Supplementary Material for Learning Ordered Representations with Nested Dropout

A. Proofs for Section 3.2

Theorem 1. *Every optimal solution of the nested dropout problem is necessarily an optimal solution of the standard autoencoder problem.*

Proof. Let the nested dropout autoencoder be of latent dimension K. Recall that the nested dropout objective function in Equation (11) is a strictly positive mixture of the K different b-truncation problems. As described in Subsection 3.1, an optimal solution to each b-truncation must be of the form $\boldsymbol{X}_b^* = \boldsymbol{T}_b\boldsymbol{\Sigma}_{\downarrow b}\boldsymbol{R}^T, \boldsymbol{\Gamma}_b^* = \boldsymbol{Q}_{\downarrow b}\boldsymbol{T}_b^{-1}$ for some invertible transformation $\boldsymbol{T}_b$. We note that the PCA decomposition is a particular optimal solution for each b that is given for the choice $\boldsymbol{T}_b = \mathbb{I}_b$. As such, the PCA decomposition exactly minimizes *every* term in the nested dropout mixture, and therefore must be a global solution of the nested dropout problem. This means that every optimal solution of the nested dropout problem must exactly minimize every term in the nested dropout mixture. In particular, one of these terms corresponds to the K-truncation problem, which is in fact the original autoencoder problem. ∎

Denote $\boldsymbol{T}_{\downarrow b} = \boldsymbol{J}_{K\to b}\boldsymbol{T}\boldsymbol{J}_{K\to b}^T$ as the b-th leading principal minor and its its bottom right corner as $t_b = T_{bb}$.

Lemma 1. *Let $\boldsymbol{T} \in \mathbb{R}^{K\times K}$ be commutative in its truncation and inversion. Then all the diagonal elements of $\boldsymbol{T}$ are nonzero, and for each $b = 2,\ldots,K$, either $\boldsymbol{A}_b = \boldsymbol{0}$ or $\boldsymbol{B}_b = \boldsymbol{0}$.*

Proof. We have $\det \boldsymbol{T}_{\downarrow b} = \det \boldsymbol{T}_{\downarrow b-1}\det(t_b - \boldsymbol{B}_b\boldsymbol{T}_{\downarrow b-1}^{-1}\boldsymbol{A}_b) \neq 0$ since $\boldsymbol{T}_{\downarrow b-1}$ is invertible. Since $\boldsymbol{T}_{\downarrow b-1}$ is also invertible, then $t_b - \boldsymbol{B}_b\boldsymbol{T}_{\downarrow b-1}^{-1}\boldsymbol{A}_b \neq 0$. As such, we write $\boldsymbol{T}_{\downarrow b}$ in terms of blocks $\boldsymbol{T}_{\downarrow b-1}, \boldsymbol{A}_b, \boldsymbol{B}_b, t_b$, and apply blockwise matrix inversion to find that $\boldsymbol{T}_{\downarrow b-1}^{-1} = \boldsymbol{T}_{\downarrow b-1}^{-1} + \boldsymbol{T}_{\downarrow b-1}^{-1}\boldsymbol{A}_b(t_b - \boldsymbol{B}_b\boldsymbol{T}_{\downarrow b-1}^{-1}\boldsymbol{A}_b)^{-1}\boldsymbol{B}_b\boldsymbol{T}_{\downarrow b-1}^{-1}$ which reduces to $\boldsymbol{A}_b\boldsymbol{B}_b = \boldsymbol{0}$. Now, assume by contradiction that $t_b = 0$. This means that either bottom row or the rightmost column of $\boldsymbol{T}_{\downarrow b}$ must be all zeros, which contradicts with the invertibility of $\boldsymbol{T}_{\downarrow b}$. ∎

Theorem 2. *Every optimal solution of the nested dropout problem must be of the form*

$$\boldsymbol{X}^* = \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{R}^T \tag{1}$$
$$\boldsymbol{\Gamma}^* = \boldsymbol{Q}\boldsymbol{T}^{-1}, \tag{2}$$

for some matrix $\boldsymbol{T} \in \mathbb{R}^{K\times K}$ that is commutative in its truncation and inversion.

