This is a repository copy of *Analysis of Regularized LS Reconstruction and Random Matrix Ensembles in Compressed Sensing*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/94762/

Version: Accepted Version

Article:

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Analysis of Regularized LS Reconstruction and Random Matrix Ensembles in Compressed Sensing

Mikko Vehkaperä *Member, IEEE*, Yoshiyuki Kabashima, and Saikat Chatterjee *Member, IEEE*

Abstract—Performance of regularized least-squares estimation in noisy compressed sensing is analyzed in the limit when the dimensions of the measurement matrix grow large. The sensing matrix is considered to be from a class of random ensembles that encloses as special cases standard Gaussian, row-orthogonal, geometric and so-called T-orthogonal constructions. Source vectors that have non-uniform sparsity are included in the system model. Regularization based on ℓ_1-norm and leading to LASSO estimation, or basis pursuit denoising, is given the main emphasis in the analysis. Extensions to ℓ_2-norm and "zero-norm" regularization are also briefly discussed. The analysis is carried out using the replica method in conjunction with some novel matrix integration results. Numerical experiments for LASSO are provided to verify the accuracy of the analytical results.

The numerical experiments show that for noisy compressed sensing, the standard Gaussian ensemble is a suboptimal choice for the measurement matrix. Orthogonal constructions provide a superior performance in all considered scenarios and are easier to implement in practical applications. It is also discovered that for non-uniform sparsity patterns the T-orthogonal matrices can further improve the mean square error behavior of the reconstruction when the noise level is not too high. However, as the additive noise becomes more prominent in the system, the simple row-orthogonal measurement matrix appears to be the best choice out of the considered ensembles.

Index Terms—Compressed sensing, eigenvalues of random matrices, compressed sensing matrices, noisy linear measurements, ℓ_1 minimization

I. INTRODUCTION

CONSIDER the standard compressed sensing (CS) [1]–[3] setup where the sparse vector $\boldsymbol{x}^0 \in \mathbb{R}^N$ of interest is observed via noisy linear measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^0 + \sigma\boldsymbol{w}, \qquad (1)$$

where $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ represents the compressive $(M \leq N)$ sampling system. Measurement errors are captured by the vector $\boldsymbol{w} \in \mathbb{R}^M$ and parameter σ controls the magnitude of the distortions. The task is then to infer $\boldsymbol{x}^0$ from $\boldsymbol{y}$, given the measurement matrix $\boldsymbol{A}$. Depending on the chosen performance metric, the level of knowledge about the statistics of the source and error vectors, or computational complexity constraints, multiple choices are available for achieving this task. One possible solution that does not require detailed information about σ or statistics of $\{\boldsymbol{x}^0, \boldsymbol{w}\}$ is regularized least-squares (LS) based reconstruction

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{R}^N} \left\{ \frac{1}{2\lambda}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + c(\boldsymbol{x}) \right\}, \qquad (2)$$

where $\|\cdot\|$ is the standard Euclidean norm, λ a non-negative design parameter and $c : \mathbb{R}^N \to \mathbb{R}$ a fixed non-negative valued (cost) function. If we interpret (2) as a maximum a posteriori probability (MAP) estimator, the implicit assumption would be that: 1) the additive noise can be modeled by a zero-mean Gaussian random vector with covariance $\lambda\boldsymbol{I}_M$, and 2) the distribution of the source is proportional to $e^{-c(\boldsymbol{x})}$. Neither is in general true for the model (1) and, therefore, reconstruction based on (2) is clearly suboptimal.

In the sparse estimation framework, the purpose of the cost function c is to penalize the trial $\boldsymbol{x}$ so that some desired property of the source is carried over to the solution $\hat{\boldsymbol{x}}$. In the special case when the measurements are *noise-free*, that is, $\sigma = 0$, the choice $\lambda \to 0$ reduces (2) to solving a constrained optimization problem

$$\min_{\hat{\boldsymbol{x}} \in \mathbb{R}^N} c(\hat{\boldsymbol{x}}) \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{A}\hat{\boldsymbol{x}}. \qquad (3)$$

It is well-known that in the noise-free case the ℓ_1-cost $c(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = \sum_j |x_j|$ leads to sparse solutions that can be found using linear programming. For the noisy case the resulting scheme is called LASSO [4] or basis pursuit denoising [5]

$$\hat{\boldsymbol{x}}_{\ell_1} = \arg\min_{\boldsymbol{x} \in \mathbb{R}^N} \left\{ \frac{1}{2\lambda}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \|\boldsymbol{x}\|_1 \right\}. \qquad (4)$$

Just like its noise-free counterpart, it is of particular importance in CS since (4) can be solved by using standard convex optimization tools such as `cvx` [6]. Due to the prevalence of reconstruction methods based on ℓ_1-norm regularization in CS, we shall keep the special case of ℓ_1-cost $c(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$ as the main example of the paper, although it is known to be a suboptimal choice in general.

A. Brief Literature Review

In the literature, compressed sensing has a strong connotation of *sparse representations*. We shall next provide a brief review of the CS literature while keeping this in mind. The theoretical works in CS can be roughly divided into two principle directions: 1) *worst case* analysis, and 2) *average / typical case* analysis. In the former approach, analytical tools that examine the algebraic properties of the sensing matrix A, such as, mutual coherence, spark or restricted isometry property (RIP) are used. The goal is then to find sufficient conditions for the chosen property of A that guarantee perfect reconstruction — at least with high probability. The latter case usually strives for sharp conditions when the reconstruction is possible when A is sampled from some random distribution. Analytical tools vary from combinatorial geometry to statistical physics methods. Both, worst case and average case analysis have their merits and flaws as we shall discuss below.

For mutual coherence, several works have considered the case of noise-free observations ($\sigma = 0$) and ℓ_1-norm minimization based reconstruction. The main objective is usually to find the conditions that need to be satisfied between the allowed sparsity level of x and the mutual coherence property of A so that exact reconstruction is possible. In particular, the authors of [7] established such conditions for the special case when A is constructed by concatenating a pair of orthonormal bases. These conditions were further refined in [8] and the extension to general matrices was reported in [9] using the concept of *spark*.

Another direction in the worst case analysis was taken in [10], where the basic setup (1) with *sparse additive noise* was considered. The threshold for exact reconstruction under these conditions was derived using RIP. By establishing a connection between the Johnson-Lindenstrauss lemma and RIP, the authors of [11] proved later that RIP holds with high probability when M grows large for a certain class of random matrices. Special cases of this ensemble are, for example, matrices whose components are independent identically distributed (IID) Gaussian or Bernoulli random variables. This translates roughly to a statement that such matrices are "good" for CS problems when ℓ_1-norm based penalty is used if the system size is sufficiently large.

In addition to the basic problem stated above, mutual coherence and RIP based worst case analysis are prominent also in the study of greedy CS algorithms and fusion strategies. Some examples are analysis of orthogonal matching pursuit [12]–[14], subspace pursuit [15], CoSaMP [16], group LASSO [17] and Fusion strategy [18]. The general weakness of these approaches is, however, that if one is interested in *typical* or *average* case performance, the results provided by the worst case analysis are often very pessimistic and loose. This consideration is tackled by the second class of analytical results we mentioned at the beginning of the review.

In a series of papers, the authors of [19]–[21] used tools from combinatorial geometry to show that in the limit of increasing system size, the ℓ_1-reconstruction has a sharp *phase transition* when the measurements are noise-free. A completely different approach based on approximate message passing (AMP) algorithm [22], [23] was introduced in [24] and shown to match the combinatorial results perfectly. Both of the above methods are mathematically rigorous and the AMP approach has the additional benefit that it provides also a low-complexity computational algorithm that matches the threshold behavior. The downside is that extending these analysis for more general ensembles, both for the measurement matrix and the source vector, seems to be quite difficult. Alternative route is to use statistical mechanics inspired tools like the *replica method* [25]–[27].

By now the replica method has been accepted in the information theory society as a mathematical tool that can tackle problems that are very difficult, or impossible, to solve using other (rigorous) approaches. Although the outcomes of the replica analysis have received considerable success (see, e.g., [28]–[34] for some results related to the present paper), one should keep in mind that mathematical rigor is still lacking in parts of the method [35]. However, recent results in mathematical physics have provided at least circumstantial evidence that the main problem of the replica method is most likely in the assumed structure of the solution [35]–[39] and not in the parts such as replica continuity that lack mathematical proof. In particular, the mistake in the original solution of the Sherrington-Kirkpatrick spin glass has now been traced to the assumption of *replica symmetric* (RS) ansatz in the saddle-point evaluation of the free energy. Indeed, the end result of the Parisi's full *replica symmetry breaking* (RSB) solution (see, e.g., [25]) has been proved to be correct [38], [39] in this case. Similar rigorous methods have also been applied in wireless communications [40] and error correction coding [41], [42], to name just a few examples[1].

B. Related Prior Work

The authors of [28] analyzed the asymptotic performance of LASSO and "zero-norm" regularized LS by extending the minimum mean square error (MMSE) estimation problem in code division multiple access (CDMA) to MAP detection in linear vector models. More specifically, the MMSE formulas obtained with the replica method [43], [44] were first assumed to be valid and then transformed to the case of MAP decoding through "hardening". Unfortunately, this approach was limited to the cases where the appropriate MMSE formulas already existed and the end result of the analysis still required quite a lot of numerical computations. The scope of the analysis was extended to a more general class of random matrices by employing the Harish-Chandra-Itzykson-Zuber (HCIZ) integral formula [45], [46] in [30]. Although the emphasis there was in the *support recovery*, also the MSE could be inferred from the given results. A slightly different scenario when the additive noise is sparse was analyzed in [47], [48]. For such a measurement model, if one replaces the squared ℓ_2-norm

---

[1]To avoid the misconception that these methods have made non-rigorous approaches obsolete, some comments are in place. Firstly, the scope of the rigorous methods tend to be much more limited than that of the non-rigorous ones. Secondly, the analysis typically give bounds for the quantities of interest rather than sharp predictions. Thirdly, it is often helpful to know the end-result obtained through some non-rigorous way, like the replica method, before applying the mathematically exact tools on the problem.

distance in (2) by ℓ_1-norm and uses also ℓ_1-regularization, perfect reconstruction becomes sometimes feasible [47], [48]. It is also possible to characterize the MSE of reconstruction outside of this region using the replica method [48].

The references above left the question open how the choice of measurement matrix affects the fidelity of the reconstruction in the noisy setup. In [49] a partial answer was obtained through information theoretic analysis. The authors showed that standard Gaussian sensing matrices incurred no loss in the noise sensitivity threshold if optimal encoding and decoding were used. Similar result was obtained earlier using the replica method in [29], and extended to more general matrix ensembles in the aforementioned paper [30]. On the other hand, generalization of the Lindeberg principle was used by the authors of [50] to show that the average cost in LASSO was universal for a class of matrices of *standard type*.

Based on the above results and the knowledge that for the noise-free case the perfect reconstruction threshold is quite universal [20], [21], [34], one might be tempted to conclude that using sensing matrices that are sampled from the standard Gaussian ensemble is the optimal choice also in the noisy case when practical algorithms such as LASSO are used. However, there is also some counter-evidence in other settings, such as the noise-free case with non-uniform sparsity [31], [32] and spreading sequence design in CDMA [51], [52] that leave the problem still interesting to investigate in more detail[2].

Albeit from a slightly different motivational point-of-view, similar endeavor was taken earlier in [60]–[63], where it was discovered that measurement matrices with specific structure are beneficial for message passing decoding in noise-free settings. These spatially coupled, or seeded, measurement matrices helped the iterative algorithm to get past local extrema and hence improved the perfect reconstruction threshold of ℓ_1-recovery significantly. Such constructions, however, turned out to be detrimental for convex relaxation based methods when compared to the standard Gaussian ensemble.

Finally we remark that the uniform sparsity model studied in [34] was extended to a non-uniform noise-free setting in [33]. The goal there was to optimize the recovery performance using weighted ℓ_1-minimization when the sparsity pattern is known. We deviate from those goals by considering a noisy setup with a more general matrix ensemble for measurements. On the other hand, we do not try to optimize the reconstruction with block-wise adaptive weights and leave such extensions as future research topics.

C. Contribution and Summary of Results

The main goal of the present paper is to extend the scope of [28] and [30] to a wider range of matrix ensembles and to non-uniform sparsities of the vector of interest. We deviate from the approach of [28], [30] by evaluating the performance directly using the replica method as in [31]–[34]. The derivations are

[2]After the initial submission of the present paper, parallel studies using completely different mathematical methods and arguing for the superiority of the orthogonal constructions have been presented in [53] and [54]. Since then, an extension to the present paper has been proposed in [55] and iterative algorithms approximating Bayesian optimal estimation for structured matrices have been devised, see for example, [56]–[59].

also akin to some earlier works on linear models [64], [65]. After obtaining the results for ℓ_1-regularization, we sketch how they can be generalized to other cases like l_2-norm and "zero-norm" based regularization.

The analysis show that under the assumption of RS ansatz (for details, see Section IV), the average MSE of reconstruction is obtained via a system of coupled fixed point equations that can be solved numerically. For the T-orthogonal case, we find that the solution depends on the sparsity pattern (how the non-zero components are located block-wise in the vector) of the source — even when such knowledge is not used in the reconstruction. In the case of *rotationally invariant* ensemble, the results are obtained as a function of the Stieltjes transform of the eigenvalue spectrum that describes the measurement matrix. For this case only the total sparsity of the source vector has influence on the reconstruction performance. The end results for the rotationally invariant case are also shown to be equivalent to those in [30], bridging the gap between two different approaches to replica analysis.

Finally, solving the MSE of the replica analysis for some practical settings reveals that the standard Gaussian ensemble is suboptimal as a sensing matrix when the system is corrupted by additive noise. For example, a random row-orthogonal measurement matrix provides uniformly better reconstructions compared to the Gaussian one. This is in contrast to the noise-free case where it is well known that the perfect reconstruction threshold is the same for the whole rotationally invariant ensemble (see, e.g., [34]). On the other hand, albeit T-orthogonal measurement matrices are able to offer lower MSE than any other ensemble we tested when the sparsity of the source is not uniform, the effect diminishes as the noise level in the system increases. This may be intuitively explained by the fact that the additive noise in the system makes it more difficult to differentiate between blocks of different sparsities when we have no prior information about it.

D. Notation and Paper Outline

Boldface symbols denote (column) vectors and matrices. Identity matrix of size $M \times M$ is written $\boldsymbol{I}_M$ and the transpose of matrix $\boldsymbol{A}$ as $\boldsymbol{A}^\mathsf{T}$. Given a variable x_k with a countable index set $\mathcal{K}$, we abbreviate $\{x_k\} = \{x_k : k \in \mathcal{K}\}$. We write $\mathrm{i} = \sqrt{-1}$ and for some (complex) function $f(z)$, denote $f(z_0) = \mathrm{extr}_z f(z)$ where z_0 is an extremum of the function f, that is, satisfies $\frac{\mathrm{d}f}{\mathrm{d}z}\big|_{z_0} = 0$. Analogous definition holds for functions of multiple variables. The indicator function satisfies $1(A) = 1$ if A is true and is zero otherwise. Dirac's delta function is written $\delta(x)$ and the Kronecker symbol δ_{ij}.

Throughout the paper we assume for simplicity that given any continuous (discrete) random variable, the respective probability density (probability mass) function exists. Same notation is used for both cases, and given a general continuous / discrete random variable (RV), we often refer to probability density function (PDF) for brevity. The *true* and *postulated* PDF of a random variable is denoted p and q, respectively. If $\boldsymbol{x}$ is a real-valued Gaussian RV with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, we write the density of $\boldsymbol{x}$ as $p(\boldsymbol{x}) = g_{\boldsymbol{x}}(\boldsymbol{\mu}; \boldsymbol{\Sigma})$.

The rest of the paper is organized as follows. The problem formulation and brief introduction to the replica trick is given

in Section II. Section III provides the end-results of replica analysis for LASSO estimation. This case is also used in the detailed replica analysis provided in Appendices A and B. Sketch of the main steps involved in the replica analysis and comparison to existing results are given in Section IV for the rotationally invariant setup. Conclusions are provided in Section V and two matrix integral results used as a part of the replica analysis are proved in Appendix C. Finally, Appendix D provides the details of the geometric ensemble.

II. PROBLEM FORMULATION AND METHODS

Consider the CS setup (1) and assume that the elements of $\boldsymbol{w}$ are IID standard Gaussian random variables, so that

$$p(\boldsymbol{y} \mid \boldsymbol{A}, \boldsymbol{x}^0) = g_{\boldsymbol{y}}(\boldsymbol{A}\boldsymbol{x}^0; \sigma^2 \boldsymbol{I}_M) \qquad (5)$$

is the conditional PDF of the observations. Recall that the notation $\boldsymbol{x}^0$ means here that the observation (1) was generated as $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^0 + \sigma\boldsymbol{w}$, that is, $\boldsymbol{x}^0$ is the true vector generated by the source. Note that in this setting the additive noise is dense and, therefore, perfect reconstruction is in general not possible [47], [48]. Let the sparse vector of interest $\boldsymbol{x}^0$ be partitioned into T equal length parts $\{\boldsymbol{x}_t^0\}_{t=1}^T$ that are statistically independent. The components in each of the blocks $t = 1, \ldots, T$ are drawn IID according to the mixture distribution

$$p_t(x) = (1 - \rho_t)\delta(x) + \rho_t \pi(x), \quad t = 1, \ldots, T, \qquad (6)$$

where $\rho_t \in [0, 1]$ is the expected fraction of non-zero elements in $\boldsymbol{x}_t^0$ that are drawn independently according to $\pi(x)$. The expected density, or sparsity, of the whole signal is thus $\rho = T^{-1}\sum_t \rho_t$. We denote the true prior according to which the data is generated by $p(\boldsymbol{x}^0; \{\rho_t\})$ and call $\{\rho_t\}$ the *sparsity pattern* of the source. For future reference, we define the following nomenclature.

Definition 1. When the system size grows without bound, namely, $M, N \to \infty$ with fixed and finite *compression rate* $\alpha = M/N$ and sparsity levels $\{\rho_t\}$, we say the CS setup approaches the large system limit (LSL).

Definition 2. Let $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ be a sensing matrix with compression rate $\alpha = M/N \leq 1$. We say that the recovery problem (2) is:

1) *T-orthogonal setup*, if $N = TM$ and the sensing matrix is constructed as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{O}_1 & \cdots & \boldsymbol{O}_T \end{bmatrix}, \qquad (7)$$

where $\{\boldsymbol{O}_t\}$ are independent and distributed uniformly on the group of orthogonal $M \times M$ matrices according to the Haar measure[3];

2) *Standard Gaussian setup*, if the elements of $\boldsymbol{A}$ are IID drawn according to $g_a(0; 1/M)$;

3) *Row-orthogonal setup*, if $\boldsymbol{O}$ is an $N \times N$ Haar matrix and the sensing matrix is constructed as $\boldsymbol{A} = \alpha^{-1/2}\boldsymbol{PO}$, where $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{I}_M & \boldsymbol{0}_{M \times (N-M)} \end{bmatrix}$ picks the first M rows of $\boldsymbol{O}$. Clearly $\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} = \alpha^{-1}\boldsymbol{I}_M$ and $\boldsymbol{A}$ has orthogonal rows.

4) *Geometric setup*, if $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{T}}$ where $\boldsymbol{U}, \boldsymbol{V}$ are independent Haar matrices and $\boldsymbol{\Sigma} \in \mathbb{R}^{M \times N}$ is a diagonal matrix whose $(m, m)th$ entry is given by $\sigma_m \propto \tau^{m-1}$ for $m = 1, \ldots, M$. The parameter $\tau \in (0, 1]$ is chosen so that given value of peak-to-average eigenvalue ratio

$$\kappa = \frac{\sigma_1^2}{\frac{1}{M}\sum_{m=1}^M \sigma_m^2} \qquad (8)$$

is met and the singular values are scaled to satisfy the power constraint $N^{-1}\sum_{m=1}^M \sigma_m^2 = 1$. For details, see Appendix D.

5) *General rotationally invariant setup*, if the decomposition $\boldsymbol{R} = \boldsymbol{A}^{\mathsf{T}}\boldsymbol{A} = \boldsymbol{O}^{\mathsf{T}}\boldsymbol{D}\boldsymbol{O}$ exists, so that $\boldsymbol{O}$ is an $N \times N$ Haar matrix and $\boldsymbol{D}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{R}$. We also assume that the empirical distribution of the eigenvalues

$$F_{\boldsymbol{R}}^M(x) = \frac{1}{M}\sum_{i=1}^M 1(\lambda_i(\boldsymbol{R}) \leq x), \qquad (9)$$

where $1(\cdot)$ is the indicator function and $\lambda_i(\boldsymbol{R})$ denotes the ith eigenvalue of $\boldsymbol{R}$, converges to some non-random limit in the LSL and satisfies $\mathsf{E}\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})/N = \mathsf{E}\operatorname{tr}(\boldsymbol{D})/N = 1$. The setups 2) – 4) are all special cases of this ensemble.

To make comparison fair between different setups, all cases above are defined so that $\mathsf{E}\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})/N = 1$. In addition, both of the orthogonal setups satisfy the condition $\alpha\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} = \boldsymbol{I}_M$.

Remark 1. The T-orthogonal sensing matrix was considered in [31], [32] under the assumption of noise-free measurements. There it was shown to improve the perfect recovery threshold when the source had non-uniform sparsity. On the other hand, the row-orthogonal setup is the same matrix ensemble that was studied in the context of CDMA in [51], [52]. There it was called Welch bound equality (WBE) spreading sequence ensemble and shown to provide maximum spectral efficiency both for Gaussian [51] and non-Gaussian [52] inputs given optimal MMSE decoding. The geometric setup is inspired by [66], where similar sensing matrix was used to examine the robustness of AMP algorithm and its variants via Monte Carlo simulations. It reduces to the row-orthogonal ensemble when $\kappa \to 1$.

A. Bayesian Framework

To enable the use of statistical mechanics tools, we reformulate the original optimization problem (2) in a probabilistic framework. For simplicity[4], we also make the additional restriction that the cost function separates as

$$c(\boldsymbol{x}) = \sum_{j=1}^N c(x_j), \qquad (10)$$

where c is a function whose actual form depends on the type of the argument (scalar or vector). Then, the postulated model

---

[3]In the following, a matrix $\boldsymbol{O}$ that has this distribution is said to be simply a *Haar matrix*.

