Latent Community Detection for Predicting Legislative Roll Call Votes

Elijahu Ben-Michael
Department of Statistics
UC Berkeley

Runjing Liu
Department of Statistics
UC Berkeley

Jake Soloff
Department of Statistics
UC Berkeley

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

[**?**] Voting records of legislators are commonly analyzed by political scientists to examine relationships between legislator political leanings, institutional structures, and legislative outcomes (Clinton et al 2004). For example, even simple dimensionality reduction techniques on voting data are able to uncover the political characteristics of individual legislators such as party affiliation (Figure 1a).

To capture further patterns, voting records are often used estimate legislator "ideal points." In ideal point modeling, each legislator and a given bill is presumed to lie in a latent ''ideological space," where the probability of a "yea" or "nay" response is a function of the bill's position and the congressman's position. The congressman's position is known as an ''ideal point" because his or her utility decreases as a bill's position deviates from this point.

These ideal points enable us to quantitatively characterize legislators and legislatures. The distribution of ideal points may reveal clusters of legislators corresponding for example to party lines, region, or caucus membership; furthermore, the distance between two ideal points or two clusters of ideal points can be used as a measure of political division. By visualizing policy preferences along a spectrum, interest groups are able to produce "ratings" of legislators according their leanings on a certain policy (Clinton et al 2004).

In this paper, we use roll call vote data from House of Representatives in the 110th Congress (2007-2009) to estimate ideal points and predict voting behavior for those representatives. In particular, we modify the Bayesian ideal point model proposed in Gerrish and Blei 2011; in their model, ideal points for each representative was drawn independently and identically distributed from a zero mean normal distribution. However, we propose that members of Congress should not be modeled as having independent ideal points but rather, a model should exploit the interactions among members of Congress.

To take into account these interactions, we posit that representatives in Congress belong to latent communities, and that these latent communities are manifested in two ways in our model: members of the same community tend to share similar caucuses, and members of the same community have similar ideal points. This connection between ideal points and caucus membership is made explicitly using a *stochastic block model* (see section 2.2 below).

By incorporating caucus membership data and connecting them to ideal points via latent communities, we hope to place more informed priors on the ideal points. Moreover, including more data will allow us to extend the one dimensional ideological space in Gerrish and Blei 2011 to higher dimensions. In doing so, we aim to produce more accurate prediction of legislative votes.

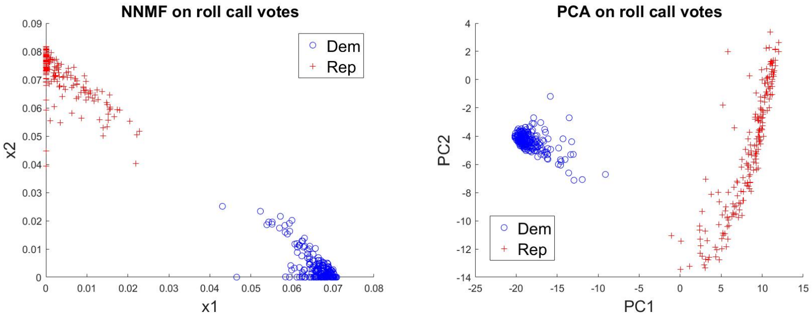
30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

1.1 Motivation

We chose to model caucus memberships because initial exploratory data analysis suggested that caucus memberships are related to a legislator's voting behavior. Figure 2 plots the number of shared caucuses between two representatives against the proportion of bills on which they voted the same way, and we see that the more caucuses two members share, the more likely they are to vote the same way.

Figure 3 shows the relationship between representatives within several caucuses in an undirected graphical model. We first used roll call vote data to infer the graph structure among the representatives in the entire House; we assumed pairwise interactions described via an Ising model in which each node denotes a binary variable of a representative voting either yes or no. The edges were inferred using neighborhood selection, and the graphs shown in figure 3 are subsets of this full graph corresponding to members of a caucus. The connectivity (#edges /#nodes(#nodes-1)) of the full graph with 448 representatives is 0.064, while the connectivity within the caucus subgraphs was much higher. This suggests that a representative more likely to be influenced by a member of his caucus than another random representative in the House.

