On Sparse Vector Recovery Performance in Structurally Orthogonal Matrices via LASSO

Chao-Kai Wen, Jun Zhang, Kai-Kit Wong, Jung-Chieh Chen, and Chau Yuen

Abstract—In this paper, we consider the compressed sensing problem of reconstructing a sparse signal from an undersampled set of noisy linear measurements. The regularized least squares or least absolute shrinkage and selection operator (LASSO) formulation is used for signal estimation. The measurement matrix is assumed to be constructed by concatenating several randomly orthogonal bases, which we refer to as structurally orthogonal matrices. Such measurement matrix is highly relevant to large-scale compressive sensing applications because it facilitates rapid computation and parallel processing. Using the replica method in statistical physics, we derive the mean-squared-error (MSE) formula of reconstruction over the structurally orthogonal matrix in the large-system regime. Extensive numerical experiments are provided to verify the analytical result. We then consider the analytical result to investigate the MSE behaviors of the LASSO over the structurally orthogonal matrix, with an emphasis on performance comparisons with matrices with independent and identically distributed (i.i.d.) Gaussian entries. We find that structurally orthogonal matrices are at least as good as their i.i.d. Gaussian counterparts. Thus, the use of structurally orthogonal matrices is attractive in practical applications.

Index Terms—Compressed sensing, LASSO, orthogonal measurement matrix, the replica method.

I. INTRODUCTION

Sparse signal reconstruction problems emerge in many engineering fields. For most applications, signals are measured from undersampled sets of noisy linear transformations. Typically, the problem of interest is the reconstruction of a *sparse* signal vector $\mathbf{x}^0 \in \mathbb{C}^{\overline{N}}$ from a set of $\overline{M}(\leq \overline{N})$ noisy measurements $\mathbf{y} \in \mathbb{C}^{\overline{M}}$, which is given by

$$\mathbf{y} = \mathbf{A}\mathbf{x}^0 + \sigma_0 \mathbf{w}, \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{\overline{M} \times \overline{N}}$ is the measurement matrix, and $\sigma_0 \mathbf{w} \in \mathbb{C}^{\overline{M}}$ is the noise vector, with σ_0 representing the noise magnitude. This problem has arisen in many areas, such as in signal processing, communications theory, information science, and statistics, and it is widely known as *compressive sensing* [1, 2].

In the past few years, many recovery algorithms have been proposed. A recent exhaustive list of such algorithms is available in [3, 4]. One popular suboptimal yet low complexity estimator is the ℓ_1-regularized least squares (LS), which is known as the least absolute shrinkage and selection operator (LASSO) [5], which

seeks $\mathbf{x}^0$ by

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{C}^N}{\arg\min} \left\{ \frac{1}{\lambda} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_1 \right\}. \tag{2}$$

In (2), $\lambda > 0$ is a design parameter, and the *complex*[1] ℓ_1-norm is defined as

$$\|\mathbf{x}\|_1 \triangleq \sum_{i=1}^{\overline{N}} |x_i| = \sum_{i=1}^{\overline{N}} \sqrt{(\text{Re}\{x_i\})^2 + (\text{Im}\{x_i\})^2}. \tag{3}$$

Problem (2) is a convex problem, for which various fast and efficient solvers have been proposed. For example, the iterative soft-thresholding method [7, Section 7.1] for solving (2) is given by

$$\hat{\mathbf{x}}^{t+1} := \eta\left(\hat{\mathbf{x}}^t + \varsigma^t \mathbf{A}^H (\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^t), \varsigma^t \lambda\right), \tag{4}$$

where t is the iteration counter, $\varsigma^t > 0$ is the chosen step size, and $\eta(x, \varsigma) \triangleq \frac{x}{|x|}(|x| - \varsigma)_+$ is a soft-thresholding function, in which $(a)_+ = a$ if $a > 0$ and is 0 otherwise. Other common algorithms that follow the iterative thresholding method include the fast iterative soft thresholding algorithm (FISTA)[2] [8] and approximate message passing [9]. An exhaustive examination of these algorithms is presented in [10].

These iterative thresholding algorithms are attractive because they require few computations per iteration and therefore enables the application of the LASSO in large-scale problems. The calculation per iteration only requires one matrix vector multiplication by $\mathbf{A}$ and another by $\mathbf{A}^H$, plus a (negligible) vector addition. In a number of applications, one often uses a very large matrix $\mathbf{A}$, which is not represented explicitly but is applied as an operator only. One such example is a randomly generated discrete Fourier transform (DFT) matrix [11–13]. With DFT matrices as the measurement matrix, fast Fourier transform (FFT) can be used to perform matrix multiplications efficiently, and storing the measurement matrix becomes unnecessary. The entries of a DFT

C. Wen is with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan. J. Zhang is with Nanjing University of Posts and Telecommunications, Nanjing, China. K. Wong is with the Department of Electronic and Electrical Engineering, University College London, London, United Kingdom. J.-C. Chen (Corresponding author) is with the Department of Optoelectronics and Communication Engineering, National Kaohsiung Normal University, Kaohsiung 802, Taiwan. E-mail: jcchen@nknucc.nknu.edu.tw. C. Yuen is with Engineering Product Development, Singapore University of Technology and Design, Singapore.

---

[1] In a real-valued setting, the ℓ_1-norm is defined as $\|\mathbf{x}\|_1 \triangleq \sum_n |x_n|$, which is different from the complex ℓ_1-norm. A simple extension of the LASSO to the complex setting involves considering the complex signal and measurements as a $2\overline{N}$-dimensional real-valued signal and $2\overline{M}$-dimensional real-valued measurements, respectively. However, several studies (e.g., [6]) have shown that the LASSO based on the complex ℓ_1-norm is superior to the simple real-valued extension when the real and imaginary components of signals tend to be either zero or nonzero simultaneously. Therefore, we consider the LASSO using the complex ℓ_1-norm definition of (2) instead of the simple real-valued extension of the LASSO.

[2] FISTA uses the following iteration [8]:

$$\hat{\mathbf{x}}^{t+1} := \eta\left(\hat{\mathbf{x}}^t + \varsigma^t \mathbf{A}^H (\mathbf{y}^t - \mathbf{A}\mathbf{z}^t), \varsigma^t \lambda\right),$$
$$\mathbf{z}^t := \hat{\mathbf{x}}^t + \frac{t-1}{t+2}(\hat{\mathbf{x}}^t - \hat{\mathbf{x}}^{t-1}).$$
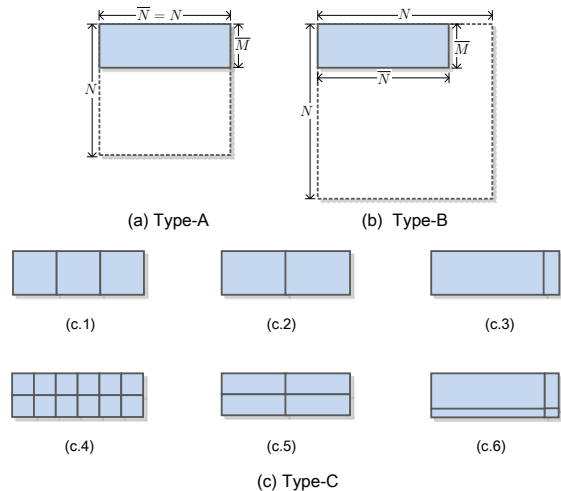
Fig. 1. Examples of structurally random orthogonal matrices.

matrix are not independent and identically distributed (i.i.d.).

In a noiseless setup (i.e., $\sigma = 0$), the measurement matrix exhibits the so-called universality property; that is, measurement matrices with i.i.d. ensembles and rotationally invariant (or row orthonormal) ensembles exhibit the same recovery capability (or phase transition) [14–17]. The universality phenomenon is subsequently extended to measurement matrices, which are constructed by concatenating several randomly square orthonormal matrices [18].

In a *noisy* setting, perfect recovery is rare. Hence, we are interested in the (average) mean squared error (MSE) of reconstruction, as defined by $\overline{N}^{-1}\langle\langle \|\mathbf{x}^0 - \hat{\mathbf{x}}\|_2^2 \rangle\rangle_{\mathbf{w},\mathbf{x}^0}$, where $\langle\langle \cdot \rangle\rangle_{\mathbf{w},\mathbf{x}^0}$ denotes the average with respect to (w.r.t.) $\mathbf{w}$ and $\mathbf{x}^0$. In [19], an analytical expression for the MSE of LASSO reconstruction was obtained when the measurement matrix was a row orthonormal matrix generated randomly. Let a *standard orthonormal matrix* be an $N \times N$ unitary matrix. Row orthonormal matrices can also be obtained by selecting a set of rows from a standard orthonormal matrix. In this case, we refer to such row orthonormal matrix as a Type-A matrix (Figure 1(a)). The work in [19] emphasized support recovery rather than the MSE of reconstruction. The superiority of row orthonormal measurement matrices over their i.i.d. Gaussian counterparts in noisy sparse recovery problems was recently revealed in [20–24].[3] This characteristic is in contrast to that of the noiseless setup. These arguments show that the choice of measurement matrices does inference the MSE of reconstruction when noise is present.

For several real-world applications, a multiprocessor version of the signal reconstruction problem (1) is a subject of interest [28, 29]. For example, each distributed node in a sensor network acquires partial measurements, which are then collected at a single fusion center where signal recovery is performed. The measurement matrix for such applications could be constructed by concatenating several randomly chosen orthonormal bases; we refer to such matrix as the Type-C matrix. As shown in

Figure 1(c), such construction presents several variations because of certain implementation considerations [11–13, 30]. Each sub-block of a measurement matrix is obtained from a partial block of a scrambled DFT matrix. Thus, recovery algorithms, e.g., (4), are able to deal with large signals in real time. Another popular application that uses the concatenations of randomly orthonormal matrices arises in wireless precoding designs [31, 32] and capacity-achieving codes [12, 13]. A natural question is as follows: "*How are the MSEs affected among the different measurement matrices?*" The Type-A and Type-B matrices belong to a class of general unitarily invariant matrices [33]. The corresponding performances of this class have been studied in [19, 20]. The authors of [20] demonstrated that the Type-A and Type-C.1 matrices (constructed by concatenating several randomly *square* orthonormal matrices) exhibit the same performance. In fact, the Type-C.1 matrix is still unitarily invariant. However, the Type-C matrix in general is not unitarily invariant. Only limited progress has been made on the measurement matrix with the general Type-C setup.

In the present study, we aim to provide an analytical characterization of the performance of the LASSO under such measurement matrices. In particular, we derive the MSE of the LASSO in the general Type-C setup by using the replica method from statistical physics [17–20, 34, 35]. Our MSE result encompasses the Type-A and Type-B matrices as special cases. Here, the Type-B matrix is constructed by selecting a set of columns *and* rows from a standard orthonormal matrix, as depicted in Figure 1(b). Computer simulations are conducted to verify the accuracy of our analysis. Then, several observations are made on the basis of such analysis. In particular, we compare the performances and behaviors of both matrices with those of random i.i.d. Gaussian matrices. We show that all the structurally orthogonal matrices (including Types A–C) perform at least as well as random i.i.d. Gaussian matrices over arbitrary setups.[4]

Notations—Throughout this paper, for any matrix $\mathbf{A}$, $[\mathbf{A}]_{i,j}$ refers to the (i,j)th entry of $\mathbf{A}$, $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$, $\mathbf{A}^H$ denotes the conjugate transpose of $\mathbf{A}$, $\mathsf{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$, and $\mathsf{vec}(\mathbf{A})$ is the column vector whose entries are the ordered stacks of columns of $\mathbf{A}$. Additionally, $\mathbf{I}_n$ denotes an n-dimensional identity matrix, $\mathbf{0}$ denotes a zero matrix with an appropriate size, $\mathbf{1}_n$ denotes an n-dimensional all-one vector, $\|\cdot\|_2$ denotes the Euclidean norm, $\mathsf{I}_{\{\text{statement}\}}$ denotes the indicator of the statement, $\langle\langle \cdot \rangle\rangle_X$ represents the expectation operator w.r.t. X, $\log(\cdot)$ is the natural logarithm, $\delta(\cdot)$ denotes Dirac's delta, $\delta_{i,j}$ denotes Kronecker's delta, $\mathsf{Extr}_x\{f(x)\}$ represents the extremization of a function $f(x)$ w.r.t. x, and $\mathsf{Q}(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}\mathrm{d}t$ is the standard Q-function. We consider the complex random variable Z as a standard Gaussian if its density function is given by $\mathcal{N}(z) \triangleq \frac{1}{\pi}e^{-|z|^2}$.

II. PROBLEM FORMULATION

We consider the sparse signal recovery setup in (1), where $\mathbf{w}$ is assumed to be the standard complex Gaussian noise vector. In

---

[3]The significance of orthogonal matrices in other problems (e.g., code-division multiple-access and multiple-input multiple-output systems) was pointed out in [25–27].

[4]To conduct a fair comparison among different setups, we properly normalize their energy consumption.
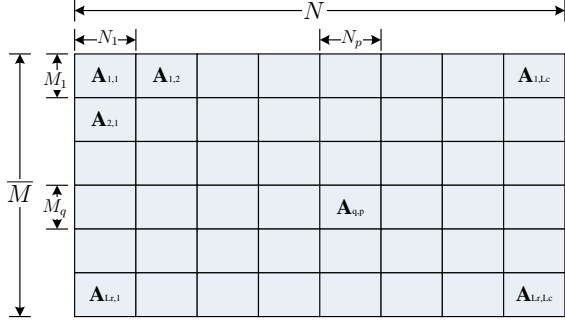
Fig. 2. Example of a structurally random matrix, where $L_c = 8$ and $L_r = 6$. Each block is obtained from an independent standard $N \times N$ orthonormal matrix by selecting M_q rows and N_p columns at random.

addition, let us suppose that

$$P_0(\mathbf{x}^0) = \prod_{n=1}^{\overline{N}} P_0(x_n), \tag{5}$$

where $P_0(x_n) = (1 - \rho_x)\delta(x_n) + \rho_x \mathcal{N}(x_n)$ for $n = 1, \ldots, \overline{N}$, and $\rho_x \in [0, 1]$ is the fraction of non-zero entries in $\mathbf{x}^0$. That is, the elements of $\mathbf{x}^0$ are sparse and i.i.d. according to $P_0(x_n)$.

