Progress Report
15-31 July 2011

Raj K. Kumar, Amir Hesam Salavati
E-mail: raj.kumar@epfl.ch, hesam.salavati@epfl.ch

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

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

September 20, 2011

1 Summary

In the past week, we read some papers about higher order neural networks since they have similar properties to our double layer constraint enforcing network. On the plus side, there are learning rules that enable us to find the hidden structure in a set of given training patterns. However, the down side is that the complexity of such learning meachanisms is prohibitive.

The other subject we read some papers about was matrix recovery and completion methods. In this problem, we are interested in infering a low rank matrix from a set of available linear measurements. In a sense, this problem is closely related to the problem of learning the constraint matrix in our neural compressed sensing problem. However, the methods given in most of the papers in this field is not realizable by simple neurons. The papers mentioned in the following give us some required conditions on the desired matrix and the measurements in order to ensure recoverability.

On a different topic, we prepared the powerpoint for the upcoming ISIT conference. We also had a pizza talk on July 29, 2011.

2 Higher Order Neural Networks

In [1], the authors discuss various applications of higher-order neural networks. A higher-order neural network is a network composed of neurons that rely on statistics of order two or more in computing the connection weights and updating neural states. Higher-order neurons are equivalents of neural networks with hidden layers. More specifically, a higher-order neuron is a neuron whose outputs not only depends on the state of its neighbors but also on their correlation. More specifically, the state of neuron i, denoted by s_i is given by:

$$s_i = f(\theta + \sum_j w_{ij} s_j + \sum_{j,k} w_{ijk} s_j s_k + \dots) \tag{1}$$

where $f(.)$ is non-linear function.

Such neurons are able to solve problems that are not solvable with a single layer ordinary neurons. In [1] two learning rules have been proposed for higher-order neurons. One possible learning rule for a network of higher-order neurons is to extend the perceptron rule. An example is given below for second order neurons:

$$\Delta w_{ijk} = [o_i - s_i] s_j s_k \tag{2}$$

where o_i is the desired output and s_i is the actual output of node i.

Another learning rule comes from extending the outer-product rule, also known as the Hebbian learning rule, to higher-order neurons. An example for the second order neurons is given below:

$$w_{ijk} = \sum_{\mu=1}^{M} [x_i^\mu - y_i][x_j^\mu - y_j][x_k^\mu - y_k] \tag{3}$$

where

$$y_i = \frac{1}{M} \sum_{\mu=1}^{M} x_i^\mu$$

Here, M is the number of patterns in the training set and x_i^μ is the i^{th} bit of pattern μ.

One disadvantage of the higher-order neural network is the complexity of learning rule in the sense that we need to keep track of $O(n^k)$ weights, where k is the highest degree of correlation we are interested in.

3 Matrix Recovery and Completion

In [2] the authors propose a novel algorithm to recover a low matrix X from its measurements $\mathcal{A}(X) = \sum_{i=1}^{d} G_i X G_i^T$, where G_i's are measurement matrices. The authors come up with necessary and sufficient conditions that ensure recoverability of X and show that if $\mathcal{A}(X)$ is a rank-expander, these conditions are satisfied. Furthermore, they show that if G_i's are i.i.d. random Gaussian matrices, they yield rank expanders with probability close to 1.

More specifically, a rank expander is defined as follows:

Definition 1. *Let $\mathcal{A} : \mathcal{R}^{n\times n} \to \mathcal{R}^{m\times m}$ be a linear operator with $m < n$. $\mathcal{A}$ is an unbalanced (ϵ, d, r_0, n)-rank expander (for $d > 0$, $0 \leq epsilon < 1$, $1 \leq r_0 \leq n$), if it satisfies the following conditions:*

1. *For every $n \times n$ Hermitian matrix X, $\mathcal{A}(X)$ is an $m \times m$ Hermitian matrix.*
2. *For every $n \times n$ Hermitian matrix X, $rank(\mathcal{A}(X)) \leq d.rank(X)$.*
3. *For all* orthogonal projections *P with $rank(P) = r \leq r_0$, we have $(1-\epsilon)rd \leq rank(\mathcal{A}(P)) < rd$.*

Now the desirable property is to have $\mathcal{P}\{rank(\mathcal{A}(X)) = \min(d \times rank(X), m)\} = 1$ *for every sufficiently low rank matrix X.* The following theorem shows that rank expanders exist:

Theorem 1. *Existence of Rank Expanders: for* any *$0 < \epsilon < 1$ there are constants C_1 and C_2 so that for any n and $r_0 \leq n$, whenever $m = \sqrt{C_1 C_2 n r_0}$ and $d = \sqrt{\frac{C_2 n}{C_1 r_0}}$ and $\{G_i\}_{i=1}^{d}$'s are independent $m \times n$ randomly $\mathcal{N}(0,1)$ Gaussian matrices, the operator $\mathcal{A}(X) = \sum_{i=1}^{d} G_i X G_i^T$ is an (ϵ, d, r_0, n)-rank expander with probability at least $1 - e^{-\Omega(n)}$.*

The following lemma gives the necessary and sufficient condition to ensure recoverability of a matrix X from its linear measurements $\mathcal{A}(X)$. The next lemma after the following one shows that rank expanders satisfy the required conditions.

