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

Benjamin Aubin[†]   AUBIN.BENJAMIN@CEA.FR
Bruno Loureiro[†]   BRUNO.LOUREIRO@CEA.FR
Antoine Baker[⋆]   ANTOINE.BAKER@ENS.FR
Florent Krzakala[⋆]   FLORENT.KRZAKALA@ENS.FR
Lenka Zdeborová[†]   LENKA.ZDEBOROVA@CEA.FR

[†] *Université Paris-Saclay, CNRS, CEA, Institut de physique théorique, 91191, Gif-sur-Yvette, France.*

[⋆] *LPENS, CNRS & Sorbonnes Universites, Ecole Normale Superieure, PSL University, Paris, France.*

Abstract

We consider the problem of compressed sensing and of (real-valued) phase retrieval with random measurement matrix. We derive sharp asymptotics for the information-theoretically optimal performance and for the best known polynomial algorithm for an ensemble of generative priors consisting of fully connected deep neural networks with random weight matrices and arbitrary activations. We compare the performance to sparse separable priors and conclude that in all cases analysed generative priors have a smaller statistical-to-algorithmic gap than sparse priors, giving theoretical support to previous experimental observations 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.

Keywords: Phase retrieval, Compressed Sensing, Generative Models, Bayesian Inference

1. Introduction

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 Donoho (2006); Candes and Tao (2006). 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) Goodfellow et al. (2014). 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. Tramel et al. (2016a,b); Bora et al. (2017); Manoel et al. (2017); Hand and Voroninski (2018); Fletcher et al. (2018); Hand et al. (2018); Mixon and Villar (2018); Aubin et al. (2019). While this direction of research seems to have many promising applications, a systematic theory of what can be efficiently achieved still falls short of the one developed over the past decade for sparse signal processing. Our aim is therefore to dialogue with the broad program of studying how generative models can help solving inverse problems using the toolbox of statistical

physics. In this manuscript, we build on a line of work allowing for theoretical analysis in the case the measurement and the weight matrices of the prior are random Manoel et al. (2017); Reeves (2017); Fletcher et al. (2018); Gabrié et al. (2018); Aubin et al. (2019).

We employ tools originally developed in the context of statistical physics to derive precise asymptotics for the information-theoretically optimal thresholds for signal recovery and for the performance of the best known polynomial algorithm 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*. Given a set of observations $\mathbf{y} \in \mathbb{R}^n$ generated from a fixed (but unknown) signal $\mathbf{x}^\star \in \mathbb{R}^d$ as

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

the goal is to reconstruct $\mathbf{x}^\star$ from the knowledge of $\mathbf{y}$, φ and $\mathrm{A} \in \mathbb{R}^{n \times d}$. Compressed sensing and phase retrieval are particular instances of this problem, corresponding to $\varphi(x) = x$ and $\varphi(x) = |x|$ respectively. Two key questions in these inverse problems are a) how many observations n are required for theoretically reconstructing the signal $\mathbf{x}^\star$, and b) how this can be done in practice - i.e. to find an efficient algorithm for reconstruction. Signal structure plays an important role in the answer to both these questions, and have been the subject of intense investigation in the literature. A typical situation is to consider signals admitting a low-dimensional representation, such as sparse signals, for which $k - d$ of the d components of $\mathbf{x}^*$ are exactly zero, see e.g. Candes et al. (2015); Netrapalli et al. (2013).

In this manuscript, we consider instead structured signals drawn from a generative model $\mathbf{x}^\star = G(\mathbf{z})$, where $\mathbf{z} \in \mathbb{R}^k$ is a low-dimensional latent representation of $\mathbf{x}^\star$. In particular, we will focus in generative multi-layer neural networks, and in order to provide a sharp asymptotic theory, we will restrict the analysis to an ensemble of random networks with known random weights:

$$\mathbf{x}^\star = G\left(\mathbf{z}\right) = \sigma^{(L)}\left(\mathrm{W}^{(L)}\sigma^{(L-1)}\left(\mathrm{W}^{(L-1)}\cdots\sigma^{(1)}\left(\mathrm{W}^{(1)}\mathbf{z}\right)\cdots\right)\right), \tag{2}$$

where $\sigma^{(l)} : \mathbb{R} \to \mathbb{R}$, $1 \leq l \leq L$ are component-wise non-linearities. As aforementioned, we take $\mathrm{A} \in \mathbb{R}^{n \times d}$ and $\mathrm{W}^{(l)} \in \mathbb{R}^{k_l \times k_{l-1}}$ to have i.i.d. Gaussian entries with zero means and variances $1/d$ and $1/k_{l-1}$ respectively, and focus on the high-dimensional regime defined by taking $n, d, k_l \to \infty$ while keeping the measurement rate $\alpha = n/d$ and the layer-wise aspect ratios $\beta_l = k_{l+1}/k_l$ constant. We stress that in this regime the depth L is of order one when compared to the width of the generative network, which scales with the input dimension d. With this observation in mind, we adopt the standard terminology in machine learning of denoting networks with $L > 1$ as *deep*. To provide a comparison with previous results for sparse signals, it is useful to define the total compression factor $\rho = k/d$. We note, however, that the comparison between generative and sparse priors herein is not based on a quantitative comparison between the reconstruction estimation errors. Indeed, since data is generated differently in both cases, such a comparison would make little sense. Instead, we compare qualitative properties of the phase diagrams, taking as a surrogate for algorithmic hardness the size of the statistical-to-algorithmic gap in these two different reconstruction problems. Our results hold for latent variables drawn from an arbitrary separable distribution $\mathbf{z} \sim P_z$, and for arbitrary activations $\sigma^{(l)}$, although for concreteness we present results for $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathrm{I}_k)$ and $\sigma^{(l)} \in \{\mathrm{linear}, \mathrm{ReLU}\}$, as it is commonly the case in practice with GANs.

Previous results on sparsity: Sparsity is probably the most widely studied type of signal structure in linear estimation and phase retrieval. It is thus instructive to recall the main results for sparse

signal reconstruction in these inverse problems in the high-dimensional regime with random measurement matrices studied in this manuscript. Optimal statistical and algorithmic thresholds have been established non-rigorously using the replica-method in a series of works Wu and Verdú (2012); Krzakala et al. (2012a); Reeves and Gastpar (2012); Zdeborová and Krzakala (2016). Later the information theoretic results, as well as the corresponding minimum mean squared error (MMSE), has been rigorously proven in Barbier et al. (2016); Reeves and Pfister (2016); Barbier et al. (2019). So far, the best known polynomial time algorithm in this context is the *Approximate Message Passing* (AMP) algorithm, the new avatar of the mean-field approach pioneered in statistical mechanics Mézard et al. (1987), that has been introduced in Donoho et al. (2009); Rangan (2011); Krzakala et al. (2012a); Schniter and Rangan (2014); Metzler et al. (2017) for these problems, and can be rigorously analysed Bayati and Montanari (2011). For both (noiseless) compressed sensing and phase retrieval, the information theoretic limit for a perfect signal recovery is given by $\alpha > \alpha_{\mathrm{IT}} = \rho_s$, 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. No polynomial algorithm achieving better performance for these problems is known. Strikingly, as discussed in Barbier et al. (2019), the behaviour 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 $\alpha_{\mathrm{alg}}^{\mathrm{sparse}} \to 1/2$ as $\rho_s \to 0$. As a consequence, *no efficient approach to real-valued compressed phase retrieval with small but order one ρ_s in the high-dimensional limit is known.*

Summary of results: In this work, we replace the sparse prior by the multi-layer generative model introduced in eq. (2). Our main contribution is specifying the interplay between the number of measurements needed for exact reconstruction of the signal, parametrised 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, parametrised by $\rho = k/d$. While the number of unknown latent variables is the same in both cases if $\rho = \rho_s$, the upshot is that generative models offer algorithmic advantages over sparsity. More precisely:

(a) We analyse the minimum mean square error (MMSE) of the optimal Bayesian estimator for the compressed sensing and phase retrieval problems with generative priors of arbitrary depth, choice of activation and prior distribution for the latent variable. We derive sufficient conditions for the existence of an *undetectable phase* in which better-than-random estimation of $\mathbf{x}^\star$ is impossible, and characterise in full generality the threshold α_c beyond which partial signal recovery becomes statistically possible.

