Supporting Information of Passed & Spurious: Descent Algorithms and Local Minima in Spiked Matrix-Tensor Models

Stefano Sarao Mannelli [1] **Florent Krzakala** [2] **Pierfrancesco Urbani** [1] **Lenka Zdeborová** [1]

A. Kac-Rice formula

A.1. p-odd Cases

In the cases in which the order of the tensor p is odd we encounter an interesting phenomenon due to the different symmetries of the two types of observation. The matrix is symmetric by inverting the sign of the signal, $\hat{x} \mapsto -\hat{x}$, while the tensor is not symmetric for odd p. This creates an asymmetry in the complexity, Fig. 6 (to be compared with Fig. 2) and causes a shift toward lower correlations of the band characterizing the non-informative minima. Therefor observing when the complexity at $m = 0$ becomes negative does not guarantee that the non-informative minima disappeared. To do so, one must check that the whole non-informative band disappears. This should be contrasted with the case of even p where a maximum of the complexity $\Sigma(m)$ is always at $m = 0$. These two definitions of the threshold have little, but not negligible, difference, see Fig. 7. Observe that as Δ_p increases the peak of the complexity decreases, since the loss Eq. (3) tends to the simple matrix-factorization problem where the landscape is characterized by two isolated minima. This implies that the two definitions become indistinguishable for large Δ_p. In the main text we use the definition taking into account the maximum (even when it is not strictly at $m = 0$) because gives a more accurate characterization of the trivialization threshold.

B. Gradient Flow

B.1. Dependence on the Initial C onditions

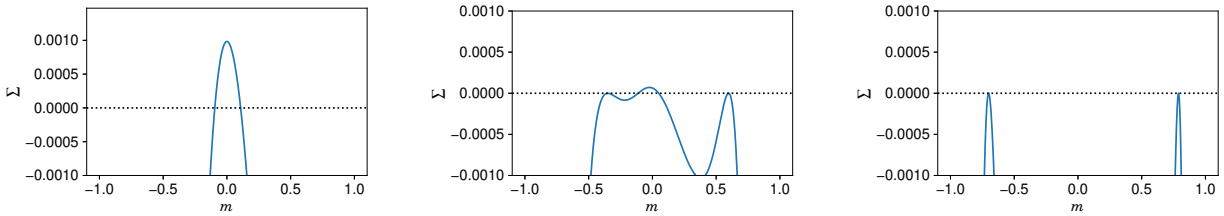
The dynamics of the gradient flow shows a dependence on the initial conditions, because formally zero correlation is a (unstable) fixed point of the GF state evolution. In practice we observe for both GF and ML-AMP that instability of the fixed point is sufficient for good performance of the algorithm. However, this makes the definition of the convergence time depend of the initial condition.

We observed from our numerical solution of the GF state evolution equations that the initial condition add a factor $a^{\log m(0)}$ to the convergence times. Thus by fitting this term and rescaling the convergence times, the different initializations collapse into a single curve, see inset of Fig. 8. Finally, the collapsed points were used to extrapolate the critical line as shown in the main text, Fig. 4.

C. AMP

C.1. From AMP to ML-AMP

In this section we consider the spiked-tensor model in a Bayesian way. We show how the Bayes-optimal AMP leads



(a) $\Delta_2 = 2$ (b) $\Delta_2 = 2/3$ (c) $\Delta_2 = 2/5$

Figure 6. Analogously to Fig. 2, the figures show the complexity, Eq. (19), as a function of the correlation with the signal for different values of parameter Δ_2 at fixed $\Delta_p = 4.0$ in the case $p = 3$.

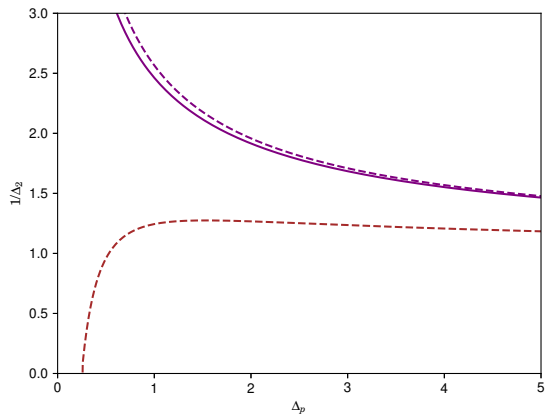


Figure 7. The thresholds representing the trivialization of the landscape (purple) and the point where the support of $\Sigma(m) \geq 0$ become disconnected (brown) for tensors of order $p = 3$. We compare the two definitions of the trivialization threshold described in Sec. A.1: the solid line considers just the positivity of the complexity Eq. (19) at $m = 0$, the dashed line considers the whole non-informative band.

