An Empirical Study of Stochastic Variational Algorithms for the Beta Bernoulli Process Supplementary Material

Amar Shah ⋆ AS793@CAM.AC.UK
David A. Knowles † DAVIDKNOWLES@CS.STANFORD.EDU
Zoubin Ghahramani ⋆ ZOUBIN@ENG.CAM.AC.UK

⋆ University of Cambridge, Department of Engineering, Cambridge, UK

† Stanford University, Department of Computer Science, Stanford, CA, USA

In this document, we provide details about the local variable approximations introduced in the main text of the paper.

Variational Inference Schemes

Mimno-SVI

The form of the local approximation in the Mimno-SVI method is

$$\begin{aligned}\log q_{\mathrm{Mimno}}(\boldsymbol{\psi}_i) &= \mathbb{E}_{q(\boldsymbol{\beta})}[\log p(\boldsymbol{\psi}_i|\boldsymbol{y}_{1:N},\boldsymbol{\beta})] \\ &= \mathbb{E}_{q(\boldsymbol{\beta})}\left[-\frac{\gamma_{\mathrm{obs}}}{2}\left\|\boldsymbol{y}_i-(\boldsymbol{z}_i\circ\boldsymbol{w}_i)\boldsymbol{\Phi}\right\|^2+\sum_k z_{ik}\log\left(\frac{\pi_k}{1-\pi_k}\right)-\frac{\gamma_w}{2}\boldsymbol{w}_i\boldsymbol{w}_i^\top\right]+\mathrm{const} \\ &= -\frac{c}{2d}\sum_k z_{ik}w_{ik}\left[w_{ik}\left(\frac{\boldsymbol{\mu}_k\boldsymbol{\mu}_k^\top}{{\tau_k}^2}+\frac{1}{\tau_k}\right)+\left(\sum_{j\neq k}z_{ij}w_{ij}\frac{\boldsymbol{\mu}_k\boldsymbol{\mu}_j^\top}{\tau_k\tau_j}\right)-2\frac{\boldsymbol{\mu}_k\boldsymbol{y}_i^\top}{\tau_k}\right] \\ &\quad+\sum_k z_{ik}(\psi(a_k)-\psi(b_k))-\frac{e}{2f}\boldsymbol{w}_i\boldsymbol{w}_i^\top+\mathrm{const}\end{aligned}$$

It is clear that $\log q_{\mathrm{Mimno}}$ is quadratic in each w_{ik} and linear in each z_{ik}, therefore a Gibbs based sampler can easily be consructed to sample from q_{Mimno}, where w_{ik} is Gaussian given all other local variables, and z_{ik} is Bernoulli given all other local variables.

MF-SSVI

The local ELBO in the MF-SSVI framework is very similar to that of the MF-SVI, the difference being that samples of the global variables are used in MF-SSVI. The local ELBO has the following form

$$\begin{aligned}\mathcal{L}_{\mathrm{local}}^{\mathrm{MF-SSVI}} &= \frac{\gamma_{\mathrm{obs}}}{2}\sum_{i,k}\theta_{ik}\boldsymbol{\phi}_k\left[2\frac{\nu_{ik}}{\kappa_{ik}}\boldsymbol{y}_i^\top-\left(\frac{{\nu_{ik}}^2}{{\kappa_{ik}}^2}+\frac{1}{\kappa_{ik}}\right)\boldsymbol{\phi}_k^\top-\sum_{j\neq k}\theta_{ij}\frac{\nu_{ij}}{\kappa_{ij}}\frac{\nu_{ik}}{\kappa_{ik}}\boldsymbol{\phi}_j^\top\right] \\ &\quad-\frac{\gamma_w}{2}\sum_{i,k}\left(\frac{{\nu_{ik}}^2}{{\kappa_{ik}}^2}+\frac{1}{\kappa_{ik}}\right)+\sum_{i,k}\theta_{ik}\left(\frac{\pi_k}{1-\pi_k}\right) \\ &\quad-\frac{1}{2}\sum_{i,k}\log(\kappa_{ik})-\sum_{i,k}\left[\theta_{ik}\log\theta_{ik}+(1-\theta_{ik})\log(1-\theta_{ik})\right].\end{aligned}$$

This is optimized as a function of $\{\theta_{ik}, \nu_{ik}, \kappa_{ik}$ using gradient descent. Once a local optimum is found, $\mathbb{E}_{q_{\mathrm{MF}}(\boldsymbol{\psi}_{1:N}|\boldsymbol{\beta}^{(t)})}[\boldsymbol{\eta}_i]$ can be computed analytically as a function of the optimized parameters and global variable samples.