For generality, we consider the measurement matrix $\mathbf{A}$ made of different blocks, as outlined in Figure 2. We refer to such matrix as the Type-C matrix. The structurally random matrix was also considered in [12, 36] in the context of compressive sensing for different purposes. In the setup, $\mathbf{A} \in \mathbb{C}^{\overline{M} \times \overline{N}}$ is constructed through vertical and horizontal concatenations of $L_r \times L_c$ blocks as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \cdots & \mathbf{A}_{1,L_c} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{L_r,1} & \cdots & \mathbf{A}_{L_r,L_c} \end{bmatrix}, \tag{6}$$

where each $\mathbf{A}_{q,p} \in \mathbb{C}^{M_q \times N_p}$ is drawn *independently* from the Haar measure of an $N \times N$ random matrix $\mathbf{W}_{q,p}$ (referred to as the *standard* orthonormal matrix in this paper) with $\mathbf{W}_{q,p}\mathbf{W}_{q,p}^H = \mathbf{I}$. To shape $\mathbf{A}_{q,p}$ in an $M_q \times N_p$-dimensional matrix, we randomly select M_q rows and N_p columns from the standard orthonormal matrix. We assume that $\mathbf{A}_{q,p}$s are independent. We denote $\mu_p = N_p/N$ and $\nu_q = M_q/N$ as the "column selection rate" and "row selection rate," respectively. We also define $\mu \triangleq \sum_p \mu_p = \overline{N}/N$ and $\nu \triangleq \sum_q \nu_q = \overline{M}/N$. To improve the flexibility of the setup, we assume that for the (q, p)th sub-block, the standard orthonormal matrix is multiplied by $\sqrt{R_{q,p}}$. By setting the values of $R_{q,p}$ appropriately, each block can be made with either zeros only or a partial orthonormal matrix.

Corresponding to the measurement matrix comparing different blocks, the $\overline{N}$ variables of $\mathbf{x}^0$ are divided into L_c blocks $\{\mathbf{x}_p : p = 1, \ldots, L_c\}$ with N_p variables in each block. The $\overline{M}$ measurements $\mathbf{y}$ are divided into L_r blocks $\{\mathbf{y}_q : q = 1, \ldots, L_r\}$ with M_q measurements in each block. Note that we have $\overline{M} = \sum_{q=1}^{L_r} M_q$ and $\overline{N} = \sum_{p=1}^{L_c} N_p$. The measurement ratio of the system is given by $\alpha = \overline{M}/\overline{N}$.

III. ANALYTICAL RESULTS

To facilitate our analysis based on tools in statistical mechanics, we adopt the approach in [18, 20] to reformulate the ℓ_1-regularized LS problem (2) in a probabilistic framework. Suppose that the posterior distribution of $\mathbf{x}$ follows the distribution

$$P_\beta(\mathbf{x}|\mathbf{y}) = \frac{1}{\mathsf{Z}_\beta(\mathbf{y}, \mathbf{A})} e^{-\beta\left(\frac{1}{\lambda}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_1\right)}, \tag{7}$$

where β is a constant and

$$\mathsf{Z}_\beta(\mathbf{y}, \mathbf{A}) = \int d\mathbf{x}\, e^{-\beta\left(\frac{1}{\lambda}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_1\right)} \tag{8}$$

is the partition function (or normalization factor) of the above distribution function. Given the posterior probability of (7), the Bayes approach of estimating $\mathbf{x}$ is given as [37]

$$\langle \mathbf{x} \rangle_{P_\beta} = \int d\mathbf{x}\, \mathbf{x} P_\beta(\mathbf{x}|\mathbf{y}). \tag{9}$$

As $\beta \to \infty$, the posterior mean estimator (9) condenses to the global minimum of (2), i.e., $\langle \mathbf{x} \rangle_{P_\beta} = \hat{\mathbf{x}}$.

In (9), $\langle \mathbf{x} \rangle_{P_\beta}$ (or equivalently $\hat{\mathbf{x}}$) is estimated from $\mathbf{y}$ given that $\mathbf{A}$ is perfectly known. Clearly, $\hat{\mathbf{x}}$ depends on $\mathbf{y}$ and is thus *random*. We are thus interested in the (average) MSE of $\hat{\mathbf{x}}$, i.e.,

$$\begin{aligned} \mathsf{mse} &= \overline{N}^{-1}\langle\!\langle \|\mathbf{x}^0 - \hat{\mathbf{x}}\|_2^2 \rangle\!\rangle_\mathbf{y} \\ &= \rho_x - 2\overline{N}^{-1}\langle\!\langle \mathrm{Re}\{\hat{\mathbf{x}}^H \mathbf{x}^0\} \rangle\!\rangle_\mathbf{y} + \overline{N}^{-1}\langle\!\langle \hat{\mathbf{x}}^H \hat{\mathbf{x}} \rangle\!\rangle_\mathbf{y}, \end{aligned} \tag{10}$$

where $\langle\!\langle \cdot \rangle\!\rangle_\mathbf{y}$ denotes an average over $\mathbf{y}$. Specifically, we define

$$\langle\!\langle f(\mathbf{y}) \rangle\!\rangle_\mathbf{y} \triangleq \int d\mathbf{y} \int d\mathbf{x}^0 f(\mathbf{y}) P(\mathbf{y}|\mathbf{x}^0)) P_0(\mathbf{x}^0), \tag{11}$$

where $P_0(\mathbf{x}^0)$ is defined by (5) and

$$P(\mathbf{y}|\mathbf{x}^0) = \frac{1}{(\pi\sigma_0^2)^{\overline{M}}} e^{-\frac{1}{\sigma_0^2}\|\mathbf{y} - \mathbf{A}\mathbf{x}^0\|_2^2} \tag{12}$$

is the conditional distribution of $\mathbf{y}$ given $\mathbf{x}^0$ under (1). Our aim is to derive an analytical result for mse.

In the analysis of mse, we assume that L_r, L_c are *finite* and consider $N, N_p, M_q \to \infty$, $N \to \infty$ while keeping $\mu_p = N_p/N$ and $\nu_q = M_q/N$ fixed and finite for $p = 1, \ldots, L_c$, $q = 1, \ldots, L_r$. For convenience, we refer to this large dimensional regime as $N \to \infty$. Notice that the MSE depends upon the measurement matrix $\mathbf{A}$. However, in the large regime $N \to \infty$, we expect (or assume) that the average MSE appears to be self-averaging. That is to say, the MSE for any typical realization of $\mathbf{A}$ coincides with its average over $\mathbf{A}$.

According to (10), the posterior distribution P_β plays a role in the MSE. In statistical mechanics, the key in finding the MSE is to compute the partition function, which is the marginal of P_β or its logarithm, known as *free energy*. Following the argument in [17, 20], we show that mse is a saddle point of free energy. In the virtue of the self-averaging property in the large-dimensional regime, we therefore compute mse by computing the *average* free energy as

$$\Phi = -\lim_{\beta, N \to \infty} \frac{1}{\beta N} \langle\!\langle \log \mathsf{Z}_\beta(\mathbf{y}, \mathbf{A}) \rangle\!\rangle_{\mathbf{y}, \mathbf{A}}. \tag{13}$$

A similar manipulation has been used in many different settings, e.g., [18–20, 35, 36]. The analysis of (13) is unfortunately still

difficult. The major difficulty in the analysis of (13) lies in the expectations over $\mathbf{y}$ and $\mathbf{A}$. Nevertheless, we can facilitate the mathematical derivation by rewriting Φ as [17, 20]

$$\Phi = -\lim_{\beta, N \to \infty} \frac{1}{\beta N} \lim_{\tau \to 0} \frac{\partial}{\partial \tau} \log \left\langle\!\left\langle Z_\beta^\tau(\mathbf{y}, \mathbf{A}) \right\rangle\!\right\rangle_{\mathbf{y}, \mathbf{A}}, \quad (14)$$

which we achieve by moving the expectation operator inside the log function. Then, we evaluate $\log \left\langle\!\left\langle Z_\beta^\tau(\mathbf{y}, \mathbf{A}) \right\rangle\!\right\rangle_{\mathbf{y}, \mathbf{A}}$ through the following steps. We interchange the limits $\tau \to 0$ and $\beta, N \to \infty$ and employ the replica symmetry (RS) ansatz to evaluate $\log \left\langle\!\left\langle Z_\beta^\tau(\mathbf{y}, \mathbf{A}) \right\rangle\!\right\rangle_{\mathbf{y}, \mathbf{A}}$ for an integer-valued τ. The obtained expression is assumed to be an analytic continuation in $\tau = 0^+$ and then used to obtain the limit $\tau \to 0$. This track, which is called the replica method, originated from the field of statistical physics [38, 39], and it is still lacking mathematical validation. Nevertheless, the replica method has been successfully applied in the information/communications theory literature [17–20, 26, 34–36, 40–46] and is thus a reasonable approach.

Details of the replica calculation are provided in Appendix A. At this point, we intend to elucidate the final analytical results (i.e., Claim 1 to be presented later). The replica analysis allows us to understand the characteristics of the errors made by the LASSO by looking at signal reconstruction via an equivalent scalar version of the linear system (1). That is,

$$y_p = \hat{m}_p x_p^0 + \sqrt{\hat{\chi}_p} z_p, \quad (15)$$

where the subscript p indicates that the equivalent linear system characterizes the signal in block p, i.e., $\mathbf{x}_p$, and N_p parallel equivalent linear systems of (15) comprise block p. The parameters $(\hat{m}_p, \hat{\chi}_p)$ are derived from the replica analysis and presented later in Claim 1, and x_p^0 is a random signal generated according to the distribution $P_0(x)$, z_p is standard complex Gaussian, and y_p can be regarded as the effective measurement.

In particular, our analysis reveals that the characteristics of the LASSO output corresponding to the signal $\mathbf{x}_p$ can be analyzed via the LASSO output of the signal x_p^0 through the effective measurement y_p, where $\hat{m}_p$ and $\hat{\chi}_p$ denote the *effective* measurement gain and *effective* noise level, respectively. Therefore, according to (2), the recovery of x_p^0 from y_p via the LASSO becomes

$$\hat{x}_p = \operatorname*{argmin}_{x_p \in \mathbb{C}} \left\{ \frac{1}{\hat{m}_p} |y_p - \hat{m}_p x_p|^2 + |x_p| \right\}. \quad (16)$$

Using [6, Lemma V.1], the optimal solution $\hat{x}_p$ of (16) is

$$\hat{x}_p = \frac{\left(|y_p| - \frac{1}{2} \right)_+ \frac{y_p}{|y_p|}}{\hat{m}_p}. \quad (17)$$

Note that $\hat{x}_p$ depends on y_p and is therefore random. Then, the MSE of $\hat{x}_p$ is given by $\left\langle\!\left\langle |x_p^0 - \hat{x}_p|^2 \right\rangle\!\right\rangle_{y_p} = \rho_x - 2 \left\langle\!\left\langle \operatorname{Re}\{\hat{x}_p^* x_p^0\} \right\rangle\!\right\rangle_{y_p} + \left\langle\!\left\langle |\hat{x}_p|^2 \right\rangle\!\right\rangle_{y_p}$, where $\left\langle\!\left\langle \cdot \right\rangle\!\right\rangle_{y_p}$ denotes an average over y_p with

$$P(y_p | x_p^0) = \frac{1}{\pi \hat{\chi}_p} e^{-\frac{1}{\hat{\chi}_p} |y_p - \hat{m}_p x_p^0|^2}. \quad (18)$$

As N_p parallel equivalent systems comprise block p, the MSE of

the LASSO reconstruction in group p is given by

$$\begin{aligned} \mathsf{mse}_p &= \frac{N_p}{\overline{N}} \left\langle\!\left\langle |x_p^0 - \hat{x}_p|^2 \right\rangle\!\right\rangle_{y_p} = \frac{\mu_p}{\mu} \left\langle\!\left\langle |x_p^0 - \hat{x}_p|^2 \right\rangle\!\right\rangle_{y_p} \\ &= \frac{1}{\mu} \left(\mu_p \rho_x - 2\mu_p \left\langle\!\left\langle \operatorname{Re}\{\hat{x}_p^* x_p^0\} \right\rangle\!\right\rangle_{y_p} + \mu_p \left\langle\!\left\langle |\hat{x}_p|^2 \right\rangle\!\right\rangle_{y_p} \right) \\ &= \frac{1}{\mu} \left(\mu_p \rho_x - 2m_p + Q_p \right), \end{aligned} \quad (19)$$

where the second equality is the result of $\mu_p = N_p/N$ and $\mu = \overline{N}/N$, and the last equality is based on the fact that $m_p \triangleq \mu_p \left\langle\!\left\langle \operatorname{Re}\{\hat{x}_p^* x_p^0\} \right\rangle\!\right\rangle_{y_p}$ and $Q_p \triangleq \mu_p \left\langle\!\left\langle |\hat{x}_p|^2 \right\rangle\!\right\rangle_{y_p}$. Using (17) and (18) and following the steps in [19, (349)–(357)], we can derive the analytical expressions of m_p and Q_p and subsequently obtain the expression of mse_p. We normalize mse_p in (19) by a factor $\frac{N_p}{\overline{N}}$. Thus, the MSE over the entire vector is $\mathsf{mse} = \sum_{p=1}^{L_c} \mathsf{mse}_p$. We summarize the results in the following claim.

