Precise asymptotics for phase retrieval and compressed sensing with random generative priors

Benjamin Aubin[†], **Bruno Loureiro**[†], **Antoine Baker**[†]
Florent Krzakala[⋆], **and Lenka Zdeborová**[†]
[†] Institut de Physique Théorique, CNRS & CEA
Université Paris Saclay, Saclay, France.
[⋆] Laboratoire de Physique de l'Ecole Normale Supérieure
CNRS & Sorbonnes Universités & École Normale Supérieure
PSL University, Paris, France.

Abstract

We consider the problem of compressed sensing and of (real-valued) phase retrieval with random measurement matrix. We analyse sharp asymptotics of the information-theoretically optimal performance and that of the best known polynomial algorithms under a generative prior consisting of a single layer neural network with a random weight matrix. We compare the performance to sparse separable priors and conclude that generative priors might be advantageous in terms of algorithmic performance. In particular, while sparsity does not allow to perform compressive phase retrieval efficiently close to its information-theoretic limit, it is found that under the random generative prior compressed phase retrieval becomes tractable.

Over the past decade the study of compressed sensing has lead to significant developments in the field of signal processing, with novel sub-Nyquist sampling strategies and a veritable explosion of work in sparse representation. A central observation is that sparsity allows one to measure the signal with fewer observations than its dimension [1, 2]. The success of neural networks in the recent years suggests another powerful and generic way of representing signals with multi-layer generative priors, such as those used in generative adversarial networks (GANs) [3]. It is therefore natural to replace sparsity by generative neural network models in compressed sensing and other inverse problems, a strategy that was successfully explored in a number of papers, e.g. [4–11]. While this direction of research seems to have many promising applications, the theory of what can be efficiently achieved still falls short of the one developed over the past decade for sparse signal processing.

Here, we aim at proving precise asymptotics for the performance in two such inverse problems: (real-valued) phase retrieval and compressed sensing. These two problems of interest can be framed as a generalised linear estimation. One is given a set of observations $\mathbf{y} \in \mathbb{R}^n$ generated from a fixed (but unknown) signal $\mathbf{x}^\star$ as

$$\mathbf{y} = \varphi\left(\mathrm{A}\mathbf{x}^\star\right), \tag{1}$$

(for a given known φ, A), and the goal is to find back $\mathbf{x}^\star \in \mathbb{R}^d$ from the knowledge of $\mathbf{y}$ and A. Compressed sensing and phase retrieval are particular instances of this problem, corresponding to $\varphi(x) = x$ and $\varphi(x) = |x|$ respectively. Structured signals - such as sparse or binary $\mathbf{x}^\star$ - have been the subject of intense investigation, see e.g. [12, 13]. A typical situation in compressed sensing is that $\mathbf{x}^\star$ is sparse, i.e. only k of the d components are non-zero.

In this manuscript, we consider structured signals from a generative-neural-network type. In order to provide a sharp asymptotic theory, we restrict the analysis to a specific situation, when the signal $\mathbf{x}^\star$ is drawn from a single-layer network with known random weight-matrix W:

$$\mathbf{x}^\star = G\left(\mathrm{W}\mathbf{z}\right), \tag{2}$$

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

where G is a component-wise non-linearity and $\mathbf{z} \in \mathbb{R}^k$ is the latent representation of the signal. We take $\mathrm{A} \in \mathbb{R}^{n\times d}$ and $\mathrm{W} \in \mathbb{R}^{d\times k}$ to have i.i.d. Gaussian entries with zero means and variances $1/d$ and $1/k$, and focus on the high-dimensional regime given by taking $n, d, k \to \infty$ while keeping the measurement rate $\alpha = n/d$ and the compression factor $\rho = k/d$ constant. The specific results we present are for $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_k)$ and for the common choices $G \in \{\mathrm{linear}, \mathrm{ReLU}\}$. We want to stress that the presented analysis is valid for arbitrary G and separable $z_k \sim P_z$, and readily generalize to multi-layer generative neural networks with random i.i.d. weight matrices, corresponding results will be presented in an extended version of this manuscript. Another case that can be treated theoretically and left for future work is when the weight matrices are random rotationally invariant matrices independent from each other [8].

Our main contribution is specifying the interplay between the number of measurements needed for exact reconstruction of the signal, parametrized by α, and its latent dimension k. Of particular interest is the comparison between a sparse and separable signal (having a fraction ρ_s of non-zero components) and the structured generative model above, parametrized by $\rho = k/d$. While the number of unknown latent variables is the same in both cases if $\rho = \rho_s$, we show that the generative model structure has algorithmic advantages over the sparse one.