Therefore, this strongly motivates taking into account interactions among the representatives in Congress. In particular, this analysis suggests that caucus memberships may at least partly explain whether two representatives will vote in a similar fashion. Therefore, we proceed in this project by utilizing caucus membership data and connecting them to ideal points using a stochastic block model; specifically, we hope that caucus memberships will inform a latent community structure among the representatives, and exploiting these interactions, we obtain better predictions of ideal points and hence better predictions of roll call votes.



(a) (b)

Figure 1: (a) Nonnegative matrix factorization on the 448×1707 matrix (448 representatives, 1707 bills) of roll call votes into two matrices of dimensions 448×2 and 2×1707. The rows of the 448×2 matrices were plotted to visualize the distribution of representatives in a 2D space, and we clearly see division along party lines. (b) Principle component analysis on the roll call vote data. The eigenvalues and eigenvectors of the 448×448 covariance matrix of representative voting data were computed, and each representative's voting profile was projected onto the space of the two eigenvectors with the two largest eigenvalues.

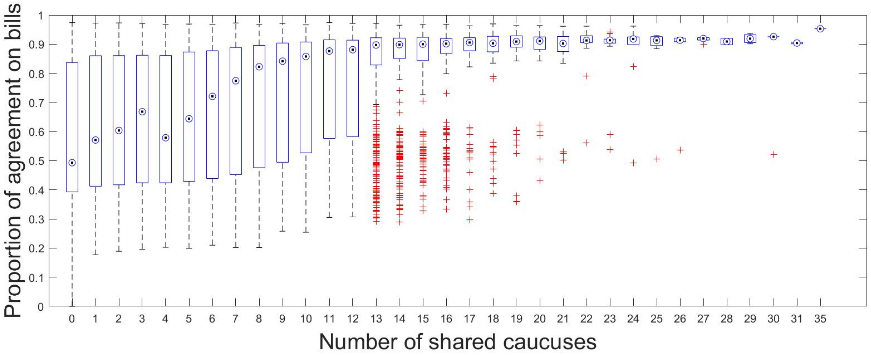
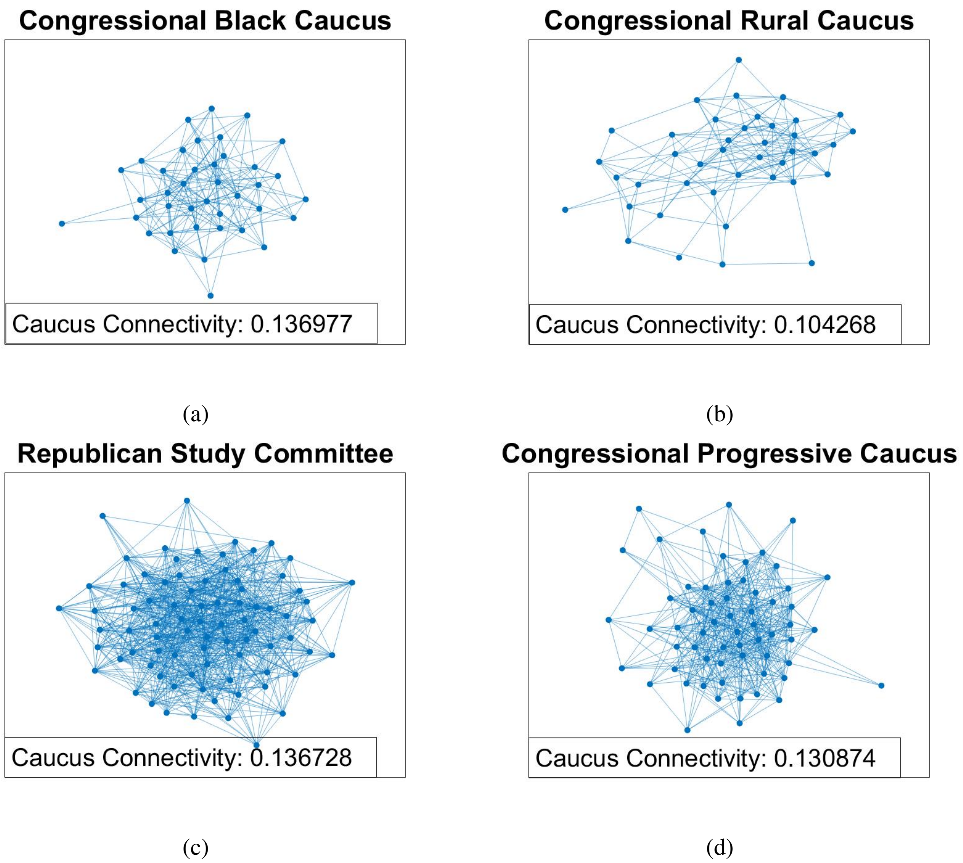


