Application of Variational EM to Wireless system

Rohit Budhiraja

Machine Learning for Wireless Communications (EE798L)

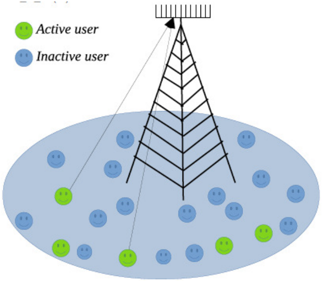
March 14, 2024

Recap of last lecture and today's agenda

- Recap of last class
  - Discussed variational EM (VEM) algorithm
- Today's agenda
  - Apply VEM algorithm to wireless system
  - Ref: Variational Approximation for Bayesian Inference [Life after the EM algorithm], IEEE Signal Processing Magazine Nov. 2008, Dimitris G. Tzikas, Aristidis C. Likas, and Nikolaos P. Galatsanos.

5G mMTC systems model (recap)

- Consider a mMTC system with M single-antenna mMTC devices and N-antenna base-station (BS)



- Only few mMTC active devices transmit data which BS need to process
- BS does not know which devices are active. All active M mMTC devices transmit simultaneously
- Total number of mMTC devices $M \gg N$ and number of active mMTC devices $K < N \ll M$
- Received signal assuming all devices are active $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$
- Tx signal $\mathbf{x} = [x_1, \cdots, x_M]^T$, rx signal $\mathbf{y} = [y_1, \cdots, y_N]^T$, and noise $\mathbf{n} = [n_1, \cdots, n_N]^T$
- Sparse transmit vector $\mathbf{x}$ contains only $K \ll M$ non-zero values $\mathbf{x} = [1, 0, , 1, , 0, 0, 0 \cdots, 0]^T$

EM for 5G mMTC systems

- 5G mMTC data model is $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ with unknown noise variance β^{-1}
- Likelihood is Gaussian $p(\mathbf{y}|\mathbf{H}, \mathbf{x}, \beta) = \mathcal{N}(\mathbf{H}\mathbf{x}, \beta^{-1}\mathbf{I})$.
- We assumed $\mathbf{x}$ as latent variable, and $\boldsymbol{\alpha}$ and β as parameters
- We assumed following sparsity-promoting prior on latent variable $\mathbf{x}$

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \mathcal{N}(0, (\mathrm{diag}(\boldsymbol{\alpha}))^{-1}) = \prod_{m=1}^{M} \mathcal{N}(0, \alpha_m^{-1})$$

- First calculated posterior distribution of $\mathbf{x}$, which we recall from lecture 10, was
  - Gaussian with $p\left(\mathbf{x}|\mathbf{y}, \mathbf{H}, \beta\right) = \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}\right)$ covariance matrix and mean (note them)

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \left(\beta\mathbf{H}^T\mathbf{H} + \mathrm{diag}(\boldsymbol{\alpha})\right)^{-1} \text{ and } \boldsymbol{\mu}_{\mathbf{x}} = \beta\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{H}^T\mathbf{y}$$

- Computed $\mathbb{E}_{\mathbf{x}}$ of log of CDLL under pos. dis. $p\left(\mathbf{x}|\mathbf{y}, \mathbf{H}, \beta\right)$ and maximized it to calculate $\boldsymbol{\alpha}$, β

$$\begin{aligned} \alpha_i^{new} &= \frac{1}{\boldsymbol{\Sigma}_{\mathbf{x}}(i,i) + (\boldsymbol{\mu}_{\mathbf{x}}(i))^2} \\ (\beta^{new})^{-1} &= \frac{1}{N}\left(||\mathbf{y} - \mathbf{H}\boldsymbol{\mu}_{\mathbf{x}}||^2 + \mathrm{Tr}[\mathbf{H}^T\mathbf{H}\boldsymbol{\Sigma}_{\mathbf{x}}]\right) \end{aligned}$$

Variational EM for 5G mMTC systems (1)

- Our 5G mMTC data model is $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ with unknown noise variance β^{-1}
- Likelihood is Gaussian $p(\mathbf{y}|\mathbf{H}, \mathbf{x}, \beta) = \mathcal{N}(\mathbf{H}\mathbf{x}, \beta^{-1}\mathbf{I})$
- We now not only assume $\mathbf{x}$ as latent variable, but also $\boldsymbol{\alpha}$ and β
  - Allows freedom in system design, which further encourages sparse solution
- We still assume the same prior on $\mathbf{x}$

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \mathcal{N}(0, (\mathrm{diag}(\boldsymbol{\alpha}))^{-1}) = \prod_{m=1}^{M} \mathcal{N}(0, \alpha_m^{-1})$$

