1}{n}\binom{n+d-1}{d}$$

if $d \geq 3$, with a few classified exceptions for (d, n, r) (see e.g. Chiantini-Ottaviani-Vannieuwenhoven '17 about secants of Veronese varieties). It implies that a GMM with a known common covariance and generic (λ_j, μ_j) is determined by its moments of order $\leq d$ up to such r.

[Others study identifiability of GMMs under different assumptions. Boils down to other secant varieties, e.g. Améndola-Ranestad-Sturmfels '17.]

Computational Algebraic Geometry: GMM Identifiability

What is the maximal number r of Gaussian components in $\mathbb{R}^n$ that is uniquely determined by the first d moments?

Symmetric CP tensor decomposition is generically unique when

$$r \dim v_d(\mathbb{P}(\mathbb{R}^n)) + (r-1) < \dim \mathbb{P}(S^d(\mathbb{R}^n)) \Leftrightarrow r < \frac{1}{n}\binom{n+d-1}{d}$$

if $d \geq 3$, with a few classified exceptions for (d, n, r) (see e.g. Chiantini-Ottaviani-Vannieuwenhoven '17 about secants of Veronese varieties). It implies that a GMM with a known common covariance and generic (λ_j, μ_j) is determined by its moments of order $\leq d$ up to such r.

[Others study identifiability of GMMs under different assumptions. Boils down to other secant varieties, e.g. Améndola-Ranestad-Sturmfels '17.]

To my knowledge, identifiability for the most general case of unknown, different and unconstrained Σ_j is currently unresolved.

Numerical Algorithm Beating the Curse of Dimensionality

To fit a GMM to data, consider minimizing the cost function

$$\text{argmin}_{\lambda_j, \mu_j, \Sigma_j} \|\mathbf{M}_d - \text{(aforementioned formula in parameters)}\|_F^2$$

Numerical Algorithm Beating the Curse of Dimensionality

To fit a GMM to data, consider minimizing the cost function

$$\operatorname{argmin}_{\lambda_j,\mu_j,\Sigma_j} \|\mathbf{M}_d - (\text{aforementioned formula in parameters})\|_F^2$$

Naively forming the terms would take $\mathcal{O}(pn^d)$ flops and $\mathcal{O}(n^d)$ in storage!

Numerical Algorithm Beating the Curse of Dimensionality

To fit a GMM to data, consider minimizing the cost function

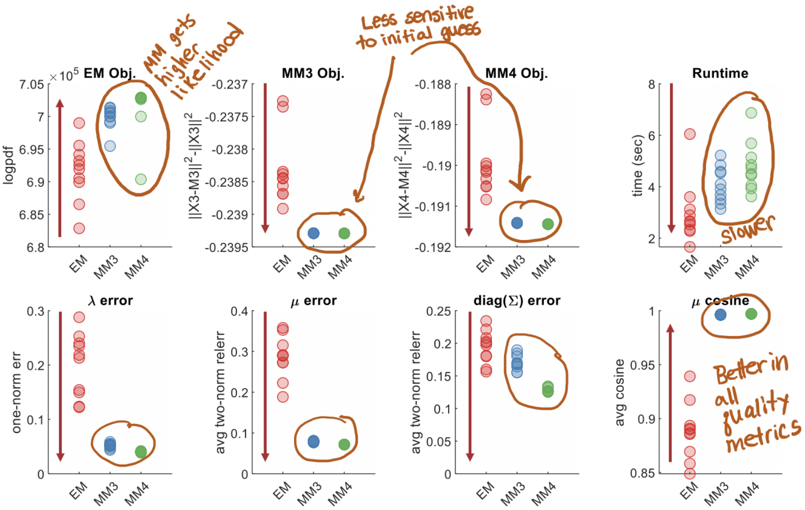
$$\operatorname{argmin}_{\lambda_j, \mu_j, \Sigma_j} \|\mathbf{M}_d - (\text{aforementioned formula in parameters})\|_F^2$$

Naively forming the terms would take $\mathcal{O}(pn^d)$ flops and $\mathcal{O}(n^d)$ in storage!

