Unknown: $O(dk)$

Gaussian mixture.

$$x = Ah + z, \quad h \in \{\vec{e}_1, \dots, \vec{e}_k\}, \quad A = [\vec{\mu}_1 | \vec{\mu}_2 | \dots] \in \mathbb{R}^{d \times k}$$

$$z \sim \mathcal{N}(0, \sigma^2 I).$$

Ignore.

$$E(xx^T) = E(x \otimes x) = E\left((Ah)(Ah)^T\right) + \cancel{\sigma^2} I$$

$$= E(A h h^T A^T) = A\, E(\vec{h}\vec{h}^T) A^T = A\, \mathrm{Diag}(\vec{w}) A^T$$

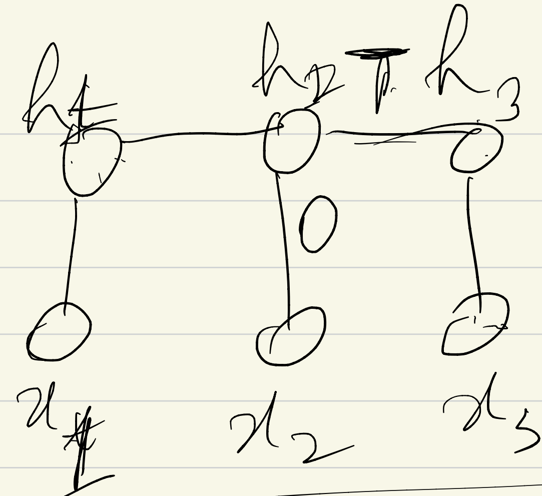
$$= E\left((Ah) \otimes (Ah)\right) = E(h \otimes h)[A, A]$$

$$= \mathrm{Diag}(\vec{w})[A, A].$$

$$= \sum_i w_i (\vec{\mu}_i \otimes \vec{\mu}_i)$$

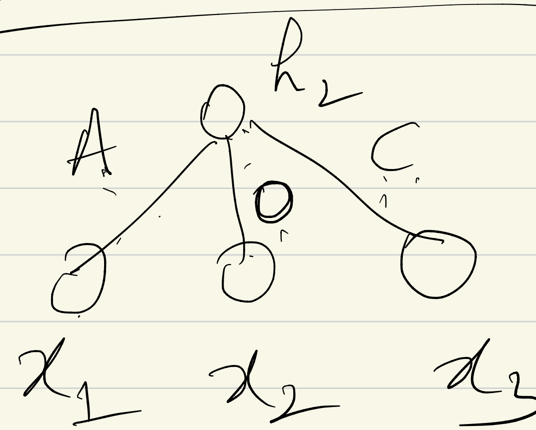
$$E\left((Ah) \otimes (Ah) \otimes (Ah)\right) = \sum_i w_i\, \vec{\mu}_i \otimes \vec{\mu}_i \otimes \vec{\mu}_i$$

CP decomposition.



$$E(\vec{x}_1 \otimes \vec{x}_2 \otimes \vec{x}_3)$$

$$= E\left(E(\vec{x}_1 \otimes \vec{x}_2 \otimes \vec{x}_3 \mid h_2)\right)$$



$$E(x_1 \mid h_2) = A h_2.$$

$$E(x_2 \mid h_2) = O h_2.$$

$$E(x_3 \mid h_2) = C h_2.$$

$$= E\left((Ah_2) \otimes (Oh_2) \otimes (Ch_2)\right)$$

$$= E(h_2 \otimes h_2 \otimes h_2)\,[A, O, C]$$

$$= \sum_i w_i \; \vec{a}_i \otimes \vec{o}_i \otimes \vec{c}_i$$

$\rightarrow$ CP-decomposition.

$$M = \sum_{i=1}^{k} \lambda_i (\vec{u}_i \otimes \vec{u}_i)$$

Uniqueness: λ_i are not repeated

u_i are <u>orthogonal</u>

How many? $k \le d$ (Rank)

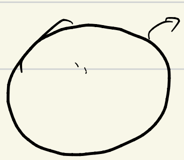
Method? power iteration

$$v \mapsto \frac{Mv}{\|Mv\|}$$ <u>converges to $\vec{u}_1$</u>

→ <u>linear map</u>

Rayleigh quotient:

$$\max_{\|u\|_2 = 1} u^T M u.$$



surface of sphere

$$T = \sum \lambda_i (\vec{u}_i \otimes \vec{u}_i \otimes \vec{u}_i)$$

Uniqueness: $\vec{u}_i$ are <u>linearly independent</u>

(sufficient but not necessary)

How many? $k > d$ (Rank)

(Kruskal condition)

Method: $T(v, v, \cdot)$

$$v \mapsto \frac{T(v, v, \cdot)}{\|T(v, v, \cdot)\|}$$

→ <u>quadratic map</u>.

Rayleigh quotient:

$$\max_{\|u_2\| = 1} T(u, u, u)$$

u_i's are orthonormal.

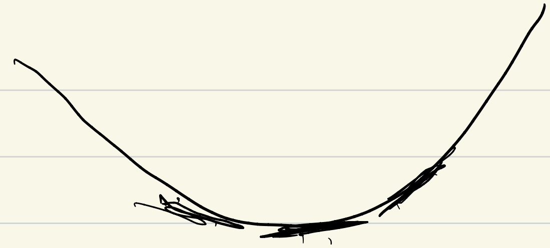
$$\max_{\vec{u}} \ \vec{u}^T M \vec{u} - \lambda(\vec{u}^T\vec{u} - 1)$$

$$2(M\vec{u} - \lambda\vec{u}) = 0.$$

$$M\vec{u} = \lambda\vec{u}. \rightarrow \text{eigen vector/value.}$$

$$\max_{u} \ T(u, u, u) - \lambda(u^T u - 1)$$

$$\min f(x)$$



if f is convex.

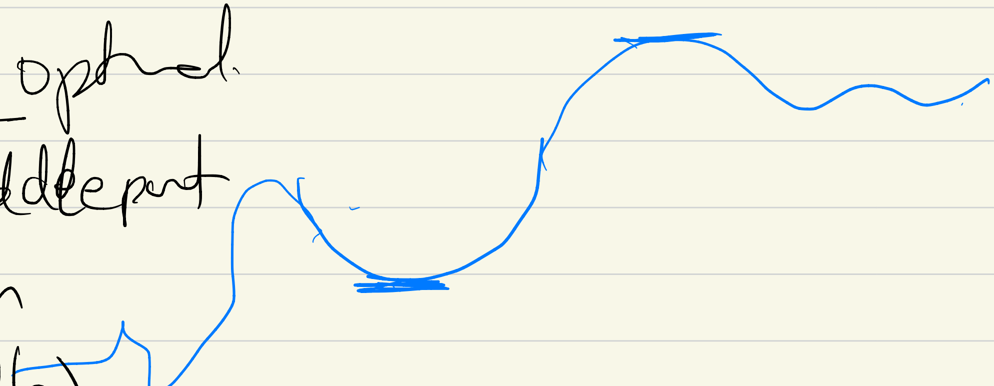
$$\nabla f(x) = 0$$ (first order stationary)

$$\nabla^2 f(x) \succ 0$$

Globally optimal.

$$\nabla^2 f(x) \prec 0.$$

If not convex. locally optimal.



Gradient descent - local min + saddle point

+ noise (stochastic gradient descent) - local min (escapes saddle) neither local min or max.

Saddle point : $\nabla^2 f(x)$ is neither +ve or -ve definite

+ve eigen vectors - locally min along these directions

-ve eigen vectors - locally max along these directions

SGD escapes saddle point with sufficient noise.