Figure 2: The distribution of agreement on bills as a function of the number of caucuses two representatives share. We see that the more caucuses people share, the more likely they are to agree on a bill.


Congressional Black Caucus
Caucus Connectivity: 0.136977
Congressional Rural Caucus
Caucus Connectivity: 0.104268
Republican Study Committee
Caucus Connectivity: 0.136728
Congressional Progressive Caucus
Caucus Connectivity: 0.130874

(a) (b)

(c) (d)

Figure 3: Graphs inferred from Neighborhood regression on House roll call vote data. Shown here are subgraphs with representatives taken from a given caucus. The caucuses are their connectivities shown here are (a) the Congressional Black Caucus, connectivity 0.137; (b) the Congressional Rural Caucus, connectivity 0.104; (c) the Republican Study Committee, connectivity 0.136; and (d) the Congressional Progressive Caucus, connectivity 0.131. In each case, the connectivity within the caucuses was higher than the connectivity of the whole graph of the House (0.064).

2 Model

2.1 Ideal Point Models

2.2 Stochastic Block Models

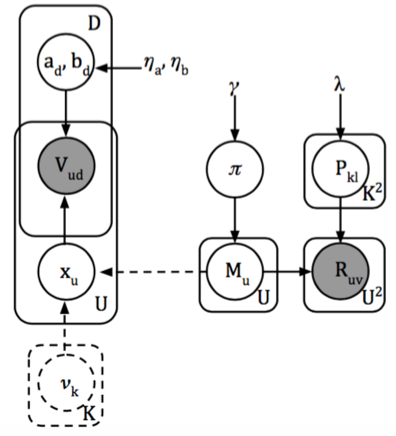


Figure 4: Graphical model for LC-IPM. Left: ideal point model. Right: stochastic block model.

3 Inference

3.1 Updates

3.2 Implementation

4 Results

5 Discussion

6 Contributions

- Bryan: pca, nbhd regression, section 1 of writeup.
- Jake: SBM, sections 2, 3, 5 and A.1 of writeup.
- Eli: dataset, IPM, experiments, sections 4 and A.2 of writeup.
- Collectively: section A.3 of writeup, poster

References

[1] Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. (2016). Variational inference: a review for statisticians. *arXiv:1601.00670.*

[2] Braun, M. & McAuliffe, J. (2010). Variational inference for large-scale models of discrete choice. *Journal of the American Statistical Association*. 105(489): 324-334.

[3] Clinton, J., Jackman, S. & Rivers D. (2004). The statistical analysis of roll call data. *American Political Science Review*. 98(2): 353-370.

[4] Gerrish, S.M. & Blei, D.M. (2011) Predicting legislative roll calls from text. *Proceedings of the 28th International Conference on Machine Learning*.

