Supplement

This supplement material provides the details of the variational inference and some implementation details. The derivation of the evidence lower bound (ELBO) is presented in Section **??**.The update equations are computed in Section **??**. Some implementation details are presented in Section **??**.

1 ELBO

This this section, the computation of each term in the ELBO will be derived step by step. All the terms, that in NUF-SLDA, are shared with NCF-SLDA. So we will write the terms in the ELBO for NUF-SLDA first and list the additional terms that are special for NCF-SLDA in the end of this section.

Following the standard mean field variational inference [**?**], we assume the fully factorized variational distribution, which is

$$q(\beta,z,\theta,B)=q(\beta)q(z)q(\theta)q(B) \tag{1}$$

This variational distribution is the same for both NUF-SLDA and NCF-SLDA, since these two model only differ in the response variables. In Equation (**??**), we have:

1. The topic mixture (topic-words distribution):
   $q(\beta)=\prod_{k=1}^{K}q(\beta_k|\lambda_k)$, where λ_k is the Dirichlet variational parameter, $\lambda_k\in\mathbb{R}^V$
2. The per word topic assignment:
   $q(z)=\prod_{j=1}^{M}\prod_{n=1}^{N}q(z_{jn}|\phi_{jn})$, where ϕ_{jn} is the Multinomial variational parameter, $\phi_{jn}\in\mathbb{R}^K$
3. The per document topic distribution :
   $q(\theta)=\prod_{j=1}^{M}q(\theta_j|\gamma_j)$, where γ_j is the Dirichlet variational parameter, $\gamma_j\in\mathbb{R}^K$.
4. The signal-noise indicator:
   $q(B)=\prod_{k=1}^{K}q(B_k|f_k)$, where f_k is the Bernoulli variational parameter. Hence, $q(B_k=1|f_k)=f_k$ and $q(B_k=0|f_k)=1-f_k$

Following the standard procedure, we apply Jensen's inequality on the log likelihood of the model.

For the first model, NUC-SLDA:

$$\begin{aligned}
&\log p(w,y|\eta,\alpha,e,\mu)\\
&=\log\int\int\sum_{z}\sum_{B}\frac{p(w,y,\beta,z,\theta,B|\eta,\alpha,e,\mu)q(\beta,z,\theta,B)}{q(\beta,z,\theta,B)}d\beta d\theta\\
&\geq\mathop{\mathbb{E}}_{q}[\log p(w,y,\beta,z,\theta,B|\eta,\alpha,e,\mu)]-\mathop{\mathbb{E}}_{q}[\log q(\beta,z,\theta,B)]\\
&=\mathscr{L}_{NUF-SLDA}.
\end{aligned} \tag{2}$$

We get the evidence lower bound (ELBO) for NUF-SLDA

$$\begin{aligned}
&\mathscr{L}_{NUF-SLDA}\\
&=\mathop{\mathbb{E}}_{q}[\log p(w,y,\beta,z,\theta,B|\eta,\alpha,e,\mu)]-\mathop{\mathbb{E}}_{q}[\log q(\beta,z,\theta,B)]\\
&=\mathop{\mathbb{E}}_{q}[\log p(w|z,\beta)]+\mathop{\mathbb{E}}_{q}[\log p(z|\theta)]+\mathop{\mathbb{E}}_{q}[\log p(\theta|\alpha)]+\mathop{\mathbb{E}}_{q}[\log p(\beta|\eta)]\\
&+\mathop{\mathbb{E}}_{q}[\log p(B|e)]+\mathop{\mathbb{E}}_{q}[\log p(y|z,B,\mu)]-\mathop{\mathbb{E}}_{q}[\log q(\beta)]-\mathop{\mathbb{E}}_{q}[\log q(z)]-\mathop{\mathbb{E}}_{q}[\log q(\theta)]-\mathop{\mathbb{E}}_{q}[\log q(B)]
\end{aligned} \tag{3}$$

Similarly, with one more response term, we can get the ELBO for the NCF-SLDA

$$
\begin{aligned}
&\mathscr{L}_{NCF-SLDA} \\
&= \mathop{\mathbb{E}}_q[\log p(w,o,y,\beta,z,\theta,B|\eta,\alpha,e,\mu,\nu)] - \mathop{\mathbb{E}}_q[\log q(\beta,z,\theta,B)] \\
&= \mathop{\mathbb{E}}_q[\log p(w|z,\beta)] + \mathop{\mathbb{E}}_q[\log p(z|\theta)] + \mathop{\mathbb{E}}_q[\log p(\theta|\alpha)] + \mathop{\mathbb{E}}_q[\log p(\beta|\eta)] + \mathop{\mathbb{E}}_q[\log p(B|e)] + \mathop{\mathbb{E}}_q[\log p(o|z,B,\nu)] \\
&+ \mathop{\mathbb{E}}_q[\log p(y|z,B,\mu)] - \mathop{\mathbb{E}}_q[\log q(\beta)] - \mathop{\mathbb{E}}_q[\log q(z)] - \mathop{\mathbb{E}}_q[\log q(\theta)] - \mathop{\mathbb{E}}_q[\log q(B)]
\end{aligned}
\tag{4}
$$

The difference, which is caused by the structured noise, is marked in cyan in $\mathscr{L}_{NCF-SLDA}$ compared to $\mathscr{L}_{NUF-SLDA}$.

Now we compute every term in the order of appearance in Equation **??**.