- Assume Gamma prior over $\boldsymbol{\alpha}$ and β, which we recall is conjugate for the precision of a Gaussian

$$\begin{aligned} p(\boldsymbol{\alpha}) &\overset{(a)}{=} \prod_{m=1}^{M} \mathrm{Gamma}(\alpha_m|a, b) = \prod_{m=1}^{M} \frac{b^a}{\Gamma(a)}(\alpha_m)^{a-1} e^{-b\alpha_m} \\ p(\beta) &= \mathrm{Gamma}(\beta|c, d) \end{aligned}$$

- Equality (a) is because we assume each α_m is independently distributed
- a, b, c, d are parameter, but we fix them, and do not calculate them
- Treat $\mathbf{x}$, $\boldsymbol{\alpha}$ and β as latent variables, such that $\mathbf{Z} = \{\mathbf{x}, \boldsymbol{\alpha}, \beta\}$ and $\boldsymbol{\theta} = \{\}$

Variational EM for 5G mMTC systems (2)

- If the current estimate for the parameters is denoted $\boldsymbol{\theta}^{old}$, then variational EM algorithm is
  1. E step: use current parameter values $\boldsymbol{\theta}^{old}$ to find posterior of latent variables

$$\begin{aligned} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{old}) &= q^*(\mathbf{Z}) = \prod_{i=1}^{M} q_i^*(\mathbf{Z}_i), \text{ where} \\ \log q_j^*(\mathbf{Z}_j) &= \mathbb{E}_{i \neq j}[\log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})] + \text{constant} \end{aligned}$$

  2. Use $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{old})$ to find expectation of CDLL evaluated for some general $\boldsymbol{\theta}$

$$\sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{old}) \log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{old})$$

  3. M step: determine the revised parameter estimate $\boldsymbol{\theta}^{new}$ by maximizing expected value of CDLL

$$\boldsymbol{\theta}^{new} = \underset{\boldsymbol{\theta}}{\text{argmax}}\, \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{old}).$$

- Since, we do not have $\boldsymbol{\theta}$, there is only variational E step, and no M step
  - Recall $\mathbf{Z} = \{\mathbf{x}, \boldsymbol{\alpha}, \beta\}$ and $\boldsymbol{\theta} = \{\}$, and $\mathbf{X} = \{\mathbf{y}, \mathbf{H}\}$
- We accordingly need to calculate $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{old})$, which is now joint distribution $p(\mathbf{x}, \boldsymbol{\alpha}, \beta|\mathbf{y}, \mathbf{H})$
  - Its calculation can be easily shown to be intractable
- Using variational (mean field) approximation, with independent $\mathbf{x}, \boldsymbol{\alpha}, \beta$, posterior is
$$p(\mathbf{x}, \boldsymbol{\alpha}, \beta|\mathbf{y}, \mathbf{H}) \approx q(\mathbf{x}, \boldsymbol{\alpha}, \beta) = q(\mathbf{x})q(\boldsymbol{\alpha})q(\beta)$$

Variational E step for calculating $q(\mathbf{x})$ (1)

- According to variational EM algorithm derived earlier, we have

$$\log q_j^*(\mathbf{Z}_j) = \mathbb{E}_{i \neq j}\left[\log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})\right] + \text{constant}$$

- In this case, it reduces to

$$\begin{aligned}
\log q(\mathbf{x}) &= \mathbb{E}_{\boldsymbol{\alpha},\beta}\left[\log p(\mathbf{y}, \mathbf{H}, \mathbf{x}, \boldsymbol{\alpha}, \beta)\right] + \text{constant} \\
&= \mathbb{E}_{\boldsymbol{\alpha},\beta}[\log p(\mathbf{y}|\mathbf{H}, \mathbf{x}, \boldsymbol{\alpha}, \beta) + \log p(\mathbf{x}|\boldsymbol{\alpha}) + \underbrace{\log p(\boldsymbol{\alpha}) + \log p(\beta)}_{\text{remove terms independent of } \mathbf{x}}] + \text{constant} \\
&= \mathbb{E}_{\boldsymbol{\alpha},\beta}[\log p(\mathbf{y}|\mathbf{H}, \mathbf{x}, \beta) + \log p(\mathbf{x}|\boldsymbol{\alpha})] + \text{constant} \\
&= \mathbb{E}_{\boldsymbol{\alpha},\beta}[\log \mathcal{N}(\mathbf{Hx}, \beta^{-1}\mathbf{I}) + \log \mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\alpha})^{-1})] \\
&= \mathbb{E}_{\boldsymbol{\alpha},\beta}\left[-\frac{\beta}{2}(\mathbf{y} - \mathbf{Hx})^T(\mathbf{y} - \mathbf{Hx}) - \frac{1}{2}\sum_{m=1}^{M} \alpha_m x_m^2\right] + \text{constant} \\
&\overset{(a)}{=} -\frac{\mathbb{E}_\beta}{2}[\beta]\left(\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{Hx} + \mathbf{x}^T\mathbf{H}^T\mathbf{Hx}\right) - \frac{1}{2}\sum_{m=1}^{M} \mathbb{E}_{\boldsymbol{\alpha}}[\alpha_m]x_m^2 + \text{constant} \\
&\overset{(b)}{=} -\frac{1}{2}\mathbf{x}^T\left[\mathbb{E}_\beta[\beta]\mathbf{H}^T\mathbf{H} + \mathbb{E}_{\boldsymbol{\alpha}}[\mathbf{A}]\right]\mathbf{x} + \mathbb{E}[\beta]\mathbf{y}^T\mathbf{Hx} + \text{constant}
\end{aligned}$$