Titsias-SSVI

Recall that the Titsias-SSVI method maintains dependence between z_{ik} and w_{ik} for each k. The local ELBO for Titsias-SSVI is

$$\begin{aligned}\mathcal{L}_{\mathrm{local}}^{\mathrm{Titsias-SSVI}} = \frac{\gamma_{\mathrm{obs}}}{2}\sum_{i,k}\theta_{ik}\boldsymbol{\phi}_k\left[2\frac{\nu_{ik}}{\kappa_{ik}}\boldsymbol{y}_i^\top - \left(\frac{{\nu_{ik}}^2}{{\kappa_{ik}}^2}+\frac{1}{\kappa_{ik}}\right)\boldsymbol{\phi}_k^\top - \sum_{j\neq k}\theta_{ij}\frac{\nu_{ij}}{\kappa_{ij}}\frac{\nu_{ik}}{\kappa_{ik}}\boldsymbol{\phi}_j^\top\right] \\ - \frac{\gamma_w}{2}\sum_{i,k}\left(\frac{{\nu_{ik}}^2}{{\kappa_{ik}}^2}+\frac{1}{\kappa_{ik}}\right) + \sum_{i,k}\theta_{ik}\left(\frac{\pi_k}{1-\pi_k}\right) \\ - \frac{1}{2}\sum_{i,k}\Big[\theta_{ik}\big(\log(\kappa_{ik})-1\big) + (1-\theta_{ik})\big(\log(\gamma_w)-1\big)\Big] \\ - \sum_{i,k}\big[\theta_{ik}\log\theta_{ik} + (1-\theta_{ik})\log(1-\theta_{ik})\big].\end{aligned}$$

Again, this function is maxized as a function of $\{\theta_{ik}, \nu_{ik}, \kappa_{ik}$ using gradient descent, and the optimized parameters along with the global variable samples are used to compute $\mathbb{E}_{q_{\mathrm{Titsias}}(\boldsymbol{\psi}_{1:N}|\boldsymbol{\beta}^{(t)})}[\boldsymbol{\eta}_i]$ analytically.

Gibbs-SSVI

The Gibbs-SVI method uses the true posterior conditional distribution for local variables

$$\begin{aligned}\log q_{\mathrm{Gibbs}}(\boldsymbol{\psi}_i) &= \log p(\boldsymbol{\psi}_i|\boldsymbol{y}_{1:N},\boldsymbol{\beta}) \\ &= -\frac{\gamma_{\mathrm{obs}}}{2}\left\|\boldsymbol{y}_i - (\boldsymbol{z}_i\circ\boldsymbol{w}_i)\boldsymbol{\Phi}\right\|^2 + \sum_k z_{ik}\log\left(\frac{\pi_k}{1-\pi_k}\right) - \frac{\gamma_w}{2}\boldsymbol{w}_i\boldsymbol{w}_i^\top + \mathrm{const} \\ &= -\frac{\gamma_{\mathrm{obs}}}{2}\sum_k z_{ik}w_{ik}\boldsymbol{\phi}_k\left[w_{ik}\boldsymbol{\phi}_k^\top + \left(\sum_{j\neq k}z_{ij}w_{ij}\boldsymbol{\phi}_j^\top\right) - 2\boldsymbol{y}_i^\top\right] \\ &\quad + \sum_k z_{ik}\log\left(\frac{\pi_k}{1-\pi_k}\right) - \frac{\gamma_w}{2}\boldsymbol{w}_i\boldsymbol{w}_i^\top + \mathrm{const}\end{aligned}$$

Just as was the case with Mimno-SVI, we notice that $\log q_{\mathrm{Gibbs}}$ is quadratic in each w_{ik} and linear in each z_{ik}, therefore a Gibbs sampler can be designed to sample from q_{Gibbs}.