(b) Fixing our attention on the natural choices of activations $\sigma \in \{\mathrm{linear}, \mathrm{ReLU}\}$, we establish the threshold α_{IT} above which perfect signal reconstruction is theoretically possible. This threshold can be intuitively understood with a simple counting argument.

(c) We analyse the performance of the associated *Approximate Message Passing* algorithm Manoel et al. (2017), conjectured to be the best known polynomial time algorithm in this setting. This allows us to establish the algorithmic threshold α_{alg} below which no known algorithm is able to perfectly reconstruct $\mathbf{x}^\star$.

As expected, the thresholds $\{\alpha_c, \alpha_{\mathrm{IT}}, \alpha_{\mathrm{alg}}\}$ are functions of the compression factor ρ, the number of layers L, the aspect ratios $\{\beta_l\}_{l=1}^L$ and the activation functions. In particular, for a fixed archi-

texture we find that the algorithmic gap $\Delta_{\mathrm{alg}} = \alpha_{\mathrm{alg}} - \alpha_{\mathrm{IT}}$ is drastically reduced with the depth L of the generative model, beating the algorithmic hindrance identified in Barbier et al. (2019) for compressive phase retrieval with sparse encoding.

2. Information theoretical analysis

2.1. Performance of the Bayes-optimal estimator

In our analysis we assume that the model generating the observations $\mathbf{y} \in \mathbb{R}^n$ is known. Therefore, the optimal estimator minimising the mean-squared-error in our setting is given by the Bayesian estimator

$$\hat{\mathbf{x}}^{\mathrm{opt}} = \underset{\hat{\mathbf{x}}}{\mathrm{argmin}} ||\hat{\mathbf{x}} - \mathbf{x}^\star||_2^2 = \mathbb{E}_{P(\mathbf{x}|\mathbf{y})}\left[\mathbf{x}\right] . \tag{3}$$

The posterior distribution of the signal given the observations is in general given by:

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{\mathcal{Z}(\mathbf{y})} P_x(\mathbf{x}) \prod_{\mu=1}^{n} \delta\left(y^\mu - \varphi\left(\sum_{j=1}^{d} \mathrm{A}_j^\mu x_j\right)\right) , \tag{4}$$

where the normalisation $\mathcal{Z}(\mathbf{y})$ is known as the *partition function*, and φ is the nonlinearity defining the estimation problem, e.g. $\varphi(x) = |x|$ for phase retrieval and $\varphi(x) = x$ for linear estimation. We note that the presented approach generalizes straightforwardly to account for the presence of noise, but we focus in this paper on the analysis of the noiseless case. For the generative model in eq. (2), the prior distribution P_x reads

$$P_x(\mathbf{x}) = \int_{\mathbb{R}^k} \mathrm{d}\mathbf{z}\, P_z(\mathbf{z}) \prod_{l=1}^{L} \int_{\mathbb{R}^{k_l}} \mathrm{d}\mathbf{h}^{(l)}\, P_{\mathrm{out}}^{(l)}\left(\mathbf{h}^{(l+1)} \middle| \mathbf{W}^{(l)}\mathbf{h}^{(l)}\right) , \tag{5}$$

where for notational convenience we denoted $\mathbf{x} \equiv \mathbf{h}^{(L+1)}$, $\mathbf{z} \equiv \mathbf{h}^{(1)}$ and defined the likelihoods $P_{\mathrm{out}}^{(l)}$ parametrising the output distribution of each layer given its input. As before, this Bayesian treatment also accounts for stochastic activation functions, even though we focus here on deterministic ones.

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 model parameters. Our main results are based on the line of works comparing, on one hand, the information-theoretically best possible reconstruction, analysing the ideal Bayesian inference decoder, regardless of the computation cost, and on the other, the best reconstruction using the most efficient known polynomial algorithm - the approximate message passing.

Our analysis builds upon the statistical physics inspired multi-layer formalism introduced in Manoel et al. (2017), who showed using the cavity and replica methods that the minimum mean-squared-error achieved by the Bayes-optimal estimator defined in eq. (3) can be written, in the limit of $n, d \to \infty$ and $\alpha = n/d = \Theta(1)$ for a generic prior distribution P_x as

$$\mathrm{mmse}(\alpha) = \lim_{d\to\infty} \frac{1}{d}\mathbb{E}||\hat{\mathbf{x}}^{\mathrm{opt}} - \mathbf{x}^\star||_2^2 = \rho_x - q_x^\star \tag{6}$$

where ρ_x is the second moment of P_x and the scalar parameter $q_x^\star \in [0, \rho_x]$ is the solution of the following *free energy* extremisation problem

$$\Phi = -\lim_{d \to \infty} \frac{1}{d} \mathbb{E}_y \log \mathcal{Z}(\mathbf{y}) = \underset{q_x, \hat{q}_x}{\mathbf{extr}} \left\{ \frac{1}{2} \hat{q}_x q_x - \alpha \Psi_y(q_x) - \Psi_x(\hat{q}_x) \right\}, \tag{7}$$

with the so-called potentials (Ψ_y, Ψ_x) given by

$$\begin{aligned} \Psi_y(t) &= \mathbb{E}_\xi \left[\int dy \, \mathcal{Z}_y\left(y; \sqrt{t}\xi, t\right) \log \mathcal{Z}_y\left(y; \sqrt{t}\xi, t\right) \right], \\ \Psi_x(r) &= \lim_{d \to \infty} \frac{1}{d} \mathbb{E}_\xi \left[\mathcal{Z}_x(\sqrt{r}\xi, r) \log \mathcal{Z}_x(\sqrt{r}\xi, r) \right], \end{aligned} \tag{8}$$

where $\xi \sim \mathcal{N}(0, 1)$ and $\mathcal{Z}_y, \mathcal{Z}_x$ are the normalisations of the auxiliary distributions

$$\begin{aligned} Q_y(x; y, \omega, V) &= \frac{1}{\mathcal{Z}_y(y; \omega, V)} \frac{e^{-\frac{1}{2V}(x-\omega)^2}}{\sqrt{2\pi V}} \delta(y - \varphi(x)), \\ Q_x(\mathbf{x}; B, A) &= \frac{P_x(\mathbf{x})}{\mathcal{Z}_x(B, A)} \prod_{j=1}^{d} e^{-\frac{A}{2}x_j^2 + Bx_j}. \end{aligned} \tag{9}$$

