Belief Propagation Is Asymptotically Equivalent to MAP Estimation for Sparse Linear Systems

Chih-Chun Wang
School of ECE
Purdue University

Dongning Guo
Dept. of EECS
Northwestern University

Allerton Conference
September 28, 2006

Outline

Linear systems with Gaussian noise: $\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{X} + \boldsymbol{N}$.

Previous results & motivation.

MAP and BP.

Main results.

Sketch of proof.

Conclusions.

The Linear System

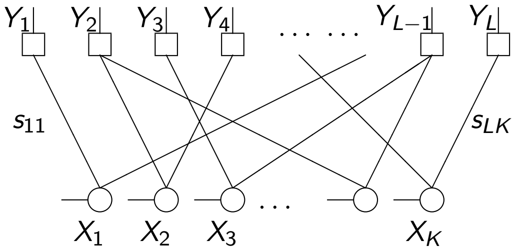
$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{X} + \boldsymbol{N}$$
$$= \sum_{k=1}^{K} \boldsymbol{S}_k A_k X_k + \boldsymbol{N}$$

- I.i.d. input symbols: $X_k \sim P_X$.
- I.i.d. amplitudes: $A_k \sim P_A$.
- $\boldsymbol{S}_{L\times K} = [A_1\boldsymbol{S}_1, \cdots, A_K\boldsymbol{S}_K]$, where $\boldsymbol{S}_k = [S_{1k}, S_{2k}, \ldots, S_{Lk}]^\top$

$$Y_l = \sum_{k=1}^{K} S_{lk} A_k X_k + N_l, \qquad l = 1, \ldots, L.$$

Applications: MIMO, CDMA, OFDM, statistics, optimization ...

Factor Graph



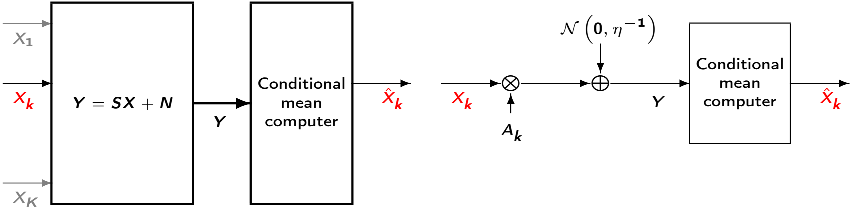
Given $(\boldsymbol{Y}, \boldsymbol{S}) = (\boldsymbol{y}, \boldsymbol{s})$, compute the maximum *a posteriori* (MAP) distribution (exact statistical inference on the graph): $p_{X_k|\boldsymbol{Y},\boldsymbol{S}}(\cdot|\boldsymbol{y}, \boldsymbol{s})$.

MMSE

The MMSE per dimension of estimating $\boldsymbol{Z}_{K\times 1}$ given $\boldsymbol{W}$:

$$\mathcal{E}\left(\boldsymbol{Z} \,|\, \boldsymbol{W}\right) = \frac{1}{K}\mathsf{E}\left\{\left\|\boldsymbol{Z} - \mathsf{E}\left\{\boldsymbol{Z} \,|\, \boldsymbol{W}\right\}\right\|^2\right\}$$

Previous Non-rigorous Result Via Statistical Mechanics



Claim (Guo & Verdú, IT-05)

Previous Result: Mutual Information

Claim (Guo & Verdú, IT '05)

Justification of Tanaka's Formula by Montanari-Tse

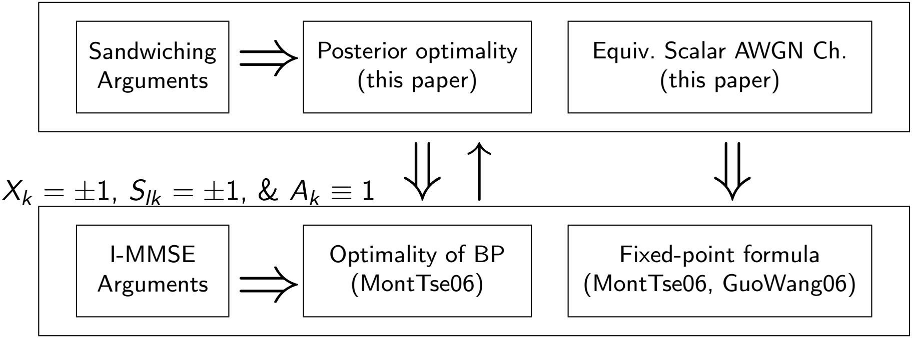
Assumptions:

- Binary input symbols;
- Equal power ($A_k \equiv 1$);
- Sparse matrix $\boldsymbol{S}$;
- $K = \beta L \to \infty$ first, then average node degree $\bar{\Gamma} \to \infty$.