Theorem (Pereira-Kileel-Kolda '22)

Given the parameters $\lambda_j, \mu_j, \Sigma_j$ and data $\mathbf{x}_i$, there is an algorithm to evaluate the above cost and its gradient in $\mathcal{O}(prn^2 + r^2n^3)$ flops and $\mathcal{O}(rn^2 + pn)$ storage. If Σ_j are diagonal, these drop to $\mathcal{O}(prn + r^2n)$ flops and $\mathcal{O}(rn + pn)$ storage.

In Practice: Method of Moments Can Outperform EM



- Randomly-generated problems with overlapping Gaussians
- $n = 100$, $r = 20$, $p = 8000$, common diagonal Σ
- Compared EM, MM3 (moments $d = 3$), MM4 (moments $d = 4$)

Sketch: Expanding Out The Inner Products

Idea is to operate on moment tensors without forming them!

$$\min_{\theta} f(\theta) \equiv \left\| \frac{1}{p} \sum_{i=1}^{p} \mathbf{x}_i^{\otimes d} - \sum_{j=1}^{m} \lambda_j \mathcal{M}_j^{(d)} \right\|^2$$

$$f(\theta) = \underbrace{\cancel{\left\| \frac{1}{p} \sum_{i=1}^{p} \mathbf{x}_i^{\otimes d} \right\|^2}}_{\text{constant}} + \left\| \sum_{j=1}^{m} \lambda_j \mathcal{M}_j^{(d)} \right\|^2 - 2 \left\langle \frac{1}{p} \sum_{i=1}^{p} \mathbf{x}_i^{\otimes d}, \sum_{j=1}^{m} \lambda_j \mathcal{M}_j^{(d)} \right\rangle$$

$$f(\theta) = C + \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j \underbrace{\left\langle \mathcal{M}_i^{(d)}, \mathcal{M}_j^{(d)} \right\rangle}_{\text{dot product of 2 moments}} - \frac{2}{p} \sum_{i=1}^{p} \sum_{j=1}^{m} \lambda_j \underbrace{\left\langle \mathbf{x}_i^{\otimes d}, \mathcal{M}_j^{(d)} \right\rangle}_{\text{dot product of moment + vector}}$$

Example Calculation: $d = 3$

$$f(\theta) = C + \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j \left\langle \mathcal{M}_i^{(d)}, \mathcal{M}_j^{(d)} \right\rangle - \frac{2}{p} \sum_{i=1}^{p}\sum_{j=1}^{m} \lambda_j \left\langle \mathbf{x}_i^{\otimes d}, \mathcal{M}_j^{(d)} \right\rangle$$

$$\mathcal{M}_j^{(3)} = \boldsymbol{\mu}_j^{\otimes 3} + 3\,\mathrm{sym}(\boldsymbol{\mu}_j \otimes \boldsymbol{\Sigma}_j)$$

$$\begin{aligned} \langle \mathbf{x}_i^{\otimes 3}, \mathcal{M}_j^{(3)} \rangle &= \left\langle \mathbf{x}_i^{\otimes 3}, \boldsymbol{\mu}_j^{\otimes 3} \right\rangle + 3\left\langle \mathbf{x}_i^{\otimes 3}, \mathrm{sym}(\boldsymbol{\mu}_j \otimes \boldsymbol{\Sigma}_j) \right\rangle \\ &= (\mathbf{x}_i^\intercal \boldsymbol{\mu}_j)^3 + 3\left\langle \mathbf{x}_i^{\otimes 3}, \boldsymbol{\mu}_j \otimes \boldsymbol{\Sigma}_j \right\rangle \\ &= (\mathbf{x}_i^\intercal \boldsymbol{\mu}_j)^3 + 3(\mathbf{x}_i^\intercal \boldsymbol{\mu}_j)(\mathbf{x}_i^\intercal \boldsymbol{\Sigma}_j \mathbf{x}_i) \end{aligned}$$