[5] Hastie, T. J., Tibshirani, R. & Wainwright, M. J. (2015). Statistical learning with sparsity: the Lasso and generalizations. *CRC Press*.

[6] Wainwright, M. J. & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning.*

A Variational updates

A.1 For SBM

After observing the symmetric matrix $R = (R_{uv})$, where R_{uv} is the number of caucuses that representatives u and v have in common, we see to find a distribution q over the latent community assignments $M = (M_u)$, the community coexpression rates $P = (P_{kl})$, and the community proportions $\pi = (\pi_k)$ which is close in relative entropy to the true posterior and lies in the factorized family $q(M)q(P)q(\pi)$. Each factor has free parameters described below and denoted with $\widehat{\text{hats}}$. The approximation q is equivalently scored by the ELBO objective $\mathcal{L}$, which we break down as:

$$\mathcal{L}(q) = \underbrace{\mathbb{E}_q\left[\log p(R \mid M, P) + \log \frac{p(P)}{q(P)}\right]}_{\mathcal{L}_{\text{data}}} + \underbrace{\mathbb{E}_q\left[-\log q(M)\right]}_{\mathcal{L}_{\text{ent}}} + \underbrace{\mathbb{E}_q\left[\log p(M \mid \pi)\right]}_{\mathcal{L}_{\text{local}}} + \underbrace{\mathbb{E}_q\left[\log \frac{p(\pi)}{q(\pi)}\right]}_{\mathcal{L}_{\text{global}}} \tag{1}$$

Variational Factors. To each u we associate variational parameters $\widehat{r}_u = (\widehat{r}_{uk})_{k=1}^K$, so

$$q(M) = \prod_{u=1}^U q(M_u \mid \widehat{r}_u) = \prod_{u=1}^U \prod_{k=1}^K \widehat{r}_{uk}^{\delta_k(M_u)}. \tag{2}$$

We define $q(\pi) \triangleq \text{Dir}(\widehat{\gamma}_1, \dots, \widehat{\gamma}_K)$ and $q(P) = \prod_{kl} q(P_{kl} \mid \widehat{\lambda}_{kl})$ where $q(P_{kl} \mid \widehat{\lambda}_{kl}) \triangleq \text{Gamma}(\widehat{\lambda}_{0kl}, \widehat{\lambda}_{1kl})$.

Computing the ELBO. Now we can write out the component terms of the ELBO more explicitly:

$$\begin{aligned}
\mathcal{L}_{\text{data}} &= \mathbb{E}_q\left[\log p(R \mid M, P) + \log \frac{p(P)}{q(P)}\right] = \sum_{kl} \mathbb{E}_q\left[\sum_{u,v} \delta_k(M_u)\delta_l(M_v) \log p(R_{uv} \mid P_{kl}) + \log \frac{p(P_{kl})}{q(P_{kl})}\right] \\
&= -\sum_{u,v} \log R_{uv}! + \sum_{k,l}\left(\lambda_0 \log \lambda_1 - \widehat{\lambda}_{0kl} \log \widehat{\lambda}_{1kl} - \log \frac{\Gamma(\lambda_0)}{\Gamma(\widehat{\lambda}_{0kl})}\right) + \sum_{k,l} \mathcal{L}_{kl}(R) \\
\mathcal{L}_{\text{ent}} &= \mathbb{E}_q\left[-\log q(M)\right] = -\sum_{u,k} \mathbb{E}_q\left[\delta_k(M_u) \log \widehat{r}_{uk}\right] = -\sum_{u,k} \widehat{r}_{uk} \log \widehat{r}_{uk} \\
\mathcal{L}_{\text{local}} &= \mathbb{E}_q\left[\log p(M \mid \pi)\right] = \sum_{u,k} \mathbb{E}_q\left[\delta_k(M_u) \log \pi_k\right] = \sum_k N_k \mathbb{E}_q\left[\log \pi_k\right] \\
\mathcal{L}_{\text{global}} &= \mathbb{E}_q\left[\log \frac{p(\pi)}{q(\pi)}\right] = \log \Gamma(C\gamma) - C \log \Gamma(\gamma) - \log \Gamma\left(\sum_k \widehat{\gamma}_k\right) + \sum_k \left\{\log \Gamma(\widehat{\gamma}_k) + (\gamma - \widehat{\gamma}_k)\mathbb{E}_q\left[\log \pi_k\right]\right\}
\end{aligned} \tag{3}$$