Note that this expression is valid for arbitrary distribution P_x, as long as the limit in Ψ_x is well-defined. In particular, it reduces to the known result in Krzakala et al. (2012b); Barbier et al. (2019) when P_x factorises. In principle, for correlated P_x such as in the generative model of eq. (5) computing Ψ_x is itself a hard problem. However, we can see eq. (5) as a chain of generalised linear models. In the limit where $k_l \to \infty$ with $\rho = k/d = \Theta(1)$, $L = \Theta(1)$ and $\beta_l = k_{l+1}/k_l = \Theta(1)$ we can apply the observation above iteratively, layer-wise, up to the input layer for which P_z factorises - and is easy to compute. This yields Manoel et al. (2017)

$$\begin{aligned} \Phi = \underset{q_x, \hat{q}_x, \{q_l, \hat{q}_l\}}{\mathbf{extr}} &\left\{ -\frac{1}{2} \hat{q}_x q_x - \frac{\rho}{2} \sum_{l=1}^{L} \beta_l q_l \hat{q}_l + \alpha \Psi_y(q_x) \right. \\ &\left. + \rho \sum_{l=2}^{L} \beta_l \Psi_{\text{out}}^{(l)}(\hat{q}_l, q_{l-1}) + \Psi_{\text{out}}^{(L+1)}(\hat{q}_x, q_L) + \rho \Psi_z(\hat{q}_z) \right\}, \tag{10} \end{aligned}$$

where we have introduced the additional potentials $(\Psi_{\text{out}}, \Psi_z)$

$$\begin{aligned} \Psi_{\text{out}}^{(l)}(r, s) &= \mathbb{E}_{\xi, \eta} \left[\mathcal{Z}_{\text{out}}^{(l)}(\sqrt{r}\xi, r, \sqrt{s}\xi, \rho_{l-1} - s) \log \mathcal{Z}_{\text{out}}^{(l)}(\sqrt{r}\xi, r, \sqrt{s}\xi, \rho_{l-1} - s) \right], \\ \Psi_z(t) &= \mathbb{E}_\xi \left[\mathcal{Z}_z(\sqrt{t}\xi, t) \log \mathcal{Z}_z(\sqrt{t}\xi, t) \right], \end{aligned} \tag{11}$$

defined in terms of the following auxiliary distributions

$$\begin{aligned} Q_{\text{out}}^{(l)}(x, z; B, A, \omega, V) &= \frac{e^{-\frac{A}{2}x^2 + Bx}}{\mathcal{Z}_{\text{out}}(B, A, \omega, V)} \frac{e^{-\frac{1}{2V}(z-\omega)^2}}{\sqrt{2\pi V}} P_{\text{out}}^{(l)}(x|z), \\ Q_z(z; B, A) &= \frac{e^{-\frac{A}{2}z^2 + Bz}}{\mathcal{Z}_z(B, A)} P_z(z), \end{aligned} \tag{12}$$

and with ρ_l the second moment of the hidden variable $\mathbf{h}^{(l)}$.

These predictions, that have also been derived with different heuristics in Reeves (2017), were rigorously proven for two-layers in Gabrié et al. (2018), while deeper architectures requires additional assumptions on the concentration of the free energies to be under a rigorous control. Eq. (10) thus reduces the asymptotics of the high-dimensional estimation problem to a low-dimensional extremisation problem over the $2(L+1)$ variables $(q_x, \hat{q}_x, \{q_l, \hat{q}_l\}_{l=1}^L)$, allowing for a mathematically sound and rigorous investigation. These parameters are also known as the *overlaps*, since they parametrise the overlap between the Bayes-optimal estimator and ground-truth signal at each layer. Solving eq. (7) provides two important statistical thresholds: the *weak recovery* threshold α_c above which better-than-random (i.e. mmse $< \rho_x$) reconstruction becomes theoretically possible and the *perfect reconstruction* threshold, above which perfect signal recovery (i.e. when mmse $= 0$) becomes possible.

Interestingly, the free energy eq. (10) also provides information about the algorithmic hardness of the problem. The above extremisation problem is closely related the state evolution of the AMP algorithm for this problem, as derived in Manoel et al. (2017), and generalized in Fletcher et al. (2018). It is conjectured to provide the best polynomial time algorithm for the estimation of $\mathbf{x}^\star$ in our considered setting. Specifically, the algorithm reaches a mean-squared error that corresponds to the local extremiser reached by gradient descent in the function (10) starting with uninformative initial conditions.

While so far we summarised results that follow from previous works, these results were up to our knowledge not systematically evaluated and analysed for the linear estimation and phase retrieval with generative priors. This analysis and its consequences is the object of the rest of this paper and constitutes the original contributions of this work.

2.2. Weak recovery threshold

Solutions for the extremisation in eq. (10) can be found by solving the fixed point equations, obtained by taking the gradient of eq. (10) with respect of the parameters $(q_x, \hat{q}_x, \{q_l, \hat{q}_l\}_{l=1}^L)$:

$$
\begin{cases}
\hat{q}_x = \alpha \Lambda_y (q_x) \\
\hat{q}_L = \beta_L \Lambda_{\text{out}} (\hat{q}_x, q_L) \\
\hat{q}_{L-1} = \beta_{L-1} \Lambda_{\text{out}} (\hat{q}_L, q_{L-1}) \\
\quad \vdots \\
\hat{q}_l = \beta_l \Lambda_{\text{out}} (\hat{q}_{l+1}, q_l) \\
\quad \vdots \\
\hat{q}_z = \beta_1 \Lambda_{\text{out}} (\hat{q}_2, q_z)
\end{cases}
\qquad
\begin{cases}
q_x = \Lambda_x (\hat{q}_x, q_L) \\
q_L = \Lambda_x (\hat{q}_L, q_{L-1}) \\
\quad \vdots \\
q_l = \Lambda_x (\hat{q}_l, q_{l-1}) \\
\quad \vdots \\
q_z = \Lambda_z (\hat{q}_z)
\end{cases}
\tag{13}
$$

where $\Lambda_y(t) = 2 \, \partial_t \Psi_y(t)$, $\Lambda_z(t) = 2 \, \partial_t \Psi_z(t)$, $\Lambda_x(t) = 2 \, \partial_r \Psi_{\text{out}}(r, s)$, $\Lambda_{\text{out}}(t) = 2 \, \partial_s \Psi_{\text{out}}(r, s)$.

The weak recovery threshold α_c is defined as the value above which one can estimate $\mathbf{x}^\star$ better than a random draw from the prior P_x. In terms of the mmse it is defined as

$$
\alpha_c = \underset{\alpha \geq 0}{\operatorname{argmax}} \{ \text{mmse}(\alpha) = \rho_x \}.
\tag{14}
$$

