Progress Report
4-15 June 2012

Amir Hesam Salavati
E-mail: hesam.salavati@epfl.ch

Supervisor: Prof. Amin Shokrollahi
E-mail: amin.shokrollahi@epfl.ch

Algorithmics Laboratory (ALGO)
Ecole Polytechnique Federale de Lausanne (EPFL)

September 7, 2012

1 Summary

In the past two weeks, I was busy preparing for the upcoming ISIT to make the presentation and the additional technical materials to put on Arxiv, both of which can be found separately on the ERC log.

I also prepared the poster for the upcoming North American School on Information Theory. The poster, which can also be found on the ERC log, focuses mainly on our recent progress in multi-level neural networks, which is going to be presented at ISIT 2012, and encourages new applications of coding theory in neurosceience.

Mr. Karbasi and I also had a meeting with Prof. Pfister from UT Austin to discuss his recent work on spatially coupled codes. This was triggered as a result of his talk in SuRI, where we witnessed a lot of similarity between their model and out clustered neural networks which we have submitted to NIPS 2012. to our benefit, they have proposed a very nice framework to analyze these models which we can easily adapt to our neural approach. I will describe their method in more details in the following. We will use this framework to expand our research to *spatially coupled neural networks* which interestingly seems very plausible according to neurophysiological data.

2 Summary of the new framework to analyze the spatially coupled codes

In [1] the authors propose a very interesting approach to prove that spatially coupled codes can achieve MAP decoding threshold over the Binary Erasure Channel (BEC). Their approach is based on introducing a *potential* which then is shown to be decreasing in each iteration of the decoding method as long as the erasure probability of the channel is less than the MAP threshold and will not decrease otherwise. Therefore, the authors prove that the MAP threshold is achievable with spatially coupled codes (note that for such codes, the boundaries are fixed to some known values). Although this might not be the first time that such an achievement is shown (as another proof can be found in [2]), the method used here is much simpler and more elegant. Furthermore, using the notion of potential, the authors provide a nice connection between MAP and BP threshold. More specifically, they show that the BP threshold corresponds to the maximum erasure probability for which the derivative of the potential is negative everywhere. The MAP threshold on the other hand corresponds to the maximum erasure probability for which the potential is positive everywhere.

2.1 Uncoupled systems

First, let us conisder the uncoupled system which is usually described with a recursion like

$$x^{(\ell+1)} = f(g(x^{(\ell)}); \epsilon) \tag{1}$$

In this equation

- "$f : [0, 1] \times [0, 1] \to [0, 1]$ is *strictly increasing in both arguments* x, ϵ."
- "$g : [0, 1] \to [0, 1]$ has positive derivative for $x \in (0, 1)$, i.e. $g'(x) > 0$.
- "f and g have continuous second derivatives on $[0, 1]$ w.r.t. all arguments."

The potential function of this system is defined as

$$U(x;\epsilon) = \int_0^x (z - f(g(z);\epsilon))g'(z)dz = xg(x) - G(x)F(g(x);\epsilon) \tag{2}$$

where $F(x;\epsilon) = \int_0^x f(z;\epsilon)dz$ and $G(x) = \int_0^x g(z)dz$. The potential function has the following properties:

- $U(x;\epsilon)$ is *strictly decreasing* in ϵ $\in (0,1]$.
- The fixed point x of the recursion (1) is equal to the stationary point of the potential function.

[**important**]The *single system threshold* $\epsilon_s^\star$ is defined to be the maximum erasure probability for which the probability of error in an *uncoupled* system converges to zero. This threshold corresponds to the maximum erasure probability for which the derivative of the potential function is positive everywhere

$$\epsilon_s^\star = \sup\{\epsilon \in [0,1] | U'(x; epsilon) > 0 \forall x \in (0,1]\} \tag{3}$$

2.2 Coupled systems

Coupled systems are usually described with a recursion as well. Often, this recursion is multi-dimensional. However, there are various *averaging* techniques to reduce it to a single dimensional recursion. For instance, the authors use equation (4), given below

$$x_i^{(\ell+1)} = \frac{1}{w}\sum_{k=0}^{w-1} f\left(\frac{1}{w}\sum_{j=0}^{w-1} g(x_{i+j-k}^{(\ell)}); \epsilon_{i-k}\right) \tag{4}$$

In this equation

- $x_i^{(\ell+1)}$ is the average error probability at round $\ell+1$ for a variable node within cluster (position) i.
- $2L+1$ is the total number of systems coupled together.
- w is the number of systems coupled with cluster i.
- f and g are similar to what has been defined previously.