1st term

$$
\begin{aligned}
&E_q[\log p(w|z,\beta)] \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\mathop{\mathbb{E}}_q[\log p(w_{jn}|z_{jn},\beta)] = \sum_{j=1}^{M}\sum_{n=1}^{N}\mathop{\mathbb{E}}_q[\log\prod_{k=1}^{K}p(w_{jn}|\beta_k)^{[z_{jn}=k]}] \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\mathop{\mathbb{E}}_q[\sum_{k=1}^{K}[z_{jn}=k]\log p(w_{jn}|\beta_k)] = \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\mathop{\mathbb{E}}_q[\log p(w_{jn}|\beta_k)] \\
&\text{since } p(w_{jn}=i|\beta_k)=\beta_{ki} \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\mathop{\mathbb{E}}_q[\log\prod_{i=1}^{V}\beta_{ki}^{[w_{jn}=i]}] = \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\sum_{i=1}^{V}[w_{jn}=i]\mathop{\mathbb{E}}_q[\log\beta_{ki}] \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\Big(\sum_{i=1}^{V}\big(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}^{V}\lambda_{kp})\big)[w_{jn}=i]\Big) = \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{V}\phi_{jnk}\big(\Psi(\lambda_{ki})-\Psi(\sum_{p}\lambda_{kp})\big)[w_{jn}=i]
\end{aligned}
\tag{5}
$$

2nd term

$$
\begin{aligned}
&\mathop{\mathbb{E}}_q[\log p(z|\theta)] \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\mathop{\mathbb{E}}_q[\log p(z_{jn}|\theta_j)] \\
&\text{since } p(z_{jn}=k|\theta_j)=\theta_{jk} \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\mathop{\mathbb{E}}_q[\log\prod_{k=1}^{K}\theta_{jk}^{[z_{jn}=k]}] = \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\mathop{\mathbb{E}}_q[[z_{jn}=k]\log\theta_{jk}] \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\big(\Psi(\gamma_{jk})-\Psi(\sum_{i=1}^{K}\gamma_{ji})\big)
\end{aligned}
\tag{6}
$$

3rd term

$$
\begin{aligned}
&\mathop{\mathbb{E}}_q[\log p(\theta|\alpha)] \\
&= \sum_{j=1}^{M}\mathop{\mathbb{E}}_q[\log p(\theta_j|\alpha)] \\
&\text{using the definition of Dirichlet distribution, and we use fixed } \alpha \\
&= \sum_{j=1}^{M}\mathop{\mathbb{E}}_q[\log\frac{\Gamma(K\alpha)}{\prod_{k=1}^{K}\Gamma(\alpha)}\prod_{k=1}^{K}\theta_k^{\alpha-1}] \\
&= \sum_{j=1}^{M}\Big(\log\Gamma(K\alpha)-K\log\Gamma(\alpha)+\sum_{k=1}^{K}(\alpha-1)\mathop{\mathbb{E}}_q[\log\theta_{jk}]\Big) \\
&= \sum_{j=1}^{M}\Big(\log\Gamma(K\alpha)-K\log\Gamma(\alpha)+\sum_{k=1}^{K}(\alpha-1)\big(\Psi(\gamma_{jk})-\Psi(\sum_{i}^{K}\gamma_{ji})\big)\Big)
\end{aligned}
\tag{7}
$$

4th term

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{q}[\log p(\beta|\eta)] \\
&= \sum_{k=1}^{K} \mathop{\mathbb{E}}_{q}[\log p(\beta_k|\eta)] \\
&\text{β follows Dirichlet distribution as the previous term} \\
&= \sum_{k=1}^{K} \Big(\log\Gamma(V\eta) - V\log\Gamma(\eta) + \sum_{v=1}^{V} (\eta-1)\big(\Psi(\lambda_{kv}) - \Psi(\sum_{i=1}^{V}\lambda_{ki})\big)\Big)
\end{aligned}
\tag{8}
$$

5th term

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{q}[\log p(B|e)] \\
&= \sum_{k=1}^{K} \mathop{\mathbb{E}}_{q}[\log p(B_k|e)] \\
&\text{Use the definition of Bernoulli distribution that} \\
&p(B_k=1|e) = e \text{ and } p(B_k=0|e) = 1-e \\
&= \sum_{k=1}^{K} \mathop{\mathbb{E}}_{q}[\log e^{B_k}(1-e)^{1-B_k}] = \sum_{k=1}^{K} \mathop{\mathbb{E}}_{q}[B_k \log e + (1-B_k)\log(1-e)] \\
&= \sum_{k=1}^{K} \Big(\log(e) \mathop{\mathbb{E}}_{q}[B_k] + \log(1-e) \mathop{\mathbb{E}}_{q}[(1-B_k)]\Big) = \sum_{k=1}^{K} \Big(f_k \log(e) + (1-f_k)\log(1-e)\Big)
\end{aligned}
\tag{9}
$$

6th term

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{q}[\log p(y|z,B,\mu)] \\
&= \sum_{j=1}^{M} \mathop{\mathbb{E}}_{q}[\log p(y_j|z_j,B,\mu)] = \sum_{j=1}^{M} \mathop{\mathbb{E}}_{q}\left[\log \frac{\exp\Big(\mu_{y_j}^T\Big((\frac{1}{N}\sum_{n=1}^{N} \mathfrak{z}_{jn}) \otimes B\Big)\Big)}{\sum_{c=1}^{C}\exp\Big(\mu_c^T\Big((\frac{1}{N}\sum_{n=1}^{N} \mathfrak{z}_{jn}) \otimes B\Big)\Big)}\right] \\
&= \sum_{j=1}^{M} \left(\mathop{\mathbb{E}}_{q}\left[\mu_{y_j}^T\Big((\frac{1}{N}\sum_{n=1}^{N} \mathfrak{z}_{jn}) \otimes B\Big)\right] - \mathop{\mathbb{E}}_{q}\left[\log\left(\sum_{c=1}^{C}\exp\Big(\mu_c^T\big((\frac{1}{N}\sum_{n=1}^{N} \mathfrak{z}_{jn}) \otimes B\big)\Big)\right)\right]\right) \\
&= \sum_{j=1}^{M} \left(\mu_{y_j}^T \mathop{\mathbb{E}}_{q}\left[(\frac{1}{N}\sum_{n=1}^{N} \mathfrak{z}_{jn})\right] \otimes \mathop{\mathbb{E}}_{q}\Big[B\Big] - \mathop{\mathbb{E}}_{q}\left[\log\left(\sum_{c=1}^{C}\exp\Big(\mu_c^T\big((\frac{1}{N}\sum_{n=1}^{N} \mathfrak{z}_{jn}) \otimes B\big)\Big)\right)\right]\right) \\
&= \sum_{j=1}^{M} \left(\mu_{y_j}^T\Big((\frac{1}{N}\sum_{n=1}^{N} \phi_{jn}) \otimes f\Big) - \mathop{\mathbb{E}}_{q}\left[\log\left(\sum_{c=1}^{C}\exp\Big(\mu_c^T\big((\frac{1}{N}\sum_{n=1}^{N} \mathfrak{z}_{jn}) \otimes B\big)\Big)\right)\right]\right)
\end{aligned}
\tag{10}
$$