Lemma 2. Any Positive Semi-Definite *(PSD) matrix X of rank at most r is the unique PSD inverse image of $\mathcal{A}(X)$,* if and only if *every non-zero Hermitian matrix W in the null space of $\mathcal{A}$ (i.e. $\mathcal{A}(W) = 0$) has at least $r+1$ negative eigenvalues.*

Lemma 3. *If $\mathcal{A}(X)$ is an (ϵ, d, r_0, n)-rank expander with $\epsilon < 1/2$ then for* every *non-zero Hermitian matrix W in the null space of $\mathcal{A}$ the number of negative eigenvalues of W is larger than $r_0/2$.*

The authors have also suggested a novel algorithm for fast recovery of matrix X, which is notably faster than the conventional Nuclear Norm Minimization (NNM).

Theorem 4. *PSD Recovery If the operator $\mathcal{A}(X) = \sum_{i=1}^{d} A_i X A_i^T$ is an (ϵ, d, r_0, n)-rank expander with $\epsilon < 1/2$, then for every $k \leq r_0(1-\epsilon)$, every PSD matrix X of rank k can be perfectly recovered from $\mathcal{A}(X)$ using the algorithm shown in figure 1.*

Algorithm 1 Reconstruct a low rank PSD matrix X from under-determined linear measurements $Y = \sum_{i=1}^{d} A_i X A_i^*$.

Input:
Constant integer $d \geq 1$.
Matrices $A_i \in \mathbb{R}^{m \times n}, 1 \leq i \leq d$, and $Y \in \mathbb{R}^{m \times m}$.
Output:
Low rank PSD matrix X.
Initialize
Compute $Y = S \Sigma S^*$, with S full column rank (SVD).
Set $P = I - SS^*$.
Set $Q := \text{Null}\big((PA_1)^T, \ldots, (PA_d)^T\big)^T$.
Compute $B_i = A_i Q$, and set $M = \sum_{i=1}^{d} B_i \otimes B_i$.
Find $X \in \mathbb{R}^{n \times n}$ with $\text{vec}(X) = (Q \otimes Q) M^{\dagger} vec(Y)$.

Figure 1: Recover the low rank matrix X from its measurements $\mathcal{A}(X) = \sum_{i=1}^{d} A_i X A_i^T$ [2].

The authors have extended the above results to the Hermitian matrices as well (instead of only PSD matrices).

Using simulations, the authors have shown the validity of the proposed algorithm in recovering low rank matrices. Although not proved analytically, in their simulation results the authors show that the proposed method also works when measurement matrices G_i are sparse.

In [3] the authors extend the conditions for recovery of low rank matrices to obtain similar recovery requirements for sparse matrices. Following this approach, they derive the best known restricted isometry property and null space conditions for matrix recovery. More specifically, the problem of interest in [3] is to recover a matrix $X_0 \in \mathcal{R}^{n_1 \times n_2}$ (with $n = n_1 \leq n_2$) from a set of *corrupted* linear measurements given by $y = \mathcal{A}(X_0) + z$. Here, z is the noise term with $\|z\|_2 \leq \epsilon$ and $\mathcal{A}(.) : \mathcal{R}^{n_1 \times n_2} \rightarrow \mathcal{R}^m$ is the measurement operator.

If X_0 is low rank, then we can solve the following optimization problem to retrieve X_0 from y:

$$\begin{aligned} &\min \|X\|_\star \\ &\text{subject to } \|\mathcal{A}(X) - y\|_2 \leq \epsilon \end{aligned} \tag{4}$$

Where $\|X\|_\star = \sum_i \sigma_i(X)$ is the nuclear norm.

The authors have proved that the extension of *any sufficient* condition for recovery of sparse vectors using ℓ_1 minimization is also sufficient for the recovery of *low rank* matrices using nuclear norm minimization.

4 Conclusions and future works

In the past two weeks, we read some papers about matrix recovery approaches and the required conditions to ensure successful recovery. Some of the concepts mentioned in these papers can help us in our work on constraint enforcing neural networks. For instance, the definition of rank expanders for matrices in [2] is in fact very similar to the definition of expander graphs. Here, instead of saying that the number of neighbors of a given subset of left nodes is larger than some value, we put the same condition on the rank of the linearly transformed matrix. The proposed linear measurement operation has the form $\mathcal{A}(X) = \sum_{i=1}^{d} G_i X G_i^T$, where $X \in \mathcal{R}^{n\times n}$ and G_i's are i.i.d random Gaussian matrices with zero mean and variance equal to one. Each measurement $G_i X G_i^T$ captures part of the information in X. Hopefully if these measurements are relatively incoherent, the information adds up and the rank of $\mathcal{A}$ increases with the number of measurements, d.

One might be able to use the above concept of rank expanders together with the theorems mentiones in [3] to prove recoverability of sparse vectors or matrices from linear measurements as well as finding proper measurement matrices.

The other important topic of interest is to use the learning rules of higher order neural networks as an initial step to find the constraint matrix in our constraint enforcing neural networks.

References

[1] C. L. Giles, T. Maxwell, *Learning, invariance, and generalization in high-order neural networks*, Applied Optics, Vol. 26, No. 23, 1987, pp. 4972-4978.

[2] A. Khajehnejad, S. Oymak, B. Hassibi, *Subspace expanders and matrix rank minimization*, Int. Symp. on Inf. Theory, 2011.

[3] S. Oymak, K. Mohan, M. Fazel, B. Hassibi, *A simplified approach to recovery conditions for low rank matrices*, Proc. IEEE Int. Symp. Inf. Theory (ISIT), 2011.