Optimal estimation is in our setting given by computing the posterior distribution of the signal given the observations. Although exact sampling from the posterior is intractable in the high dimensional regime, it is still possible to track the behaviour of the minimum-mean-squared-error estimator as a function of the parameters (α, ρ). Our main result are based on the line of works comparing, on one hand, the information-theoretically best possible reconstruction, using an ideal Bayesian inference decoder, regardless of the computation cost, and on the other, the best reconstruction using the most efficient known polynomial algorithm.

Sparsity: In the case of a separable prior, and in particular a sparse one, the setting of this paper has been the subject of many studies using the non-rigorous replica method, e.g. [14]. Later the information theoretic results, as well as the corresponding minimum mean squared error (MMSE), has been established fully rigorously in [15], together with the performance of the associated *approximate message passing* (AMP) algorithm [16]. For both the linear estimation and phase retrieval, the information theoretic limit for a perfect recovery is simply $\alpha > \alpha_{\mathrm{IT}} = \rho_s$, with ρ_s being the fraction of non-zero components of the signal $\mathbf{x}^\star$.

The ability of AMP to exactly reconstruct the signal, however, is different. A non-trivial line $\alpha_{\mathrm{alg}}^{\mathrm{sparse}}(\rho_s) > \alpha_{\mathrm{IT}}$ appears below which AMP fails, and no other polynomial algorithm is known. Strikingly, as discussed in [15], the behavior of the sparse linear estimation and phase retrieval is drastically different: while $\alpha_{\mathrm{alg}}^{\mathrm{sparse}}(\rho_s)$ is going to zero as $\rho_s \to 0$ for sparse linear estimation hence allowing for compressed sensing, it is not the case for the phase retrieval, for which $\alpha_{\mathrm{alg}}^{\mathrm{sparse}} \to 1/2$ in this limit. As a consequence, no efficient approach to real-valued compressed phase retrieval, in the limit considered here, is known.

Generative priors: In the case of a generative model prior, eq. (2), the computation of the information theoretic and algorithmic limits is more involved, and requires the generalization of both the replica method and of the approximate message passing algorithm. This generalization was developed for the multi-layer estimation problem with random matrices in [6], and proven rigorously for the single-layer prior in [17]. Neither of these works analysed the questions we are investigating here. We thus evaluate the corresponding equations and interpret them with the purpose of comparing sparse and generative priors for compressed sensing and phase retrieval.

From [6, 17] and our own rederivation it follows that the Bayes-optimal estimator $\hat{\mathbf{x}}^{\mathrm{opt}}$ achieves in the limit of $n, d, k \to \infty$ and $\alpha = n/d = \Theta(1)$, $\rho = k/d = \Theta(1)$ the minimum mean-squared-error

$$\mathrm{mmse}(\alpha, \rho) = \lim_{d\to\infty} \frac{1}{d}\mathbb{E}||\hat{\mathbf{x}}^{\mathrm{opt}} - \mathbf{x}^\star||_2^2 = \rho_x - q_x^\star \tag{3}$$

where ρ_x is the second moment of $P_{x^\star}$ and the scalar parameter $q_x^\star \in [0, \rho_x]$ is the solution of the following extremization problem

$$f(\alpha, \rho) = \underset{(q_x, \hat{q}_x, q_z, \hat{q}_z)}{\mathrm{extr}} \left\{ \frac{1}{2}\hat{q}_x q_x + \frac{\rho}{2}\hat{q}_z q_z - \alpha\Psi_y\left(q_x\right) + \Psi_{\mathrm{out}}\left(\hat{q}_x, q_z\right) - \rho\Psi_z(\hat{q}_z) \right\}. \tag{4}$$

The so-called potentials $(\Psi_y, \Psi_{\text{out}}, \Psi_z)$ are scalar functions depending on the choice of model and of the generative prior, and are given by

$$\Psi_y(q_x) = \mathbb{E}_\xi \left[\int \mathrm{d}y \; \mathcal{Z}_y\left(y, \sqrt{q_x}\xi, q_x\right) \log \mathcal{Z}_y\left(y, \sqrt{q_x}\xi, q_x\right) \right] , \tag{5}$$