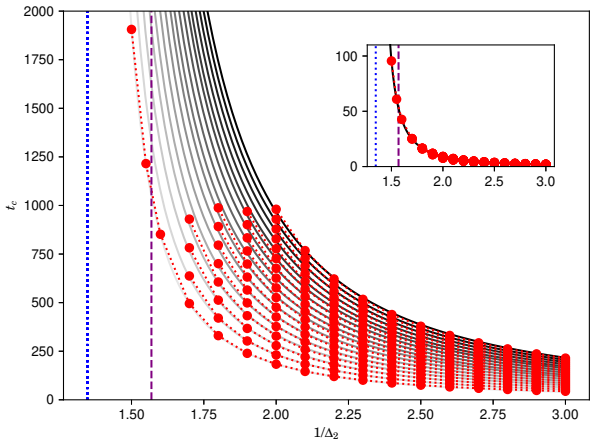


Figure 8. The time corresponding to convergence close to the signal is shown for $\Delta_p = 4.0$ in the case $p = 3$. Different shades of grey correspond to different initial conditions, from $m(0) = 10^{-10}$ (light grey) to $m(0) = 10^{-42}$ (dark grey). The different initializations collapse to a single line when the time is rescaled by $a^{\log m(0)}$ with $a = 1.3$, see inset. In the figure we fit only the case $m(0) = 10^{-10}$ with a power law and use the same parameters for all the other fits with a vertical translation. The divergence point extrapolated is $1/\Delta_2^{\rm GF} = 1.35$ and is represented by the vertical dotted line, while the dashed line identifies the landscape trivialization predicted with the Kac-Rice formula, $1/\Delta_2^{\rm triv} = 1.57$.

to the Maximum Likelihood AMP using a temperature-like parameter T. We will introduce the algorithm AMP for a generic T, and show that as $T \to 0$ we recover ML-AMP as presented in the main text. The probability distribution we consider is

$$P(X|Y,T) \propto e^{-\mu\|x\|^2} \prod_{i<j} e^{-\frac{1}{2T\Delta_2}\left(Y_{ij}-\frac{x_i x_j}{\sqrt{N}}\right)^2} \prod_{i_1<\cdots<i_p} e^{-\frac{1}{2T\Delta_p}\left(T_{i_1\dots i_p}-\frac{\sqrt{(p-1)!}}{N^{(p-1)/2}}x_{i_1}\dots x_{i_p}\right)^2}. \tag{33}$$

The scheme for deriving AMP estimating marginals of such a probability distribution can be found in (Lesieur et al., 2017; Sarao Mannelli et al., 2018) and consist in making a Gaussian assumption on the distribution of the messages in the belief propagation (BP) algorithm and neglecting the node-dependence in the messages. A final consideration to be used in order to derive the algorithm is that the spherical constrain can be imposed by setting $\frac{1}{N}\sum_i(\hat{x}_i{}^2 + \sigma_i) = 1$ at every iteration. The resulting AMP algorithm will iterate on the following equations:

$$B_i^t = \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}} \sum_{k_2<\cdots<k_p} \frac{T_{ik_2\dots k_p}}{T\Delta_p} \hat{x}_{k_2}^t \dots \hat{x}_{k_p}^t + \frac{1}{\sqrt{N}} \sum_k \frac{Y_{ik}}{T\Delta_2} \hat{x}_k^t - \mathrm{r}_t \hat{x}_i^{t-1} \tag{34}$$

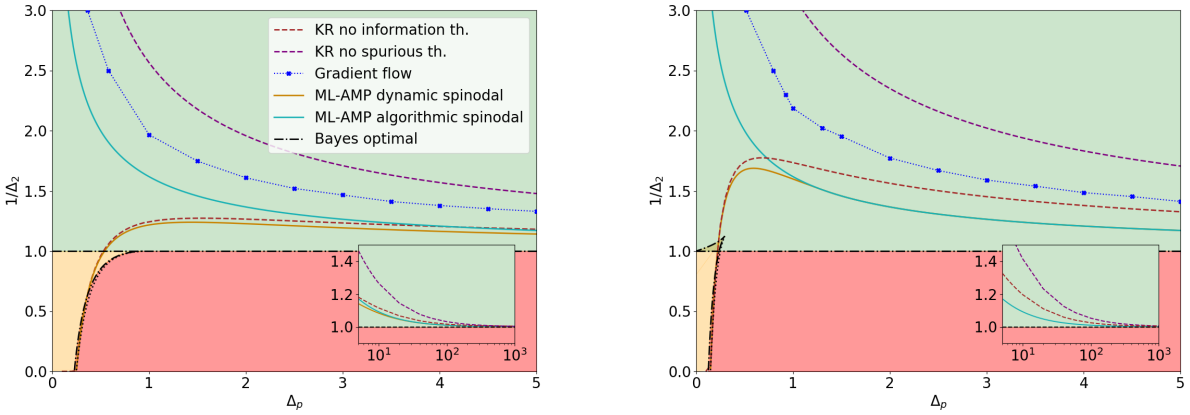
$$\hat{x}_i^{t+1} = 2\frac{B_i^t}{1+\sqrt{1+\frac{4}{N}||B^t||_2^2}}, \tag{35}$$