[4]This assumption is in fact not necessary for the replica analysis. However, if the source vector has independent elements and the regularization function decouples element-wise, the numerical evaluation of the saddle-point equations is a particularly simple task.

prior (recall our notational convention from Section I-D) of the source is defined as

$$q_\beta(\boldsymbol{x}) = \frac{1}{z_\beta} \mathrm{e}^{-\beta c(\boldsymbol{x})}, \qquad (11)$$

where $z_\beta = \int \mathrm{e}^{-\beta c(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} < \infty$ is a normalization constant. Hence, $c(\boldsymbol{x})$ needs to be such that the above integral is convergent for given finite M, N and $\beta > 0$. The purpose of the non-negative parameter β (inverse temperature) is to enable MAP detection as will become clear later. Note that (11) encodes no information about the sparsity pattern of the source and is mismatched from the true prior $p(\boldsymbol{x}^0; \{\rho_t\})$. From an algorithmic point-of-view, this means that the system operator has no specific knowledge about the underlying sparsity structure or does not want to utilize it due to increased computational complexity. We also define a postulated PDF for the measurement process

$$q_\beta(\boldsymbol{y} \mid \boldsymbol{A}, \boldsymbol{x}) = g_{\boldsymbol{y}}\left(\boldsymbol{A}\boldsymbol{x}; \frac{\lambda}{\beta}\boldsymbol{I}_M\right), \qquad (12)$$

so that unless $\lambda/\beta = \sigma^2$, the observations are generated according to a different model than what the reconstruction algorithm assumes. Note that λ is the same parameter as in the original problem (2).

Due to Bayes' theorem, the (mismatched) posterior density of $\boldsymbol{x}$ based on the postulated distributions reads

$$\begin{aligned} &q_\beta(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A}) \\ &= \frac{1}{Z_\beta(\boldsymbol{y}, \boldsymbol{A})} \exp\left[-\beta\left(\frac{1}{2\lambda}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + c(\boldsymbol{x}) \right) \right], \end{aligned} \quad (13)$$

where $Z_\beta(\boldsymbol{y}, \boldsymbol{A})$ is the normalization factor or *partition function* of the above PDF. We could now estimate $\boldsymbol{x}$ based on (13), for example by computing the posterior mean $\langle \boldsymbol{x} \rangle_\beta$, where we used the notation

$$\langle h(\boldsymbol{x}) \rangle_\beta = \int h(\boldsymbol{x}) q_\beta(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A}) \mathrm{d}\boldsymbol{x} \qquad (14)$$

for some given $\beta > 0$ and trial function h of $\boldsymbol{x}$. The specific case that maximizes the a posteriori probability for given λ (and σ^2) is the *zero-temperature configuration*, obtained by letting $\beta \to \infty$. In this limit (13) reduces to a uniform distribution over $\boldsymbol{x}$ that provides the global minimum of $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2/(2\lambda) + c(\boldsymbol{x})$. If the problem has a unique solution, we have $\langle \boldsymbol{x} \rangle_{\beta \to \infty} = \hat{\boldsymbol{x}}$, where $\hat{\boldsymbol{x}}$ is the solution of (2). Thus, the behavior of regularized LS reconstruction can be obtained by studying the density (13). This is a standard problem in statistical mechanics if we interpret $q_\beta(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A})$ as the Boltzmann distribution of a spin glass, as described next.

B. Free Energy, The Replica Trick and Mean Square Error

The key for finding the statistical properties of the reconstruction (2) is the normalization factor or partition function $Z_\beta(\boldsymbol{y}, \boldsymbol{A})$. Based on the statistical mechanics approach, our goal is to assess the (normalized) *free energy*

$$f_\beta(\boldsymbol{y}, \boldsymbol{A}) = -\frac{1}{\beta N} \ln Z_\beta(\boldsymbol{y}, \boldsymbol{A}) \qquad (15)$$

in the LSL where $M, N \to \infty$ with $\alpha = M/N$ fixed, and obtain the desired statistical properties from it. However, the

formulation above is problematic since f_β depends on the observations $\boldsymbol{y}$ and the measurement process $\boldsymbol{A}$. One way to circumvent this difficulty is to notice that the law of large numbers guarantees that for $\forall \epsilon > 0$, the probability that $|f_\beta(\boldsymbol{y}, \boldsymbol{A}) - \mathsf{E}\{f_\beta(\boldsymbol{y}, \boldsymbol{A})\}| > \epsilon$ tends to vanish in the LSL for any finite and positive λ, σ^2. This leads to computation of the average free energy $f_\beta = \mathsf{E}\{f_\beta(\boldsymbol{y}, \boldsymbol{A})\}$ instead of (15) and is called *self-averaging* in statistical mechanics.

Concentrating on the average free energy f_β avoids the explicit dependence on $\{\boldsymbol{y}, \boldsymbol{A}\}$. Unfortunately, assessing the necessary expectations is still difficult and we need some further manipulations to turn the problem into a tractable one. The first step is to rewrite the average free energy in the zero-temperature limit as

$$f = -\lim_{\beta, N \to \infty} \frac{1}{\beta N} \lim_{n \to 0^+} \frac{\partial}{\partial n} \ln \mathsf{E}\{[Z_\beta(\boldsymbol{y}, \boldsymbol{A})]^n\}. \qquad (16)$$

So-far the development has been rigorous if $n \in \mathbb{R}$ and the limits are unique and exist[5]. The next step is to employ the *replica trick* to overcome the apparent road block of evaluating the necessary expectations as a function of real-valued parameter n.

Replica Trick. Consider the free energy in (16) and let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^0 + \boldsymbol{w}$ be a fixed observation vector. Assume that the limits commute, which in conjunction with the expression

$$\begin{aligned} &[Z_\beta(\boldsymbol{y}, \boldsymbol{A}; \lambda)]^n \\ &= \int \exp\left(-\frac{\beta}{2\lambda} \sum_{a=1}^n \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^a\|^2 - \beta c(\boldsymbol{x}^a) \right) \prod_{a=1}^n \mathrm{d}\boldsymbol{x}^a \end{aligned} \quad (17)$$

for $n = 1, 2, \ldots$ allows the evaluation of the expectation in (16) as a function of $n \in \mathbb{R}$. The obtained functional expression is then utilized in taking the limit of $n \to 0^+$.

It is important to note that as written above, the validity of the analytical continuation remains an open question and the replica trick is for this part still lacking mathematical validation. However, as remarked in Introduction, the most serious problem in practice seems to arise from the simplifying assumptions one makes about how the correlations between the variables $\{\boldsymbol{x}^a\}$ behave in the LSL. The simplest case is the *RS ansatz*, that is described by the *overlap matrix* $\boldsymbol{Q} \in \mathbb{R}^{(n+1) \times (n+1)}$ of the form

$$\boldsymbol{Q} = [Q^{[a,b]}]_{a,b=0}^n = \begin{bmatrix} r & m & \cdots & m \\ m & Q & q & \\ \vdots & q & \ddots & \\ m & & & Q \end{bmatrix} \qquad (18)$$

with slightly non-standard indexing that is common in literature related to replica analysis. The elements of $\boldsymbol{Q}$ are defined as overlaps, or empirical correlations, $Q^{[a,b]} = N^{-1}\boldsymbol{x}^a \cdot \boldsymbol{x}^b$. The implication of RS ansatz is that the replica indexes $a = 1, 2, \ldots, n$ can be arbitrarily permuted without changing the end result when $M = \alpha N \to \infty$. This seems a

---

[5]In principle, the existence of a unique thermodynamic limit can be checked using the techniques introduced in [37]. However, since the replica method itself is already non-rigorous we have opted to verify the results in the end using numerical simulations.

priori reasonable since the replicas introduced in (17) were identical and in no specific order. But as also mentioned in Introduction, the RS assumption is not always correct. Sometimes the discrepancy is easy to fix, for example as in the case of random energy model [67], while much more intricate methods like Parisi's full RSB solution are needed for other cases [25]. For the purposes of the present paper, we restrict ourselves to the RS case and check accuracy of the end result w.r.t. simulations. Although this might seem mathematically somewhat unsatisfying approach, we believe that the RS results are useful for practical purposes due to their simple form and can server as a stepping stone for possible extensions to the RSB cases.

Finally, let us consider the problem of finding the MSE of reconstruction (2). Using the notation introduced earlier, we may write

$$
\begin{aligned}
\text{mse} &= N^{-1}\mathsf{E}\big\{\langle\|\boldsymbol{x}^0 - \boldsymbol{x}\|^2\rangle_{\beta\to\infty}\big\} \\
&= \rho - 2\mathsf{E}\big\{N^{-1}\langle\boldsymbol{x}\rangle_{\beta\to\infty}^{\mathsf{T}}\boldsymbol{x}^0\big\} + \mathsf{E}\big\{N^{-1}\langle\|\boldsymbol{x}\|^2\rangle_{\beta\to\infty}\big\} \\
&= \rho - 2m + Q,
\end{aligned}
\tag{19}
$$

where $\langle\cdots\rangle_\beta$ was defined in (14) and $\mathsf{E}\{\cdots\}$ denotes the expectation w.r.t. variables in (1). Thus, if we can compute m and Q using the replica method, the MSE of reconstruction follows immediately from (19). As shown above, this amounts to computing the overlap matrix (18).

III. RESULTS FOR LASSO RECONSTRUCTION

In this section we provide the results of the replica analysis for LASSO reconstruction (4) for the ensembles introduced in Definition 2. For simplicity, we let the non-zero elements of the source be standard Gaussian, that is, $\pi(x) = g_x(0; 1)$ in (6). Recall also that LASSO is the special case of regularization $c(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$ in the general problem (2). Replica symmetric ansatz is assumed in the derivations given in Appendices A and B. Casual reader finds a sketch of replica analysis along with some generalizations for different choices of the cost function $c(\boldsymbol{x})$ in Section IV. Further interpretation of the result and connections to the earlier work in [28] are also discussed there. After the analytical results, we provide some numerical examples in the following subsection.

A. Analytical Results

The first result shows that when the measurement matrix is of the T-orthogonal form, the MSE over the whole vector may depend, not just on the average sparsity ρ but also on the block-wise sparsities $\{\rho_t\}$.

Proposition 1. *Consider the T-orthogonal setup described in Definition 2 and let* mse_t *denote the MSE of the LASSO reconstruction in block* $t = 1,\ldots,T$. *The average MSE over the entire vector of interest reads*

$$
\text{mse} = \frac{1}{T}\sum_{t=1}^{T}\text{mse}_t = \frac{1}{T}\sum_{t=1}^{T}(\rho_t - 2m_t + Q_t).
\tag{20}
$$

Then, under RS ansatz,

$$
m_t = 2\rho_t \mathcal{Q}\left(\frac{1}{\sqrt{\hat\chi_t + \hat m_t^2}}\right),
\tag{21}
$$

$$
Q_t = -\frac{2(1-\rho_t)}{\hat m_t^2}r(\hat\chi_t) - \frac{2\rho_t}{\hat m_t^2}r(\hat\chi_t + \hat m_t^2),
\tag{22}
$$

where $\mathcal{Q}(x) = \int_x^\infty \mathrm{d}z e^{-z^2/2}/\sqrt{2\pi}$ *is the standard Q-function and we denoted*

$$
r(h) \triangleq \sqrt{\frac{h}{2\pi}}e^{-\frac{1}{2h}} - (1+h)\mathcal{Q}\left(\frac{1}{\sqrt{h}}\right),
\tag{23}
$$

$$
\hat m_t \triangleq \frac{1}{\lambda + \sum_{k\neq t}\Lambda_k^{-1}},
\tag{24}
$$

for notational convenience. The parameters $\{\Lambda_t\}$ *and* $\{\hat\chi_t\}$ *are the solutions to the set of coupled equations*

$$
\Lambda_t = \left(\frac{1}{R_t} - 1\right)\hat m_t
\tag{25}
$$

$$
\hat\chi_t = \frac{(\rho_t - 2m_t + Q_t)\Lambda_t^2}{(1 - R_t)^2}
$$
$$
+ \sum_{s=1}^{T}\Delta_{s,t}(\rho_s - 2m_s + Q_s - \sigma^2 R_s^2),
\tag{26}
$$

where we also used the auxiliary variables $R_t = \chi_t\hat m_t$ *with*

$$
\chi_t \triangleq \frac{2(1-\rho_t)}{\hat m_t}\mathcal{Q}\left(\frac{1}{\sqrt{\hat\chi_t}}\right) + \frac{2\rho_t}{\hat m_t}\mathcal{Q}\left(\frac{1}{\sqrt{\hat\chi_t + \hat m_t^2}}\right),
\tag{27}
$$

$$
\Delta_{s,t} \triangleq \frac{R_s R_t \Lambda_s \Lambda_t}{(1 - 2R_s)(1 - 2R_t)}\left(1 + \sum_{k=1}^{T}\frac{R_k^2}{1 - 2R_k}\right)^{-1}
$$
$$
- \frac{\Lambda_t^2}{1 - 2R_t}\delta_{st},
\tag{28}
$$

and Kronecker delta symbol δ_{ij} *to simplify the notation.*

Proof: See Appendix A. ∎

The connection of the MSE provided in Proposition 1 and the formulation given in (19) is as follows. Here $m = T^{-1}\sum_t m_t$ and $Q = T^{-1}\sum_t Q_t$ due to the assumption of block-wise sparsity, as described in Section II. The parameters $\{\Lambda_t, \hat\chi_t\}$ have to be solved for all $t = 1,\ldots,T$, i.e., we have a set of $2T$ non-linear equations of $2T$ variables. Note that for the purpose of solving these equations, the parameters $\{\hat m, R_t, \Delta_{s,t}, m_t, Q_t\}$ are just notation and do not act as additional variables in the problem. Except for m_t and Q_t, the rest of the variables can in fact be considered to be "auxiliary". They arise in the replica analysis when we assess the expectations w.r.t. randomness of the measurement process, additive noise and vector of interest. Hence, one may think that the replica trick transformed the task of computing difficult expectations to a problem of finding solutions to a set of coupled fixed point equations defined by some auxiliary variables. In terms of computational complexity, this is a very fair trade indeed.

The implication of Proposition 1 is that the performance of the T-orthogonal ensemble is in general dependent on the details of the sparsity pattern $\{\rho_t\}$ — in a rather complicated way. Similar result was reported for the noise-free case in [32],

where the T-orthogonal ensemble was shown to provide superior reconstruction threshold compared to rotationally invariant cases when the source vector had non-uniform sparsity.

For future convenience, we next present the special case of uniform sparsity as an example. Since now $\rho = \rho_t$ for all $t = 1, \ldots, T$, we have only two coupled fixed point equations to solve. The auxiliary parameters also simplify significantly and we have the following result.

Example 1. Consider the T-orthogonal setup and assume uniform sparsity $\rho_t = \rho$ for all $t = 1, \ldots, T$. The per-component MSE of reconstruction can be obtained by solving the set of equations

$$\Lambda = \frac{1}{\lambda}\left[2(1-\rho)\mathcal{Q}\left(\frac{1}{\sqrt{\hat{\chi}}}\right) + 2\rho\mathcal{Q}\left(\frac{1}{\hat{\chi} + \hat{m}^2}\right)\right]^{-1} - \frac{T}{\lambda}, \quad (29)$$

$$\hat{\chi} = \Lambda^2\left(\frac{\rho - 2m + Q}{(1-R)^2} - \frac{\rho - 2m + Q - \sigma^2 R^2}{1 - 2R + TR^2}\right), \quad (30)$$

where we introduced the definitions

$$R \triangleq \frac{1}{T + \lambda\Lambda}, \quad (31)$$

$$\hat{m} \triangleq \frac{\Lambda}{T + \lambda\Lambda - 1}, \quad (32)$$

for notational simplicity.

For completeness, we provide next a result similar to Proposition 1 for the rotationally invariant case. It shows that the performance of this matrix ensemble does not depend on the specific values of $\{\rho_t\}$ but only on the expected sparsity level $\rho = T^{-1}\sum_t \rho_t$ of the vector of interest. The MSE is given as a function of the Stieltjes transform and its first order derivative of the asymptotic eigenvalue distribution $F_{\boldsymbol{AA}^\mathsf{T}}$ of the measurement matrix. As shown in Section IV-B, the end result is essentially the same as the HCIZ integral formula based approach in [30], but the derivation and form of the proposition are chosen here to match the previous analysis. It should be remarked, however, that Proposition 1 cannot be obtained from [30] and the T-orthogonal ensemble requires a special treatment.

Proposition 2. *Recall the rotationally invariant setup given in Definition 2. The average MSE of reconstruction for LASSO in this case is given by* $\mathsf{mse} = \rho - 2m + Q$, *where the parameters m and Q are as in (21) and (22) with the block index t omitted. To obtain the MSE, the following set of equations*

$$\chi = \frac{2(1-\rho)}{\hat{m}}\mathcal{Q}\left(\frac{1}{\sqrt{\hat{\chi}}}\right) + \frac{2\rho}{\hat{m}}\mathcal{Q}\left(\frac{1}{\sqrt{\hat{\chi} + \hat{m}^2}}\right), \quad (33)$$

$$\hat{\chi} = (\rho - 2m + Q)\left(\frac{1}{\chi^2} + \Lambda'\right)$$
$$\qquad - \alpha\sigma^2\left[G_{\boldsymbol{AA}^\mathsf{T}}(-\lambda\Lambda) - (\lambda\Lambda)\cdot G'_{\boldsymbol{AA}^\mathsf{T}}(-\lambda\Lambda)\right]\Lambda', \quad (34)$$

$$\Lambda = \frac{1}{\chi}\left(1 - \alpha\left[1 - (\lambda\Lambda)\cdot G_{\boldsymbol{AA}^\mathsf{T}}(-\Lambda\lambda)\right]\right), \quad (35)$$

need to be solved given the definitions

$$\Lambda' \triangleq \frac{\partial\Lambda}{\partial\chi} = -\left[\frac{1-\alpha}{\Lambda^2} + \alpha\lambda^2 G'_{\boldsymbol{AA}^\mathsf{T}}(-\Lambda\lambda)\right]^{-1}, \quad (36)$$

$$\hat{m} \triangleq \frac{1}{\chi} - \Lambda = \frac{\alpha}{\chi}\left[1 - (\lambda\Lambda)\cdot G_{\boldsymbol{AA}^\mathsf{T}}(-\Lambda\lambda)\right]. \quad (37)$$

The function $G_{\boldsymbol{AA}^\mathsf{T}}(s)$ is the Stieltjes transform of the asymptotic distribution $F_{\boldsymbol{AA}^\mathsf{T}}(x)$ and $G'_{\boldsymbol{AA}^\mathsf{T}}(s)$ is the derivative of $G_{\boldsymbol{AA}^\mathsf{T}}(s)$ w.r.t. the argument.

Proof: See Appendix B. ∎

Remark 2. Due to (27) and (33), the rotationally invariant and T-orthogonal setups have exactly the same form for variables $\{m, Q, \chi\}$. Hence, the choice of random matrix ensemble does not affect these variables, except for adding the indexes $t = 1, \ldots, T$ to them in the case of T-orthogonal setup.

Notice that in contrast to Definition 2, the above proposition uses the eigenvalue distribution of $\boldsymbol{AA}^\mathsf{T}$ instead of $\boldsymbol{A}^\mathsf{T}\boldsymbol{A}$. This is more convenient in the present setup since $M \leq N$, so we do not have to deal with the zero eigenvalues. Compared to the T-orthogonal case, here we have a fixed set of three equations and unknowns to solve, regardless of $\{\rho_t\}$ and the partition of the source vector. Thus, if one knows the Stieltjes transform $G_{\boldsymbol{AA}^\mathsf{T}}(s)$ and it is (once) differentiable with respect to the argument, the required parameters can be solved numerically from the coupled equations given in Proposition 2. For some $G_{\boldsymbol{AA}^\mathsf{T}}(s)$, however, the equations can be reduced analytically to simpler forms that allow for efficient numerical evaluation of the MSE, as seen in the following two examples.

Example 2. Recall the *row-orthogonal setup*. For this case $\boldsymbol{D} = \alpha^{-1}\boldsymbol{I}_M$ and, thus, the Stieltjes transform of $F_{\boldsymbol{AA}^\mathsf{T}}$ reads

$$G_{\boldsymbol{AA}^\mathsf{T}}(s) = \frac{1}{\alpha^{-1} - s}, \quad (38)$$

so that $G'_{\boldsymbol{AA}^\mathsf{T}}(s) = (\alpha^{-1} - s)^{-2}$. Plugging the above in Proposition 2 provides

$$\Lambda = \frac{\lambda - \alpha^{-1}\chi + \sqrt{-4\lambda\chi + (\lambda + \alpha^{-1}\chi)^2}}{2\lambda\chi}, \quad (39)$$

$$\hat{m} = \frac{\lambda + \alpha^{-1}\chi - \sqrt{-4\lambda\chi + (\lambda + \alpha^{-1}\chi)^2}}{2\lambda\chi}, \quad (40)$$

$$\Lambda' = -\left[\frac{1-\alpha}{\Lambda^2} + \frac{\alpha\lambda^2}{(\alpha^{-1} + \lambda\Lambda)^2}\right]^{-1}, \quad (41)$$

$$\hat{\chi} = (\rho - 2m + Q)\left(\frac{1}{\chi^2} + \Lambda'\right) - \frac{\sigma^2}{(\alpha^{-1} + \lambda\Lambda)^2}\Lambda', \quad (42)$$

so that the MSE can be obtained by solving the set $\{\chi, \hat{\chi}\}$ of equations given by (33) and (34). As expected, for the special case of uniform sparsity $\rho_t = \rho$ and $\alpha = 1/T$, the row-orthogonal and T-orthogonal setups give always the same MSE (see Remark 5 in Appendix B).

In the above example, we first used (38) in (35) and (37) to solve Λ and $\hat{m}$ as functions of χ. Plugging then G' into (36) provides immediately Λ', and $\hat{\chi}$ follows similarly from (34). However, if the form of G is more cumbersome, it may be more convenient to compute the parameters in slightly different order as demonstrated in the next example that considers a generalized version of the standard Gaussian CS setup.

