Statistical Physics of Computation - Exercises

Emanuele Troiani, Vittorio Erba, Yizhou Xu

September 2024

Week 11

11.1 The spiked-tensor model

Consider the following inference problem. You observe a symmetric tensor

$$y_{ijk} = \sqrt{\frac{2\lambda}{N^2}} x_i^* x_j^* x_k^* + \xi_{ijk} \tag{1}$$

with $1 \leq i \leq j \leq k \leq N$. $x^* \in \mathbb{R}^N$ is the ground truth, generated with a factorized prior $P_0(x) = \prod_{i=1}^N P_0(x_i)$. ξ_{ijk} is a symmetric tensor, whose components $1 \leq i \leq j \leq k \leq N$ are all independent standard Gaussian random variables (mean zero, variance one). $\lambda > 0$ acts as a signal-to-noise (SNR) ratio.

The aim of this exercise is to derive the state equation for the order parameter

$$m = \frac{1}{N} \mathbb{E}_{y,x_*} \mathbb{E}_{x \sim P_{\text{post}}(\cdot|y)} x_*^T x \tag{2}$$

as a function of the SNR λ.

1. Show that the posterior distribution for the problem can be written as

$$P(x|y) = \frac{1}{Z(y)} \left[\prod_{i=1}^N P_0(x_i) \right] \left[\prod_{i \leq j \leq k} \frac{e^{-\frac{\lambda}{N^2} x_i^2 x_j^2 x_k^2 + \sqrt{\frac{2\lambda}{N^2}} x_i x_j x_k y_{ijk}}}{\sqrt{2\pi}} \right] \tag{3}$$

for a given $Z(y)$. How is $Z(y)$ defined for this measure?

The posterior is the same as for the spiked-matrix model, or spike-Wigner model, we saw in class, see ex 9.1.1. In particular, the output channel term is given by a product of Gaussian densities evaluated in y_{ijk}, each with mean $\sqrt{2\lambda/N^2} x_i x_j x_k$ and variance one. One then expands the square at the exponent, and reabsorbs all x-independent terms in the partition function. The partition function equals

$$Z(y) = \int dx_1 \dots dx_N \left[\prod_{i=1}^N P_0(x_i) \right] \left[\prod_{i \leq j \leq k} \frac{e^{-\frac{2\lambda}{2N^2} x_i^2 x_j^2 x_k^2 + \sqrt{\frac{2\lambda}{N^2}} x_i x_j x_k y_{ijk}}}{\sqrt{2\pi}} \right]. \tag{4}$$

2. Show that the averaged replicated partition function equals

$$\mathbb{E}_y[Z(y)^n] = \int dy_{ijk}\, e^{-\frac{1}{2}\sum_{i\le j\le k} y_{ijk}^2} \prod_{\alpha=0}^{n} \int dx^{(\alpha)} \left(\prod_{i=1}^{N} P_0\left(x_i^{(\alpha)}\right) \right) \times \left(\prod_{i\le j\le k} \frac{e^{-\frac{\lambda}{N^2}\left(x_i^{(\alpha)}\right)^2\left(x_j^{(\alpha)}\right)^2\left(x_k^{(\alpha)}\right)^2 + \sqrt{\frac{2\lambda}{N^2}} x_i^{(\alpha)} x_j^{(\alpha)} x_k^{(\alpha)} y_{ijk}}}{\sqrt{2\pi}} \right) \tag{5}$$

where you should notice that we are taking the product over $n+1$ replicas.

This is the same as exercise 9.1.2, with the minimal modifications of dealing with rank-3 tensors instead of rank-2 tensors, i.e. matrices. Also consider the additional factor 2 in λ.

3. Integrate over the disorder, i.e. the observation y, to get at leading order in N

$$\mathbb{E}_y[Z(y)^n] = \int \prod_{\alpha,i} P_0\left(x_i^{(\alpha)}\right) \mathrm{d}x_i^{(\alpha)}\ \exp\left(\frac{\lambda N}{3} \sum_{\alpha<\beta} \left(\sum_i \frac{x_i^{(\alpha)} x_i^{(\beta)}}{N} \right)^3 \right) \tag{6}$$