$$\sigma^{t+1} = \frac{2}{1+\sqrt{1+\frac{4}{N}||B^t||_2^2}}. \tag{36}$$

with $||\cdots||_2^2$ the ℓ_2-norm and r_t the Onsager reaction term

$$\mathrm{r}_t = \frac{1}{\Delta_2 T^2}\frac{1}{N}\sum_k \sigma_k^t + \frac{p-1}{\Delta_p T^2}\frac{1}{N}\sum_k \sigma_k^t \left(\frac{1}{N}\sum_k \hat{x}_k^t \hat{x}_k^{t-1}\right)^{p-2}. \tag{37}$$

In the limit $T \to 0$ AMP defined by Eqs. (34-37) is equivalent to ML-AMP, Eqs. (25-28). To see this we define the rescaled variables $\hat{\sigma}^t \doteq \sigma^t/T$, $\tilde{B}^t \doteq T\,B^t$ and $\tilde{\mathrm{r}}_t \doteq T\mathrm{r}_t$. Taking the limit $T \to 0$ the expression for $\hat{x}_i^{t+1}$ Eq. (35) and the expression for $\hat{\sigma}_i^{t+1}$ Eq. (36) simplify as Eq. (26) and as Eq. (26) respectively. Dropping the tildes we obtain ML-AMP as presented in the main text.



(a) $p = 3$ (b) $p = 4$

Figure 9. The phase diagram already describe in Fig. 1 with two additional lines. The dashed brown line is the limit predicted by Kac-Rice formula where the support of the $\Sigma(m) \geq 0$ becomes disconnected (above the line). The full orange line is related to the ML-AMP algorithm, is called *dynamical spinodal*, below it the algorithm converenges to $m = 0$ even if initialized in the solution. In the insets we show the large Δ_p behaviour of the thresholds, where we can observe that the lines merge at infinity.

C.2. State evolution

The generic T version of AMP has a slightly more complicated SE that depends of two order parameters: the already introduced $m^t = \frac{1}{N}\sum_i \hat{x}_i^t x_i^*$ and $q^t = \frac{1}{N}\sum_i (\hat{x}_i^t)^2$ the self-overlap of the estimator. The SE equations are:

$$m^{t+1} = 2\frac{z^t(T)}{1+\sqrt{1+4y^t(T)}}\,, \tag{38}$$

$$q^{t+1} = 4\frac{y^t(T)}{\left(1+\sqrt{1+4y^t(T)}\right)^2} \tag{39}$$

and

$$\text{MSE}^t = 1 - 2m^t + q^t\,, \tag{40}$$

with $y^t(T) = \left(z^t(T)\right)^2 + \left(\frac{1}{T^2}\frac{q^t}{\Delta_2} + \frac{1}{T^2}\frac{(q^t)^{p-1}}{\Delta_p}\right)$ and $z^t(T) = \frac{1}{T}\frac{m^t}{\Delta_2} + \frac{1}{T}\frac{(m^t)^{p-1}}{\Delta_p}$.

Given $\frac{1}{N}||\hat{x}^0||_2^2 \neq 0$, in the limit $T \to 0$ AMP SE Eqs. (38-39) simplify, to a single equation corresponding to ML-AMP SE Eq. (41). This is seen by taking the limit for Eq. (39) which gives $q^t = 1\ \forall t > 0$, implying $\text{MSE}^t = 2(1-m^t)$. Then, using the result for q^t, we show that Eq. (38) tends to Eq. (30).

C.3. Derivation of spinodals

From SE Eq. (30) we can obtain analytical equations for the *spinodals*, the threshold of stability of the different ML-AMP fixed points. We have $\hat{x}^{t+1} = f_{SE}(z^t)$ with

$$f_{SE}(z) = \frac{z}{\sqrt{z^2+\gamma}}, \tag{41}$$

with $\gamma = 1/\Delta_2 + 1/\Delta_p$ and $z = m/\Delta_2 + m^{p-1}/\Delta_p$. Observe that: $f'_{SE}(z) = \frac{\gamma}{(z^2+\gamma)^{\frac{3}{2}}}$. We can now define either $\Delta_p \equiv \Delta_p(z;\Delta_2,\gamma) = \frac{f_{SE}(z)^{p-1}}{z - \frac{f_{SE}(z)}{\Delta_2}}$ or $\Delta_2 \equiv \Delta_2(z;\Delta_p,\gamma) = \frac{f_{SE}(z)}{z - \frac{f_{SE}(z)^{p-1}}{\Delta_p}}$.