Example 3. Consider the rotationally invariant setup where the elements of $\boldsymbol{A}$ are zero-mean IID with variance $1/M$. The

eigenvalue spectrum of $\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}$ is then given by the Marčenko-Pastur law

$$G_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}(s) = \frac{-1 + \alpha^{-1} - s - \sqrt{-4s + (-1 + \alpha^{-1} - s)^2}}{2s}, \tag{43}$$

that provides together with Proposition 2

$$\Lambda = \frac{1}{\chi} - \frac{1}{\lambda + \alpha^{-1}\chi}, \tag{44}$$

$$\hat{m} = \frac{1}{\lambda + \alpha^{-1}\chi}, \tag{45}$$

$$\Lambda' = -\frac{1}{\chi^2} + \frac{\alpha^{-1}}{(\lambda + \alpha^{-1}\chi)^2}, \tag{46}$$

$$G'_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}(-\lambda\Lambda) = -\frac{1}{\alpha\lambda^2}\left(\frac{1 - \alpha}{\Lambda^2} + \frac{1}{\Lambda'}\right). \tag{47}$$

Plugging the above to Proposition 2 and solving $\{\chi, \hat{\chi}\}$ yields the MSE of reconstruction.

In this case, $G'_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}(s)$ is of more complex form than in Example 2, and the approach used there does not provide as simple solution as before. However, since now Λ has a particularly convenient form, we may use the definition $\Lambda' = \frac{\partial \Lambda}{\partial \chi}$ to write Λ' as a function of χ. This can be then used in (36) to obtain G' indirectly.

Remark 3. As shown in Section IV-B, Examples 2 and 3 provide the same average reconstruction error as reported in [30], which also implies that the IID case matches [28]. The benefit of Example 3 compared to [28] is that there are no integrals or expectations left to solve. Example 2, on the other hand, proved directly that for the special case of uniform sparsity $\rho_t = \rho$ and $\alpha = 1/T$, the row-orthogonal and T-orthogonal setups give always the same MSE.

The final example demonstrates the capabilities of the analytical framework for a "non-standard" ensemble that has singular values defined by geometric progression.

Example 4. Recall the geometric setup where the singular values $\sigma_m \propto \tau^{m-1}, m = 1, \ldots, M$ satisfy peak-to-average condition (8). Equivalent limiting eigenvalue spectrum of $\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}$ in the large system limit is described by the Stieltjes transform

$$G_{\boldsymbol{A}\boldsymbol{A}^T}(s) = \frac{1}{s\gamma}\ln\left(\frac{A - s}{Ae^{-\gamma} - s}\right) - \frac{1}{s}, \tag{48}$$

$$G'_{\boldsymbol{A}\boldsymbol{A}^T}(s) = -\frac{1}{s}G_{\boldsymbol{A}\boldsymbol{A}^T}(s) - \frac{1}{s}\frac{A(e^{-\gamma} - 1)}{\gamma(Ae^{-\gamma} - s)(A - s)}, \tag{49}$$

where $A = \kappa/\alpha$ and γ satisfies

$$\kappa = \frac{\gamma}{1 - e^\gamma}, \tag{50}$$

for some given κ. For details on how to generate the geometric ensemble for simulations and how the Stieltjes transform arises for this setup, see Appendix D.

B. Numerical Examples

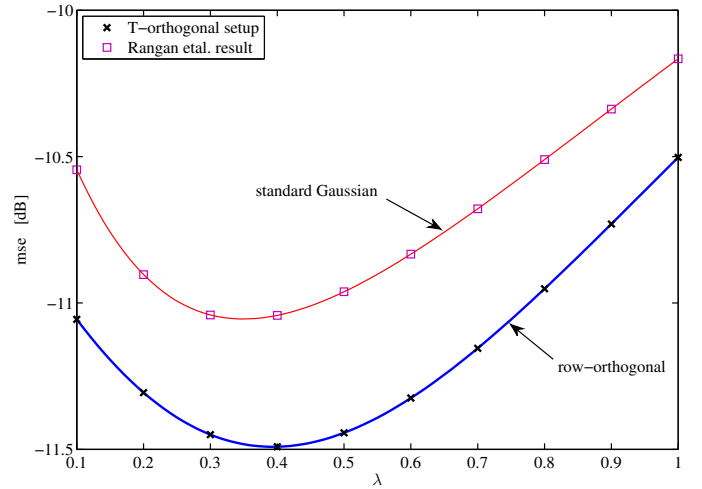Having obtained the theoretical performance of various matrix ensembles, we now examine the behavior of the MSE

Fig. 1. Average MSE in decibels $\mathsf{mse} \to 10\log_{10}(\mathsf{mse})$ dB as a function of the tuning parameter λ. Solid lines given by Examples 2 and 3. Markers for the standard Gaussian setup are obtained from the analytical results provided in [28, Section V] and by Example 1 for the T-orthogonal case. All results should thus be considered to be in the LSL. As predicted by the analysis, the lines and markers match perfectly. Parameter values: $\alpha = 1/3, \rho = 0.15, \sigma^2 = 0.1$.

in some chosen setups numerically. First we consider the case of uniform density and the MSE of reconstruction as a function of the tuning parameter λ, as shown in Fig. 1. The solid lines depict the performances of row-orthogonal and standard Gaussian setups, as given by Examples 2 and 3. The markers, on the other hand, correspond to the result obtained by Rangan et al. [28, Section V-B] and Example 1 given in this paper. As expected, the solid lines and markers match perfectly although the analytical results are represented in a completely different form. It is important to notice, however, that we plot here the MSE (in decibels) while normalized mean square error (in decibels) mse/ρ is used in [28]. Also, the definition of signal-to-noise ratio SNR_0 there would correspond to value ρ/σ^2 in this paper. Comparing the two curves in Fig. 1 makes it clear that the orthogonal constructions provide superior MSE performance compared to the standard Gaussian setup. It is also worth pointing out that the optimal value of λ depends on the choice of the matrix ensemble.

In the next experiment we consider the case of "localized sparsity" where all non-zero elements are concentrated in one subvector $\boldsymbol{x}_t$, namely, $\rho_t = \rho T$ for some $t \in \{1, \ldots, T\}$ and $\rho_s = 0 \ \forall s \neq t$. For simplicity, we take the simplest case of $T = 2$ and choose the overall sparsity to be $\rho = 0.2$. The average mean square error vs. inverse noise variance $1/\sigma^2$ of this case is depicted in Fig. 2. For clarity of presentation, the variables related to both axes are given in a decibel form, that is, $x \to 10\log_{10}(x)$ dB, where $x \in \{\mathsf{mse}, \sigma^{-2}\}$. The tuning parameter λ is chosen for each point so that the lowest possible MSE is obtained for all considered methods. Due to the simple form of the analytical results, this is easy to do numerically. Examining Fig. 2 reveals a surprising phenomenon, namely, for small noise variance the T-orthogonal setup gives the lowest average MSE while for more noisy setups it is the row-orthogonal ensemble that achieves the best reconstruction performance. The universality of this behavior
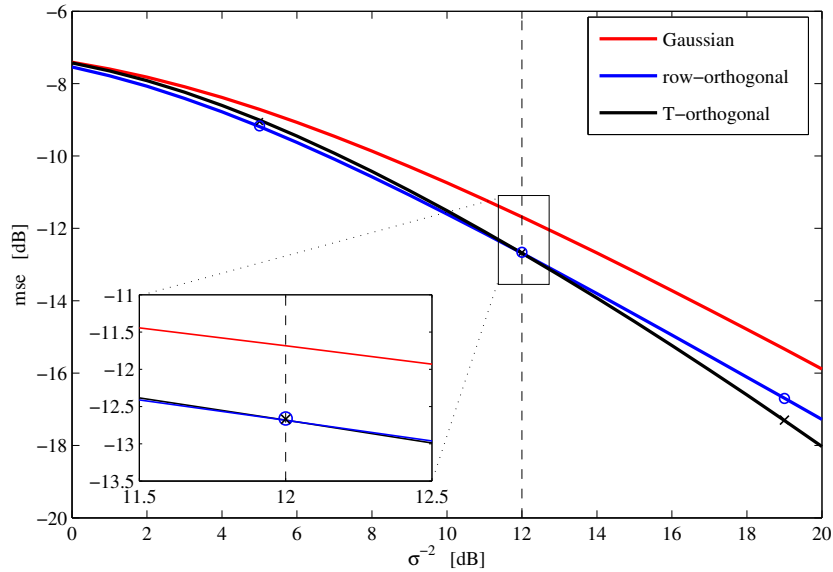
Fig. 2. Average MSE vs. inverse noise variance $1/\sigma^2$, where both quantities are presented in decibels, that is, $x \to 10\log_{10}(x)$ dB. For each point, the value of λ that minimizes the MSE is chosen numerically. The parameter values are $\alpha = 1/2$, $\rho = 0.2$ and localized sparsity is considered, that is, either $\rho_1 = 0.4$ and $\rho_2 = 0$ or vice versa. The dashed line at $\sigma^{-2} = 12$ dB represents the point where the MSE performance of T-orthogonal and row-orthogonal ensembles approximately cross each other. Markers at $\sigma^{-2} = 5, 12, 19$ dB are obtained by using cvx for reconstruction and averaging over $100\,000$ realizations of the problem.
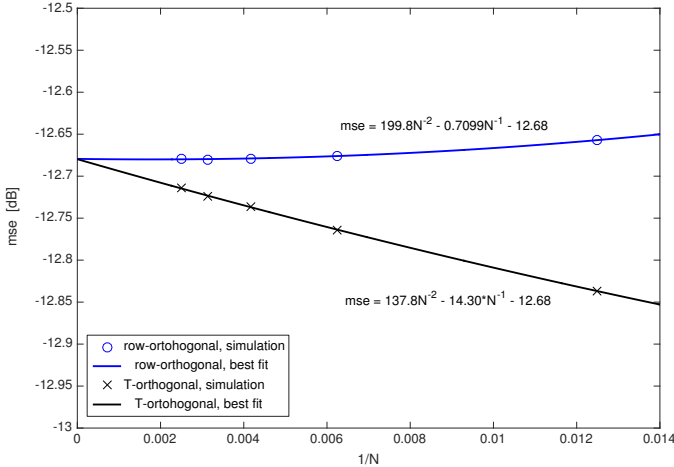

Fig. 3. Experimental assessment of the MSE for the point $\sigma^{-2} = 12$ dB in Fig. 2. Markers correspond to simulated values using SpaSM with path-following optimization over λ. Problem sizes $N = 80, 160, \ldots, 400$, each averaged over 10^6 realizations. The experimental data were fitted with a quadratic function of $1/N$ and plotted with solid lines. Extrapolation for $M = \alpha N \to \infty$ provides the estimates for the asymptotic MSE that agrees well with the replica prediction for the RS ansatz mse $= -12.685$ dB.

for other parameter values is left as an open question for future research. The point where these two ensembles give approximately the same MSE for the given setup is located at $\sigma^{-2} = -12$ dB. Hence, for optimal performance, if one is given the choice of these three matrix ensembles, the row-orthogonal setup should be chosen when on the left hand side of the dashed line and T-orthogonal setup otherwise. At any point in the figure, however, the standard Gaussian setup gives the worst reconstruction in the MSE sense and should never been chosen if such freedom is presented.

To illustrate how the experimental points in Fig. 2 were

obtained, we have plotted in Fig. 3 the average MSE of the point $\sigma^{-2} = 12$ dB, for which the asymptotic prediction given by the replica method is mse $= -12.685$ dB. To obtain more accurate results, the experimental data is averaged now over 10^6 realizations and estimates are obtained by using SpaSM [68] that provides an efficient MATLAB implementation of the least angle regression (LARS) algorithm [69]. The simulated data is fitted with a quadratic function of N^{-1} and the estimates for the asymptotic MSEs are obtained by extrapolating $M = \alpha N \to \infty$. The end result for for both the row-orthogonal and T-orthogonal ensembles is mse $= -12.68$ dB, showing that the replica analysis provides a good approximation of the MSE for large systems. One can also observe that the simulations approach the asymptotic result relatively fast and for realistic system sizes the match is already very good. Albeit the convergence behavior depends somewhat on the noise variance σ^2, the present figure is a typical example of what was observed in our simulations.

Finally, we use the geometric setup to examine the robustness of LASSO against varying peak-to-average ratio κ introduced in (8). Substituting the formulas given in Example 4 to Proposition 2 provides the MSE of reconstruction for LASSO as given in Fig. 4. The system parameters are $\alpha = 1/2$, $\rho = 0.2$ and σ^2 is chosen to match the markers in Fig. 2. As mentioned earlier, the geometric setup reduces to the row-orthogonal one when $\kappa \to 1$, which can verified by comparing the $\kappa = 1$ values to the markers of the row-orthogonal curve in Fig. 2. We have also included additional simulation points at $\kappa = 5$, obtained by using SpaSM as in Fig. 3. One can observe that the performance degradation with increasing κ for the LASSO problem is relatively graceful compared to the abrupt transition of GAMP observed in [66]. Hence, the algorithmic considerations are indeed very important, as studied therein.
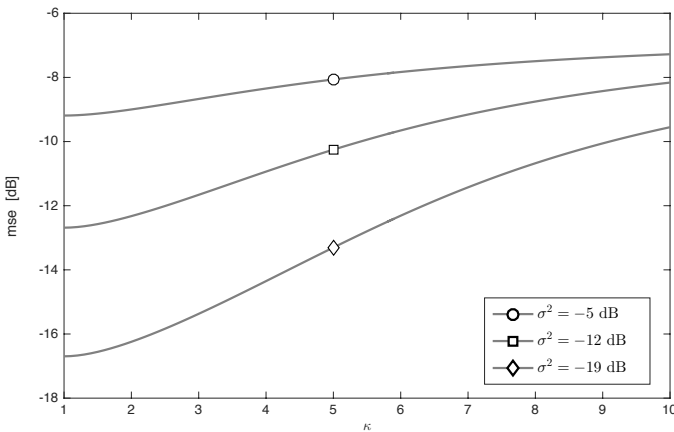
Fig. 4. Average MSE vs. the peak-to-average ratio κ for the geometric setup. Markers are obtained by extrapolating simulated values as in Fig. 3 and optimal λ is used for each point. Parameter values are $\alpha = 1/2, \rho = 0.2$, and the points at $\kappa = 1$ match the markers of the row-orthogonal setup in Fig. 2.

We also observe that as expected, the MSE is a monotonic increasing function of κ for all σ^2, highlighting the fact that sensing matrices with "flat" eigenvalue distributions are in general good for reconstruction of noisy sparse signals.

IV. EXTENSIONS AND A SKETCH OF A DERIVATION WITH THE REPLICA METHOD

Although the LASSO reconstruction examined in the previous section is one of the most important special cases of regularized LS problem (2), one might be interested in expanding the scope of analysis to some other cases. In fact, up to a certain point, the evaluation of free energy (16) is independent of the choice of regularization $c(\boldsymbol{x})$ as well as the marginal distribution of the non-zero components of the source vector. For this end, let us assume in the following that the source vector has elements drawn independently according to (6) where $\pi(x)$ is a suitable PDF with zero-mean, unit variance and finite moments. Note, however, that in order to obtain the final saddle-point conditions, one needs to make a choice about both c and π in the end.

A. Sketch of a Replica Analysis

To provide a brief sketch how the results in previous section were obtained and elucidate where the choice of regularization (10) affects the analysis, let us consider for simplicity the rotationally invariant ensemble with source vector that has uniform sparsity, i.e., we set $T = 1$ and $\rho = \rho_1$. By Appendices A and B we know that (16) can be written under RS ansatz as

$$f = -\lim_{n \to 0^+} \frac{\partial}{\partial n} \lim_{\beta, N \to \infty} \frac{1}{\beta N} \ln \Xi_{\beta, N}(n), \qquad (51)$$

where n is the number of replicas in the system. To characterize $\Xi_{\beta, N}(n)$, we consider the simplest case of RS overlap matrix (18) and remind the reader that albeit this choice may seem intuitively reasonable, it is known to be incorrect in some cases (see Introduction for further discussion).

For given replica symmetric $\boldsymbol{Q}$, we may compute its probability weight[6]

$$p_{\beta, N}(\boldsymbol{Q}; n) \propto \int [\mathcal{V}_\beta(\hat{\boldsymbol{Q}}; n)]^N e^{N\beta n(\hat{Q}Q - \hat{\chi}\chi - 2\hat{m}m)/2} d\hat{\boldsymbol{Q}}, \qquad (52)$$

where $\chi = \beta(Q - q)$. With some abuse of notation, we used above $\boldsymbol{Q}$ as a shorthand for the set $\{\chi, Q, m\}$ and similarly $\hat{\boldsymbol{Q}}$ for the auxiliary parameters $\{\hat{\chi}, \hat{Q}, \hat{m}\}$. Given that the prior of $\boldsymbol{x}^0$ factorizes and the regularization that separates as (10) the auxiliary term $\mathcal{V}_\beta(\hat{\boldsymbol{Q}}; n)$ is given in (53) at the top of the next page where $\mathrm{D}z = dz\, e^{z^2/2}/\sqrt{2\pi}$ and $p(x^0)$ is assumed to be of the form (6) with $\rho = \rho_1$ and $T = 1$. It is important to realize that the replica analysis could, in principle, be done with general forms of $c(\boldsymbol{x})$ and $p(\boldsymbol{x})$ but then (53) would have vectors in place of scalars. This creates computational problems as explained in short. Note also that we have taken here a slightly different route (on purpose) compared to the analysis carried out in the Appendices. The approach there is more straightforward mathematically but the methods presented here give some additional insight to the solution and provide connection to the results given in [28] and [30].

Next, define a function

$$H_{\beta, \lambda}(\sigma^2, \nu) = -\frac{1 + \ln(\beta \nu) - \alpha \ln \lambda}{2}$$
$$+ \frac{1}{2} \operatorname*{extr}_{\Lambda} \{\Lambda(\beta \nu) - (1 - \alpha) \ln \Lambda - \alpha \mathsf{E} \ln(\Lambda \beta \sigma^2 + \Lambda \lambda + x)\}, \qquad (54)$$

where $\alpha = M/N$, $\nu > 0$ and the expectation of the ln-term is w.r.t. the asymptotic eigenvalue distribution $F_{\boldsymbol{A}\boldsymbol{A}^\top}(x)$. Then

$$\Xi_{\beta, N}(n) = \int d\boldsymbol{Q}\, p_{\beta, N}(\boldsymbol{Q}; n)$$
$$\times e^{N H_{\beta, \lambda}(n\sigma^2, n(r - 2m + q) + Q - q) + N(n-1)H_{\beta, \lambda}(0, Q-q)}, \qquad (55)$$

where the exponential term involving function H arises from the expectation of (17) w.r.t. noise $\boldsymbol{w}$ and sensing matrix $\boldsymbol{A}$ given $\{\boldsymbol{x}^a\}$ and $\boldsymbol{Q}$. The relevant matrix integral is proved in Appendix C-B and more details can be found in the derivations following Lemma 1 in Appendix A. In the limit of vanishing temperature and increasing system size $\beta, N \to \infty$, saddle point method[7] may be employed to assess the integrals over $\boldsymbol{Q}$ and $\hat{\boldsymbol{Q}}$. Taking also the partial derivative w.r.t. n and then letting $n \to 0$ provides

$$f = \operatorname*{extr}_{\chi, Q, m, \hat{\chi}, \hat{Q}, \hat{m}} \left\{ \hat{m}m - \frac{\hat{Q}Q}{2} + \frac{\hat{\chi}\chi}{2} \right.$$
$$+ \iint p(x^0)\phi(z\sqrt{\hat{\chi}} + \hat{m}x^0; \hat{Q})dx^0 \mathrm{D}z \qquad (56)$$
$$\left. - \lim_{\beta \to \infty} \frac{1}{\beta} \lim_{n \to 0} \frac{\partial}{\partial n} H_{\beta, \lambda}(n\sigma^2, n(\rho - 2m + q) + Q - q) \right\},$$

---

[6]We have omitted a term vanishing multiplicative terms and $-n^2 \beta \hat{\chi}q/2$ in the exponent since it does not affect the free energy, see (134) in Appendix A.

[7]For more information, see for example [70, Ch. 6] and [71, Ch. 12.7], or for a gentle introduction [72]. Note that in our case when $\beta, N \to \infty$, the correction terms in front of the exponentials after saddle point approximation vanish due to the logarithm and division by βN in (51). With some abuse of notation, we have simply omitted them in the paper to avoid unnecessary distractions.

$$\mathcal{V}_\beta(\hat{\boldsymbol{Q}}; n) = \int p(x^0) \left(\prod_{a=0}^n \mathrm{d}x^a \right) \exp \left(-\beta \sum_{a=1}^n c(x^a) \right) \exp \left[-\frac{\beta\hat{Q}}{2} \sum_{a=1}^n (x^a)^2 + \beta\hat{m}x^0 \sum_{a=1}^n x^a + \frac{1}{2} \left(\beta\sqrt{\hat{\chi}} \sum_{a=1}^n x^a \right)^2 \right]$$

$$= \iint p(x^0) \left\{ \int \exp \left(\beta \left[-\frac{\hat{Q}}{2} x^2 + (z\sqrt{\hat{\chi}} + \hat{m}x^0)x - c(x) \right] \right) \mathrm{d}x \right\}^n \mathrm{d}x^0 \mathrm{D}z, \tag{53}$$

where we defined a scalar function

$$\phi(y; \hat{Q}) = \min_x \left\{ \frac{\hat{Q}}{2} x^2 - yx + c(x) \right\}. \tag{57}$$

Note that due to the limit $n \to 0$ in (56), the function H has the arguments $\sigma^2 = 0$ and $\nu = Q - q$ when it comes to solving the extremization over Λ in (54). The value of Λ which provides a solution for this is denoted Λ^* in the following. We also remark that (57) is the only part of free energy that directly depends on the choice of cost function c. If we would have chosen $p(\boldsymbol{x})$ that is not a product distribution and c that did not separate as in (10), we would need to consider here a multivariate optimization over $\boldsymbol{x}$. In fact, the dimensions should grow without bound by the assumptions of the analysis and, hence, the problem would be infeasible in general. In practice one could consider some large but finite setting and use it to approximate the infinite dimensional limit, or concentrate on forms of $c(\boldsymbol{x})$ and $p(\boldsymbol{x})$ that have only a few local dependencies. For simplicity we have chosen to restrict ourselves to the case where the problem fully decouples into a single dimensional one.