$$\Psi_{\text{out}}(\hat{q}_x, q_z) = \mathbb{E}_{\xi,\eta} \left[\mathcal{Z}_{\text{out}}(\sqrt{\hat{q}_x}\xi, \hat{q}_x, \sqrt{q_z}\xi, \rho_z - q_z) \log \mathcal{Z}_{\text{out}}(\sqrt{\hat{q}_x}\xi, \hat{q}_x, \sqrt{q_z}\xi, \rho_z - q_z) \right] , \tag{6}$$

$$\Psi_z(\hat{q}_z) = \mathbb{E}_\xi \left[\mathcal{Z}_z(\sqrt{\hat{q}_z}\xi, \hat{q}_z) \log \mathcal{Z}_z(\sqrt{\hat{q}_z}\xi, \hat{q}_z) \right] , \tag{7}$$

with $\xi, \eta \sim \mathcal{N}(0,1)$ and auxiliary functions

$$\mathcal{Z}_y\left(y, \omega, V\right) = \int \frac{\mathrm{d}x}{\sqrt{2\pi V}} e^{-\frac{1}{2V}(x-\omega)^2} \delta\left(y - \varphi(x)\right) , \quad \mathcal{Z}_z\left(B, A\right) = \int \mathrm{d}z \; P_z(z) e^{-\frac{A}{2}z^2 + Bz} ,$$

$$\mathcal{Z}_{\text{out}}(B, A, \omega, V) = \int \mathrm{d}x \; e^{-\frac{A}{2}x^2 + Bx} \int \frac{\mathrm{d}z}{\sqrt{2\pi V}} e^{-\frac{1}{2V}(z-\omega)^2} \; \delta(x - G(z)) \, . \tag{8}$$

This reduces the asymptotics of the high-dimensional estimation problem to a low-dimensional extremization problem. Solving eq. (4) provides the information theoretical thresholds for perfect recovery (i.e. when mmse = 0). Interestingly, it also provides information about the algorithmic hardness of the problem. The above extremization problem is closely related the state evolution of the AMP algorithm for this problem, as derived in [6]. This algorithm is conjectured to provide the best polynomial time algorithm for the estimation of $\mathbf{x}^\star$ in the considered setting and limit. Specifically, the algorithm reaches a mean-squared error that corresponds to the local extremizer reached by gradient descent in the function (4) starting with uninformative initial conditions.

Phase diagrams: Below we summarize the findings of the above theory in the form of so-called phase diagrams. These are diagrams in the (ρ, α) plane quantifying the quality of reconstruction of the signal for the corresponding problem. For both the phase retrieval and compressed sensing problems we distinguish the following regions of parameters and the respective thresholds separating them: *Undetectable* region where the best achievable error is as bad as a random guess from the prior as if no measurements $\mathbf{y}$ were available. *Weak recovery* region where the optimal reconstruction error is better than the one of a random guess from the prior, but exact reconstruction cannot be achieved. *Hard* region where exact reconstruction can be achieved information-theoretically, but no efficient algorithm achieving it is known. The so-called *easy* region where the aforementioned AMP algorithm for this problem achieves exact reconstruction of the signal.

We locate the corresponding phase transitions in the following manner: For the weak recovery threshold α_c, we notice that the fixed point corresponding to an error as bad as a random guess corresponds to the values of the order parameters $(q_x, q_z) = (0, 0)$. This is an extremizer of the free energy (4) when the prior P_z has zero mean and the non-linearity φ is an even function. This condition is satisfied for the phase retrieval problem with generative priors that leads to zero-mean distributions on the components of the signal, but is not achieved for the other analyzed cases. In case this uninformative fixed point exists, we investigate its stability under the state evolution of the AMP algorithm, thus defining the threshold α_c. For $\alpha < \alpha_c$ the fixed point is stable, implying the algorithm is not able to find an estimator better than random guess. In contrast, for $\alpha > \alpha_c$ the AMP algorithm provides an estimator better than random guess. For phase retrieval with linear generative model in the setting of the present paper this analysis leads to the threshold $\alpha_c = \rho/[2(1+\rho)]$. If there exists a region where the performance of the AMP algorithm and the information-theoretic one do not agree we call it the *hard* region. The hard region is delimited by threshold α_{IT} and α_{alg}. Numerically we find these two thresholds by following the state evolution from two different initializations: one close to the ground truth $\mathbf{x}^\star$, referred as *informative*, and a random initialization, referred as *uninformative*. When these different initializations provide a different iterative fixed point we mark the presence of the hard phase.

