We thank the reviewers for their insightful comments.

Rev 1: *Projection step*: Thanks for reading the paper carefully, and for your kind assessment. We agree with you that linear convergence of Net-PGD only holds as long as the projection step can be solved exactly (proof lines [517,543]). Performing such a projection can be challenging in general, and resolving this is an important open problem. However, we make two comments. First, our theory goes through even if we relax the projection requirement to be approximate, as long as the approximation error is additively bounded, i.e. $\|x - v^t\| \leq \|x - v^*\| + \epsilon$ for some global parameter $\epsilon > 0$. Second, there could be other relaxations; a recent (June 2019) preprint by Gamez, Eftekhari, and Cevher gives a polynomial-time ADMM-type algorithm for inverse imaging with (trained) generative priors, as long as the mapping $z \rightarrow x$ is near-isometric. A similar result may be possible in our (untrained) setting, but its proof will require some care.

Tightness of sample complexity bounds:We have not attempted to derive Fano-style information-theoretic lower bounds, but intuitively, our sample complexity result is not too loose. If we assume k_1=k_2. . .=k_L(as in [10]), then our derived sample complexity matches N_w=$\sum_{i=1}^{L-1} k_i k_{i+1}$(no. of unknown parameters of network prior) up to log factors. Our result is asymmetric in k_1 which makes sense as k_1 is dim. of latent code, but this could be an artifact of proof technique.

Rev 2: *Novelty of proofs*: We agree that IHT-style algorithms and proofs are not new; however, we emphasize that the proof of Lemma 1 (RIP of Gaussian matrices) is novel for deep *untrained* network priors. Moreover, compressive phase retrieval is a nonlinear forward model, and to show linear convergence we require Lemma 2, which is also novel. Theorems 1 and 2 strictly use Lemmas 1 and 2 to complete the algorithmic guarantees.

Validity of Definition 1 and approximation error: While the true validity of any model can be questioned, we point to prior work in ([10] and [9]) which establish this architecture as a useful prior for natural images. Empirically, we show in Column 2 of Figs. 1a,1b,2a,2b, that all test images are well-reconstructed using this prior. If $x^* = G(\mathbf{w}^*; z) + \varepsilon_o$ where G is a DIP parameterized by $\{k_l, d_l\}$, then the modeling error ε_o gets reflected in the final reconstruction accuracy: $\|\widehat{x} - x^*\| \leq (1 + 2\beta/\gamma)\, \|\varepsilon_o\| + \varepsilon/\gamma$ whenever $\|y - A\widehat{x}\| \leq \min_{x \in \mathcal{S}} \|y - Ax\| + \varepsilon$ (follows from combination of Lemma 1 and Lemma 4.3 of [6]), as long as sample requirements for Lemma 1 are satisfied. We will append this result.

Improvements over learned image priors and AMP: Since our network prior is untrained, we do not claim accuracy benefits over learning-based methods; the (significant) benefit of our approach is that it does not require large training datasets. We will certainly add additional comparisons to AMP; however, BM3D-AMP appears to perform worse than TVAL3 (Figs. 1,2 of concurrent work in [17]) in extremely low sample regimes such as those in this paper.

Advantage of Net-PGD v/s Net-GD: We do not know how to analyze NetGD, since the output of the intermediate iterations is not guaranteed to lie in the range of untrained generators (which our theoretical analysis requires). In our experience, the running times of Net-GD and Net-PGD are comparable; even though Net-PGD requires solving subproblems in each iteration, the overall iteration complexity is lower. We will clarify this in the revision. Step size requirements will be appended to Theorems 1 and 2, as indicated on Lines [524] and [552].

Rev 3: *Novelty over Bora et al, '17, Oymak et al '17*: We respectfully push back against novelty criticisms in general, and specifically when contrasted against these two papers. Please allow us to clarify possible misunderstandings.

First, we emphasize that our second application (compressive phase retrieval) is a **non-linear** inverse problem. To our knowledge, we are the first to formally consider deep image priors in nonlinear recovery problems (and phase retrieval in particular) whereas these previous papers only address linear inverse problems.

Second, we emphasize the DIP model is **not** a generative prior model a la Bora et al '17. They assume a trained network (and optimize over the latent code), while we assume an untrained network with a fixed, random latent code and optimize over all *network weights*. This obviously is a much more challenging problem experimentally, but also theoretically. Therefore, our results and techniques, used to establish Lemma 1 and particularly Lemma 2, are more involved than those in Bora et al, '17.

Third, our motivation, techniques, and results are very different from the approach in the seminal work of Oymak et al, '17. They focus on linear inverse problems and priors defined by *convex* constraint sets; moreover, their focus is on getting *sharp* bounds which they succeed to do using their Gaussian widths analysis. In contrast, our proofs are significantly simpler and shorter (albeit potentially sub-optimal; see our response to Reviewer 1 above).

Validating the RIP result: It is well-known that RIP is empirically difficult to verify for any given measurement matrix (for the normal sparsity case, it is known to be NP-hard). Moreover, RIP is a sufficient but not necessary condition for successfully solving any inverse problem. The tradition in the literature has been to experimentally measure sample complexity and show improvement over handcrafted priors, which we have presented in Figures 1c and 2c.

More empirical results: We will gladly add more experiments to validate local linear convergence of Alg.1. (see right, log scale) and Alg.2. Please also note that we show superior empirical performance of Alg. 2 over a state-of-art (Sparta) for compressive phase retrieval (Fig.2), validating our theory.