A Dictionary Learning Approach for Factorial Gaussian Models - Supplemental Material

Anonymous Author(s)
Affiliation
Address
email

1 Proofs

To reduce the notation clutter we drop tilde's, although we still refer to the SC-FM parameters, and we use the regular factorial model notation where the indicator variable $r_t^k \in [M]$, for $k \in [K]$. Conforming with that notation we set the last columns of all the emission matrices to be the shared component, such that $\mu_M^k = s, \forall k \in [K]$.

Definition 1. *Let x_l denote l'th column of X^c, so $x_l := X^c(:,l) = \sum_{k=1}^{K}\sum_{m=1}^{M-1}\mu_m^k r_{m,l}^k + \sum_{k=1}^{K} s\ r_{M,l}^k$, where $r_{m,l}^k, l \in [M^K]$ denotes the m'th entry of an indicator vector of length M where only the m'th entry is one and the rest is zero, for the k'th emission matrix and l'th possible combination.*

Definition 2. *We define the index sets for three different 'types' of terms. $\mathcal{D}1 := \left\{l \in [M^K] : \sum_{k=1}^{K} r_{M,l}^k = 0\right\}$, which corresponds to the terms of the form $\sum_{k=1}^{K}\sum_{m=1}^{M-1}\mu_m^k r_{m,l}^k$, $l \in \mathcal{D}1$; $\mathcal{D}2 := \{l \in [M^K] : \sum_{k=1}^{K} r_{M,l}^k = K\}$, which corresponds to the term Ks; $\mathcal{D}3 := [M^K]\setminus\{\mathcal{D}1 \cup \mathcal{D}2\}$, which corresponds to the terms of the form $\sum_{k=1}^{K}\sum_{m=1}^{M-1}\mu_m^k r_{m,l}^k + \sum_{k=1}^{K} s\ r_{M,l}^k$, $l \in \mathcal{D}3$.*

Definition 3. *Let $v(x_{l'}) : \mathbb{R}^L \to \mathbb{R}^{M^K}$ denote a vector valued function with the argument $x_{l'}$, such that $v(x_{l'}) = \omega\left([\langle x_1, x_{l'}\rangle, \langle x_2, x_{l'}\rangle, \ldots, \langle x_l, x_{l'}\rangle, \ldots, \langle x_{M^K}, x_{l'}\rangle]\right)$, where $\omega : [M^K] \to [M^K]$ is an ascending sorting mapping such that $v_1(x_{l'}) \le v_2(x_{l'}) \le \cdots \le v_{M^K}(x_{l'})$, where $v_l(x_{l'})$ is the l'th smallest element in $v(x_{l'})$ vector.*

Lemma 1. *If $\left\langle \mu_{m''}^{k''}, s\right\rangle \le \left\langle \mu_m^k, \mu_{m'}^{k'}\right\rangle$, $\forall (k,k',k'') \in [K]$, and $\forall (m,m',m'') \in [M-1]$, i.e. for any component μ_m^k, the least correlated component is s, and $\left\langle \mu_m^k, s\right\rangle \le \langle s, s\rangle$, $\forall k \in [K]$, $m \in [M-1]$, i.e., the shared component s has a non-trivial magnitude (e.g. all zeros vector doesn't satisfy this condition), then*

$$Ks = \operatorname*{arg\,min}_{x_{l'}, l' \in [M^K]} \sum_{l=1}^{(M-1)^K} v_l(x_{l'}),\ \textit{for}\ M > 2, K \ge 1. \tag{1}$$

Proof: The general inner product expression is as follows:

$$\begin{aligned}\langle x_l, x_{l'}\rangle &= \left\langle \sum_{k=1}^{K}\sum_{m=1}^{M-1}\mu_m^k r_{m,l}^k + \sum_{k=1}^{K} s\, r_{M,l}^k, \sum_{k'=1}^{K}\sum_{m'=1}^{M-1}\mu_{m'}^{k'} r_{m',l'}^{k'} + \sum_{k'=1}^{K} s\, r_{M,l'}^{k'} \right\rangle \\ &= \sum_{k,k'=1}^{K}\sum_{m,m'=1}^{M-1}\left\langle \mu_m^k, \mu_{m'}^{k'}\right\rangle r_{m,l}^k r_{m',l'}^{k'} + \sum_{k,k'=1}^{K}\sum_{m=1}^{M-1}\left\langle \mu_m^k, s\right\rangle r_{m,l}^k r_{M,l'}^{k'} \\ &\quad + \sum_{k,k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{M,l}^k r_{m',l'}^{k'} + \sum_{k,k'=1}^{K}\langle s, s\rangle r_{M,l}^k r_{M,l'}^{k'}. \end{aligned} \tag{2}$$