$$\left\langle \mathbf{a}^{\otimes 3}, \mathrm{sym}(\mathcal{B}) \right\rangle = \left\langle \mathbf{a}^{\otimes 3}, \mathcal{B} \right\rangle$$

$$\left\langle \mathbf{a}^{\otimes 3}, \mathbf{b}^{\otimes 3} \right\rangle = (\mathbf{a}^\intercal \mathbf{b})^3$$

$$\left\langle \mathbf{a}^{\otimes 3}, \mathbf{b} \otimes \mathbf{C} \right\rangle = \mathbf{a}^\intercal \mathbf{b}\, \mathbf{a}^\intercal \mathbf{C} \mathbf{a}$$

Computing terms $\langle \mathbf{M}_i^{(d)}, \mathbf{M}_j^{(d)} \rangle$ more involved (Bell polynomials).

PART II: *Mixture of Products & Coupled Incomplete CP Decompositions*

Other Noise Models

Implicit moment tensor decomposition can be applied to mixtures with other noise models (e.g., Poisson noise).

- Focus on conditionally-independent mixtures in $\mathbb{R}^n$:

$$\mathcal{D} = \sum_{j=1}^{r} w_j \mathcal{D}_j = \sum_{j=1}^{r} w_j \bigotimes_{i=1}^{n} \mathcal{D}_{ij}$$

  for some distributions $\mathcal{D}_{ij}$ on $\mathbb{R}$. That is, conditional on latent variable, standard coords in $\mathbb{R}^n$ are indep. *[e.g. Hall-Zhou '03]*

- Make no parametric assumptions on $\mathcal{D}_{ij}$.

Precise Formulation

$$\mathcal{D} = \sum_{j=1}^{r} w_j \mathcal{D}_j = \sum_{j=1}^{r} w_j \bigotimes_{i=1}^{n} \mathcal{D}_{ij}.$$

Denote dth joint moment and componentwise moments by

$$\mathbf{M}^d = \mathbb{E}_{X \sim \mathcal{D}}[X^{\otimes d}] \in \mathbb{R}^{n^d} \quad \text{and} \quad \mathbf{m}_j^d = \mathbb{E}_{X \sim \mathcal{D}_j}[X^{*d}] \in \mathbb{R}^n,$$

where $*$ denotes entrywise power.

Precise Formulation

$$\mathcal{D} = \sum_{j=1}^{r} w_j \mathcal{D}_j = \sum_{j=1}^{r} w_j \bigotimes_{i=1}^{n} \mathcal{D}_{ij}.$$

Denote dth joint moment and componentwise moments by

$$\mathbf{M}^d = \mathbb{E}_{X\sim\mathcal{D}}[X^{\otimes d}] \in \mathbb{R}^{n^d} \quad \text{and} \quad \mathbf{m}_j^d = \mathbb{E}_{X\sim\mathcal{D}_j}[X^{*d}] \in \mathbb{R}^n,$$

where $*$ denotes entrywise power.

Problem

Given data $V \subset \mathbb{R}^n$ from a conditionally-independent mixture $\mathcal{D}$ of r components. We want to estimate the mixing weights w and componentwise moments $\{(\mathbf{m}_j^1, \mathbf{m}_j^2, \dots) : j = 1, \dots, r\}$ from the sample moment tensors $\widehat{\mathbf{M}}^1, \widehat{\mathbf{M}}^2, \dots$ without parametrizing $\mathcal{D}$.

