1-bit Quantized Variational Bayesian Soft Symbol Decoder for Massive MIMO Systems

Sai Subramanyam Thoota

SPC Lab, Department of ECE
Indian Institute of Science

March 30, 2019

Table of contents

System Model & Problem Statement

- Uplink of a single cell massive MIMO wireless communication system
  - N_r receive antennas at the base station (BS)
  - K single transmit antenna users
- Complex baseband received signal at the t^{th} symbol interval $\mathbf{y}_c^{(t)} \in \{\pm 1 \pm j\}^{N_r \times 1}$ at the BS:

$$\mathbf{y}_c^{(t)} = \mathcal{Q}\left(\mathbf{z}_c^{(t)}\right) = \mathcal{Q}\left(\mathbf{H}_c \mathbf{x}_c^{(t)} + \mathbf{w}_c^{(t)}\right)$$

  where
  - $\mathbf{z}_c^{(t)} \in \mathbb{C}^{N_r \times 1}$ is the unquantized received signal
  - $\mathbf{H}_c = [\mathbf{h}_c(1), \ldots, \mathbf{h}_c(K)] \in \mathbb{C}^{N_r \times K}$ is the channel matrix
  - Each entry of $\mathbf{H}_c$ distributed as i.i.d. $\mathcal{CN}(0, 1)$
  - $\mathbf{w}_c^{(t)} \in \mathbb{C}^{N_r \times 1}$ is the additive white Gaussian noise distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$
  - $\mathbf{x}_c^{(t)} = [x_c^{(t)}(1), \ldots, x_c^{(t)}(K)]^T \in \left\{\pm \frac{1}{\sqrt{2}} \pm \frac{j}{\sqrt{2}})\right\}^{K \times 1}$ is the QPSK modulated transmit symbols of the K users at the t^{th} symbol interval
- We assume perfect knowledge of the channel state information at the transmitter (CSIT)
  - Extensions (TBD): Imperfect CSIT, pilot contamination effects in a multicell scenario on the BER performance

System Model & Problem Statement contd.

- Converting the signal into real domain, we get

$$\begin{aligned}\mathbf{y}^{(t)} = \mathcal{Q}\left(\mathbf{z}_{re}^{(t)}\right) &= \mathcal{Q}\left(\mathbf{H}_{re}\mathbf{x}_{re}^{(t)} + \mathbf{w}^{(t)}\right)\\ &= \mathcal{Q}\left(\mathbf{H}\mathbf{b}^{(t)} + \mathbf{w}^{(t)}\right)\end{aligned}$$

  where
  - $\mathbf{b}^{(t)} = \sqrt{2}\mathbf{x}_{re}^{(t)}$, $\mathbf{H} = \frac{1}{\sqrt{2}}\mathbf{H}_{re} \in \mathbf{C}^{2N_r \times 2K}$
  - $\mathbf{b} \in \{\pm 1\}^{2K \times 1}$
  - $\mathbf{w}^{(t)} \sim \mathcal{CN}(\mathbf{0}, \frac{\sigma^2}{2}\mathbf{I}_{2N_r})$
- LTE/LTE-Advanced, 5G wireless communication systems have a channel encoder/decoder at the Tx/Rx
  - Instead of detecting the hard QPSK symbols, we need soft decisions (beliefs/probabilities) at the output of the demodulator
- Goal is to obtain the soft decisions of the transmit symbols **b** and to decode the transmit bits from the 1-bit quantized received signal
  - Currently, QPSK modulation is considered because of the 1-bit quantization
  - Further extensions: Can any space-time coding scheme be designed to decode higher order modulation schemes?

Variational Bayesian Inference

- 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 approximation for the 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*

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

$$\begin{aligned}
\mathcal{L}(q) &= \int \prod_i q_i \left\{ \ln p(\mathbf{Z}, \mathbf{X}) - \sum_i \ln q_i \right\} d\mathbf{X} \\
&= \int q_j \int \ln p(\mathbf{Z}, \mathbf{X}) \prod_{i \neq j} q_i d\mathbf{X}_i d\mathbf{X}_j - \int q_j \ln q_j d\mathbf{X}_j - \sum_{i \neq j} \int q_i \ln q_i d\mathbf{X}_i \\
&= \int q_j \ln \tilde{p}(\mathbf{Z}, \mathbf{X}_j) d\mathbf{X}_j - \int q_j \ln q_j d\mathbf{X}_j + \text{const.} \\
&= -\mathrm{KL}\left(q_j \| \tilde{p}(\mathbf{Z}, \mathbf{X}_j) \right) + \text{const.}
\end{aligned} \tag{1}$$