Having seen the most general equation, let's consider special cases for $x_{l'}$ separately:

- $v(x_{l'''}) = v(Ks)$, $l''' \in \mathcal{D}2$:

$$\begin{aligned}\langle x_l, x_{l'''}\rangle &= \sum_{k,k'''=1}^{K}\sum_{m=1}^{M-1}\left\langle \mu_m^k, s\right\rangle r_{m,l}^k r_{M,l'''}^{k'''} + \sum_{k,k'''=1}^{K}\langle s, s\rangle r_{M,l}^k r_{M,l'''}^{k'''} \\ &= K\sum_{k=1}^{K}\sum_{m=1}^{M-1}\left\langle \mu_m^k, s\right\rangle r_{m,l}^k + K\sum_{k=1}^{K}\langle s, s\rangle r_{M,l}^k \end{aligned} \tag{3}$$

  We observe that there are $(M-1)^K$ terms which are of the form $\sum_{k=1}^{K}\sum_{m=1}^{M-1}\left\langle \mu_m^k, s\right\rangle r_{m,l}^k$. We conclude that,

$$v_l(Ks) = K\sum_{k=1}^{K}\sum_{m=1}^{M-1}\left\langle \mu_m^k, s\right\rangle r_{m,\omega(l)}^k,\ l \in [(M-1)^K], \tag{4}$$

  where $\omega : M^K \to M^K$ is the sorting mapping of $v(Ks)$ function.

- $v(x_{l'}) = v\left(\sum_{k'=1}^{K}\sum_{m'=1}^{M-1}\mu_{m'}^{k'} r_{m',l'}^{k'}\right)$, for $l' \in \mathcal{D}1$:

$$\langle x_l, x_{l'}\rangle = \sum_{k,k'=1}^{K}\sum_{m,m'=1}^{M-1}\left\langle \mu_m^k, \mu_{m'}^{k'}\right\rangle r_{m,l}^k r_{m',l'}^{k'} + \sum_{k,k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{M,l}^k r_{m',l'}^{k'}$$

  Notice that we only have one term of the form $K\sum_{k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{m',l'}^{k'},\ l' \in \mathcal{D}1$. Consequently,

$$v_1(x_{l'}) = K\sum_{k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{m',l'}^{k'},\ l' \in \mathcal{D}1.$$

  And the larger elements for $2 \le l \le (M-1)^K$ are of the form:

$$\begin{aligned} v_l(x_{l'}) &= \sum_{k,k'=1}^{K}\sum_{m,m'=1}^{M-1}\left\langle \mu_m^k, \mu_{m'}^{k'}\right\rangle r_{m,\omega(l)}^k r_{m',l'}^{k'} + \sum_{k,k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{M,\omega(l)}^k r_{m',l'}^{k'} \\ &= \sum_{k,k'=1}^{K}\sum_{m'=1}^{M-1}\left(\sum_{m=1}^{M-1}\left\langle \mu_m^k, \mu_{m'}^{k'}\right\rangle r_{m,\omega(l)}^k r_{m',l'}^{k'} + \left\langle s, \mu_{m'}^{k'}\right\rangle r_{M,\omega(l)}^k r_{m',l'}^{k'}\right) \\ &\ge K\sum_{k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{m',l'}^{k'} = v_l(Ks),\ l' \in \mathcal{D}1,\ 2 \le l \le (M-1)^K, \end{aligned}$$

  where the inequality is due to the first incoherence condition in the Lemma. Therefore we conclude that $\sum_{l=1}^{(M-1)^K} v_l(Ks) \le \sum_{l=1}^{(M-1)^K} v_l(x_{l'}), \forall l' \in \mathcal{D}1$.

