– – – Response to Reviewer 1 – – –

Re: Definition of ω The noise level ω is defined in Theorem 2 and it is analogous to the effective SNR $\sigma^2\sqrt{s\log(n)/N}$ which governs sharp transitions in Sparse PCA (e.g. Amini & Wainwright, 2009). In our case the effective SNR ω naturally depends also the inner layer dimensions and depth of the generative network. Up to log factors and polynomials in d, ω takes the form $\sigma^2\sqrt{k\log(n)/N}$ (see for example the informal Theorem 1). We will add a remark after the main Theorem 2 and Proposition 1 explaining how to interpret ω as an SNR making a connection with the bounds in previous works (in Section 2).

Re: Symbols not properly defined We regret the omission of the definitions of n_d and other symbols. We will make sure they are properly defined and fix the typos in the camera ready.

– – – Response to Reviewer 2 – – –

Re: Proof of convergence We suspect our analysis could indeed be translated into a convergence proof. Doing so would require significant technical enhancements, including establishing convexity (or a similar property) locally around the global minimizer, along with careful estimates to ensure a proper step size. These are significant technical improvements, and we leave them for future work.

– – – Response to Reviewer 3 – – –

Re: Novelty of the proof ingredients The extension from previous works is substantial at a technical level. Both [29] and [31] solve quadratic objectives, but in our paper, the objective is quartic. Generally, quartic objectives are challenging to analyze because the fourth power of Gaussians have tails too thick for uniform concentration results. Our work demonstrates how to deal with these terms while maintaining optimal sample complexity. Additionally, unlike in [29] and [31], we control sub-exponential matrix noise with multiple structures. We will add a paragraph to the paper discussing exactly these points.

Re: Title We will change the title to "Nonasymptotic Guarantees for Spiked Matrix Recovery with Generative Priors"

Re: Comparability to Sparse PCA Signal recovery problems where multiple signal structures hold simultaneously (e.g. low-rank AND sparse matrices) have been notoriously difficult, leading to no tractable algorithms at optimal sample complexity. Consequently, one would expect that enforcing low-rank AND generative priors would be comparably difficult. In this work, we indeed show that this combination of structural priors is not inherently difficult. This would lead practitioners to invest in building and using generative priors, as studied in this paper.

Re: Claiming computational-statistical guarantees The central property that allows the "no computational-statistical gap" statement is the fact that the optimization landscape is benign. Specifically, we show that the direction given by the gradient (almost everywhere) is a descent direction (with nonzero directional derivative). It is true that we do not provide a proof of convergence of a particular algorithm, but we establish the the conditions are appropriate for such an algorithm to converge. See response to Reviewer 2 for further comments on a possible convergence proof.

Re: gradient algorithm might not find descent direction The theorem asserts that there is a linear descent direction (with rate bounded away from zero) for any direction within the Clarke subdifferential (which almost everywhere is precisely the gradient). Thus, the descent claim applies in any direction a subgradient descent algorithm would take. Additionlly, this result ensures that there are no spurious maxima or saddles that can not be escaped.

– – – Response to Reviewer 4 – – –

Re: Novelty of the proof ingredients See response to Reviewer 3.

Re: Polynomial scaling in d **of the rates** We focused on studying optimal sample complexity with respect to the intrinsic dimensionality of the signal. We aimed for a result where the dependence on depth d is polynomial (it could have been exponential: for example straightforward bounds on the Lipschitness of a network are exponential in the parameter d!) Optimizing the proof for superior dependence on d would not drastically alter the fundamental theoretical advance, and it would require a much more cumbersome proof. As we show in the numerical experiments the bounds are quite conservative and the actual dependence on the depth is much better in practice.

Re: Empirical results This paper is a theoretical paper of a fundamental sample complexity improvement when using generative priors. We agree it would be exciting to see demonstrations in practical domain-specific applications, and we hope this result inspires authors to go through the difficult process of training models in these settings.

Re: Clarity, typos and odd formulations We will simplify the presentation of the main results by putting notation and assumptions outside. We will add a comment after the main Theorem 2 and Proposition 1 explaining their consequences and interpretation.