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 $q_j(\mathbf{X}_j) = \tilde{p}(\mathbf{X}, \mathbf{X}_j)$

- 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

Quantized Variational Bayesian Soft Symbol Decoder (QVBSSD)

Bayesian Network Model for the massive MIMO wireless communication system

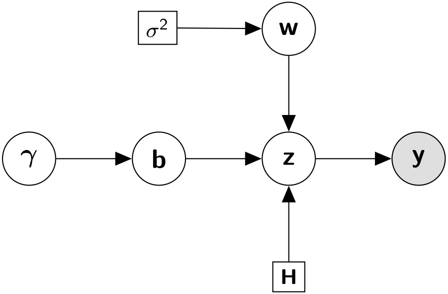


Figure: Graphical model for the quantized MU-MIMO wireless communication system

- The transmit bits are parameterized by γ

$$p\left(b_n = 1 | \gamma_n\right) = \gamma_n = 1 - p\left(b_n = -1 | \gamma_n\right)$$

- The joint distribution of the observations and the latent variables is given by

$$p\left(\mathbf{y}, \mathbf{z}, \mathbf{b}, \gamma | \mathbf{H}; \sigma^2\right) = p\left(\mathbf{y}|\mathbf{z}\right) p\left(\mathbf{z}|\mathbf{H}, \mathbf{b}; \sigma^2\right) p\left(\mathbf{b}|\gamma\right) p\left(\gamma\right)$$

- The conditional distributions of the latent variables is computed as

$$p(\mathbf{z}|\mathbf{H},\mathbf{b};\sigma^2) = \frac{1}{(\pi\sigma^2)^{2N_r}} \exp\left(-\frac{1}{\sigma^2}\|\mathbf{z}-\mathbf{H}\mathbf{b}\|^2\right)$$

$$p(\mathbf{b}|\boldsymbol{\gamma}) = \prod_{n=1}^{2K} p(b_n|\gamma_n) = \prod_{n=1}^{2K} \gamma_n^{\frac{1+b_n}{2}} (1-\gamma_n)^{\frac{1-b_n}{2}}$$

$$p(\boldsymbol{\gamma}) = \prod_{n=1}^{2K} \mathbb{1}(\gamma_n \in [0,1])$$

- The conditional distribution of the quantized received signal given the unquantized signal is given as

$$p(\mathbf{y}|\mathbf{z}) = \mathbb{1}\left(\mathbf{z} \in \left[\mathbf{z}^{(lo)}, \mathbf{z}^{(hi)}\right]\right)$$

where $\mathbf{z}^{(lo)}$ and $\mathbf{z}^{(hi)}$ are the lower and upper thresholds corresponding to the observations $\mathbf{y}$

- Approximation of the posterior distribution using a factorized distribution as:

$$p\left(\mathbf{z}, \mathbf{b}, \gamma | \mathbf{y}, \mathbf{H}; \sigma^2\right) = q_{\mathbf{z}}\left(\mathbf{z}\right) \prod_{n=1}^{2K} q_{b_n}\left(b_n\right) q_{\gamma_n}\left(\gamma_n\right)$$

- Computation of $q_{\mathbf{z}}(\mathbf{z})$:

$$\begin{aligned} \ln q_{\mathbf{z}}(\mathbf{z}) &\propto \left\langle \ln p\left(\mathbf{y}|\mathbf{z}\right) + \ln p\left(\mathbf{z}|\mathbf{H}, \mathbf{b}; \sigma^2\right)\right\rangle_{q_{\mathbf{b}}(\mathbf{b})} \\ &\propto \left\langle \ln \mathbb{1}\left(\mathbf{z} \in \left[\mathbf{z}^{(lo)}, \mathbf{z}^{(hi)}\right]\right) - \frac{1}{\sigma^2}\left\|\mathbf{z} - \mathbf{H}\mathbf{b}\right\|^2 \right\rangle_{q_{\mathbf{b}}(\mathbf{b})} \end{aligned}$$