From eq. (6), it is clear that an uninformative solution mmse $= \rho_x$ of eq. (10) corresponds to a fixed point $q_x = 0$. For both the phase retrieval and linear estimation, evaluating the right-hand side of

eqs. (13) at $q_x = 0$ we can see that $\hat{q}_x^\star = 0$ is a fixed point if σ is an odd function and if

$$\mathbb{E}_{P_z}[z] = 0, \qquad \text{and} \qquad \mathbb{E}_{Q_{\text{out}}^{(l),0}}[x] = 0, \tag{15}$$

where $Q_{\text{out}}^{(l),0}(x, z) = Q_{\text{out}}^{(l)}(x, z; 0, 0, 0, \rho_{l-1})$. These conditions reflect the intuition that if the prior P_z or the likelihoods $P_{\text{out}}^{(l)}$ are biased towards certain values, this knowledge helps the statistician estimating better than a random guess. If these conditions are satisfied, then α_c can be obtained as the point for which the fixed point $q_x = 0$ becomes unstable. The stability condition is determined by the eigenvalues of the Jacobian of eqs. (13) around the fixed point $(q_x^\star, \hat{q}_x^\star, \{q_l^\star, \hat{q}_l^\star\}_{l=1}^L) = 0$. More precisely, the fixed point becomes unstable as soon as one eigenvalue of the Jacobian is bigger than one. Expanding the update functions around the fixed point and using the conditions in eq. (15),

$$\Lambda_y(t) \underset{t \ll 1}{=} \frac{1}{\rho_x^2} \int dy \, \mathcal{Z}_y(y; 0, \rho_x) \left(\mathbb{E}_{Q_y^0}[\rho_x - x^2] \right)^2 t + \mathcal{O}\left(t^{3/2} \right),$$

$$\Lambda_x^{(l)}(r, s) \underset{r,s \ll 1}{=} \left(\mathbb{E}_{Q_{\text{out}}^{(l),0}}[x^2] \right)^2 r + \frac{1}{\rho_{l-1}^2} \left(\mathbb{E}_{Q_{\text{out}}^{(l),0}}[xz] \right)^2 s + \mathcal{O}\left(r^{3/2}, s^{3/2} \right), \tag{16}$$

$$\Lambda_{\text{out}}^{(l)}(r, s) \underset{r,s \ll 1}{=} \left(\mathbb{E}_{Q_{\text{out}}^{(l),0}}[xz] \right)^2 r + \frac{1}{\rho_{l-1}^2} \left(\mathbb{E}_{Q_{\text{out}}^{(l),0}}[z^2] - \rho_{l-1} \right)^2 s + \mathcal{O}\left(r^{3/2}, s^{3/2} \right),$$

$$\Lambda_z(t) \underset{t \ll 1}{=} \left(\mathbb{E}_{P_z}[z^2] \right)^2 t + \mathcal{O}\left(t^{3/2} \right).$$

For a generative prior with depth L, the Jacobian is a cumbersome sparse $(L+1) \times (L+1)$ matrix, with all the entries given by the six partial derivatives above. For the sake of conciseness we only write it here for $L = 1$:

$$\begin{pmatrix} 0 & \left(\mathbb{E}_{Q_{\text{out}}^0}[x^2] \right)^2 & \frac{1}{\rho_z^2} \left(\mathbb{E}_{Q_{\text{out}}^0}[xz] \right)^2 & 0 \\ \frac{\alpha}{\rho_x^2} \int dy \, \mathcal{Z}_y^0 \left(\mathbb{E}_{Q_y^0}[\rho_x - x^2] \right)^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \left(\mathbb{E}_{P_z}[z^2] \right)^2 \\ 0 & \beta \left(\mathbb{E}_{Q_{\text{out}}^0}[xz] \right)^2 & \frac{\beta}{\rho_z^2} \left(\mathbb{E}_{Q_{\text{out}}^0}[z^2] - \rho_z \right)^2 & 0 \end{pmatrix}. \tag{17}$$

Note that this holds for any choice of $P_{\text{out}}^{(l)}$ and latent space distribution P_z, as long as conditions eq. (15) hold. For the phase retrieval with a linear generative model for instance $P^{(l)}(x|z) = \delta(x - z)$, we find $\alpha_c = \frac{1}{2} \frac{1}{1 + \rho^{-1}}$. For a linear network of depth L this generalises to

$$\alpha_c = \frac{1}{2} \left(1 + \sum_{l=1}^L \prod_{k=0}^{l-1} \beta_{L-k} \right)^{-1}. \tag{18}$$

The linear estimation problem has exactly the same threshold, but without the global $1/2$ factor. Since $\rho, \beta_l \geq 0$, it is clear that α_c is decreasing in the depth L of the network. This analytical formula is verified by numerically solving eqs. (13), see Figs. 3 and 4. For other choices of activation satisfying condition (15) (e.g. the sign function), we always find that depth helps in the weak recovery of the signal.

2.3. Perfect recovery threshold

We now turn our attention to the perfect recovery threshold, above which perfect signal reconstruction becomes statistically possible. Formally, it can be defined as

$$\alpha_{\text{IT}} = \underset{\alpha \geq 0}{\text{argmin}}\{\text{mmse}(\alpha) = 0\}, \tag{19}$$

and corresponds to the global minimum of the free energy in eq. (10). Numerically, the perfect recovery threshold is found by solving the fixed point equations (13) from an informed initialisation $q_x \approx \rho_x$, corresponding to mmse ≈ 0 according to eq. (6). The resulting fixed point is then checked to be a minimiser of the free energy eq. (10). Different from α_c, it cannot be computed analytically for an arbitrary architecture. However, for the compressed sensing and phase retrieval problems with $\sigma \in \{\text{linear}, \text{ReLU}\}$ generative priors, α_{IT} can be analytically computed by generalising a simple argument based on the invertibility of the linear system of equations at each layer, originally used in the usual compressive sensing Candes and Tao (2006); Tao (2009).

First, consider the linear estimation problem with a deep linear generative prior, i.e. $\mathbf{y} = A\mathbf{x}^\star \in \mathbb{R}^n$ with $\mathbf{x}^\star = W^{(L)} \dots W^{(1)}\mathbf{z} \in \mathbb{R}^d$ and $A, \{W^{(l)}\}_{l=1}^L$ i.i.d. Gaussian matrices, that are full rank with high probability. For $n > d$, the system $\mathbf{y} = A\mathbf{x}^\star$ is overdetermined as there are more equations than unknowns. Hence the information theoretical threshold has to verify $\alpha_{\text{IT}} = \frac{n_{\text{IT}}}{d} \leq 1$. For $L = 0$ (i.e. $\mathbf{x}^\star$ is Gaussian i.i.d.), we have exactly $\alpha_{\text{IT}}^{(0)} = 1$ as the prior does not give any additional information for solving the linear system. For $L \geq 1$ though, at each level $l \in [1 : L]$, we need to solve successively $\mathbf{h}^{(l)} \in \mathbb{R}^{k_l}$ in the linear system $\mathbf{y} = AW^{(L)} \cdots W^{(l)}\mathbf{h}^{(l)}$. Again as $AW^{(L)} \cdots W^{(l)} \in \mathbb{R}^{n \times k_l}$, if $n > k_l$ the system is over-constrained. Hence the information theoretical threshold for this equation is such that $\forall l \in [1 : L], n_{\text{IT}}^{(l)} \leq k_l \Leftrightarrow \alpha_{\text{IT}}^{(l)} \leq \prod_{k=1}^{l} \frac{1}{\beta_{L-k+1}}$. And note that $\rho \equiv \prod_{k=1}^{L} \frac{1}{\beta_{L-k+1}}$. Hence, the information theoretical threshold is obtained by taking the smallest of the above values $\alpha_{\text{IT}}^{(l)}$:

$$\alpha_{\text{IT}} = \min_{l \in [0:L]} \alpha_{\text{IT}}^{(l)} = \min\left(1, \left\{\prod_{k=1}^{l} \frac{1}{\beta_{L-k+1}}\right\}_{l=1}^{L-1}, \rho\right). \tag{20}$$

This result generalises to the real-valued phase retrieval problem. First, we note that by the data processing inequality taking $\mathbf{y} = |A\mathbf{x}^\star|$ cannot increase the information about $\mathbf{x}^\star$, and therefore the transition in phase retrieval cannot be *better* than for compressed sensing. Secondly, an inefficient algorithm exists that achieve the same performance as compressed sensing for the real valued phase retrieval: one just needs to try all the possible 2^m assignments for the sign, and then solve the corresponding compressed sensing problem. This strategy that will work as soon as the compressed sensing problem is solvable. Eq. (20) is thus valid for the real phase retrieval problem as well.