- $v(x_{l''}) = v\left(\sum_{k'=1}^{K}\sum_{m'=1}^{M-1}\mu_{m'}^{k'}r_{m',l''}^{k'} + \sum_{k'=1}^{K} s r_{M,l''}^{k'}\right)$, for $l'' \in \mathcal{D}3$. In this case none of the terms vanish in equation (2):

$$\begin{aligned}
v_l(x_{l''}) = &\sum_{k,k'=1}^{K}\sum_{m,m'=1}^{M-1}\left\langle \mu_m^k, \mu_{m'}^{k'}\right\rangle r_{m,\omega(l)}^{k} r_{m',l''}^{k'} + \sum_{k,k'=1}^{K}\sum_{m=1}^{M-1}\left\langle \mu_m^k, s\right\rangle r_{m,\omega(l)}^{k} r_{M,l''}^{k'} \\
&+ \sum_{k,k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{M,\omega(l)}^{k} r_{m',l''}^{k'} + \sum_{k,k'=1}^{K}\left\langle s, s\right\rangle r_{M,\omega(l)}^{k} r_{M,l''}^{k'} \\
\geq &K\sum_{k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{m',l'}^{k'} = v_l(Ks),\ l' \in \mathcal{D}1,\ l'' \in \mathcal{D}3, \text{and},\ 2 \leq l \leq (M-1)^K,
\end{aligned}$$

where the inequality is due to the incoherence conditions. Therefore, we conclude that $\sum_{l=1}^{(M-1)^K} v_l(Ks) \leq \sum_{l=1}^{(M-1)^K} v_l(x_{l''})\ \forall l'' \in \mathcal{D}3$. Together with the conclusion, $\sum_{l=1}^{(M-1)^K} v_l(Ks) \leq \sum_{l=1}^{(M-1)^K} v_l(x_{l'}), \forall l' \in \mathcal{D}1$, we see that the claim in the Lemma is true. □

1.1 Finding Ks term in $M = 2$ case:

Lemma 2. *If $\left\langle \mu_{m''}^{k''}, s\right\rangle \leq \left\langle \mu_m^k, \mu_{m'}^{k'}\right\rangle$, $\forall (k,k',k'') \in [K]$, and $\forall (m,m',m'') \in [1]$, i.e. for any component μ_m^k, the least correlated component is s, $\left\langle \mu_m^k, s\right\rangle \leq \langle s, s\rangle$, $\forall k \in [K]$, $m \in [1]$, i.e., the shared component s has a non-trivial magnitude (e.g. all zeros vector doesn't satisfy this condition), and $\left\langle \mu_m^k, \mu_m^{k'}\right\rangle \leq \langle s, s\rangle$, $\forall k,k' \in [K]$, $m \in [1]$, then*

$$Ks = \underset{x_{l'}, l' \in [2^K]}{\arg\max}\left(\sum_{l=2}^{2^K-1} v_l(x_{l'})\right), \text{ for } K \geq 1. \tag{5}$$

Proof: We know from Lemma 3 in the paper that Ks is a minimizer of the term $v_1(x_{l'})$. Going through separate cases like we did for the Proof of Lemma 3,

- $v(x_{l'''}) = v(Ks)$, $l''' \in \mathcal{D}2$:

$$\begin{aligned}
\langle x_l, x_{l'''}\rangle = &\sum_{k,k'''=1}^{K}\sum_{m=1}^{M-1}\left\langle \mu_m^k, s\right\rangle r_{m,l}^{k} r_{M,l'''}^{k'''} + \sum_{k,k'''=1}^{K}\langle s, s\rangle r_{M,l}^{k} r_{M,l'''}^{k'''} \\
= &K\sum_{k=1}^{K}\sum_{m=1}^{M-1}\left\langle \mu_m^k, s\right\rangle r_{m,l}^{k} + K\sum_{k=1}^{K}\langle s, s\rangle r_{M,l}^{k} \\
= &K\sum_{k=1}^{K}\left\langle \mu_1^k, s\right\rangle r_{1,l}^{k} + K\sum_{k=1}^{K}\langle s, s\rangle r_{M,l}^{k}
\end{aligned}$$