Theorem (Entropy optimality, Montanari and Tse, '06)

Overview of Previous and New Results

Arbitrary P_X, P_A, P_S



Useful tools:

Sparse Linear Systems

The Doubly-Poisson Ensemble:
$S_{lk} \sim aP_S + (1-a)\delta(0)$ where $a = \bar{\Gamma}/K$.

The *large-system limit*: $K, L \to \infty$ with $K/L \to \beta$.

The *large-sparse-system limit*:

$$\lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}^2 = o(K) \to \infty}}$$

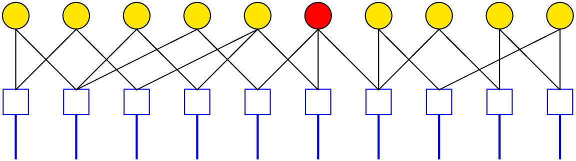
— asymptotically free of short cycles.

β is s.t. the solution is unique for

$$\eta^{-1} = 1 + \beta\,\mathcal{E}\left(AX \,|\, \sqrt{\eta}\,AX + N, A\right)$$

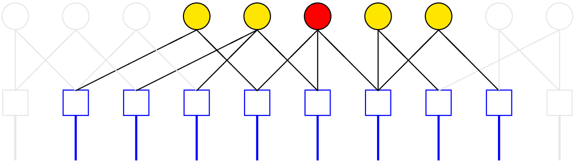
Can be generalized to many other ensembles (e.g., LDPC-type).

MAP

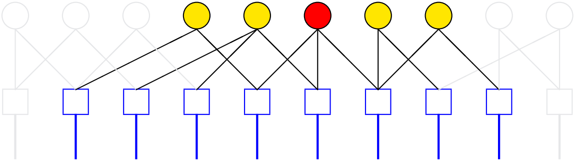


Belief Propagation

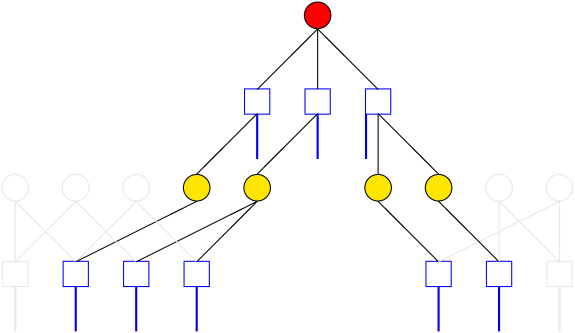
1 iteration



Belief Propagation: Supporting Tree Representation



1 iteration



Iterative Update Formulas

Consider binary X_k first. Define $LLR(X|Z) = \log \frac{\mathsf{P}\{X=+1|Z\}}{\mathsf{P}\{X=-1|Z\}}$.

$$U_{l\to k}^{(t)} = LLR\Big(X_k \,\Big|\, Y_l, \Big\{LLR(X_i) = V_{i\to l}^{(t-1)}\Big\}_{i\neq k}\Big)$$

$$V_{k\to l}^{(t)} = V_k^{(0)} + \sum_{j\neq l} U_{j\to k}^{(t)}.$$

In general, $LLR(X=x|Z) = \log \frac{\mathsf{P}\{X=x|Z\}}{\mathsf{P}\{X=x_0|Z\}}$ (equiv. to $\mathsf{P}\{X=\cdot|Z\}$).

$$U_{l\to k}^{(t)}(x) = LLR\Big(X_k = x \,\Big|\, Y_l, \Big\{LLR(X_i=\cdot) = V_{i\to l}^{(t-1)}(\cdot)\Big\}_{i\neq k}\Big)$$

$$V_{k\to l}^{(t)}(x) = V_k^{(0)}(x) + \sum_{j\neq l} U_{j\to k}^{(t)}(x).$$

Can be generalized to continuous X_k. In general, one may pass any sufficient statistics of the extrinsic information, e.g., LLR, posterior.

If cycle-free, BP computes the exact posterior.

Posterior Optimality of Belief Propagation

Theorem (Symbol-wise optimality)

Equivalent Scalar Gaussian Channels

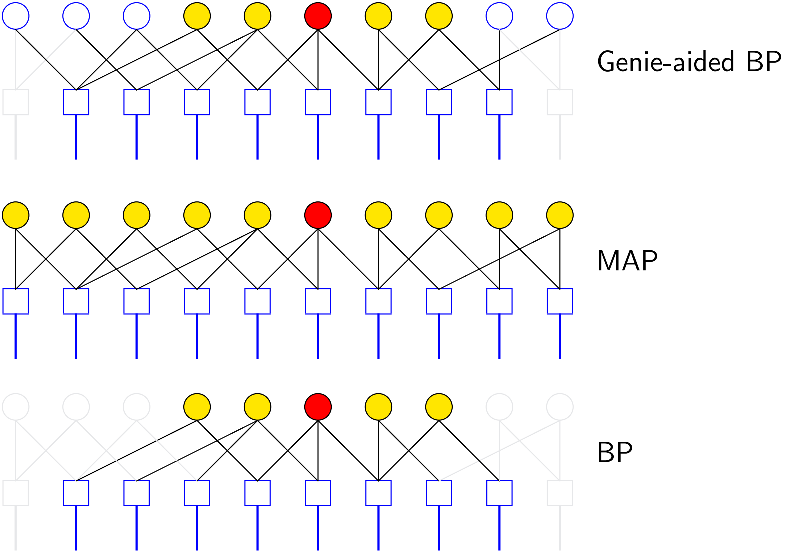
Let $p_{X|Y,A}(\cdot|\cdot,\cdot)$ be the posterior of the scalar Gaussian channel