For the second term in Equation **??**,

$$
\begin{aligned}
&-\mathop{\mathbb{E}}_{q}\left[\log\Big(\sum_{c=1}^{C}\exp\Big(\mu_c^T\big((\frac{1}{N}\sum_{n=1}^{N}\mathfrak{z}_{jn})\otimes B\big)\Big)\Big)\right] \\
&\geq -\log\mathop{\mathbb{E}}_{q}\left[\Big(\sum_{c=1}^{C}\exp\Big(\mu_c^T\big((\frac{1}{N}\sum_{n=1}^{N}\mathfrak{z}_{jn})\otimes B\big)\Big)\Big)\right] = -\log\mathop{\mathbb{E}}_{q}\left[\Big(\sum_{c=1}^{C}\prod_{n=1}^{N}\exp\Big(\mu_c^T\big((\frac{1}{N}\mathfrak{z}_{jn})\otimes B\big)\Big)\Big)\right] \\
&\text{since } \mathfrak{z}_{jn} \text{ is the vector presentation of the topic assignment} \\
&\text{So } \mu_c^T(\mathfrak{z}_{jn}\otimes B) = \prod_{e=1}^{K}(\mu_{le}B_e)^{[z_{jn}=e]}\text{, plug this in} \\
&= -\log\mathop{\mathbb{E}}_{q}\left[\Big(\sum_{c=1}^{C}\prod_{n=1}^{N}\exp\big(\frac{1}{N}\prod_{e=1}^{K}(\mu_{le}B_e)^{[z_{jn}=e]}\big)\Big)\right] = -\log\mathop{\mathbb{E}}_{q}\left[\Big(\sum_{c=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}[z_{jn}=e]\exp\big(\frac{1}{N}\mu_{le}B_e\big)\Big)\Big)\right] \\
&= -\log\Big(\sum_{c=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\mathop{\mathbb{E}}_{q}[\exp\big(\frac{1}{N}\mu_{le}B_e\big)]\Big)\Big) \\
&\text{Since B is Bernoulli Distributed. So } \exp\big(\frac{1}{N}\mu_{le}B_e\big) = B_e\exp(\frac{1}{N}\mu_{le}) + 1 - B_e \\
&= -\log\Big(\sum_{c=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\mathop{\mathbb{E}}_{q}[B_e\exp(\frac{1}{N}\mu_{le}) + 1 - B_e]\Big)\Big) \\
&= -\log\Big(\sum_{c=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big(f_e\exp(\frac{1}{N}\mu_{le}) + 1 - f_e\big)\Big)\Big)
\end{aligned}
\tag{11}
$$

Plug the above result back to Equation **??**

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{q}[\log p(y|z,B,\mu)] \\
&= \sum_{j=1}^{M}\left(\mu_{y_j}^T\Big((\frac{1}{N}\sum_{n=1}^{N}\phi_{jn})\otimes f\Big) - \log\Big(\sum_{c=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big(f_e\exp(\frac{1}{N}\mu_{le}) + 1 - f_e\big)\Big)\Big)\right)
\end{aligned}
\tag{12}
$$

7th term

$$
\begin{aligned}
&-\mathop{\mathbb{E}}_{q}[\log q(\beta)] \\
&= \sum_{k=1}^{K}\big(-\log\Gamma(\sum_{p=1}^{V}\lambda_{kp}) + \sum_{i=1}^{V}\log\Gamma(\lambda_{ki}) - \sum_{i=1}^{V}(\lambda_{ki}-1)(\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V}\lambda_{kp}))\big)
\end{aligned}
\tag{13}
$$

8th term

$$
-\mathop{\mathbb{E}}_{q}[\log q(z)] = -\sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\log\phi_{jnk} \tag{14}
$$

9th term

$$
-\mathop{\mathbb{E}}_{q}[\log q(\theta)] = \sum_{j=1}^{M}\Big(-\log\Gamma(\sum_{p=1}^{K}\gamma_{jp}) + \sum_{i=1}^{K}\log\Gamma(\gamma_{ji}) - \sum_{i=1}^{K}(\gamma_{ji}-1)\big(\Psi(\gamma_{ji}) - \Psi(\sum_{p=1}^{K}\gamma_{jp})\big)\Big) \tag{15}
$$

10th term

$$
-\mathop{\mathbb{E}}_{q}[\log q(B)] = \sum_{k=1}^{K} -\mathop{\mathbb{E}}_{q}[\log q(B_k)] = \sum_{k=1}^{K}\big(-f_k\log f_k - (1-f_k)\log(1-f_k)\big) \tag{16}
$$

Recall the ELBO for NCF-SLDA:

$$\begin{aligned}
&\mathscr{L}_{NCF-SLDA} \\
&= \mathbb{E}_q[\log p(w|z,\beta)] + \mathbb{E}_q[\log p(z|\theta)] + \mathbb{E}_q[\log p(\theta|\alpha)] + \mathbb{E}_q[\log p(\beta|\eta)] + \mathbb{E}_q[\log p(B|e)] \\
&+ \mathbb{E}_q[\log p(o|z,B,\nu)] + \mathbb{E}_q[\log p(y|z,B,\mu)] - \mathbb{E}_q[\log q(\beta)] - \mathbb{E}_q[\log q(z)] - \mathbb{E}_q[\log q(\theta)] - \mathbb{E}_q[\log q(B)]
\end{aligned} \tag{17}$$

The additional term in NCF-SLDA compared to $NUF-SLDA$ is:

$$\begin{aligned}
&\mathbb{E}_q[\log p(o|z,B,\nu)] \\
&= \sum_{j=1}^{M} \Bigg(\nu^T \Big((\frac{1}{N}\sum_{n=1}^{N} \phi_{jn}) \otimes (1-f) \Big) - \log \Bigg(\prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big((1-f_e)\exp(\frac{1}{N}\nu_e) + f_e \big) \Big) + 1 \Bigg) \Bigg)
\end{aligned} \tag{18}$$

2 Compute the update equations

We need to update the variational parameters: λ for topic mixture, ϕ for topic assignment, γ for per document topic distribution, f for the signal-noise indicator; and the parameter for the label regression : μ for both NUF-SLDA and NCF-SLDA. The parameter for the noise label regression ν need to be updated for NCF-SLDA. With the additional term in NCF-SLDA, the update for ϕ, f will be different between there two models. However, the update for λ, γ and μ are the same for both models. We will recall this again before derive the update equation for each parameter.