Proof. Consider an optimal solution $\boldsymbol{X}^*, \boldsymbol{\Gamma}^*$ of the nested dropout problem. For *each* b-truncation, as established in the proof of Theorem 1, it must hold that

$$\boldsymbol{X}_b^* = \boldsymbol{T}_b\boldsymbol{J}_{K\to b}\boldsymbol{\Sigma}\boldsymbol{R}^T \tag{3}$$
$$\boldsymbol{\Gamma}_b^* = \boldsymbol{Q}\boldsymbol{J}_{K\to b}^T\boldsymbol{T}_b^{-1}. \tag{4}$$

However, it must also be true that $\boldsymbol{X}_b = \boldsymbol{X}_{\downarrow b}, \boldsymbol{\Gamma}_b = \boldsymbol{\Gamma}_{\downarrow b}$ by the definition of the nested dropout objective in Equation (11). The first equation thus gives that $\boldsymbol{T}_b\boldsymbol{J}_{K\to b} = \boldsymbol{J}_{K\to b}\boldsymbol{T}_K$, and therefore $\boldsymbol{T}_b = \boldsymbol{J}_{K\to b}\boldsymbol{T}_K\boldsymbol{J}_{K\to b}^T = \boldsymbol{T}_{\downarrow b}$. This establishes the fact that the optimal solution for each b-truncation problem simply draws the b-th leading principal minor from the *same* "global" matrix $\boldsymbol{T} := \boldsymbol{T}_K$. The second equation implies that for every b, it holds that $\boldsymbol{J}_{K\to b}\boldsymbol{T}^{-1}\boldsymbol{J}_{K\to b}^T = (\boldsymbol{J}_{K\to b}\boldsymbol{T}\boldsymbol{J}_{K\to b}^T)^{-1}$ and as such $\boldsymbol{T}$ is commutative in its truncation and inversion. ∎

Theorem 3. *Under the orthonormality constraint $\boldsymbol{\Gamma}^T\boldsymbol{\Gamma} = \mathbb{I}_K$, there exists a unique optimal solution for the nested dropout problem, and this solution is exactly the set of the K top eigenvectors of the covariance of $\boldsymbol{Y}$, ordered by eigenvalue magnitude. Namely, $\boldsymbol{X}^* = \boldsymbol{\Sigma}\boldsymbol{R}^T, \boldsymbol{\Gamma}^* = \boldsymbol{Q}$.*

Proof. The orthonormality constraint implies $(\boldsymbol{T}^{-1}\boldsymbol{Q})^T\boldsymbol{Q}\boldsymbol{T}^{-1} = \mathbb{I}_K$ which gives $\boldsymbol{T}^T = \boldsymbol{T}^{-1}$. Hence every row and every column must have unit norm. We also have have that for every $b = 1,\ldots,K$

$$
\begin{aligned}
\boldsymbol{T}_{\downarrow b}^T &= (\boldsymbol{J}_{K\to b}\boldsymbol{T}\boldsymbol{J}_{K\to b}^T)^T & (5)\\
&= \boldsymbol{J}_{K\to b}\boldsymbol{T}^T\boldsymbol{J}_{K\to b}^T & (6)\\
&= \boldsymbol{J}_{K\to b}\boldsymbol{T}^{-1}\boldsymbol{J}_{K\to b}^T & (7)\\
&= (\boldsymbol{J}_{K\to b}\boldsymbol{T}\boldsymbol{J}_{K\to b}^T)^{-1} & (8)\\
&= \boldsymbol{T}_{\downarrow b}^{-1} & (9)
\end{aligned}
$$

where in the last equation we applied Lemma 1 to Theorem 2. As such, every leading principal minor is also orthonormal. For the sake of contradiction, assume there exist some $m, n, m \neq n$ such that $T_{mn} \neq 0$. Without loss of generality assume $m < n$. Then $\sum_{p=1}^{n-1} T_{mp}^2 < 1$, but this violates the orthonormality of $\boldsymbol{T}_{n-1}$. Thus it must be that the diagonal elements of $\boldsymbol{T}$ are all identically 1, and therefore $\boldsymbol{T} = \mathbb{I}_K$. The result follows. ■