One has

$$\begin{aligned}
&\prod_{i\le j\le k} \int dy_{ijk}\, \frac{e^{-\frac{1}{2}y_{ijk}^2}}{\sqrt{2\pi}} e^{-\sum_\alpha \frac{2\lambda}{2N^2}\left(x_i^{(\alpha)}\right)^2\left(x_j^{(\alpha)}\right)^2\left(x_k^{(\alpha)}\right)^2 + \sum_\alpha \sqrt{\frac{2\lambda}{N^2}} x_i^{(\alpha)} x_j^{(\alpha)} x_k^{(\alpha)} y_{ijk}} \\
&= \prod_{i\le j\le k} e^{-\sum_\alpha \frac{2\lambda}{2N^2}\left(x_i^{(\alpha)}\right)^2\left(x_j^{(\alpha)}\right)^2\left(x_k^{(\alpha)}\right)^2 + \frac{2\lambda}{2N^2}\sum_{\alpha,\beta} x_i^{(\alpha)} x_j^{(\alpha)} x_k^{(\alpha)} x_i^{(\beta)} x_j^{(\beta)} x_k^{(\beta)}} \\
&= e^{\frac{\lambda}{N^2}\sum_{\alpha\ne\beta}\sum_{i\le j\le k} x_i^{(\alpha)} x_i^{(\beta)} x_j^{(\alpha)} x_j^{(\beta)} x_k^{(\alpha)} x_k^{(\beta)}} \\
&\approx e^{\frac{\lambda}{N^2}\sum_{\alpha\ne\beta}\frac{1}{3!}\sum_{i,j,k} x_i^{(\alpha)} x_i^{(\beta)} x_j^{(\alpha)} x_j^{(\beta)} x_k^{(\alpha)} x_k^{(\beta)}} \\
&= e^{\frac{\lambda}{N^2}\sum_{\alpha\ne\beta}\frac{1}{3!}\left(\sum_i x_i^{(\alpha)} x_i^{(\beta)}\right)^3} \\
&= e^{\frac{\lambda}{3N^2}\sum_{\alpha<\beta}\left(\sum_i x_i^{(\alpha)} x_i^{(\beta)}\right)^3}
\end{aligned} \tag{7}$$

4. Introduce the appropriate order parameters and obtain

$$\mathbb{E}_y[Z(y)^n] = \int \prod_{\alpha<\beta} \mathrm{d}\hat{q}_{\alpha\beta}\, \mathrm{d}q_{\alpha\beta}\ \exp\left(N I_{\text{energy}}(q_{\alpha\beta}, \hat{q}_{\alpha\beta}) + N I_{\text{entropy}}(\hat{q}_{\alpha\beta}) \right) \tag{8}$$

where we defined

$$I_{\text{energy}}(q_{\alpha\beta}, \hat{q}_{\alpha\beta}) = \frac{\lambda}{3} \sum_{\alpha<\beta} q_{\alpha\beta}^3 - \sum_{\alpha<\beta} q_{\alpha\beta}\hat{q}_{\alpha\beta} \tag{9}$$

and

$$I_{\text{entropy}}(\hat{q}_{\alpha\beta}) = \log\left(\int \prod_\alpha P_0(x_\alpha)\, \mathrm{d}x_\alpha\ \exp\left\{ \sum_{\alpha<\beta} \hat{q}_{\alpha\beta} x_\alpha x_\beta \right\} \right) \tag{10}$$

where we stress that here the integral over dx_α runs over the real numbers for all $\alpha = 0, \dots, n$.

Same as exercise 9.1.4.

5. Argue that the RS free entropy

$$\phi = \lim_{N\to\infty} \mathbb{E}_y \left[\frac{1}{N}\log Z(y)\right] \tag{11}$$

can be expressed as

$$\phi = \text{extr}_{q,\hat{q}} \left[\frac{1}{2}\left(\frac{\lambda}{3}q^3 - q\hat{q}\right) + \int Dz P_0\left(x_0\right) \mathrm{d}x_0 \log\left(\int P_0\left(x\right)\mathrm{d}x\ \exp\left\{-\frac{\hat{q}}{2}x^2 + \sqrt{\hat{q}}xz + \hat{q}xx_0\right\}\right)\right] \tag{12}$$

In the RS ansatz and at leading order in n, we have

$$I_{\text{energy}}(q_{\alpha\beta}, \hat{q}_{\alpha\beta}) = \frac{\lambda}{3}\frac{n(n+1)}{2}q^3 - \frac{n(n+1)}{2}q\hat{q} \approx \frac{n}{2}\left(\frac{\lambda}{3}q^3 - q\hat{q}\right). \tag{13}$$

The entropic term is the same as in the matrix model, see ex 9.1.7 and 9.1.8.

We now consider the case $P_0(x) = N(x, 0, 1)$, i.e. over each component the prior is a standard Gaussian distribution.

