ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE

School of Computer and Communication Sciences

Handout final Statistical Physics for Communication and Computer Science
Final Project - due date: July 3 2017 June 02, 2017

In the course we discussed in some detail the AMP algorithm in the framework of compressive sensing when the prior is unknown and we use the Lasso estimator. The aim of this project is to explore another estimation problem called *rank one factorization* with a *known prior*. Various versions of this problem have applications in community detection and also is of interest in statistics.

1 Formulation of matrix factorization problem: Bayesian setting

Let $\underline{s}^{(0)} = (s_1^{(0)}, \cdots, s_n^{(0)})^T$ be a binary column vector with $s_i^{(0)} \in \{+1, -1\}$. These are the hidden variables that we will have to estimate. The observation model is

$$y_{ij} = \sqrt{\frac{\lambda}{n}} s_i^{(0)} s_j^{(0)} + z_{ij}, \qquad i, j = 1, \cdots, n$$

where z_{ij} is a *symmetric* matrix with iid Gaussian elements such that $z_{ij} \sim \mathcal{N}(0,1)$ for $i \neq j$ and $z_{ii} \sim \mathcal{N}(0,2)$. In vector form we have $\mathbf{y} = \sqrt{\frac{\lambda}{n}} \underline{s}^{(0)} \underline{s}^{(0)T} + \mathbf{z}$, so we have noisy observations of a rank one matrix. We assume further that the prior on the binary vector components is iid uniform on $\{-1, +1\}$. Thefore it is natural to use a Bayesian setting. The posterior is

$$p(\underline{s}|\mathbf{Y}) = \frac{\exp\{-\frac{1}{4}\sum_{i,j=1}^n (Y_{ij} - \sqrt{\frac{\lambda}{n}} s_i s_j)^2\}}{\sum_{\underline{s}} \exp\{-\frac{1}{4}\sum_{i,j=1}^n (Y_{ij} - \sqrt{\frac{\lambda}{n}} s_i s_j)^2\}}$$

Expanding the squares and simplifying terms independent of $\underline{s}$ distribution can also be written as

$$p(\underline{s}|\mathbf{z}, \underline{s}^{(0)}) = \frac{1}{Z} \exp\{-\frac{1}{2} \sum_{i,j=1}^n (\frac{\lambda}{n} s_i^{(0)} s_j^{(0)} s_i s_j + \sqrt{\frac{\lambda}{n}} z_{ij} s_i s_j)\} \tag{1}$$

with the partition function

$$Z = \sum_{\underline{s}} \exp\{-\frac{1}{2} \sum_{i,j=1}^n (\frac{\lambda}{n} s_i^{(0)} s_j^{(0)} s_i s_j + \sqrt{\frac{\lambda}{n}} z_{ij} s_i s_j)\}$$

This is a random spin system on a complete graph with random iid "coupling constants" z_{ij} and $s_i^{(0)}$. The average MSE of an estimator $\hat{s}_i$ is by definition

$$MSE = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mathbf{z}, \underline{s}^{(0)}} (\hat{s}_i - s_i^{(0)})^2$$

and it is minimized by the MMSE estimator $\hat{s}_i^{\text{mmse}} = \langle s_i \rangle$ which is the magnetization. Instead of computing this optimal estimator we set in for a suboptimal solution using the AMP estimator.

2 AMP equations

Consider first the "scalar" estimation problem $y = \sqrt{\gamma} s^{(0)} + z$ where $s^{(0)} = \pm 1$, γ is the snr and $z \sim \mathcal{N}(0,1)$.

(i) Show that the MMSE estimator for this scalar problem is

$$\hat{s} = \langle s \rangle = \tanh(\gamma s^{(0)} + z)$$

and that the MMSE: $\mathbf{E}_{z,s^{(0)}}[(s - \langle s \rangle)^2] = \mathrm{mmse}(\gamma)$ where

$$\mathrm{mmse}(\gamma) \equiv 1 - \mathbb{E}_z[(\tanh(\gamma + \sqrt{\gamma} z))^2]$$

(ii) Write down the BP equations for the model (1). Use the same parametrizations as in class used for binary spins.

(iii) Deduce in the large n limit the AMP equations by going through the necessary simplications. *Hint*: silmilar derivations are done in Chap 7 on the Sherrington-Kirkpatrick model that we did not treat in class this year. You should find the following AMP (TAP-like) equations for the AMP estimates at time t: $\hat{s}_i^t = m_i^t$

$$m_i^t = \tanh\Big\{\sqrt{\frac{\lambda}{n}}(Y\underline{m}^{t-1})_i - m_i^{t-2}\frac{\lambda}{n}\sum_{j=1}^{n}(1-(m_j^{t-1})^2)\Big\}. \tag{2}$$

(iii) Explain heuristically how to find the correct state evolution equation for the $\mathrm{MSE}^t_{\mathrm{amp}} \equiv 1 - q_t^2$ at iteration t. This equation is:

$$q_{t+1}^2 = 1 - \mathrm{mmse}(\lambda q_t), \qquad q_0 = 1 - \epsilon$$

Note that in this problem we do not initialize with $\sigma_0 = 1$ otherwise state evolution does not start, so we take ϵ small positive representing a small prior bias on the signal. (Correspondingly you also have to be careful how you initialize AMP above).

3 Numerical implememtation

Independently from your derivations above, we ask you to implement numerically the AMP algorithm and state evolution.

(i) Investigate the fixed points and threshold λ_c of state evolution.

(ii) Run AMP above and below the threshold and check that the empirical $\mathrm{MSE}_{\mathrm{amp}}$ agrees with the one predicted by state evolution (as a function of λ).

(iii) In the heuristic derivation of state evolution one supposes that

$$\sqrt{\frac{\lambda}{n}}(Y\underline{m}^{t-1})_i - m_i^{t-2}\frac{\lambda}{n}\sum_{j=1}^{n}(1-(m_j^{t-1})^2)$$

has a Gaussian distribution (as if the CLT applied). Check this assertion numerically. Check also that this is not true for $\sqrt{\frac{\lambda}{n}}(Y\underline{m}^{t-1})_i$ if the Onsager term is not present.