Approximate Message Passing: Applications to Communications Receivers

Phil Schniter

THE OHIO STATE UNIVERSITY

(With support from NSF grant CCF-1018368, NSF grant CCF-1218754, and DARPA/ONR grant N66001-10-1-4090)

TrellisWare, Feb. 2014
The Generalized Linear Model:

- Consider observation $y \in \mathbb{C}^M$ of unknown vector $x \in \mathbb{C}^N$ that is
  - sent through known linear transform $A$, generating hidden $z = Ax$, then
  - observed through a probabilistic measurement channel $p_{y|z}(y|z)$. 
Our goal is to infer $x$ from $y$.

- When $p_x$ and $p_{y|z}$ are both Gaussian, the MMSE/MAP estimator is linear and easy to state in closed-form. The more interesting case is when $p_x$ and/or $p_{y|z}$ are non-Gaussian.

- Equally interesting is when $M \ll N$: Compressive sensing tells us that $K$-sparse $x \in \mathbb{C}^N$ can be accurately recovered from $M \gtrsim O(K \log N/K)$ measurements when $A$ is information-preserving (e.g., satisfies $2K$-RIP).

- There are many applications of estimation under the generalized linear model in engineering, biology, medicine, finance, etc.
Example Applications:

- **Pilot-aided channel estimation** / “compressed channel sensing”
  
  \( x \): sparse channel impulse response (length \( N \))
  
  \( y \): pilot observations (\( M < N \) with sparse channel)
  
  \( A \): built from pilot symbols and other aspects of linear-modulation

- **Imaging** (medical, radar, etc.)
  
  \( x \): spatial-domain image (rasterized)
  
  \( y \): noisy measurements (AWGN, Gaussian, phaseless, etc.)
  
  \( A \): typically Fourier-based (details are application dependent)

- **Binary linear classification and feature selection**
  
  \( x \): prediction vector (\( \perp \) to class-separating hyperplane, sparse)
  
  \( y \): binary experimental outcomes (e.g., \{sick, healthy\})
  
  \( A \): each row contains per-experient features (e.g., age, weight, etc.)
Generalized Approximate Message Passing (GAMP):

- Suppose we are interested in computing the MMSE or MAP estimate of $x$ from $y$ (under known $A$, $p_x$, $p_{y|z}$).

- For general $A$, $p_x$, and $p_{y|z}$, this is difficult... in fact NP hard.

- However, for sufficiently large and dense $A$, and separable $p_x$ and $p_{y|z}$, there is a remarkable new iterative algorithm that gets close: GAMP. S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” arXiv:1010.5141, Oct. 2010.

  - In the large-system limit ($M, N \to \infty$ with fixed $M/N$) when $A$ is drawn iid sub-Gaussian, and $p_x$ and $p_{y|z}$ are separable (i.e., independent r.v.s), GAMP’s performance is characterized by a state evolution whose fixed points, when unique, coincide with the MMSE or MAP optimal estimates.

  - In practice, $A$ is finite sized and structured (e.g., Fourier). Still, for any $A$, the fixed-points of the GAMP iterations correspond to the critical points of the MAP optimization objective, $\max_x \left\{ \ln p_{y|z}(y|Ax) + \ln p_x(x) \right\}$. 
A Revolution in Loopy Belief Propagation:

- The GAMP algorithm can be derived as an approximation of the sum-product (in the MMSE case) or max-product (in the MAP case) loopy BP algorithms.

- The approximation makes use of the central limit theorem and Taylor series approximations that hold in the large-system limit.

- An interesting observation is that, because $A$ is dense, the factor graph is extremely loopy. Loosely speaking, these loops are OK because (for normalized $A$) they get “weaker” as the problem gets larger.

- Note: Rigorous analyses of GAMP are based on the algorithm itself, not on the loopy-BP approximations.

