Appendix A. Sampling New Features in Gibbs iPM^{3}F with Linear Cost

For detailed derivation and analysis of Gibbs sampling in IBP models with a linear Gaussian likelihood, we recommend the extended version[6] of (Doshi-Velez et al., 2009). Below we concentrate on our discovery of a more computationally efficient way to sample new latent features, which achieves linear complexity with respect to the Poisson truncation level κ.

As indicated in Eq. (26), when sampling $Z_i^\nu = \mathbf{1}_{k_i}^\top$ for each row of Z, we need to calculate $|\Sigma_{ijk_i}|$ and $\boldsymbol{\omega}_{ijk_i}^\top \Sigma_{ijk_i}^{-1} \boldsymbol{\omega}_{ijk_i}$ for all candidate k_i values, be they either $\mathbb{Z}_{\geq 0}$ or $\{0, 1, 2, \ldots, \kappa\}$ for the truncated case. The calculation of determinant for a k-by-k matrix is of complexity $O(k^3)$ with LU decomposition, and this would bring about an overall complexity of $O(\kappa^4)$ and hence restrict our choice of the truncation level. However, we may take advantage of the special form of $\Sigma_{ijk_i}^{-1}$ to reduce the computational cost to $O(\kappa)$.

Specifically, we find that for matrices X_k of the following form:

$$X_k = (a-b)I_{k\times k} + b\mathbf{1}_{k\times k} = \begin{bmatrix} a & b & \cdots & b \\ b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & a \end{bmatrix}_{k\times k} \tag{29}$$

we have

$$\begin{aligned} \delta_k &\triangleq \det(X_k) = a\delta_{k-1} - (k-1)b\delta'_{k-1} \\ \delta'_k &\triangleq \det\left(\begin{bmatrix} b & b\mathbf{1}_{k-1}^\top \\ b\mathbf{1}_{k-1} & X_{k-1} \end{bmatrix}\right) = \delta_k + (b-a)\delta_{k-1} \end{aligned} \tag{30}$$

and by solving these recursions we obtain the closed form solution

$$\delta_k = (a-b)^{k-1}(kb + a - b) \quad (k \geq 0) \tag{31}$$

Furthermore, according to Cramer's rule, we have

$$X_k^{-1} = \frac{1}{\delta_k} \begin{bmatrix} \delta_{k-1} & -\delta'_{k-1} & \cdots & -\delta'_{k-1} \\ -\delta'_{k-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\delta'_{k-1} \\ -\delta'_{k-1} & \cdots & -\delta'_{k-1} & \delta_{k-1} \end{bmatrix}_{k\times k} \tag{32}$$

and accordingly,

$$\begin{aligned} \mathbf{1}_k^\top X_k^{-1} \mathbf{1}_k &= \frac{1}{\delta_k}(k\delta_{k-1} - k(k-1)\delta'_{k-1}) \\ &= \frac{k}{kb + a - b} \quad (k \geq 0) \end{aligned} \tag{33}$$

Then by taking $b_{ij} = \sum_{r=1}^{L-1} \frac{C^2}{4\lambda_{ijr}}$ and $a_{ij} - b_{ij} = \frac{1}{\sigma^2}$, we may apply Eq. (31) to calculate $|\Sigma_{ijk_i}^{-1}|$ and by setting $\xi_{ij} \triangleq -\left(\frac{C}{2}\sum_{r=1}^{L-1} T_{ij}^r\left(1 + \frac{\Delta_{ij}^r}{\lambda_{ijr}}\right)\right)$, we get

$$\boldsymbol{\omega}_{ijk_i}^\top \Sigma_{ijk_i}^{-1} \boldsymbol{\omega}_{ijk_i} = \xi_{ij}^2 \mathbf{1}_{k_i}^\top \Sigma_{ijk_i} \mathbf{1}_{k_i}$$

and thus may apply Eq. (33) for its calculation.

With b_{ij} and ξ_{ij} already at hand and updated from previous steps, the incremental cost of calculating Eq. (31) and (33), and hence Eq. (26), for $k_i = 0, 1, \ldots, \kappa$ is thus reduced to $O(\kappa)$.

Appendix B. Asymptotic Complexity

Below we discuss the asymptotic computational complexity of each iteration in our Gibbs sampling methods. Specifically, to draw samples from each conditional distribution, there are typically two individual costs, one for the calculation of the sufficient statistics, e.g., Δ_{ij}^r in Eq. (22), B_j and b_j in Eq. (23), etc., and the other for the actual drawing of the samples from the corresponding distribution. Normally, the first one is linear w.r.t. the number of the observed entries $|\mathcal{I}|$, while the second one is independent of $|\mathcal{I}|$ but linear to the number of samples to be drawn, i.e. the number of parameters in the model. We list the results in Table 3 and 4.

Note that we use the Cholesky decomposition $B_j^{-1} = R_j^\top R_j$ both to calculate b_j (23) where $B_j\mathbf{v}$ is calculated as $R_j \backslash (R_j^\top \backslash \mathbf{v})$[7] and to draw samples from $\mathcal{N}(b_j, B_j)$ as $b_j + R_j \backslash \mathbf{x}$ where $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, I)$.

Table 3. Asymptotic complexity for Gibbs M^{3}F.

Step	Asymptotic complexity
Sample $\boldsymbol{\lambda}$	$O(\|\mathcal{I}\|LK)$
Sample V (and U likewise)	$O(\|\mathcal{I}\|(L+K^2)) + O(MK^3)$
Calculate $\{B_j^{-1}\}_{j=1}^M$	$O(\|\mathcal{I}\|(L+K^2))$
Cholesky Decomposition	$O(MK^3)$
Calculate $\{b_j\}_{j=1}^M$	$O(\|\mathcal{I}\|(L+K)) + O(MK^2)$
Draw $\{V_j\}_{j=1}^M$ from $\mathcal{N}$	$O(MK^2)$
Sample $\boldsymbol{\theta}$	$O(\|\mathcal{I}\|(L+K)) + O(NL)$

Table 4. Asymptotic complexity for Gibbs iPM^{3}F.

Sampler for	Asymptotic complexity
$\boldsymbol{\lambda}, V, \boldsymbol{\theta}$	same as Gibbs M^{3}F (Table 3)
$\{Z_{ik}\}_{i=1,k=1}^{N,K}$	$O(\|\mathcal{I}\|(L+K)) + O(NK)$
$\{Z_i^\nu\}_{i=1}^N : \{k_i\}_{i=1}^N$	$O(\|\mathcal{I}\|\kappa) + O(N\kappa)$
$\{V^{i\nu}\}_{i=1}^N$	tight: $O(\sum_{ij\in\mathcal{I}} k_i^3) + O(M\sum_i k_i)$ loose: $O(\|\mathcal{I}\|\kappa^3) + O(MN\kappa)$

[6] http://people.csail.mit.edu/finale/papers/aistats09_tr.pdf

[7] "\" is the "*matrix left division*" operator in MATLAB.