Estimating λ

The addition terms that we introduced do not add any extra conditioning on β, hence, it does not affect λ. The part of the new ELBO with $\lambda$for both of our models is the same as the part of the ELBO for λ for LDA [**?**]. Hence, updating λ is the same as LDA. We will still present it here for completeness, so this martial is self-contained.

The part of the bound which varies with λ

$$\begin{aligned}
&\mathscr{L}_\lambda \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{V} \phi_{jnk} \big(\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V} \lambda_{kp}) \big) [w_{jn} = i] + \sum_{k=1}^{K} \Big(\sum_{v=1}^{V} (\eta - 1) \big(\Psi(\lambda_{kv}) - \Psi(\sum_{i=1}^{V} \lambda_{ki}) \big) \Big) \\
&+ \sum_{k=1}^{K} \Big(-\log\Gamma(\sum_{p=1}^{V} \lambda_{kp}) + \sum_{i=1}^{V} \log\Gamma(\lambda_{ki}) - \sum_{i=1}^{V} (\lambda_{ki} - 1)(\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V} \lambda_{kp})) \Big) \\
&= \sum_{k=1}^{K} \Bigg(\sum_{i=1}^{V} \Big(\sum_{j}^{M}\sum_{n}^{N} \phi_{jnk}[w_{jn} = i] + \eta - \lambda_{ki} \Big) \Psi(\lambda_{ki}) - \sum_{i=1}^{V} \Big(\sum_{j}^{M}\sum_{n}^{N} \phi_{jnk}[w_{jn} = i] + \eta - \lambda_{ki} \Big) \Psi(\sum_{p} \lambda_{kp}) \\
&- \log\Gamma(\sum_{p=1}^{V} \lambda_{kp}) + \sum_{i=1}^{V} \log\Gamma(\lambda_{ki}) \Bigg)
\end{aligned} \tag{19}$$

Compute the partial derivative, we get:

$$\begin{aligned}
&\frac{\partial \mathscr{L}_\lambda}{\partial \lambda_{ki}} \\
&= -\Psi(\lambda_{ki}) + \Big(\sum_{j=1}^{M}\sum_{n=1}^{N} \phi_{jnk}[w_{jn} = i] + \eta - \lambda_{ki} \Big) \Psi'(\lambda_{ki}) \\
&+ \Psi(\sum_{p=1} \lambda_{kp}) - \Big(\sum_{j=1}^{M}\sum_{n=1}^{N} \phi_{jnk}[w_{jn} = i] + \eta - \lambda_{ki} \Big) \Psi'(\sum_{p=1} \lambda_{ki}) - \Psi(\sum_{p=1} \lambda_{kp}) + \Psi(\lambda_{ki}) \\
&= \Big(\sum_{j=1}^{M}\sum_{n=1}^{N} \phi_{jnk}[w_{jn} = i] + \eta - \lambda_{ki} \Big) \Psi'(\lambda_{ki}) - \Big(\sum_{j=1}^{M}\sum_{n=1}^{N} \phi_{jnk}[w_{jn} = i] + \eta - \lambda_{ki} \Big) \Psi'(\sum_{p=1} \lambda_{ki})
\end{aligned} \tag{20}$$

. Set the derivative to 0, we can get the update equation for λ_{ki}

$$\lambda_{ki} = \eta + \sum_{j=1}^{M}\sum_{n=1}^{N} \phi_{jnk}[w_{jn} = i] \tag{21}$$

Estimating ϕ

How to update of ϕ is different between NCF-SLDA and NUF-SLDA. The derivation procedure is similar as SLDA, however, the details differ which is caused by the factorization construction. We will present The update for ϕ for NCF-SLDA first, since the NCF-SLDA is more complex.

For NCF-SLDA, the part of the bound with ϕ.

$$\begin{aligned}
&\mathscr{L}_{\phi}^{(NCF-SLDA)} \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{V}\phi_{jnk}\big(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}\lambda_{kp})\big)[w_{jn}=i]+\sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\big(\Psi(\gamma_{jk})-\Psi(\sum_{i=1}^{K}\gamma_{ji})\big) \\
&+\sum_{j=1}^{M}\left(\mu_{y_j}^{T}\Big((\frac{1}{N}\sum_{n=1}^{N}\phi_{jn})\cdot f\Big)-\log\Big(\sum_{l=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big(f_e\exp(\frac{1}{N}\mu_{le})+1-f_e\big)\Big)\Big)\right) \\
&+\sum_{j=1}^{M}\left(\nu^{T}\Big((\frac{1}{N}\sum_{n=1}^{N}\phi_{jn})\cdot(1-f)\Big)-\log\Big(\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big((1-f_e)\exp(\frac{1}{N}\nu_e)+f_e\big)\Big)+1\Big)\right) \\
&-\sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\log\phi_{jnk}
\end{aligned} \tag{22}$$

reorganize them a little

$$\begin{aligned}
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\left(\sum_{i=1}^{V}\phi_{jnk}\big(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}\lambda_{kp})\big)[w_{jn}=i]+\phi_{jnk}\big(\Psi(\gamma_{jk})-\Psi(\sum_{i=1}^{K}\gamma_{ji})\big)-\phi_{jnk}\log\phi_{jnk}\right) \\
&+\sum_{j=1}^{M}\left((\mu_{y_j}\cdot f)^{T}\Big(\frac{1}{N}\sum_{n=1}^{N}\phi_{jn}\Big)-\log\Big(\sum_{l=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big(f_e\exp(\frac{1}{N}\mu_{le})+1-f_e\big)\Big)\Big)\right) \\
&+\sum_{j=1}^{M}\left(\nu^{T}\Big((\frac{1}{N}\sum_{n=1}^{N}\phi_{jn})\cdot(1-f)\Big)-\log\Big(\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big((1-f_e)\exp(\frac{1}{N}\nu_e)+f_e\big)\Big)+1\Big)\right) \\
&-\sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\log\phi_{jnk}
\end{aligned}$$