where $N_k = \sum_u \widehat{r}_{uk}$, $N_{kl} = \sum_{uv} \widehat{r}_{uk}\widehat{r}_{vl}$, $S_{kl} = \sum_{uv} \widehat{r}_{uk}\widehat{r}_{vl} R_{uv}$, and

$$\mathcal{L}_{kl}(R) = (S_{kl} + \lambda_0 - \widehat{\lambda}_{0kl})\mathbb{E}_q[\log P_{kl}] - (N_{kl} + \lambda_1 - \widehat{\lambda}_{1kl})\mathbb{E}_q[P_{kl}],$$

and the posterior expectations can also be computed explicitly as

$$\mathbb{E}_q[P_{kl}] = \frac{\widehat{\lambda}_{0kl}}{\widehat{\lambda}_{1kl}},\ \mathbb{E}_q[\log P_{kl}] = \psi(\widehat{\lambda}_{0kl}) - \log \widehat{\lambda}_{1kl},\ \mathbb{E}_q\left[\log \pi_k\right] = \psi\left(\widehat{\gamma}_k\right) - \psi\left(\sum_l \widehat{\gamma}_l\right)$$

CAVI Updates. The simplest approach to variational inference maximizes the ELBO $\mathcal{L}$ via coordinate-ascent, i.e. choosing the best value of a variational parameter with all others fixed. Iteratively applying these updates, the variational approximation q improves at every step toward some local optimum. Conditional conjugacy yields closed form updates for the global variational parameters.

- **Global Update to $q(\pi)$.** We have $\widehat{\gamma}_k = \gamma + N_k$.

- **Global Update to $q(P)$.** We have $\widehat{\lambda}_{0kl} = \lambda_0 + S_{kl}$ and $\widehat{\lambda}_{1kl} = \lambda_1 + N_{kl}$.
- **Local Update to $q(M)$.** Differentiating the ELBO with respect to $\widehat{r}_{uk}$,

  $$0 = \frac{\partial \mathcal{L}}{\partial \widehat{r}_{uk}} = -\log \widehat{r}_{uk} - 1 + \mathbb{E}_q[\log \pi_k] + \sum_{v \neq u} \sum_l \widehat{r}_{vl} \left(R_{uv} \mathbb{E}_q[\log P_{kl}] - \mathbb{E}_q[P_{kl}] \right).$$

  Thus, we take

  $$\widehat{r}_{uk} \propto_k \exp \left(\mathbb{E}_q[\log \pi_k] + \sum_{v \neq u} \sum_l \widehat{r}_{vl} \left(R_{uv} \mathbb{E}_q[\log P_{kl}] - \mathbb{E}_q[P_{kl}] \right) \right).$$

A.2 For IPM

We observe the votes matrix $V = (V_{ud})$ where V_{ud} is the vote of congressperson u on bill d. We have the ideal point for congressperson u, $x_u \in \mathbb{R}^s$, and the discrimination and difficulty for bill d, $a_d, b_d \in \mathbb{R}^s$. The variational distribution is fully factorized $\prod_{u=1}^{U} \prod_{d=1}^{D} q(x_u) q(a_d) q(b_d)$ where $q(x_u) \triangleq \text{Normal}(\hat{\tau}_u, \hat{\sigma}_\tau^2 I_S)$, $q(x_u) \triangleq \text{Normal}(\hat{\kappa}_{au}, \hat{\sigma}_{\kappa_a}^2 I_S)$, and $q(x_u) \triangleq \text{Normal}(\hat{\kappa}_{bu}, \hat{\sigma}_{\kappa_b}^2 I_S)$.

