Supplemental Material for
"A Bayesian Approach to Multimodal Visual Dictionary Learning"

Go Irie[†], Dong Liu[‡], Zhenguo Li[‡], Shih-Fu Chang[‡]
[†] NTT Corporation, Kanagawa, Japan
[‡] Columbia University, New York, USA
[†]irie.go@lab.ntt.co.jp [‡]{dongliu, zgli, sfchang}@ee.columbia.edu

Derivation of Collapsed Gibbs Sampling

We discuss the derivation of collapsed Gibbs sampling distributions for our CD-BCC, (see Section 3.1 in the original paper). The joint distribution of our CD-BCC is:

$$\begin{aligned} &p(\mathbf{X}, \mathbf{R}, \boldsymbol{\omega}, \boldsymbol{z}^x, \boldsymbol{z}^w, \boldsymbol{\mu}, \boldsymbol{\Theta}, \pi, \kappa, \lambda; \gamma, \eta, \zeta, \Sigma_0, \Sigma_x, \beta, \phi) \\ &= \left\{ p(\pi; \gamma/V) \prod_v^V p(\mu_v; \Sigma_0) \prod_i^N p(\omega_i|\pi) p(x_i|\mu_{\omega_i}; \Sigma_x) \right\} \\ &\times \left\{ p(\kappa; \zeta/K) \prod_v^V p(z_v^x|\kappa) \right\} \left\{ p(\lambda; \eta/L) \prod_j^W p(z_j^w|\lambda) \right\} \\ &\times \left\{ \prod_{k,l}^{K\times L} p(\theta_{k,l}; \beta, \phi) \prod_{v,j}^{V\times W} p(r_{v,j}|\theta_{z_v^x, z_j^w}) \right\} \end{aligned} \tag{1}$$

The goal is to derive the sampling distributions Eq. (7-9) shown in Appendix.

Marginal Distribution

We first derive a marginal distribution $p(\mathbf{X}, \mathbf{R}, \boldsymbol{\omega}, \boldsymbol{z}^x, \boldsymbol{z}^w; \gamma, \eta, \zeta, \Sigma_0, \Sigma_x, \beta, \phi)$ by integrating out π, κ, λ, μ_v, and $\theta_{k,l}$ in Eq. (1).

Integrating out π, κ, and λ: in Eq. (1), the part related to π is only $p(\pi; \gamma/V) \prod_i^N p(\omega_i|\pi)$. Because $p(\pi; \gamma/V)$ and $p(\omega_i|\pi)$ are Dirichlet and multinomial distributions respectively,

$$\begin{aligned} &\int p(\pi; \gamma/V) \prod_i^N p(\omega_i|\pi) d\pi \\ &= \frac{\Gamma(\gamma)}{\prod_v^V \Gamma(\gamma/V)} \int \prod_v^V \pi^{\gamma/V-1} \prod_v^V \pi^{m_v} d\pi \\ &= \frac{\Gamma(\gamma)}{\prod_v^V \Gamma(\gamma/V)} \int \prod_v^V \pi^{m_v+\gamma/V-1} d\pi \\ &= \frac{\Gamma(\gamma)}{\prod_v^V \Gamma(\gamma/V)} \frac{\prod_v^V \Gamma(m_v + \gamma/V)}{\Gamma(N+\gamma)} \end{aligned} \tag{2}$$

Similarly, κ and λ can also be integrated out as:

$$\begin{aligned} &\int p(\kappa; \zeta/K) \prod_v^V p(z_v^x|\kappa) d\kappa \\ &= \frac{\Gamma(\zeta)}{\prod_k^K \Gamma(\zeta/K)} \frac{\prod_k^K \Gamma(m_k^x + \zeta/K)}{\Gamma(V+\zeta)} \end{aligned} \tag{3}$$

$$\begin{aligned} &\int p(\lambda; \eta/L) \prod_j^W p(z_j^w|\lambda) d\lambda \\ &= \frac{\Gamma(\eta)}{\prod_l^L \Gamma(\eta/L)} \frac{\prod_l^L \Gamma(m_l^w + \eta/L)}{\Gamma(W+\eta)} \end{aligned} \tag{4}$$