For $\mathrm{i}=1\ldots K$, we set:

$$h_{\mathrm{i}}=\sum_{l=1}^{C}\left(\prod_{\substack{n=1\\ n\neq n_{now}}}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big(f_e\exp(\frac{1}{N}\mu_{le})+1-f_e\big)\Big)\big(f_{\mathrm{i}}\exp(\frac{1}{N}\mu_{l\mathrm{i}})+1-f_{\mathrm{i}}\big)\right)$$

$$\sum_{c=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big(f_e\exp(\frac{1}{N}\mu_{le})+1-f_e\big)\Big)=\sum_{\mathrm{i}=1}^{K}h_{\mathrm{i}}\phi_{jn_{now}\mathrm{i}}$$

$$g_{\mathrm{i}}=\prod_{\substack{n=1\\ n\neq n_{now}}}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big((1-f_e)\exp(\frac{1}{N}\nu_e)+f_e\big)\Big)\big((1-f_{\mathrm{i}})\exp(\frac{1}{N}\nu_{\mathrm{i}})+f_{\mathrm{i}}\big)$$

So,

$$-\log\left(\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big((1-f_e)\exp(\frac{1}{N}\nu_e)+f_e\big)\Big)+1\right)=-\log(\sum_{\mathrm{i}=1}^{K}g_{\mathrm{i}}\phi_{jn_{now}\mathrm{i}}+1)$$

Now we write the part of the bound which is affected by n_{now}. Let n_{now} be the n^{th} word in the j^{th} document.

$$\begin{aligned}
&\mathscr{L}_{\phi_{jn}}^{(NCF-SLDA)} \\
&= \sum_{k=1}^{K}\left(\sum_{i=1}^{V}\phi_{jnk}\big(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}^{V}\lambda_{kp})\big)[w_{jn}=i]+\phi_{jnk}\big(\Psi(\gamma_{jk})-\Psi(\sum_{i=1}^{K}\gamma_{ji})\big)-\phi_{jnk}\log\phi_{jnk}\right) \\
&+(\mu_{y_j}\otimes f)^{T}\frac{1}{N}\phi_{jn}-\log(h^{T}\phi_{jn})+(\nu\otimes(1-f))^{T}\frac{1}{N}\phi_{jn}-\log(g^{T}\phi_{jn}+1)
\end{aligned} \tag{23}$$

In the same manner as [?], suppose we have a previous value ϕ_{jn}^{old}. We know $\log(x) \leq \mathfrak{x}^{-1}x + \log(\mathfrak{x}) - 1$ and the equality holds if and only if $\mathfrak{x} = x$. Set $x = h^T\phi_{jn}$ and $\mathfrak{x} = h^T\phi_{jn}^{old}$, and apply it again by setting $x = (g^T\phi_{jn} + 1)$ and $\mathfrak{x} = (g^T\phi_{jn}^{old} + 1)$

We have:

$$\begin{aligned}
&\mathscr{L}_{\phi_{jn}}^{(NCF-SLDA)} \\
&\geq \sum_{k=1}^{K}\Big(\sum_{i=1}^{V}\phi_{jnk}\big(\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V}\lambda_{kp})\big)[w_{jn} = i] + \phi_{jnk}\big(\Psi(\gamma_{jk}) - \Psi(\sum_{i=1}^{K}\gamma_{ji})\big) - \phi_{jnk}\log\phi_{jnk}\Big) \\
&+ (\mu_{y_j}\otimes f)^T\frac{1}{N}\phi_{jn} - (h^T\phi_{jn}^{old})^{-1}h^T\phi_{jn} - \log(h^T\phi_{jn}^{old}) + 1 \\
&+ (\nu\otimes(1-f))^T\frac{1}{N}\phi_{jn} - (g^T\phi_{jn}^{old} + 1)^{-1}(g^T\phi_{jn} + 1) - \log(g^T\phi_{jn}^{old} + 1) + 1 \\
&= \mathscr{L}_{\phi_{jn}}^{\prime(NCF-SLDA)}
\end{aligned} \tag{24}$$

Compute the derivative assuming a given ϕ^{old}

$$\begin{aligned}
&\frac{\partial\mathscr{L}_{\phi_{jn}}^{\prime(NCF-SLDA)}}{\partial\phi_{jnk}} \\
&= \sum_{i=1}^{V}\big(\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V}\lambda_{kp})\big)[w_{jn} = i] + \big(\Psi(\gamma_{jk}) - \Psi(\sum_{i=1}^{K}\gamma_{ji})\big) - 1 - \log\phi_{jnk} \\
&+ \frac{1}{N}\mu_{y_jk}f_k - (h^T\phi_{jn}^{old})^{-1}h_k + \frac{1}{N}\nu_k(1-f_k) - (g^T\phi_{jn}^{old} + 1)^{-1}g_k
\end{aligned} \tag{25}$$

So the final fix point update for ϕ_{ink} with NCF-SLDA is

$$\begin{aligned}
\phi_{jnk} = \exp\Big(&\sum_{i=1}^{V}\big(\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V}\lambda_{kp})\big)[w_{jn} = i] + \big(\Psi(\gamma_{jk}) - \Psi(\sum_{i=1}^{K}\gamma_{ji})\big) - 1 + \frac{1}{N}\mu_{y_jk}f_k - (h^T\phi_{jn}^{old})^{-1}h_k \\
&+ \frac{1}{N}\nu_k(1-f_k) - (g^T\phi_{jn}^{old} + 1)^{-1}g_k\Big)
\end{aligned} \tag{26}$$

All additional terms in NCF-SLDA are marked cyan in the above equations. Follow the same procedure (removing all the cyan parts above), we can get the update equation for NUF-SLDA as:

$$\phi_{jnk} = \exp\Big(\sum_{i=1}^{V}\big(\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V}\lambda_{kp})\big)[w_{jn} = i] + \big(\Psi(\gamma_{jk}) - \Psi(\sum_{i=1}^{K}\gamma_{ji})\big) - 1 + \frac{1}{N}\mu_{y_jk}f_k - (h^T\phi_{jn}^{old})^{-1}h_k\Big) \tag{27}$$

Estimating γ

The update of γ will be the same for NUF-SLDA, NCF-SLDA and LDA. Since the addiction terms do not affect the part of the bound containing γ. We will present the computations details here to be self-contained.