One can finally generalise this analysis for a non-linear generative prior with ReLU activation at each layer, i.e. $\mathbf{x}^\star = \text{relu}\left(W^{(L)}\text{relu}\left(\cdots W^{(1)}\mathbf{z}\right)\cdots\right)$. Noting that on average $\mathbf{x}$ has half of zero entries and half of i.i.d. Gaussian entries, the system can be reorganised and simplified $\mathbf{y} = \tilde{A}\tilde{\mathbf{x}}$, with $\tilde{\mathbf{x}} \in \mathbb{R}^{d/2}$ the extracted vector of $\mathbf{x}$ with on average $d/2$ strictly positive entries and the corresponding reduced matrix $\tilde{A} \in \mathbb{R}^{n \times d/2}$, is over-constrained for $n > d/2$ and hence the information theoretical

threshold verifies $\alpha_{\text{IT}} = \frac{n_{\text{IT}}}{d} \leq \frac{1}{2}$. Noting that this observation remains valid for generative layers, we will have on average at each layer an input vector $\mathbf{h}^{(l)}$ with half of zero entries and half of Gaussian distributed entries - except at the very first layer for which the input $\mathbf{z} \in \mathbb{R}^k$ is dense. Repeating the above arguments yields the following perfect recovery threshold

$$\alpha_{\text{IT}} = \min\left(\frac{1}{2}, \left\{ \frac{1}{2} \prod_{k=1}^{l} \frac{1}{\beta_{L-k+1}} \right\}_{l=1}^{L-1}, \rho \right). \tag{21}$$

for both the linear estimation and phase retrieval problems. Both these results are consistent with the solution of the saddle-point eqs. (13) with a informed initialisation, see Figs. 4-6.

2.4. Algorithmic threshold

The discussion so far focused on the statistical limitations for signal recovery, regardless of the cost of the reconstruction procedure. In practice, however, one is concerned with the algorithmic costs for reconstruction. In the high-dimensional regime we are interested, where the number of observations scale with the number of parameters in the model, only (low)-polynomial time algorithms are manageable in practice. Remarkably, the formula in eq. (10) also provides useful information about the algorithmic hindrances for the inverse problems under consideration. Indeed, with a corresponding choice of iteration schedule and initialisation, the fixed point equations eq. (10) are identical to the state evolution describing the asymptotic performance of an associated *Approximate Message Passing* (AMP) algorithm Manoel et al. (2017); Fletcher et al. (2018). Moreover, the AMP aforementioned is the *best known* polynomial time algorithm for the estimation problem under consideration, and it is conjectured to be the optimal polynomial algorithm in this setting.

The AMP state evolution corresponds to initialising the overlap parameters $(q_x, q_l) \approx 0$ and updating, at each time step t the hat variables $\hat{q}_x^t = \alpha \Lambda_y(q_x^t)$ before the overlaps $q_x^{t+1} = \Lambda_x(\hat{q}_x^t, q_L^t)$, etc. In Fig. 1 we illustrate this equivalence by comparing the MSE obtained by iterating eqs. (13) with the averaged MSE obtained by actually running the AMP algorithm from Manoel et al. (2017) for a specific architecture and implemented with the tramp python package Baker et al. (2020). In particular even though the AMP state evolution is not yet rigorously proven, we see a strong agreement of our analytical results with AMP simulations.

Note that, by construction, the performance of the Bayes-optimal estimator corresponds to the global minimum of the scalar potential in eq. (10). If this potential is convex, eqs. (13) will converge to the global minimum, and the asymptotic performance of the associated AMP algorithm will be optimal. However, if the potential has also a local minimum, initialising the fixed point equations will converge to the different minima depending on the initialisation. In this case, the MSE associated to the AMP algorithm (corresponding to the local minimum) differs from the Bayes-optimal one (by construction the global minimum). In the later setting, we define the *algorithmic threshold* as the threshold above which AMP is able to perfectly reconstruct the signal - or equivalently for which mmse $= 0$ when eqs. (13) are iterated from $q_x^{t=0} = q_l^{t=0} = \epsilon \ll 1$. Note that by definition $\alpha_{\text{IT}} < \alpha_{\text{alg}}$, and we refer to $\Delta_{\text{alg}} = \alpha_{\text{alg}} - \alpha_{\text{IT}}$ as the algorithmic gap. See Fig. 2 for an illustration of the evolution of the free energy landscape for increasing α.

Studying the existence of an algorithmic gap for the linear estimation and phase retrieval problems, and how it depends on the architecture and depth of the generative prior, is the subject of the next section.
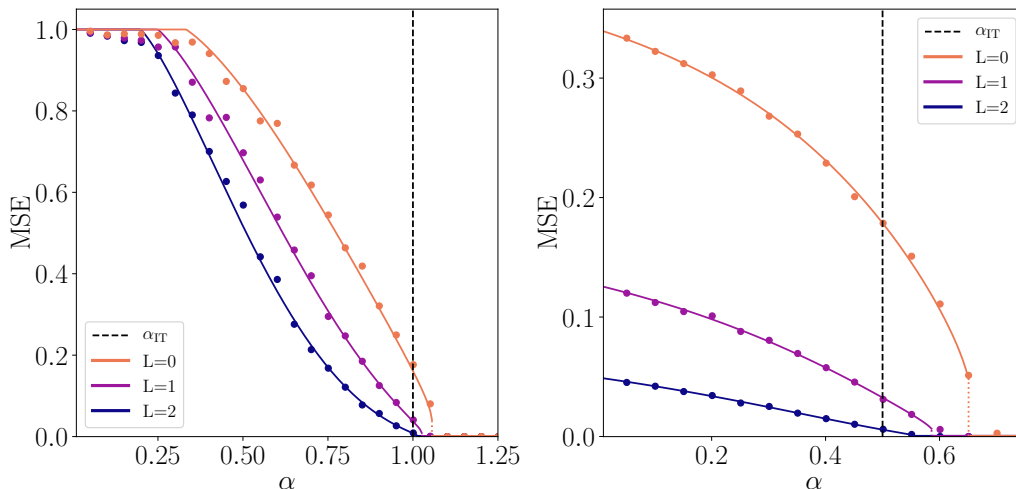
Figure 1: Mean squared error obtained by running the AMP algorithm (dots) from Manoel et al. (2017) and implemented with the `tramp` package Baker et al. (2020), for $d = 2.10^3$ averaged on 10 samples, compared to the MSE obtained from the state evolution eqs. (13) with uninformative initialisation $q_x = q_l \approx 0$ (solid line) for the phase retrieval problem with linear (**left**) and relu (**right**) generative prior networks. Different curves correspond to different depths L, with fixed $\rho = 2$ and layer-wise aspect ratios $\beta_l = 1$. The dashed vertical line corresponds to α_{IT}. To illustrate for instance in the linear case (**left**), $(\alpha_c^{L=0}, \alpha_c^{L=1}, \alpha_c^{L=2}) = (1/3, 1/4, 1/5)$, $\alpha_{\text{IT}} = 1$ and $(\alpha_{\text{alg}}^{L=0}, \alpha_{\text{alg}}^{L=1}, \alpha_{\text{alg}}^{L=2}) = (1.056, 1.026, 1.011)$.

