Derivation of SLDA Variational Inference

November 5, 2014

Derivation of Variational Spatial LDA

Terminology

- Bold $\boldsymbol{var}$ for observed variables

Variational Distribution

$$q(\phi, \pi, d, z|\lambda, \gamma, \rho, \xi) = \prod_{k=1}^{K} q(\phi_k|\lambda) \prod_{m=1}^{M} q(\pi_m|\gamma) \prod_{n=1}^{N} q(d_n|\rho) q(z_n|\xi) \tag{1}$$

where

$\phi_k \in R^V$ distribution over word of topic k

$\pi_m \in R^K$ distribution over topics of document m

$d_n \in R^M$ indicator vector where if $d_{nm} = 1$ means word n belongs to document m

$z_n \in R^K$ indicator vector where if $z_{nk} = 1$ means word n is assigned to topic k

Lower bound of log likelihood:

$$\begin{aligned}
p(\boldsymbol{w}, \boldsymbol{c}|\alpha, \beta, \eta, \sigma) &= \log \int_\phi \int_\pi \sum_d \sum_z p(\phi, \pi, d, z, \boldsymbol{w}, \boldsymbol{c}|\alpha, \beta, \eta, \sigma) &(2)\\
&= \log \int_\phi \int_\pi \sum_d \sum_z q(\phi, \pi, d, z|\lambda, \gamma, \rho, \xi) \frac{p(\phi, \pi, d, z, \boldsymbol{c}, \boldsymbol{w}|\alpha, \beta, \eta, \sigma)}{q(\phi, \pi, d, z|\lambda, \gamma, \rho, \xi)} \,\mathrm{d}\phi \,\mathrm{d}\pi \,\mathrm{d}d \,\mathrm{d}z &(3)\\
&= \log E_q \frac{p(\phi, \pi, d, z, \boldsymbol{c}, \boldsymbol{w}|\alpha, \beta, \eta, \sigma)}{q(\phi, \pi, d, z|\lambda, \gamma, \rho, \xi)} &(4)\\
&\geq E_q \log p(\phi, \pi, d, z, \boldsymbol{c}, \boldsymbol{w}|\alpha, \beta, \eta, \sigma) - E_q \log q(\phi, \pi, d, z|\lambda, \gamma, \rho, \xi) = L &(5)
\end{aligned}$$

where $E_q \equiv E_{q(\phi,\pi,d,z|\lambda,\gamma,\rho,\xi)}$

Instead of maximizing p, we estimate it by maximizing the lower bound L

First Term in L

$$\begin{aligned}
E_q[\log p(\phi, \pi, d, z, \boldsymbol{c}, \boldsymbol{w}|\alpha, \beta, \eta, \sigma)] &= E_q[\log p(\phi|\beta)\, p(\pi|\alpha)\, p(d|\eta)\, p(\boldsymbol{c}|c_d^{\mathrm{d}}, \sigma)\, p(z|\pi_d)\, p(\boldsymbol{w}|\phi_d) &(6)\\
&= E_q[\log p(\phi|\beta)] + E_q[\log p(\pi|\alpha)] + E_q[\log p(d|\eta)] &(7)\\
&+ E_q[\log p(\boldsymbol{c}|c_d^{\mathrm{d}}, \sigma)] + E_q[\log p(z|\pi_d)] + E_q[\log p(\boldsymbol{w}|\phi_z)] &(8)\\
& &(9)
\end{aligned}$$

For each sub-terms:

$$\begin{aligned}
E_q[\log p(\phi|\beta)] &= E_q[\log \prod_{k=1}^{K} p(\phi_k|\beta)] &(10)\\
&= \sum_{k=1}^{K} E_q \left[\log \Gamma \left(\sum_{v=1}^{V} \beta_v \right) - \sum_{v=1}^{V} \log \Gamma(\beta_v) + \sum_{v=1}^{V} (\beta_v - 1) \log(\phi_{kv}) \right] &(11)\\
&= \sum_{k=1}^{K} \left[\log \Gamma(\sum_{v=1}^{V} \beta_v) - \sum_{v=1}^{V} \log \Gamma(\beta_v) + \sum_{v=1}^{V} (\beta_v - 1)(\Psi(\lambda_{kv}) - \Psi(\sum_{v'=1}^{V} \lambda_{kv'})) \right] &(12)
\end{aligned}$$