The part of the ELBO which varies with γ is:

$$\begin{aligned}
&\mathscr{L}_{\gamma} \\
&= \sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{jnk}\big(\Psi(\gamma_{jk}) - \Psi(\sum_{i=1}^{K}\gamma_{ji})\big) + \sum_{j=1}^{M}\Big(\sum_{k=1}^{K}(\alpha-1)\big(\Psi(\gamma_{jk}) - \Psi(\sum_{i=1}^{K}\gamma_{ji})\big)\Big) \\
&+ \sum_{j=1}^{M}\Big(-\log\Gamma(\sum_{p=1}^{K}\gamma_{jp}) + \sum_{i=1}^{K}\log\Gamma(\gamma_{ji}) - \sum_{i=1}^{K}(\gamma_{ji}-1)\big(\Psi(\gamma_{ji}) - \Psi(\sum_{p=1}^{K}\gamma_{jp})\big)\Big) \\
&= \sum_{j=1}^{M}\sum_{k=1}^{K}\Big(\Big(\sum_{n=1}^{N}\phi_{jnk} + \alpha - \gamma_{jk}\Big)\Psi(\gamma_{jk}) - \Big(\sum_{n=1}^{N}\phi_{jnk} + \alpha - \gamma_{jk}\Big)\Psi(\sum_{i=1}^{K}\gamma_{ji})\Big) + \sum_{j=1}^{M}\Big(-\log\Gamma(\sum_{p=1}^{K}\gamma_{jp}) + \sum_{i=1}^{K}\log\Gamma(\gamma_{ji})\Big)
\end{aligned} \tag{28}$$

Compute the derivative:

$$\begin{aligned}
&\frac{\partial\mathscr{L}_{\gamma}}{\partial\gamma_{jk}} \\
&= \Big(\sum_{n=1}^{N}\phi_{jnk} + \alpha - \gamma_{jk}\Big)\Psi'(\gamma_{jk}) - \sum_{i=1}^{K}\Big(\sum_{n=1}^{N}\phi_{jni} + \alpha - \gamma_{ji}\Big)\Psi'(\sum_{i=1}^{K}\gamma_{ji})
\end{aligned} \tag{29}$$

Set the derivative to 0 and the update equation for γ_{jk} is:

$$\gamma_{jk} = \alpha + \sum_{n=1}^{N} \phi_{jnk} \tag{30}$$

Estimating f

The part of the ELBO which varies with f is different between NUF-SLDA and NCF-SLDA. We will present the one for NCF-SLDA first:

$$\begin{aligned}
&\mathscr{L}_f^{NCF-SLDA} \\
=&\sum_{k=1}^{K}\Big(f_k \log(e) + (1-f_k)\log(1-e)\Big) \\
&+\sum_{j=1}^{M}\Bigg(\mu_{y_j}^T\Big(\big(\frac{1}{N}\sum_{n=1}^{N}\phi_{jn}\big)\otimes f\Big) - \log\bigg(\sum_{l=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big(f_e \exp(\frac{1}{N}\mu_{le}) + 1 - f_e\big)\Big)\bigg)\Bigg) \\
&+\sum_{j=1}^{M}\Bigg(\nu^T\Big(\big(\frac{1}{N}\sum_{n=1}^{N}\phi_{jn}\big)\otimes(1-f)\Big) - \log\bigg(\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big((1-f_e)\exp(\frac{1}{N}\nu_e) + f_e\big)\Big)+1\bigg)\Bigg) \\
&+\sum_{k=1}^{K}\big(-f_k\log f_k - (1-f_k)\log(1-f_k)\big)
\end{aligned} \tag{31}$$

Let $\mathfrak{F} = \sum_{l=1}^{C}\prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big(f_e\exp(\frac{1}{N}\mu_{le}) + 1 - f_e\big)\Big)$, and $\mathfrak{G} = \prod_{n=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jne}\big((1-f_e)\exp(\frac{1}{N}\nu_e) + f_e\big)\Big)+1$ Then we compute the partial derivative:

$$\begin{aligned}
&\frac{\partial \mathscr{L}_f^{NCF-SLDA}}{\partial f_k} \\
=&\log(e) - \log(1-e) \\
&+\sum_{j=1}^{M}(\mu_{y_jk} - \nu_k)\frac{1}{N}\sum_{n=1}^{N}\phi_{jnk} \\
&-\sum_{j=1}^{M}\Bigg(\mathfrak{F}^{-1}\sum_{l=1}^{C}\bigg(\sum_{n=1}^{N}\Big(\phi_{jnk}\big(\exp(\frac{1}{N}\mu_{lk}) - 1\big)\prod_{m=1 m\neq n}^{N}\Big(\sum_{e=1}^{K}\phi_{jme}\big(f_e\exp(\frac{1}{N}\mu_{le}) + 1 - f_e\big)\Big)\Big)\bigg)\Bigg) \\
&-\sum_{j=1}^{M}\Bigg(\mathfrak{G}^{-1}\cdot\sum_{n=1}^{N}\bigg(\phi_{jnk}(1 - exp(\frac{1}{N}\nu_e))\prod_{m=1 m\neq n}^{N}\Big(\sum_{e=1}^{K}\phi_{jme}\big((1-f_e)exp(\frac{1}{N}\nu_e) + f_e\big)\Big)\bigg)\Bigg) \\
&-1 - \log f_k + 1 + \log(1-f_k) \\
=&\log(e) - \log(1-e) \\
&+\sum_{j=1}^{M}(\mu_{y_jk} - \nu_k)\frac{1}{N}\sum_{n=1}^{N}\phi_{jnk} \\
&-\sum_{j=1}^{M}\Bigg(\mathfrak{F}^{-1}\sum_{l=1}^{C}\bigg(\prod_{m=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jme}\big(f_e\exp(\frac{1}{N}\mu_{le}) + 1 - f_e\big)\Big)\sum_{n=1}^{N}\frac{\phi_{jnk}\big(\exp(\frac{1}{N}\mu_{lk}) - 1\big)}{\sum_{e=1}^{K}\phi_{jne}\big(f_e\exp(\frac{1}{N}\mu_{le}) + 1 - f_e\big)}\bigg)\Bigg) \\
&-\sum_{j=1}^{M}\Bigg(\mathfrak{G}^{-1}\cdot\prod_{m=1}^{N}\Big(\sum_{e=1}^{K}\phi_{jme}\big((1-f_e)exp(\frac{1}{N}\nu_e) + f_e\big)\Big)\cdot\sum_{n=1}^{N}\bigg(\frac{\phi_{jnk}(1 - exp(\frac{1}{N}\nu_e))}{\sum_{e=1}^{K}\phi_{jne}\big((1-f_e)exp(\frac{1}{N}\nu_e) + f_e\big)}\bigg)\Bigg) \\
&+\log\frac{1-f_k}{f_k}
\end{aligned} \tag{32}$$

There is no closed form solution for the derivative to be zero. Conjugate gradient method is used for the estimation using Equation **??** and **??**.