We can get an interpretation of the remaining parameters as follows. First, let

$$\langle h(x); y \rangle_\beta = \frac{1}{Z_\beta(y)} \int h(x) \mathrm{e}^{-\beta[\hat{Q}x^2/2 - yx + c(x)]} \mathrm{d}x \tag{58}$$

be a posterior mean of some function $h(x)$. Note that the structure of this scalar estimator is essentially the same as the vector valued counterpart given in (13) and (14). If we denote the x that minimizes (57) by $\hat{x}(y; \hat{Q})$ then clearly

$$\langle h(x); y \rangle_{\beta \to \infty} = h(\hat{x}(y; \hat{Q})), \tag{59}$$

and also

$$\hat{x}(y; \hat{Q}) = -\frac{\partial}{\partial y} \phi(y; \hat{Q}). \tag{60}$$

We thus obtain from (53) with a little bit calculus that as $N, \beta \to \infty$ and $n \to 0$, the terms that yield the mean square error $\mathrm{mse} = \rho - 2m + Q$ are given by

$$m = \iint x^0 \hat{x}(z\sqrt{\hat{\chi}} + \hat{m}x^0; \hat{Q}) p(x^0) \mathrm{d}x^0 \mathrm{D}z, \tag{61}$$

$$Q = \iint \left[\hat{x}(z\sqrt{\hat{\chi}} + \hat{m}x^0; \hat{Q}) \right]^2 p(x^0) \mathrm{d}x^0 \mathrm{D}z, \tag{62}$$

where we now have $y = z\sqrt{\hat{\chi}} + \hat{m}x^0$ and z is a standard Gaussian RV. Similarly, the extremum condition for the variable χ reads

$$\chi = \frac{1}{\sqrt{\hat{\chi}}} \int z \int \hat{x}(z\sqrt{\hat{\chi}} + \hat{m}x^0; \hat{Q}) p(x^0) \mathrm{d}x^0 \mathrm{D}z$$

$$= \iint \frac{\partial}{\partial(z\sqrt{\hat{\chi}})} \hat{x}(z\sqrt{\hat{\chi}} + \hat{m}x^0; \hat{Q}) p(x^0) \mathrm{d}x^0 \mathrm{D}z, \tag{63}$$

where the latter equality is obtained using integration by parts formula. For many cases of interest, these equations can be evaluated analytically, or at least numerically, and they provide the single body representation of the variables in (19). If one substitutes (6) with $\pi(x) = g_x(0; 1)$ in the above formulas along with $c(x) = |x|$, some (tedious) calculus shows that the results of the previous section are recovered. An alternative way of obtaining the parameters is provided in Appendices A and B.

Given $\{\hat{m}, \hat{Q}, \hat{\chi}\}$, one may now obtain the MSE of the original reconstruction described in Section II by considering an equivalent scalar problem, namely,

$$\mathrm{mse} = \rho - 2m + Q$$

$$= \iint \left| x^0 - \hat{x}(z\sqrt{\hat{\chi}} + \hat{m}x^0; \hat{Q}) \right|^2 p(x^0) \mathrm{d}x^0 \mathrm{D}z. \tag{64}$$

We may thus conclude that the minimizing x in (57) is the $\hat{x}$ above, which can be interpreted as the output of a regularized LS estimator that postulates $y = \hat{Q}x + \hat{Q}^{1/2}z$, or equivalently,

$$y = x + \hat{Q}^{-1/2}z, \tag{65}$$

while the true model is $y = \hat{m}x^0 + z\sqrt{\hat{\chi}}$, or equivalently,

$$y = x^0 + z\frac{\sqrt{\hat{\chi}}}{\hat{m}}. \tag{66}$$

In the notation of [30] (resp. [28]), we thus have the relations $\xi \leftrightarrow \hat{Q}$ ($\leftrightarrow \lambda_p$) and $\eta \leftrightarrow \hat{m}^2/\hat{\chi}$ ($\leftrightarrow \mu^{-1}$) between the parameters. The above also implies $\hat{m} = \hat{Q}$ as is indeed verified later in (70). We shall expand on the connection between this paper and [28], [30] in Section IV-B.

Let us denote Λ^* for the solution of extremization in (54) under condition $\sigma^2 = 0$, namely,

$$\Lambda^* - \frac{1}{\chi} = -\frac{\alpha}{\chi} \left[1 - (\lambda\Lambda^*) \cdot G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*) \right], \tag{67}$$

which is the same condition as (35) in Proposition 2. Then we can plug in (56) the identity

$$\lim_{\beta \to \infty} \frac{1}{\beta} \lim_{n \to 0} \frac{\partial}{\partial n} H_{\beta,\lambda}(n\sigma^2, \nu(n))$$

$$= -\frac{\alpha\sigma^2\Lambda^*}{2} G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*) + \frac{\rho - 2m + Q}{2} \left(\Lambda^* - \frac{1}{\chi} \right), \tag{68}$$

where we used the fact that in the LSL, $r \to \rho$ in (18) by the weak law of large numbers. The RS free energy thus becomes

$$f = \mathop{\mathrm{extr}}_{\chi, Q, m, \hat{\chi}, \hat{Q}, \hat{m}} \left\{ \hat{m}m - \frac{\hat{Q}Q}{2} + \frac{\hat{\chi}\chi}{2} \right.$$

$$+ \iint p(x^0) \phi(z\sqrt{\hat{\chi}} + \hat{m}x^0; \hat{Q}) \mathrm{d}x^0 \mathrm{D}z \tag{69}$$

$$\left. + \frac{\alpha\sigma^2\Lambda^*}{2} G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*) - \frac{\rho - 2m + Q}{2} \left(\Lambda^* - \frac{1}{\chi} \right) \right\},$$

where the parameters $\{m, Q, \chi\}$ satisfy (61)–(63) Note that so-far we have not made any assumptions about the details of the function $c(x)$, which means that (69) is valid for any type of regularization that separates as given in (10). In fact, we can go even further and solve the partial derivatives w.r.t. variables m and Q, which reveals that

$$\hat{Q} = \hat{m} = \frac{1}{\chi} - \Lambda^*. \tag{70}$$

With some additional effort, one also finds that

$$\hat{\chi} = (\rho - 2m + Q)\left(\frac{1}{\chi^2} + \frac{\partial \Lambda^*}{\partial \chi}\right)$$
$$- \alpha\sigma^2\big[G_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}(-\lambda\Lambda^*) - (\lambda\Lambda^*)\cdot G'_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}(-\lambda\Lambda^*)\big]\frac{\partial \Lambda^*}{\partial \chi} \tag{71}$$

holds for the RS free energy with

$$\frac{\partial \Lambda^*}{\partial \chi} = -\left[\frac{1-\alpha}{(\Lambda^*)^2} + (\alpha\lambda^2)\cdot G'_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}(-\lambda\Lambda^*)\right]^{-1}. \tag{72}$$

It is now easy to see that for the rotationally invariant setup, our initial assumption of uniform sparsity, i.e., $T = 1$ and $\rho = \rho_1$ is not necessary and the same set of equations is obtained for arbitrary sparsity pattern $\{\rho_t\}$ that satisfies $\rho = T^{-1}\sum_t \rho_t$. Similarly, we may obtain the equivalent representation for the T-orthogonal setup considered in Proposition 1.

Remark 4. Comparing (67) together with (70)–(72) to the saddle-point conditions (34)–(37) given in Proposition 2 shows that the choice of c or the marginal PDF of the non-zero elements in the source vector in (6) has no direct impact on the expressions that provide the variables $\{\hat{m}, \hat{Q}, \hat{\chi}\}$. Hence, the form of these conditions is the same for all setups where the sensing matrix is from the same ensemble. The parameters $\{m, Q, \chi\}$ on the other hand are affected by the choice of regularization and source distribution. As stated in Remark 2, the effect of sensing matrix ensemble is the reverse, namely, for fixed c and $p(\boldsymbol{x}^0)$, the form of $\{m, Q, \chi\}$ is always the same while $\{\hat{m}, \hat{Q}, \hat{\chi}\}$ can have different form depending on the choice of the measurement matrix.

B. Alternative Representation of Rotationally Invariant Case and Comparison to Existing Results

The saddle-point condition for the rotationally invariant case is described in Proposition 2 in terms of the Stieltjes transform of $F_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}$. In [30] the HCIZ-formula is used, which makes it natural to express the results in terms of the R-transform of $F_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}$. Furthermore, different sets of auxiliary variables are used in these two papers. In this section we sketch an alternative representation of Proposition 2 that is equivalent to [30] up to some minor scaling factors. This also implies that apart from minor differences in scalings, our results for the IID case are also equivalent to [28] as explained in [30, Sec. IV-C] and shown in Fig. 1.

Let us first consider the conditions enforced by the matrix integration formula (54) through Λ. By the remark following (57), we know that the relevant terms can also be obtained

from (195) by setting $\sigma^2 = 0$ and $\chi = \beta(Q - q)$, namely

$$H_{\beta,\lambda}(\sigma^2 = 0, \nu = Q - q)$$
$$\simeq \frac{1}{2}\operatorname*{extr}_\Lambda\left\{\Lambda\chi - \int \ln(x + \lambda\Lambda)\mathrm{d}F_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}(x)\right\}, \tag{73}$$

where we have omitted the terms that do not depend on Λ. The solution to the extremization then provides the condition (we write $\Lambda = \Lambda^*$ here for simplicity)

$$\chi = \lambda \int \frac{1}{x + \lambda\Lambda}\mathrm{d}F_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}(x) = \lambda G_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}(-\lambda\Lambda) \tag{74}$$

$$\iff \Lambda = -\frac{1}{\lambda}G_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}^{-1}\left(\frac{\chi}{\lambda}\right), \tag{75}$$

where $G_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}^{-1}\big(G_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}(s)\big) = s$ is the functional inverse of the Stieltjes transform. Note that this also implies

$$G_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}(-\lambda\Lambda) = G_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}\left(G_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}^{-1}\left(\frac{\chi}{\lambda}\right)\right) = \frac{\chi}{\lambda}. \tag{76}$$

Using the definition $\mathsf{R}_X(z) = G_X^{-1}(-z) - z^{-1}$ of the R-transform in (75) yields

$$\frac{1}{\lambda}\mathsf{R}_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}\left(-\frac{\chi}{\lambda}\right) = \frac{1}{\chi} - \Lambda. \tag{77}$$

On the other hand, we know from (70) that a solution to (69) satisfies $\hat{Q} = \hat{m} = \chi^{-1} - \Lambda$, so that

$$\hat{Q} = \hat{m} = \frac{1}{\lambda}\mathsf{R}_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}\left(-\frac{\chi}{\lambda}\right) \tag{78}$$

is the saddle-point condition for $\hat{Q}$ and $\hat{m}$ in terms of the R-transform. Note that compared to the Stieltjes-transform that is related to $\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}$ at the saddle-point solution, the R-transform describes the eigenvalue spectrum of $\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}$. The condition (78) matches [30, (131)] apart from a slightly different placements of regularization parameters so that $\hat{m} \leftrightarrow \xi$ as already remarked earlier. The above also suggests that apart from scalings by the regularization parameter $\chi \leftrightarrow \mathsf{E}[\sigma^2(Y; \xi)]$, which can also be inferred from [28, Lemma 9].

Finally, we know from the above developments and [30, Appendix B] that $\hat{\chi} \leftrightarrow f^\star$ should hold if the results are equal. To this end, let us examine the last line of (69), namely,

$$\frac{\alpha\sigma^2\Lambda}{2}G_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}(-\lambda\Lambda) - \frac{\rho - 2m + Q}{2}\left(\Lambda - \frac{1}{\chi}\right), \tag{79}$$

and substitute (75)–(77) there. Considering the end result as a function of χ, we obtain

$$\varphi(\chi) = \frac{\alpha\sigma^2}{2}\left[\frac{1}{\chi} - \frac{1}{\lambda}\mathsf{R}_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}\left(-\frac{\chi}{\lambda}\right)\right]\frac{\chi}{\lambda}$$
$$+ \frac{\rho - 2m + Q}{2\lambda}\mathsf{R}_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}\left(-\frac{\chi}{\lambda}\right) \tag{80}$$

$$\simeq \frac{1}{2}\left(\frac{\rho - 2m + Q}{\lambda} - \frac{\alpha\sigma^2\chi}{\lambda^2}\right)\mathsf{R}_{\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}}\left(-\frac{\chi}{\lambda}\right), \tag{81}$$

where (81) is obtained by omitting the terms that do not depend on χ. We are interested in the point where the partial

derivative in (69) w.r.t. χ vanishes, that is,

$$
\begin{aligned}
\hat{\chi} &= -2\frac{\partial}{\partial \chi}\varphi(\chi) \\
&= \left(\frac{\rho - 2m + Q}{\lambda^2}\right)\mathsf{R}'_{\boldsymbol{A}^\mathsf{T}\boldsymbol{A}}\left(-\frac{\chi}{\lambda}\right) \\
&\quad + \left(\frac{\alpha\sigma^2}{\lambda^2}\right)\frac{\partial}{\partial \chi}\left[\chi\mathsf{R}_{\boldsymbol{A}^\mathsf{T}\boldsymbol{A}}\left(-\frac{\chi}{\lambda}\right)\right] \\
&= \left(\frac{\alpha\sigma^2}{\lambda^2}\right)\mathsf{R}_{\boldsymbol{A}^\mathsf{T}\boldsymbol{A}}\left(-\frac{\chi}{\lambda}\right) \\
&\quad + \frac{1}{\lambda^2}\mathsf{R}'_{\boldsymbol{A}^\mathsf{T}\boldsymbol{A}}\left(-\frac{\chi}{\lambda}\right)\left(\rho - 2m + Q - \frac{\chi\alpha\sigma^2}{\lambda}\right). \quad (82)
\end{aligned}
$$

Thus we have a formula for $\hat{\chi}$ as a function of χ and mse, expressed in terms of the R-transform (and its derivative R'). Comparing to [30, (195)] we see that the expressions are the same apart from minor scalings. The differences can be explained by noticing that in [30]: 1) the noise variance is $\sigma^2 = 1$, 2) $\lambda = \gamma^{-1}$ by definition, and 3) the matrices are square so that $\alpha = 1$. Thus, we conclude that $\hat{\chi} \leftrightarrow f^\star$ and Proposition 2 is indeed identical to [30], just expressed differently.

C. Regularization with ℓ_2-norm and "zero-norm"

As a first example of regularization other than ℓ_1-norm, consider the case when

$$
c(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|^2 = \sum_{j=1}^{N}\frac{x_j^2}{2}. \quad (83)
$$

This regularization implies that in the MAP-framework, the desired signal is postulated to have a standard Gaussian distribution. It is thus not surprising that such an assumption reduces the estimate (14) to the standard linear form

$$
\hat{\boldsymbol{x}} = \boldsymbol{A}^\mathsf{T}(\boldsymbol{A}\boldsymbol{A}^\mathsf{T} + \lambda\boldsymbol{I}_M)^{-1}\boldsymbol{y}, \quad (84)
$$

which is independent of the parameter β. For the replica analysis one obtains from (65) and (66)

$$
\hat{x}(y; \hat{Q}) = \frac{\hat{Q}}{1 + \hat{Q}}y \quad \stackrel{\hat{Q}=\hat{m}}{=} \quad \frac{\hat{m}x^0 + z\sqrt{\hat{\chi}}}{1 + \hat{m}}, \quad (85)
$$

which can be interpreted as mismatched linear MMSE estimation of x^0 from observation (66). From (61)–(63) we obtain the following simple result.

Example 5. Let the distribution of the non-zero elements $\pi(x)$ have zero-mean, unit variance and finite moments. For rotationally invariant setup and general problem (2) with the ℓ_2-regularization we have

$$
\mathsf{mse} = \frac{\rho + \hat{\chi}}{(1 + \hat{m})^2}, \quad (86)
$$

$$
\chi = \frac{1}{1 + \hat{m}}. \quad (87)
$$

Comparing to [30, (135)–(138)], we see that the results indeed match as discussed in Section IV-B. The value of the parameter λ that minimizes the MSE is $\lambda^* = \sigma^2/\rho$. The marginal density $\pi(x)$ of the non-zero elements (6) has no impact on the MSE. The choice of the sensing matrix, on the other hand, does

affect the MSE. In this special case though it is the same for the T-orthogonal and row-orthogonal setups — also for non-uniform sparsities.

The benefit of ℓ_1 and ℓ_2-norm regularizations is that both are of polynomial complexity. Implementation of (84) is trivial and for solving (4) one may use standard convex optimization tools like cvx [6]. However, one may wonder if there are better choices for regularization when the goal is to reconstruct a sparse vector. If we take the noise-free case as the guide, instead of say ℓ_1-norm, we should have a direct penalty on the number of non-zero elements in the source vector. We may achieve this by so-called "zero-norm" based regularization. One way to write the corresponding cost function is[8]

$$
c(\boldsymbol{x}) = \sum_{j=1}^{N}1\big(x_j \in \mathbb{R} \setminus \{0\}\big) \quad (88)
$$

$$
= \text{ number of non-zero elements in } \boldsymbol{x}. \quad (89)
$$

If the postulated and true scalar outputs are given by (65) and (66), respectively, we obtain

$$
\hat{x}(y; \hat{Q}) = y \cdot 1\big(|y| > \sqrt{2\hat{Q}}\big), \quad (90)
$$

which is just hard thresholding estimator of scalar input (66), given mismatched model (65). To proceed further, we need to fix the marginal distribution $\pi(x)$ of the non-zero components in (6). For the special case of Gaussian distribution, some algebra provides the following result.

Example 6. Let $\pi(x) = g_x(0; 1)$, that is, consider the case of Gaussian marginals (6) with the rotationally invariant setup and general problem (2) with the "zero-norm" regularization given by (88). Define a function

$$
r_0(h) = \mathrm{e}^{-h}\sqrt{\frac{h}{\pi}} + \mathcal{Q}(\sqrt{2h}). \quad (91)
$$

Then, the average MSE for the rotationally invariant setup $\mathsf{mse} = \rho - 2m + Q$ is obtained from

$$
m = 2\rho r_0\left(\frac{\hat{m}}{\hat{m}^2 + \hat{\chi}}\right), \quad (92)
$$

$$
\begin{aligned}
Q &= 2(1 - \rho)\left(\frac{\hat{\chi}}{\hat{m}^2}\right)r_0\left(\frac{\hat{m}}{\hat{\chi}}\right) \\
&\quad + 2\rho\left(\frac{\hat{m}^2 + \hat{\chi}}{\hat{m}^2}\right)r_0\left(\frac{\hat{m}}{\hat{m}^2 + \hat{\chi}}\right), \quad (93)
\end{aligned}
$$

using the condition

$$
\chi = \frac{2(1 - \rho)}{\hat{m}}r_0\left(\frac{\hat{m}}{\hat{\chi}}\right) + \frac{2\rho}{\hat{m}}r_0\left(\frac{\hat{m}}{\hat{m}^2 + \hat{\chi}}\right) \quad (94)
$$

and equations (34)–(37) in Proposition 2. Furthermore, by Remark 4, the T-orthogonal case can also be obtained easily

---

[8]As remarked also in [28, Section V-C], "zero-norm" regularization does not satisfy the requirement that the normalization constant in (11) is well-defined for any finite M, N and $\beta > 0$. Hence, appropriate limits should be considered for mathematically rigorous treatment. However, using similar analysis as given in [32, Section 3.5], it is possible to show that the RS solution for the "zero-norm" regularization is in fact always unstable due to the discontinuous nature of (90). For this reason, we skip the formalities of taking appropriate limits and report the results as they arise by directly using the form given in (88).
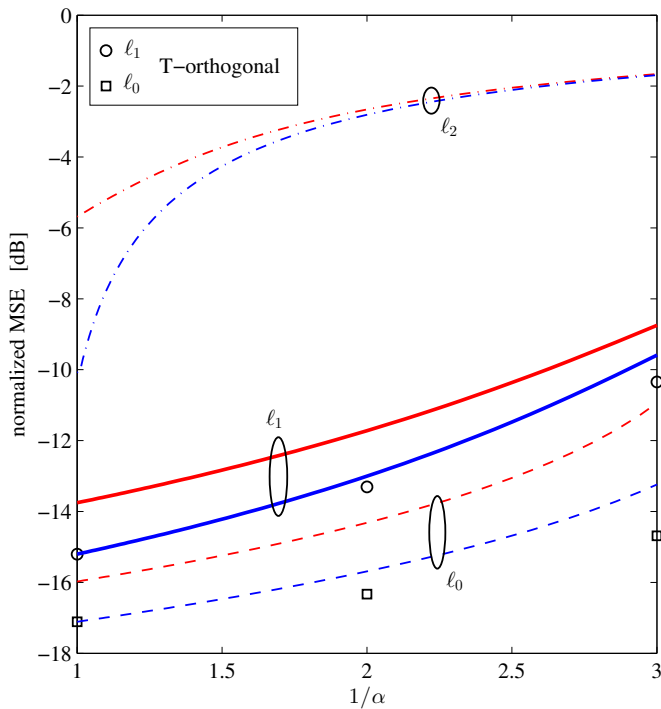
Fig. 5. Normalized mean square error mse/ρ in decibels vs. inverse compression rate $1/\alpha$ as given by Examples 5 and 6. Rotationally invariant ensemble with arbitrary sparsity pattern and T-orthogonal case with localized sparsity, namely, $\rho_t = \rho T$ for some $t \in \{1, \ldots, T\}$ and $\rho_s = 0 \ \forall s \neq t$. Thin red lines for IID sensing matrix, thick blue lines for row-orthogonal case. Noise variance $\sigma^2 = 0.01$ and the parameter λ is chosen so that the MSE is minimized. The red lines match the corresponding curves in [28, Fig. 3].

by first adding the block index t to all variables and then replacing (21) and (22) by the formulas given here. The last modification is to write (27) simply as $R_t = \chi_t \hat{m}_t$ with the χ_t given in (94).

To illustrate the above analytical results, we have plotted the *normalized mean square error* $10 \log_{10}(\mathsf{mse}/\rho)$ dB as a function of *inverse compression rate* $1/\alpha$ in Fig. 5. The axes are chosen so that the curves can be directly compared to [28, Fig. 3]. Note, however, that we plot only the region $1/3 \leq \alpha \leq 1$ (in contrast to $1/3 \leq \alpha \leq 2$ there) since this corresponds to the assumption of CS setup and one cannot construct a row-orthogonal matrix for $\alpha > 1$. It is clear that for all estimators, using row-orthogonal ensemble for measurements is beneficial compared to the standard Gaussian setup in this region. Furthermore, if the source has non-uniform sparsity and ℓ_0 or ℓ_1 regularization is used, the T-orthogonal setup (the black markers in the figure) provides an additional gain in average MSE for the compression ratios $\alpha = 1/2$ and $\alpha = 1/3$.