Claim 1: Consider a Type-C matrix as the measurement matrix. Let mse_p denote the MSE of the LASSO reconstruction in block $p = 1, \ldots, L_c$, and define

$$g_{\mathsf{c}}(\zeta) \triangleq \zeta e^{-\frac{1}{4\zeta}} - \sqrt{\pi \zeta} \mathsf{Q}\left(\frac{1}{\sqrt{2\zeta}} \right), \quad (20a)$$

$$\acute{g}_{\mathsf{c}}(\zeta) \triangleq e^{-\frac{1}{4\zeta}} - \sqrt{\frac{\pi}{4\zeta}} \mathsf{Q}\left(\frac{1}{\sqrt{2\zeta}} \right). \quad (20b)$$

Then, as $N \to \infty$, the average MSE over the entire vector becomes $\mathsf{mse} = \sum_{p=1}^{L_c} \mathsf{mse}_p$, where $\mathsf{mse}_p = (\mu_p \rho_x - 2m_p + Q_p)/\mu$ with

$$m_p = \mu_p \rho_x \acute{g}_{\mathsf{c}}(\hat{m}_p^2 + \hat{\chi}_p), \quad (21a)$$

$$Q_p = \mu_p \left(\frac{1 - \rho_x}{\hat{m}_p^2} g_{\mathsf{c}}(\hat{\chi}_p) + \frac{\rho_x}{\hat{m}_p^2} g_{\mathsf{c}}(\hat{m}_p^2 + \hat{\chi}_p) \right). \quad (21b)$$

In (21), we define

$$\hat{m}_p \triangleq \frac{\sum_{q=1}^{L_r} \Delta_{q,p}}{\chi_p}, \quad (22)$$

where

$$\Delta_{q,p} = \nu_q \frac{\frac{R_{q,p}}{\Gamma_{q,p}^\star}}{\lambda + \sum_{l=1}^{L_c} \frac{R_{q,l}}{\Gamma_{q,l}^\star}}, \quad (23)$$

$$\chi_p = \mu_p \left(\frac{1 - \rho_x}{\hat{m}_p} \acute{g}_{\mathsf{c}}(\hat{\chi}_p) + \frac{\rho_x}{\hat{m}_p} \acute{g}_{\mathsf{c}}(\hat{m}_p^2 + \hat{\chi}_p) \right). \quad (24)$$

The parameters $\Gamma_{q,p}^\star$ and $\hat{\chi}_p = \sum_{q=1}^{L_r} \hat{\chi}_{q,p}$ are the solutions of the coupled equations

$$\Gamma_{q,p}^\star = \frac{1 - \Delta_{q,p}}{\chi_p}, \quad (25a)$$

$$\hat{\chi}_{q,p} = \sum_{r=1}^{L_c} \left(\mathsf{mse}_r - \frac{\sigma_0^2 \chi_r}{\lambda} \right) \Gamma_{q,p,r}' + \frac{\mathsf{mse}_p}{\chi_p^2} - \frac{\sigma_0^2}{\lambda} \Gamma_{q,p}^\star, \quad (25b)$$

where

$$\begin{aligned} \Gamma_{q,p,r}' = \Bigg(\frac{1}{\nu_q} \frac{\Delta_{q,p} \Delta_{q,r} \Gamma_{q,p}^\star \Gamma_{q,r}^\star}{(1 - 2\Delta_{q,p})(1 - 2\Delta_{q,r})} &\left(1 + \sum_{l=1}^{L_c} \frac{1}{\nu_q} \frac{\Delta_{q,l}^2}{1 - 2\Delta_{q,l}} \right)^{-1} \\ &- \frac{\Gamma_{q,r}^{\star 2}}{1 - 2\Delta_{q,r}} \delta_{p,r} \Bigg). \end{aligned} \quad (26)$$

Proof: See Appendix A. □

Note that except for $\{m_p, Q_p\}$, the remaining parameters in Claim 1 originate from the replica analysis and are auxiliary. The parameters $\{\Gamma^\star_{q,p}, \hat{\chi}_{q,p}\}$ should be solved in (25) for all p, q.

We must point out that Claim 1 provides not only a new finding but also a unified formula that embraces previously known results [19, 20]. For example, the MSE of the LASSO under the Type-A measurement matrix in [19] can be obtained if we set $L_c = L_r = 1$ and $\mu_1 = 1$ in Claim 1. Apparently, by setting $\mu_1 < 1$, we are also able to further study the MSE of LASSO under the Type-B measurement matrix. In the next section, we discuss the MSEs of the LASSO under the Type-A and Type-B measurement matrices. We also compare their performances and behaviors with those of random i.i.d. Gaussian matrices.

Another existing result is related to the Type-C.1 measurement matrix [20], in which $L_r = 1$ and $\mu_p = 1$ for $p = 1, \ldots, L_c$. In [20], a Type-C.1 orthogonal matrix was referred to as the T-orthogonal matrix because the matrix was constructed by concatenating T independent standard orthonormal matrices. In addition, in [20], only the real-valued setting was considered; the signal $\mathbf{x}^0$, the measurements $\mathbf{y}$, and the measurement matrix $\mathbf{A}$ were all real-valued settings. In this case, the ℓ_1-norm is defined as $\|\mathbf{x}\|_1 \triangleq \sum_n |x_n|$, which is different from the definitation of the complex ℓ_1-norm (see footnote 1). In the real-valued setting, the analytical MSE expression of the LASSO in Claim 1 holds while g_c and $\acute{g}_\mathsf{c}$ in (20) are replaced with

$$g_\mathsf{r}(\zeta) \triangleq -2\left(\sqrt{\frac{\zeta}{2\pi}} e^{-\frac{1}{2\zeta}} - (1+\zeta)\mathsf{Q}\left(\frac{1}{\sqrt{\zeta}}\right)\right), \qquad (27a)$$

$$\acute{g}_\mathsf{r}(\zeta) \triangleq 2\mathsf{Q}\left(\frac{1}{\sqrt{\zeta}}\right). \qquad (27b)$$

As mentioned previously, the replica method still lacks mathematical rigor. In particular, the RS assumption is not always correct, and in some cases, the replica symmetry breaking (RSB) solution is needed to yield satisfactory results. However, convex optimization problems are generally believed to not exhibit RSB [47]. Our problem of interest (2) is convex. In addition, as demonstrated in the next section, the RS solution for the MSE of the LASSO agrees with the numerical simulations. We therefore restrict the analysis to the RS case, although an investigation into the RSB solution for the current setting may be required.

IV. DISCUSSIONS

To improve our understanding of the performance under the general Type-C setup, we first compare the performances of the Type-A and Type-B measurement matrices with those of random i.i.d. Gaussian matrices. Recall that in the Type-B setup, the matrix $\mathbf{A}$ is constructed by randomly selecting $\overline{M}$ rows and $\overline{N}$ columns from the standard orthonormal matrix. To achieve a fair comparison, we normalize all cases of the measurement matrices such that $\langle\langle\mathrm{tr}(\mathbf{A}\mathbf{A}^H)\rangle\rangle_\mathbf{A} = \overline{M}$ (referred to as the power constraint of the measurement matrix). If the elements of $\mathbf{A}$ are i.i.d. Gaussian random variables with zero mean and variance $1/\overline{N}$, then the power constraint of the measurement matrix is satisfied. We call this matrix the i.i.d. Gaussian matrix. For the orthogonal matrices, we set the gain factor $R_{q,p} = N/\overline{N} = 1/\mu, \forall q, p$ to satisfy this power constraint.
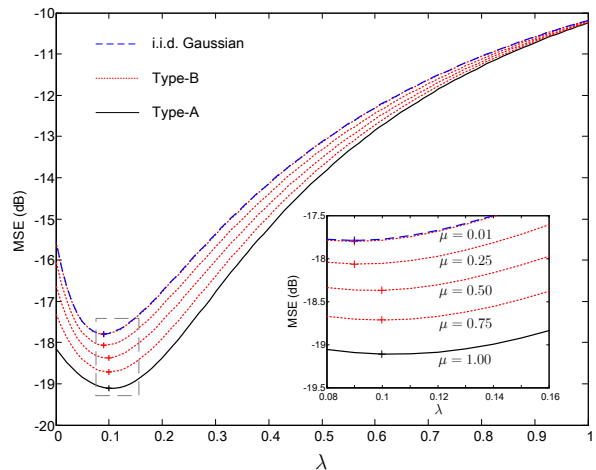

Fig. 3. Average MSE against the regularization parameter λ for $\alpha = 0.5$, $\sigma_0^2 = 10^{-2}$, and $\rho_x = 0.15$. The dashed line is the MSE of the i.i.d. Gaussian setup, the solid line is the MSE of the Type-A setup, and the dotted lines are the MSE of the Type-B setup with the selection rate $\mu = 0.75, 0.5, 0.25, 0.01$ (from bottom to top). Since the measurement ratio is fixed, we vary the row selection rate according to $\nu = \alpha \times \mu$ for each μ. Markers correspond to the lowest MSE w.r.t. the regularization parameter.

As argued in [21], for *noisy* measurements (1), the singular value distribution of the measurement matrix plays a key role in LASSO reconstruction, and measurement matrices with similar singular value characteristics exhibit similar MSE behaviors. In the Type-B setup, the empirical distribution of the eigenvalues of $\mathbf{A}\mathbf{A}^H$ was given by [48, Theorem 3.1]. In particular, the work [49, Theorem 3] proved that if $\overline{M}$ is of order $o(N/\log N)$ and $\overline{M} \to \infty$, $\overline{M}/\overline{N} \to \alpha > 0$, then the empirical distribution converges in probability to the Marchenko-Pastur law. That is, the empirical distribution of the Type-B matrix with a *small* column selection rate is roughly similar to that of the i.i.d. Gaussian matrix. This characteristic implies that the MSEs of the LASSO under the preceding setups are similar.

In Figure 3, we compare the theoretical MSEs of the LASSO for various regularization parameters λ under different types of measurement matrices. The solid, dotted, and dashed lines correspond to the MSEs under the Type-A setup, Type-B, and i.i.d. Gaussian setups, respectively. The theoretical MSE of the LASSO under i.i.d. Gaussian matrices is given by [19, (133)], whereas those under the Type-A and Type-B matrices can be obtained by substituting the corresponding parameters of the Type-A and Type-B setups into Claim 1. The column selection rated from bottom to top are $\mu = 1$ (Type-A) and $\mu = 0.75, 0.5, 0.25, 0.01$ (Type-B). As expected, the MSE of the LASSO under the Type-B setup with a small column selection rate is similar to that under the i.i.d. Gaussian matrix. Furthermore, Figure 3 demonstrates that the MSE performance of the Type-B setup degrades with decreasing column selection rate μ while they are at least as good as their random i.i.d. Gaussian counterparts. Markers show the lowest MSE w.r.t. the regularization parameter λ. Moreover, the optimal value of λ depends on the matrix ensemble, although such dependence is not overly sensitive.

Next, we study the MSE of the LASSO under the Type-C measurement matrices. Recall that Type-C matrices are at-

TABLE I
COMPARISON BETWEEN THE EXPERIMENTAL AND THEORETICAL MSEs OF LASSO UNDER THE FOUR MEASUREMENT MATRICES FOR $\lambda = 0.1$, $\sigma_0^2 = 10^{-2}$, AND $\rho_x = 0.15$.

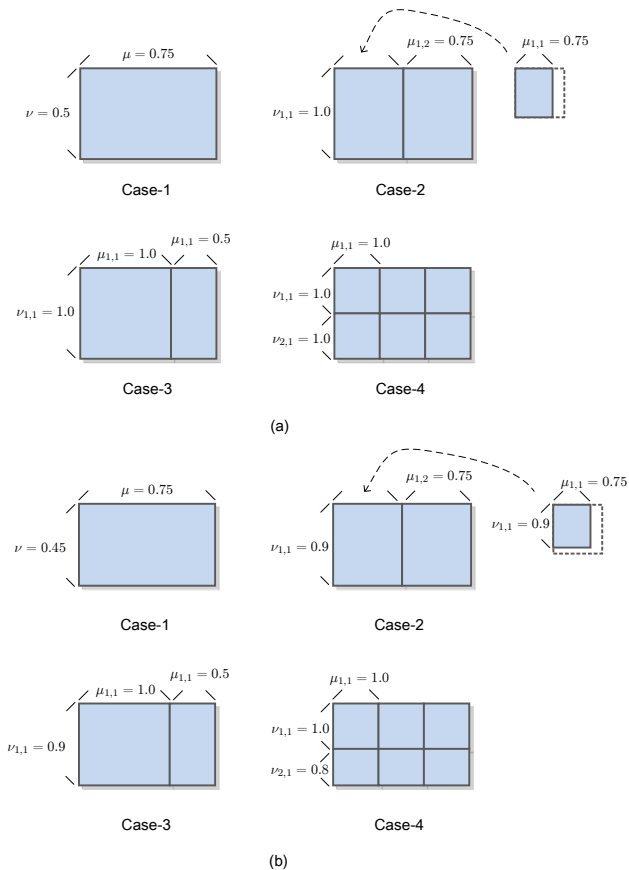|		Case-1	Case-2	Case-3	Case-4
(a) $\overline{M}/\overline{N} = 0.5/0.75$	Theory (dB)	-21.09	-21.09	-21.10	-20.72
	Experiment (dB)	-21.09	-21.09	-21.11	-20.72
(b) $\overline{M}/\overline{N} = 0.45/0.75$	Theory (dB)	-20.32	-20.32	-20.36	-19.96
	Experiment (dB)	-20.32	-20.32	-20.36	-19.96

Fig. 4. The four examples of constructing measurement matrices with (a) $\overline{M} = 0.5 \times 2^{15}$ and $\overline{N} = 0.75 \times 2^{15}$, and with (b) $\overline{M} = 0.45 \times 2^{15}$ and $\overline{N} = 0.75 \times 2^{15}$. Each block is obtained from an independent scrambled DFT matrix. The original DFT matrices of the four cases have dimensions $N = 2^{15}$, $N = 2^{14}$, $N = 2^{14}$, and $N = 2^{13}$, respectively.

tractive because of implementation considerations [11, 30]. For example, to recover $0.75 \times 4,096 = 3,072$ sparse signals with $0.5 \times 4,096 = 2,048$ measurements by using DFT operators, we employ at least four approaches, as shown in Figure 4(a). Specifically, 1) the Type-B measurement matrix is derived by selecting $3,072$ columns and $2,048$ rows from the $4,096 \times 4,096$ DFT matrix, and 2) the Type-C.2 measurement matrix is concatenated by two matrices with $\mathbf{A}_{1,1}, \mathbf{A}_{1,2} \in \mathbb{C}^{2048 \times 1536}$. The two matrices are obtained from partially scrambled $2,048 \times 2,048$ DFT matrices with the column selection rate $\mu = 0.75$. 3) The Type-C.3 measurement matrix is concatenated by two matrices with

$\mathbf{A}_{1,1} \in \mathbb{C}^{2048 \times 2048}$ and $\mathbf{A}_{1,2} \in \mathbb{C}^{2048 \times 1024}$. The two matrices are taken from randomly scrambled $2,048 \times 2,048$ DFT matrices, and the additional column selection with rate $\mu_{1,2} = 0.5$ is used to obtain $\mathbf{A}_{1,2}$. 4) The Type-C.4 measurement matrix is concatenated by six randomly scrambled $1,024 \times 1,024$ DFT matrices, namely, $\mathbf{A}_{1,1}, \mathbf{A}_{1,2}, \mathbf{A}_{1,3}, \mathbf{A}_{2,1}, \mathbf{A}_{2,2}, \mathbf{A}_{2,3} \in \mathbb{C}^{1024 \times 1024}$. In contrast to the implementtion of the Type-B setup, the implementations of the Type-C.2, Type-C.3, and Type-C.4 setups can exploit parallelism or distributed computation, whith the Type-C.4 setup achieving the best structure for parallel computations. Therefore, a natural problem is how their MSEs are affected among the different measurement matrices.