$$\begin{aligned}
E_q[\log p(\pi|\alpha)] &= E_q[\log \prod_{m=1}^{M} p(\pi_m|\alpha)] && (13)\\
&= \sum_{m=1}^{M} E_q \log(p(\pi_m|\alpha)) && (14)\\
&= \sum_{m=1}^{M} E_q \left[\log(\Gamma\left(\sum_{k=1}^{K}\alpha_k\right)) - \sum_{k=1}^{K}\log(\Gamma(\alpha_k)) + \sum_{k=1}^{K}(\alpha_k - 1)\log(\pi_{mk})\right] && (15)\\
&= \sum_{m=1}^{M} \left[\log(\Gamma(\left(\sum_{k=1}^{K}\alpha_k\right)) - \sum_{k=1}^{K}\log(\Gamma(\alpha_k)) + \sum_{k=1}^{K}(\alpha_k - 1)(\Psi(\gamma_{mk}) - \Psi(\sum_{k'=1}^{K}\gamma_{mk'}))\right] && (16)\\
& && (17)
\end{aligned}$$

The following derivations skip some steps for the purpose of shorter writings.

$$\begin{aligned}
E_q[\log p(d|\eta)] &= \sum_{n=1}^{N} E_q \log(p(d_n|\eta)) && (18)\\
&= \sum_{n=1}^{N} E_q\left[\sum_{m=1}^{M} d_{mn}\log(\eta)\right] && (19)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M} \rho_{mn}\log(\eta) && (20)\\
& && (21)
\end{aligned}$$

$c^{\mathrm{d}}_{d_n}$*: location of document d (in its image).*

$$\begin{aligned}
E_q[\log p(\boldsymbol{c}|c^{\mathrm{d}}_d, \sigma)] &= \sum_{n=1}^{N} E_q \log p(c_n|c^{\mathrm{d}}_{d_n}) && (22)\\
&= \sum_{n=1}^{N} E_q\left[\sum_{m=1}^{M} d_{mn}\log p(c_n|c^{\mathrm{d}}_{d_{mn}})\right] && (23)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M} E_q[d_{mn}\log p(c_n|c^{\mathrm{d}}_{d_{mn}})] && (24)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M} \rho_{mn}(\log\delta_{g^{\mathrm{d}}_{d_{mn}}} - \frac{(x^{\mathrm{d}}_{d_{mn}} - x_n)^2 + (y^{\mathrm{d}}_{d_{mn}} - y_n)^2}{\sigma^2}) && (25)\\
& && (26)
\end{aligned}$$

$$\begin{aligned}
E_q[\log p(z|\pi_d)] &= \sum_{n=1}^{N} E_q\left[\sum_{m=1}^{M}\sum_{k=1}^{K} d_{mn} z_{kn}\log(\pi_{mk})\right] && (27)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K} E_q[d_{mn} z_{kn}\log(\pi_{mk})] && (28)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K} \rho_{mn}\xi_{kn}(\Psi(\gamma_{mk}) - \Psi(\sum_{m'=1}^{M}\gamma_{m'k})) && (29)
\end{aligned}$$

$$
\begin{aligned}
E_q[\log p(\boldsymbol{w}|\phi_z)] &= \sum_n^N E_q\left[\sum_m^M \sum_k^K \sum_v^V d_{mn} z_{kn} w_n^v log(\phi_{kv})\right] &(30)\\
&= \sum_n^N \sum_m^M \sum_k^K \sum_v^V E_q[d_{mn} z_{kn} w_n^v log(\phi_{kv})] &(31)\\
&= \sum_n^N \sum_m^M \sum_k^K \sum_v^V \rho_{mn} \xi_{kn} w_n^v (\Psi(\lambda_{k\boldsymbol{v}}) - \Psi(\sum_{k'}^K \lambda_{k'\boldsymbol{v}})) &(32)\\
& &(33)
\end{aligned}
$$

The second term in L

This uses the log of the joint distribution under variational parameters in (1)

$$
\begin{aligned}
E_q \log q(\phi|\lambda) &= \sum_k^K E_q \log q(\phi_k|\lambda) &(34)\\
&= \sum_k^K \left[\log \Gamma(\sum_v^V \lambda_{kv}) - \sum_v^V \log(\Gamma(\lambda_{kv})) + \sum_v^V (\lambda_{kv} - 1) E_q[\log(\phi_{kv})]\right] &(35)\\
&= \sum_k^K \left[\log \Gamma\left(\sum_v^V \lambda_{kv}\right) - \sum_v^V \log(\Gamma(\lambda_{kv})) + \sum_v^V (\lambda_{kv} - 1)(\Psi(\lambda_{kv}) - \Psi(\sum_{v'}^V \lambda_{kv'})\right] &(36)\\
& &(37)
\end{aligned}
$$