7. Show that the free entropy simplifies to

$$\phi(q, \hat{q}) = \frac{\lambda}{6}q^3 - \frac{1}{2}q\hat{q} + \frac{\hat{q}}{2} - \frac{1}{2}\log(1+\hat{q}) \tag{14}$$

By performing all the Gaussian integrals we get

$$\phi(q, \hat{q}) = \frac{\lambda}{6}q^3 - \frac{1}{2}q\hat{q} + \frac{\hat{q}}{2} - \frac{1}{2}\log(1+\hat{q}) \tag{15}$$

See also ex 10.1.1 and 10.1.2.

8. Show that the state equation for $m = q$ is

$$m = \frac{\lambda m^2}{\lambda m^2 + 1}. \tag{16}$$

The saddle point equations for the free entropy of point 11.1.7 are $\hat{q} = \lambda q^2$ and

$$q = 1 - \frac{1}{1+\hat{q}} = \frac{\hat{q}}{1+\hat{q}} \tag{17}$$

which we can combine to get the solution, using also $m = q$ from Nishimori's identities.

9. Solve the state equation. In particular, make sure you answer clearly to the following subquestions as a function of λ:

   (a) How many solutions there are?

   (b) Which of the solutions is the dominant one for each value of λ?

   (c) If there is any phase transition, specify at which value of λ it happens and its order.

(d) What is the interpretation of the phase transitions, if any, in terms of the inference problem?

Hint: not all questions above have nice analytical answers. Use your favorite numerical tools to explore this question, and feel free to provide approximate numerical answers. In your submitted homework, include all plots you used to motivate your answer.
It may be useful to generate a plot of all solutions m to the state equation as a function of λ, where you highlight which of the solutions is dominant.

We have the trivial solution $m_0 = 0$. If $m \neq 0$, we have

$$\lambda m^2 - \lambda m + 1 = 0 \tag{18}$$

giving

$$m_\pm = \frac{\lambda \pm \sqrt{\lambda^2 - 4\lambda}}{2\lambda}\,. \tag{19}$$

The two solutions are real for $\lambda > 4$. Thus, for $\lambda < 4$ we have only the trivial solution, and for $\lambda > 4$ we have 3 coexisting solutions. Notice that $m_\pm(\lambda \geq 4)$ are both bounded away from zero for any finite λ, so that if there is a phase transition at finite λ, it will necessarily be of the first order.
We expect that the dominant solution for $\lambda \gg 4$ will be

$$m_+ = \frac{\lambda + \sqrt{\lambda^2 - 4\lambda}}{2\lambda}\,, \tag{20}$$

as for $\lambda \to \infty$ it is the only solution that converges to $m = 1$, and is the only solution which is monotone increasing (we expect that as the SNR increases, m increases). We can check whether this is the case explicitly by computing the free entropy in the three solutions. We have

$$\phi(m) = \frac{\lambda}{6}m^3 - \frac{\lambda}{2}m^3 + \frac{\lambda m^2}{2} - \frac{1}{2}\log(1+\lambda m^2) = -\frac{\lambda}{3}m^3 + \frac{\lambda}{2}m^2 - \frac{1}{2}\log(1+\lambda m^2) \tag{21}$$

It is easy to check numerically that $\phi(m_0) = 0$, that $\phi(m_-) < 0$ for all $\lambda > 4$, and that $\phi(m_+) > 0$ for $\lambda \gtrapprox 4.4$. Thus, we have a first order phase transition at $\lambda \approx 4.4$. Before the trivial solution is dominant, after the m_+ solution is dominant.
Thus

a) For $\lambda < 4$ there is only one trivial solution, while for $\lambda \geq 4$ there are three distinct solutions.
b) For $\lambda \lessapprox 4.4$ we have $m_* = m_0 = 0$, while for $\lambda \gtrapprox 4.4$ we have $m_* = m_+(\lambda)$.
c) The phase transition is of the first order, as the order parameter m is discontinuous.
d) The phase transition can be interpreted as a weak recovery transition, the minimal value of λ for which $m_* \neq 0$.

10. Suppose you were to derive and run the AMP algorithm for this problem. How do you expect the magnetization of the iterates of AMP evolve iteration by iteration? How is the magnetization initialized?

AMP is tracked by state evolution, which is just the following iterative scheme for the state equations

$$m_{t+1} = \frac{\lambda m_t^2}{\lambda m_t^2 + 1} \tag{22}$$

with initialization $m_{t=0} = 0^+$.

11. To which value of the magnetization will AMP converge for any finite λ when initialized for $m = 0^+$, i.e. very small but still positive magnetization? Discuss for which value of λ the problem has an impossible phase, for which ones has an hard phase, and for which ones has an easy phase.

