An Introduction to Approximate Inference with Applications to Massive MIMO Communication Systems

Sai Subramanyam Thoota

SPC Lab, Department of ECE
Indian Institute of Science

September 7, 2019

Outline

Introduction

- Evaluation of posterior distribution of latent variables given the observed data is a central task in statistical inference
- Computationally intractable due to the integrals over the latent variables
- Approximation techniques
  - Stochastic: Sampling methods like MCMC
  - Computationally demanding
  - Deterministic: Variational inference, Expectation propagation
  - Approximate solutions but faster
- Variational inference
- Expectation propagation

Variational Bayes

- Consider a Bayesian model
  - Observations $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$
  - Latent variables $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$
- Goal is to find an approximate posterior distribution $p(\mathbf{X}|\mathbf{Z})$ and the model evidence $p(\mathbf{Z})$
  - Exact computations are computationally intractable

$$\ln p(\mathbf{Z}) = \mathcal{L}(q) + \mathrm{KL}(q\|p)$$

where

$$\mathcal{L}(q) \triangleq \int q(\mathbf{X}) \ln \left\{ \frac{p(\mathbf{Z}, \mathbf{X})}{q(\mathbf{X})} \right\} d\mathbf{X}$$

$$\mathrm{KL}(q\|p) = -\int q(\mathbf{X}) \ln \left\{ \frac{p(\mathbf{X}|\mathbf{Z})}{q(\mathbf{X})} \right\} d\mathbf{X} \geq 0$$

- Need to find a distribution $q(\mathbf{X})$ which will maximize the evidence lower bound (ELBO) $\mathcal{L}(q)$
  - Maximum occurs when $q(\mathbf{X}) = p(\mathbf{X}|\mathbf{Z}) \implies$ computational intractability
  - Impose structure on q and minimize the KL divergence

- Factorized distributions for q
  - Approximation framework developed in physics called *mean field theory*

$$q(\mathbf{X}) = \prod_{i=1}^{M} q_i(\mathbf{X}_i)$$

$$\begin{aligned}
\mathcal{L}(q) &= \int \prod_i q_i \left\{ \ln p(\mathbf{Z}, \mathbf{X}) - \sum_i \ln q_i \right\} d\mathbf{X} \\
&= \int q_j \int \ln p(\mathbf{Z}, \mathbf{X}) \prod_{i \neq j} q_i \, d\mathbf{X}_i \, d\mathbf{X}_j - \int q_j \ln q_j \, d\mathbf{X}_j - \sum_{i \neq j} \int q_i \ln q_i \, d\mathbf{X}_i \\
&= \int q_j \ln \tilde{p}(\mathbf{Z}, \mathbf{X}_j) d\mathbf{X}_j - \int q_j \ln q_j \, d\mathbf{X}_j + \text{const.} \\
&= -\text{KL}\left(q_j \| \tilde{p}(\mathbf{Z}, \mathbf{X}_j) \right) + \text{const.}
\end{aligned}$$

where

$$\ln \tilde{p}(\mathbf{Z}, \mathbf{X}_j) \triangleq \mathbb{E}_{i \neq j} \left[\ln p(\mathbf{Z}, \mathbf{X}) \right] + \text{const.}$$

- To maximize $\mathcal{L}(q)$, need to minimize the KL divergence in (1)
  - Minimum occurs when $q_j(\mathbf{X}_j) = \tilde{p}(\mathbf{Z}, \mathbf{X}_j)$

- Optimal q_j is given by

$$
\begin{aligned}
q_j^\star(\mathbf{X}_j) &= \text{const} \times \exp\left(\mathbb{E}_{i \neq j}\left[\ln p(\mathbf{Z}, \mathbf{X})\right]\right) \\
&= \frac{\exp\left(\mathbb{E}_{i \neq j}\left[\ln p(\mathbf{Z}, \mathbf{X})\right]\right)}{\int \exp\left(\mathbb{E}_{i \neq j}\left[\ln p(\mathbf{Z}, \mathbf{X})\right]\right) d\mathbf{X}_j}
\end{aligned}
$$

- Can obtain the parameters of the distribution $q_j^\star(\mathbf{Z}_j)$ by inspection
- Fix $q_{i \neq j}$ and obtain the parameters of q_j and iterate for all j