Computing the ELBO. We can write the ELBO as

$$\begin{aligned}
\mathcal{L}(q) &= H(q) + \sum_u \mathbb{E}_q \left[\log p(x_u) \right] + \sum_d \mathbb{E}_q \left[\log p(a_d) \right] + \sum_d \mathbb{E}_q \left[\log p(b_d) \right] + \mathbb{E}_q[\log p(V|x, a, b)] \\
H(q) &= \left(US \log 2\pi e \hat{\sigma}_\tau^2 + DS \log 2\pi e \hat{\sigma}_{\kappa_a}^2 + DS \log 2\pi e \hat{\sigma}_{\kappa_b}^2 \right) /2 \\
\mathbb{E}_q \left[\log p(x_u) \right] &= \mathbb{E}_q \left[-\frac{S}{2} \log 2\pi\sigma_x^2 - \frac{1}{2\sigma_x^2} \|x_u - \nu\|_2^2 \right] = -\frac{S}{2} - \frac{1}{2\sigma_x^2} \hat{\sigma}_\tau^2 S + \|\hat{\tau}_u - \nu\|_2^2 \\
\mathbb{E}_q \left[\log p(a_d) \right] &= \mathbb{E}_q \left[-\frac{S}{2} \log 2\pi\sigma_a^2 - \frac{1}{2\sigma_a^2} \|a_d - \eta_a\|_2^2 \right] = -\frac{S}{2} - \frac{1}{2\sigma_a^2} \hat{\sigma}_{\kappa_a}^2 S + \|\hat{\kappa}_{ad} - \eta_a\|_2^2 \\
\mathbb{E}_q \left[\log p(b_d) \right] &= \mathbb{E}_q \left[-\frac{S}{2} \log 2\pi\sigma_b^2 - \frac{1}{2\sigma_b^2} \|b_d - \eta_b\|_2^2 \right] = -\frac{S}{2} - \frac{1}{2\sigma_b^2} \hat{\sigma}_{\kappa_b}^2 S + \|\hat{\kappa}_{bd} - \eta_b\|_2^2
\end{aligned}$$

We can deal with the last expectation by using the 2nd order delta method (Braun McAullife 2008) which takes

$$\mathbb{E}[f(V)] \approx f(\mathbb{E}[V]) + \frac{1}{2} \text{trace} \left(\nabla^2 \mathbb{E}[V] \text{Cov}(V) \right).$$

Letting $u(i), d(i)$ be the users and documents for data point i, and applying this gives the approximation to the ELBO contribution from the likelihood

$$\begin{aligned}
\mathbb{E}_q[\log p(V|x, a, b)] &= \sum_{i=1}^n \mathbb{E}_q[V_i(a_{d(i)} \cdot (X_{u(i)} - b_{d(i)}))] + \mathbb{E}_q[\log(1 - \sigma(a_{d(i)} \cdot (X_{u(i)} - b_{d(i)})))] \\
&\approx \sum_{i=1}^n V_i(\hat{\kappa}_{ad(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}) - \log(1 + \exp(\hat{\kappa}_{ad(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}) \\
&\quad - \frac{1}{2} \sigma''(\kappa_{a\hat{d}(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)})(\hat{\sigma}_{\kappa_a}^2 \|\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}\|_2^2 + (\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2) \|\hat{\kappa}_{ad(i)}\|^2)
\end{aligned}$$