    For $\lambda < 4$, it can only converge to $m = 0$. For $\lambda > 4$, we study the iteration at first order around $m = 0^+$. We have

$$m_{t+1} \approx \lambda m_t^2 \implies m_t = \lambda^{-1} (\lambda m_0)^{2^t} \tag{23}$$

    from which we see that for any fixed $\lambda > 4$, we can always pick a m_0 small enough, i.e. $m_0 = 0^+$, such that $\lambda m_0 < 1$. This implies that the state evolution of AMP converges always to $m = 0$ for any finite λ.
    Thus, the problem has an impossible phase for $\lambda \lessapprox 4.4$, where AMP converges to the only solution $m = 0$ of the state equations, which is also the BO solution. It has instead a big hard phase $\lambda \gtrapprox 4.4$ for which the BO estimator achieves good performance $m > 0$, while AMP remains stuck at $m = 0$. Finally, there is no easy phase. The spiked-tensor problem is particularly hard for algorithms.

11.2 Mixed spiked-tensor model

We now consider a mixed inference problem for generic factorized prior P_0. We observe both a matrix $y^{(2)}$ and a tensor $y^{(3)}$

$$\begin{aligned} y_{ij}^{(2)} &= \sqrt{\frac{\lambda_2}{N}} x_i^* x_j^* + \xi_{ij}^{(2)} \\ y_{ijk}^{(3)} &= \sqrt{\frac{2\lambda_3}{N^2}} x_i^* x_j^* x_k^* + \xi_{ijk}^{(3)} \end{aligned} \tag{24}$$

1. Argue without doing additional computations that the RS free entropy associated to the posterior distribution equals

$$\begin{aligned} \phi = \operatorname{extr}_{q,\hat{q}} \Bigg[&\frac{1}{2} \left(\frac{\lambda_2}{2} q^2 + \frac{\lambda_3}{3} q^3 - q\hat{q} \right) \\ &+ \int Dz P_0(x_0) \, \mathrm{d}x_0 \log \left(\int P_0(x) \, \mathrm{d}x \, \exp \left\{ -\frac{\hat{q}}{2} x^2 + \sqrt{\hat{q}} x z + \hat{q} x x_0 \right\} \right) \Bigg] \end{aligned} \tag{25}$$

   and for Gaussian priors equals

$$\phi = \operatorname{extr}_{q,\hat{q}} \left[\frac{1}{2} \left(\frac{\lambda_2}{2} q^2 + \frac{\lambda_3}{3} q^3 - q\hat{q} \right) + \frac{\hat{q}}{2} - \frac{1}{2} \log(1 + \hat{q}) \right] \tag{26}$$

   In the mixed model, the posterior distribution is just the prior term, times the output channel of the matrix model, times the output channel of the tensor model. The disorders in the matrix and tensor parts are independent, so the averages can be performed separately and lead to an energetic term which is just the sum of the matrix and the tensor energetic terms. The entropic term is the same, as it depend only on the unchanged prior and on the unchanged definition of the order parameter q.
   Note about the grading: many of you were very quick on this answer, not providing much detail at all. Even if all computations are known and can be skipped, there are at least the following points that should be clearly made in this answer:

- that the posterior distribution is the usual prior term, times the output channel of the matrix observation, times the out channel of the tensor observation.
- that the disorder ξ for the matrix and tensor output channel are independent, so that when averaging over the disorder I can average them separately.
- that the final energetic term is just then the sum of the 2 previous cases.
- that the prior term does not change.

Missing any of these points produces a partial answer!

2. From now on consider the Gaussian prior. Show that the state equation for $m = q$ equals

$$m = \frac{\lambda_2 m + \lambda_3 m^2}{1 + \lambda_2 m + \lambda_3 m^2} \tag{27}$$

The saddle-point equations are

$$\hat{q} = \lambda_2 q + \lambda_3 q^2 \tag{28}$$

and

$$q = 1 - \frac{1}{1+\hat{q}} = \frac{\hat{q}}{1+\hat{q}} \tag{29}$$

giving the answer. We report for completeness the free entropy

$$\begin{aligned} \phi(m) &= \frac{1}{2}\left(\frac{\lambda_2}{2}m^2 + \frac{\lambda_3}{3}m^3 - \lambda_2 m^2 - \lambda_3 m^3\right) + \frac{\lambda_2 m + \lambda_3 m^2}{2} - \frac{1}{2}\log(1 + \lambda_2 m + \lambda_3 m^2) \\ &= -\frac{\lambda_2}{4}m^2 - \frac{\lambda_3}{3}m^3 + \frac{\lambda_2}{2}m + \frac{\lambda_3}{2}m^2 - \frac{1}{2}\log(1 + \lambda_2 m + \lambda_3 m^2) \end{aligned} \tag{30}$$