- The Exponential family

$$p(\mathbf{x}|\boldsymbol{\eta}) = h(\mathbf{x})g(\boldsymbol{\eta}) \exp\left\{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{x})\right\}$$

  where $\boldsymbol{\eta}$ is the natural parameter vector

- Maximum likelihood and sufficient statistics:

$$-\nabla \ln g(\boldsymbol{\eta}_{ML}) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{u}(\mathbf{x}_n)$$

- Differentiating $\int p(\mathbf{x}|\boldsymbol{\eta})d\mathbf{x} = 1$ gives

$$-\nabla \ln g(\boldsymbol{\eta}) = \mathbb{E}\{\mathbf{u}(\mathbf{x})\}$$

- To obtain an approximate posterior by minimizing $\mathrm{KL}(p \,\|\, q)$ instead of $\mathrm{KL}(q \,\|\, p)$

$$\mathrm{KL}(p \,\|\, q) = -\ln g(\boldsymbol{\eta}) - \boldsymbol{\eta}^T \mathbb{E}_{p(\mathbf{z})}\left[\mathbf{u}(\mathbf{z})\right] + \mathrm{const}$$

- Minimum occurs when

$$\mathbb{E}_{q(\mathbf{z})}\left[\mathbf{u}(\mathbf{z})\right] = \mathbb{E}_{p(\mathbf{z})}\left[\mathbf{u}(\mathbf{z})\right]$$

  which is called moment matching
- Joint distribution of data $\mathcal{D}$ and hidden variables $\boldsymbol{\theta}$

$$p(\mathcal{D}, \boldsymbol{\theta}) = \prod_i f_i(\boldsymbol{\theta})$$

- Approximation to the posterior distribution (from an exponential family)

$$q(\boldsymbol{\theta}) = \frac{1}{Z} \prod_i \tilde{f}_i(\boldsymbol{\theta})$$

- EP proceeds in a similar fashion as variational Bayes
  - Initialize the factors and cycle through them refining one at a time

- To obtain $\tilde{f}_j(\boldsymbol{\theta})$ by ensuring that the product

$$q^{\text{new}}(\boldsymbol{\theta}) \propto \tilde{f}_j(\boldsymbol{\theta}) \prod_{i \neq j} \tilde{f}_i(\boldsymbol{\theta})$$

  is as close as possible to $f_j(\boldsymbol{\theta}) \prod_{i \neq j} \tilde{f}_i(\boldsymbol{\theta})$

- Remove the factor $\tilde{f}_j(\boldsymbol{\theta})$ from the current approximation

$$q^{\setminus j}(\boldsymbol{\theta}) = \frac{q(\boldsymbol{\theta})}{\tilde{f}_j(\boldsymbol{\theta})}$$

- Combine it with $f_j(\boldsymbol{\theta})$ to give a distribution $\frac{1}{Z_j} f_j(\boldsymbol{\theta}) q^{\setminus j}(\boldsymbol{\theta})$
- Find $\tilde{f}_j(\boldsymbol{\theta})$ by minimizing

$$\text{KL}\left(\frac{f_j(\boldsymbol{\theta}) q^{\setminus j}(\boldsymbol{\theta})}{Z_j} \,\Big\|\, q^{\text{new}}(\boldsymbol{\theta}) \right)$$

- Since q is from an exponential family, the minimization can be done using moment matching

$$\tilde{f}_j(\boldsymbol{\theta}) = K \frac{q^{\text{new}}(\boldsymbol{\theta})}{q^{\setminus j}(\boldsymbol{\theta})}$$

- Disadvantage: No convergence guarantee in general. For exponential family, it may converge to a stationary point

Joint Channel Estimation and Data Detection for Uplink Massive MIMO Systems with Low Resolution ADCs

Motivation

- Large number of antennas results in an increased hardware complexity and circuit power consumption.
  - Power consumption of ADCs increases exponentially with respect to the number of bits per sample
- Non-linearity due to quantization necessitates novel signal processing algorithms
- Most of the existing literature on low resolution ADCs assume perfect CSI at the receiver, which is not realistic.