Fig. 1 depicts the compressed sensing problem with linear (left) and ReLU (right) generative priors. We depict the phase transitions defined above. On the left hand side we compare to the algorithmic phase transition known from [14] for sparse separable prior with fraction $1-\rho$ of zero entries and ρ of Gaussian entries of zero mean presenting an algorithmically hard phase for $\rho < \alpha < \alpha_{\text{alg}}^{\text{sparse}}(\rho)$. In case of compressed sensing with linear generative prior we do not observe any hard phase and exact recovery is possible for $\alpha \geq \min(\rho, 1)$ due to invertibility (or the lack of there-of) of the matrix product AW. With ReLU generative prior we have $\alpha_{\rm IT} = \min(\rho, 1/2)$ and the hard phase exists and has interesting properties: The $\rho \to \infty$ limit corresponds to the separable prior, and thus in this limit $\alpha_{\rm alg}(\rho \to \infty) = \alpha_{\rm alg}^{\rm sparse}(\rho_s = 1/2)$. Curiously we observe $\alpha_{\rm alg} > \alpha_{\rm IT}$ for all $\rho \in (0, \infty)$ except at $\rho = 1/2$. Moreover the size of the hard phase is very small for $\rho < 1/2$ when compared to the one for compressed sensing with separable priors, suggesting that exploring structure in terms of generative models might be algorithmically advantageous over sparsity.

Fig. 2 depicts the phase diagram for the phase retrieval problem with linear (left) and ReLU (right) generative priors. The information-theoretic transition is the same as the one for compressed sensing, while numerical inspection shows that $\alpha_{\rm alg}^{\rm PR} > \alpha_{\rm alg}^{\rm CS}$ for all $\rho \neq 0, 1/2, 1$. In the left hand side we depict also the algorithmic transition corresponding to the sparse separable prior with non-zero components being Gaussian of zero mean, $\alpha_{\rm alg}^{\rm sparse}(\rho_s)$, as taken from [15]. Crucially, in that case the algorithmic transition to exact recovery does not fall bellow $\alpha = 1/2$ even for very small (yet finite) ρ_s, thus effectively disabling the possibility to sense compressively. In contrast, with both the linear and ReLU generative priors we observe $\alpha_{\rm alg}(\rho \to 0) \to 0$. More specifically, the theory for the linear prior implies that $\alpha_{\rm alg}/\rho(\rho \to 0) \to \alpha_{\rm alg}^{\rm sparse}(\rho_s = 1) \approx 1.128$ with the hard phase being largely reduced. Again the hard phase disappears entirely for $\rho = 1$ for the linear model and $\rho = 1/2$ for ReLU. We note that the limit $\alpha_{\rm alg}(\rho \to \infty)$ corresponds to the algorithmic transition for separable priors, concretely for the linear prior $\alpha_{\rm alg}(\rho \to \infty) \to \alpha_{\rm alg}^{\rm sparse}(\rho_s = 1) \approx 1.128$.

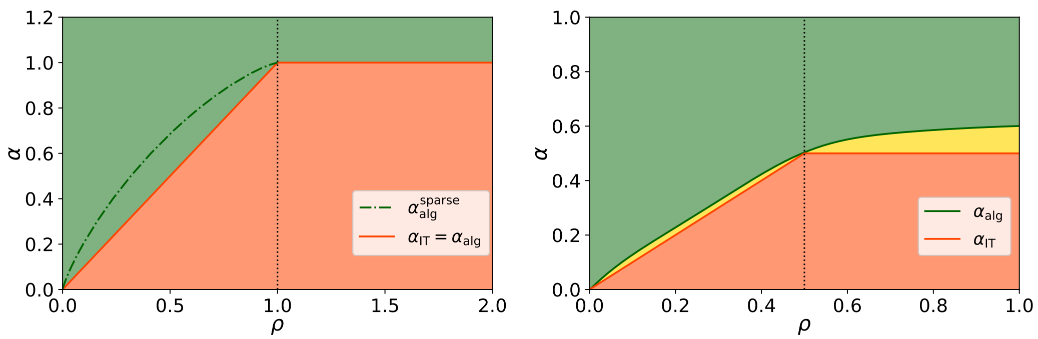

Figure 1: Phase diagrams for the compressed sensing problem with **(left)** linear generative prior and **(right)** ReLU generative prior, in the plane (ρ, α). The $\alpha_{\rm IT}$ (red line) represents the information theoretic transition for perfect reconstruction and $\alpha_{\rm alg}$ (green line) the algorithmic transition to perfect reconstruction. In the left part we depict for comparison the algorithmic phase transition for sparse separable prior $\alpha_{\rm alg}^{\rm sparse}$ (dashed-dotted green line). Colored areas correspond respectively to the *weak recovery* (orange), *hard* (yellow) and *easy* (green) phases.