CAVI Updates. There are no closed form updates for $\hat{\tau}_u$, $\hat{\kappa}_{ad}$, and $\hat{\kappa}_{bd}$, so we maximize $\mathcal{L}$ numerically when updating these parameters. Let $V(u)$ be the set of votes for user u, and similarly let $V(d)$ be the set of votes on bill d. Also let $\rho_{ud} = \sigma(\widehat{\kappa_{ad(i)}} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}))$. The gradients are

$$\begin{aligned}
\nabla_{\hat{\tau}_u}\mathcal{L} &= -\frac{1}{\sigma_x^2}(\hat{\tau}_u - \nu) + \sum_{i\in V(u)} (V_i - \rho_{ud(i)})\hat{\kappa}_{ad(i)} - \sigma'(\hat{\kappa}_{ad(i)} \cdot (\hat{\tau}_u - \hat{\kappa}_{bd(i)}))\hat{\sigma}_{ka}^2(\hat{\tau}_u - \hat{\kappa}_{bd(i)}) \\
&\quad - \frac{1}{2}\sigma''(\hat{\kappa}_{ad(i)} \cdot (\hat{\tau}_u - \hat{\kappa}_{bd(i)})(\hat{\sigma}_{\kappa_a}^2 \|\hat{\tau}_u - \hat{\kappa}_{bd(i)}\|_2^2 + (\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2)\|\hat{\kappa}_{ad(i)}\|^2)\hat{\kappa}_{ad(i)} \\
\nabla_{\hat{\kappa}_{ad}}\mathcal{L} &= -\frac{1}{\sigma_a^2}(\hat{\kappa}_{ad} - \eta_a) + \sum_{i\in V(d)} (V_i - \rho_{u(i)d})(\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}) - \sigma'(\hat{\kappa}_{ad} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd})(\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2)\hat{\kappa}_{ad} \\
&\quad - \frac{1}{2}\sigma''(\hat{\kappa}_{ad} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd})(\hat{\sigma}_{\kappa_a}^2 \|\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}\|_2^2 + (\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2)\|\hat{\kappa}_{ad}\|^2)(\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}) \\
\nabla_{\hat{\kappa}_{bd}}\mathcal{L} &= \frac{1}{\sigma_b^2}(\hat{\kappa}_{bd} - \eta_b) - \sum_{i\in V(d)} (V_i - \rho_{u(i)d})\hat{\kappa}_{ad} + \sigma'(\hat{\kappa}_{ad} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd})\hat{\sigma}_{ka}^2(\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}) \\
&\quad + \frac{1}{2}\sigma''(\hat{\kappa}_{ad} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd})(\hat{\sigma}_{\kappa_a}^2 \|\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}\|_2^2 + (\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2)\|\hat{\kappa}_{ad}\|^2)\hat{\kappa}_{ad}
\end{aligned}$$

For each parameter we solve this optimization problem using L-BFGS. Finally, there are closed form updates for the variational variance parameters by taking the derivative and setting to zero

$$\begin{aligned}
\hat{\sigma}_\tau^2 &= \frac{US}{\frac{US}{\sigma_x^2} + \sum_{i=1}^n \sigma'(\widehat{\kappa_{ad(i)}} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}))(S\hat{\sigma^2}_{\kappa_a} + \|\hat{\kappa}_{ad(i)}\|_2^2)} \\
\hat{\sigma}_{\kappa_a}^2 &= \frac{DS}{\frac{DS}{\sigma_a^2} + \sum_{i=1}^n \sigma'(\widehat{\kappa_{ad(i)}} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}))(S(\hat{\sigma^2}_\tau + \hat{\sigma}_{\kappa_b}^2) + \|\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}\|_2^2)} \\
\hat{\sigma}_\tau^2 &= \frac{DS}{\frac{DS}{\sigma_b^2} + \sum_{i=1}^n \sigma'(\widehat{\kappa_{ad(i)}} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}))(S\hat{\sigma^2}_{\kappa_a} + \|\hat{\kappa}_{ad(i)}\|_2^2)}
\end{aligned}$$