Integrating out μ_v: the part related to μ_v is only $\prod_v^V p(\mu_v; \Sigma_0) \prod_i^N p(x_i|\mu_{\omega_i}; \Sigma_x)$. Considering that both of the distributions are Gaussian,

$$\begin{aligned} &\prod_v^V \int p(\mu_v; \Sigma_0) \prod_i^N p(x_i|\mu_{\omega_i}; \Sigma_x) d\mu_v \\ &= \prod_v^V C(\Sigma_0^{-1}) C(\Sigma_x^{-1})^{m_v} \\ &\times \int \exp\left[-\frac{1}{2} \mu_v^T \Sigma_0^{-1} \mu_v - \frac{1}{2} \sum_i^N (x_i - \mu_{\omega_i})^T \Sigma_x^{-1} (x_i - \mu_{\omega_i}) \right] d\mu_v \\ &= \prod_v^V C(\Sigma_0^{-1}) C(\Sigma_x^{-1})^{m_v} \\ &\times \exp\left[-\frac{1}{2} \text{trace}\left(\Sigma_x^{-1} S_v^x - \Sigma_v \hat{\mu}_v \hat{\mu}_v^T \right) \right] \\ &\times \int \exp\left[-\frac{1}{2} (\mu_v - \bar{\mu}_v)^T \Sigma_v^{-1} (\mu_v - \bar{\mu}_v) \right] d\mu_v \\ &= \prod_v^V \frac{C(\Sigma_0^{-1}) C(\Sigma_x^{-1})^{m_v}}{C(\Sigma_v^{-1})} \\ &\times \exp\left[-\frac{1}{2} \text{trace}\left(\Sigma_x^{-1} S_v^x - \Sigma_v \hat{\mu}_v \hat{\mu}_v^T \right) \right] \end{aligned} \tag{5}$$

where, $C(\Sigma^{-1}) \equiv (2\pi)^{-d/2}|\Sigma|^{-1/2}$,
$\bar{x}_v \equiv \frac{1}{m}\sum_{i\ s.t.\ \omega_i=v} x_i$, $S^x_v \equiv \sum_{i\ s.t.\ \omega_i=v} x_i x_i^T$,
$\hat{\mu}_v \equiv m_v \Sigma_x^{-1}\bar{x}_v$, $\bar{\mu}_v \equiv \Sigma_v\hat{\mu}_v$, and $\Sigma_v^{-1} \equiv \Sigma_0^{-1} + m_v\Sigma_x^{-1}$.
Integrating out $\theta_{k,l}$: the part related to $\theta_{k,l}$ is $\prod_{k,l}^{K\times L} p(\theta_{k,l};\beta,\phi)\prod_{v,j}^{V\times W} p(r_{v,j}|\theta_{z^x_v,z^w_j})$. Considering that $p(\theta_{k,l};\beta,\phi)$ and $p(r_{v,j}|\theta_{z^x_v,z^w_j})$ are Gamma and Poisson distributions respectively,

$$\begin{aligned}
&\prod_{k,l}^{K\times L}\int p(\theta_{k,l};\beta,\phi)\prod_{v,j}^{V\times W} p(r_{v,j}|\theta_{z^x_v,z^w_j})d\theta_{k,l}\\
&=\prod_{k,l}^{K\times L}\frac{R^{k,l}\phi^{-\beta}}{\Gamma(\beta)}\int\left(\theta_{k,l}^{\beta-1}e^{-\theta_{k,l}/\phi}\right)\left(\theta_{k,l}^{m_{k,l}}e^{-n_{k,l}\theta_{k,l}}\right)d\theta_{k,l}\\
&=\prod_{k,l}^{K\times L}\frac{R^{k,l}\phi^{-\beta}}{\Gamma(\beta)}\int\left(\theta_{k,l}^{m_{k,l}+\beta-1}e^{-\frac{n_{k,l}\phi+1}{\phi}\theta_{k,l}}\right)d\theta_{k,l}\\
&=\prod_{k,l}^{K\times L}\frac{R^{k,l}\phi^{-\beta}}{\Gamma(\beta)}\Gamma(m_{k,l}+\beta)\left(\frac{\phi}{n_{k,l}\phi+1}\right)^{(m_{k,l}+\beta)} \qquad (6)
\end{aligned}$$