V. Conclusions

The main emphasis of the present paper was in the analysis of ℓ_1-regularized least-squares estimation, also known as LASSO or basis pursuit denoising, in the noisy compressed sensing setting. Extensions to ℓ_2-norm and "zero-norm" regularization were briefly discussed. Using the replica method from statistical physics, the mean square error behavior of

reconstruction was derived in the limit when the system size grows very large. By introducing some novel results for taking an expectation of a matrix in an exponential form, the previous results concerning standard Gaussian measurement matrix was extended to more general ensembles. As specific examples, row-orthogonal, geometric and T-orthogonal random matrices were considered in addition to the Gaussian one. The assumption about uniform sparsity of the source was also relaxed and blockwise sparsity levels were allowed.

The analytical results show that while the MSE of reconstruction depends only on the average sparsity level of the source for rotationally invariant cases, the performance of T-orthogonal setup depends on the individual sparsities of the sub-blocks. In case of uniform sparsity, row-orthogonal and T-orthogonal setups have provably the same performance. It was also found that while the row-orthogonal, geometric and Gaussian setups each fall under the category of rotationally invariant ensemble, that is known to have a unique perfect recovery threshold in a noise-free setting, with additive noise the MSE performance of these ensembles can be very different.

The numerical experiments revealed the fact that under all considered settings, the standard Gaussian ensemble performed always worse than the orthogonal constructions. The MSE for the geometric ensemble was found to be an increasing function of the peak-to-average ratio of the eigenvalues of the sensing matrix, suggesting that spectrally uniform sensing matrices are beneficial for recovery. When the sparsity was non-uniform, the ranking of the orthogonal constructions depended on the noise level. For highly noisy measurements, the row-orthogonal measurement matrix was found to provide the best overall MSE, while relatively clean measurements benefited from the T-orthogonal sensing matrices. These findings show that the choice of random measurement matrix does have an impact in the MSE of the reconstruction when noise is present. Furthermore, if the source does not have a uniform sparsity, the effect becomes even more varied and complex.

A natural extension of the current work is to consider Bayesian optimal recovery and its message passing approximation for the matrix ensembles that differ from the standard ones. Indeed, since the initial submission of the present paper, such algorithms have been developed and shown to benefit of matrices with structure, see for example, [56]–[59].

Appendix A
Replica Analysis of T-Orthogonal Setup

A. Free Energy

Recall the T-orthogonal setup from Definition 2 and let the partition of $\boldsymbol{A}$ be one that matches that of $\boldsymbol{x}$, i.e.,

$$\boldsymbol{A}\boldsymbol{x} = \sum_{t=1}^{T} \boldsymbol{O}_t \boldsymbol{x}_t. \tag{95}$$

We then recall (16), invoke the replica trick introduced in Section II-B, and assume that the limits commute. The normalized free energy of the system reads thus

$$f = -\frac{1}{T} \lim_{n \to 0^+} \frac{\partial}{\partial n} \lim_{\beta, M \to \infty} \frac{1}{\beta M} \ln \Xi_{\beta, M}(n), \tag{96}$$

where we denoted

$$\Xi_{\beta,M}(n) = \mathsf{E}_{\boldsymbol{w},\{\boldsymbol{O}_t\}} \int p(\boldsymbol{x}^0; \{\rho_t\}) \exp\left(-\beta \sum_{a=1}^n c(\boldsymbol{x}^a)\right)$$

$$\times \exp\left(-\frac{\beta}{2\lambda} \sum_{a=1}^n \left\|\sigma\boldsymbol{w} - \sum_{t=1}^T \boldsymbol{O}_t\Delta\boldsymbol{x}_t^a\right\|^2\right) \prod_{a=0}^n d\boldsymbol{x}^a \quad (97)$$

for notational convenience. The form of (97) implies that $\{\boldsymbol{x}^a\}$ are independently drawn according to (11), $\boldsymbol{x}^0$ has the same distribution as $\boldsymbol{x}_t$, i.e., elements drawn according to (6) for each block $t = 1, \ldots, T$, and $\Delta\boldsymbol{x}_t^a = \boldsymbol{x}_t^0 - \boldsymbol{x}_t^a$ for $a = 1, \ldots, n$. The outer expectation is w.r.t. the additive noise and measurement matrices.

For each set of random vectors $\{\Delta\boldsymbol{x}_t^a\}_{a=1}^n$, let us now construct a matrix $\boldsymbol{S}_t \in \mathbb{R}^{n \times n}$ for all $t = 1, \ldots, T$ whose (a,b)th element represents the empirical covariance

$$S_t^{[a,b]} = \frac{1}{M}\Delta\boldsymbol{x}_t^a \cdot \Delta\boldsymbol{x}_t^b$$

$$= \frac{\|\boldsymbol{x}_t^0\|^2}{M} - \frac{\boldsymbol{x}_t^0 \cdot \boldsymbol{x}_t^b}{M} - \frac{\boldsymbol{x}_t^a \cdot \boldsymbol{x}_t^0}{M} + \frac{\boldsymbol{x}_t^a \cdot \boldsymbol{x}_t^b}{M} \quad (98)$$

$$= Q_t^{[0,0]} - Q_t^{[0,b]} - Q_t^{[a,0]} + Q_t^{[a,b]}. \quad (99)$$

We also construct a similar set of matrices $\{\boldsymbol{Q}_t\}_{t=1}^T$ whose elements $\{Q_t^{[a,b]}\}$ have the obvious definitions. The rotational symmetry of distributions $\boldsymbol{w}_t$ and $\{\boldsymbol{Q}_t\}$ indicates that

$$\mathsf{E}_{\boldsymbol{w},\{\boldsymbol{O}_t\}} \exp\left(-\frac{\beta}{2\lambda} \sum_{a=1}^n \left\|\sigma\boldsymbol{w} - \sum_{t=1}^T \boldsymbol{O}_t\Delta\boldsymbol{x}_t^a\right\|^2\right) \quad (100)$$

becomes a function of $\{Q_t^{[a,b]}\}$ for any fixed set of $\{\Delta\boldsymbol{x}_t^a\}_{a=1}^n$. In addition, inserting a set of trivial identities

$$1 = M^{n(n+1)/2} \int \prod_{0 \leq a \leq b \leq n} \delta(MQ_t^{[a,b]} - \boldsymbol{x}_t^a \cdot \boldsymbol{x}_t^b)dQ_t^{[a,b]}$$

$$\quad (101)$$

for $t = 1, 2, \ldots, T$ into (97) and performing the integration over $\{\boldsymbol{x}^a\}$ yields an expression that allows for saddle point evaluation of $\Xi_{\beta,M}(n)$ with respect to $\{Q_t^{[a,b]}\}$.

To proceed with the analysis, we make the RS assumption which states that at the dominant saddle point, the overlap matrices $\{\boldsymbol{Q}_t\}$ are invariant under the rotation of the replica indexes $a = 1, 2, \ldots, n$ (see also (18))

$$Q_t^{[0,0]} = r_t, \quad (102)$$

$$Q_t^{[0,b]} = Q_t^{[a,0]} = m_t \quad \forall a, b \geq 1, \quad (103)$$

$$Q_t^{[a,a]} = Q_t \quad \forall a \geq 1, \quad (104)$$

$$Q_t^{[a,b]} = q_t \quad \forall a \neq b \geq 1. \quad (105)$$

Under the RS assumption, the matrices $\boldsymbol{S}_t$ introduced above have also a simple form

$$\boldsymbol{S}_t = S_t^{[1,2]}\mathbf{1}_n\mathbf{1}_n^{\mathsf{T}} + (S_t^{[1,1]} - S_t^{[1,2]})\boldsymbol{I}_n, \quad (106)$$

where $\mathbf{1}_n = [1 \cdots 1]^{\mathsf{T}} \in \mathbb{R}^n$ is an all-ones vector and

$$S_t^{[1,1]} = r_t - 2m_t + Q_t \quad (107)$$

$$S_t^{[1,2]} = r_t - 2m_t + q_t. \quad (108)$$

When $\boldsymbol{O}_t$ are independent Haar matrices, the vectors $\boldsymbol{O}_t\Delta\boldsymbol{x}_t^a$ are distributed on the M-dimensional hyper spheres

of radius $\|\boldsymbol{O}_t\Delta\boldsymbol{x}_t^a\| = \|\Delta\boldsymbol{x}_t^a\| = \sqrt{MQ_t}$ that are centered at the origin (see Appendix C). As $\boldsymbol{O}_t$ are sampled independently among $t = 1, 2, \ldots, T$, the cross-correlations for all $t_1 \neq t_2$ reduce to

$$\mathsf{E}_{\{\boldsymbol{O}_t\}}\{(\boldsymbol{O}_{t_1}\Delta\boldsymbol{x}_{t_1}^a)^{\mathsf{T}}(\boldsymbol{O}_{t_2}\Delta\boldsymbol{x}_{t_2}^b)\}$$

$$= (\Delta\boldsymbol{x}_{t_1}^a)^{\mathsf{T}}\mathsf{E}_{\{\boldsymbol{O}_t\}}\{\boldsymbol{O}_{t_1}^{\mathsf{T}}\boldsymbol{O}_{t_2}\}\Delta\boldsymbol{x}_{t_2}^b$$

$$= (\Delta\boldsymbol{x}_{t_1}^a)^{\mathsf{T}}\mathbf{0}_M\Delta\boldsymbol{x}_{t_2}^b = 0, \quad (109)$$

where $\mathbf{0}_M$ is the $M \times M$ zero matrix. On the other hand, given $t_1 = t_2 = t$ we obtain

$$M^{-1}\mathsf{E}_{\{\boldsymbol{O}_t\}}\{(\boldsymbol{O}_t\Delta\boldsymbol{x}_t^a)^{\mathsf{T}}(\boldsymbol{O}_t\Delta\boldsymbol{x}_t^b)\}$$

$$= M^{-1}(\Delta\boldsymbol{x}_t^a)^{\mathsf{T}}\mathsf{E}_{\{\boldsymbol{O}_t\}}\{\boldsymbol{O}_t^{\mathsf{T}}\boldsymbol{O}_t\}\Delta\boldsymbol{x}_t^b$$

$$= M^{-1}(\Delta\boldsymbol{x}_t^a)^{\mathsf{T}}\boldsymbol{I}_M\Delta\boldsymbol{x}_t^b = S_t^{[a,b]}. \quad (110)$$

Since $S_t^{[a,b]}$ are in general non-zero for any pairs of replica indexes $a, b = 1, 2, \ldots, n$, the expectation in (97) is nontrivial to compute. However, these correlations can be decoupled by linearly transforming the variables using a matrix

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad (111)$$

that satisfies $\boldsymbol{E}^{\mathsf{T}}\boldsymbol{E} = \boldsymbol{E}\boldsymbol{E}^{\mathsf{T}} = \boldsymbol{I}_n$, and $\boldsymbol{e}_1 = \mathbf{1}_n/\sqrt{n}$ by definition. More precisely, if we let $\begin{bmatrix} \Delta\tilde{\boldsymbol{x}}_t^1 & \cdots & \Delta\tilde{\boldsymbol{x}}_t^n \end{bmatrix} = \begin{bmatrix} \Delta\boldsymbol{x}_t^1 & \cdots & \Delta\boldsymbol{x}_t^n \end{bmatrix}\boldsymbol{E}$ be the transformed vectors,

$$\frac{1}{M}\mathsf{E}_{\{\boldsymbol{O}_t\}}\left\{\left(\begin{bmatrix} \boldsymbol{O}_t\Delta\tilde{\boldsymbol{x}}_t^1 & \cdots & \boldsymbol{O}_t\Delta\tilde{\boldsymbol{x}}_t^n \end{bmatrix}\right)^{\mathsf{T}}\right.$$

$$\left. \times \begin{bmatrix} \boldsymbol{O}_t\Delta\tilde{\boldsymbol{x}}_t^1 & \cdots & \boldsymbol{O}_t\Delta\tilde{\boldsymbol{x}}_t^n \end{bmatrix}\right\}$$

$$= \frac{1}{M}\mathsf{E}_{\{\boldsymbol{O}_t\}}\left\{\left(\boldsymbol{O}_t\begin{bmatrix} \Delta\boldsymbol{x}_t^1 & \cdots & \Delta\boldsymbol{x}_t^n \end{bmatrix}\boldsymbol{E}\right)^{\mathsf{T}}\right.$$

$$\left. \times \left(\boldsymbol{O}_t\begin{bmatrix} \Delta\boldsymbol{x}_t^1 & \cdots & \Delta\boldsymbol{x}_t^n \end{bmatrix}\boldsymbol{E}\right)\right\}$$

$$= \frac{1}{M}\left(\begin{bmatrix} \Delta\boldsymbol{x}_t^1 & \cdots & \Delta\boldsymbol{x}_t^n \end{bmatrix}\boldsymbol{E}\right)^{\mathsf{T}}$$

$$\times \mathsf{E}_{\{\boldsymbol{O}_t\}}\{\boldsymbol{O}_t^{\mathsf{T}}\boldsymbol{O}_t\}\begin{bmatrix} \Delta\boldsymbol{x}_t^1 & \cdots & \Delta\boldsymbol{x}_t^n \end{bmatrix}\boldsymbol{E}$$

$$= \frac{1}{M}\left(\begin{bmatrix} \Delta\boldsymbol{x}_t^1 & \cdots & \Delta\boldsymbol{x}_t^n \end{bmatrix}\boldsymbol{E}\right)^{\mathsf{T}}\begin{bmatrix} \Delta\boldsymbol{x}_t^1 & \cdots & \Delta\boldsymbol{x}_t^n \end{bmatrix}\boldsymbol{E}$$

$$\triangleq \tilde{\boldsymbol{S}}_t. \quad (112)$$

But $\tilde{\boldsymbol{S}}_t \in \mathbb{R}^{n \times n}$ is just

$$\tilde{\boldsymbol{S}}_t = \boldsymbol{E}^{\mathsf{T}}\boldsymbol{S}_t\boldsymbol{E}$$

$$= \mathrm{diag}(nS_t^{[1,2]} + S_t^{[1,1]} - S_t^{[1,2]},$$

$$\underbrace{S_t^{[1,1]} - S_t^{[1,2]}, \ldots, S_t^{[1,1]} - S_t^{[1,2]}}_{n-1 \text{ times}}), \quad (113)$$

since $\boldsymbol{e}_a^{\mathsf{T}}\mathbf{1}_n = \boldsymbol{e}_a^{\mathsf{T}}\boldsymbol{e}_1 = 0$ for all $a = 2, \ldots, n$ and, thus,

$$\frac{1}{M}\mathsf{E}_{\{\boldsymbol{O}_t\}}\{(\boldsymbol{O}_t\Delta\tilde{\boldsymbol{x}}_t^a)^{\mathsf{T}}(\boldsymbol{O}_t\Delta\tilde{\boldsymbol{x}}_t^b)\}$$

$$= \begin{cases} 0 & \text{if } a \neq b, \\ n(r_t - 2m_t + q_t) + Q_t - q_t & \text{if } a = b = 1, \\ Q_t - q_t & \text{if } a = b = 2, \ldots, n, \end{cases}$$

$$\quad (114)$$

holds. The above shows that for given $\{\Delta\boldsymbol{x}_t^a\}_{a=1}^n$, the set of vectors $\{\boldsymbol{O}_t\Delta\tilde{\boldsymbol{x}}_t^a\}$ are independent among $t = 1, \ldots, T$ and also uncorrelated in the space of replicas, as indicated

by (114). This is in contrast to the original set $\{O_t \Delta x_t^a\}$ whose replica space correlation structure (110) is much more cumbersome to deal with.

To proceed with the analysis, we first notice that since $EE^\mathsf{T} = I_n$, the quadratic term in (97) can be expressed as

$$\sum_{a=1}^{n} \left\| \sum_{t=1}^{T} O_t \Delta x_t^a \right\|^2 = \sum_{a=1}^{n} \left\| \sum_{t=1}^{T} O_t \Delta \tilde{x}_t^a \right\|^2 \qquad (115)$$

using the uncorrelated random vectors $\{O_t \Delta \tilde{x}_t^a\}$. For notational convenience, we define next an auxiliary matrix $E' = \begin{bmatrix} e'_1 & \cdots & e'_n \end{bmatrix} = E^\mathsf{T}$ so that

$$\Delta x_t^a = \begin{bmatrix} \Delta \tilde{x}_t^1 & \cdots & \Delta \tilde{x}_t^n \end{bmatrix} e'_a, \qquad a = 1, \ldots, n, \quad (116)$$

where $\{e'_a\}$ again forms an orthonormal set that is independent of t. Then, after the transformation (116), the linear term in (97) becomes

$$\sum_{t=1}^{T} \sum_{a=1}^{n} w^\mathsf{T} O_t \Delta x_t^a = \sum_{t=1}^{T} \sum_{a=1}^{n} w^\mathsf{T} O_t \begin{bmatrix} \Delta \tilde{x}_t^1 & \cdots & \Delta \tilde{x}_t^n \end{bmatrix} e'_a$$

$$= \sqrt{n} w^\mathsf{T} \sum_{t=1}^{T} O_t \Delta \tilde{x}_t^1, \qquad (117)$$

where the we used the fact that

$$\sum_{a=1}^{n} e'_a = E^\mathsf{T} \mathbf{1}_n = \begin{bmatrix} \sqrt{n} & 0 & \cdots & 0 \end{bmatrix}^\mathsf{T}. \qquad (118)$$

Combining the above findings and re-arranging implies that (97) can be equivalently expressed as

$$\Xi_{\beta,M}(n)$$

$$= \mathsf{E}_{w,\{O_t\}} \int \exp\left(-\frac{\beta}{2\lambda} \left\| \sqrt{n\sigma^2} w - \sum_{t=1}^{T} O_t \Delta \tilde{x}_t^1 \right\|^2 \right)$$

$$\times \exp\left(-\frac{\beta}{2\lambda} \sum_{a=2}^{n} \left\| \sum_{t=1}^{T} O_t \Delta \tilde{x}_t^a \right\|^2 \right)$$

$$\times p(x^0; \{\rho_t\}) \exp\left(-\beta \sum_{a=1}^{n} c(x^a) \right) \prod_{a=0}^{n} dx^a. \qquad (119)$$

Next, recall the definition of the matrix Q_t whose elements are as given in (99). From the identity of (101) we obtain the probability weight for $Q_t, t = 1, \ldots, T$ as

$$p_{\beta,M}(Q_t; n) = \frac{1}{z_{\beta,M}^n} \int p(x_t^0) dx_t^0 \prod_{a=1}^{n} \left(e^{-\beta c(x_t^a)} dx_t^a \right)$$

$$\times \prod_{0 \le a \le b \le n} \delta(x_t^a \cdot x_t^b - M Q_t^{[a,b]}), \qquad (120)$$

where $c(x_t^a)$ is interpreted as in (10) to be a sum of scalar regularization functions and the normalization constant is given by $z_{\beta,M} = z_\beta M^{-(n+1)/2}$, where z_β is as in (11).

Then we proceed as follows:

1) Fix the matrices $\{Q_t\}_{t=1}^{T}$ so that the lengths $\tilde{S}_t^{[a,a]}$, $a = 1, \ldots, n$ in (114) are constant and, thus, $\{\Delta \tilde{x}_t^a\}$ have fixed (squared) lengths. Then, assuming M grows without bound, average over the joint distribution of w, $\{O_t\}$ and $\{\Delta \tilde{x}_t^a\}$, given $\{Q_t\}_{t=1}^{T}$.

2) Average the obtained result w.r.t. $p_{\beta,M}(Q_t; n)$ as given in (120) when $\beta \to \infty$.

The first step may be achieved by separately averaging over the replicas $a = 1, \ldots, n$ using the following result. Note this is always possible since we consider the setting of large M and hence $n \ll M$.

Lemma 1. *Let $\{u_t\}_{t=1}^{T}$ be a set of length-M vectors that satisfy $\|u_t\|^2 = M\nu_t$ for some given non-negative reals $\{\nu_t\}$. Let $\{O_t\}$ a set of independent Haar matrices and define*

$$e^{MG_{\beta,\lambda}(\sigma^2, \{\nu_t\})} = \mathsf{E}_{w,\{O_t\}} e^{-\frac{\beta}{2\lambda} \|\sigma w - \sum_{t=1}^{T} O_t u_t\|^2}, \quad (121)$$

where w is a standard Gaussian random vector. Then, for large M

$$G_{\beta,\lambda}(\sigma^2, \{\nu_t\}) = -\frac{1}{2} \left(T - \ln \lambda + \sum_{t=1}^{T} \ln(\beta \nu_t) \right)$$

$$+ \frac{1}{2} \operatorname*{extr}_{\{\Lambda_t\}} \left\{ \sum_{t=1}^{T} \left[\Lambda_t(\beta \nu_t) - \ln \Lambda_t \right] - \ln \left(\lambda + \beta\sigma^2 + \sum_{t=1}^{T} \frac{1}{\Lambda_t} \right) \right\}, \qquad (122)$$

where we have omitted terms of the order $O(1/M)$.