$$
\begin{aligned}
E_q \log q(\pi|\gamma) &= \sum_m^M E_q \log q(\pi_m|\gamma) &(38)\\
&= \sum_m^M \left[\log \Gamma\left(\sum_k^K \gamma_{km}\right) - \sum_k^K \log \Gamma(\gamma_{km}) + \sum_k^K (\gamma_{km} - 1) E_q[\log(\pi_{km})]\right] &(39)\\
&= \sum_m^M \left[\log \Gamma\left(\sum_k^K \gamma_{km}\right) - \sum_k^K \log \Gamma(\gamma_{km}) + \sum_k^K (\gamma_{km} - 1)(\Psi(\gamma_{km}) - \Psi(\sum_{k'}^K \gamma_{k'm}))\right] &(40)\\
& &(41)
\end{aligned}
$$

$$
\begin{aligned}
E_q \log q(d|\rho) &= \sum_n^N E_q[\log p(d_n|\rho)] &(42)\\
&= \sum_n^N \sum_m^M E_q[d_{mn} \log(\rho_{mn})] &(43)\\
&= \sum_n^N \sum_m^M \rho_{mn} \log(\rho_{mn}) &(44)\\
& &(45)
\end{aligned}
$$

$$
\begin{aligned}
E_q \log q(z|\xi) &= \sum_n^N E_q \log q(z_n|\xi) &(46)\\
&= \sum_n^N \sum_k^K E_q[z_{kn} \log(\xi_{kn})] &(47)\\
&= \sum_n^N \sum_k^K \xi_{kn} \log(\xi_{kn}) &(48)\\
& &(49)
\end{aligned}
$$

Putting it all together

$$
\begin{aligned}
L &= \sum_{k=1}^{K}\left[\log\Gamma(\sum_{v=1}^{V}\beta_v) - \sum_{v=1}^{V}\log\Gamma(\beta_v) + \sum_{v=1}^{V}(\beta_v-1)(\Psi(\lambda_{kv}) - \Psi(\sum_{v'=1}^{V}\lambda_{kv'}))\right] && (50)\\
&+ \sum_{m=1}^{M}\left[\log(\Gamma(\sum_{k=1}^{K}\alpha_k)) - \sum_{k=1}^{K}\log(\Gamma(\alpha_k)) + \sum_{k=1}^{K}(\alpha_k-1)(\Psi(\gamma_{mk}) - \Psi(\sum_{k'=1}^{K}\gamma_{mk'}))\right] && (51)\\
&+ \sum_{n=1}^{N}\sum_{m=1}^{M}\rho_{mn}\log(\eta) && (52)\\
&+ \sum_{n=1}^{N}\sum_{m=1}^{M}\rho_{mn}(\log\delta_{g^{\mathrm{d}}_{d_{mn}}} - \frac{(x^{\mathrm{d}}_{d_{mn}}-x_n)^2 + (y^{\mathrm{d}}_{d_{mn}}-y_n)^2}{\sigma^2}) && (53)\\
&+ \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K}\rho_{mn}\xi_{kn}(\Psi(\gamma_{mk}) - \Psi(\sum_{m'=1}^{M}\gamma_{m'k})) && (54)\\
&+ \sum_{n}^{N}\sum_{m}^{M}\sum_{k}^{K}\sum_{v}^{V}\rho_{mn}\xi_{kn}w_n^v(\Psi(\lambda_{k\boldsymbol{v}}) - \Psi(\sum_{k'}^{K}\lambda_{k'\boldsymbol{v}})) && (55)\\
&- \sum_{k}^{K}\left[\log\Gamma\left(\sum_{v}^{V}\lambda_{kv}\right) - \sum_{v}^{V}\log(\Gamma(\lambda_{kv})) + \sum_{v}^{V}(\lambda_{kv}-1)(\Psi(\lambda_{kv}) - \Psi(\sum_{v'}^{V}\lambda_{kv'})\right] && (56)\\
&- \sum_{m}^{M}\left[\log\Gamma\left(\sum_{k}^{K}\gamma_{km}\right) - \sum_{k}^{K}\log\Gamma(\gamma_{km}) + \sum_{k}^{K}(\gamma_{km}-1)(\Psi(\gamma_{km}) - \Psi(\sum_{k'}^{K}\gamma_{k'm}))\right] && (57)\\
&- \sum_{n}^{N}\sum_{m}^{M}\rho_{mn}\log(\rho_{mn}) && (58)\\
&- \sum_{n}^{N}\sum_{k}^{K}\xi_{kn}\log(\xi_{kn}) && (59)\\
& && (60)
\end{aligned}
$$