We first conduct extensive numerical experiments to verify the theoretical results in Claim 1. In the experiments, we use the example of the four cases mentioned above while enlarging the dimensions proportionally so that $\overline{M} = 0.5 \times 2^{15}$ and $\overline{N} = 0.75 \times 2^{15}$. Unlike those works employing the linear programming method [50] for the LASSO problem, we use FISTA (see footnote 2) in conjunction with the FFT operators. This approach allows us to deal with signal sizes as large as 10^5 on a typical personal computer in about a few seconds. In this regard, orthogonal matrices are obviously highly relevant for large-scale compressive sensing applications; thus, the theoretical result based on the assumption of $N \to \infty$ is useful. The experimental average MSEs of the LASSO under the four measurement matrices are listed in Table I, which also lists the theoretical MSE estimates by Claim 1 for comparison. The experimental average MSE is obtained by averaging over 10,000 independent realizations. The selected column and row sets at each realization are changed randomly. In the same table, we repeat the previous experiment but use a different measurement rate with $\overline{M} = 0.45 \times 2^{15}$ and $\overline{N} = 0.75 \times 2^{15}$. The corresponding four cases are depicted in Figure 4(b). For all cases presented in the tables, the differences between the two estimates are only evident in the last digits. Therefore, we confirm that Claim 1 provides an excellent estimate of the MSE of the LASSO in large systems.

Next, we use the theoretical expression to examine the behaviors of MSEs under the Type-C measurement matrices. In Figure 5, we provide and compare the MSEs of the LASSO as a function of the regularization parameter λ for the four cases shown in Figure 4. The MSEs under the Type-A matrices and i.i.d. Gaussian counterparts are also plotted as references. Note that the i.i.d. Gaussian matrices are generated directly according to the dimension setting and that they follow the power constraint. The Type-A setup always achieves the best MSE result, whereas the i.i.d. Gaussian setup yields the worst MSE result. However,
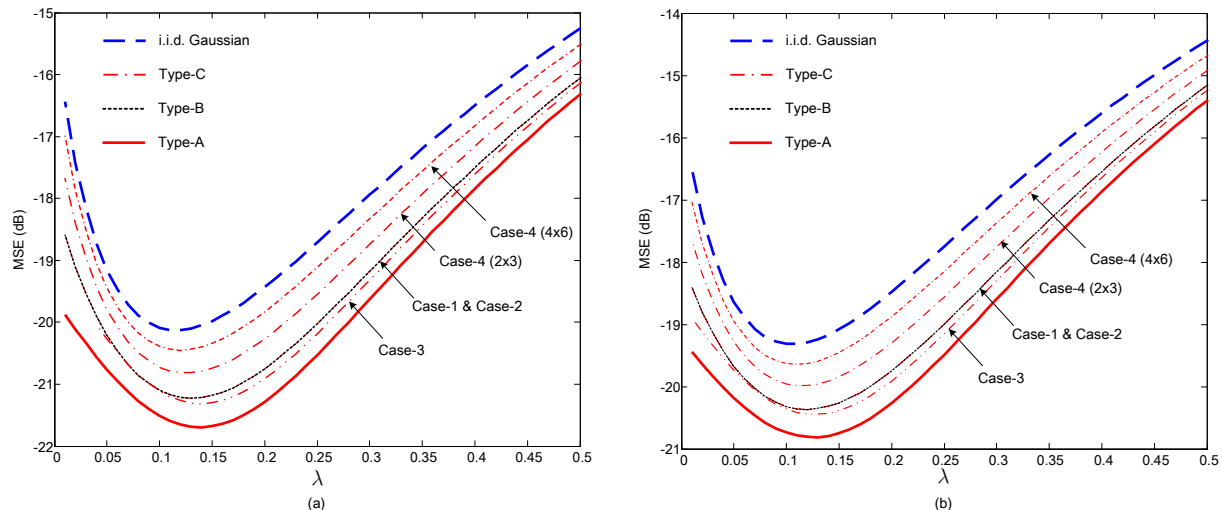
Fig. 5. Average MSE against the regularization parameter λ for $\sigma_0^2 = 10^{-2}$ and $\rho_x = 0.15$, (a) $\alpha = 0.5/0.75$, and (b) $\alpha = 0.45/0.75$. The dashed line is the MSE of the i.i.d. Gaussian setup, the solid line is the MSE of the Type-A setup, and the dotted lines denote the MSE of the Type-B setup with the selection rate $\mu = 0.75, 0.5, 0.25, 0.01$ (from bottom to top). Markers correspond to the lowest MSE w.r.t. the regularization parameter.

the Type-A setup is not always useful if the corresponding size of the FFT operators is not available in some DSP chips. Moreover, Case-1 and Case-2 always demonstrate the same MSE behaviors. This finding motivates us to obtain the following observation that can match the performance of the Type-B matrix via concatenating orthonormal bases.

Observation 1: Consider a Type-B measurement matrix with column and row selection rates μ and ν, respectively. The MSE of the LASSO is identical to that under the horizontal concatenation of L_c matrices, in which each matrix originates from a partial orthonormal matrix with column and row selection rates μ and $L_c\nu$. For a meaningful construction, L_c should be subjected to $L_c\nu \le 1$.

In the application of this observation, let us consider two examples. First, consider the Type-B measurement matrix with $\mu = 1.0$ and $\nu = 0.25$. With the application of Observation 1, the MSE of the LASSO under the row orthonormal measurement matrix is identical to that under the measurement matrix of $\mathbf{A} = [\mathbf{A}_{1,1} \mathbf{A}_{1,2} \mathbf{A}_{1,3} \mathbf{A}_{1,4}]$, with each $\mathbf{A}_{1,p}$ being a square orthogonal matrix. This argument was also revealed in [20]. The columns of each $\mathbf{A}_{1,p}$ are orthogonal so that no interference occurs within each square orthogonal matrix. The interference resulting from the other sub-block of the measurement matrix reduces the MSE. Thus, we can infer that the matrix constructed by concatenating many square orthonormal matrices should lose its advantage over the i.i.d. Gaussian matrix. In other words, if the measurement ratio is small, i.e., $\overline{M} \ll \overline{N}$, the MSEs of the LASSO under the row orthonormal measurement matrix and those under the i.i.d. Gaussian matrix should be comparable. This inference also seems reasonable from the aspect of eigenvalue spectrum [21]; that is, when $\overline{M} \ll \overline{N}$, an i.i.d. Gaussian matrix comparies approximately orthogonal rows and behaves similarly to a row orthonormal matrix.

Another example is the Type-B measurement matrix with $\mu = 0.6$ and $\nu = 0.3$. With Observation 1, the MSE of the LASSO under this matrix is identical to that under the measurement

matrix of $\mathbf{A} = [\mathbf{A}_{1,1} \mathbf{A}_{1,2} \mathbf{A}_{1,3}]$, with each $\mathbf{A}_{1,p}$ being a partial orthogonal matrix with $\mu_{1,p} = 0.6$ and $\nu_{1,p} = 0.9$. In this case, the columns of each $\mathbf{A}_{1,p}$ are not orthogonal but are *nearly* orthogonal. Therefore, we can expect some performance degradation. Case-1 and Case-2 in Figure 5 clearly exhibit the same MSE behaviors, but Case-2 features a better structure for parallel computation and entails less requirement with regard to the size of the FFT operator.

In Figure 5, we also observe that the measurement matrix constructed by the vertical *and* horizontal concatenation of several blocks, i.e., Case-4, achieves the worst performance among the structurally orthogonal matrices. In fact, if we continue to increase the number of concatenation blocks, e.g., $L_r \times L_c = 4 \times 6$, then their MSEs decrease accordingly. Nevertheless, they are at least as good as their random i.i.d. Gaussian counterparts. This observation therefore provides us another way to match the random i.i.d. Gaussian matrix by vertically and horizontally concatenating orthonormal bases.

By comparing the four cases shown in Figure 5, we notice that if the Type-A matrix is not available, Case-3 provides the best MSE result. This observation, together with the previous experiments, indicates that to construct a measurement matrix that achieves good MSE performance in LASSO formulation, one should follow the example of Case-3. That is to say, a row-orthogonal matrix that can best fit the dimension of the measurement matrix should first be used. Then, the remaining part should be horizontally concatenated.

V. CONCLUSION

We investigated the MSE performance of the estimation of a sparse vector with an undersampled set of noisy linear transformations when the measurement matrix is constructed by concatenating several randomly chosen orthonormal bases and when LASSO is adopted. Using the replica method, we derived the theoretical MSE result. Extensive numerical experiments demonstrated excellent agreement with the theoretical result. Our numerical results

also revealed the fact that structurally orthogonal matrices are at least as good as i.i.d. Gaussian matrices. In particular, we derived the following observations.

- Type-A matrices (or row-orthogonal matrices) achieve the best MSE performance among all the other types of structurally orthogonal matrices and perform significantly better than i.i.d. Gaussian matrices.
- The advantage of row-orthogonal matrices over i.i.d. Gaussian matrices is retained even when a random set of columns is removed (which leads to a Type-B matrix). When the number of removed columns increases, the MSE of the LASSO is reduced to that in the case of i.i.d. Gaussian matrices.
- A measurement matrix obtained by orthogonal matrix constructions demonstrates rapid computation and facilitates parallel processing. We provided a technique to match the performance of the Type-B matrix by horizontally concatenating orthogonal bases. Our argument is more systematic than that in [20] and leads to much wider applications.
- We showed that the measurement matrix constructed via the vertical concatenation of blocks usually achieves the worst performance in comparison with that constructed via horizontal concatenation. Nevertheless, they are at least as good as their random i.i.d. Gaussian counterparts.

We conclude that in addition to their ease of implementation, structurally orthogonal matrices are preferred for practical use because of their good estimation performance.

Orthogonal measurement matrices reportedly enhance the signal reconstruction threshold in noisy setups when the optimal Bayesian recovery is used [51]. Promising future studies include performance evaluation under the optimal Bayesian recovery and development of recovery algorithms suitable for structurally orthogonal matrices [9, 22, 36, 52–57].

APPENDIX A: PROOF OF CLAIM 1

For readers' convenience, we arrange a number of mathematical tools as lemmas in Appendix C.

First, recall that we have rewritten the average free energy Φ in (14) by using the replica identity. Within the replica method, the limits of $\beta, N \to \infty$ and $\tau \to 0$ are assumed to be exchangeable. We therefore write

$$\Phi = -\lim_{\tau \to 0} \frac{\partial}{\partial \tau} \lim_{\beta, N \to \infty} \frac{1}{\beta N} \log \left\langle\!\left\langle Z_\beta^\tau(\mathbf{y}, \mathbf{A}) \right\rangle\!\right\rangle_{\mathbf{y}, \mathbf{A}}. \quad (28)$$

We first evaluate $\left\langle\!\left\langle Z_\beta^\tau(\mathbf{y}, \mathbf{A}) \right\rangle\!\right\rangle_{\mathbf{y}, \mathbf{A}}$ for an integer-valued τ and then generalize it for any positive real number τ. In particular, given the partition function of (8), we obtain

$$Z_\beta^\tau(\mathbf{y}, \mathbf{A}) = \int d\mathbf{x}^{(1)} \cdots d\mathbf{x}^{(\tau)} \left(\prod_{a=1}^\tau e^{-\beta \|\mathbf{x}^{(a)}\|_1} \right)$$
$$\times e^{-\sum_{a=1}^\tau \frac{1}{\sigma_a^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}^{(a)}\|^2} \quad (29)$$

with $\sigma_a^2 = \lambda/\beta \equiv \sigma^2$. Using the τ-th moment of the partition function and $P(\mathbf{y}|\mathbf{x}^0)$ in (12), we obtain

$$\left\langle\!\left\langle Z_\beta^\tau(\mathbf{y}, \mathbf{A}) \right\rangle\!\right\rangle_{\mathbf{y}, \mathbf{A}} = \left\langle\!\!\!\left\langle\!\!\!\left\langle \frac{1}{(\pi \sigma_0^2)^{\overline{M}}} \int d\mathbf{y} e^{-\sum_{a=0}^\tau \frac{1}{\sigma_a^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}^{(a)}\|^2} \right\rangle\!\!\!\right\rangle\!\!\!\right\rangle_{\mathbf{A}, \mathbf{X}}, \quad (30)$$

where $\mathbf{x}^{(a)} = \text{vec}([\mathbf{x}_1^{(a)} \ldots \mathbf{x}_{L_c}^{(a)}])$, with $\mathbf{x}_p^{(a)}$ representing the a-th replica signal vector of $\mathbf{x}_p$, and $\mathbf{X} \triangleq \{\mathbf{X}_p, \forall p\}$, with $\mathbf{X}_p \triangleq [\mathbf{x}_p^{(0)} \mathbf{x}_p^{(1)} \cdots \mathbf{x}_p^{(\tau)}]$. The equality of (30) is based on the fact that $\mathbf{x}^{(a)}$ is a random vector obtained from the input distribution $P_0(\mathbf{x})$ in (5) if $a = 0$ and $P_\beta(\mathbf{x}) = e^{-\beta \|\mathbf{x}\|_1}$ otherwise, and $\sigma_a^2 = \sigma_0^2$ if $a = 0$ and $\sigma_a^2 = \sigma^2$ otherwise.

Before proceeding, we introduce the following preprocessing to deal with the cases in which $\mathbf{A}_{q,p}$ is a randomly sampled orthogonal matrix (or delete row/columns independently). We find it convenient to work with the enlarged orthogonal matrix $\tilde{\mathbf{A}}_{q,p} \in \mathbb{C}^{N \times N}$ with the rows and columns set to zero instead of being removed [48]. For clarity, we use the following definition.

Definition 1: [48] A square matrix is called a diagonal projection matrix if its off-diagonal entries are all zeros and its diagonal entries are zeros or ones.