Based on the results above, the marginal distribution is represented as

$$\begin{aligned}
&p(\mathbf{X},\mathbf{R},\boldsymbol{\omega},\boldsymbol{z}^x,\boldsymbol{z}^w;\gamma,\eta,\zeta,\Sigma_0,\Sigma_x,\beta,\phi)\\
&=\frac{\Gamma(\gamma)}{\prod_v^V\Gamma(\gamma/V)}\frac{\prod_v^V\Gamma(m_v+\gamma/V)}{\Gamma(N+\gamma)}\\
&\times\prod_v^V\frac{C(\Sigma_0^{-1})C(\Sigma_x^{-1})^{m_v}}{C(\Sigma_v^{-1})}\\
&\times\prod_v^V\exp\left[-\frac{1}{2}\mathrm{trace}\left(\Sigma_x^{-1}S^x_v-\Sigma_v\hat{\mu}_v\hat{\mu}_v^T\right)\right]\\
&\times\frac{\Gamma(\zeta)}{\prod_k^K\Gamma(\zeta/K)}\frac{\prod_k^K\Gamma(m^x_k+\zeta/K)}{\Gamma(V+\zeta)}\\
&\times\frac{\Gamma(\eta)}{\prod_l^L\Gamma(\eta/L)}\frac{\prod_l^L\Gamma(m^w_l+\eta/L)}{\Gamma(W+\eta)}\\
&\times\prod_{k,l}^{K\times L}\frac{R^{k,l}\phi^{-\beta}}{\Gamma(\beta)}\Gamma(m_{k,l}+\beta)\left(\frac{\phi}{n_{k,l}\phi+1}\right)^{(m_{k,l}+\beta)}
\end{aligned} \qquad (7)$$

Collapsed Gibbs Sampling

First, ignoring parts regarded to be constant in Eq. (7)

$$\begin{aligned}
&p(\mathbf{X},\mathbf{R},\boldsymbol{\omega},\boldsymbol{z}^x,\boldsymbol{z}^w;\gamma,\eta,\zeta,\Sigma_0,\Sigma_x,\beta,\phi)\\
&\propto\prod_v^V\Gamma(m_v+\gamma/V)\\
&\times\prod_v^V\frac{|\Sigma_v|^{1/2}}{|\Sigma_x|^{m_v/2}}\exp\left[-\frac{1}{2}\mathrm{trace}\left(\Sigma_x^{-1}S^x_v-\Sigma_v\hat{\mu}_v\hat{\mu}_v^T\right)\right]\\
&\times\prod_l^L\Gamma(m^w_l+\eta/L)\\
&\times\prod_k^K\Gamma(m^x_k+\zeta/K)\\
&\times\prod_{k,l}^{K\times L}R^{k,l}\Gamma(m_{k,l}+\beta)\left(\frac{\phi}{n_{k,l}\phi+1}\right)^{(m_{k,l}+\beta)} \qquad (8)
\end{aligned}$$

Based on this equation, we derive collapsed Gibbs sampler for ω_i, z^x_v, and z^w_j. For simplicity, we denote $p(\mathbf{X},\mathbf{R},\boldsymbol{\omega},\boldsymbol{z}^x,\boldsymbol{z}^w) \equiv p(\mathbf{X},\mathbf{R},\boldsymbol{\omega},\boldsymbol{z}^x,\boldsymbol{z}^w;\gamma,\eta,\zeta,\Sigma_0,\Sigma_x,\beta,\phi)$ hereafter.
Sampling for ω_i: Focusing on only i-th image descriptor corresponds to only t-th Gaussian component,