    Proof: Proof is given in Appendix C-A. ∎

Since $\{O_t \Delta \tilde{x}_t^a\}_{a=1}^{n}$ are uncorrelated, we can apply the above result to (119) separately for all replica indexes $a = 1, \ldots, n$ in order to evaluate the expectations w.r.t. w and $\{O_t\}$. Thus, for $n \ll M$, we get (123) at the top of the next page, where n is now just a parameter in $G_{\beta,\lambda}$ and is not enforced to be an integer by the function itself. The next step is to compute the integral over $\{Q_t\}$. With some abuse of notation, we start by using the Dirac's delta identity (181) to write

$$p_{\beta,M}(Q_t; n) = \frac{1}{z_{\beta,M}^n} \int \left(\prod_{0 \le a \le b \le n} \frac{d\tilde{Q}_t^{[a,b]}}{2\pi i} \right)$$

$$\times \exp\left(M \sum_{0 \le a \le b \le n} \tilde{Q}_t^{[a,b]} Q_t^{[a,b]} \right) \mathcal{V}_{\beta,M}(\tilde{Q}_t; n), \quad (124)$$

where $\{\tilde{Q}_t\}_{t=1}^{T}$ is a set of transform domain matrices whose elements are $\{\tilde{Q}_t^{[a,b]}\}$ and

$$\mathcal{V}_{\beta,M}(\tilde{Q}_t; n) = \int p(x_t^0) dx_t^0 \prod_{a=1}^{n} \left(e^{-\beta c(x_t^a)} dx_t^a \right)$$

$$\times \exp\left(-\sum_{0 \le a \le b \le n} \tilde{Q}_t^{[a,b]} x_t^a \cdot x_t^b \right). \quad (125)$$

Note that n has to be an integer in (125) and the goal is thus to write it in a form where n can be regarded as a real valued non-negative variable. One can then verify that since $\tilde{Q}_t^{[0,0]}$ is connected only to the zeroth replica, $\tilde{Q}_t^{[0,0]} \to 0$ when $n \to 0$. Therefore, it plays no role in the evaluation of the asymptotic free energy and consequently, the MSE. This is indeed a common feature of replica symmetric solution and similar conclusion can be found, e.g., in [30], [43] and [44]. To simplify notation, we therefore omit $\tilde{Q}_t^{[0,0]}$ from further consideration.

$$\frac{1}{M}\ln\Xi_{\beta,M}(n)$$

$$= \frac{1}{M}\ln\int\prod_{t=1}^{T}\big[p_{\beta,M}(\boldsymbol{Q}_t;n)\mathrm{d}\boldsymbol{Q}_t\big]\mathsf{E}_{\boldsymbol{w},\{\boldsymbol{O}_t\}}\Big\{\mathrm{e}^{-\frac{\beta}{2\lambda}\|\sqrt{n\sigma^2}\boldsymbol{w}-\sum_{t=1}^{T}\boldsymbol{O}_t\Delta\tilde{\boldsymbol{x}}_t^1\|^2}\Big\}\Big[\mathsf{E}_{\boldsymbol{w},\{\boldsymbol{O}_t\}}\Big\{\mathrm{e}^{-\frac{\beta}{2\lambda}\|\sum_{t=1}^{T}\boldsymbol{O}_t\Delta\tilde{\boldsymbol{x}}_t^2\|^2}\Big\}\Big]^{n-1}$$

$$= \frac{1}{M}\ln\int\prod_{t=1}^{T}\big[p_{\beta,M}(\boldsymbol{Q}_t;n)\mathrm{d}\boldsymbol{Q}_t\big]\mathrm{e}^{MG_{\beta,\lambda}(n\sigma^2,\{n(r_t-2m_t+q_t)+Q_t-q_t\})}\mathrm{e}^{M(n-1)G_{\beta,\lambda}(0,\{Q_t-q_t\})} \tag{123}$$

With some foresight we now impose the RS assumption on $\{\tilde{\boldsymbol{Q}}_t\}$ via auxiliary parameters $\{\hat{Q}_t,\hat{m}_t,\hat{\chi}_t\}$ as

$$\tilde{Q}_t^{[a,0]}=\tilde{Q}_t^{[0,b]}=-\beta\hat{m}_t, \qquad \forall a,b\geq 1, \tag{126}$$

$$\tilde{Q}_t^{[a,a]}=\frac{\beta\hat{Q}_t-\beta^2\hat{\chi}_t}{2}, \qquad \forall a\geq 1, \tag{127}$$

$$\tilde{Q}_t^{[a,b]}=-\beta^2\hat{\chi}_t, \qquad \forall a\neq b\geq 1. \tag{128}$$

Recalling that the elements of $\boldsymbol{x}_t^0$ are IID according to (6) with $\pi(x)=g_x(0;1)$ simplifies the function $\mathcal{V}_{\beta,M}$ under the RS assumption to $\mathcal{V}_{\beta,M}(\tilde{\boldsymbol{Q}}_t;n)=[\mathcal{V}_\beta(\hat{\boldsymbol{Q}}_t;n)]^M$ where

$$\mathcal{V}_\beta(\hat{\boldsymbol{Q}}_t;n)=\int\Big(\prod_{a=1}^n\mathrm{d}x_t^a\Big)\exp\Big(-\beta\sum_{a=1}^n c(x_t^a)\Big)$$

$$\times\Big\{(1-\rho_t)\exp\Big[-\frac{\beta\hat{Q}_t}{2}\sum_{a=1}^n(x_t^a)^2+\frac{1}{2}\Big(\beta\sqrt{\hat{\chi}_t}\sum_{a=1}^n x_t^a\Big)^2\Big]$$

$$+\rho_t\exp\Big[-\frac{\beta\hat{Q}_t}{2}\sum_{a=1}^n(x_t^a)^2+\frac{1}{2}\Big(\beta\sqrt{\hat{\chi}_t+\hat{m}_t^2}\sum_{a=1}^n x_t^a\Big)^2\Big]\Big\}, \tag{129}$$

and $\hat{\boldsymbol{Q}}_t$ should be read as a shorthand for the set $\{\hat{\chi}_t,\hat{Q}_t,\hat{m}_t\}$. To assess the integrals in (129) w.r.t. the replicated variables we first decouple the quadratic terms that have summations inside by using (182). By the fact that all integrals for $a=1,2,\ldots,n$ are identical we obtain

$$\mathcal{V}_\beta(\hat{\boldsymbol{Q}}_t;n)$$

$$= (1-\rho_t)\int\Big\{\int\mathrm{e}^{-\beta[\hat{Q}_t x_t^2/2-z_t\sqrt{\hat{\chi}_t}x_t+c(x_t)]}\mathrm{d}x_t\Big\}^n\mathrm{D}z_t$$

$$+\rho_t\int\Big\{\int\mathrm{e}^{-\beta[\hat{Q}_t x_t^2/2-z_t\sqrt{\hat{\chi}_t+\hat{m}_t^2}x_t+c(x_t)]}\mathrm{d}x_t\Big\}^n\mathrm{D}z_t, \tag{130}$$

where $\mathrm{D}z_t=\mathrm{d}z_t\mathrm{e}^{-z_t^2/2}/\sqrt{2\pi}$ is the standard Gaussian measure. For large β we may then employ the saddle-point integration w.r.t. x_t. If we now specialize to LASSO reconstruction (4) so that the per-element regularization function is $c(x)=|x|$, we may define

$$\phi(y;\hat{Q})=\min_{x\in\mathbb{R}}\Big\{\frac{\hat{Q}}{2}x^2-yx+|x|\Big\}$$

$$= \begin{cases} -\dfrac{(|y|-1)^2}{2\hat{Q}}, & |y|>1, \\ 0 & \text{otherwise,} \end{cases} \tag{131}$$

where the second equality follows by the fact that the x that minimizes the cost in (131) is given by[9]

$$\hat{x}(y;\hat{Q})=\begin{cases} \dfrac{y-1}{\hat{Q}}, & \text{if } y>1; \\ 0, & \text{if } |y|\leq 1; \\ \dfrac{y+1}{\hat{Q}}, & \text{if } y<-1. \end{cases} \tag{132}$$

The saddle-point method then provides the following expression

$$\mathcal{V}_\beta(\hat{\boldsymbol{Q}}_t;n)=(1-\rho_t)\int\exp\Big[-\beta n\phi\big(z_t\sqrt{\hat{\chi}_t};\hat{Q}_t\big)\Big]\mathrm{D}z_t$$

$$+\rho_t\int\exp\Big[-\beta n\phi\big(z_t\sqrt{\hat{\chi}_t+\hat{m}_t^2};\hat{Q}_t\big)\Big]\mathrm{D}z_t. \tag{133}$$

Note that the structure of the equations does not force n to be an integer anymore, so we assume that analytical continuation can be used to take the limit $n\to 0$. This provides $\mathcal{V}_\beta(\hat{\boldsymbol{Q}}_t;n)\to 1$ for the data dependent part of the probability weight (124), which is consistent with (125).

Returning to (124) and denoting $\chi_t=\beta(Q_t-q_t)$, we have under RS ansatz

$$p_{\beta,M}(\boldsymbol{Q}_t;n)$$

$$= \frac{1}{z_{\beta,M}^n}\int\exp\Big[\beta M\Big(n\frac{\hat{Q}_tQ_t-\hat{\chi}_t\chi_t}{2}-n\hat{m}_tm_t-n^2\beta\frac{\hat{\chi}_tq_t}{2}$$

$$+\frac{1}{\beta}\log\mathcal{V}_\beta(\hat{\boldsymbol{Q}}_t;n)\Big)\Big]\mathrm{d}\hat{\boldsymbol{Q}}_t, \tag{134}$$

where $\mathrm{d}\hat{\boldsymbol{Q}}$ is a short-hand for $\mathrm{d}\hat{\chi}_t\mathrm{d}\hat{Q}_t\mathrm{d}\hat{m}_t$. It is important to recognize that we have now managed to write the components of the free energy in a functional form of n where the limit $n\to 0$ can be taken, at least in principle. Applying the saddle-point method to integrate w.r.t. $\hat{\boldsymbol{Q}}$ and $\boldsymbol{Q}$ as $\beta,M\to\infty$ and changing the order of extremization and partial derivation, we get (135) at the top of the next page. Here we used the fact that $\beta^{-1}G_{\beta,\lambda}(0,\{Q_t-q_t\})\to 0$ as $\beta\to\infty$ for $\chi,\Lambda\in\mathbb{R}$ and

$$-\frac{1}{\beta}\lim_{n\to 0}\frac{\partial}{\partial n}\log\mathcal{V}_\beta(\hat{\boldsymbol{Q}}_t;n)$$

$$= (1-\rho_t)\int\mathrm{D}z_t\phi\big(z_t\sqrt{\hat{\chi}_t};\hat{Q}_t\big)$$

$$+\rho_t\int\mathrm{D}z_t\phi\big(z_t\sqrt{\hat{\chi}_t+\hat{m}_t^2};\hat{Q}_t\big). \tag{136}$$

[9] Note that $\hat{x}(y;\hat{Q})$ can be interpreted as soft thresholding of observation y. Compare the above also to (60). For further discussion on the relevance and interpretation of this function, see Section IV.

$$f = \frac{1}{T} \operatorname*{extr}_{\mathbf{Q},\hat{\mathbf{Q}}} \left\{ \sum_{t=1}^{T} \left(\hat{m}_t m_t - \frac{\hat{Q}_t Q_t}{2} + \frac{\hat{\chi}_t \chi_t}{2} + (1 - \rho_t) \int \mathrm{D}z_t \phi\big(z_t \sqrt{\hat{\chi}_t}; \hat{Q}_t\big) + \rho_t \int \mathrm{D}z_t \phi\big(z_t \sqrt{\hat{\chi}_t + \hat{m}_t^2}; \hat{Q}_t\big) \right) \right.$$
$$\left. - \lim_{\beta \to \infty} \lim_{n \to 0} \frac{\partial}{\partial n} \frac{1}{\beta} G_{\beta,\lambda}(n\sigma^2, \{n(\rho_t - 2m_t + q_t) + Q_t - q_t\}) \right\} \tag{135}$$

We also used above the fact that $r_t \to \rho_t$ for large M and that the term $\frac{1}{z_{\beta,M}^n}$ in (120) is irrelevant for the analysis because

$$\frac{1}{M} \lim_{n \to 0} \frac{\partial}{\partial n} \ln z_{\beta,M}^n \xrightarrow{M \to \infty} 0. \tag{137}$$

By the chain rule

$$\frac{1}{\beta} \lim_{n \to 0} \frac{\partial}{\partial n} G_{\beta,\lambda}(n\sigma^2, \{\nu_t(n)\}_{t=1}^T)$$
$$= \frac{\sigma^2}{\beta} \lim_{n \to 0} \frac{\partial G_{\beta,\lambda}(n\sigma^2, \{\nu_t(n)\}_{t=1}^T)}{\partial(n\sigma^2)}$$
$$+ \frac{1}{\beta} \lim_{n \to 0} \sum_{t=1}^{T} \left(\frac{\partial \nu_t(n)}{\partial n} \right) \frac{\partial G_{\beta,\lambda}(n\sigma^2, \{\nu_t(n)\}_{t=1}^T)}{\partial \nu_t(n)}, \tag{138}$$

so that by plugging $\nu_t(n) = n(\rho_t - 2m_t + q_t) + Q_t - q_t$ to (138) we have

$$\frac{1}{\beta} \lim_{n \to 0} \frac{\partial}{\partial n} G_{\beta,\lambda}(n\sigma^2, \{\nu_t(n)\}_{t=1}^T)$$
$$= -\frac{\sigma^2}{2} \left(\lambda + \sum_{t=1}^{T} \frac{1}{\Lambda_t^*} \right)^{-1} + \sum_{t=1}^{T} \frac{\rho_t - 2m_t + q_t}{2} \left(\Lambda_t^* - \frac{1}{\chi_t} \right), \tag{139}$$

where $\{\Lambda_t^*\}$ denotes the solution to the extremization problem in (122), given $\sigma^2 = 0$.

To solve the last integrals in (135), let us denote

$$r(h) = \sqrt{\frac{h}{2\pi}} e^{-\frac{1}{2h}} - (1 + h)\mathcal{Q}\left(\frac{1}{\sqrt{h}} \right). \tag{140}$$

With some calculus one may verify that for $h > 0$

$$\int \phi\big(z_t \sqrt{h}; \hat{Q}_t\big) \mathrm{D}z_t = \frac{r(h)}{\hat{Q}_t}, \tag{141}$$

which implies that combining all of the above, the free energy has the form (142) given at the top of the next page. To obtain this result, we used the fact that denoting

$$R_t(\lambda, \{\Lambda_t\}) = \frac{1}{\Lambda_t} \left(\lambda + \sum_{s=1}^{T} \frac{1}{\Lambda_s} \right)^{-1}, \tag{143}$$

the extremization in (122) implies the condition

$$\Lambda_t^* - \frac{1}{\chi_t} = -\frac{R_t(\lambda, \{\Lambda_t^*\})}{\chi_t}, \tag{144}$$

between the variables χ_t and $\{\Lambda_t\}$. Furthermore, in order to have a meaningful solution to (142) and (144) for $\sigma > 0$, we also need to have $\chi_t = \beta(Q_t - q_t)$ positive and finite[10] for

all values of $\beta > 0$, which means $\chi_t^{-1}(\rho_t - 2m_t + q_t) \to \chi_t^{-1}(\rho_t - 2m_t + Q_t)$ as $\beta \to \infty$.

B. Saddle-Point Conditions

Using the short-hand notation

$$R_t = R_t(\lambda, \{\Lambda_t^*\}), \tag{145}$$

the partial derivatives w.r.t. $\{m_t, Q_t\}$ in (142) provide the saddle-point conditions

$$\hat{Q}_t = \hat{m}_t = \frac{R_t}{\chi_t} = \frac{1}{\chi_t} - \Lambda_t^*. \tag{146}$$

By the fact that

$$\frac{\partial}{\partial x} r(h) = -\left(\frac{\partial h}{\partial x} \right) \mathcal{Q}\left(\frac{1}{\sqrt{h}} \right), \tag{147}$$

we may assess the partial derivatives w.r.t. the variables $\{\hat{Q}_t, \hat{m}_t, \hat{\chi}_t\}$ as well to obtain

$$m_t = 2\rho_t \mathcal{Q}\left(\frac{1}{\sqrt{\hat{\chi}_t + \hat{m}_t^2}} \right), \tag{148}$$

$$Q_t = -\frac{2(1 - \rho_t)}{\hat{m}_t^2} r(\hat{\chi}_t) - \frac{2\rho_t}{\hat{m}_t^2} r(\hat{\chi}_t + \hat{m}_t^2), \tag{149}$$

$$\chi_t = \frac{2(1 - \rho_t)}{\hat{m}_t} \mathcal{Q}\left(\frac{1}{\sqrt{\hat{\chi}_t}} \right) + \frac{2\rho_t}{\hat{m}_t} \mathcal{Q}\left(\frac{1}{\sqrt{\hat{\chi}_t + \hat{m}_t^2}} \right), \tag{150}$$

where we used the identity $\hat{Q}_t = \hat{m}_t$ to simplify the results. The MSE of the reconstruction for $\boldsymbol{x}_t$ thus becomes

$$\mathsf{mse}_t = \rho_t - 2m_t + Q_t$$
$$= \rho_t - 4\rho_t \mathcal{Q}\left(\frac{1}{\sqrt{\hat{\chi}_t + \hat{m}_t^2}} \right)$$
$$- \frac{2(1 - \rho_t)}{\hat{m}_t^2} r(\hat{\chi}_t) - \frac{2\rho_t}{\hat{m}_t^2} r(\hat{\chi}_t + \hat{m}_t^2). \tag{151}$$

Finally, recalling that Λ_t^* is a function of $\{\chi_t\}$, we obtain from the partial derivative of χ_t

$$\hat{\chi}_t = \frac{\mathsf{mse}_t}{\chi_t^2} + \sum_{s=1}^{T} (\mathsf{mse}_s - \sigma^2 R_s^2) \Delta_{s,t}, \tag{152}$$

where we denoted $\Delta_{s,t} = \frac{\partial \Lambda_s}{\partial \chi_t}$ for the partial derivative of Λ_s w.r.t. χ_t.

To solve the equation for $\hat{\chi}_t$, we need an expression for $\Delta_{s,t}$. We do this via the inverse function theorem that relates the Jacobian matrices as

$$\boldsymbol{\Delta} = \frac{\partial(\Lambda_1, \ldots, \Lambda_T)}{\partial(\chi_1, \ldots, \chi_T)} = \left[\frac{\partial(\chi_1, \ldots, \chi_T)}{\partial(\Lambda_1, \ldots, \Lambda_T)} \right]^{-1}. \tag{153}$$

---

[10]The case of $\chi_t \to 0$ is in fact relevant for the noise-free scenario $\sigma = 0$ and corresponds to the perfect recovery condition $\rho_t = m_t = Q_t \implies \mathsf{mse}_t = \rho_t - 2m_t + Q_t = 0$, which automatically satisfies $Q_t = q_t$ as well. Furthermore, for this scenario $\hat{Q} = \hat{m} \to \infty$ as $\beta \to \infty$, while in the noisy case they are always positive and finite parameters.

$$f = \frac{1}{T} \underset{\{m_t, Q_t, \chi_t, \hat{m}_t, \hat{Q}_t, \hat{\chi}_t\}}{\text{extr}} \left\{ \sum_{t=1}^{T} \left[\hat{m}_t m_t - \frac{\hat{Q}_t Q_t}{2} + \frac{\hat{\chi}_t \chi_t}{2} + \frac{1 - \rho_t}{\hat{Q}_t} r(\hat{\chi}_t) + \frac{\rho_t}{\hat{Q}_t} r(\hat{\chi}_t + \hat{m}_t^2) \right] \right.$$
$$\left. + \frac{\sigma^2}{2} \left(\lambda + \sum_{t=1}^{T} \frac{1}{\Lambda_t^*} \right)^{-1} + \sum_{t=1}^{T} \frac{\rho_t - 2m_t + Q_t}{2} \left(\frac{1}{\chi_t} - \Lambda_t^* \right) \right\} \qquad (142)$$

Here the (i,j)th element of the Jacobian $\frac{\partial(\chi_1, \ldots, \chi_T)}{\partial(\Lambda_1, \ldots, \Lambda_T)}$ is given by

$$\frac{\partial \chi_i}{\partial \Lambda_j} = \frac{\partial}{\partial \Lambda_j} \frac{(1 - R_i)}{\Lambda_i}$$
$$= -\frac{(1 - R_i)}{\Lambda_i^2} \delta_{ij} - \frac{1}{\Lambda_i} \frac{\partial R_i}{\partial \Lambda_j}$$
$$= -\frac{(1 - 2R_i)}{\Lambda_i^2} \delta_{ij} - \frac{R_i R_j}{\Lambda_i \Lambda_j}. \qquad (154)$$

In other words, denoting $\boldsymbol{b} = [R_1/\Lambda_1 \;\; \cdots \;\; R_T/\Lambda_T]^{\mathsf{T}}$ and defining $\boldsymbol{C}$ to be diagonal matrix whose (t,t)th entry is given by $(1 - 2R_t)\Lambda_t^{-2}$, we obtain by (153) and the matrix inversion lemma the desired Jacobian as

$$\boldsymbol{\Delta} = -(\boldsymbol{C} + \boldsymbol{b}\boldsymbol{b}^{\mathsf{T}})^{-1} = -\boldsymbol{C}^{-1} + \frac{(\boldsymbol{C}^{-1}\boldsymbol{b})(\boldsymbol{C}^{-1}\boldsymbol{b})^{\mathsf{T}}}{1 + \boldsymbol{b}^{\mathsf{T}}\boldsymbol{C}^{-1}\boldsymbol{b}}, \quad (155)$$

which means that

$$\Delta_{s,t} = \frac{R_s R_t \Lambda_s \Lambda_t}{(1 - 2R_s)(1 - 2R_t)} \left(1 + \sum_{k=1}^{T} \frac{R_k^2}{1 - 2R_k} \right)^{-1}$$
$$- \frac{\Lambda_t^2}{1 - 2R_t} \delta_{st}. \qquad (156)$$

Combining all the results completes the derivation.

APPENDIX B
REPLICA ANALYSIS OF ROTATIONALLY INVARIANT SETUP

The derivation in this Appendix provides an end result that is essentially the same as the HCIZ-formula based [45], [46] approach used in Section IV and Appendix B in [30]. In our case the difference is that the source does not need to have IID elements, but can have a block structure. Furthermore, our analytical approach is slightly different to the one in [30] since we do not seek to find first a decoupling result for finite β and then use hardening arguments as in [28] to obtain the final result when $\beta \to \infty$. Both end results are equivalent as shown in Section IV-B.