Goal

- Joint channel estimation and data detection in the uplink of a coded massive MIMO wireless communication system with low resolution ADCs

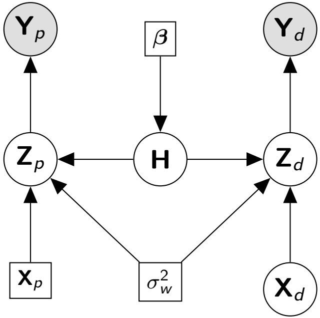
Contributions

- Joint channel estimation and data detection as a statistical inference problem in a probabilistic graphical model.
- MIMO channel, data symbols and unquantized outputs as latent variables whose marginal distributions are inferred
- Variational Bayesian inference to compute the marginal distributions of the channel and data

Received Signal

$$\mathbf{Y}_p = \mathcal{Q}(\mathbf{Z}_p) = \mathcal{Q}(\mathbf{H}\mathbf{X}_p + \mathbf{W}_p), \quad (1)$$
$$\mathbf{Y}_d = \mathcal{Q}(\mathbf{Z}_d) = \mathcal{Q}(\mathbf{H}\mathbf{X}_d + \mathbf{W}_d), \quad (2)$$

Bayesian Network Model



Quantized Variational Bayesian Joint Channel Estimation and Data Detection

- Fully factorized approximation of the posterior distribution is shown below:

$$p\left(\mathbf{Z}_p, \mathbf{Z}_d, \mathbf{X}_d, \mathbf{H} | \mathbf{Y}_p, \mathbf{Y}_d, \mathbf{X}_p; \boldsymbol{\beta}, \sigma_w^2\right) \approx q\left(\mathbf{Z}_p\right) q\left(\mathbf{Z}_d\right) q\left(\mathbf{X}_d\right) q\left(\mathbf{H}\right), \quad (3)$$

where

$$q\left(\mathbf{H}\right) = \prod_{n=1}^{N_{RX}} \prod_{k=1}^{K} q\left(h_{nk}\right), q\left(\mathbf{X}_d\right) = \prod_{k=1}^{K} \prod_{t=1}^{\tau_d} q\left(x_{d,kt}\right) \quad (4)$$

$$q\left(\mathbf{Z}_d\right) = \prod_{n=1}^{N_{RX}} \prod_{t=1}^{\tau_d} q\left(z_{d,nt}\right), q\left(\mathbf{Z}_p\right) = \prod_{n=1}^{N_{RX}} \prod_{t=1}^{\tau_p} q\left(z_{p,nt}\right). \quad (5)$$

- Conditional probability distributions of the observations and the latent variables are given as

$$p\left(\mathbf{Z}_p|\mathbf{X}_p,\mathbf{H};\sigma_w^2\right)\propto\exp\left(-\frac{1}{\sigma_w^2}\sum_{t=1}^{\tau_p}\|\mathbf{z}_{p,t}-\mathbf{H}\mathbf{x}_{p,t}\|^2\right),\tag{6}$$

$$p\left(\mathbf{Z}_d|\mathbf{X}_d,\mathbf{H};\sigma_w^2\right)\propto\exp\left(-\frac{1}{\sigma_w^2}\sum_{t=1}^{\tau_d}\left\|\mathbf{z}_{d,t}-\mathbf{H}\mathbf{x}_{d,t}\right\|^2\right),\tag{7}$$

$$p\left(\mathbf{H}|\boldsymbol{\beta}\right)\propto\exp\left(-\sum_{k=1}^{K}\frac{1}{\beta_k}\|\mathbf{h}_k\|^2\right),\tag{8}$$

$$p\left(\mathbf{Y}_d|\mathbf{Z}_d\right)=\mathbb{1}\left(\mathbf{Z}_d\in[\mathbf{Z}_d^{(lo)},\mathbf{Z}_d^{(hi)}]\right),\tag{9}$$

$$p\left(\mathbf{Y}_p|\mathbf{Z}_p\right)=\mathbb{1}\left(\mathbf{Z}_p\in[\mathbf{Z}_p^{(lo)},\mathbf{Z}_p^{(hi)}]\right),\tag{10}$$

- Computation of $q(h_{nk})$:

$$\ln q\left(h_{nk}\right) \propto \left\langle \ln p\left(\mathbf{Z}_p|\mathbf{X}_p, \mathbf{H}; \sigma_w^2\right) + \ln p\left(\mathbf{Z}_d|\mathbf{X}_d, \mathbf{H}; \sigma_w^2\right) + \ln p\left(\mathbf{H}|\boldsymbol{\beta}\right)\right\rangle, \tag{11}$$

