1 Proofs

Proof of Proposition 4. If Z is not degenerated, then Laplace’s method yields Eq. (10). By collecting from Eq. (10) the terms that depend on Z, we obtain

\[ p(Z | X, K) \propto p(Z, X | \hat{\Pi}, K) |F_{\hat{\Pi}}|^{-1/2}(1 + O(N^{-1})). \]  

(101)

If p(Z | X, K) is degenerated, we consider the transformation (11). Here, the transformed prior \( \tilde{p}(\Pi | K_0) \) would differ from the original prior \( p(\Pi | K_0) \). However, since the mapping \( \Pi \rightarrow \tilde{\Pi}_{K_0} \) is onto A1 and the prior is strictly positive in the whole space of \( \Pi \in A_4 \), \( \tilde{p}(\Pi | K_0) \) is also strictly positive, including \( \tilde{\Pi}_{K_0} = \arg\max_{\Pi_{K_0}} \ln p(X, \tilde{Z}_{K_0} | \Pi_{K_0}, K_0) \). Consequently, we can again use Laplace’s method for \( \ln p(X, \tilde{Z}_{K_0} | \hat{\Pi}_{K_0}, K_0) \), and by collecting the terms that depend on Z, we obtain

\[ p(X | Z, K) \propto p(X, \tilde{Z}_{K_0} | \hat{\Pi}_{K_0}, K_0) |F_{\hat{\Pi}_{K_0}}|^{-1/2}(1 + O(N^{-1})) \]

(102)

\[ \propto p_{K_0}(\tilde{Z}_{K_0}, K_0)(1 + O(N^{-1})). \]  

(103)

This concludes the proof.

Proof of Theorem 2. First, we prove the case that \( p(Z | X, K) \) is not degenerated. In that case, Laplace’s approximation yields Eq. (10) in probability, and substituting Eq. (10) into (7) gives (8).

If \( \kappa(p(Z | X, K)) = K' < K \), Proposition 4 gives us that \( p(Z | X, K) \propto p_{K'}(Z)(1 + O(N^{-1})). \) Since

\[ \mathbb{E}_{p(Z|X,K)}[\ln p(X, Z | K)] = \mathbb{E}_{p_{K'}}[\ln p(X, Z | K)] + O(1) \]

and

\[ H(p(Z | X, K)) = (1 + O(N^{-1}))H(p_{K'}) + (1 + O(N^{-1}))\ln(1 + O(N^{-1})) \]

\[ = H(p_{K'}) + O(1), \]

1
\[ \ln p(X | K) \] is rewritten by

\[
\begin{align*}
E_{p_{K'}}[\ln p(X, Z | K)] + H(p_{K'}) + O(1) \\
= E_{p_{K'}}[\mathcal{L}(Z_{K'}, \hat{\Pi}_{K'}, K')] + H(p_{K'}) + O(1)
\end{align*}
\]

(104) (105)

Here, since the projection \( T_{K'} : Z \rightarrow Z_{K'} \) is continuous and onto (A1), we can describe \( p_{K'}(Z) \) as the density of \( Z_{K'} \) by using a change of variables, which we denote by \( \tilde{p}_{K'}(Z_{K'}) \). Now, we can rewrite the first term as the integral over \( Z_{K'} \), i.e.,

\[
E_{p_{K'}}[\mathcal{L}(Z_{K'}, \hat{\Pi}_{K'}, K')] = \int \mathcal{L}(T_{K'}(Z), \hat{\Pi}_{K'}, K') p_{K'}(T_{K'}(Z)) |Z|
\]

(106)

Similarly, \( gFIC(K') \) is rewritten using Proposition 4 as

\[
gFIC(K') = E_{p_{K'}}[\mathcal{L}(Z_{K'}, \hat{\Pi}_{K'}, K')] + H(p_{K'}) + O(1)
\]

(108)

Again, the first term is written as

\[
E_{p_{K'}}[\mathcal{L}(Z_{K'}, \hat{\Pi}_{K'}, K')] = \int \mathcal{L}(Z_{K'}, \hat{\Pi}_{K'}, K') p_{K'}(Z_{K'}) |Z_{K'}|
\]