Maximize with respect to each term:

$$
\begin{aligned}
L_{\rho_{mn}} &= \sum_{n=1}^{N}\sum_{m=1}^{M}\rho_{mn}\log(\eta) && (61)\\
&+ \sum_{n=1}^{N}\sum_{m=1}^{M}\rho_{mn}(\log\delta_{g^{\mathrm{d}}_{d_{mn}}} - \frac{(x^{\mathrm{d}}_{d_{mn}}-x_n)^2 + (y^{\mathrm{d}}_{d_{mn}}-y_n)^2}{\sigma^2}) && (62)\\
&+ \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K}\rho_{mn}\xi_{kn}(\Psi(\gamma_{mk}) - \Psi(\sum_{m'=1}^{M}\gamma_{m'k}) && (63)\\
&+ \sum_{n}^{N}\sum_{m}^{M}\sum_{k}^{K}\sum_{v}^{V}\rho_{mn}\xi_{kn}w_n^v(\Psi(\lambda_{k\boldsymbol{v}}) - \Psi(\sum_{k'}^{K}\lambda_{k'\boldsymbol{v}}) && (64)\\
&- \sum_{n}^{N}\sum_{m}^{M}[\rho_{mn}\log(\rho_{mn})] + \lambda\left(\sum_{m}^{M}\rho_{mn} - 1\right) && (65)
\end{aligned}
$$

$$\frac{\partial L}{\partial \rho_{mn}} = \log(\eta) \tag{66}$$

$$+ \log \delta_{g^{\mathrm{d}}_{d_{mn}}} - \frac{(x^{\mathrm{d}}_{d_{mn}} - x_n)^2 + (y^{\mathrm{d}}_{d_{mn}} - y_n)^2}{\sigma^2} \tag{67}$$

$$+ \sum_{k=1}^{K} \xi_{kn} \left[\Psi(\gamma_{mn}) - \Psi(\sum_{m'=1}^{M} \gamma_{m'n}) \right] \tag{68}$$

$$+ \sum_{k}^{K} \sum_{v}^{V} \xi_{kn} w_n^v (\Psi(\lambda_{kv}) - \Psi(\sum_{k'}^{K} \lambda_{k'\boldsymbol{v}}) \tag{69}$$

$$- \log(\rho_{mn}) - 1 + \lambda \stackrel{set}{=} 0 \tag{70}$$

$$\tag{71}$$

Solving yields:

$$\boxed{\rho_{mn} \propto \eta \delta_{g^{\mathrm{d}}_{d_{mn}}} \frac{\exp\left(\sum_{k=1}^{K} \xi_{kn} \left[\Psi(\gamma_{mn}) - \Psi\left(\sum_{m'=1}^{M} \gamma_{m'n} \right) \right] + \sum_{k}^{K} \sum_{v}^{V} \left[\xi_{kn} w_n^v \left(\Psi(\lambda_{kv}) - \Psi\left(\sum_{k'}^{K} \lambda_{k'\boldsymbol{v}} \right) \right) \right] \right)}{\exp\left(\frac{(x^{\mathrm{d}}_{d_{mn}} - x_n)^2 + (y^{\mathrm{d}}_{d_{mn}} - y_n)^2}{\sigma^2} \right)}}$$

$$L_{\xi_{kn}} = \sum_{n=1}^{N} \sum_{m=1}^{M} \sum_{k=1}^{K} \rho_{mn} \xi_{kn} (\Psi(\gamma_{mk}) - \Psi(\sum_{m'=1}^{M} \gamma_{m'k})) \tag{72}$$

$$+ \sum_{n}^{N} \sum_{m}^{M} \sum_{k}^{K} \sum_{v}^{V} \rho_{mn} \xi_{kn} w_n^v (\Psi(\lambda_{k\boldsymbol{v}}) - \Psi(\sum_{k'}^{K} \lambda_{k'\boldsymbol{v}})) \tag{73}$$

$$- \sum_{n}^{N} \sum_{k}^{K} \xi_{kn} \log(\xi_{kn}) + \lambda(\sum_{k}^{K} \xi_{kn} - 1) \tag{74}$$