Incomplete Tensor Decomposition: Up To $d = 3$

Unknowns $\{w_j, \mathbf{m}_j^1, \mathbf{m}_j^2, \mathbf{m}_j^3\}_{j\in[r]}$.

$$\begin{aligned}
\mathbf{M}^1 &= \textstyle\sum_{j=1}^r w_j \mathbf{m}_j^1 && \in \mathbb{R}^n \\
P(\mathbf{M}^2) &= P\big(\textstyle\sum_{j=1}^r w_j (\mathbf{m}_j^1)^{\otimes 2}\big) && \in S^2(\mathbb{R}^n) \\
P_{(2)}(\mathbf{M}^2) &= \textstyle\sum_{j=1}^r w_j \mathbf{m}_j^2 && \in \mathbb{R}^n \\
P(\mathbf{M}^3) &= P\big(\textstyle\sum_{j=1}^r w_j (\mathbf{m}_j^1)^{\otimes 3}\big) && \in S^3(\mathbb{R}^n) \\
P_{(2,1)}(\mathbf{M}^3) &= P\big(\textstyle\sum_{j=1}^r w_j \mathbf{m}_j^2 \otimes \mathbf{m}_j^1\big) && \in \mathbb{R}^{n\times n} \\
P_{(3)}(\mathbf{M}^3) &= \textstyle\sum_{j=1}^r w_j \mathbf{m}_j^3 && \in \mathbb{R}^n.
\end{aligned}$$

For general d, this becomes a coupled system of partially symmetric incomplete CP tensor decomposition problems.

[building on Guo-Nie-Yang '22]

Identifiability Bound

Theorem (Zhang-K. '23)

Let $\mathcal{D}$ be a conditionally independent mixture with positive weights $w \in \mathbb{R}^n$ and Zariski-generic means $A \in \mathbb{R}^{n \times r}$. Let $d_1, d_2 \in \mathbb{N}$ be distinct such that $2 < d_1 < n$, $r \leq \binom{\lfloor (n-1)/2 \rfloor}{\lfloor d_1/2 \rfloor}$ and $r \leq \binom{n}{d_2}$. Then w and A are uniquely determined from the equations

$$P\Big(\sum_{j=1}^{r} w_j {m^1}_j^{\otimes d_1}\Big) = P\Big(\mathbf{M}^{d_1}\Big), \quad P\Big(\sum_{j=1}^{r} w_j {m^1}_j^{\otimes d_2}\Big) = P\Big(\mathbf{M}^{d_2}\Big),$$

up to possible sign flips on each ${m^1}_j$ if d_1 and d_2 are both even.

Numerical Optimization

Use least squares cost function:

$$f^{[d]}(w, A; V) = \sum_{i=1}^{d} \tau_i \left\| P\left(\underbrace{\frac{1}{p} \sum_{\ell=1}^{p} v_\ell^{\otimes i}}_{\widehat{\mathbf{M}}^i} - \underbrace{\sum_{j=1}^{r} w_j a_j^{\otimes i}}_{\mathbf{M}^i} \right) \right\|^2 .$$

Residuals are multilinear in w (obvious) and each row of the mean matrix A (less obvious).

Algorithm 1 Baseline ALS algorithm for solving means and weights

function $\textsc{SolveMeanAndWeight}$(data $\boldsymbol{V}$, initialization $(\boldsymbol{w}, \boldsymbol{A})$, order d, hyperparameters $\boldsymbol{\tau}$)
  Compute $\{\boldsymbol{G}_s^{A,A}\}_{s=1}^{d}$ and $\{\boldsymbol{G}_s^{A,V}\}_{s=1}^{d}$
  while not converged **do**
    $\boldsymbol{A}, \{\boldsymbol{G}_s^{A,A}\}_{s=1}^{d}, \{\boldsymbol{G}_s^{A,V}\}_{s=1}^{d} \leftarrow \textsc{UpdateMean}(\{\boldsymbol{G}_s^{A,A}\}_{s=1}^{d}, \{\boldsymbol{G}_s^{A,V}\}_{s=1}^{d}, \boldsymbol{w}, \boldsymbol{A}, d, \boldsymbol{\tau})$
    $\boldsymbol{w} \leftarrow \textsc{UpdateWeight}(\{\boldsymbol{G}_s^{A,A}\}_{s=1}^{d}, \{\boldsymbol{G}_s^{A,V}\}_{s=1}^{d}, d, \boldsymbol{\tau})$
return $\boldsymbol{w}$, $\boldsymbol{A}$