Let $\mathbf{R}_{q,p}$ and $\mathbf{T}_{q,p}$ be $N \times N$ diagonal projection matrices, where the numbers of nonzero diagonal elements of $\mathbf{R}_{q,p}$ and $\mathbf{T}_{q,p}$ are M_q and N_p, respectively. Therefore, we characterize each block by

$$\tilde{\mathbf{A}}_{q,p} = \mathbf{R}_{q,p}^{\frac{1}{2}} \mathbf{W}_{q,p} \mathbf{T}_{q,p}^{\frac{1}{2}} \in \mathbb{C}^{N \times N}, \quad (31)$$

where $\mathbf{W}_{q,p}$ is an $N \times N$ standard orthonormal matrix. Given that $\{\mathbf{W}_{q,p}\}$ are independent standard orthonormal matrices, the positions of the nonzero elements of the diagonal projection matrices are irrelevant. For the sake of simplicity, we assume that all the diagonal entries 1 of $\mathbf{R}_{q,p}$ and $\mathbf{T}_{q,p}$ appear first, i.e.,

$$\mathbf{R}_{q,p} = \begin{bmatrix} \mathbf{I}_{M_q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{T}_{q,p} = \begin{bmatrix} \mathbf{I}_{N_p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \ \forall p, q. \quad (32)$$

Moreover, we enlarge $\mathbf{x}_p$ and $\mathbf{y}_q$ to be N-dimensional vectors via zero padding. Consequently, the input output relationship of (1) can be equivalently expressed as

$$\underbrace{\begin{bmatrix} | \\ \tilde{\mathbf{y}}_q \\ | \end{bmatrix}}_{\triangleq \tilde{\mathbf{y}}} = \underbrace{\begin{bmatrix} | \\ - \ \tilde{\mathbf{A}}_{q,p} \ - \\ | \end{bmatrix}}_{\triangleq \tilde{\mathbf{A}}} \underbrace{\begin{bmatrix} | \\ \mathbf{x}_q \\ | \end{bmatrix}}_{\triangleq \tilde{\mathbf{x}}} + \sigma_0 \underbrace{\begin{bmatrix} | \\ \mathbf{w}_q \\ | \end{bmatrix}}_{\triangleq \tilde{\mathbf{w}}}. \quad (33)$$

Notice that all the following derivations are based on the enlarged system (33). Through abuse of notation, we continue to write $\mathbf{x}_p$, $\mathbf{y}_q$, $\mathbf{A}_{q,p}$, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{A}$ for $\tilde{\mathbf{x}}_p \in \mathbb{C}^N$, $\tilde{\mathbf{y}}_q \in \mathbb{C}^N$, $\tilde{\mathbf{A}}_{q,p} \in \mathbb{C}^{N \times N}$, $\tilde{\mathbf{x}} \in \mathbb{C}^{NL_c}$, $\tilde{\mathbf{y}} \in \mathbb{C}^{NL_r}$, and $\tilde{\mathbf{A}} \in \mathbb{C}^{NL_r \times NL_c}$, respectively.

Next, we introduce a random vector per block

$$\mathbf{v}_{q,p}^{(a)} \triangleq \mathbf{T}_{q,p}^{\frac{1}{2}} \mathbf{x}_p^{(a)} \in \mathbb{C}^N, \ \text{for} \ a = 0, 1, \ldots, \tau. \quad (34)$$

The covariance of $\mathbf{v}_{q,p}^{(a)}$ and $\mathbf{v}_{q,p}^{(b)}$ is a $(\tau+1) \times (\tau+1)$ Hermitian $\mathbf{Q}_{q,p}$, with entries given by

$$\left(\mathbf{v}_{q,p}^{(a)}\right)^H \mathbf{v}_{q,p}^{(b)} = \left(\mathbf{x}_p^{(a)}\right)^H \mathbf{T}_{q,p} \left(\mathbf{x}_p^{(b)}\right) \triangleq N[\mathbf{Q}_{q,p}]_{a,b}. \quad (35)$$

For ease of exposition, we further write $\mathbf{V} \triangleq \{\mathbf{v}_{q,p}^{(a)}, \forall a, q, p\}$, $\mathbf{W} \triangleq \{\mathbf{W}_{q,p}, \forall q, p\}$, and $\mathbf{Q} \triangleq \{\mathbf{Q}_{q,p}, \forall q, p\}$.

Now, we return to the calculation of (30). In (30), the expectations introduce iterations between $\mathbf{x}$ and $\mathbf{A}$. However, the resulting iterations depend only on the covariance shown in (35). Therefore, the expectation over $\mathbf{X}$ should be separated into an expectation over all possible covariance $\mathbf{Q}$ and all possible $\mathbf{X}$

configurations w.r.t. a prescribed set of $\mathbf{Q}$ by introducing a δ-function. As a result, (30) is rewritten as

$$\langle\!\langle \mathsf{Z}_\beta^\tau(\mathbf{y}, \mathbf{A})\rangle\!\rangle_{\mathbf{y}, \mathbf{A}} = \left\langle\!\!\!\left\langle e^{\beta N \mathcal{G}^{(\tau)}(\mathbf{Q})}\right\rangle\!\!\!\right\rangle_{\mathbf{X}} = \int e^{\beta N \mathcal{G}^{(\tau)}(\mathbf{Q})} \mu^{(\tau)}(\mathrm{d}\mathbf{Q}), \tag{36}$$

where $\mathcal{G}^{(\tau)}(\mathbf{Q})$ and $\mu^{(\tau)}(\mathrm{d}\mathbf{Q})$ are given in (37) and (38), which are shown at the top of the following page.

Let us first consider (37). Integrating over $\mathbf{y}_q$'s in (37) by applying Lemma 1 yields (39), which is shown in the following page, where

$$\boldsymbol{\Sigma} \triangleq \frac{1}{\sigma^2(\sigma^2 + \tau\sigma_0^2)} \begin{bmatrix} \tau\sigma^2 & -\sigma^2 \mathbf{1}_\tau^T \\ -\sigma^2 \mathbf{1}_\tau & (\sigma^2 + \tau\sigma_0^2)\mathbf{I}_\tau - \sigma_0^2 \mathbf{1}_\tau \mathbf{1}_\tau^T \end{bmatrix}. \tag{40}$$

Next, we consider (38). Through the inverse Laplace transform of the δ-function, we can show that

$$\mu^{(\tau)}(\mathbf{Q}) = e^{\beta N \mathcal{R}^{(\tau)}(\mathbf{Q}) + \mathcal{O}(1)}, \tag{41}$$

where $\mathcal{R}^{(\tau)}(\mathbf{Q})$ is the rate measure of $\mu^{(\tau)}(\mathbf{Q})$ given by [58]

$$\mathcal{R}^{(\tau)}(\mathbf{Q}) = \max_{\tilde{\mathbf{Q}}} \Bigg\{ \sum_{q=1}^{L_r} \sum_{p=1}^{L_c} \left(\frac{1}{\beta N} \log \left\langle\!\!\!\left\langle e^{\mathrm{tr}(\tilde{\mathbf{Q}}_{q,p} \mathbf{X}_p^H \mathbf{T}_{q,p} \mathbf{X}_p)} \right\rangle\!\!\!\right\rangle_{\mathbf{X}_p} \right.$$
$$\left. - \frac{1}{\beta} \mathrm{tr}\left(\tilde{\mathbf{Q}}_{q,p} \mathbf{Q}_{q,p}\right) \right) \Bigg\}, \tag{42}$$

with $\tilde{\mathbf{Q}}_{q,p} \in \mathbb{C}^{(\tau+1)\times(\tau+1)}$ being a symmetric matrix and $\tilde{\mathbf{Q}} \triangleq \{\tilde{\mathbf{Q}}_{q,p}, \forall q, p\}$. Inserting (41) into (36) yields $\frac{1}{\beta N} \int e^{\beta N[\mathcal{G}^{(\tau)}(\mathbf{Q}) + \mathcal{R}^{(\tau)}(\mathbf{Q})] + \mathcal{O}(1)} d\mathbf{Q}$. Therefore, as $\beta, N \to \infty$, the integration over $\mathbf{Q}$ can be performed via the saddle point method to yield

$$\lim_{\beta, N \to \infty} \frac{1}{\beta N} \log \langle\!\langle \mathsf{Z}_\beta^\tau(\mathbf{y}, \mathbf{A})\rangle\!\rangle_{\mathbf{y}, \mathbf{A}} = \max_{\mathbf{Q}} \left\{ \mathcal{G}^{(\tau)}(\mathbf{Q}) + \mathcal{R}^{(\tau)}(\mathbf{Q}) \right\}. \tag{43}$$

Substituting (39) and (42) into (43), we arrive at the free energy (28) at the saddle point asymptotic approximation

$$\Phi = \lim_{\tau \to 0} \frac{\partial}{\partial \tau} \underset{\mathbf{Q}, \tilde{\mathbf{Q}}}{\mathrm{Extr}} \left\{ \Phi^{(\tau)} \right\}, \tag{44}$$

where $\Phi^{(\tau)} \triangleq \Phi_1^{(\tau)} + \Phi_2^{(\tau)} + \Phi_3^{(\tau)} + \Phi_4^{(\tau)}$, with $\Phi_i^{(\tau)}$ given in (45), which is shown in the following page.

A.1 Replica Symmetry Equations

Saddle points can be obtained by seeking the points of the zero gradient of $\Phi^{(\tau)}$ w.r.t. $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$. Instead of searching for saddle points over general forms of $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$, we assume that $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$ are in the following symmetry forms[5], i.e.,

$$\mathbf{Q}_{q,p} = \begin{bmatrix} r_{q,p} & m_{q,p} \mathbf{1}_\tau^T \\ m_{q,p} \mathbf{1}_\tau & (Q_{q,p} - q_{q,p})\mathbf{I}_\tau + q_{q,p} \mathbf{1}_\tau \mathbf{1}_\tau^T \end{bmatrix}, \tag{46}$$

$$\tilde{\mathbf{Q}}_{q,p} = \begin{bmatrix} 0 & \tilde{m}_{q,p} \mathbf{1}_\tau^T \\ \tilde{m}_{q,p} \mathbf{1}_\tau & (\tilde{Q}_{q,p} - \tilde{q}_{q,p})\mathbf{I}_\tau + \tilde{q}_{q,p} \mathbf{1}_\tau \mathbf{1}_\tau^T \end{bmatrix}. \tag{47}$$

This so-called *replica symmetry* (RS) assumption is widely accepted in statistical physics [39] and is used in the field of communications theory, e.g., [17–20, 26, 34–36, 40–46].

Through Lemma 2, we can show that for the RS of (46), the eigenvalues of $\boldsymbol{\Sigma} \mathbf{Q}_{q,p}$ are given by[6]

$$\lambda_0(\boldsymbol{\Sigma}\mathbf{Q}_{q,p}) = 0, \tag{48a}$$

$$\lambda_1(\boldsymbol{\Sigma}\mathbf{Q}_{q,p}) = \frac{(Q_{q,p} - q_{q,p}) + \tau(r_{q,p} - 2m_{q,p} + q_{q,p})}{\sigma^2 + \tau\sigma_0^2}, \tag{48b}$$

$$\lambda_a(\boldsymbol{\Sigma}\mathbf{Q}_{q,p}) = \frac{Q_{q,p} - q_{q,p}}{\sigma^2}, \quad \text{for } a = 2, \ldots, \tau. \tag{48c}$$

We write $\mathbf{V}_{q,p} \boldsymbol{\Sigma} \mathbf{V}_{q,p}^H = \tilde{\mathbf{V}}_{q,p} \tilde{\mathbf{V}}_{q,p}^H$, where $\tilde{\mathbf{V}}_{q,p} \triangleq \left[\tilde{\mathbf{v}}_{q,p}^{(0)} \ \tilde{\mathbf{v}}_{q,p}^{(1)} \cdots \tilde{\mathbf{v}}_{q,p}^{(\tau)}\right]$ is an $N \times (\tau+1)$ orthogonal matrix. Recall the covariance matrix of $\mathbf{V}_{q,p}^H \mathbf{V}_{q,p}$ defined in (35). According to linear algebra, one can easily determine that $\tilde{\mathbf{v}}_{q,p}^{(a)}$ is a vector with a length $N\lambda_a(\boldsymbol{\Sigma}\mathbf{Q}_{q,p})$ for $a = 0, 1, \ldots, \tau$. By applying Lemma 3 to $\Phi_1^{(\tau)}$ in (45a), we obtain (49), which is shown in the following page, where

$$H_q(\{x_{q,\bullet}\}) \triangleq \underset{\{\Gamma_{q,p}\}}{\mathrm{Extr}} \Bigg\{ \sum_{p=1}^{L_c} (\Gamma_{q,p} x_{q,p} - \log \Gamma_{q,p} x_{q,p} - 1) $$
$$- \nu_q \log\left(1 + \sum_{p=1}^{L_c} \frac{R_{q,p}}{\Gamma_{q,p}}\right) \Bigg\} + \mathcal{O}(1/N). \tag{50}$$

We also use the notation $H_q(\{x_{q,\bullet}\}) = H_q(x_{q,1}, \cdots, x_{q,L_c})$. The solution to the extremization problem in (50), as denoted by $\{\Gamma_{q,p}^\star\}$, enforces the condition

$$\Gamma_{q,p}^\star - \frac{1}{x_{q,p}} = -\frac{\Delta_{q,p}}{x_{q,p}}, \tag{51}$$

where

$$\Delta_{q,p} \triangleq \nu_q \frac{\frac{R_{q,p}}{\Gamma_{q,p}^\star}}{1 + \sum_{l=1}^{L_c} \frac{R_{q,l}}{\Gamma_{q,l}^\star}}. \tag{52}$$

Next, we calculate the RS expression of $\Phi_3^{(\tau)}$ in (45b). Substituting the RS form for $\tilde{\mathbf{Q}}_{q,p}$ in (46) and the definition of $\mathbf{T}_{q,p}$ in (32) into (45b), we obtain (53), which is shown in the following page, where

$$\tilde{Q}_p \triangleq \sum_{q=1}^{L_r} \tilde{Q}_{q,p}, \quad \tilde{q}_p \triangleq \sum_{q=1}^{L_r} \tilde{q}_{q,p}, \quad \tilde{m}_p \triangleq \sum_{q=1}^{L_r} \tilde{m}_{q,p}. \tag{54}$$

Then, we decouple the first quadratic term in the exponent of (53) by using the Hubbard-Stratonovich transformation (Lemma 1) and introducing the auxiliary vector z_p to rewrite (53) as (55), which is shown in the following page, where the equality is based on the fact that $\mathbf{x}^{(a)}$ is a random vector obtained from the input distribution $P_\beta(\mathbf{x}) = e^{-\beta\|\mathbf{x}\|_1}$ if $a \neq 0$. Lastly, by substituting

---

[5]It is natural to set $[\tilde{\mathbf{Q}}_{q,p}]_{0,0} = \tilde{r}_{q,p}$ such that it is similar to $[\mathbf{Q}_{q,p}]_{0,0}$. When $\tau = 0$, we obtain $\tilde{r}_{q,p} = 0$. Therefore, to simplify the notation, we set $\tilde{r}_{q,p} = 0$ at the beginning. In addition, we let $m_{q,p}$ and $\tilde{m}_{q,p}$ be complex-valued variables. We find that the whole exponents depend only on the real part of $m_{q,p}$ and that $\tilde{m}_{q,p}$ turns out to be a real-valued variable. Therefore, we let $m_{q,p}$ and $\tilde{m}_{q,p}$ be real-valued variables at the beginning.