$$\begin{aligned} &\propto -\frac{1}{\sigma_w^2}\Bigg\{\Bigg(\sum_{t=1}^{\tau_p}|x_{p,kt}|^2 + \sum_{t=1}^{\tau_d}\left\langle|x_{d,kt}|^2\right\rangle + \frac{\sigma_w^2}{\beta_k}\Bigg)|h_{nk}|^2 \\ &- 2\Re\Bigg(\Bigg(\sum_{t=1}^{\tau_p}\Bigg[\left\langle z_{p,nt}\right\rangle^* x_{p,kt} - x_{p,kt}\sum_{\substack{k'=1\\k'\neq k}}^{K} x_{p,k't}^* \left\langle h_{nk'}\right\rangle^*\Bigg] \\ &+ \sum_{t=1}^{\tau_d}\Bigg[\left\langle z_{d,nt}\right\rangle^*\left\langle x_{d,kt}\right\rangle - \left\langle x_{d,kt}\right\rangle\sum_{\substack{k'=1\\k'\neq k}}^{K}\left\langle x_{d,k't}\right\rangle^*\left\langle h_{nk'}\right\rangle^*\Bigg]\Bigg)h_{nk}\Bigg)\Bigg\} \end{aligned} \tag{12}$$

- The structure of (12) resembles that of a complex normal distribution

- Computation of $q(x_{d,kt})$:

$$q(x_{d,kt} = s) = \frac{\exp\Big(-\frac{1}{\sigma_w^2} f(s)\Big)}{\sum_{s'} \exp\Big(-\frac{1}{\sigma_w^2} f(s')\Big)}, \tag{13}$$

where s belongs to a symbol from the $M-$QAM, and

$$\begin{aligned} f(s) &= \big\langle \, \|\mathbf{h}_k\|^2 \, \big\rangle \, |s|^2 \\ &\quad + 2\Re\Big[\Big(\sum_{\substack{k'=1 \\ k' \neq k}}^{K} \big\langle \mathbf{h}_{k'} \big\rangle^H \big\langle \mathbf{h}_k \big\rangle \big\langle x_{d,k't} \big\rangle^* - \big\langle \mathbf{z}_{d,t} \big\rangle^H \big\langle \mathbf{h}_k \big\rangle \Big) s \Big]. \end{aligned}$$

- Follows a Boltzmann distribution

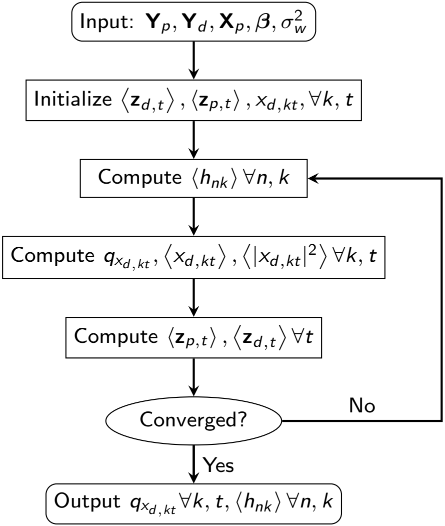
- Computation of $q(\mathbf{z}_{d,t})$ and $q(\mathbf{z}_{p,t})$:

$$\ln q\left(\mathbf{z}_{d,t}\right) \propto \left\langle \ln \mathbb{1}\left(\mathbf{z}_{d,t} \in [\mathbf{z}_{d,t}^{(lo)}, \mathbf{z}_{d,t}^{(hi)}]\right) - \frac{1}{\sigma_w^2} \left\| \mathbf{z}_{d,t} - \mathbf{H}\mathbf{x}_{d,t} \right\|^2 \right\rangle \tag{14}$$

- The structure in (14) is that of a truncated complex normal distribution with mean given below.

$$\left\langle \mathbf{z}_{d,t} \right\rangle = \boldsymbol{\mu}_{\mathbf{z}_{d,t}} + \frac{\phi\left(\frac{\mathbf{z}_{d,t}^{(lo)} - \boldsymbol{\mu}_{\mathbf{z}_{d,t}}}{\sigma_w/\sqrt{2}}\right) - \phi\left(\frac{\mathbf{z}_{d,t}^{(hi)} - \boldsymbol{\mu}_{\mathbf{z}_{d,t}}}{\sigma_w/\sqrt{2}}\right)}{\Phi\left(\frac{\mathbf{z}_{d,t}^{(hi)} - \boldsymbol{\mu}_{\mathbf{z}_{d,t}}}{\sigma_w/\sqrt{2}}\right) - \Phi\left(\frac{\mathbf{z}_{d,t}^{(lo)} - \boldsymbol{\mu}_{\mathbf{z}_{d,t}}}{\sigma_w/\sqrt{2}}\right)} \frac{\sigma_w}{\sqrt{2}},$$

Flow Diagram



THANK YOU!