Implicit Tensor Computations

- $\widehat{\mathbf{M}}^d$ needs $\mathcal{O}(n^d)$ to store and $\mathcal{O}(pn^d)$ flops to compute.
- Implicit: compute the normal equations for the least square solves without forming any higher-order tensors. Flops: $\mathcal{O}(npr + nr^3)$. Storage: $\mathcal{O}(n(r+p))$.
- Main thing is to efficiently evaluate the kernel:

$$K_d(x, y) = \langle P(x^{\otimes d}), P(y^{\otimes d}) \rangle$$

at $K_d(a_j, a_{j'})$ and $K_d(a_j, v_\ell)$.

Implicit Tensor Computations

- $\widehat{\mathbf{M}}^d$ needs $\mathcal{O}(n^d)$ to store and $\mathcal{O}(pn^d)$ flops to compute.
- Implicit: compute the normal equations for the least square solves without forming any higher-order tensors. Flops: $\mathcal{O}(npr + nr^3)$. Storage: $\mathcal{O}(n(r+p))$.
- Main thing is to efficiently evaluate the kernel:

$$K_d(x, y) = \langle P(x^{\otimes d}), P(y^{\otimes d}) \rangle$$

at $K_d(a_j, a_{j'})$ and $K_d(a_j, v_\ell)$.

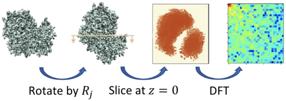
Lemma

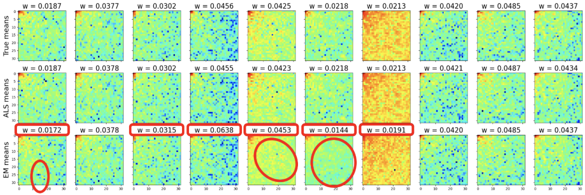
Let e_d be the elementary symmetric polynomial of degree d. Then $K_d(x, y) = d! \cdot e_d(x * y)$.

We evaluate e_d via Newton-Gerard identities, relating e_d to power sums.

Example: Clustering X-Ray Free Electron Laser Images

$$I_j := |\mathcal{PF}(\phi \circ R_j)| : \mathbb{R}^2 \to \mathbb{R}$$





- Here $n = 1024$, $r = 30$, $p = 20000$.
- Noise is pixelwise Poisson. Our algorithm doesn't know this, but EM does.
- We take ~ 40 min to converge. Error 0.9% in weights, 0.5% in means.
- EM is initialized with best of 30 k-means runs. We then run EM three times with different seeds. It takes $\sim 50 - 70$ min. Error in means is $> 13\%$.

CONCLUSIONS

Summary

- Moment formulas for general Gaussian Mixture Models and a tensor-based algorithm avoiding exponential cost in order d.
- Variations for non-Gaussian mixtures, with product structure.
- Competitive with non-tensor approaches; in some cases better.

References

Joao M. Pereira, Joe Kileel, Tamara G. Kolda, “Tensor moments of Gaussian mixture models: theory and applications”, *preprint 2022*.

Yifan Zhang, Joe Kileel, “Moment estimation for nonparametric mixture models through implicit tensor decomposition”, *SIAM Journal on Mathematics of Data Science 2023*.

Yulia Alexandr, Joe Kileel, Bernd Sturmfels, “Moment varieties for mixtures of products”, *ACM International Symposium on Symbolic and Algebraic Computation 2023*.

References

Joao M. Pereira, Joe Kileel, Tamara G. Kolda, "Tensor moments of Gaussian mixture models: theory and applications", *preprint 2022*.

Yifan Zhang, Joe Kileel, "Moment estimation for nonparametric mixture models through implicit tensor decomposition", *SIAM Journal on Mathematics of Data Science 2023*.

Yulia Alexandr, Joe Kileel, Bernd Sturmfels, "Moment varieties for mixtures of products", *ACM International Symposium on Symbolic and Algebraic Computation 2023*.

THANK YOU!