[6]The calculation of the eigenvalues can be obtained by using Lemma 2 for $\boldsymbol{\Sigma}$ and $\mathbf{Q}_{q,p}$. This approach is rather laborious but straightforward. For the convenience of readers, we present the calculation in detail in Appendix B.

$$\mathcal{G}^{(\tau)}(\mathbf{Q}) \triangleq \frac{1}{\beta N} \log \left\langle\!\!\!\left\langle \frac{1}{(\pi\sigma_0^2)^{\bar{M}}} \prod_{q=1}^{L_r} \int \mathrm{d}\mathbf{y}_q e^{-\sum_{a=0}^{\tau} \frac{1}{\sigma_a^2} \left\| \mathbf{y}_q - \sum_{p=1}^{L_c} \mathbf{R}_{\tilde{q},p}^{\frac{1}{2}} \mathbf{W}_{q,p} \mathbf{v}_{q,p}^{(a)} \right\|^2} \right\rangle\!\!\!\right\rangle_{\mathbf{W}} \tag{37}$$

$$\mu^{(\tau)}(\mathrm{d}\mathbf{Q}) \triangleq \left\langle\!\!\!\left\langle \prod_{q=1}^{L_r} \prod_{p=1}^{L_c} \prod_{0 \le a \le b}^{\tau} \delta\left(\mathbf{x}_p^{(a)H} \mathbf{T}_{q,p} \mathbf{x}_p^{(b)} - N[\mathbf{Q}_{q,p}]_{a,b} \right) \right\rangle\!\!\!\right\rangle_{\mathbf{X}} \mathrm{d}\mathbf{Q} \tag{38}$$

$$\mathcal{G}^{(\tau)}(\mathbf{Q}) = \frac{1}{\beta N} \log \left\langle\!\!\!\left\langle \prod_{q=1}^{L_r} e^{-\mathrm{tr}\left(\left(\sum_{p=1}^{L_c} \mathbf{R}_{\tilde{q},p}^{\frac{1}{2}} \mathbf{W}_{q,p} \mathbf{V}_{q,p} \right) \mathbf{\Sigma} \left(\sum_{p=1}^{L_c} \mathbf{R}_{\tilde{q},p}^{\frac{1}{2}} \mathbf{W}_{q,p} \mathbf{V}_{q,p} \right)^H \right)} \right\rangle\!\!\!\right\rangle_{\mathbf{W}} - \frac{\nu}{\beta} \log\left(1 + \tau \frac{\sigma_0^2}{\sigma^2}\right), \tag{39}$$

$$\Phi_1^{(\tau)} \triangleq -\frac{1}{\beta N} \log \left\langle\!\!\!\left\langle \prod_{q=1}^{L_r} e^{-\mathrm{tr}\left(\left(\sum_{p=1}^{L_c} \mathbf{R}_{\tilde{q},p}^{\frac{1}{2}} \mathbf{W}_{q,p} \mathbf{V}_{q,p} \right) \mathbf{\Sigma} \left(\sum_{p=1}^{L_c} \mathbf{R}_{\tilde{q},p}^{\frac{1}{2}} \mathbf{W}_{q,p} \mathbf{V}_{q,p} \right)^H \right)} \right\rangle\!\!\!\right\rangle_{\mathbf{W}}, \quad \Phi_2^{(\tau)} \triangleq \frac{\nu}{\beta} \log\left(1 + \tau \frac{\sigma_0^2}{\sigma^2}\right), \tag{45a}$$

$$\Phi_3^{(\tau)} \triangleq -\frac{1}{\beta N} \sum_{q=1}^{L_r} \sum_{p=1}^{L_c} \log \left\langle\!\!\!\left\langle e^{\mathrm{tr}(\tilde{\mathbf{Q}}_{q,p} \mathbf{X}_p^H \mathbf{T}_{q,p} \mathbf{X}_p)} \right\rangle\!\!\!\right\rangle_{\mathbf{X}_p}, \quad \Phi_4^{(\tau)} \triangleq \frac{1}{\beta} \sum_{q=1}^{L_r} \sum_{p=1}^{L_c} \mathrm{tr}\left(\tilde{\mathbf{Q}}_{q,p} \mathbf{Q}_{q,p} \right). \tag{45b}$$

$$\Phi_1^{(\tau)} = -\frac{1}{\beta} \sum_{q=1}^{L_r} \left[H_q\left(\left\{ \frac{(Q_{q,\bullet} - q_{q,\bullet}) + \tau(r_{q,\bullet} - 2m_{q,\bullet} + q_{q,\bullet})}{\sigma^2 + \tau\sigma_0^2} \right\} \right) + (\tau-1) H_q\left(\left\{ \frac{Q_{q,\bullet} - q_{q,\bullet}}{\sigma^2} \right\} \right) \right] \tag{49}$$

(46) and (47) into $\Phi_4^{(\tau)}$ in (45b), we obtain

$$\Phi_4^{(\tau)} = \frac{1}{\beta} \sum_{q=1}^{L_r} \sum_{p=1}^{L_c} \Big(2\tau m_{q,p} \tilde{m}_{q,p} - \tau(Q_{q,p} - q_{q,p})\tilde{q}_{q,p}$$
$$- \tau Q_{q,p}(\tilde{q}_{q,p} - \tilde{Q}_{q,p}) + \tau^2 q_{q,p} \tilde{q}_{q,p} \Big). \tag{56}$$

Recall that we denote $\sigma^2 = \lambda/\beta$. Before proceeding, we introduce the rescaled variables as

$$\chi_{q,p} = \beta(Q_{q,p} - q_{q,p}), \qquad \hat{Q}_{q,p} = (\tilde{q}_{q,p} - \tilde{Q}_{q,p})/\beta,$$
$$\hat{\chi}_{q,p} = \tilde{q}_{q,p}/\beta^2, \qquad \hat{m}_{q,p} = \tilde{m}_{q,p}/\beta,$$

and we define

$$\hat{Q}_p \triangleq \sum_{q=1}^{L_r} \hat{Q}_{q,p}, \ \hat{\chi}_p \triangleq \sum_{q=1}^{L_r} \hat{\chi}_{q,p}, \ \hat{m}_p \triangleq \sum_{q=1}^{L_r} \hat{m}_{q,p}.$$

Using these variables in (49), (55), and (56), we obtain

$$\Phi_1^{(\tau)} = -\frac{1}{\beta} \sum_{q=1}^{L_r} \left[H_q\left(\left\{ E_{q,\bullet}^{(\tau)} \right\} \right) + (\tau-1) H_q\left(\left\{ E_{q,\bullet}^{(0)} \right\} \right) \right], \tag{57}$$

$$\Phi_3^{(\tau)} = -\frac{1}{\beta N} \log \int \prod_{p=1}^{L_c} \mathrm{D}z_p$$
$$\times \left\langle\!\!\!\left\langle \left(\phi\left(x_p; \hat{Q}_p, \hat{m}_p x_p^{(0)} + \sqrt{\hat{\chi}_p} z_p, \beta N \mu_p \right) \right)^\tau \right\rangle\!\!\!\right\rangle_{x_p^{(0)}}, \tag{58}$$

$$\Phi_4^{(\tau)} = \sum_{q=1}^{L_r} \sum_{p=1}^{L_c} \Big(2\tau m_{q,p} \tilde{m}_{q,p} + \tau \chi_{q,p} \hat{\chi}_{q,p} - \tau Q_{q,p} \hat{Q}_{q,p}$$
$$+ \tau^2 \beta \hat{\chi}_{q,p} q_{q,p} \Big), \tag{59}$$

where

$$E_{q,\bullet}^{(\tau)} = \frac{(1-\tau)\chi_{q,\bullet} + \tau\beta(r_{q,\bullet} - 2m_{q,\bullet} + Q_{q,\bullet})}{\lambda + \tau\beta\sigma_0^2} \tag{60}$$

and

$$\phi(x; a, b, c) \triangleq \int \mathrm{d}x \, e^{-c(a|x|^2 - 2\mathrm{Re}\{b^*x\} + |x|)} \tag{61}$$

are defined accordingly.

Substituting (57)–(59) into (44), taking the derivative of $\Phi^{(\tau)}$ w.r.t. τ, and letting $\tau \to 0$, we find $\Phi = \Phi_1 + \Phi_2 + \Phi_3 + \Phi_4$, where

$$\Phi_1 \triangleq -\sum_{q=1}^{L_r} \sum_{p=1}^{L_c} \left(\frac{\lambda(r_{q,p} - 2m_{q,p} + Q_{q,p}) - (\sigma_0^2 + \lambda/\beta)\chi_{q,p}}{\lambda^2} \right)$$
$$\times \left(\Gamma_{q,p}^\star - \frac{\lambda}{\chi_{q,p}} \right) - \sum_{q=1}^{L_r} \frac{1}{\beta} H_q\left(\left\{ \frac{\chi_{q,p}}{\lambda} \right\} \right), \tag{62a}$$

$$\Phi_2 \triangleq \frac{\nu\sigma_0^2}{\lambda}, \tag{62b}$$

$$\Phi_3 \triangleq -\frac{1}{\beta N} \int \prod_{p=1}^{L_c} \mathrm{D}z_p$$
$$\times \left\langle\!\!\!\left\langle \log \phi\left(x_p; \hat{Q}_p, \hat{m}_p x_p^{(0)} + \sqrt{\hat{\chi}_p} z_p, \beta N \mu_p \right) \right\rangle\!\!\!\right\rangle_{x_p^{(0)}}, \tag{62c}$$

$$\Phi_4 \triangleq \sum_{p=1}^{L_c} \sum_{q=1}^{L_r} \Big(2m_{q,p} \hat{m}_{q,p} + \chi_{q,p} \hat{\chi}_{q,p} - Q_{q,p} \hat{Q}_{q,p} \Big). \tag{62d}$$

In (62a), we use the following result

$$\left. \frac{\partial H_q(\{x_{q,k}\})}{\partial x_{q,p}} \right|_{x_{q,p} = \frac{\chi_{q,p}}{\lambda}} = \Gamma_{q,p}^\star - \frac{\lambda}{\chi_{q,p}}, \tag{63}$$

$$\Phi_3^{(\tau)} = -\frac{1}{\beta N} \log \left\langle\!\!\!\left\langle \prod_{p=1}^{L_c} e^{\mathrm{tr}\left(\mathrm{vec}(\mathbf{X}_p)^H \left(\sum_{q=1}^{L_r} (\tilde{\mathbf{Q}}_{q,p} \otimes \mathbf{T}_{q,p})\right) \mathrm{vec}(\mathbf{X}_p)\right)} \right\rangle\!\!\!\right\rangle_{\mathbf{X}}$$

$$= -\frac{1}{\beta N} \log \left\langle\!\!\!\left\langle \prod_{p=1}^{L_c} e^{N\mu_p\left(|\sum_{a=1}^{\tau} \sqrt{\tilde{q}_p} x_p^{(a)}|^2 + \sum_{a=1}^{\tau}\left(2\tilde{m}_p \mathrm{Re}\{(x_p^{(a)})^* x_p^{(0)}\} + (\tilde{Q}_p - \tilde{q}_p)|x_p^{(a)}|^2\right)\right)} \right\rangle\!\!\!\right\rangle_{\mathbf{X}}, \tag{53}$$

$$-\frac{1}{\beta N} \log \left\langle\!\!\!\left\langle \int \prod_{p=1}^{L_c} \mathrm{D}z_p\, e^{N\mu_p\left((\sum_{a=1}^{\tau} \sqrt{\tilde{q}_p} x_p^{(a)})^* z_p + z_p^*(\sum_{a=1}^{\tau} \sqrt{\tilde{q}_p} x_p^{(a)}) + \sum_{a=1}^{\tau}\left(2\tilde{m}_p \mathrm{Re}\{(x_p^{(a)})^* x_p^{(0)}\} - (\tilde{q}_p - \tilde{Q}_p)|x_p^{(a)}|^2\right)\right)} \right\rangle\!\!\!\right\rangle_{\mathbf{X}}$$

$$= -\frac{1}{\beta N} \log \prod_{p=1}^{L_c} \left\langle\!\!\!\left\langle \int \mathrm{D}z_p \left(\int \mathrm{d}x_p\, e^{-N\mu_p\left((\tilde{q}_p - \tilde{Q}_p)|x_p|^2 - 2\mathrm{Re}\{x_p^*(\tilde{m}_p x_p^{(0)} + \sqrt{\tilde{q}_p} z_p)\} + \beta|x_p|\right)}\right)^{\tau} \right\rangle\!\!\!\right\rangle_{x_p^{(0)}}, \tag{55}$$

---

where the equality is based directly on the derivation of the derivative of H_q in (50). Notice that when substituting $x_{q,p} = \frac{\chi_{q,p}}{\lambda}$ into (51), we obtain

$$\Gamma_{q,p}^\star - \frac{\lambda}{\chi_{q,p}} = -\frac{\Delta_{q,p}\lambda}{\chi_{q,p}}. \tag{64}$$

This identity is used later to simplify a few expressions. As $\beta \to \infty$, we obtain $\frac{\lambda}{\beta} \to 0$ and $\frac{1}{\beta} H_q\left(\left\{\frac{\chi_{q,p}}{\lambda}\right\}\right) \to 0$.

