Multiuser Detection of M-QAM Symbols via Bit-Level Equalization and Soft Detection

Darryl Lin and Teng Joon Lim
Department of Electrical & Computer Engineering
University of Toronto
Outline

- **Multi-User Detection With Prior Knowledge**
  - Needed for turbo multi-user detection
  - A number of heuristic techniques
- **Our proposal:**
  - Use variational inference as unifying design tool
  - Several turbo MUD methods can be derived like this
  - Can also deal with Gray-coded M-QAM in a natural way
Decoding in an Interference Channel

- At tx’er: Coding - Interleaving - QAM bit-to-symbol mapping - Channelization (CDMA)

Inf. bits $u_k$ → Encoder → Code bits $c_k$ → $\pi_k$ → $b_k$ → QAM → $d_k$ → $s_k$ → $x_k$ → User $k$’s transmitter
Decoding in an Interference Channel

Channel: Multi-access interference, unfaded.
At rx’er: Iterative decoding and multi-user detection.
Optimal decoding and detection too complex.
Iterative Decoding Basics

“Straightforward” iterative decoding by sum-product algorithm requires full-blown APP updates in multi-user channel.

\[ d_k = k\text{-th user's channel symbol} \]
Iterative Decoding Basics

- Need sub-optimal MUD to generate approximate posterior symbol probabilities.
- For example, use Wang/Poor’s LMMSE-based detector:
  - Compute LMMSE filter output per iteration per user, $\hat{d}_k$.
  - Assume $\hat{d}_k = \mu_k d_k + \eta_k$, where $\eta_k$ is Gaussian. With known channels and AWGN variance, both $\mu_k$ and variance of $\eta_k$ can be found.
Iterative Decoding Basics

- Then APP of symbol $d_k$ can be found, assuming
  $$P(d_k|\mathbf{r}) = P(d_k|\hat{d}_k)$$
- From symbol APP, bit APPs can be found by summing over $2^{L-1}$ terms, in $2^L$-ary modulation:
  $$P(b_{1k}^1 = 0|\mathbf{r}) = \sum_{d_k : b_{1k}^1 = 0} P(d_k|\mathbf{r})$$
- Other turbo MUDs for $M$-ary modulation can be defined, using similar heuristic assumptions.
  - E.g. interference cancellation.
Variational Inference As A Unified Approach

- Variational Inference approximates the posterior distribution $p(d|r)$ with a simpler one $Q(d)$
- The parameters (mean, variance, etc.) of the $Q$ function are chosen to minimize the KL divergence b/w $Q(d)$ and $p(d|r)$:

$$F(\lambda) = \int Q(d) \log \left( \frac{Q(d)}{p(r|d)p(d)} \right) dd$$

where $\lambda$ denotes the set of parameters for the $Q$ function.
Variational Inference

- Good choices of $Q$ result in major simplifications of the original inference problem (finding $P(d_k | r)$)
  - Mean field approximation...
    \[
    Q(d) = \prod_{k=1}^{K} Q_k(d_k)
    \]
  - ...or Gaussianity
    \[
    Q(d) \propto \exp\left[ (d - \mu)^T \Sigma^{-1} (d - \mu) \right]
    \]
Variational Inference

- We can also replace prior distribution $p(d)$ with postulated form, i.e. $p(d)$ needn’t be the true expression.

- **Key**: KL divergence must be in closed form; optimal parameters must be obtainable.

- Previous Results: Obtained LMMSE-based turbo MUD, and IC-based turbo MUD with suitable choices of Q function and priors.
Discrete SISO MUD

Make mean field assumption, and let $b_k \in \{0, 1\}$

\[
P(b) = \prod_{k=1}^{K} \xi_k^{b_k} (1 - \xi_k)^{1-b_k}
\]

\[
p(r|b) = \mathcal{N}(Hb, \sigma^2 I)
\]

\[
Q(b) = \prod_{k=1}^{K} \gamma_k^{b_k} (1 - \gamma_k)^{1-b_k}
\]

where $\xi_k$ is the prior probability of $b_k = 1$, and $\gamma_k$ is the posterior prob. of $b_k = 1$. 