Recall the rotationally invariant setup as given in Definition 2. Let $p(\boldsymbol{x}^0; \{\rho_t\})$ be the distribution of the source vector $\boldsymbol{x}^0 \in \mathbb{R}^N$ and assume that each of the sub-vectors $\boldsymbol{x}_t^0$ has $\hat{M}$ elements drawn independently according to (6). Clearly we have to have $N = \hat{M}T$ but it is not necessary to have $M = \hat{M}$ as in the case of T-orthogonal setup. Define

$$\Xi_{\beta,N}(n) = \mathsf{E}_{\boldsymbol{w}, \boldsymbol{A}} \int p(\boldsymbol{x}^0; \{\rho_t\}) \exp \left(-\beta \sum_{a=1}^{n} c(\boldsymbol{x}^a) \right)$$
$$\times \exp \left(-\frac{\beta}{2\lambda} \sum_{a=1}^{n} \|\sigma \boldsymbol{w} - \boldsymbol{A} \Delta \boldsymbol{x}^a\|^2 \right) \prod_{a=0}^{n} \mathrm{d}\boldsymbol{x}^a, \quad (157)$$

where $\Delta \boldsymbol{x}^a = \boldsymbol{x}^0 - \boldsymbol{x}^a \in \mathbb{R}^N$ for $a = 1, \ldots, n$, so that the counterpart of (96) reads

$$f = -\lim_{n \to 0^+} \frac{\partial}{\partial n} \lim_{\beta, N \to \infty} \frac{1}{\beta N} \ln \Xi_{\beta,N}(n). \qquad (158)$$

The goal is then to assess the normalized free energy (158) by following the same steps as given in Appendix A.

Let us construct matrices $\boldsymbol{S}_t \in \mathbb{R}^{n \times n}$ and $\boldsymbol{Q}_t$ for all $t = 1, \ldots, T$ with elements as given in (98) and (99). Also define the "empirical mean" matrices $\boldsymbol{S} = T^{-1} \sum_t \boldsymbol{S}_t$, $\boldsymbol{Q} = T^{-1} \sum_t \boldsymbol{Q}_t$, that have the respective elements $S^{[a,b]}$ and $Q^{[a,b]}$ and invoke the RS assumption (102)–(105). We then make the transformation $\{\Delta \boldsymbol{x}_t^a\} \to \{\Delta \tilde{\boldsymbol{x}}_t^a\}$ as with the T-orthogonal setup so that the empirical correlations of $\{\Delta \tilde{\boldsymbol{x}}_t^a\}$ satisfy (114). Note that this means that given $\{\boldsymbol{Q}_t\}$, the transformed vectors $\Delta \tilde{\boldsymbol{x}}^a = \left[(\Delta \tilde{\boldsymbol{x}}_1^a)^{\mathsf{T}} \;\; \cdots \;\; (\Delta \tilde{\boldsymbol{x}}_T^a)^{\mathsf{T}} \right]^{\mathsf{T}}$ satisfy

$$\|\Delta \tilde{\boldsymbol{x}}^a\|^2 = \hat{M} \sum_{t=1}^{T} \tilde{S}_t^{[a,b]} = N \tilde{S}^{[a,b]}, \qquad (159)$$

where $\tilde{S}^{[a,b]} = T^{-1} \sum_t \tilde{S}_t^{[a,b]}$. Combining the above provides the counterpart of (119) as

$$\Xi_{\beta,N}(n)$$
$$= \mathsf{E}_{\boldsymbol{w}, \{\boldsymbol{O}_t\}} \int \left(\prod_{a=0}^{n} \mathrm{d}\boldsymbol{x}^a \right) p(\boldsymbol{x}^0; \{\rho_t\}) \exp \left(-\beta \sum_{a=1}^{n} c(\boldsymbol{x}^a) \right)$$
$$\times \exp \left(-\frac{\beta}{2\lambda} \|\sqrt{n\sigma^2} \boldsymbol{w} - \boldsymbol{A} \Delta \tilde{\boldsymbol{x}}^1\|^2 - \frac{\beta}{2\lambda} \sum_{a=2}^{n} \|\boldsymbol{A} \Delta \tilde{\boldsymbol{x}}^a\|^2 \right). \qquad (160)$$

We then need the following small result to proceed.

Lemma 2. *Consider the case where $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is sampled from the rotationally invariant setup given in Definition 2. Let $\{\boldsymbol{u}_t\}_{t=1}^{T}$ be a fixed set of length-$\hat{M}$ vectors satisfying $\|\boldsymbol{u}_t\|^2 = \hat{M}\nu_t$ for some given non-negative values $\{\nu_t\}$ and $N = T\hat{M}$. Denote $\boldsymbol{u} \in \mathbb{R}^N$ for the vector obtained by stacking $\{\boldsymbol{u}_t\}$ and define*

$$\mathrm{e}^{N H_{\beta,\lambda}(\sigma^2, \{\nu_t\})} = \mathsf{E}_{\boldsymbol{w}, \boldsymbol{A}} \mathrm{e}^{-\frac{\beta}{2\lambda} \|\sigma \boldsymbol{w} - \boldsymbol{A} \boldsymbol{u}\|^2}, \qquad (161)$$

where $\boldsymbol{w}$ is a standard Gaussian random vector. Then, for large N

$$H_{\beta,\lambda}(\sigma^2, \nu) = H_{\beta,\lambda}(\sigma^2, \{\nu_t\})$$
$$= \frac{1}{2} \underset{\Lambda}{\text{extr}} \left\{ \Lambda(\beta\nu) - (1 - \alpha) \ln \Lambda \right.$$
$$\left. - \alpha \int \ln(\Lambda \beta \sigma^2 + \Lambda \lambda + x) \mathrm{d}F_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}(x) \right\}$$
$$- \frac{1 + \ln(\beta\nu) - \alpha \ln \lambda}{2}, \qquad (162)$$

where $\nu = T^{-1}\sum_{t=1}^T \nu_t$ and we omitted terms of the order $O(1/N)$.

Proof: Proof is given in Appendix C-B. ∎

Notice that the H-function in (162) depends on the parameters $\{\nu_t\}$ only through the "empirical mean" $\nu = T^{-1}\sum_{t=1}^T \nu_t$. This will translate later to the fact that the performance of rotationally invariant setup depends on the sparsities $\{\rho_t\}$ only through $\rho = T^{-1}\sum_t \rho_t$. With the above in mind, we may obtain the probability weight $p_{\beta,N}(\boldsymbol{Q}; n)$ of $\boldsymbol{Q}$ by using (101) with suitable variable substitutions. Applying then Lemma 2 to (160) provides

$$
\begin{aligned}
&\frac{1}{N}\ln \Xi_{\beta,N}(n) \\
&= \frac{1}{N}\ln \int p_{\beta,N}(\boldsymbol{Q}; n)e^{NH_{\beta,\lambda}(n\sigma^2, n(r-2m+q)+Q-q)} \\
&\quad \times e^{N(n-1)H_{\beta,\lambda}(0, Q-q)}\mathrm{d}\boldsymbol{Q},
\end{aligned}
\tag{163}
$$

where $r = T^{-1}\sum_t r_t, m = T^{-1}\sum_t m_t, Q = T^{-1}\sum_t Q_t$, and $q = T^{-1}\sum_t q_t$ are the "averaged" versions of the RS variables $\{r_t, m_t, Q_t, q_t\}$. The probability weight of $\boldsymbol{Q}$ reads

$$
\begin{aligned}
p_{\beta,N}(\boldsymbol{Q}; n) = N^{n(n+1)/2}\int \left(\prod_{0\le a\le b\le n}\frac{\mathrm{d}\tilde{Q}^{[a,b]}}{2\pi\mathrm{i}}\right) \\
\times \exp\left(N\sum_{0\le a\le b\le n}\tilde{Q}^{[a,b]}Q^{[a,b]}\right)\mathcal{V}_{\beta,N}(\tilde{\boldsymbol{Q}}; n),
\end{aligned}
\tag{164}
$$

where $\tilde{\boldsymbol{Q}}$ is a $(n+1)\times(n+1)$ transform domain matrix whose elements are $\{\tilde{Q}^{[a,b]}\}$ and

$$
\begin{aligned}
\mathcal{V}_{\beta,N}(\tilde{\boldsymbol{Q}}; n) = \int p(\boldsymbol{x}^0)\mathrm{d}\boldsymbol{x}^0 \prod_{a=1}^n \left(e^{-\beta\|\boldsymbol{x}^a\|_1}\mathrm{d}\boldsymbol{x}^a\right) \\
\times \exp\left(-\sum_{0\le a\le b\le n}\tilde{Q}^{[a,b]}\boldsymbol{x}^a\cdot\boldsymbol{x}^b\right).
\end{aligned}
\tag{165}
$$

We then get directly using the arguments from Appendix A that (165) becomes in the limit $N\to\infty$

$$
\begin{aligned}
\mathcal{V}_{\beta}(\hat{\boldsymbol{Q}}; n) = (1-\rho)\int \exp\left[-\beta n\phi(z\sqrt{\hat{\chi}}; \hat{Q})\right]\mathrm{D}z \\
+\rho\int \exp\left[-\beta n\phi(z\sqrt{\hat{\chi}+\hat{m}^2}; \hat{Q})\right]\mathrm{D}z,
\end{aligned}
\tag{166}
$$

where $\rho = T^{-1}\sum_t \rho_t$ is the expected sparsity of the entire source vector $\boldsymbol{x}$. Therefore, the details of how the non-zero elements are distributed on different sub-blocks $\{\boldsymbol{x}_t\}$ is irrelevant for the rotationally invariant case.

Combining everything above and denoting $\chi = T^{-1}\sum_t \beta(Q_t - q_t)$ implies that the free energy for the rotationally invariant case reads

$$
\begin{aligned}
f = \operatorname*{extr}_{\{m,Q,\chi,\hat{m},\hat{Q},\hat{\chi}\}} \Bigg\{&\hat{m}m - \frac{\hat{Q}Q}{2} + \frac{\hat{\chi}\chi}{2} \\
&+\frac{1}{\hat{Q}}\left[(1-\rho)r(\hat{\chi}) - \rho r(\hat{\chi}+\hat{m}^2)\right] \\
&+\frac{\alpha\sigma^2\Lambda^*}{2}G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*) + \frac{\rho-2m+Q}{2}\left(\frac{1}{\chi}-\Lambda^*\right)\Bigg\},
\end{aligned}
\tag{167}
$$

where we used the Stieltjes transform of $F_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(x)$,

$$
G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(s) = \int \frac{1}{x-s}\mathrm{d}F_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(x),
\tag{168}
$$

along with the chain rule

$$
\begin{aligned}
&\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to0}\frac{\partial}{\partial n}H_{\beta,\lambda}(n\sigma^2, \nu(n)) \\
&= \lim_{\beta\to\infty}\frac{\sigma^2}{\beta}\lim_{n\to0}\frac{\partial H_{\beta,\lambda}(n\sigma^2, \nu(n))}{\partial(n\sigma^2)} \\
&\quad + \lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to0}\left(\frac{\partial\nu(n)}{\partial n}\right)\frac{\partial H_{\beta,\lambda}(n\sigma^2, \nu(n))}{\partial\nu(n)} \\
&= -\frac{\alpha\sigma^2}{2}\int\frac{\Lambda^*}{\lambda\Lambda^*+x}\mathrm{d}F_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(x) + \frac{\rho-2m+Q}{2}\left(\Lambda^*-\frac{1}{\chi}\right) \\
&= -\frac{\alpha\sigma^2\Lambda^*}{2}G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*) + \frac{\rho-2m+Q}{2}\left(\Lambda^*-\frac{1}{\chi}\right),
\end{aligned}
\tag{169}
$$

where $\nu(n) = n(r-2m+q)+Q-q$. Here Λ^* is the solution to the extremization in (162), given $\sigma^2 = 0$, and satisfies the condition

$$
\Lambda^* - \frac{1}{\chi} = -\frac{\alpha}{\chi}\left[1-(\lambda\Lambda^*)\cdot G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*)\right] = -\frac{\hat{R}(\Lambda^*)}{\chi},
\tag{170}
$$

where

$$
\hat{R}(\Lambda^*) = \alpha\left[1-(\lambda\Lambda^*)\cdot G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*)\right].
\tag{171}
$$

Finally, we need to resolve the saddle point conditions in (167). The partial derivatives w.r.t. $\{m, Q\}$ provide

$$
\hat{Q} = \hat{m} = \frac{1}{\chi} - \Lambda^* = \frac{\hat{R}(\Lambda^*)}{\chi},
\tag{172}
$$

while the partial derivatives w.r.t. $\{\hat{Q}, \hat{m}, \hat{\chi}\}$ are of the same format as in (148)–(150) but without indexes t. Finally, recalling that Λ^* depends on χ

$$
\frac{\partial}{\partial\chi}G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*) = -\lambda\left(\frac{\partial\Lambda}{\partial\chi}\right)G'_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*),
\tag{173}
$$

where $G'_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}$ denotes the derivative of $G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}$ w.r.t. the argument, gives

$$
\begin{aligned}
\hat{\chi} = \mathsf{mse}\left(\frac{1}{\chi^2} + \frac{\partial\Lambda^*}{\partial\chi}\right) \\
-\alpha\sigma^2\left[G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*) - (\lambda\Lambda^*)\cdot G'_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*)\right]\frac{\partial\Lambda^*}{\partial\chi},
\end{aligned}
\tag{174}
$$

in which

$$
\frac{\partial\Lambda^*}{\partial\chi} = -\left[\frac{1-\alpha}{(\Lambda^*)^2} + (\alpha\lambda^2)\cdot G'_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*)\right]^{-1}.
\tag{175}
$$

To obtain the last formula we used the fact that

$$
\begin{aligned}
\frac{\partial\chi}{\partial\Lambda^*} &= -\frac{1}{(\Lambda^*)^2}(1-\hat{R}) - \frac{1}{\Lambda^*}\frac{\partial\hat{R}}{\partial\Lambda^*} \\
&= -\left[\frac{1-\alpha}{(\Lambda^*)^2} + (\alpha\lambda^2)\cdot G'_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\lambda\Lambda^*)\right].
\end{aligned}
\tag{176}
$$

Remark 5. Consider the row-orthogonal setup where

$$
G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(s) = \frac{1}{\alpha^{-1}-s}.
\tag{177}
$$

For this case, the extremization in (162) can also be written in the form

$$\Lambda - \frac{1}{\beta\nu} = -\frac{1}{\beta\nu}\left(\frac{1}{\alpha^{-1} + \Lambda(\lambda + \beta\sigma^2)}\right)$$
$$\xrightarrow{\sigma=0} -\frac{1}{\beta\nu}\left(\frac{1}{\alpha^{-1} + \Lambda\lambda}\right). \qquad (178)$$

We may then plug (177) and (178) to (167) and compare the end result with (142)–(144). It is clear that the two free energies are exactly the same if we set $\alpha = 1/T$ and $\rho = \rho_t$ so that $\nu = \nu_t$ and $\Lambda = \Lambda_t$ for all $t = 1, \ldots, T$. Therefore, also the saddle point solutions of row-orthogonal and T-orthogonal setups match for this special case and the MSE is the same.

APPENDIX C
USEFUL MATRIX INTEGRALS

A. T-Orthogonal Setup

Let $\{O_t\}_{t=1}^T$ be a set of independent $M \times M$ Haar matrices and $\{\Delta x_t\}_{t=1}^T$ a set of (fixed) length-M vectors that satisfy $\|\Delta x_t\|^2 = M\nu_t$ for some given non-negative values $\{\nu_t\}$. Given $\{\Delta x_t\}$ and $\{\nu_t\}$, the vector $u_t = O_t\Delta x_t$ is uniformly distributed on a surface of a sphere that has a fixed radius $\sqrt{M\nu_t}$ for each $t = 1, \ldots, T$. Thus, the joint PDF of $\{u_t\}$ reads

$$p^{(M)}(\{u_t\}; \{\nu_t\})$$
$$= \frac{1}{Z(\{\nu_t\})}\prod_{t=1}^T \delta(\|u_t\|^2 - M\nu_t) \qquad (179)$$
$$= \frac{(4\pi i)^{-T}}{Z(\{\nu_t\})}\int \prod_{t=1}^T \left(e^{-\frac{\Lambda_t}{2}(\|u_t\|^2 - M\nu_t)}d\Lambda_t\right), \qquad (180)$$

where $Z(\{\nu_t\})$ is the normalization factor, $\{\Lambda_t\}$ is a set of complex numbers and we used the identity

$$\delta(t - a) = \frac{1}{4\pi i}\int_{c-i\infty}^{c+i\infty} e^{-\frac{\Lambda}{2}(t-a)}d\Lambda, \qquad (181)$$

where $a, c, t \in \mathbb{R}, \Lambda \in \mathbb{C}$. Using the Gaussian integration formula

$$\frac{1}{(2\pi)^{N/2}}\int e^{-\frac{1}{2}z^\mathsf{T} M z + b^\mathsf{T} z}dz = \frac{1}{\sqrt{\det(M)}}e^{\frac{1}{2}b^\mathsf{T} M^{-1} b}, \qquad (182)$$

where $b, z \in \mathbb{R}^N$ and M is symmetric positive definite, the normalization factor becomes

$$Z(\{\nu_t\}) = \frac{1}{(4\pi i)^T}\int \prod_{t=1}^T \left(e^{\frac{\Lambda_t}{2}M\nu_t}e^{-\frac{1}{2}\Lambda_t\|u_t\|^2}du_t d\Lambda_t\right)$$
$$= \left(\frac{(2\pi)^{M/2}}{4\pi i}\right)^T \int \prod_{t=1}^T \left(e^{\frac{M}{2}(\Lambda_t\nu_t - \ln\Lambda_t)}d\Lambda_t\right). \quad (183)$$

Since the argument of the exponent in (183) is a complex analytic function of $\{\Lambda_t\}$ and we are interested in the large-M asymptotic, the saddle-point method further simplifies the normalization factor to the form

$$\frac{1}{M}\ln Z(\{\nu_t\}) = \frac{1}{2}\sum_{t=1}^T \operatorname*{extr}_{\Lambda_t}\{\Lambda_t\nu_t - \ln\Lambda_t\} + O(M^{-1})$$
$$= \sum_{t=1}^T \frac{1 + \ln\nu_t}{2} + O(M^{-1}), \qquad (184)$$

where the second equality is obtained by solving the extremization problem. Substituting (184) back to (180) provides an expression for $p^{(M)}(\{u_t\}; \{\nu_t\})$.

Recall the T-orthogonal setup given in Definition 2. Fix the parameters M, β, λ and define

$$G_{\beta,\lambda}^{(M)}(\sigma^2, \{\nu_t\})$$
$$= \frac{1}{M}\ln \mathsf{E}_{w,\{O_t\}} e^{-\frac{\beta}{2\lambda}\|\sigma w - \sum_{t=1}^T O_t\Delta x_t\|^2}$$
$$= \frac{1}{M}\ln \mathsf{E}_w \int p^{(M)}(\{u_t\}; \{\nu_t\})$$
$$\times e^{-\frac{1}{2\lambda}\|\sqrt{\beta\sigma^2}w - \sqrt{\beta}\sum_{t=1}^T u_t\|^2}\prod_{t=1}^T du_t, \quad (185)$$

where $\{O_t\}, \{\Delta x_t\}, \{u_t\}$ and $\{\nu_t\}$ are as before. Applying the Gaussian integration formula (182) from right-to-left along with the expressions (180) and (184) provides

$$G_{\beta,\lambda}^{(M)}(\sigma^2, \{\nu_t\})$$
$$= \frac{1}{M}\ln \mathsf{E}_w \int \prod_{t=1}^T \left(d\Lambda_t e^{\frac{M}{2}\Lambda_t\nu_t}\right)\int dk e^{a(k,w)}$$
$$\times \int \prod_{t=1}^T \left(e^{-\frac{1}{2}\Lambda_t\|u_t\|^2 - i\sqrt{\beta}k^\mathsf{T} u_t}du_t\right)$$
$$+ \frac{1}{2}\ln\frac{\lambda}{2\pi} - \frac{1}{M}\ln Z(\{\nu_t\}), \qquad (186)$$

where $k \in \mathbb{R}^M$, the normalization factor is given in (184) and we denoted

$$a(k, w) = -\frac{\lambda}{2}\|k\|^2 + i\sqrt{\beta\sigma^2}k^\mathsf{T} w. \qquad (187)$$

Using next Gaussian integration repeatedly to assess the expectations w.r.t. $\{u_t\}, k$ and w yields (188) at the top of the next page. We then change the integration variables as $\Lambda_t \to \beta\Lambda_t$, take the limit $M \to \infty$ and employ saddle-point integration. Omitting all terms that vanish in the large-M limit provides the final expression

$$G_{\beta,\lambda}(\sigma^2, \{\nu_t\}) = -\frac{1}{2}\left(T - \ln\lambda + \sum_{t=1}^T \ln(\beta\nu_t)\right)$$
$$+ \frac{1}{2}\operatorname*{extr}_{\{\Lambda_t\}}\left\{\sum_{t=1}^T [\Lambda_t(\beta\nu_t) - \ln\Lambda_t] - \ln\left(\lambda + \beta\sigma^2 + \sum_{t=1}^T \frac{1}{\Lambda_t}\right)\right\}. \qquad (189)$$

Finally, we remark that the extremization in $G_{\beta,\lambda}(\sigma^2, \{\nu_t\})$ as given above enforces the condition

$$\beta\nu_t(\sigma^2, \beta, \lambda) = \frac{1}{\Lambda_t}\left(1 - \frac{\Lambda_t^{-1}}{\lambda + \beta\sigma^2 + \sum_{t=1}^T \Lambda_t^{-1}}\right), \qquad (190)$$

implying $\Lambda_t \in \mathbb{R} \setminus \{0\}$ for all $\{\beta, \lambda, \sigma^2\}$ and $t = 1, \ldots, T$. Thus, the expression (189) together with the condition (190) provides the solution to the integration problem defined in (185). Furthermore, for the special case of $\sigma = 0$ we have

$$\beta\nu_t(\sigma^2 = 0, \beta, \lambda) = \frac{1}{\Lambda_t}\left(1 - \frac{\Lambda_t^{-1}}{\lambda + \sum_{k=1}^T \Lambda_k^{-1}}\right), \qquad (191)$$

$$G_{\beta,\lambda}^{(M)}(\sigma^2,\{\nu_t\}) = \frac{1}{M}\ln \mathsf{E}_{\boldsymbol{w}} \int \prod_{t=1}^{T}\left(\mathrm{d}\Lambda_t e^{\frac{M}{2}\Lambda_t\nu_t - \frac{M}{2}\ln\Lambda_t}\right)\int\exp\left[-\frac{1}{2}\left(\lambda + \sum_{t=1}^{T}\frac{\beta}{\Lambda_t}\right)\|\boldsymbol{k}\|^2 + \mathrm{i}\sqrt{\beta\sigma^2}\boldsymbol{w}^\mathsf{T}\boldsymbol{k}\right]\mathrm{d}\boldsymbol{k}$$

$$+\frac{1}{2}\ln\frac{\lambda}{2\pi} - \frac{1}{M}\ln Z(\{\nu_t\})$$

$$= \frac{1}{M}\ln\int\left(\prod_{t=1}^{T}\mathrm{d}\Lambda_t\right)\exp\left\{\frac{M}{2}\left[\sum_{t=1}^{T}\Lambda_t\nu_t - \sum_{t=1}^{T}\ln\Lambda_t - \ln\left(\lambda + \sum_{t=1}^{T}\frac{\beta}{\Lambda_t}\right)\right]\right\}$$

$$\times\frac{1}{(2\pi)^{M/2}}\int\exp\left\{-\frac{1}{2}\left[1+\beta\sigma^2\left(\lambda+\sum_{t=1}^{T}\frac{\beta}{\Lambda_t}\right)^{-1}\right]\|\boldsymbol{w}\|^2\right\}\mathrm{d}\boldsymbol{w} + \frac{1}{2}\ln\lambda - \frac{1}{M}\ln Z(\{\nu_t\})$$

$$= \frac{1}{M}\ln\int\exp\left\{\frac{M}{2}\left[\sum_{t=1}^{T}\Lambda_t\nu_t - \sum_{t=1}^{T}\ln\Lambda_t - \ln\left(\lambda+\beta\sigma^2+\sum_{t=1}^{T}\frac{\beta}{\Lambda_t}\right)\right]\right\}\prod_{t=1}^{T}\mathrm{d}\Lambda_t$$

$$+\frac{1}{2}\ln\lambda - \frac{1}{M}\ln Z(\{\nu_t\}) \tag{188}$$

---

so that $\nu_t(\sigma^2 = 0, \beta \to \infty, \lambda) \to 0$ and $\beta^{-1}G_{\beta,\lambda}(\sigma^2 = 0, \{\nu_t\}) \xrightarrow{\beta\to\infty} 0$. This is fully compatible with the earlier result obtained in [32], as expected.