GAMP Heuristics (Sum-Product Case):

1. Message from $y_i$ node to $x_j$ node:

$$p_{i \rightarrow j}(x_j) \propto \int \prod_{r \neq j} p_{y_i | z_i}(y_i | \sum_r a_{ir} x_r) p_{i \leftarrow r}(x_r)$$

$$\approx \mathcal{N} \text{ via CLT}$$

$$\approx \int_{z_i} p_{y_i | z_i}(y_i | z_i) \mathcal{N}(z_i; \hat{z}_i(x_j), \nu_i^2(x_j)) \approx \mathcal{N}$$

To compute $\hat{z}_i(x_j), \nu_i^2(x_j)$, the means and variances of $\{p_{i \leftarrow r}\}_{r \neq j}$ suffice, thus Gaussian message passing!

Remaining problem: we have $2MN$ messages to compute (too many!).

2. Exploiting similarity among the messages $\{p_{i \leftarrow j}\}_{i=1}^{M}$, AMP employs a Taylor-series approximation of their difference whose error vanishes as $M \to \infty$ for dense $A$ (and similar for $\{p_{i \rightarrow j}\}_{j=1}^{N}$ as $N \to \infty$). Finally, need to compute only $O(M+N)$ messages!

The resulting algorithm requires two matrix-vector multiplications per iteration, and converges in typically $\lesssim 25$ iterations.
GAMP Extensions:

- Standard GAMP assumes \textit{known, separable} $p_x$ and $p_{y|z}$.

- However, in practice . . .
  - Densities $p_x$ and $p_{y|z}$ are usually \textit{unknown}.
  - Often, they are also \textit{non-separable} (i.e., elements of $x$ are statistically dependent; same for $y|z$)

- We have developed an EM-based methodology to learn $p_x$ and $p_{y|z}$ \textit{online} and subsequently leverage this information for near-optimal Bayesian inference.
  

- We also have developed a “turbo” methodology that handles probabilistic dependencies among the elements of $x$ and the elements of $y|z$.
  
Some Communications Applications of (EM/turbo) GAMP:

1. **Communications over wideband channels**
   - joint channel-estimation/equalization/decoding

2. **Communications over underwater channels**
   - joint channel-tracking/equalization/decoding

3. **Communications in impulsive noise**
   - joint channel-estimation/equalization/impulse-mitigation/decoding
1. Comms over Wideband Channels:

- At large communication bandwidths, channel impulse responses are \textit{sparse}.
- Below left shows channel taps $\mathbf{x} = [x_0, \ldots, x_{L-1}]$, where
  - $x_n = x(nT)$ for bandwidth $T^{-1} = 256$ MHz,
  - $x(t) = h(t) \ast P_{RC}(t)$, and
  - $h(t)$ is generated randomly using 802.15.4a outdoor NLOS specs.
Simplified Channel Model:

First, let’s simplify things to talk concretely about sparse channels. . .

Consider a discrete-time channel that is

- **block-fading** with block size $N$,
- **frequency-selective** with $L$ taps (where $L < N$),
- **sparse** with $S$ non-zero complex-Gaussian taps (where $0 < S \leq L$),

where *both the channel coefficients and support are unknown* to the receiver.

Important questions:

1. What is the **capacity** of this channel?
2. How can we build a **practical** comm system that operates near this capacity?
Noncoherent Capacity of the Sparse Channel:

For the unknown \( N \)-block-fading, \( L \)-length, \( S \)-sparse channel described earlier, we established that [1]

1. In the high-SNR regime, the ergodic capacity obeys

\[
C_{\text{sparse}}(\text{SNR}) = \frac{N - S}{N} \log(\text{SNR}) + \mathcal{O}(1).
\]

2. To achieve the prelog factor \( R_{\text{sparse}} = \frac{N - S}{N} \), it suffices to use
   
   • pilot-aided OFDM (with \( N \) subcarriers, of which \( S \) are pilots)
   • with joint channel estimation and data decoding.

Key points:

• The effect of unknown channel support manifests only in the \( \mathcal{O}(1) \) offset.
• [1] uses constructive proofs, but the decoder proposed there is not practical.