Discrete SISO MUD

- The parameters of the Q function that appear in the KL divergence are \( \{\gamma_1, \ldots, \gamma_K\} \).
- KL divergence has closed form which can be minimized using coordinate descent.
- As the MUD part of a turbo MUD, one form of this receiver is the IC-based turbo MUD of Alexander, et al.
- But M-QAM not easy to handle -- more than one parameter per user!
Gray Mapping for PAM

\[ d = b^{(3)} b^{(2)} b^{(1)} + 2b^{(3)} b^{(2)} + 4b^{(3)} \]

\[ b^{(q)} \in \{-1, +1\} \]

\[(q\text{-th bit in symbol})\]

- In general, for \(2^L\)-ary PAM, we have
  \[ d = \sum_{l=1}^{L} 2^{l-1} \prod_{q=l}^{L} b^{(q)} \]

- This multi-linear transformation enables the extension of the variational approach to \(M\)-ary modulation.
M-QAM Turbo MUD

- For K users’ symbols in a vector:

\[
d = \sum_{l=1}^{L} 2^{l-1} \prod_{q=l}^{L} b^{(q)}\]

where \(\prod_{q} b^{(q)}\) denotes element-wise product.

- Received signal is

\[
r = H d + n = H \sum_{l=1}^{L} 2^{l-1} \prod_{q=l}^{L} b^{(q)} + n\]
M-QAM Turbo MUD

- So we have
  \[ p(r|d) = p(r|b^{(1)}, \ldots, b^{(L)}) \]
  which is known (except for noise variance).

- By the mean-field approximation, we have
  \[ p(d) = \prod_{q=1}^{L} p(b^{(q)}) = \prod_{q=1}^{L} \prod_{k=1}^{K} p(b_{k}^{(q)}) \]
  which is known from the FEC decoder in a turbo MUD.
In variational inference, we want to minimize KL divergence b/w $Q(d)$ and $p(r|d)p(d)$. For tractability, let $Q(d)$ be factorizable i.e. assume all bits are conditionally independent:

$$Q(d) = \prod_{q=1}^{L} Q(b^{(q)}) = \prod_{q=1}^{L} \prod_{k=1}^{K} Q(b^{(q)}_{k})$$

Then approx. marginal distribution can be found w/o summation or integration. Gaussian form also suitable.
M-QAM Turbo MUD

- Assuming binary distributions for $p(b_k^{(q)})$ and $Q(b_k^{(q)})$ we can find the KL divergence as a function of $m_k^{(q)} = E_Q(b_k^{(q)})$ and $\tilde{b}_k^{(q)} = E_p(b_k^{(q)})$.
- Setting the derivative of divergence w.r.t. $m_k^{(q)}$ to zero gives coordinate-descent updates (eq. 20 in the paper):

$$\log \frac{1 + m_k^{(q)}}{1 - m_k^{(q)}} = \log \frac{1 + \tilde{b}_k^{(q)}}{1 - \tilde{b}_k^{(q)}} + \text{IC-like update}$$
M-QAM Turbo MUD

SISO FEC Decoder

\[ m^{(q)} = \text{Bit APP's from MUD} \]
\[ \tilde{b}^{(q)} = \text{Bit priors to MUD} \]
Unknown Noise

- By including $\sigma^2$ as an unknown variable to be estimated with variational inference, and using a “point distribution”, we get a variational EM algorithm.

- Other unknowns e.g. channel can also be incorporated
  - But more unknowns usually means worse performance.
Simulation

- Random spreading, $N = 64$
- 4-PAM
- Rate 1/2 conv. code with generators 10011 and 11101.
Conclusions

- **Most Important:**
  - Traditional view of optimal MUD as too complex led to signal processing approaches e.g. MMSE, int. cancellation, adaptive filters, etc.
  - But these don’t allow obvious link to FEC decoder in turbo MUD.
  - Variational approach keeps probabilistic inference viewpoint of optimal MUD, but uses distributions that are non-exact.
Conclusions

- Variational inference can be used as a unifying concept in detection
  - Different choices of $p(d)$ and $Q(d)$ can lead to various familiar detectors.
  - New viewpoint can lead to improved detectors e.g. variational EM.
- M-QAM (Gray coding) can be handled systematically, without first finding symbol APP’s.