Variational E step for calculating $q(\mathbf{x})$ (2)

- Here $\mathbf{A} = \text{diag}(\boldsymbol{\alpha})$
- Equality (a): Mean field approximation according to which $\boldsymbol{\alpha}$ and β are independently distributed
- Equality (b): By rearranging the terms and by ignoring terms independent of $\mathbf{x}$
- We have from the last slide

$$\begin{aligned} q(\mathbf{x}) &= -\frac{1}{2}\mathbf{x}^T \left[\mathbb{E}_\beta[\beta]\mathbf{H}^T\mathbf{H} + \mathbb{E}_{\boldsymbol{\alpha}}[\mathbf{A}]\right]\mathbf{x} + \mathbb{E}[\beta]\mathbf{y}^T\mathbf{H}\mathbf{x} + \text{constant} \\ &= -\frac{1}{2}\mathbf{x}^T \left[\mathbb{E}_\beta[\beta]\mathbf{H}^T\mathbf{H} + \mathbb{E}_{\boldsymbol{\alpha}}[\mathbf{A}]\right]\mathbf{x} + \mathbb{E}[\beta]\mathbf{x}^T\mathbf{H}^T\mathbf{y} + \text{constant} \end{aligned}$$

- By comparing it with standard Gaussian expression,[1] we get

$$\begin{aligned} q(\mathbf{x}) &= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\Sigma}_\mathbf{x}), \text{ with} \\ \boldsymbol{\Sigma}_\mathbf{x} &= [\mathbb{E}_\beta[\beta]\mathbf{H}^T\mathbf{H} + \mathbb{E}_\alpha[\mathbf{A}]]^{-1}, \qquad \boldsymbol{\mu}_\mathbf{x} = \mathbb{E}_\beta[\beta]\boldsymbol{\Sigma}_\mathbf{x}\mathbf{H}^T\mathbf{y} \end{aligned}$$

- These updates are same as the EM-SBL updates derived earlier
  - Extra expectation over $\boldsymbol{\alpha}$ and β because we assumed them as random variables
- To calculate the expectations, we need posterior $q(\boldsymbol{\alpha})$ and $q(\beta)$ which we now calculate

[1] $-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_\mathbf{x})^T\boldsymbol{\Sigma}_\mathbf{x}^{-1}(\mathbf{x}-\boldsymbol{\mu}_\mathbf{x}) = -\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}_\mathbf{x}^{-1}\mathbf{x} + \mathbf{x}^T\boldsymbol{\Sigma}_\mathbf{x}^{-1}\boldsymbol{\mu}_x$+constant

Variational E step for calculating $q(\alpha)$ (1)

- For calculating $q(\boldsymbol{\alpha})$, we again take expectation of log of CDLL

$$\begin{aligned}\log q(\boldsymbol{\alpha}) &= \mathbb{E}_{\mathbf{x},\beta}[\log p(\mathbf{y},\mathbf{H},\mathbf{x},\boldsymbol{\alpha},\beta)] + \text{constant}\\ &= \mathbb{E}_{\mathbf{x},\beta}[\log p(\mathbf{y}|\mathbf{H},\mathbf{x},\beta) + \log p(\mathbf{x}|\boldsymbol{\alpha}) + \log p(\boldsymbol{\alpha}) + \log p(\beta)] + \text{constant}\end{aligned}$$