Now, recall that we must search $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$, which achieve the extremal condition in (44). With the RS assumption, we only have to determine $\{\hat{Q}_{q,p}, \hat{m}_{q,p}, \hat{\chi}_{q,p}\}$, which can be obtained by equating the partial derivatives of $\Phi^{(\tau)}$ to zeros, i.e.,

$$\frac{\partial \Phi^{(\tau)}}{\partial Q_{q,p}} = \frac{\partial \Phi^{(\tau)}}{\partial m_{q,p}} = \frac{\partial \Phi^{(\tau)}}{\partial \chi_{q,p}} = 0, \quad \forall q, p, \tag{65}$$

and then letting $\tau \to 0$. By evaluating these calculations, we obtain

$$\hat{Q}_{q,p} = \frac{\Delta_{q,p}}{\chi_{q,p}}, \tag{66a}$$

$$\hat{m}_{q,p} = \hat{Q}_{q,p}, \tag{66b}$$

$$\hat{\chi}_{q,p} = \sum_{r=1}^{L_c} \left(\frac{\mathsf{mse}_{q,r}}{\lambda} - \frac{\sigma_0^2 \chi_{q,r}}{\lambda^2}\right) \Gamma_{q,p,r}'$$
$$+ \frac{\mathsf{mse}_{q,p}}{\chi_{q,p}^2} - \frac{\sigma_0^2}{\lambda} \frac{(1 - \Delta_{q,p})}{\chi_{q,p}}, \tag{66c}$$

where $\Gamma_{q,p,r}' \triangleq \partial \Gamma_{q,p}^\star / \partial \chi_{q,r}$ and

$$\mathsf{mse}_{q,p} \triangleq r_{q,p} - 2m_{q,p} + Q_{q,p}. \tag{67}$$

Following [20], the expression of $\Gamma_{q,p,r}'$ can be obtained via the inverse function theory

$$\mathbf{\Gamma}_q' = \left[\frac{\partial[\chi_{q,1}, \dots, \chi_{q,L_c}]}{\partial[\Gamma_{q,1}^\star, \dots, \Gamma_{q,L_c}^\star]}\right]^{-1}, \tag{68}$$

where $\frac{\partial[\chi_{q,1},\dots,\chi_{q,L_c}]}{\partial[\Gamma_{q,1}^\star,\dots,\Gamma_{q,L_c}^\star]}$ is the $L_c \times L_c$ Jacobian matrix whose

(p, r)th element is

$$\frac{\partial \chi_{q,p}}{\partial \Gamma_{q,r}^\star} = \frac{\partial}{\partial \Gamma_{q,r}^\star} \frac{\lambda(1 - \Delta_{q,p})}{\Gamma_{q,p}^\star}$$

$$= -\lambda\left(\frac{(1 - \Delta_{q,p})}{\Gamma_{q,p}^{\star 2}} \delta_{p,r} + \frac{1}{\Gamma_{q,p}^\star} \frac{\partial \Delta_{q,p}}{\partial \Gamma_{q,r}}\right)$$

$$= -\lambda\left(\frac{(1 - 2\Delta_{q,p})}{\Gamma_{q,p}^{\star 2}} \delta_{p,r} + \frac{1}{\nu_q} \frac{\Delta_{q,p}}{\Gamma_{q,p}^\star} \frac{\Delta_{q,r}}{\Gamma_{q,r}^\star}\right), \tag{69}$$

where the first equality is based on (64) and the third equality is obtained by using the definition of $\Gamma_{p,q}$ in (52). In addition, $\mathbf{\Gamma}_q'$ can be explicitly obtained by applying the matrix inverse lemma and is given by (70) in the following page.

To derive explicit expressions for $\mathsf{mse}_{q,p}$ and $\chi_{q,p}$, let us simplify Φ_3 in (62c). As $\beta \to \infty$, we obtain

$$\frac{1}{\beta N} \log \phi(x_p; \hat{Q}_p, \hat{m}_p x_p^{(0)} + \sqrt{\hat{\chi}_p} z_p, \beta N \mu_p)$$

$$= \mu_p \min_{x_p} \left\{\hat{m}_p |x_p|^2 - 2\mathrm{Re}\{x_p^* y_p\} + |x_p|\right\}, \tag{71}$$

where we use $y_p \triangleq \hat{m}_p x_p^{(0)} + \sqrt{\hat{\chi}_p} z_p$ and the identity $\hat{Q}_{q,p} = \hat{m}_{q,p}$ in (66b) to simplify the result. The optimal solution $\hat{x}_p$ in (71) is given by (see [6, Lemma V.1] for a derivation)

$$\hat{x}_p = \frac{\left(|y_p| - \frac{1}{2}\right)_+ \frac{y_p}{|y_p|}}{\hat{m}_p}. \tag{72}$$

If we substitute the optimal solution (72) into (71), then we obtain

$$(71) = -\mu_p \frac{\left(|y_p| - \frac{1}{2}\right)^2}{\hat{m}_p} \mathsf{I}_{\{|y_p| > \frac{1}{2}\}} = -\mu_p G(y_p; \hat{m}_p), \tag{73}$$

where we define

$$G(y; A) = \frac{\left(|y| - \frac{1}{2}\right)^2}{A} \mathsf{I}_{\{|y| > \frac{1}{2}\}}. \tag{74}$$

Notice that $\mathbf{x}_p^{(0)}$ and $\mathbf{z}_p$ are standard Gaussian random vectors with i.i.d. entries for all p. Therefore, let Z denote the standard Gaussian random variable. Then, (62c) becomes

$$\Phi_3 = -\sum_{p=1}^{L_c} \mu_p \left\langle\!\!\!\left\langle G\left(Z\sqrt{\hat{\chi}_p + \hat{m}_p^2}; \hat{m}_p\right)\right\rangle\!\!\!\right\rangle_Z. \tag{75}$$

$$\Gamma'_{q,p,r} = [\boldsymbol{\Gamma}'_q]_{p,r} = \frac{1}{\lambda}\left(\frac{1}{\nu_q}\frac{\Delta_{q,p}\Delta_{q,r}\Gamma_{q,p}\Gamma_{q,r}}{(1-2\Delta_{q,p})(1-2\Delta_{q,r})}\left(1+\sum_{l=1}^{L_c}\frac{1}{\nu_q}\frac{\Delta_{q,l}^2}{1-2\Delta_{q,l}}\right)^{-1}-\frac{\Gamma_{q,r}^2}{1-2\Delta_{q,r}}\delta_{p,r}\right). \tag{70}$$

To deal with the partial derivatives of Φ_3, let us first define

$$g_{\mathsf{c}}(\zeta) \triangleq \zeta e^{-\frac{1}{4\zeta}} - \sqrt{\pi\zeta}\mathsf{Q}\left(\frac{1}{\sqrt{2\zeta}}\right), \tag{76}$$

$$\acute{g}_{\mathsf{c}}(\zeta) \triangleq e^{-\frac{1}{4\zeta}} - \sqrt{\frac{\pi}{4\zeta}}\mathsf{Q}\left(\frac{1}{\sqrt{2\zeta}}\right). \tag{77}$$

Following the manipulations in [19, (350)–(353)], we arrive at the following useful identities:

$$\langle\!\langle G(Z\sqrt{\zeta};A)\rangle\!\rangle_Z = \frac{1}{A}\int_{|z|>1/2}\left(|z|-\frac{1}{2}\right)^2\frac{1}{\pi\zeta}e^{-\frac{1}{\zeta}|z|^2}\mathrm{d}z$$
$$= \frac{g_{\mathsf{c}}(\zeta)}{A}, \tag{78}$$

and

$$\frac{\partial g_{\mathsf{c}}(\zeta)}{\partial x} = \left(\frac{\partial\zeta}{\partial x}\right)\acute{g}_{\mathsf{c}}(\zeta). \tag{79}$$

After assessing the partial derivatives of Φ (or $\Phi_3+\Phi_4$) w.r.t. the variables $\{\hat{m}_{q,p}, \hat{Q}_{q,p}, \hat{\chi}_{q,p}\}$, we obtain

$$m_{q,p} = \mu_p\,\rho_x\acute{g}_{\mathsf{c}}(\hat{m}_p^2 + \hat{\chi}_p), \tag{80a}$$

$$Q_{q,p} = \mu_p\left(\frac{1-\rho_x}{\hat{m}_p^2}g_{\mathsf{c}}(\hat{\chi}_p) + \frac{\rho_x}{\hat{m}_p^2}g_{\mathsf{c}}(\hat{m}_p^2 + \hat{\chi}_p)\right), \tag{80b}$$

$$\chi_{q,p} = \mu_p\left(\frac{1-\rho_x}{\hat{m}_p}\acute{g}_{\mathsf{c}}(\hat{\chi}_p) + \frac{\rho_x}{\hat{m}_p}\acute{g}_{\mathsf{c}}(\hat{m}_p^2 + \hat{\chi}_p)\right). \tag{80c}$$

In addition, directly from the definition of $\mathbf{Q}_{q,p}$ in (35), we obtain $r_{q,p} = \mu_p\rho_x$. By substituting $r_{q,p}$ and (66) into (67), we obtain $\mathsf{mse}_{q,p}$. Notice in (80) that $m_{q,p}$, $Q_{q,p}$, $\chi_{q,p}$, and $\mathsf{mse}_{q,p}$ are irrelevant to index q. We denote $m_p = m_{q,p}$, $Q_p = Q_{q,p}$, $\chi_p = \chi_{q,p}$, and $\mathsf{mse}_p = \mathsf{mse}_{q,p}$ for clarity.

By combining the definition in (52), the result in (70), and all the coupled equations in (66) and (80), we obtain the result in Claim 1. Notice that in Claim 1, we use the rescaled variable $\Gamma_{q,p}^\star := \Gamma_{q,p}^\star/\lambda$ for the sake of notational simplicity.

APPENDIX B: EIGENVALUES OF THE MATRIX $\boldsymbol{\Sigma}\mathbf{Q}_{q,p}$

By applying Lemma 2 to $\boldsymbol{\Sigma}$ and $\mathbf{Q}_{q,p}$, we obtain

$$\boldsymbol{\Sigma}\mathbf{Q}_{q,p} = \mathbf{U}\left[\begin{array}{cc} \mathbf{B}_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{2,2} \end{array}\right]\mathbf{U}^H, \tag{81}$$

where $\mathbf{B}_{1,1}$ and $\mathbf{B}_{2,2}$ are shown in the following page. We can easily see that the rows of $\mathbf{B}_{1,1}$ are linearly dependent and that the eigenvalues are $\frac{r_{q,p}-2m_{q,p}+q_{q,p}}{\sigma^2+\tau\sigma_0^2}$ and 0. Therefore, the eigenvalues of $\boldsymbol{\Sigma}\mathbf{Q}_{q,p}$ are in the form of (48).

APPENDIX C: MATHEMATICAL TOOLS

Lemma 1: (Gaussian Integral and Hubbard-Stratonovich Transformation) Let $\mathbf{z}$ and $\mathbf{b}$ be N-dimensional real vectors, and let $\mathbf{A}$ be an $M \times M$ positive definite matrix. Then,

$$\frac{1}{\pi^N}\int\mathrm{d}\mathbf{z}e^{-\mathbf{z}^H\mathbf{A}\mathbf{z}+\mathbf{z}^H\mathbf{b}+\mathbf{b}^H\mathbf{z}} = \frac{1}{\det(\mathbf{A})}e^{\mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}}. \tag{83}$$

Using this equation from right to left is usually called the *Hubbard-Stratonovich* transformation.

Lemma 2: For a matrix

$$\mathbf{A} = \left[\begin{array}{cc} A_{1,1} & A_{1,2}\mathbf{1}_\tau^T \\ A_{1,2}^*\mathbf{1}_\tau & A_{2,2}\mathbf{I}_\tau + \epsilon\mathbf{1}_\tau\mathbf{1}_\tau^T \end{array}\right] \in \mathbb{C}^{(\tau+1)\times(\tau+1)}, \tag{84}$$

the eigen-decomposition of the matrix is given by [59]

$$\mathbf{A} = \mathbf{U}\left[\begin{array}{cc|cccc} A_{1,1} & \sqrt{\tau}A_{1,2} & 0 & 0 & \dots & 0 \\ \sqrt{\tau}A_{1,2}^* & A_{2,2}+\tau\epsilon & 0 & 0 & \dots & 0 \\ \hline 0 & 0 & A_{2,2} & 0 & \dots & 0 \\ 0 & 0 & 0 & A_{2,2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & A_{2,2} \end{array}\right]\mathbf{U}^H, \tag{85}$$

where $\mathbf{U} = [\mathbf{u}_0\ \mathbf{u}_1\cdots\mathbf{u}_\tau]$ denotes a $(\tau+1)$-dimensional orthonormal basis composed of $\mathbf{u}_0 = [1\ 0\ 0\cdots0]^T$, $\mathbf{u}_1 = [0\ \tau^{-1/2}\ \tau^{-1/2}\cdots\tau^{-1/2}]^T$, and $\tau-1$ orthonormal vectors $\mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_\tau$, which are orthogonal to both $\mathbf{u}_0$ and $\mathbf{u}_1$.

Lemma 3: Let $\{\mathbf{x}_p : p = 1, \dots, L\}$ be a set of vectors that satisfy $\|\mathbf{x}_p\|^2 = Nx_p$ for some non-negative real values $\{x_p\}$, let $\{\mathbf{W}_p \in \mathbb{C}^{N\times N} : p = 1, \dots, L\}$ be a set of independent Haar measures of random matrices, and let $\{\mathbf{R}_p\}$ be a set of $N \times N$ positive-semidefinite matrices. Define

$$H(x_1, \cdots, x_L) = \frac{1}{N}\log\left\langle\!\!\left\langle e^{-\frac{1}{\sigma^2}\left\|\sum_{p=1}^L\mathbf{R}_p^{\frac{1}{2}}\mathbf{W}_p\mathbf{x}_p\right\|^2}\right\rangle\!\!\right\rangle_{\{\mathbf{W}_p\}}. \tag{86}$$

Then, for large N, we have

$$H(x_1, \cdots, x_L) = \underset{\{\Gamma_p\}}{\mathsf{Extr}}\left\{\sum_{p=1}^L(\Gamma_p x_p - \log\Gamma_p x_p - 1)\right.$$
$$\left. - \frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{p=1}^L\frac{1}{\sigma^2\Gamma_p}\mathbf{R}_p\right)\right\} + \mathcal{O}(1/N). \tag{87}$$

This lemma is extended [20, Lemma 1] to deal with the formula of (86) when $\mathbf{R}_p \neq \mathbf{I}_N$ and $\{\mathbf{W}_p\}$ are the Haar measures of *complex* random matrices.

