Comments on Why Generalized BP Serves So Remarkably in 2-D Channels

Ori Shental[1] Noam Shental[2] Shlomo Shamai (Shitz)[3] Ido Kanter[4] Anthony J. Weiss[5] Yair Weiss[6]

***Abstract*— Generalized belief propagation (GBP) algorithm has been shown recently to infer the a-posteriori probabilities of finite-state input two-dimensional (2-D) Gaussian channels with memory in a practically accurate manner, thus enabling near-optimal estimation of the transmitted symbols and the Shannon-theoretic information rates. In this note, a rationalization of this excellent performance of GBP is addressed.**

A $N \times N$ 2-D finite-state input channel with memory is of the form

$$\mathsf{y}_{k,l} = \mathsf{x}_{k,l} + \mathsf{v}_{k,l} + \sum_{(i,j)\in\langle k,l\rangle} \alpha_{i,j}\mathsf{x}_{i,j} \quad \forall k,l = 1,\ldots,N,$$

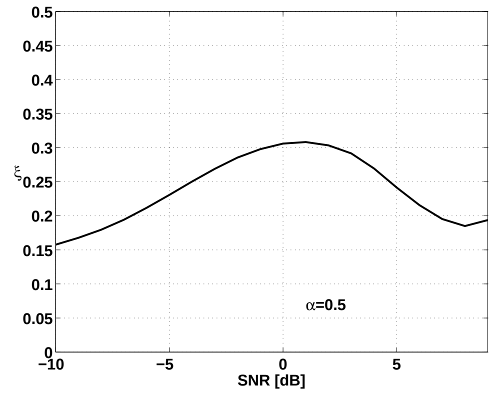
where $\mathsf{y}_{k,l}$, the channel's output observation at symbol $(k,l) \in \mathbb{Z}^2$, is the sum of the finite alphabet input symbol $\mathsf{x}_{k,l}$ and two additional terms. The first term $\mathsf{v}_{k,l}$ represents the common ambient AWGN, while the second term is the scaled interference caused by adjacent symbols to (k,l), denoted by $\langle k,l\rangle$. The parameter $\alpha_{i,j}$ ($|\alpha_{i,j}| \leq 1$) controls the interference attenuation.

In recent papers, Shental *et al.* have shown that the fully tractable generalized belief propagation (GBP) algorithm, applied for 2-D dispersive channels, yields practically accurate a-posteriori probabilities (APP) [1]. The marginal probabilities are then used for symbol detection [2], and also for estimating the information rate, via the symbol-wise Guo-Shamai-Verdú theorem [1]. The information rate was also estimated in a complementary approach using the joint APP, via the Shannon-McMillan-Breiman theorem, and the statistical mechanics notion of free energy [3].

In GBP messages are sent between regions (clusters) of nodes, as opposed to loopy belief propagation (LBP) where messages are passed along the graph's edges. The region graph approach to GBP begins by defining 'basic' regions, which completely cover the graph and include all linked nodes. Let $\mathcal{R}$ be this set of regions, their intersections, the intersections of the intersection, and so on. Based on the set $\mathcal{R}$, a region graph of the graphical model can be composed, along which messages are passed in an analogous way to LBP. A smart choice of the basic regions will be such that it encompasses all nodes along the shortest loops. Since inference within a basic region is performed exactly, we may avoid the short loops, which probably caused LBP not to converge.

Since the graphical models of the examined 2-D channels contain interactions between nearest neighbors and next nearest neighbors (NNN), a natural choice of overlapping regions in this case is a sliding 3×3 square of nodes. Our empirical study shows that the GBP message passing algorithm yields remarkable inference results in 2-D channels, even for harsh interference conditions (*i.e.*, large α values). As rigorous analysis of the workings of GBP is a challenging open problem, we suggest a qualitative explanation for its remarkable performance.

[1]Center for Magnetic Recording Research, University of California, San Diego, USA, oshental@ucsd.edu. [2]Physics of Complex Systems, Weizmann Institute of Science, Israel, fenoam@wisemail.weizmann.ac.il. [3]Electrical Engineering, Technion, Israel, sshlomo@ee.technion.ac.il. [4]Physics, Bar-Ilan University, Israel, kanter@mail.biu.ac.il. [5]Electrical Engineering-Systems, Tel Aviv University, Israel, ajw@eng.tau.ac.il. [6]School of Computer Science and Engineering, Hebrew University of Jerusalem, Israel, yweiss@cs.huji.ac.il.



We conjecture that the reason for the success of GBP stems from the 'local' nature of the system (but not too 'local', hence message passing is inevitable). More specifically, we claim that the correlation length, ξ, which measures the typical range over which symbols are correlated, is of the order of the chosen (*e.g.*, 3×3) basic region over the whole relevant range of SNR and α values. According to statistical mechanics' transfer matrix theory, the correlation length is calculated as $\xi = 1/\log(\lambda_0/\lambda_1)$, where λ_0 and λ_1 are the largest and second largest eigenvalues of the transfer matrix, respectively. The transfer matrix $\mathbf{M}$, in the case of 2-D channels with NNN is a $2^9\times2^9$ matrix, where M_{ij} is proportional to the probability that a 3×3 block is in a certain state i and its adjacent block is in state j.

Thus, one can take each two adjacent 3×3 blocks in the 2-D channel's corresponding graph and calculate the correlation length as if they were replicated indefinitely. The above figure presents the (mean) correlation length (in 3×3 block units) over all pairs of adjacent blocks, as a function of SNR (with $\alpha = 0.5$ and $\langle k,l\rangle$ describing the hexagonal Wyner cellular model). One observes that this value is lower than 1 for all SNR. The reason for the low correlation length, even at high SNR (and α) values, is the positive effect of external fields operating on each node, originated from the observations.

References

[1] O. Shental, N. Shental, S. Shamai (Shitz), I. Kanter, A. J. Weiss, and Y. Weiss, "Finite-state input two-dimensional Gaussian channels with memory: Estimation and information via graphical models and statistical mechanics," submitted for publication.

[2] O. Shental, N. Shental, A. J. Weiss, and Y. Weiss, "Generalized belief propagation receiver for near-optimal detection of two-dimensional channels with memory," in *Proc. IEEE Information Theory Workshop (ITW)*, San Antonio, Texas, USA, Oct. 2004.

[3] O. Shental, N. Shental, and S. Shamai (Shitz), "On the achievable information rates of two-dimensional channels with memory," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Adelaide, Australia, Sept. 2005.