- $\mathbf{z}$ is truncated Gaussian distributed with mean

$$\langle \mathbf{z} \rangle = \mathbf{H}\langle \mathbf{b} \rangle + \frac{\phi\left(\frac{\mathbf{z}^{(lo)} - \mathbf{H}\langle \mathbf{b}\rangle}{\sigma/\sqrt{2}}\right) - \phi\left(\frac{\mathbf{z}^{(hi)} - \mathbf{H}\langle \mathbf{b}\rangle}{\sigma/\sqrt{2}}\right)}{\Phi\left(\frac{\mathbf{z}^{(hi)} - \mathbf{H}\langle \mathbf{b}\rangle}{\sigma/\sqrt{2}}\right) - \Phi\left(\frac{\mathbf{z}^{(lo)} - \mathbf{H}\langle \mathbf{b}\rangle}{\sigma/\sqrt{2}}\right)} \frac{\sigma}{\sqrt{2}}$$

where ϕ and Φ are the pdf & cdf of the standard normal distribution

- Computation of $q_{\gamma_n}(\gamma_n)$:

$$\begin{aligned}
\ln q_{\gamma_n}(\gamma_n) &\propto \langle \ln p(\mathbf{b}|\boldsymbol{\gamma}) + \ln p(\boldsymbol{\gamma})\rangle_{\{q_{\gamma_m}(\gamma_m)\}_{\substack{m=1\\ m\neq n}}^{2K} q_{\mathbf{b}}(\mathbf{b})} \\
&\propto \left\langle \sum_{k=1}^{2K} \frac{1+b_k}{2}\ln\gamma_k + \frac{1-b_k}{2}\ln(1-\gamma_k) + \ln \mathbb{1}(\gamma_k \in [0,1]) \right\rangle_{\{q_{\gamma_m}(\gamma_m)\}_{\substack{m=1\\ m\neq n}}^{2K} q_{\mathbf{b}}(\mathbf{b})} \\
&\propto \frac{1+\langle b_n\rangle}{2}\ln\gamma_n + \frac{1-\langle b_n\rangle}{2}\ln(1-\gamma_n) + \ln \mathbb{1}(\gamma_n \in [0,1])
\end{aligned}$$

- γ_n is beta distributed with parameters $\frac{3+\langle b_n\rangle}{2}$, and $\frac{3-\langle b_n\rangle}{2}$, and mean

$$\begin{aligned}
\langle\gamma_n\rangle &= \frac{3+\langle b_n\rangle}{6} \\
\langle \ln\gamma_n - \ln(1-\gamma_n)\rangle &= \psi\left(\frac{3+\langle b_n\rangle}{2}\right) - \psi\left(\frac{3-\langle b_n\rangle}{2}\right)
\end{aligned}$$

where $\psi(.)$ is the digamma function[1]

[1] $\psi(x) = \frac{d}{dx}\ln\Gamma(x)$

- Computation of $q_{b_n}(b_n)$

$$\ln q_{b_n}(b_n) \propto \left\langle \ln p\left(\mathbf{z}|\mathbf{H},\mathbf{b};\sigma^2\right) + \ln p\left(\mathbf{b}|\boldsymbol{\gamma}\right)\right\rangle_{q_{\mathbf{z}}(\mathbf{z}),q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}),\{q_{b_m}(b_m)_{\substack{m=1\\m\neq n}}^{2K}}$$

$$\propto \left\langle -\frac{1}{\sigma^2}\left\|\mathbf{z}-\mathbf{H}\mathbf{b}\right\|^2 + \sum_{k=1}^{2K}\frac{1+b_k}{2}\ln\gamma_k + \frac{1-b_k}{2}\ln(1-\gamma_k)\right\rangle_{q_{\mathbf{z}}(\mathbf{z}),q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}),\{q_{b_m}(b_m)\}_{\substack{m=1\\m\neq n}}^{2K}}$$

$$\propto -\frac{1}{\sigma_n^2}\left(b_n^2 - 2b_n\mu_n\right) \tag{2}$$

where

$$\mu_n = \frac{\langle\mathbf{z}\rangle^T\mathbf{H}_{(:,n)} - \mathbf{H}_{(:,n)}^T\mathbf{H}\langle\mathbf{b}\rangle + \sum_{m=1}^{2N_r}\mathbf{H}_{(m,n)}^2\langle b_n\rangle + \frac{\sigma^2}{4}\langle\ln\gamma_n - \ln(1-\gamma_n)\rangle}{\sum_{m=1}^{2N_r}\mathbf{H}_{(m,n)}^2} \tag{3}$$

$$\sigma_n^2 = \frac{\sigma^2/2}{\sum_{m=1}^{2N_r}\mathbf{H}_{(m,n)}^2} \tag{4}$$

