We are grateful for the detailed reviews and comments that will help us to make our paper clearer.

Referee #1

We thank for remarks (2), (3), (8) that we will fix.

Remark 1: Δ_2 and Δ_p are the variances, not standard deviations. The loss comes from terms such as $\exp[-(Y_{ij} - \sqrt{N}\sigma_i\sigma_j)^2/(2\Delta_2)]$ where the cross-term appearing in the loss is $Y_{ij}\sqrt{N}\sigma_i\sigma_j/\Delta_2$.

Remark 4: We thank the referee for pointing our imprecision. We will add the references to the original work: "S. O. Rice, Mathematical analysis of random noise, Bell System Tech. J., 24 (1945), 46–156"; "M. Kac, On the average number of real roots of a random algebraic equation (1943)" summarized very nicely in "Adler, Robert J., and Jonathan E. Taylor. Random fields and geometry. Springer 2009". The references we gave originally is the application of the Kac-Rice method in the high-dimensional setting.

Remark 5: Bounding the variance would be another way to justify that the annealed calculation is tight. We have not done that, but will look into it. Having the quenched complexity equal to the annealed complexity is only asymptotic and in density. It implies that the annealed complexity bound is tight, but it does not itself imply a very strong bound on the variance. We agree with the referee, that this passage should be clarified to explain better why this is important and why this means that the annealed complexity bound is actually tight. We will adjust the final version.

Remark 6: The reasoning of the referee is indeed what we had in mind. We will rephrase the argument in the final version following the suggestion of the reviewer.

Remark 7: We thank the reviewer for the remark we will clarify in the final version. In the formula we select a specific reference frame where we can easily represent the Hessian. In this frame the basis element $\boldsymbol{e}_1$ is the vector aligned with the part of the signal that is tangent to the estimator. Calling $\boldsymbol{\sigma}^*$ the signal and $\boldsymbol{\sigma}$ the estimator, and saying that the estimator is an element in the basis, we have the relation: $\boldsymbol{\sigma}^* = \sqrt{1-m^2}\,\boldsymbol{e}_1 + m\,\boldsymbol{\sigma}$ with $m = \langle \boldsymbol{\sigma}, \boldsymbol{\sigma}^* \rangle \in [-1, +1]$.

Referee #3

We summarise the state-of-the-art for the model and the technical contributions:

The literature on this specific model is so far limited. As far as we know it was introduced in [21], and studied in [20].

The main theoretical techniques the CHSCK equations and the (annealed = log of expectation) Kac-Rice method were used in [20]. In this paper we derive the replica symmetric quenched (= expectation of the log) Kac-Rice equations justifying (non-rigorously) that the annealed bound is tight in the context we use it, this is one technically involved contribution. Another technically involved contribution (non-rigorous, but conjectured exact) is the analysis of the large time solution of the CHSCK equations that has not been done in existing literature, it has only been studied in related optimization problems, not in inference problems where the spike is to be recovered. Finally the interpretation of the Kac-Rice calculation in terms of the relation of the effective Hessian and the behaviour of the dynamics is also original. For a summary of the main non-technically-involved but in our opinion broadly interesting contributions see the answer to Referee #4.

Referee #4

Remark 1: We consider gradient flow from random initialization because among the algorithms we are able to analyze this is closest to what is done in practice for training neural networks. Our point is not to search for the best algorithm for this specific model – we believe that would be the AMP algorithm (in agreement with conjectures of previous works). Our main interest is to understand the trajectory Gradient Flow (GF) takes in the non-convex high-dimensional landscape. While understanding the behaviour of variants of GF is also a very interesting question, we start with the most basic algorithm that has non-trivial behaviour.

Remark 2: Referee's assessment of what was known from previous work, notably ref. [20], is correct. We summarised the technical contributions of the present work with respect to existing literature in the answer to Referee #3. Additionally, the reason why we are persuaded the present paper is significant is, as Referee #1 puts it "The basic idea is in retrospect simple and may well play a part in analysing a range of other models." Indeed, our reasoning leading to the formula for Δ_2^{GF} is in retrospect not restricted to the present model, and ends up way simpler than the full CHSCK analysis or the full Kac-Rice complexity calculation. The mechanism of converging to the threshold states and then escaping from them can also be tested numerically even in models that are not amenable to analytic description. This makes the results of the present work widely testable and applicable to other settings than the present model. We shall highlight these in the final version.