3. Phase diagrams

In this section we summarise the previous discussions in plots in the (ρ, α)-plane, hereafter named *phase diagrams*. Phase diagrams quantify the quality of signal reconstruction for a fixed architecture $(\beta_1, \ldots, \beta_{L-1})$ [1] as a function of the compression ρ. Moreover, it allows a direct visual comparison between the phase diagram for a sparse Gaussian prior and the multi-layer generative prior. For both the phase retrieval and compressed sensing problems we distinguish the following regions of parameters limited by the thresholds of sec. 2:

- *Undetectable* region where the best achievable error is as bad as a random guess from the prior as if no measurement **y** were available. Corresponds to $\alpha < \alpha_c$.

- *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. Corresponds to $\alpha_c < \alpha < \alpha_{\text{IT}}$.

- *Hard* region where exact reconstruction can be achieved information-theoretically, but no efficient algorithm achieving it is known. Corresponds to $\alpha_{\text{IT}} < \alpha < \alpha_{\text{alg}}$

---

1. Note that β_L is fixed from the knowledge of $(\rho, \beta_1, \ldots, \beta_{L-1})$.

- The so-called *easy* region where the aforementioned AMP algorithm for this problem achieves exact reconstruction of the signal. Corresponds to $\alpha > \alpha_{\text{alg}}$.

As already explained, 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 $q_x, q_l = 0$. This is an extremiser of the free energy (7) when the prior P_z has zero mean and the non-linearity φ is an even function. This condition is satisfied for both the linear estimation and the phase retrieval problem with linear generative priors that leads to zero-mean distributions on the components of the signal, but is not achieved for a generative prior with ReLU activation, since it biases estimation. 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 derived in eq. (18). 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}.

The statistical and algorithmic thresholds defined above admit an alternative and instructive description in terms of free energy landscape, see Fig. 2. Consider a fixed ρ: for small α the free energy eq. (10) has a single global minimum with small overlap (high MSE) with the ground truth solution $\mathbf{x}^\star$, referred as the *uninformative* fixed point. At a value α_{sp}, known as the *first spinodal transition*, a second local minimum appears with higher overlap (smaller MSE) with the ground truth, referred as *informative* fixed point. The later fixed point becomes a global minimum of the free energy at $\alpha_{\text{IT}} > \alpha_{\text{sp}}$, while the uninformative fixed point becomes a local minimum. A second spinodal transition occurs at α_{alg} when the informed fixed point becomes unstable. Numerically, the informed and uninformative fixed points can be reached by iterating the saddle-point equations from different initial conditions. When the two are present, the informed fixed point can be reached by iterating from $q_x \approx \rho_x$, which corresponds to a minimum overlap with the ground truth $\mathbf{x}^\star$, and the uninformative fixed point from $q_x \approx 0$, corresponding to no initial overlap with the signal. In the noiseless linear estimation and phase retrieval studied here we observe $\alpha_{\text{IT}} = \alpha_{\text{sp}}$.

3.1. Single-layer generative prior

First, we consider the case where the signal is generated from a single-layer generative prior, $\mathbf{x}^\star = \sigma(\mathbf{W}\mathbf{z})$ with $\mathbf{z} \sim \mathcal{N}(0, \mathrm{I}_k)$. We analyse both the compressed sensing and the phase retrieval problem, for $\sigma \in \{\text{linear}, \text{ReLU}\}$. In this case the only free parameters of the model are (ρ, α), and therefore the phase diagram fully characterises the recovery in these inverse problems. The aim is to compare with the phase diagram of a sparse prior with density $\rho_s = \rho$ of nonzero components.

Fig. 3 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 Krzakala et al. (2012a) 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 the 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 $\alpha_{\text{IT}} = \min(\rho, 1/2)$ and the hard

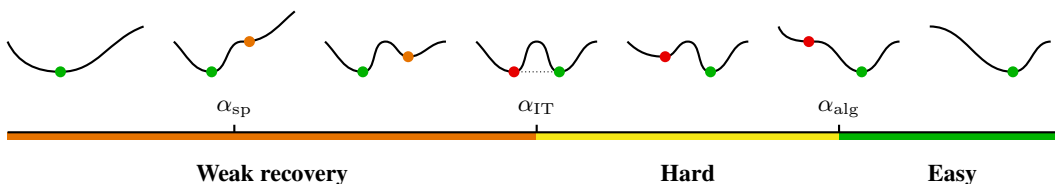Figure 2: Illustration of the free energy landscape as a function of the overlap with the ground truth solution, when one increases α. For small $\alpha < \alpha_{\mathrm{sp}}$, there exists a unique global minimum, whose overlap with the solution is small (high MSE). At $\alpha = \alpha_{\mathrm{sp}}$, a *local* minimum (orange dot) with higher overlap (small MSE) appears. By definition, the global minimum corresponds to the MMSE of the problem, which is the MSE attained by the Bayes-optimal estimator (green dot). For $\alpha < \alpha_{\mathrm{IT}}$ the accessible solution, i.e the global minimum (green dot) has a high MSE while a better solution exists but has a higher free energy (weak recovery phase). At $\alpha = \alpha_{\mathrm{IT}}$ the two minima are global and have the same free energy. Between $\alpha_{\mathrm{IT}} < \alpha < \alpha_{\mathrm{alg}}$ (hard phase), the local minimum with higher MSE corresponds to the performance of the AMP estimator (red dot). Above α_{alg} only the small MSE minima survive and the AMP estimator is able to achieve the Bayes-optimal performance (easy phase).

phase exists and has interesting properties: The $\rho \to \infty$ limit corresponds to the separable prior, and thus in this limit $\alpha_{\mathrm{alg}}(\rho \to \infty) = \alpha_{\mathrm{alg}}^{\mathrm{sparse}}(\rho_s = 1/2)$. Curiously we observe $\alpha_{\mathrm{alg}} > \alpha_{\mathrm{IT}}$ 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. 4 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 $\alpha_{\mathrm{alg}}^{\mathrm{PR}} > \alpha_{\mathrm{alg}}^{\mathrm{CS}}$ 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, $\alpha_{\mathrm{alg}}^{\mathrm{sparse}}(\rho_s)$, as taken from Barbier et al. (2019). 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 $\alpha_{\mathrm{alg}}(\rho \to 0) \to 0$. More specifically, the theory for the linear prior implies that $\alpha_{\mathrm{alg}}/\rho(\rho \to 0) \to \alpha_{\mathrm{alg}}^{\mathrm{sparse}}(\rho_s = 1) \approx 1.128$ 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.