(109)

Since Eq. (107) and (110) are the same, this concludes Eq. (8).

\[ \Box \]

**Proof of Proposition 6.** Proposition 4 shows that, if \( Z \) is non-degenerated,

\[
p(Z | X, K) \propto p(X, Z | \hat{\Pi})|F_{\hat{\Pi}}|^{-1/2} \propto \prod_{n} p(x_n, z_n | \tilde{\Pi})|\hat{\Pi}||F_{\hat{\Pi}}|^{-1/2N}
\]

(111) (112)

Since \( \ln |F_{\hat{\Pi}}| = O(1), |\hat\Pi|^{-1/2N} \) quickly diminishes to 1 for \( N \to \infty \).

\[ \Box \]

**Proof of Proposition 7.** For technical reasons, we redefine the estimators as follows:

\[
\hat{\Pi} \equiv \arginf_{\Pi} g_N(\Pi) = \arginf_{\Pi} \frac{1}{N} \ln p(X, Z | \Pi),
\]

(113)

\[
\tilde{\Pi} \equiv \arginf_{\Pi} G_N(\Pi) = \arginf_{\Pi} E_{\hat{\Pi}}[\frac{1}{N} \ln p(X, Z | \Pi)].
\]

(114)

According to A5, \( g_N(\Pi) \) is continuous and concave, and it uniformly converges to \( G_N(\Pi) \), i.e.,

\[
\sup_{\Pi \in \mathcal{P}} |g_N(\Pi) - G_N(\Pi)| \overset{P}{\to} 0.
\]

(115)

This suffices to show the consistency (for example, see Theorem 5.7 in van der Vaart (1998)).

\[ \Box \]
2 The gFAB Algorithm of BPCA

Let $Q$ be restricted as Gaussian with mean field. Then, the form of $q$ is determined as $q(Z) = \prod_n N(z_n | \mu_n, \Omega)$. By substituting the BPCA’s likelihood into Eq. (17), we obtain

$$\max_{q \in Q} E_q \left[ -\frac{\lambda}{2} \|X - ZW^T\|_{\text{Fro}}^2 - \frac{1}{2} \|Z\|_{\text{Fro}}^2 - \frac{1}{2} \ln |F| \right] + \frac{ND}{2} \ln \lambda + H(q) + \text{const}.$$ 

$$\geq \max_{q \in Q} E_q \left[ -\frac{\lambda}{2} \|X - ZW^T\|_{\text{Fro}}^2 - \frac{1}{2} \|Z\|_{\text{Fro}}^2 - D(\ln |N^{-1}Z^TZ| + K \ln \lambda) \right] + \frac{ND}{2} \ln \lambda + \frac{N}{2} \ln |\Omega| + \text{const}.$$ 

$$= \max_{q \in Q} \frac{\lambda}{2} (\tr(W^TWS - XW\mathbb{E}[Z^T])) - \frac{1}{2} \tr(S) + \frac{(N - K)D}{2} \ln |S| + \frac{(N - K)D}{2} \ln \lambda + \frac{N}{2} \ln |\Omega| + \text{const}.$$ 

where $S = \mathbb{E}[Z^TZ] = N\Omega + \sum_n \mu_n^T \mu_n$. Note that the equality in the second-to-last line holds at $N \to \infty$.

**Update $q$** By setting the derivatives to zero, we obtain the following update rules:

$$\mu_n^{\text{new}} = \lambda x_n W \Omega,$$  \hspace{1cm} (116)

$$\Omega^{\text{new}} = (I + \lambda W^TW + D(S + I)^{-1})^{-1}. \hspace{1cm} (117)$$

Note that we use $S$ as an auxiliary variable.

**Update $\Pi$** Update rules of $W$ and $\lambda$ are almost the same as those of the EM algorithm, which are:

$$W^{\text{new}} = X^T[\mu_1, \ldots, \mu_N]S^{-1},$$  \hspace{1cm} (118)

$$\lambda^{\text{new}} = \frac{(N - K)D}{\mathbb{E}[\|X - ZW^T\|_{\text{Fro}}^2]}.$$  \hspace{1cm} (119)

**References**