- Recalling that $p(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \text{Gamma}(\alpha_m|a,b) = \prod_{m=1}^{M} \frac{b^a}{\Gamma(a)}(\alpha_m)^{a-1}e^{-b\alpha_m}$
- Also, recall $p(\mathbf{x}|\boldsymbol{\alpha}) = \prod_{m=1}^{M} \mathcal{N}(0,\alpha_m^{-1})$. By retaining terms which are dependent on $\boldsymbol{\alpha}$, we get

$$\begin{aligned}\log q(\boldsymbol{\alpha}) &= \mathbb{E}_{\mathbf{x},\beta}[\log p(\mathbf{x}|\boldsymbol{\alpha}) + \log p(\boldsymbol{\alpha})] + \text{constant}\\ &= \frac{1}{2}\left(\sum_{m=1}^{M}\log\alpha_m - \sum_{m=1}^{M}\alpha_m\mathbb{E}[|x_m|^2]\right) + (a-1)\sum_{m=1}^{M}\log\alpha_m - b\sum_{m=1}^{M}\alpha_m + \text{constant}\\ &= \left(a-\frac{1}{2}\right)\sum_{m=1}^{M}\log\alpha_m - \sum_{m=1}^{M}\alpha_m\left(\frac{1}{2}\mathbb{E}[|x_m|^2] + b\right) + \text{constant}\\ &= (\tilde{a}-1)\sum_{m=1}^{M}\log\alpha_m - \sum_{m=1}^{M}\alpha_m\tilde{b}_m + \text{constant}\end{aligned}$$

- Here $\tilde{a} = a + 1/2$ and $\tilde{b}_m = \frac{1}{2}\mathbb{E}[|x_m|^2] + b$

Variational E step for calculating $q(\alpha)$ (2)

- We recall the standard form of a Gamma distribution

$$\begin{aligned} p(\boldsymbol{\alpha}) &= \prod_{m=1}^{M} \text{Gamma}(\alpha_m | a, b) = \prod_{m=1}^{M} \frac{b^a}{\Gamma(a)} (\alpha_m)^{a-1} e^{-b\alpha_m} \\ \Rightarrow \log(p(\boldsymbol{\alpha})) &= (a-1) \sum_{m=1}^{M} \log \alpha_m - \sum_{m=1}^{M} b_m \alpha_m + \text{constant} \qquad (1) \end{aligned}$$

- Our posterior expression $q(\boldsymbol{\alpha})$ from last slide is

$$\log q(\boldsymbol{\alpha}) = \tilde{a} \sum_{m=1}^{M} \log \alpha_m - \sum_{m=1}^{M} \tilde{b}_m \alpha_m + \text{constant} \qquad (2)$$

- Comparing (1) and (2) , posterior of $\boldsymbol{\alpha}$ is $q(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \text{Gamma}(\alpha_m | \tilde{a}, \tilde{b}_m)$, where

$$\begin{aligned} \tilde{a} &= a + \frac{1}{2} \\ \tilde{b}_m &= b + \frac{1}{2} \mathbb{E}[|x_m|^2] \end{aligned}$$

Variational E step for calculating $q(\beta)$

- Similarly, the posterior of β is $q(\beta) = \text{Gamma}(\beta|\tilde{c}, \tilde{d})$, with

$$\begin{aligned} \tilde{c} &= c + \frac{N}{2} \\ \tilde{d} &= d + \frac{1}{2}\mathbb{E}[||\mathbf{y} - \mathbf{H}\mathbf{x}||^2] \end{aligned}$$

Variational EM algorithm for 5G mMTC system

1. Calculate posterior $q(\mathbf{x})$ as follows

$$\begin{aligned} q(\mathbf{x}) &= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}), \text{ with} \\ \boldsymbol{\Sigma}_{\mathbf{x}} &= [\mathbb{E}_{\beta}[\beta]\mathbf{H}^T\mathbf{H} + \mathbb{E}_{\alpha}[\mathbf{A}]]^{-1}, \qquad \boldsymbol{\mu}_{\mathbf{x}} = \mathbb{E}_{\beta}[\beta]\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{H}^T\mathbf{y} \end{aligned}$$

2. Calculate $q(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \text{Gamma}(\alpha_m|\tilde{a}, \tilde{b}_m)$ with $\mathbb{E}_{\alpha_m}[\alpha_m] = \tilde{a}/\tilde{b}_m$
   - To calculate $\tilde{b}_m$, we will use posterior q($\mathbf{x}$)
3. Calculate posterior of β is $q(\beta) = \text{Gamma}(\beta|\tilde{c}, \tilde{d})$, with $\mathbb{E}_{\beta}[\beta] = \tilde{c}/\tilde{d}$
   - To calculate $\tilde{d}$, we will use posterior q($\mathbf{x}$)
4. Keep repeating above step till convergence

Proof of pudding is in the eating