3.2. Multi-layer generative prior

From the discussion above, we conclude that generative priors are algorithmically advantageous over sparse priors, allowing compressive sensing for the phase retrieval problem. We now investigate how the role of depth of the prior in this discussion. As before, we analyse both the linear estimation and phase retrieval problems, fixing $\sigma^{(l)} \equiv \sigma \in \{\mathrm{linear}, \mathrm{ReLU}\}$ at every layer $1 \leq l \leq L$. Different from the $L = 1$ case discussed above, for $L > 1$ we have other $L - 1$ free parameters characterising the layer-wise compression factors $(\beta_1, \ldots, \beta_{L-1})$.
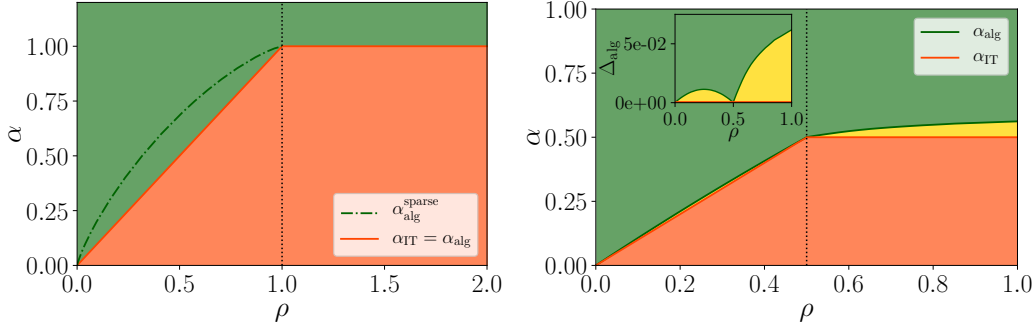
Figure 3: Phase diagrams for the compressed sensing problem with **(left)** linear generative prior and **(right)** ReLU generative prior, in the plane (ρ, α). The α_{IT} (red line) represents the information theoretic transition for perfect reconstruction and α_{alg} (green line) the algorithmic transition to perfect reconstruction. In the left part we depict for comparison the algorithmic phase transition for sparse separable prior $\alpha_{\text{alg}}^{\text{sparse}}$ (dashed-dotted green line). The inset in the right part depicts the difference $\Delta_{\text{alg}} = \alpha_{\text{alg}} - \alpha_{\text{IT}}$. Colored areas correspond respectively to the *weak recovery* (orange), *hard* (yellow) and *easy* (green) phases. The behaviour of the free energy landscape for increasing α and fixed ρ is illustrated in Fig. 2.
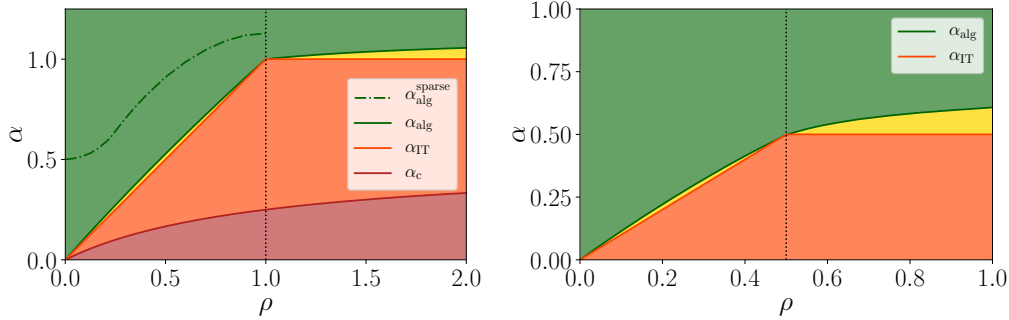

Figure 4: The same as Fig. 3 for the phase retrieval problem with **(left)** linear generative prior and **(right)** ReLU generative prior. A major result is that while with sparse separable priors (green dashed-dotted line) compressed phase retrieval is algorithmically hard for $\alpha < 1/2$, with 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).

First, we fix β_l and investigate the role played by depth. Fig. 5 depicts the phase diagrams for compressed sensing (left) and phase retrieval (right) with ReLU activation with varying depth, and a fixed architecture $\beta_l = 3$ for $1 \leq l \leq L$ and note that all these curves share the same $\alpha_{\text{IT}} = \min(0.5, \rho)$. It is clear that depth improves even more the small gap already observed for

a single-layer generative prior. The algorithmic advantage of multi-layer generative priors in the phase retrieval problem has been previously observed in a similar setting in Hand et al. (2018).

Next, we investigate the role played by the layer-wise compression factor β_l. Fig. 6 depicts the phase diagrams for the compressed sensing (left) and phase retrieval (right) with ReLU activation for fixed depth $L = 2$, and varying $\beta \equiv \beta_1$. According to the result in eq. (20), we have $\alpha_{\text{IT}} = \min\left(1/2, \rho, 1/2\beta\right)$. It is interesting to note that there is a trade-off between compression $\beta < 2$ and the algorithmic gap, in the following sense. For $\rho < 0.5$ fixed, α_{IT} decreases with decreasing $\beta \ll 1$: compression helps perfect recovery. However, the algorithmic gap Δ_{alg} becomes wider for fixed $\rho < 0.5$ and decreasing $\beta \ll 1$.

These observations also hold for a linear generative model. In Fig. 7 we have a closer look by plotting the algorithmic gap $\Delta_{\text{alg}} \equiv \alpha_{\text{alg}} - \alpha_{\text{IT}}$ in the phase retrieval problem. On the left, we fix $L = 4$ and plot the gap for increasing values of $\beta \equiv \beta_l$, leading to increasing Δ_{alg}. On the right, we fix $\beta = 2$ and vary the depth, observing a monotonically decreasing Δ_{alg}.
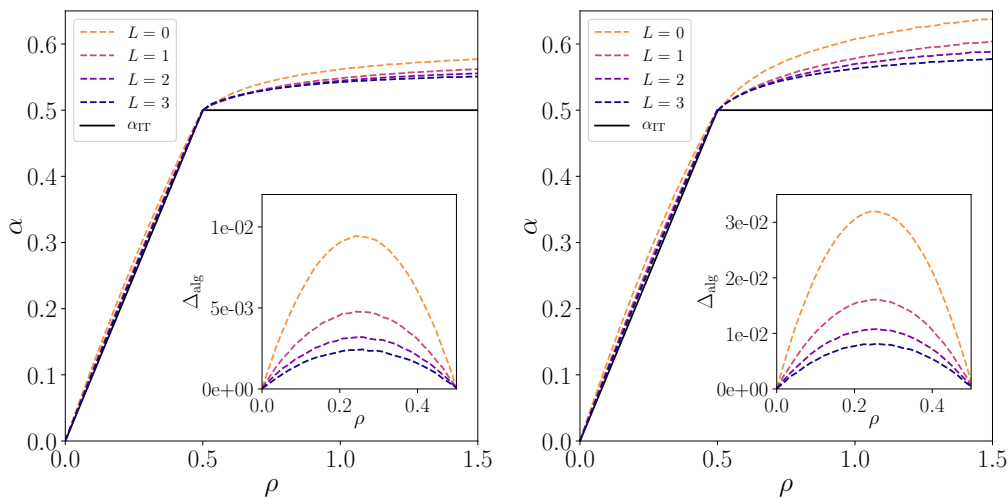

Figure 5: Phase diagrams for the compressed sensing (**left**) and phase retrieval (**right**) problems for different depths of the prior, with ReLU activation and fixed layer-wise compression $\beta_l = 3$. Dashed lines represent the algorithmic threshold α_{alg} and solid lines the perfect recovery threshold α_{IT}. We note that the algorithmic gap Δ_{alg} (shown in insets) decreases with the network depth L.