B. Rotationally Invariant Setup

Let us consider the case where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is sampled from an ensemble that allows the decomposition $\boldsymbol{R} = \boldsymbol{A}^\mathsf{T}\boldsymbol{A} = \boldsymbol{O}^\mathsf{T}\boldsymbol{D}\boldsymbol{O}$ where $\boldsymbol{O}$ is an $N \times N$ Haar matrix and $\boldsymbol{D} = \mathrm{diag}(d_1,\ldots,d_N)$ contains the eigenvalues of $\boldsymbol{R}$. This is the case of rotationally invariant setup given in Definition 2. Furthermore, let $\{\Delta\boldsymbol{x}_t\}_{t=1}^{T}$ be a set of (fixed) length-$\hat{M}$ vectors satisfying $\|\Delta\boldsymbol{x}_t\|^2 = \hat{M}\nu_t$ for some given non-negative values $\{\nu_t\}$ and $N = T\hat{M}$. For notational convenience, we write $\Delta\boldsymbol{x} \in \mathbb{R}^N$ for the vector obtained by stacking $\{\Delta\boldsymbol{x}_t\}$. The counterpart of (185) reads then

$$H_{\beta,\lambda}^{(N)}(\sigma^2,\{\nu_t\})$$
$$= \frac{1}{N}\ln\mathsf{E}_{\boldsymbol{w},\boldsymbol{A}}e^{-\frac{\beta}{2\lambda}\|\sigma\boldsymbol{w}-\boldsymbol{A}\Delta\boldsymbol{x}\|^2},$$
$$= -\frac{\alpha}{2}\ln\left(1+\frac{\beta\sigma^2}{\lambda}\right)$$
$$+\frac{1}{N}\ln\mathsf{E}_{\boldsymbol{R}}\exp\left[-\frac{1}{2}\left(\frac{\beta}{\lambda+\beta\sigma^2}\right)\Delta\boldsymbol{x}^\mathsf{T}\boldsymbol{R}\Delta\boldsymbol{x}\right], \tag{192}$$

where the second equality follows by using Gaussian integration formula (182) to average over the additive noise term $\boldsymbol{w}$. Recall next the fact that $\boldsymbol{R} = \boldsymbol{O}^\mathsf{T}\boldsymbol{D}\boldsymbol{O}$ and denote $\boldsymbol{u} = \boldsymbol{O}\Delta\boldsymbol{x}$. Since $\boldsymbol{O}$ are Haar matrices and

$$\|\boldsymbol{O}\Delta\boldsymbol{x}\|^2 = T\hat{M}\sum_{t=1}^{T}\frac{\nu_t}{T} = N\nu, \tag{193}$$

where ν is the "empirical average" over $\{\nu_t\}$, we get by the same arguments as in Appendix C-A an expression for $H_{\beta,\lambda}^{(N)}(\sigma^2,\nu)$ as given in (194) at the top of the next page. Considering next the limit of large M and N, we replace the summation in (194) by an integral over the empirical distribution of the eigenvalues (9), so that the outer expectation w.r.t. $\boldsymbol{D}$ becomes an expectation over all empirical eigenvalue distributions of $\boldsymbol{R}$. But when $M, N \to \infty$ with a finite and

fixed ratio $\alpha = M/N$, this expectation is by assumption w.r.t. a probability measure that has a single non-zero point corresponding to the limiting deterministic eigenvalue distribution $F_{\boldsymbol{A}^\mathsf{T}\boldsymbol{A}}$. Finally, using saddle point method to integrate over Λ, we obtain

$$H_{\beta,\lambda}(\sigma^2,\nu)$$
$$= \frac{1}{2}\operatorname*{extr}_{\Lambda}\left\{\Lambda(\beta\nu) - \int\ln\left(\Lambda + \frac{1}{\lambda+\beta\sigma^2}x\right)\mathrm{d}F_{\boldsymbol{A}^\mathsf{T}\boldsymbol{A}}(x)\right\}$$
$$-\frac{\alpha}{2}\ln\left(1+\frac{\beta\sigma^2}{\lambda}\right) - \frac{1+\ln(\beta\nu)}{2} \tag{195}$$
$$= \frac{1}{2}\operatorname*{extr}_{\Lambda}\left\{\Lambda(\beta\nu) - (1-\alpha)\ln\Lambda\right.$$
$$\left. -\alpha\int\ln(\Lambda\beta\sigma^2+\Lambda\lambda+x)\mathrm{d}F_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(x)\right\}$$
$$-\frac{1+\ln(\beta\nu)-\alpha\ln\lambda}{2}, \tag{196}$$

where the second equality is obtained by changing the integral measure and simplifying. For the case $\sigma^2 = 0$, the extremization then provides the condition

$$\Lambda - \frac{1}{\beta\nu} = -\frac{\alpha}{\beta\nu}\left(1 - \int\frac{\Lambda\beta\sigma^2+\Lambda\lambda}{\Lambda\beta\sigma^2+\Lambda\lambda+x}\mathrm{d}F_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(x)\right)$$
$$\xrightarrow{\sigma^2=0} \Lambda - \frac{1}{\beta\nu} = -\frac{\alpha}{\beta\nu}\left[1-(\Lambda\lambda)G_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(-\Lambda\lambda)\right], \tag{197}$$

where we used again the Stieltjes transformation (168) of $F_{\boldsymbol{A}\boldsymbol{A}^\mathsf{T}}(x)$.

APPENDIX D
GEOMETRIC ENSEMBLE

Recall that the geometric singular value ensemble is generated as $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\mathsf{T}$ where $\boldsymbol{U}$ and $\boldsymbol{V}$ are independent Haar matrices. The diagonal elements of $\boldsymbol{\Sigma}$ are the singular values $\sigma_m = \sqrt{a(\kappa)}\tau^{m-1}, m = 1,\ldots,M$ of $\boldsymbol{A}$ with $\tau \in (0,1]$ and $a(\kappa) > 0$ such that $N^{-1}\sum_{m=1}^{M}\lambda_m = 1$ where $\lambda_m = \sigma_m^2$ are the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\mathsf{T}$. Alternatively, we may write $\lambda_{i+1} = a(\kappa)e^{-\gamma_M(i/M)}, i = 0,1,\ldots,M-1$

$$H_{\beta,\lambda}^{(N)}(\sigma^2,\nu) = \frac{1}{N}\ln \mathsf{E}_{\boldsymbol{D}}\int \mathrm{d}\Lambda e^{\frac{\Lambda}{2}N\nu}\int \exp\left[-\frac{1}{2}\boldsymbol{u}^{\mathsf{T}}\left(\Lambda \boldsymbol{I}_N + \frac{\beta}{\lambda+\beta\sigma^2}\boldsymbol{D}\right)\boldsymbol{u}\right]\mathrm{d}\boldsymbol{u}$$
$$-\frac{\alpha}{2}\ln\left(1+\frac{\beta\sigma^2}{\lambda}\right)-\frac{1+\ln\nu}{2}+O(N^{-1})$$
$$=\frac{1}{N}\ln\mathsf{E}_{\boldsymbol{D}}\int\exp\left\{\frac{N}{2}\left[\Lambda(\beta\nu)-\frac{1}{N}\sum_{n=1}^{N}\ln\left(\Lambda+\frac{1}{\lambda+\beta\sigma^2}d_n\right)\right]\right\}\mathrm{d}\Lambda$$
$$-\frac{\alpha}{2}\ln\left(1+\frac{\beta\sigma^2}{\lambda}\right)-\frac{1+\ln(\beta\nu)}{2}+O(N^{-1}) \tag{194}$$

where $\gamma_M = -2M\ln\tau \geq 0$. Letting $M\to\infty$ provides the continuous limit function for the eigenvalues

$$\lambda(t) = A(\kappa)e^{-\gamma t}, \qquad t\in[0,1), \tag{198}$$

where $\gamma > 0$ satisfies

$$\kappa = \frac{\lambda(0)}{\int_0^1 \lambda(t)dt} = \frac{\gamma}{1-e^{-\gamma}}, \tag{199}$$

for the given peak-to-average ratio κ. The normalization condition $N^{-1}\sum_{m=1}^{M}\lambda_m = 1$ becomes now

$$\alpha\int_0^1 A(\kappa)e^{-\gamma t}dt = 1 \iff A(\kappa) = \frac{\kappa}{\alpha}, \tag{200}$$

which means that $\lambda(t)\in[\frac{\kappa}{\alpha}e^{-\gamma},\frac{\kappa}{\alpha}]$.

The function (198) describes the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}$ in the large system limit. Since the order of the eigenvalues and associated eigenvectors does not affect the performance of the reconstruction, we may also consider sampling randomly and uniformly $t\in[0,1)$ and assigning the corresponding eigenvalues according to (198). Then, by construction the limit of (9) for this ensemble is given by $F_{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}(Ae^{-\gamma t})=1-t$, $t\in[0,1)$ or more conveniently

$$F_{\boldsymbol{A}\boldsymbol{A}^T}(x) = \begin{cases} 1+\gamma^{-1}\ln x - \gamma^{-1}\ln A, & \text{if } x\in(Ae^{-\gamma},A], \\ 0, & \text{otherwise,} \end{cases} \tag{201}$$

where we wrote for simplicity $A=A(\kappa)$. This is also called the reciprocal distribution whose density reads

$$f_{\boldsymbol{A}\boldsymbol{A}^T}(x) = \begin{cases} \frac{1}{\gamma x}, & \text{if } x\in(Ae^{-\gamma},A], \\ 0, & \text{otherwise.} \end{cases} \tag{202}$$

For the analysis, one can obtain the Stieltjes transform of (202) directly from the definition (168), as given in Example 4. The sensing matrices for the geometric setup in finite size simulations, on the other hand, can be constructed as follows:

1) Generate $M\times N$ matrix $\boldsymbol{X}$ with IID standard normal elements and calculate the singular value decomposition $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathsf{T}}$. For the Gaussian ensemble, $\boldsymbol{U}$ and $\boldsymbol{V}$ are independent Haar matrices.

2) Find numerically the value of τ that meets the peak-to-average constraint (8) and set

$$a(\kappa) = \frac{1}{N^{-1}\sum_{m=1}^{M}\tau^{2(m-1)}}, \tag{203}$$

so that the average power constraint is satisfied.

3) Replace $\boldsymbol{S}$ by $\boldsymbol{\Sigma}$ to create a sensing matrix $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{T}}$. Note that permutations of the diagonal elements in $\boldsymbol{\Sigma}$ has no impact on the reconstruction performance.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[6] CVX Research, Inc., "CVX: Matlab software for disciplined convex programming," http://cvxr.com/cvx.

[7] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[8] M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[9] D. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via l_1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Nov. 2003.

[10] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203 – 4215, dec. 2005.

[11] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr Approx*, vol. 28, no. 3, pp. 253–263, 2008.

[12] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[13] M. A. Davenport and W. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.

[14] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[15] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[16] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[17] X. Lv, G. Bi, and C. Wan, "The group lasso for stable recovery of block-sparse signal representations," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1371–1382, Apr. 2011.

[18] S. K. Ambat, S. Chatterjee, and K. V. S. Hari, "Fusion of algorithms for compressed sensing," *IEEE Trans. Signal Process.*, vol. 61, no. 14, pp. 3699–3704, Jul. 2013.

[19] D. Donoho and J. Tanner, "Counting faces of randomly projected polytopes when the projection radically lowers dimension," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 1–53, 2009.

[20] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete & Computational Geometry*, vol. 43, no. 3, pp. 522–541, 2010.

[21] ——, "Precise undersampling theorems," *Proc. IEEE*, vol. 98, no. 6, pp. 913–924, Jun. 2010.

[22] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, pp. 18 914–18 919, 2009.

[23] A. Montanari, "Graphical models concepts in compressed sensing," in *Compressed sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012, pp. 394–438.

[24] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[25] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*. Singapore: World Scientific, 1987.

[26] V. Dotsenko, *Introduction to the Replica Theory of Disordered Statistical Systems*. New York: Cambridge University Press, 2001.

[27] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing*. New York: Oxford University Press, 2001.

[28] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 58, no. 3, pp. 1902–1923, Mar. 2012.

[29] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. Annual Allerton Conf. Commun., Contr., Computing*, Sep. 30 - Oct. 2 2009, pp. 52–59.

[30] A. M. Tulino, G. Caire, S. Verdú, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, Jul. 2013.

[31] M. Vehkaperä, Y. Kabashima, S. Chatterjee, E. Aurell, M. Skoglund, and L. Rasmussen, "Analysis of sparse representations using bi-orthogonal dictionaries," in *Proc. IEEE Inform. Theory Workshop*, Sep. 3–7 2012.

[32] Y. Kabashima, M. Vehkaperä, and S. Chatterjee, "Typical l_1-recovery limit of sparse vectors represented by concatenations of random orthogonal matrices," *J. Stat. Mech.*, vol. 2012, no. 12, p. P12003, 2012.

[33] T. Tanaka and J. Raymond, "Optimal incorporation of sparsity information by weighted l_1-optimization," in *Proc. IEEE Int. Symp. Inform. Theory*, Jun. 2010, pp. 1598–1602.

[34] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on l_p-norm minimization," *J. Stat. Mech.*, vol. 2009, no. 9, p. L09003, 2009.

[35] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians, Cavity and Mean Field Models*. Berlin Heidelberg: Springer-Verlag, 2003.

[36] F. Guerra and F. L. Toninelli, "Quadratic replica coupling in the Sherrington-Kirkpatrick mean field spin glass model," *J. Math. Phys.*, vol. 43, no. 7, pp. 3704–3716, 2002.

[37] ——, "The thermodynamic limit in mean field spin glass models," *Commun. Math. Phys.*, vol. 230, no. 1, pp. 71–79, 2002.

[38] F. Guerra, "Broken replica symmetry bounds in the mean field spin glass model," *Commun. Math. Phys.*, vol. 233, no. 1, pp. 1–12, 2003.

[39] M. Talagrand, "The Parisi formula," *Annals of Math*, vol. 163, no. 1, pp. 221–263, 2006.

[40] S. B. Korada and N. Macris, "Tight bounds on the capacity of binary input random CDMA systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5590–5613, Nov. 2010.

[41] A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3221–3246, Sep. 2005.

[42] S. Kudekar and N. Macris, "Sharp bounds for optimal decoding of low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4635–4650, Oct. 2009.

[43] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[44] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

[45] Harish-Chandra, "Differential operators on a semisimple lie algebra," *Amer. J. Math.*, vol. 79, no. 1, pp. 87–120, 1957.

[46] C. Itzykson and J. B. Zuber, "Planar approximation 2," *J. Math. Phys.*, vol. 21, no. 3, pp. 411–421, 1980.

[47] J. Wright and Y. Ma, "Dense error correction via ℓ_1-minimization," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3540–3560, Jul. 2010.

[48] M. Vehkaperä, Y. Kabashima, and S. Chatterjee, "Statistical mechanics approach to sparse noise denoising," in *Proc. European Sign. Proc. Conf.*, Sep. 9–13 2013.

[49] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6241–6263, Oct. 2012.

[50] S. B. Korada and A. Montanari, "Applications of the Lindeberg principle in communications and statistical learning," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2440–2450, Apr. 2011.

[51] P. Viswanath, V. Anantharam, and D. N. C. Tse, "Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1968–1983, Sep. 1999.

[52] K. Kitagawa and T. Tanaka, "Optimization of sequences in CDMA systems: A statistical-mechanics approach," *Computer Networks*, vol. 54, no. 6, pp. 917–924, 2010.

[53] S. Oymak and B. Hassibi, "A case for orthogonal measurements in linear inverse problems," in *Proc. IEEE Int. Symp. Inform. Theory*, 2014, pp. 3175–3179.

[54] C. Thrampoulidis and B. Hassibi, "Isotropically random orthogonal matrices: Performance of LASSO and minimum conic singular values," in *Proc. IEEE Int. Symp. Inform. Theory*, 2015, pp. 556–560.

[55] C.-K. Wen, J. Zhang, K.-K. Wong, J.-C. Chen, and C. Yuen, "On sparse vector recovery performance in structurally orthogonal matrices via LASSO," arXiv:1410.7295 [cs.IT], Oct. 2014.

[56] Y. Kabashima and M. Vehkaperä, "Signal recovery using expectation consistent approximation for linear observations," in *Proc. IEEE Int. Symp. Inform. Theory*, 2014, pp. 226–230.

[57] B. Cakmak, O. Winther, and B. H. Fleury, "S-AMP: Approximate message passing for general matrix ensembles," in *Proc. IEEE Inform. Theory Workshop*, 2014, pp. 192–196.

[58] J. Ma, X. Yuan, and L. Ping, "On the performance of turbo signal recovery with partial DFT sensing matrices," *IEEE Trans. Signal Process.*, vol. 22, no. 10, pp. 1580–1584, Oct. 2015.

[59] M. Opper, B. Cakmak, and O. Winther, "A theory of solving TAP equations for Ising models with general invariant random matrices," arXiv:1509.01229 [cond-mat.dis-nn], Sep. 2015.

[60] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Proc. Annual Allerton Conf. Commun., Contr., Computing*, Sep. 29 - Oct. 1 2010, pp. 347–353.

[61] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *Proc. IEEE Int. Symp. Inform. Theory*, Jul. 1 - 6 2012, pp. 1231–1235.

[62] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *J. Stat. Mech.*, vol. 2012, no. 08, p. P08009, 2012.

[63] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, p. 021005, May 2012.

[64] K. Takeda, S. Uda, and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," *Europhys. Lett.*, vol. 76, no. 6, p. 1193, 2006.

[65] Y. Kabashima, "Inference from correlated patterns: A unified theory for perceptron learning and linear vector channels," *J. Phys. Conf. Ser.*, vol. 95, no. 1, p. 012001, 2008.

[66] S. Rangan, A. K. Fletcher, P. Schniter, and U. Kamilov, "Inference for generalized linear models via alternating directions and Bethe free energy minimization," arXiv:1501.01797 [cs.IT], Jan. 2015.

[67] M. Mézard and A. Montanari, *Information, Physics, and Computation*. New York: Oxford University Press, 2009.

[68] K. Sjöstrand and B. Ersbøll, "SpaSM: A Matlab toolbox for sparse statistical modeling," http://www2.imm.dtu.dk/projects/spasm/.

[69] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 04 2004.

[70] S. A. Orszag and C. M. Bender, *Advanced Mathematical Methods for Scientists and Engineers*. McGraw-Hill, 1978.

[71] G. B. Arfken, H. J. Weber, and F. E. Harris, *Mathematical Methods for Physicists*, 7th ed. Elsevier, 2013.

[72] C. Goutis and G. Casella, "Explaining the saddlepoint approximation," *The American Statistician*, vol. 53, no. 3, pp. 216–224, 1999.

Mikko Vehkaperä received the Ph.D. degree from Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2010. Between 2010–2013 he was a post-doctoral researcher at School of Electrical Engineering, and the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Sweden, and 2013–2015 an Academy of Finland Postdoctoral Researcher at Aalto University School of Electrical Engineering, Finland. He is now a lecturer (assistant professor) at University of Sheffield, Department of Electronic and Electrical Engineering, United Kingdom. He held visiting appointments at Massachusetts Institute of Technology (MIT), US, Kyoto University and Tokyo Institute of Technology, Japan, and University of Erlangen-Nuremberg, Germany. His research interests are in the field of wireless communications, information theory and signal processing. Dr. Vehkaperä was a co-recipient for the Best Student Paper Award at IEEE International Conference on Networks (ICON2011) and IEEE Sweden Joint VT-COM-IT Chapter Best Student Conference Paper Award 2015.

Yoshiyuki Kabashima received the B.Sci., M.Sci., and Ph.D. degrees in physics from Kyoto University, Japan, in 1989, 1991, and 1994, respectively. From 1993 until 1996, he was with the Department of Physics, Nara Women's University, Japan. In 1996, he moved to the Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Japan, where he is currently a professor. His research interests include statistical mechanics, information theory, and machine learning. Dr. Kabashima received the 14th Japan IBM Science Prize in 2000, the Young Scientist Award from the Ministry of Education, Culture, Sports, Science, and Technology, Japan, in 2006, and the 11th Ryogo Kubo Memorial Prize in 2007.

Saikat Chatterjee is an assistant professor and docent in the Dept of Communication Theory, KTH-Royal Institute of Technology, Sweden. He is also part of Dept of Signal Processing, KTH. Before moving to Sweden, he received Ph.D. degree in 2009 from Indian Institute of Science, India. He has published more than 80 papers in international journals and conferences. He was a co-author of the paper that won the best student paper award at ICASSP 2010. His current research interests are signal processing, machine learning, coding, speech and audio processing, and computational biology.