Proximity Variational Inference

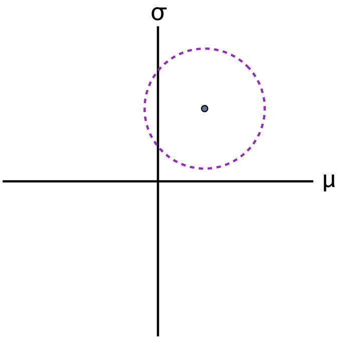
Jaan Altosaar, Rajesh Ranganath, David Blei
Princeton University, Columbia University

$$\mathcal{L}(\lambda) = \mathbb{E}_{q(\mathbf{z};\lambda)}[\log p(\mathbf{x}, \mathbf{z}) - \log q(\mathbf{z}; \lambda)]$$

Gradient ascent using proximity operators

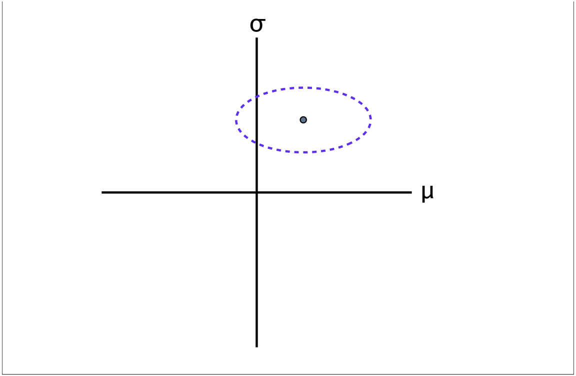
$$
\begin{aligned}
U(\boldsymbol{\lambda}_{t+1}) = & \mathcal{L}(\boldsymbol{\lambda}_t) + \nabla\mathcal{L}(\boldsymbol{\lambda}_t)^\top(\boldsymbol{\lambda}_{t+1} - \boldsymbol{\lambda}_t) \\
& - \frac{1}{2\rho}(\boldsymbol{\lambda}_{t+1} - \boldsymbol{\lambda}_t)^\top(\boldsymbol{\lambda}_{t+1} - \boldsymbol{\lambda}_t)
\end{aligned}
$$

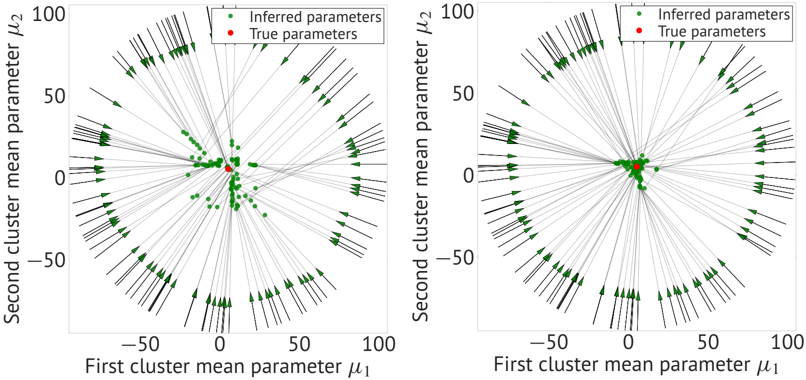
$$
\Rightarrow \boldsymbol{\lambda}^*_{t+1} = \boldsymbol{\lambda}_t + \rho\nabla\mathcal{L}(\boldsymbol{\lambda}_t)
$$



Proximity operators for variational inference

$$
\begin{aligned}
U(\boldsymbol{\lambda}_{t+1}) = &\mathcal{L}(\boldsymbol{\lambda}_t) + \nabla\mathcal{L}(\boldsymbol{\lambda}_t)^\top(\boldsymbol{\lambda}_{t+1} - \boldsymbol{\lambda}_t) \\
&- \frac{1}{2\rho}(\boldsymbol{\lambda}_{t+1} - \boldsymbol{\lambda}_t)^\top(\boldsymbol{\lambda}_{t+1} - \boldsymbol{\lambda}_t) \\
&- kd(f(\boldsymbol{\lambda}_t), f(\boldsymbol{\lambda}_{t+1}))
\end{aligned}
$$

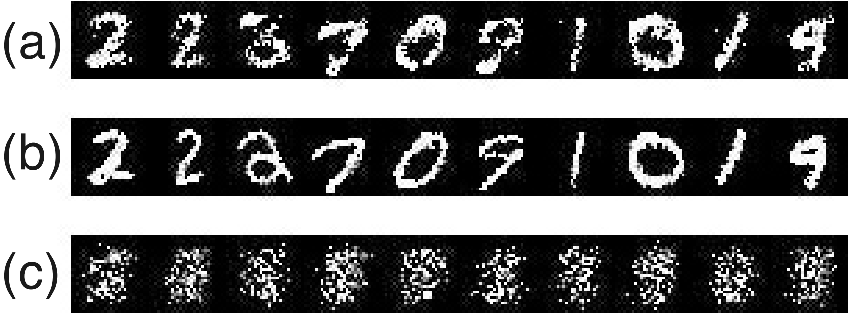




Variational Inference

Proximity Variational Inference

Binarized MNIST



(a) Proximity Variational Inference (b) Data (c) Variational Inference