$$\tag{75}$$

$$\frac{\partial L}{\partial \xi_{kn}} = \sum_{m}^{M} \rho_{mn} \left(\Psi(\gamma_{mk}) - \Psi\left(\sum_{m'=1}^{M} \gamma_{m'k} \right) \right) \tag{76}$$

$$+ \sum_{m}^{M} \sum_{v}^{V} \rho_{mn} w_n^v \left(\Psi(\lambda_{k\boldsymbol{v}}) - \Psi\left(\sum_{k'}^{K} \lambda_{k'\boldsymbol{v}} \right) \right) \tag{77}$$

$$- \log(\xi_{kn}) - 1 + \lambda \stackrel{set}{=} 0 \tag{78}$$

Similarly, solving yields:

$$\boxed{\xi_{kn} \propto \exp\left(\sum_{m}^{M} \rho_{mn} \left(\Psi(\gamma_{mk}) - \Psi\left(\sum_{m'=1}^{M} \gamma_{m'k} \right) \right) + \sum_{m}^{M} \sum_{v}^{V} \rho_{mn} w_n^v \left(\Psi(\lambda_{k\boldsymbol{v}}) - \Psi\left(\sum_{k'}^{K} \lambda_{k'\boldsymbol{v}} \right) \right) \right)}$$

$$
\begin{aligned}
L_{\gamma_{mk}} &= \sum_{m=1}^{M}\sum_{k=1}^{K}(\alpha_k - 1)\left(\Psi(\gamma_{mk}) - \Psi\left(\sum_{k'=1}^{K}\gamma_{mk'}\right)\right) && (79)\\
&+ \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K}\rho_{mn}\xi_{kn}\left(\Psi(\gamma_{mk}) - \Psi\left(\sum_{m'=1}^{M}\gamma_{m'k}\right)\right) && (80)\\
&- \sum_{m}^{M}\left[\log\Gamma\left(\sum_{k}^{K}\gamma_{km}\right) - \sum_{k}^{K}\log\Gamma(\gamma_{km}) + \sum_{k}^{K}(\gamma_{km} - 1)\left(\Psi(\gamma_{km}) - \Psi\left(\sum_{k'}^{K}\gamma_{k'm}\right)\right)\right] && (81)\\
&= \sum_{m=1}^{M}\sum_{k=1}^{K}\left(\Psi(\gamma_{mk}) - \Psi\left(\sum_{k'=1}^{K}\gamma_{mk'}\right)\right)\left(\alpha_k + \sum_{n=1}^{N}\rho_{mn}\xi_{kn} - \gamma_{km}\right) && (82)\\
&- \sum_{m}^{M}\log\Gamma\left(\sum_{k}^{K}\gamma_{km}\right) + \sum_{m}^{M}\sum_{k}^{K}\log\Gamma(\gamma_{km}) && (83)\\
& && (84)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial\gamma_{mk}} &= \Psi'(\gamma_{mk})(\alpha_k + \sum_{n=1}^{N}\rho_{mn}\xi_{kn} - \gamma_{mk}) && (85)\\
&- \Psi'\left(\sum_{k'=1}^{K}\gamma_{mk'}\right)\sum_{k'}^{K}(\alpha_{k'} + \sum_{n=1}^{N}\rho_{mn}\xi_{k'n} - \gamma_{mk'}) \overset{set}{=} 0 && (86)
\end{aligned}
$$

Or:

$$
(\alpha_{mk} + \sum_{n=1}^{N}\rho_{mn}\xi_{kn} - \gamma_{mk}) = \frac{\Psi'\left(\sum_{k'=1}^{K}\gamma_{mk'}\right)}{\Psi'(\gamma_{mk})}\sum_{k'}^{K}\left(\alpha_{k'} + \sum_{n=1}^{N}\rho_{mn}\xi_{k'n} - \gamma_{mk'}\right) \tag{87}
$$

Sum over k, yields:

$$
\sum_{k}^{K}\left(\alpha_k + \sum_{n=1}^{N}\rho_{mn}\xi_{kn} - \gamma_{mk}\right) = \left(\sum_{k}^{K}\frac{\Psi'\left(\sum_{k'=1}^{K}\gamma_{mk'}\right)}{\Psi'(\gamma_{mk})}\right)\sum_{k'}^{K}\left(\alpha_{k'} + \sum_{n=1}^{N}\rho_{mn}\xi_{k'n} - \gamma_{mk'}\right) \tag{88}
$$