All additional terms which are specific for NCF-SLDA are marked by cyan. Following the same computational procedure (remove the cyan terms), we can get:

$$
\begin{aligned}
&\mathscr{L}_f^{NUF-SLDA} \\
&= \sum_{k=1}^{K} \Big(f_k \log(e) + (1 - f_k) \log(1 - e) \Big) \\
&+ \sum_{j=1}^{M} \Bigg(\mu_{y_j}^T \Big((\frac{1}{N} \sum_{n=1}^{N} \phi_{jn}) \otimes f \Big) - \log \Bigg(\sum_{l=1}^{C} \prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big(f_e \exp(\frac{1}{N} \mu_{le}) + 1 - f_e \big) \Big) \Bigg) \Bigg) \\
&+ \sum_{k=1}^{K} \big(- f_k \log f_k - (1 - f_k) \log(1 - f_k) \big),
\end{aligned} \tag{33}
$$

and

$$
\begin{aligned}
&\frac{\partial \mathscr{L}_f^{NUF-SLDA}}{\partial f_k} \\
&= \log(e) - \log(1 - e) \\
&+ \sum_{j=1}^{M} \mu_{y_j k} \frac{1}{N} \sum_{n=1}^{N} \phi_{jnk} \\
&- \sum_{j=1}^{M} \Bigg(\mathfrak{F}^{-1} \sum_{l=1}^{C} \Bigg(\prod_{m=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jme} \big(f_e \exp(\frac{1}{N} \mu_{le}) + 1 - f_e \big) \Big) \sum_{n=1}^{N} \frac{\phi_{jnk} \big(\exp(\frac{1}{N} \mu_{lk}) - 1 \big)}{\sum_{e=1}^{K} \phi_{jne} \big(f_e \exp(\frac{1}{N} \mu_{le}) + 1 - f_e \big)} \Bigg) \Bigg) \\
&+ \log \frac{1 - f_k}{f_k}
\end{aligned} . \tag{34}
$$

Equation **??** and **??** are used for the conjugate gradient method for estimating f with NUF-SLDA. In the implementation, NUF-SLDA is initialed with SLDA, then f is estimated after the initialization and all other parameters are updated with NUF-SLDA.

Estimating μ

Introducing the signal-noise indicator B will change the ELBO terms which varies with μ, hence, the update method for μ will be different from SLDA. However, the noise response does not effect μ, which means that the update method for μ will be the same for NUF-SLDA and NCF-SLDA. For both NUF-SLDA and NCF-SLDA, the part of ELBO which varies with μ is:

$$
\begin{aligned}
&\mathscr{L}_\mu \\
&= \sum_{j=1}^{M} \Bigg(\mu_{y_j}^T \Big((\frac{1}{N} \sum_{n=1}^{N} \phi_{jn}) \otimes f \Big) \\
&- \log \Big(\sum_{l=1}^{C} \prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big(f_e \exp(\frac{1}{N} \mu_{le}) + 1 - f_e \big) \Big) \Big) \Bigg)
\end{aligned} \tag{35}
$$

Recall: $\mathfrak{F} = \sum_{l=1}^{C} \prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big(f_e \exp(\frac{1}{N} \mu_{le}) + 1 - f_e \big) \Big)$

The derivative:

$$
\begin{aligned}
&\frac{\partial \mathscr{L}}{\partial \mu_{ck}} \\
&= \sum_{j=1}^{M} \Big(\frac{1}{N} \sum_{n=1}^{N} \phi_{jnk} f_k [y_j = c] \Big) - \sum_{j=1}^{M} \Bigg(\mathfrak{F}^{-1} \prod_{m=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jme} \big(f_e \exp(\frac{1}{N} \mu_{le}) + 1 - f_e \big) \Big) \sum_{n=1}^{N} \Big(\frac{\phi_{jnk} f_k \exp(\frac{1}{N} \mu_{ck}) \frac{1}{N}}{\sum_{e=1}^{K} \phi_{jne} \big(f_e \exp(\frac{1}{N} \mu_{le}) + 1 - f_e \big)} \Big) \Bigg)
\end{aligned} \tag{36}
$$

There is no closed form solution for the derivative to be zero. Conjugate gradient is used to estimate μ using Equation **??** and **??**.

Estimating ν

ν is only used in NCF-SLDA model which is the parameter for the noise label regression. The part of ELBO which varies with ν is

$$
\begin{aligned}
&\mathscr{L}_{\nu}^{NCF-SLDA} \\
&= \sum_{j=1}^{M} \Bigg(\nu^T \Big((\frac{1}{N} \sum_{n=1}^{N} \phi_{jn}) \otimes (1-f) \Big) - \log \Bigg(\prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big((1-f_e) \exp(\frac{1}{N} \nu_e) + f_e \big) \Big) + 1 \Bigg) \Bigg)
\end{aligned} \tag{37}
$$

. Let $\mathfrak{G} = \prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big((1-f_e) \exp(\frac{1}{N} \nu_e) + f_e \big) \Big) + 1$

The derivative is:

$$
\begin{aligned}
&\frac{\partial \mathscr{L}^{NCF-SLDA}}{\partial \nu_k} \\
&= \sum_{j=1}^{M} (\frac{1}{N} \sum_{n=1}^{N} \phi_{jnk}) \cdot (1 - fk) \\
&- \sum_{j=1}^{M} \Bigg(\mathfrak{G}^{-1} \prod_{m=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jme} \big((1-f_e) \exp(\frac{1}{N} \nu_e) + f_e \big) \Big) \cdot \sum_{n=1}^{N} \Big(\frac{-\phi_{jnk} f_k exp(\frac{1}{N} \nu_e) \frac{1}{N}}{\sum_{e=1}^{K} \phi_{jne} \big((1-f_e) \exp(\frac{1}{N} \nu_e) + f_e \big)} \Big) \Bigg)
\end{aligned} \tag{38}
$$

.