The authors show that equation (4) can be rewritten in matrix format

$$\mathbf{x}^{(\ell+1)} = \mathbf{A}_2^\top \mathbf{f}\left(\mathbf{A}_2 \mathbf{g}(\mathbf{x}^{(\ell)}); \epsilon\right) \tag{5}$$

where $\mathbf{x}^{(\ell)} = [..., x_i^{(\ell)}, ..]$ is the vector of probabilities, $\mathbf{A}_2 \in \mathbb{R}^{(2L+1)\times(2L+w)}$ is the coupling matrix (given below), $\mathbf{f}$ and $\mathbf{g}$ are point-wise vector functions i.e. $\mathbf{f}_i(x) = f(x_i)$.

$$A_2 = \frac{1}{w}\begin{bmatrix} \overbrace{1 \quad 1 \quad \cdots \quad 1}^{w} & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 \quad 1 \quad \cdots \quad 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & & \\ 0 \quad 0 \quad \cdots \quad 0 & 0 & 1 & 1 & \cdots & 1 & 0 \\ 0 \quad 0 \quad \cdots \quad 0 & 0 & 0 & 1 & \cdots & 1 & 1 \end{bmatrix}$$

Note that in contrast to [2], the recursion for the coupled system is defined as the average of f functions, instead of being defined as the f function of the average (which is the more natural definition). However, this seems to has no effect on the proof or threshold saturation.

So far, we were talking about a *two-sided* coupled system, i.e. the one in which the boundaries are not fixed. If we get fix the boundaries, we will get a *one-sided* system. The recursion would be essentially very similar to (5), but with a different coupling matrix. For a one-sided system, one should enforce the boundary conditions after each iteration.

In [2], the authors prove that the recursion (5) for a two-sided coupled system and the corresponding recursion for the one-sided setup are *component-wise* decreasing in each iteration and will converge to a well-defined fixed point (not necessarily zero of course). See Lemma 14 in [2] for more details.

The authors build upon the above result and use the notion of potential to show the achievability of MAP threshold. The potential function for the coupled system is defined as

$$U(\mathbf{x};\epsilon) = \int_C (\mathbf{z} - \mathbf{A}^\top \mathbf{f}(\mathbf{A}\mathbf{g}(\mathbf{z});\epsilon))\mathbf{g}'(\mathbf{z})d\mathbf{z} = \mathbf{g}(\mathbf{x})^\top \mathbf{x} - G(\mathbf{x})F(\mathbf{g}(\mathbf{x});\epsilon) \tag{6}$$

where $\mathbf{g}'(\mathbf{x}) = diag([g'(x_i)]$, $G(\mathbf{x}) = \int_C \mathbf{g}(\mathbf{z}).d\mathbf{z} = \sum_i G(x_i)$, $F(\mathbf{x}) = \int_C \mathbf{f}(\mathbf{z}).d\mathbf{z} = \sum_i F(x_i)$ and C is the integration path.

The main contribution of [1] is then mentioned in **Lemma 4** in which the authors show that if $\mathbf{x} \in [0,1]^n$ is a ***non-decreasing*** vector generated by *averaging* $\mathbf{z} \in [0,1]^n$ over a sliding window of size w, then the operation $\mathbf{Sx}$ yields: $U(\mathbf{Sx};\epsilon) - U(\mathbf{x};\epsilon) = -U(x_{i_0};\epsilon)$. Here, S is the shift operation such that $[\mathbf{Sx}]_1 = 0$ and $[\mathbf{Sx}]_i = \mathbf{x}_{i-1}$. Therefore, one iteration of the decoding procedure reduces the potential *if* $U(x_{i_0};\epsilon)$ is positive. Which happens for all $\epsilon < \epsilon^\star$. Furthermore, the authors also show that for all this to happen, the window size w should be large enough. More specifically, in **Theorem 1**, they prove that we must have $w > K_{f,g}/\Delta E(\epsilon)$, where $K_{f,g}$ is an upper bound on $\|U\text{"}(\mathbf{x};\epsilon)\|_\infty$.

3 Future works

In the upcoming weeks, we are going to apply the method introduced in [1] to our neural model in order to expand the model to coupled neural networks. This hopefully will not only improve the performance but also provide us with a nice tool to analytically investigate the performance of our network.

References

[1] A. Yedla, Y. Y. Jian, P. S. Nguyen, H. D. Pfister , *A simple proof of threshold saturation for coupled scalar recursions*, Proc. ISTC 2012.

[2] S. Kudekar, T. J. Richardson, R. L. Urbanke, *Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC*, IEEE Trans. Inform. Theory, vol. 57, no. 2, pp. 803-834, 2011.