Statistical Physics of Computation 2025 - Exercises

Vittorio Erba, Emanuele Troiani

Week 7

7.1 Approximate Message Passing for spiked-Wigner model

We are still looking at spiked-Wigner matrices Y of dimension N

$$Y = \sqrt{\frac{\lambda}{N}} x_\star x_\star^\top + Z \tag{1}$$

where Z is a symmetric matrix with standard Gaussian elements and each element of $x_\star$ has independently prior $P_\star$. We are in particular interested in Bayes Optimal recovery of the spike vector $x_\star$: we are given Y and the knowledge of the generative process. This will be achieved through the Approximate Message Passing (AMP) algorithm, an iterative algorithm that will iterate $\hat{x}^t$, starting with a sample from the prior $\hat{x}^0 \sim P_\star$ as follows

$$\begin{aligned} u^t &= \sqrt{\frac{\lambda}{N}} Y \hat{x}^t - b_{t-1} \hat{x}^{t-1} \\ \hat{x}^{t+1} &= \eta_t(u^t) \end{aligned} \tag{2}$$

where b_t is zero for $t = 0$ and otherwise

$$b_t = \frac{1}{N} \sum_{i=1}^{N} \nabla \eta_t(u^t)_i \tag{3}$$

A remarkable property of AMP is that it has, for $N \gg 1$, a low-dimensional, step-by-step characterization through the so called state evolution equations of the order parameters $m^t = u^t x_\star^\top / N$ and $q^t = \|u^t\|^2 / N$. This algorithm works for a large class of denoisers η_t, and it *can* be Bayes-Optimal with the choice $\eta_t(u^t) = \eta(u^t, \sigma_t)$ where $\sigma_t = \lambda \|\hat{x}^t\|^2 / N$ and

$$\eta(u, \sigma) = \frac{\int_{\mathbb{R}^N} \mathrm{d}x \; P_\star(x) \; x \; e^{-\frac{\sigma}{2}||x||^2 + u^\top x}}{\int_{\mathbb{R}^N} \mathrm{d}x \; P_\star(x) e^{-\frac{\sigma}{2}||x||^2 + u^\top x}} \,. \tag{4}$$

With this choice of denoise, we could show also that $q^t = m^t$ at all times (i.e. Nishimori conditions are satisfied, which is a necessary condition for Bayes-Optimality), with

$$m^{t+1} = \frac{\lambda m^t}{\lambda m^t + 1} \,. \tag{5}$$

We will not do that explicitly, as it is quite a gymnastics of Gaussian integration.

Notice that:

- AMP can be Bayes-Optimal with this choice of denoiser because the state evolution tracking the magnetization of the iterate u against the truth $x_\star$ are the same as the state equation we derived in week 5.
- Even with the correct denoiser, AMP may fail to be Bayes-Optimal. The BO estimator solves the state equations "globally" (picks the solution with largest free entropy), while AMP solves state evolution iteratively.
- The denoiser uses $\sigma_t = \lambda \|\hat{x}^t\|^2/N$ at each time. State evolution guarantees that for $N \gg 1$ then $\sigma_t \approx \lambda q_t$. Thus, we can either compute σ_t at each step from the iterate, or we can compute it directly by state evolution, independently from the iterate. Asymptotically the two options are the same.

Consider the case of Gaussian prior on the components of $x_\star$:

$$(x_\star)_i \sim \mathcal{N}(0,1)\,. \tag{6}$$

1. Show that the denoiser becomes

$$\eta(u,\sigma)_i = \frac{u_i}{\sigma+1} \tag{7}$$

For this prior one can do the integral explicitly

$$\eta(u,\sigma) = \frac{\int_{\mathbb{R}^N} \mathrm{d}x\; x\; e^{-\frac{\sigma+1}{2}||x||^2 + u^\top x}}{\int_{\mathbb{R}^N} \mathrm{d}x\; e^{-\frac{\sigma+1}{2}||x||^2 + u^\top x}} = \frac{u}{\sigma+1} \tag{8}$$

2. Code AMP (2) for this case and plot the average magnetization m^* at convergence (how do you check for convergence?) as a function of λ for $N = 1000$, averaged over 5 realizations of $x_\star, Y$. Given that the problem enjoys a $\mathbb{Z}_2$ symmetry, we take the magnetization in absolute value for this case, as we could get magnetized along $x_\star$ or $-x_\star$ depending on the initialization. You can verify that your code works by comparing it with the state equation from week 6: if $\lambda < 1$ then $m^* = 0$ and $m^* = 1 - 1/\lambda$ otherwise.

   You can see the code in the companion notebook.

3. In week 7 we showed that the equilibrium magnetization m^* is a fixed point of the state evolution iteration

$$m^{t+1} = \frac{\lambda m^t}{\lambda m^t + 1}\,. \tag{9}$$

   Of course $m^* = 0$ is a fixed point of the iteration. When is it stable as a function of λ, meaning that if $m^0 = m^* + \epsilon$ then $m^{t\to\infty} \to m^*$?