Similarly as μ, conjugate gradient method is used to estimate ν.

3 Implementation Details

In the implementation, the multidimensional minimizer with BFGS (Broyden Fletcher Goldfarb Shanno) is used from GSL(GNU Scientific Library). f is a Bernoulli prior which should be bounded in $[0,1]$, however, this constrain cannot be set with BFGS. Hence, we use $f_k = \frac{exp(g_k)}{exp(g_k)+1}$ instead. With this logistic trick, f is bounded in $(0,1)$, hence we convert the problem from estimating f to estimating g.

So instead of using Equation **??** and **??** with NCF-SLDA. The following two equations are used for BFGS.

$$
\begin{aligned}
&\mathscr{L}_{f}^{NCF-SLDA} \\
&= \sum_{k=1}^{K} \Big(\frac{exp(g_k)}{exp(g_k)+1} \log(e) + \frac{1}{exp(g_k)+1} \log(1-e) \Big) \\
&+ \sum_{j=1}^{M} \Bigg(\mu_{y_j}^T \Big((\frac{1}{N} \sum_{n=1}^{N} \phi_{jn}) \cdot \frac{exp(g)}{exp(g)+1} \Big) \\
&- \log \Bigg(\sum_{l=1}^{C} \prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big(\frac{exp(g_e)}{exp(g_e)+1} \exp(\frac{1}{N} \mu_{le}) + 1 - \frac{exp(g_e)}{exp(g_e)+1} \big) \Big) \Bigg) \Bigg) \\
&+ \sum_{j=1}^{M} \Bigg(\nu^T \Big((\frac{1}{N} \sum_{n=1}^{N} \phi_{jn}) \otimes (1 - \frac{exp(g)}{exp(g)+1}) \Big) \\
&- \log \Bigg(\prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big((1 - \frac{exp(g_e)}{exp(g_e)+1}) \exp(\frac{1}{N} \nu_e) + \frac{exp(g_e)}{exp(g_e)+1} \big) \Big) + 1 \Bigg) \Bigg) \\
&+ \sum_{k=1}^{K} \big(- \frac{exp(g_k)}{exp(g_k)+1} \log \frac{exp(g_k)}{exp(g_k)+1} + \frac{1}{exp(g_k)+1} \log(exp(g_k)+1) \big)
\end{aligned} \tag{39}
$$

The partial derivative is taken over g_k instead of f_k. Compared to Equation **??**, apart from replacing f_k with $\frac{exp(g_k)}{exp(g_k)+1}$. The partial derivative of $\frac{exp(g_k)}{exp(g_k)+1}$ need to be multiplied with, according to the chain rule, which is the term marked red in the following equation.

Let $\mathfrak{F} = \sum_{l=1}^{C} \prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big(\frac{exp(g_e)}{exp(g_e)+1} \exp(\frac{1}{N} \mu_{le}) + 1 - \frac{exp(g_e)}{exp(g_e)+1} \big) \Big)$, and $\mathfrak{G} = \prod_{n=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jne} \big((1 - \frac{exp(g_e)}{exp(g_e)+1}) \exp(\frac{1}{N} \nu_e) + \frac{exp(g_e)}{exp(g_e)+1} \big) \Big) + 1$

$$
\begin{aligned}
&\frac{\partial \mathscr{L}_f^{NCF-SLDA}}{\partial g_k} \\
=&\Bigg(\log(e) - \log(1-e) \\
&+ \sum_{j=1}^{M} (\mu_{y_j k} - \nu_k) \frac{1}{N} \sum_{n=1}^{N} \phi_{jnk} \\
&- \sum_{j=1}^{M} \Bigg(\mathfrak{F}^{-1} \sum_{l=1}^{C} \Bigg(\prod_{m=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jme} \big(\frac{exp(g_e)}{exp(g_e)+1} \exp(\frac{1}{N}\mu_{le}) + 1 - \frac{exp(g_e)}{exp(g_e)+1} \big) \Big) \\
&\cdot \sum_{n=1}^{N} \frac{\phi_{jnk} \big(\exp(\frac{1}{N}\mu_{lk}) - 1 \big)}{\sum_{e=1}^{K} \phi_{jne} \big(\frac{exp(g_e)}{exp(g_e)+1} \exp(\frac{1}{N}\mu_{le}) + 1 - \frac{exp(g_e)}{exp(g_e)+1} \big)} \Bigg) \Bigg) \\
&- \sum_{j=1}^{M} \Bigg(\mathfrak{G}^{-1} \cdot \prod_{m=1}^{N} \Big(\sum_{e=1}^{K} \phi_{jme} \big((1 - \frac{exp(g_e)}{exp(g_e)+1}) exp(\frac{1}{N}\nu_e) + \frac{exp(g_e)}{exp(g_e)+1} \big) \Big) \\
&\cdot \sum_{n=1}^{N} \Bigg(\frac{\phi_{jnk}(1 - exp(\frac{1}{N}\nu_e))}{\sum_{e=1}^{K} \phi_{jne} \big((1 - \frac{exp(g_e)}{exp(g_e)+1}) exp(\frac{1}{N}\nu_e) + \frac{exp(g_e)}{exp(g_e)+1} \big)} \Bigg) \Bigg) \\
&+ \log \Big((1 - \frac{exp(g_k)}{exp(g_k)+1}) / (\frac{exp(g_k)}{exp(g_k)+1}) \Big) \Bigg) \cdot \Big(\frac{exp(g_k)}{exp(g_k)+1} - \big(\frac{exp(g_k)}{exp(g_k)+1} \big)^2 \Big)
\end{aligned} \tag{40}
$$

For NUF-SLDA, it follows the same steps, and we can get the part of the new ELBO and its derivative by removing the additional terms in Equation **??** and **??**, which are marked cyan.

References

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[2] C. Wang, D. M. Blei, and L. Fei-Fei. Simultaneous image classification and annotation. In *CVPR*, 2009.