A.3 For LC-IPM

The factorization for q is the same, with one more factor $q(\nu) = \prod_k q(\nu_k)$ where $q(\nu_k) \triangleq \mathcal{N}(\widehat{\mu}_k, \widehat{\sigma}_\mu^2)$. Due to the factorization in the LC-IPM generative model, the only contribution to the ELBO from IPM which changes is that corresponding to (x_u). This becomes

$$\begin{aligned}
\mathcal{L}_x = \mathbb{E}_q\left[\log\frac{p(x|\nu, M)}{q(x)}\right] &= \mathbb{E}_q\left[\log\prod_{uk}\phi(x_u|\nu_k)^{\delta_k(M_u)}\right] + H(q) \\
&= \sum_{uk}\widehat{r}_{uk}\mathbb{E}_q\left[\log\phi(x_u|\nu_k)\right] + H(q)
\end{aligned}$$

In particular, the gradient of the ELBO w.r.t. the responsibility $\widehat{r}_{uk}$ is

$$\begin{aligned}
0 = \frac{\partial\mathcal{L}_{\text{SBM}}}{\partial\widehat{r}_{uk}} + \frac{\partial\mathcal{L}_x}{\partial\widehat{r}_{uk}} &= -\log\widehat{r}_{uk} - 1 + \mathbb{E}_q\left[\log\pi_k\right] + \mathbb{E}_q\left[\log\phi(x_u|\nu_k)\right] \\
&+ \sum_{v\neq u}\sum_l \widehat{r}_{vl}\left(R_{uv}\mathbb{E}_q[\log P_{kl}] - \mathbb{E}_q[P_{kl}]\right)
\end{aligned}$$

so the update is

$$\widehat{r}_{uk} \propto_k \exp\left(\mathbb{E}_q[\log\pi_k] + \sum_{v\neq u}\sum_l \widehat{r}_{vl}\left(R_{uv}\mathbb{E}_q[\log P_{kl}] - \mathbb{E}_q[P_{kl}]\right) + \mathbb{E}_q\left[\log\phi(x_u|\nu_k)\right]\right).$$

To determine the updates for the variational mean $\widehat{\mu}_k$ corresponding to ν_k, we need the ELBO term

$$\mathcal{L}_\nu = \mathbb{E}_q\left[\log\frac{p(\nu)}{q(\nu)}\right] = \sum_k \mathbb{E}_q\left[\log p(\nu_k)\right] + KH(q(\nu_1)) = -\frac{1}{2\sigma_\nu^2}\sum_k\|\widehat{\mu}_k - \varpi\|^2 + \frac{KS}{2}\log\left(2\pi e\widehat{\sigma}_\mu^2\right) + \text{const.}$$

Setting the gradient of the ELBO w.r.t. $\widehat{\mu}_k$ equal to zero, we obtain

$$0 = \frac{\partial(\mathcal{L}_\nu + \mathcal{L}_x)}{\partial\widehat{\mu}_k} = \frac{1}{\sigma_x^2}\sum_u \widehat{r}_{uk}(\widehat{\tau}_u - \widehat{\mu}_k) - \frac{1}{\sigma_\nu^2}(\widehat{\mu}_k - \varpi) = \frac{1}{\sigma_x^2}\sum_u \widehat{r}_{uk}\widehat{\tau}_u - \left(\frac{N_k}{\sigma_x^2} + \frac{1}{\sigma_\nu^2}\right)\widehat{\mu}_k + \frac{1}{\sigma_\nu^2}\varpi$$

and thus

$$\widehat{\mu}_k = \left(\frac{\sum_u \widehat{r}_{uk}\widehat{\tau}_u}{\sigma_x^2} + \frac{\varpi}{\sigma_\nu^2}\right)\widehat{\sigma}^2_{\hat{\mu}_k};\ \text{where } \widehat{\sigma}^2_{\hat{\mu}_k} = \left(\frac{N_k}{\sigma_x^2} + \frac{1}{\sigma_\nu^2}\right)^{-1}.$$