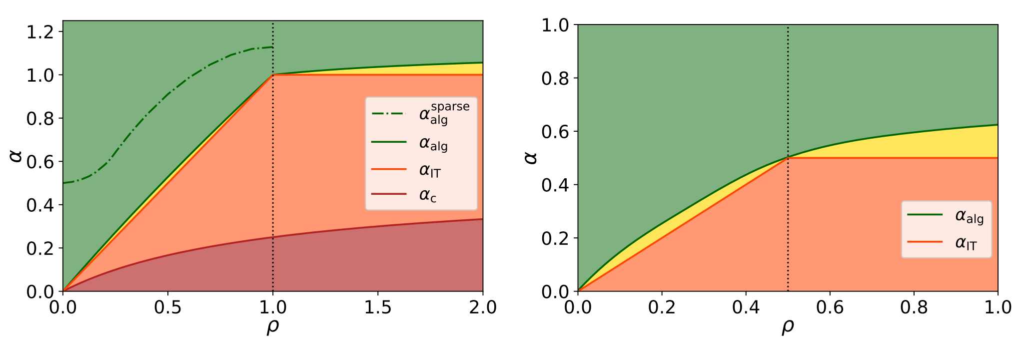

Figure 2: The same as Fig. 1 for the phase retrieval problem with **(left)** linear generative prior and **(right)** ReLU generative prior. A major result is that while with sparse separable priors (green dashed-dotted line) compressed phase retrieval is algorithmically hard for $\alpha < 1/2$, with generative priors compressed phase retrieval is tractable down to vanishing α (green line). In the left part we depict additionally the *weak recovery* transition $\alpha_c = \rho/[2(1+\rho)]$ (dark red line). It splits the *no-exact-recovery* phase into the *undetectable* (dark red) and the *weak-recovery* region (orange).

Acknowledgments

This work is supported by the ERC under the European Union's Horizon 2020 Research and Innovation Program 714608-SMiLe, as well as by the French Agence Nationale de la Recherche under grant ANR-17-CE23-0023-01 PAIL. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. We thank Google Cloud for providing us access to their platform through the Research Credits Application program. Additional funding is acknowledged by AM from 'Chaire de recherche sur les modèles et sciences des données', Fondation CFM pour la Recherche-ENS.

References

[1] David L Donoho et al. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[2] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006.

[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[4] Eric W Tramel, Andre Manoel, Francesco Caltagirone, Marylou Gabrié, and Florent Krzakala. Inferring sparsity: Compressed sensing using generalized restricted Boltzmann machines. In *2016 IEEE Information Theory Workshop (ITW)*, pages 265–269. IEEE, 2016.

[5] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 537–546. JMLR. org, 2017.

[6] Andre Manoel, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Multi-layer generalized linear estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2098–2102. IEEE, 2017.

[7] Paul Hand and Vladislav Voroninski. Global guarantees for enforcing deep generative priors by empirical risk. In *Conference On Learning Theory*, pages 970–978, 2018.

[8] Alyson K Fletcher, Sundeep Rangan, and Philip Schniter. Inference in deep networks in high dimensions. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1884–1888. IEEE, 2018.

[9] Paul Hand, Oscar Leong, and Vlad Voroninski. Phase retrieval under a generative prior. In *Advances in Neural Information Processing Systems*, pages 9136–9146, 2018.

[10] Dustin G Mixon and Soledad Villar. Sunlayer: Stable denoising with generative networks. *arXiv preprint arXiv:1803.09319*, 2018.

[11] Benjamin Aubin, Bruno Loureiro, Antoine Maillard, Florent Krzakala, and Lenka Zdeborová. The spiked matrix model with generative priors. *arXiv preprint arXiv:1905.12385*, 2019.

[12] Emmanuel J Candes, Yonina C Eldar, Thomas Strohmer, and Vladislav Voroninski. Phase retrieval via matrix completion. *SIAM review*, 57(2):225–251, 2015.

[13] Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, pages 2796–2804, 2013.

[14] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová. Statistical-Physics-Based Reconstruction in Compressed Sensing. *Physical Review X*, 2(2), May 2012.

[15] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019.

[16] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[17] Marylou Gabrié, Andre Manoel, Clément Luneau, Nicolas Macris, Florent Krzakala, Lenka Zdeborová, et al. Entropy and mutual information in models of deep neural networks. In *Advances in Neural Information Processing Systems*, pages 1821–1831, 2018.