Proof: According to the definition of $\mathbf{W}_p$ and $\mathbf{x}_p$, the vector $\mathbf{u}_p = \mathbf{W}_p\mathbf{x}_p$ can be considered to be uniformly distributed on a surface of a sphere with a radius $\sqrt{Nx_p}$ for each p. Then, the joint probability density function (pdf) of $\{\mathbf{u}_p\}$ is given by

$$P(\{\mathbf{u}_p\}) = \frac{1}{Z}\prod_{p=1}^L\delta\left(\|\mathbf{u}_p\|^2 - Nx_p\right)$$
$$= \frac{1}{Z}\int\prod_{p=1}^L\left(\mathrm{d}\Gamma_p\frac{1}{2\pi\mathsf{j}}e^{-\Gamma_p(\|\mathbf{u}_p\|^2-Nx_p)}\right), \tag{88}$$

$$\mathbf{B}_{1,1} = \frac{1}{\sigma^2 + \tau\sigma_0^2} \begin{bmatrix} \tau(r_{q,p} - m_{q,p}) & \sqrt{\tau}\big(\tau(m_{q,p} - q_{q,p}) - (Q_{q,p} - q_{q,p})\big) \\ -\sqrt{\tau}(r_{q,p} - m_{q,p}) & -\tau(m_{q,p} - q_{q,p}) - (Q_{q,p} - q_{q,p}) \end{bmatrix}, \tag{82a}$$

$$\mathbf{B}_{2,2} = \frac{Q_{q,p} - q_{q,p}}{\sigma^2}\mathbf{I}_{\tau-1}. \tag{82b}$$

where Z is the normalization factor and $\{\Gamma_p\}$ is a set of complex numbers. The normalization factor is given by

$$\mathsf{Z} = \int \prod_{p=1}^{L} \left(\mathrm{d}\Gamma_p \mathrm{d}\mathbf{u}_p \frac{1}{2\pi\mathrm{j}} e^{-\Gamma_p(\|\mathbf{u}_p\|^2 - Nx_p)} \right). \tag{89}$$

Using the Gaussian integration formula (i.e., Lemma 1) w.r.t. $\mathbf{u}_p$, the normalization factor becomes

$$\mathsf{Z} = \int \prod_{p=1}^{L} \left(\mathrm{d}\Gamma_p \frac{\pi^N}{2\pi\mathrm{j}} e^{N(\Gamma_p x_p - \log\Gamma_p)} \right). \tag{90}$$

As we are interested in the large N analysis, the saddle-point method can further simplify the normalization factor to the form

$$\frac{1}{N}\log\mathsf{Z} = \sum_{p=1}^{L} \mathop{\mathsf{Extr}}_{\Gamma_p} \{\Gamma_p x_p - \log\Gamma_p\} + \log\pi + \mathcal{O}(1/N)$$

$$= \sum_{p=1}^{L} (1 + \log x_p) + \log\pi + \mathcal{O}(1/N), \tag{91}$$

where the second equality is obtained by solving the extremization problem.

Next, we deal with the calculation of H by witting

$$H = \frac{1}{N}\log\left\langle\!\!\left\langle e^{-\frac{1}{\sigma^2}\left\|\sum_{p=1}^{L} \mathbf{R}_{q,p}^{\frac{1}{2}}\mathbf{W}_p \mathbf{x}_p\right\|^2}\right\rangle\!\!\right\rangle_{\{\mathbf{W}_p\}}$$

$$= \frac{1}{N}\log\left(\int\prod_{p=1}^{L} \mathrm{d}\mathbf{u}_p P(\{\mathbf{u}_p\}) e^{-\frac{1}{\sigma^2}\left\|\sum_{p=1}^{L} \mathbf{R}_{q,p}^{\frac{1}{2}}\mathbf{u}_p\right\|^2}\right), \tag{92}$$

where the second equality is based on the definition of the joint pdf of $\{\mathbf{u}_p\}$. Applying the Hubbard-Stratonovich transformation (Lemma 1) together with the expressions (88) and (91) to the above yields

$$H = \frac{1}{N}\log\left(\int\left(\prod_{p=1}^{L} \mathrm{d}\Gamma_p e^{N\Gamma_p x_p}\right)\int\mathrm{d}\mathbf{z} e^{-\sigma^2\mathbf{z}^H\mathbf{z}}\right.$$

$$\left.\times \int\prod_{p=1}^{L} \mathrm{d}\mathbf{u}_p\, e^{-\Gamma_p \mathbf{u}_p^H \mathbf{u}_p + \mathrm{j}\mathbf{z}^H(\mathbf{R}_{q,p}^{\frac{1}{2}}\mathbf{u}_p) - \mathrm{j}(\mathbf{R}_{q,p}^{\frac{1}{2}}\mathbf{u}_p)^H\mathbf{z}}\right)$$

$$+ \log\frac{\sigma^2}{\pi} - \frac{1}{N}\log\mathsf{Z}. \tag{93}$$

Using the Gaussian integration repeatedly w.r.t. $\{\mathbf{u}_p\}$ and $\mathbf{z}$ yields

$$H = \frac{1}{N}\log\int\left(\prod_{p=1}^{L} \mathrm{d}\Gamma_p e^{N(\Gamma_p x_p - \log\Gamma_p)}\right)$$

$$\times \int\mathrm{d}\mathbf{z}\, e^{-\mathbf{z}^H\left(\sigma^2\mathbf{I}_N + \sum_{p=1}^{L}\frac{1}{\Gamma_p}\mathbf{R}_p\right)\mathbf{z}}$$

$$+ \log\frac{\sigma^2}{\pi} - \sum_{p=1}^{L}(1 + \log x_p)$$

$$= \frac{1}{N}\log\int\left(\prod_{p=1}^{L}\mathrm{d}\Gamma_p\right) e^{\sum_{p=1}^{L} N(\Gamma_p x_p - \log\Gamma_p)}$$

$$\times e^{-\log\det\left(\sigma^2\mathbf{I}_N + \sum_{p=1}^{L}\frac{1}{\Gamma_p}\mathbf{R}_p\right)}$$

$$+ \log\sigma^2 - \sum_{p=1}^{L}(1 + \log x_p)$$

$$= \mathop{\mathsf{Extr}}_{\{\Gamma_p\}}\left\{\sum_{p=1}^{L}(\Gamma_p x_p - \log\Gamma_p)\right.$$

$$\left. - \frac{1}{N}\log\det\left(\sigma^2\mathbf{I}_N + \sum_{p=1}^{L}\frac{1}{\Gamma_p}\mathbf{R}_p\right)\right\}$$

$$+ \log\sigma^2 - \sum_{p=1}^{L}(1 + \log x_p) + \mathcal{O}(1/N)$$

$$= \mathop{\mathsf{Extr}}_{\{\Gamma_p\}}\left\{\sum_{p=1}^{L}(\Gamma_p x_p - \log\Gamma_p x_p - 1)\right.$$

$$\left. - \frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{p=1}^{L}\frac{1}{\sigma^2\Gamma_p}\mathbf{R}_p\right)\right\} + \mathcal{O}(1/N), \tag{94}$$

where the second equality is obtained by taking the Gaussian integration with respect to $\mathbf{z}$, and the third equality is obtained by applying the saddle-point method. $\qquad\square$

REFERENCES

[1] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

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

[3] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[4] K. Hayashi, M. Nagahara, and T. Tanaka, "A user's guide to compressed sensing for communications systems," *IEICE Trans. Commun.*, vol. E96-B, no. 3, pp. 685–712, Mar. 2013.

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

[6] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4290–4308, Jul. 2013.

[7] N. Parikh and S. Boyd, *Proximal Algorithms*. Foundations and Trends in Optimization, 2014.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage thresholding algorithm for linear inverse problems," *SIAM J. on Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[9] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, 2009.

[10] A. Maleki and R. G. Baraniuk, "Least favorable compressed sensing problems for the first order methods," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 134–138.

[11] T. T. Do, T. D. Tran, and L. Gan, "Fast compressive sampling with structurally random matrices," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, 2008, pp. 3369–3372.

[12] J. Barbier, F. Krzakala, and C. Schülke, "Approximate message-passing with spatially coupled structured operators, with applications to compressed sensing and sparse superposition codes," *J. Stat. Mech.*, vol. 2015, no. 5, p. P05013, 2015.

[13] J. Barbier and F. Krzakala, "Approximate message-passing decoder and capacity-achieving sparse superposition codes," preprint, 2015. [Online]. Available: http://arxiv.org/abs/1503.08040

[14] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Nat. Acad. Sci.*, 2005.

[15] ——, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data an alysis and signal processing," *Philos. Trans. Roy. Soc. London A, Math. Phys. Sci.*, vol. 367, no. 1906, pp. 4273–4293, Nov. 2011.

[16] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and message passing algorithms," *Ann. Appl. Probab.*, vol. 25, no. 2, pp. 753–822, 2015.

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

[18] Y. Kabashima, M. M. Vehkaperä, and S. Chatterjee, "Typical l1-recovery limit of sparse vectors represented by concatenations of random orthogonal matrices," *J. Stat. Mech.*, vol. 2012, no. 12, p. P12003, 2012.

[19] 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.

[20] M. Vehkaperä, Y. Kabashima, and S. Chatterjee, "Analysis of regularized LS reconstruction and random matrix ensembles in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2100–2124, Apr. 2016.

[21] S. Oymak and B. Hassibi, "A case for orthogonal measurements in linear inverse problems," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Honolulu, HI, 2014, pp. 3175–3179.

[22] J. Ma, X. Yuan, and L. Ping, "Turbo compressed sensing with partial DFT sensing matrix," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 158–161, Feb. 2015.

[23] ——, "On the performance of turbo signal recovery with partial DFT sensing matrices," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1580–1584, Oct. 2015.

[24] C. Thrampoulidis and B. Hassibi, "Isotropically random orthogonal matrices: Performance of lasso and minimum conic singular values," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Hong Kong, 2015, pp. 556–560.

[25] K. Takeda, S. Uda, and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," *Europhysics Letters*, vol. 76, pp. 1193–1199, 2006.

[26] A. Hatabu, K. Takeda, and Y. Kabashima, "Statistical mechanical analysis of the Kronecker channel model for multiple-input multipleoutput wireless communication," *Phys. Rev. E*, vol. 80, pp. 061 124(1–12), 2009.

[27] 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.

[28] S. Patterson, Y. C. Eldar, and I. Keidar, "Distributed compressed sensing for static and time-varying networks," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4931–4946, Oct. 2014.

[29] P. Han, J. Zhu, R. Niu, and D. Baron, "Multi-processor approximate message passing using lossy compression," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016.

[30] S. M. J. E. Fowler and E. W. Tramel, *Block-based compressed sensing of images and video*. Foundations and Trends in Signal Processing, 2012.

[31] R. Couillet, J. Hoydis, M. Debbah, "Random beamforming over quasi-static and fading channels: A deterministic equivalent approach," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6392–6425, Oct. 2012.

[32] X. Yuan, J. Ma, and L. Ping, "Energy-spread-transform based MIMO systems: Iterative equalization, evolution analysis, and precoder optimization," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5237–5250, Sep. 2014.

[33] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publishers Inc., 2004.

[34] T. Tanaka and J. Raymond, "Optimal incorporation of sparsity information by weighted l1-optimization," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2010, pp. 1598–1602.

[35] 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. Inf. Theory*, vol. 58, no. 3, pp. 1902–1923, Mar. 2012.

[36] 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. P08009, 2012.

[37] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

[38] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *J. Physics F: Metal Physics*, vol. 5, pp. 965–974, 1975.

[39] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. ser. Number 111 in Int. Series on Monographs on Physics. Oxford U.K.: Oxford Univ. Press, 2001.

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

[41] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large N analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.

[42] D. Guo and S. Verdú, "Randomly spread CDMA: asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 1982–2010, Jun. 2005.

[43] R. R. Müller, "Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2821–2828, Nov. 2003.

[44] C. K. Wen and K. K. Wong, "Asymptotic analysis of spatially correlated MIMO multiple-access channels with arbitrary signaling inputs for joint and separate decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 252–268, Jan. 2007.

[45] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, "On an achievable rate of large Rayleigh block-fading MIMO channels with no CSI," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6517–6541, Oct. 2013.

[46] M. A. Girnyk, M. Vehkaperä, and L. K. Rasmussen, "Large-system analysis of correlated MIMO channels with arbitrary signaling in the presence of interference," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1536–1276, Apr. 2014.

[47] B. Zaidel, R. Müller, A. Moustakas, R. de Miguel, "Vector precoding for Gaussian MIMO broadcast channels: Impact of replica symmetry breaking," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1413–1440, Mar. 2012.

[48] B. Farrell, "Limiting empirical singular value distribution of restrictions of discrete fourier transform matrices," *Journal of Fourier Analysis and Applications*, vol. 17, no. 4, pp. 733–753, 2011.

[49] T. Jiang, "Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles," *Probability Theory and Related Fields*, vol. 144, no. 1, pp. 221–246, May 2009.

[50] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[51] C. K. Wen and K. K. Wong, "Analysis of compressed sensing with spatially-coupled orthogonal matrices," preprint, 2014. [Online]. Available: http://arxiv.org/abs/1402.3215.

[52] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2168–2172.

[53] B. Çakmak, O. Winther, and B. H. Fleury, "S-AMP: Approximate message passing for general matrix ensembles," in *2014 IEEE Information Theory Workshop (ITW)*, Hobart, Australia, Nov. 2014, pp. 192–196.

[54] Y. Kabashima and M. Vehkaperä, "Signal recovery using expectation consistent approximation for linear observations," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Honolulu, HI, 2014, pp. 226–230.

[55] T. Liu, C.-K. Wen, S. Jin, X. You, "Generalized turbo signal recovery for nonlinear measurements and orthogonal sensing matrices," preprint, 2015. [Online]. Available: http://arxiv.org/abs/1512.04833.

[56] M. Opper, B. Çakmak, and O. Winther, "A theory of solving TAP equations for Ising models with general invariant random matrices," *Journal of Physics A: Mathematical and Theoretical*, vol. 49, no. 11, p. 114002 (24pp), Feb. 2016.

[57] J. Ma and Li Ping, "Orthogonal AMP," preprint, 2016. [Online]. Available: http://arxiv.org/abs/1602.06509.

[58] C. K. Wen, K. K. Wong, and J. C. Chen, "Spatially correlated MIMO multiple-access systems with macrodiversity: Asymptotic analysis via statistical physics," *IEEE Trans. Commun.*, vol. 55, no. 3, pp. 477–488, Mar. 2007.

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