$$\begin{aligned}
&p(\omega_i=t|\mathbf{X},\mathbf{R},\boldsymbol{\omega}_{-i},\boldsymbol{z}^x,\boldsymbol{z}^w)\\
&\propto p(\omega_i=t,\mathbf{X},\mathbf{R},\boldsymbol{\omega}_{-i},\boldsymbol{z}^x,\boldsymbol{z}^w)\\
&\propto\Gamma(m_t+\gamma/V)\\
&\times\frac{|\Sigma_t|^{1/2}}{|\Sigma_x|^{m_t/2}}\exp\left[-\frac{1}{2}\mathrm{trace}\left(\Sigma_x^{-1}S^x_t-\Sigma_t\hat{\mu}_t\hat{\mu}_t^T\right)\right]\\
&\times\prod_{k,l}^{K\times L}R^{k,l}\Gamma(m_{k,l}+\beta)\left(\frac{\phi}{n_{k,l}\phi+1}\right)^{(m_{k,l}+\beta)}\\
&\propto(m_{t,-i}+\gamma/V)\\
&\times\frac{|\Sigma_t|^{1/2}}{|\Sigma_x|^{m_t/2}}\exp\left[-\frac{1}{2}\mathrm{trace}\left(\Sigma_x^{-1}S^x_t-\Sigma_t\hat{\mu}_t\hat{\mu}_t^T\right)\right]\\
&\times\prod_{k,l}^{K\times L}R^{k,l}\Gamma(m_{k,l}+\beta)\left(\frac{\phi}{n_{k,l}\phi+1}\right)^{(m_{k,l}+\beta)} \qquad (9)
\end{aligned}$$

Sampling for z^x_v and z^w_j: Focusing on only v-th Gaussian component assigned to a-th row cluster,

$$
\begin{aligned}
& p(z_v^x = a | \mathbf{X}, \mathbf{R}, \boldsymbol{\omega}, \boldsymbol{z}_{-v}^x, \boldsymbol{z}^w) \\
& \propto p(z_v^x = a, \mathbf{X}, \mathbf{R}, \boldsymbol{\omega}, \boldsymbol{z}_{-v}^x, \boldsymbol{z}^w) \\
& \propto \Gamma(m_a^x + \zeta/K) \\
& \times \left\{ \prod_l^L R^{a,l} \Gamma(m_{a,l} + \beta) \left(\frac{\phi}{n_{a,l}\phi + 1} \right)^{(m_{a,l}+\beta)} \right\} \\
& \propto (m_{a,-v}^x + \zeta/K) \\
& \times \left\{ \prod_l^L R^{a,l} \Gamma(m_{a,l} + \beta) \left(\frac{\phi}{n_{a,l}\phi + 1} \right)^{(m_{a,l}+\beta)} \right\} \quad (10)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& p(z_j^w = b | \mathbf{X}, \mathbf{R}, \boldsymbol{\omega}, \boldsymbol{z}^x, \boldsymbol{z}_{-j}^w) \\
& \propto (m_{b,-j}^w + \eta/L) \\
& \times \left\{ \prod_k^K R^{k,b} \Gamma(m_{k,b} + \beta) \left(\frac{\phi}{n_{k,b}\phi + 1} \right)^{(m_{k,b}+\beta)} \right\} \quad (11)
\end{aligned}
$$

Finally, Eq. (9), (10), and (11) are the sampling distributions shown in the original paper.

Derivation of Non-parametric Extension

We next discuss the derivation of the infinite version of CD-BCC in Section 3.3 in the original paper. In our CD-BCC, V and L are dimensions of two Dirichlet-multinomial conjugate pairs inside first and third curly brackets in Eq. (1) respectively. Let us consider a stochastic process $p(\omega_i | \omega_{1:i-1}; \gamma)$. Using multinomial $p(\omega_i = t | \pi)$ and a posterior $p(\pi | \omega_{1:i-1})$ (Dirichlet),

$$
\begin{aligned}
p(\omega_i | \omega_{1:i-1}; \gamma) &= \int p(\omega_i = t | \pi) p(\pi | \omega_{1:i-1}) d\pi \\
&\propto \int \pi_t \left(\prod_t^V \pi_t^{m_t + \gamma/V - 1} \right) d\pi \\
&= \frac{m_t + \gamma/V}{\sum_v^V (m_v + \gamma/V)} \\
&= \frac{m_t + \gamma/V}{i - 1 + \gamma} \qquad (12)
\end{aligned}
$$

By taking its infinite limit $V \to \infty$,

$$
p(\omega_i = t | \omega_{1:i-1}; \gamma) \propto \begin{cases} \frac{m_t}{i-1+\gamma} & (m_t > 0) \\ \frac{\gamma}{i-1+\gamma} & (otherwise) \end{cases} \qquad (13)
$$

This is the same with Eq. (5) in the original paper. Eq. (6) can also be derived by the completely the same procedure.