- $v(x_{l'}) = v\left(\sum_{k'=1}^{K}\sum_{m'=1}^{M-1}\mu_{m'}^{k'}r_{m',l'}^{k'}\right)$, for $l' \in \mathcal{D}1$:

$$\begin{aligned}
\langle x_l, x_{l'}\rangle = &\sum_{k,k'=1}^{K}\sum_{m,m'=1}^{M-1}\left\langle \mu_m^k, \mu_{m'}^{k'}\right\rangle r_{m,l}^{k} r_{m',l'}^{k'} + \sum_{k,k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{M,l}^{k} r_{m',l'}^{k'} \\
= &\sum_{k,k'=1}^{K}\left\langle \mu_1^k, \mu_1^{k'}\right\rangle r_{m,l}^{k} r_{m',l'}^{k'} + \sum_{k,k'=1}^{K}\left\langle s, \mu_1^{k'}\right\rangle r_{M,l}^{k} r_{1,l'}^{k'} \\
\leq &K\sum_{k=1}^{K}\left\langle \mu_1^k, s\right\rangle r_{1,l}^{k} + K\sum_{k=1}^{K}\langle s, s\rangle r_{M,l}^{k} = \langle x_l, x_{l'''}\rangle
\end{aligned}$$

- $v(x_{l''}) = v\left(\sum_{k'=1}^{K}\sum_{m'=1}^{M-1}\mu_{m'}^{k'} r_{m',l''}^{k'} + \sum_{k'=1}^{K} s r_{M,l''}^{k'}\right)$, for $l'' \in \mathcal{D}3$:

$$\begin{aligned}
\langle x_l, x_{l''}\rangle =& \sum_{k,k'=1}^{K}\sum_{m,m'=1}^{M-1}\left\langle \mu_m^k, \mu_{m'}^{k'}\right\rangle r_{m,(l)}^{k} r_{m',l''}^{k'} + \sum_{k,k'=1}^{K}\sum_{m=1}^{M-1}\left\langle \mu_m^k, s\right\rangle r_{m,(l)}^{k} r_{M,l''}^{k'} \\
&+ \sum_{k,k'=1}^{K}\sum_{m'=1}^{M-1}\left\langle s, \mu_{m'}^{k'}\right\rangle r_{M,(l)}^{k} r_{m',l''}^{k'} + \sum_{k,k'=1}^{K}\langle s, s\rangle r_{M,(l)}^{k} r_{M,l''}^{k'} \\
=& \sum_{k,k'=1}^{K}\left\langle \mu_1^k, \mu_1^{k'}\right\rangle r_{1,(l)}^{k} r_{1,l''}^{k'} + \sum_{k,k'=1}^{K}\left\langle \mu_1^k, s\right\rangle r_{1,(l)}^{k} r_{M,l''}^{k'} \\
&+ \sum_{k,k'=1}^{K}\left\langle s, \mu_1^{k'}\right\rangle r_{M,(l)}^{k} r_{1,l''}^{k'} + \sum_{k,k'=1}^{K}\langle s, s\rangle r_{M,(l)}^{k} r_{M,l''}^{k'} \\
\leq& K\sum_{k=1}^{K}\left\langle \mu_1^k, s\right\rangle r_{1,l}^{k} + K\sum_{k=1}^{K}\langle s, s\rangle r_{M,l}^{k} = \langle x_l, x_{l'''}\rangle,
\end{aligned}$$

where all of the inequalities are due to the incoherence conditions. Therefore, we conclude that $\sum_{l=2}^{2^K-1} v_l(Ks) \geq \sum_{l=2}^{2^K-1} v_l(x_{l''})\ \forall l'' \in \mathcal{D}3$, and $\sum_{l=2}^{2^K-1} v_l(Ks) \geq \sum_{l=2}^{2^K-1} v_l(x_{l'})$, $\forall l' \in \mathcal{D}1$, we see that the claim in the Lemma is true. $\square$