   To check the stability we look at small perturbation around zero. Say at time zero we have $m^0 = \epsilon$. Then at the next iteration we have $m^1 = \lambda\epsilon$ at leading order in ϵ. The fixed point is stable if the perturbation shrinks as you iterate, so if $\lambda < 1$. Notice that a second solution exist for $\lambda > 1$, and in that case $m^* = 0$ is always instable.

4. This may be a problem: if you initialize in the prior, at infinitely large N you should expect the magnetization to start *exactly* at zero. This means that despite $m = 0$ being an unstable fixed point of the iteration, AMP is still stuck there even for $\lambda > 1$, where recovery can in principle happen with a better initialization. As you should have seen in

your code, this is typically not a problem though, as in a computer N is always finite, albeit large. Show that initializing in the prior we expect $m^0 \sim 1/\sqrt{N}$ (i.e. that finite size correction break the $\mathbb{Z}_2$ symmetry of the problem).

The magnetization will be

$$m^0 = \frac{1}{N}\sum_{i=1}^{N}(x_\star)_i(x_1)_i \tag{10}$$

where $x_\star$ and x_1 are two samples from the prior. At large d, $m^0 = 0$ at leading order by the law of large numbers:

$$m^0 = \mathbb{E}[x_\star]\mathbb{E}[x_1] \tag{11}$$

where we dropped the i index. In this particular case the mean of the prior is zero, so we have to go to the next order. Notice that for each i, $(x_\star)_i(x_1)_i$ is an independent random variable with mean zero and variance $\mathbb{E}[x_\star^2]$, which is finite. For the central limit theorem, this means that m^0 will then be a centered Gaussian with variance $\mathcal{O}(N/N^2) = \mathcal{O}(1/N)$, giving our result.

5. We now would like to find an interpretation for what is AMP doing, i.e. to understand which estimator it is implementing. Start by showing that at the fixed point with positive magnetization, $\hat{x}^*$ is proportional to an eigenvector of Y.

   The denoiser η is a linear function, so at the fixed point we have

$$\begin{aligned} u^* &= \sqrt{\frac{\lambda}{N}}Y\hat{x}^* - \frac{\hat{x}^*}{\lambda m^* + 1} \\ \hat{x}^* &= \frac{u^*}{\lambda m^* + 1} \end{aligned} \tag{12}$$

   After a rewriting, we have

$$\hat{x}^* \propto \sqrt{\frac{\lambda}{N}}Y\hat{x}^* \tag{13}$$

   which is the equation for an eigenvector of Y, up to a rescaling.

6. We now want to show that the eigenvector from before is the one associated to the largest eigenvalue. Suppose at time $t = 0$ we are just a small perturbation away from the $m^* = 0$ fixed point

$$\hat{x}^0 = \delta\hat{x}\,. \tag{14}$$

   Show that at time $t = 1$ you have

$$\hat{x}^1 = \sqrt{\frac{\lambda}{N}}Y\delta\hat{x}\,. \tag{15}$$

   First, notice that $\sigma_\star = 0$. This gives us

$$u^1 = \sqrt{\frac{\lambda}{N}}Y\delta\hat{x}^0 \tag{16}$$

   which then gives the result

7. Now let's continue the iteration, and for simplicity we neglect the Onsager correction b_t at all times. Argue that that after enough iterations our estimate will be aligned with the top eigenvector of Y (assume that the leading eigenvalue of Y is non-degenerate).

   If we neglect the Onsager term, AMP is implementing a fixed point iteration similar to Lanczos algorithm. We thus expect AMP to converge to the top eigenvector. For clarity we give an illustration of this phenomenon. Say you start with a vector u and iteratively multiply it k times by Y. Call λ_p, v_p the eigenvalues and eigenvectors of Y, sorted in decreasing order. We can then write

$$u = \sum_{p=1}^{N} \alpha_p v_p \,, \qquad Y = \sum_{p=1}^{N} \lambda_p v_p v_p^\top \,, \qquad Y^k = \sum_{p=1}^{N} \lambda_p^k v_p v_p^\top \,. \tag{17}$$

   After k iterations we would have

$$Y^k u = \sum_{p=1}^{N} \alpha_p \lambda_p^k v_p \tag{18}$$

   The overlap with the top eigenvector v_1 will be

$$v_1^\top Y^k u = \sum_{p=1}^{N} \alpha_p \lambda_p^k v_1^\top v_p = \alpha_1 \lambda_1^k \|v_1\| + \sum_{p=2}^{N} \alpha_p \left(\frac{\lambda_p}{\lambda_1}\right)^k v_1^\top v_p \,. \tag{19}$$

   Since $\lambda_p < \lambda_1$, $Y^k u$ will converge exponentially fast to the eigenspace of v_1.

8. Thus, without Onsager's term we see that AMP converges to a rescaled version of the top eigenvector of Y. Let's verify that this is the case also with the correct Onsager's term numerically.

   We verify this in the notebook