$$Y = AX + \frac{1}{\sqrt{\eta}}N$$

where $X \sim P_X$, $N \sim \mathcal{N}(0,1)$, and η is determined by

Theorem (Symbol-wise equivalent channel)

Sandwiching Proof



Uniqueness of Fixed Point

Density evolution formula

$$\eta_t^{-1} = 1 + \beta\,\mathcal{E}\left(AX \,|\, \sqrt{\eta_{t-1}}AX + N, A\right)$$

At least one fixed point.

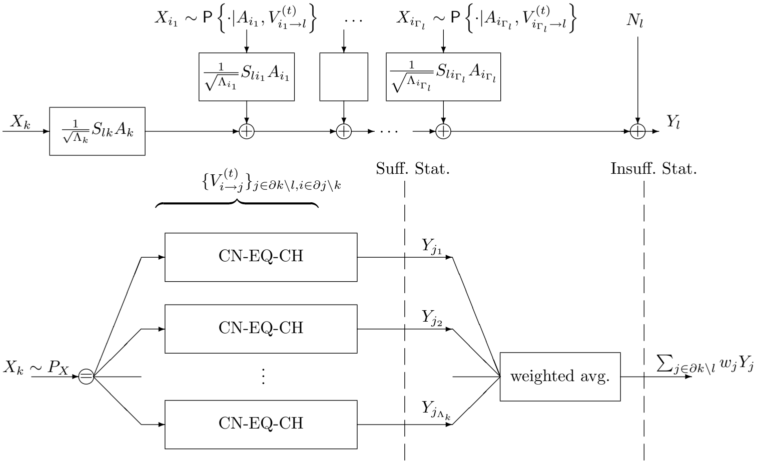
We have assumed small enough β to guarantee uniqueness.

The threshold depends on P_X and P_A.

Relation to I-MMSE

Theorem (Guo-Shamai-Verdú)

Equivalent Channel Via Interference Cancellation



Corollaries

Theorem (The symbol MMSE)

Generalization to Broader Ensembles

Symbol degree Λ_k, chip degree Γ_l.

Chip-semi-regularity:

$$\lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}^2=o(K)\to\infty}} \mathsf{P}\left\{|\Gamma_l - \bar{\Gamma}| > \epsilon\bar{\Gamma}\right\} = 0, \quad \forall\,\epsilon > 0, \forall\, l$$

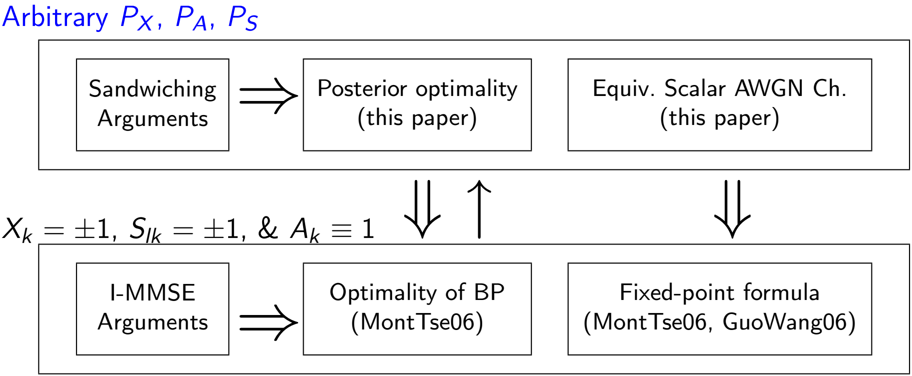
Symbol balanced condition: $\exists\, a < \infty$ such that

$$\lim_{K\to\infty} \frac{1}{K}\left|\{k \mid \Lambda_k > a\bar{\Gamma}\}\right| = 0.$$

Examples

- The graph-based ensembles from regular/irregular LDPC codes;
- The symbol-irregular, chip-Poisson ensembles (Montanari-Tse).

Conclusion


Arbitrary P_X, P_A, P_S
Sandwiching Arguments
Posterior optimality (this paper)
Equiv. Scalar AWGN Ch. (this paper)
$X_k = \pm 1$, $S_{lk} = \pm 1$, & $A_k \equiv 1$
I-MMSE Arguments
Optimality of BP (MontTse06)
Fixed-point formula (MontTse06, GuoWang06)