Or

$$
\sum_{k}^{K}\left(\alpha_k + \sum_{n=1}^{N}\rho_{mn}\xi_{kn} - \gamma_{mk}\right) = 0 \tag{89}
$$

Replacing it in ... yields:

$$
\boxed{\gamma_{mk} = \alpha_k + \sum_{n=1}^{N}\rho_{mn}\xi_{kn}}
$$

$$
\begin{aligned}
L_{\lambda_{kv}} &= \sum_{k=1}^{K}\left[\sum_{v=1}^{V}(\beta_v - 1)(\Psi(\lambda_{kv}) - \Psi(\sum_{v'=1}^{V}\lambda_{kv'}))\right] && (90)\\
&+ \sum_{n}^{N}\sum_{m}^{M}\sum_{k}^{K}\sum_{v}^{V}\rho_{mn}\xi_{kn}w_n^v(\Psi(\lambda_{k\boldsymbol{v}}) - \Psi(\sum_{k'}^{K}\lambda_{k'\boldsymbol{v}})) && (91)\\
&- \sum_{k}^{K}\left[\log\Gamma\left(\sum_{v}^{V}\lambda_{kv}\right) - \sum_{v}^{V}\log(\Gamma(\lambda_{kv})) + \sum_{v}^{V}(\lambda_{kv} - 1)(\Psi(\lambda_{kv}) - \Psi(\sum_{v'}^{V}\lambda_{kv'})\right] && (92)\\
&= \sum_{k=1}^{K}\sum_{v=1}^{V}\left(\Psi(\lambda_{kv}) - \Psi(\sum_{v'=1}^{V}\lambda_{kv'})\right)\left(\beta_v + \sum_{n}^{N}\sum_{m}^{M}\rho_{mn}\xi_{kn}w_n^v - \lambda_{kv}\right) && (93)\\
&- \sum_{k}^{K}\left\{\log\Gamma(\sum_{v}^{V}\lambda_{kv}) - \sum_{v}^{V}\log(\Gamma(\lambda_{kv}))\right\} && (94)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \lambda_{kv}} &= \Psi'(\lambda_{kv})(\beta_v + \sum_{n}^{N}\sum_{m}^{M}\rho_{mn}\xi_{kn}w_n^v - \lambda_{kv}) && (95)\\
&- \Psi'(\sum_{v'=1}^{V}\lambda_{kv'})\sum_{v'}^{V}(\beta_{v'} + \sum_{n}^{N}\sum_{m}^{M}\rho_{mn}\xi_{kn}w_n^{v'} - \lambda_{kv'}) \overset{set}{=} 0 && (96)
\end{aligned}
$$

Similar to above, solving yields:

$$
\boxed{\lambda_{kv} = \beta_v + \sum_{n}^{N}\sum_{m}^{M}\rho_{mn}\xi_{kn}w_n^v}
$$

$$
\begin{aligned}
L_{\beta_v} &= \sum_{k=1}^{K}\left[\log\Gamma(\sum_{v=1}^{V}\beta_v) - \sum_{v=1}^{V}\log\Gamma(\beta_v) + \sum_{v=1}^{V}(\beta_v - 1)(\Psi(\lambda_{kv}) - \Psi(\sum_{v'=1}^{V}\lambda_{kv'}))\right] + \sum_{k}^{K}\lambda_k\left(\sum_{v}^{V}\beta_v - 1\right) && (97)\\
& && (98)
\end{aligned}
$$

$$
\begin{aligned}
L_{\alpha} &= \sum_{m=1}^{M}\left[\log(\Gamma(\sum_{k=1}^{K}\alpha_k)) - \sum_{k=1}^{K}\log(\Gamma(\alpha_k)) + \sum_{k=1}^{K}(\alpha_k - 1)(\Psi(\gamma_{mk}) - \Psi(\sum_{k'=1}^{K}\gamma_{mk'}))\right] && (99)\\
& && (100)
\end{aligned}
$$

$$
\frac{\partial L}{\partial \alpha_i} = M\left[\Psi\left(\sum_{k}^{K}\alpha_k\right) + \Psi\left(\alpha_i\right)\right] + \sum_{m}^{M}\left[\Psi(\gamma_{mi}) - \Psi(\sum_{k'=1}^{K}\gamma_{mk'})\right] \quad (101)
$$

Hessian:

$$
\frac{\partial^2 L}{\partial \alpha_i \partial \alpha_j} = M\left[\Psi'\left(\sum_{k}^{K}\alpha_k\right) - I\left[i = j\right]\Psi'(\alpha_i)\right] \quad (102)
$$