We now propose a communication scheme that...

- is practical, with decode complexity $\mathcal{O}(N \log_2 N + N|\mathcal{S}|)$ per block,
- (empirically) achieves the optimal prelog factor $R_{\text{sparse}} = \frac{N-S}{N}$,
- significantly outperforms "compressed channel sensing" (CCS) schemes.

Our scheme uses...

- a conventional transmitter: pilot-aided BICM OFDM,
- a novel receiver: based on GAMP.
To jointly infer all random variables, we perform loopy-BP via the sum-product algorithm, using GAMP approximations in the GAMP sub-graph.
Numerical Results — Perfectly Sparse Channel:

Transmitter:

- LDPC codewords with length $\sim 10000$ bits.
- $2^M$-QAM with $2^M \in \{4, 16, 64, 256\}$ and multi-level Gray mapping.
- OFDM with $N = 1024$ subcarriers.
- $P$ pilot subcarriers and/or $T$ training MSBs.

Channel:

- Length $L = 256 = N/4$.
- Sparsity $S = 64 = L/4$.

Reference Schemes:

- Pilot-aided LASSO was implemented using SPGL1 with genie-aided tuning.
- Pilot-aided LMMSE, support-aware MMSE, and info-bit+support-aware MMSE channel estimates were also tested.
BER & Outage vs SNR (with $P=L$ pilots and $T=0$ MSBs):

Key points:

- GAMP outperforms both LASSO and the support genie (SG).
- GAMP performs nearly as well as the info-bit+support-aware genie (BSG).
- With $P = L$, all approaches yield prelog factor $R = \frac{N-L}{N} = \frac{3}{4}$, which falls short of the optimal $R_{\text{sparse}} = \frac{N-S}{N} = \frac{15}{16}$.
BER & Outage vs SNR (with $P=0$ pilots & $T=SM$ training MSBs):

\[
\log_{10}(\text{BER}) \text{ (256 QAM, 3.75 bpcu, 20dB SNR)}
\]

Key points:

- GAMP favors $P=0$ pilot subcarriers and $T=SM$ training MSBs.
  - Precisely the necc/suff redundancy of the capacity-maximizing system!

- GAMP achieves the sparse-channel’s capacity-prelog factor, $R_{\text{sparse}} = \frac{N-S}{N}$.
In reality, channel taps are not perfectly sparse, nor i.i.d:

- For example, consider channel taps $\mathbf{x} = [x_0, \ldots, x_{L-1}]$, where
  - $x_n = x(nT)$ for bandwidth $T^{-1} = 256$ MHz,
  - $x(t) = h(t) \ast p_{RC}(t)$, and
  - $h(t)$ is generated randomly using 802.15.4a outdoor NLOS specs.

- The tap distribution varies as the lag increases, becoming more heavy-tailed.
- The big taps are clustered together in lag, as are the small ones.
Proposed channel model:

- Saleh-Valenzuela (e.g., 802.15.4a) models are accurate but difficult to exploit in receiver design.

- We propose a structured-sparse channel model based on a 2-state Gaussian Mixture model with discrete-Markov-chain structure on the state:

\[
p(x_j | d_j) = \begin{cases} 
\mathcal{CN}(x_j; 0, \mu_j^0) & \text{if } d_j = 0 \quad \text{"small"} \\
\mathcal{CN}(x_j; 0, \mu_j^1) & \text{if } d_j = 1 \quad \text{"big"}
\end{cases}
\]

\[
\Pr\{d_{j+1} = 1\} = p_{j}^{10} \Pr\{d_j = 0\} + (1 - p_{j}^{01}) \Pr\{d_j = 1\}
\]

- Our model is parameterized by the lag-dependent quantities:

  \{\mu_j^1\} : big-state power-delay profile
  
  \{\mu_j^0\} : small-state power-delay profile
  
  \{p^{01}_j\} : big-to-small transition probabilities
  
  \{p^{10}_j\} : small-to-big transition probabilities

- Can learn these statistical params from observed realizations via the EM alg.
Factor graph for pilot-aided BICM-OFDM:

To jointly infer all random variables, we perform loopy-BP via the sum-product algorithm, using GAMP approximations in the GAMP sub-graph.
Numerical results:

Transmitter:

- OFDM with \( N = 1024 \) subcarriers.
- 16-QAM with multi-level Gray mapping
- LDPC codewords with length \( \sim 10000 \) yielding spectral efficiency of 2 bpcu.
- \( P \) “pilot subcarriers” and \( T \) “training MSBs.”

Channel:

- 802.15.4a outdoor-NLOS (not our Gaussian-mixture model!)
- Length \( L = 256 = N/4 \).

Reference Channel Estimation / Equalization Schemes:

- soft-input soft-output (SISO) versions of LMMSE and LASSO.
- perfect-CI genie.
BER versus $E_b/N_o$ for $P = 224$ pilots and $T = 0$ training MSBs:

Our scheme shows 4dB improvement over (turbo) LASSO.
Our scheme only 0.5dB from perfect-CSI genie!
BER versus $E_b/N_o$ for $P = 0$ pilots and $T = 448$ training MSBs:

Use of training MSBs gives 1dB improvement over use of pilot subcarriers!
2. Communications over Underwater Channels:

- SPACE-08 Underwater Experiment 2920156F038_C0_S6
- Time-varying channel response estimated using WHOI M-sequence:

- The channel is nearly over-spread: $f_d T_s L = 20 \times \frac{1}{10000} \times 400 = 0.8$ !
- Can’t afford to ignore structure of temporal variations!
• Channel taps are modeled as independent Bernoulli-Gaussian processes:
  – each tap’s amplitude follows a temporal Gauss-Markov chain
  – each tap’s on/off state follows a temporal discrete-Markov chain
Performance versus SNR:

Settings:
- experimentally measured underwater channel
- 16-QAM
- 1024 total tones
- 0 pilot tones
- 256 training MSBs
- LDPC length 10k
- LDPC rate 0.5

Exploiting the persistence in channel support and channel amplitudes was critical in this difficult underwater application.
3. Communications in Impulsive Noise:

- In many wireless and power-line communication systems, the (time-domain) noise is not Gaussian but *impulsive*.

- The marginal noise statistics are well captured by a 2-state Gaussian mixture (i.e., Middleton class-A) model.

- Noise burstiness is well captured by a discrete Markov chain on the noise state.
Factor Graph for pilot-aided BICM-OFDM:
Numerical Results — Uncoded Case:

Settings:
- 5 channel taps
- GM noise
- 256 total tones
- 15 pilot tones
- 80 null tones
- 4-QAM

Proposed “joint channel/impulsive-noise/symbol” estimation (JCIS) scheme gives \( \sim 15 \) dB gain over previous state-of-the-art and performs within 1 dB of MFB!
Numerical Results — Coded Case:

Settings:
- 10 channel taps
- GM noise
- 1024 total tones
- 150 pilot tones
- 0 null tones
- 16-QAM
- LDPC
- Rate 0.5
- Length 60k

Proposed “joint channel/impulsive-noise/symbol/bit” estimation (JCISB) scheme gives $\sim15$ dB gain over traditional DFT-based receiver!
Conclusions:

- Inference in the generalized linear model yields an important but challenging class of problems.

- The generalized approximate message passing (GAMP) is an important new tool for solving such problems (under sufficiently large and dense transforms).

- Problems of this form manifest in BICM-OFDM comms receivers, where one wants to optimally decode bits in the presence of unknown channels, symbols, and noise.

- Often, the channel and noise processes have interesting statistical structures (e.g., sparsity, clustering, time-variation) and decoding performance can be dramatically improved when these structures are properly exploited.

- For such problems, GAMP can be “plugged into” the standard “turbo” receiver architecture to yield near-optimal performance with manageable complexity.