4. Conclusion and perspectives

In this manuscript we analysed how generative priors from an ensemble of random multi-layer neural networks impact signal reconstruction in the high-dimensional limit of two important inverse problems: real-valued phase retrieval and linear estimation. More specifically, we characterised the phase diagrams describing the interplay between number of measurements needed at a given signal compression ρ, for a range of shallow and multi-layer architectures for the generative prior. We observed that although present, the algorithmic gap significantly decreases with depth in the studied architectures. This is particularly striking when compared with sparse priors at $\rho \ll 1$, for which the
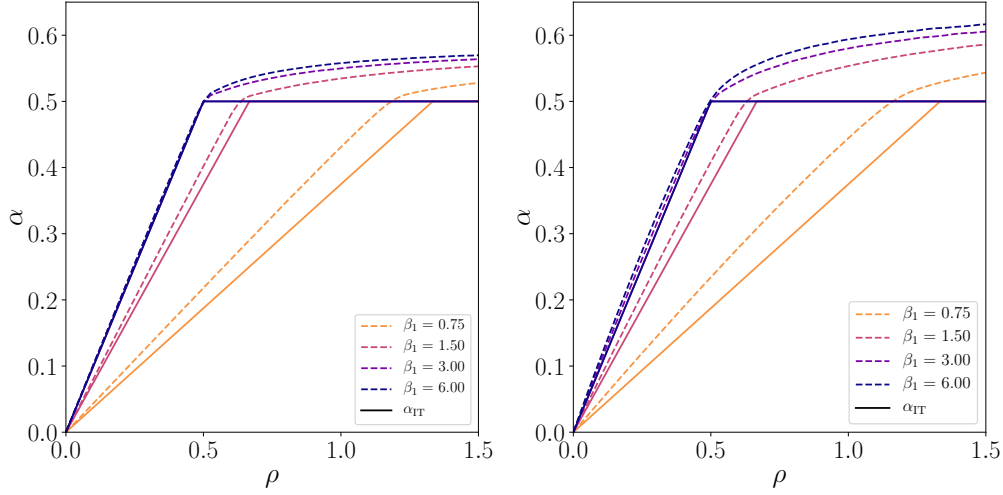
Figure 6: Phase diagrams for the compressed sensing (**left**) and phase retrieval (**right**) problems with $L = 2$ and ReLU activation for different values of the layer-wise compression factor β_1. Dashed lines represent the algorithmic threshold α_{alg} and solid lines the perfect recovery threshold α_{IT}. We note that for a given $\rho < 0.5$, α_{IT} is decreasing with $\beta \ll 1$. However, the algorithmic gap Δ_{alg} (shown in the inset) grows for decreasing β. Note that for $\beta_1 \geq 2$ the hard phase is hardly visible at $\rho = 0.5$, even though it disappears only in the large width limit, for both compressed sensing and phase retrieval settings.
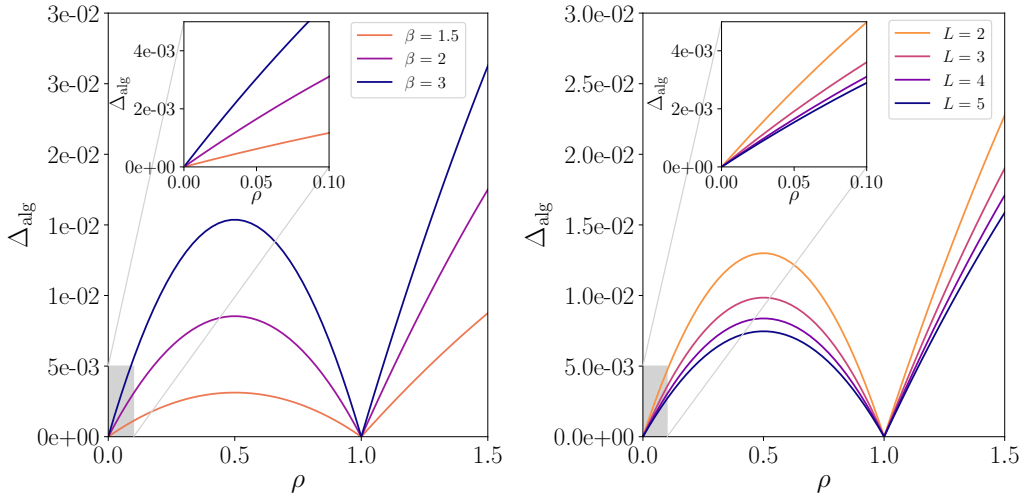

Figure 7: Algorithmic gap $\Delta_{\text{alg}} = \alpha_{\text{alg}} - \alpha_{\text{IT}}$ for small ρ and linear activation, as a function of (**left**) the compression $\beta \equiv \beta_l$ for fixed depth $L = 4$ and of (**right**) depth for a fixed compression $\beta = 2$.

69

algorithmic gap is considerably wider. In practice, this means generative models given by random multi-layer neural networks allow for efficient compressive sensing in these problems.

In this work we have only considered independent random weight matrices for both the estimation layer and for the generative model. Ideally, one would like to introduce correlations in a setting closer to reality to show that the smaller computation-to-statistical gap also appears in real-life tasks. The hurdle is that in those cases one does not know what is the theoretically optimal performance nor what are the optimal polynomial algorithms, so that one cannot evaluate the computation-to-statistical empirically in those cases. Yet another tractable case is the study of random rotationally invariant or unitary sensing matrices, as in Kabashima (2008); Fletcher et al. (2018); Barbier et al. (2018); Dudeja et al. (2019). In a different direction, it would be interesting to observe the phenomenology from this work in an experimental setting, for instance using a generative model, such as GANs or VAEs, trained on a real dataset to improve the performance of approximate message passing algorithms in a practical task.

Acknowledgments

References

Benjamin Aubin, Bruno Loureiro, Antoine Maillard, Florent Krzakala, and Lenka Zdeborová. The spiked matrix model with generative priors. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8364–8375. Curran Associates, Inc., 2019.

Antoine Baker, Benjamin Aubin, Florent Krzakala, and Lenka Zdeborová. Tramp: Compositional inference with tree approximate message passing, 2020.

Jean Barbier, Mohamad Dia, Nicolas Macris, and Florent Krzakala. The mutual information in random linear estimation. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 625–632. IEEE, 2016.

Jean Barbier, Nicolas Macris, Antoine Maillard, and Florent Krzakala. The mutual information in random linear estimation beyond iid matrices. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1390–1394. IEEE, 2018.

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

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.

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.

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.

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

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

David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

Rishabh Dudeja, Junjie Ma, and Arian Maleki. Information theoretic limits for phase retrieval with subsampled Haar sensing matrices. *arXiv preprint arXiv:1910.11849*, 2019.

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.

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.

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.

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

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

Yoshiyuki Kabashima. Inference from correlated patterns: a unified theory for perceptron learning and linear vector channels. In *Journal of Physics: Conference Series*, volume 95, page 012001. IOP Publishing, 2008.

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 2012a.

Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(08):P08009, 2012b.

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.

Christopher A Metzler, Manoj K Sharma, Sudarshan Nagesh, Richard G Baraniuk, Oliver Cossairt, and Ashok Veeraraghavan. Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset. In *2017 IEEE International Conference on Computational Photography (ICCP)*, pages 1–16. IEEE, 2017.

Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

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

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

Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. IEEE, 2011.

G. Reeves and H. D. Pfister. The replica-symmetric prediction for compressed sensing with gaussian matrices is exact. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 665–669, July 2016. doi: 10.1109/ISIT.2016.7541382.

Galen Reeves. Additivity of information in multilayer networks via additive gaussian noise transforms. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1064–1070. IEEE, 2017.

Galen Reeves and Michael Gastpar. Compressed sensing phase transitions: Rigorous bounds versus replica predictions. In *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2012.

Philip Schniter and Sundeep Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2014.

Terence Tao. Compressed sensing, or: the equation $Ax = b$, revisited. Clay-Mahler Lecture Series, 2009.

Eric W Tramel, Angélique Drémeau, and Florent Krzakala. Approximate message passing with restricted Boltzmann machine priors. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(7):073401, 2016a.

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, 2016b.

Yihong Wu and Sergio Verdú. Optimal phase transitions in compressed sensing. *IEEE Transactions on Information Theory*, 58(10):6241–6263, 2012.

Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.