As remarked in (Sarao Mannelli et al., 2018), the spinodals are given by the following condition:

$$\begin{aligned} 0 = \frac{d\log\Delta_2}{dm} \propto \frac{d\log\Delta_2}{dz} = \\ = \frac{z\left[(p-2)\gamma\left(\frac{z}{\sqrt{z^2+\gamma}}\right)^{p-1} - z^3\Delta_p\right]}{z(z^2+\gamma)\left[\Delta_p z^2 - z\left(\frac{z}{\sqrt{z^2+\gamma}}\right)^{p-1}\right]}. \end{aligned} \tag{42}$$

A trivial solution is given by $z \to 0$ corresponding to stability of the non-informative solution $m = 0$, and gives the *algorithmic spinodal* for the cases $p \in \{3,4\}$. This solution and has a very simple equation for every p: $\Delta_2 = 1/\sqrt{\gamma}$ giving Eq. (32), already presented in the main text. An interesting implication of Eq. (32) is that it is independent from the value of p, it is in some sense *universal* among the $2+p$-models.

The expression for the stability of the informative solution, *dynamical spinodal*, is less straightforward, but analytical progresses can be done in the cases $p = 3$ and $p = 6$ (using Cardano formula) and in the case $p = 4$ for which it is

p	Δ_2	Δ_p
4	$\frac{2}{3} \simeq 0.667$	$\frac{4}{3} \simeq 1.333$
5	0.470	0.451
6	0.384	0.305
7	0.322	0.220
8	0.279	0.172
9	0.246	0.147
10	0.220	0.121

Table 1. Table of the values of tricritical points for $p \geq 4$.

equivalent to a second order polynomial

$$z^2 + \gamma = \left(\frac{(p-2)\gamma}{\Delta_p} \right)^{\frac{2}{(p-1)}} = \left(\frac{2\gamma}{\Delta_4} \right)^{\frac{2}{3}}, \tag{43}$$

that admits a single solution in $\mathbb{R}^+$:

$$z = \sqrt{\left(\frac{2\gamma}{\Delta_4} \right)^{\frac{2}{3}} - \gamma}\,. \tag{44}$$

An important point in the phase diagram is where the algorithmic and dynamical spinodals meet, this is called the *tricritical point*. Its value is obtained for different p, numerically (for $p > 4$) and analytically (for $p = 4$), and is reported in Table 1. The case $p = 3$ does not show any tricritical point for any finite Δ_p, the two lines eventually meet at $\Delta_p = \infty$ when the spiked matrix problem is recovered.

For the cases $p > 4$ we observe additionally the zero temperature analog of what is called *hybrid phase* in AMP in Bayes-optimal regime (Ricci-Tersenghi et al.). The hybrid phase is illustrated in Fig. 10. This phase is defined as a region where the ML-AMP algorithm initialized at random converges to a solution with positive correlation but that is less correlated then the solution achievable starting from the solution. In these cases Eq. (32) does not correspond to the algorithmic spinodal but it is just the stability of the non-informative solution.

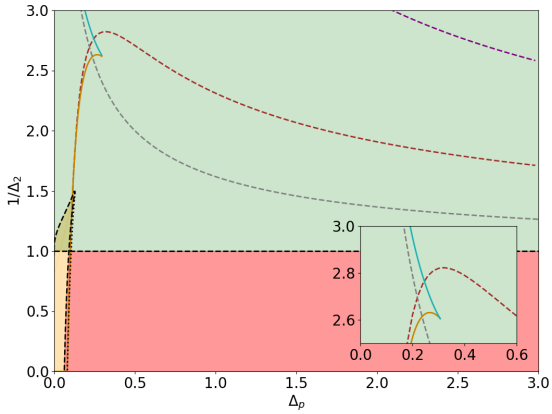


Figure 10. Phase diagram as shown in Fig. 9 for the case $p = 6$. The difference between $p = 3, 4$ and $p > 4$ is that a new phase appears, called hybrid hard phase, where two fixed points of ML-AMP aligned with the signal are present and the convergence to one or the other depends on the initialization. The region is highlighted in the inset. In the phase diagram the grey dashed line represent the threshold above which the non-informative fixed point becomes unstable.