- b_n is Gaussian distributed with mean and variance given in (3) and (4) respectively

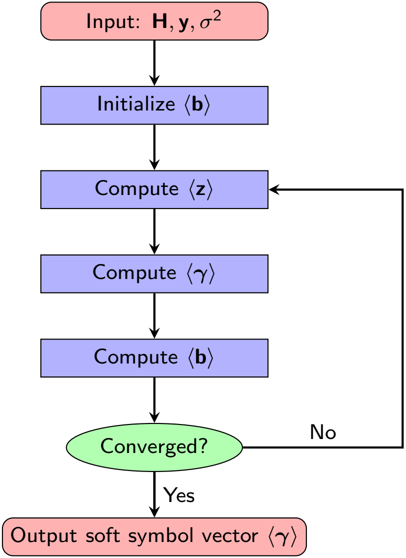
- Truncated Gaussian approximation for b_n
  - $\langle \ln \gamma_n - \ln(1-\gamma_n)\rangle$ needs $\langle b_n \rangle$ to be in the interval $(-3, 3)$ as the digamma function $\psi(x)$ cannot be evaluated for $x \leq 0$
  - Adding and subtracting a constant term to (2), we approximate the Gaussian distribution with a truncated Gaussian distribution with mean given by [2]

$$\langle b_n \rangle = \mu_n + \frac{\phi\left(-\frac{3+\mu_n}{\sigma_n}\right) - \phi\left(\frac{3+\mu_n}{\sigma_n}\right)}{\Phi\left(\frac{3+\mu_n}{\sigma_n}\right) - \Phi\left(-\frac{3+\mu_n}{\sigma_n}\right)}\sigma_n$$

where ϕ and Φ are the pdf & cdf of the standard normal distribution

[2] Infinite loop due to this sometimes. Currently removed it with a heuristic hack using an approach taken in simulated annealing i.e., decreasing the limits of the truncated Gaussian if needed. But not sure whether the infinite loop will be caught in some other corner cases. Need to check this in more detail

QVBSSD Flow Diagram



Preliminary Simulation Results

Simulation Setup

- Number of users $K = 30$
- Number of receive antennas $N_r = 100$
- Channel Code:
  - LDPC Code rate 1/3
  - Dimension of parity check matrix: 2304×3456
- QPSK symbols normalized to have average energy of 1

Preliminary Simulation Results contd.

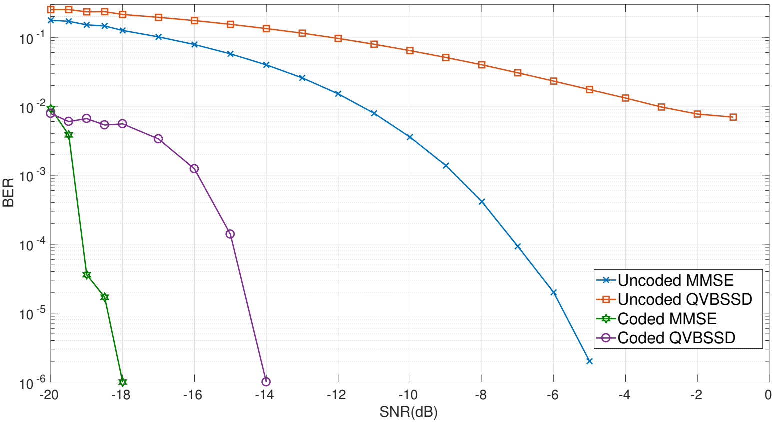


Figure: $N_r = 100, K = 30$, LDPC Code Rate 1/3

Future Work

- Pilot contamination and imperfect CSIT effects on the BER performance
- Extension to higher order modulation schemes (new signaling scheme)
- DNN implementation of QVBSSD
- Sparse signal recovery extension of QVBSSD
  - Building up a use-case. Just thinking aloud!
  - In MMTC, most of the devices are stationary whose channels have a larger coherence period. Can we include the channels as the columns of the dictionary and try to recover the transmit symbols which are the rows of the sparse MMV?
  - Channel errors $\implies$ dictionary errors. Alternating optimization of the dictionary (channel matrix) and the transmit symbols...
- BER analysis???

THANK YOU!