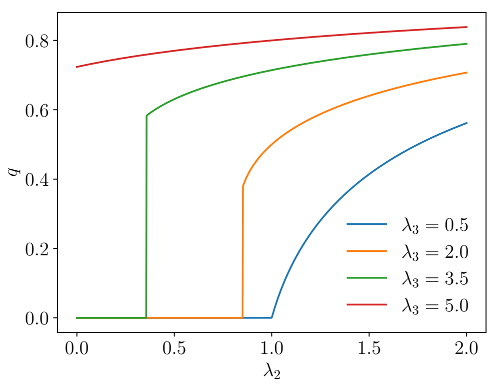
3. How does the dominant solution m_* of the state equation behave in the two limit $\lambda_3 = 0$ and $\lambda_2 = 0$?

For $\lambda_3 = 0$ we have the matrix model we studied in the lecture. We know that $m_* = 0$ up to $\lambda_2 = 1$, and then $m_* > 0$ with a second order phase transition for $\lambda_2 > 1$.
For $\lambda_2 = 0$ we have the tensor model studied in the exercise above. We know that $m_* = 0$ up to roughly $m_* \approx 4.4$, and then $m_* > 0$ with a first order phase transition.

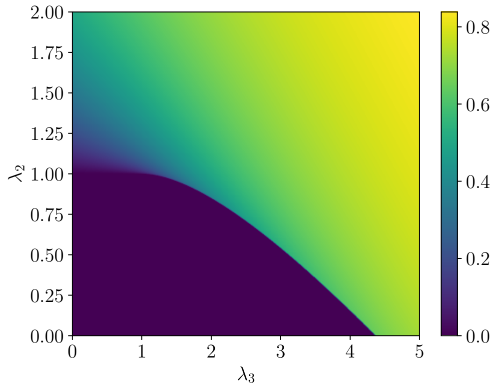
4. Study numerically the phase diagram. To do that:

   (a) Fix a value of λ_3, and then using your favorite tool plot as a function of λ_2 the dominant solution m to the state equation. Repeat for several values of λ_3.

   (b) Combine the information obtained from all the plots above into a single phase diagram, i.e. a plot in the (λ_2, λ_3) plane where for each value of (λ_2, λ_3) you specify whether $m_* = 0$ or $m_* > 0$, and you highlight the order of the phase transitions separating the two regions. The plot can be qualitative, i.e. the location of the transitions may be approximate as long as the overall phenomenology is well described. Even a hand-drawn qualitative phase diagram can be acceptable for this question if clearly drawn.



(a) Overlap as a function of λ_2 for different values of λ_3.



(b) Phase diagram / value of the magnetization as a function of λ_2, λ_3

Figure 1: Overlaps as a function of the SNR ratio λ for the binary prior.

Hints: discard a priori all solutions for which $m \notin [0, 1]$. A good choice for λ_3 in point (a) is $\lambda_3 = [0.01, 0.5, 1, 1.5, 2, 3.5, 4, 4.3, 4.4, 4.5, 5]$, but feel free to explore more values.

You can see the curves of point (a) in Figure 1a. We can see that increasing λ_3 the kind of phase transition changes drastically. For all $0 < \lambda_3 < 1$ there is a second order phase transition at $\lambda_2 = 1$ from zero to non-zero m. This is what we observed in the spiked matrix model. For $1 < \lambda_3 \lessapprox 4.4$ there is a first order phase transition at $0 < \lambda_2^c(\lambda_3) < 1$ from zero to finite m, as we observed in the first part of this homework. For $\lambda_3 \gtrapprox 4.4$ there is no phase transition in λ_2. The point (λ_2, λ_3) is a tricritical point, where a first order transition boundary vanishes into a second order phase transition.
In Figure 1b you can see the phase diagram requested in point (b). The region surrounding $(\lambda_2, \lambda_3) = (0, 0)$ has $m = 0$, so that all phase transitions observed are weak recovery transitions, meaning that the BO estimator goes non-analytically from being completely uninformed to having a non-zero overlap with the ground though. Depending on the precise path in the phase diagram, the weak recovery can be either a first or second order phase transition. As expected in noisy estimation problems there is no perfect recovery transition, as there is always some residual noise that we cannot eliminate unless $\lambda \to \infty$. The the second order phase transition happens for $0 < \lambda_3 < 1$ and $\lambda_2 = 1$. The point $(\lambda_2, \lambda_3) = (1, 1)$ is a tricritical point, the point at which the first order transition line becomes a second order transition line.