Asymptotic Analysis of MAP Estimation via the Replica Method and Applications to Compressed Sensing

Citation: Rangan, Sundeep, Alyson K. Fletcher, and Vivek K Goyal. "Asymptotic Analysis of MAP Estimation via the Replica Method and Applications to Compressed Sensing." IEEE Transactions on Information Theory 58.3 (2012): 1902–1923.

As Published: http://dx.doi.org/10.1109/tit.2011.2177575

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/73161

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Asymptotic Analysis of MAP Estimation via the Replica Method and Applications to Compressed Sensing

Sundeep Rangan, Alyson K. Fletcher, and Vivek K Goyal

***Abstract*—The replica method is a non-rigorous but well-known technique from statistical physics used in the asymptotic analysis of large, random, nonlinear problems. This paper applies the replica method, under the assumption of replica symmetry, to study estimators that are maximum a posteriori (MAP) under a postulated prior distribution. It is shown that with random linear measurements and Gaussian noise, the replica-symmetric prediction of the asymptotic behavior of the postulated MAP estimate of an n-dimensional vector "decouples" as n scalar postulated MAP estimators. The result is based on applying a hardening argument to the replica analysis of postulated posterior mean estimators of Tanaka and of Guo and Verdú.**

The replica-symmetric postulated MAP analysis can be readily applied to many estimators used in compressed sensing, including basis pursuit, lasso, linear estimation with thresholding, and zero norm-regularized estimation. In the case of lasso estimation the scalar estimator reduces to a soft-thresholding operator, and for zero norm-regularized estimation it reduces to a hard-threshold. Among other benefits, the replica method provides a computationally-tractable method for precisely predicting various performance metrics including mean-squared error and sparsity pattern recovery probability.

***Index Terms*—Compressed sensing, Laplace's method, large deviations. least absolute shrinkage and selection operator (lasso), nonlinear estimation, non-Gaussian estimation, random matrices, sparsity, spin glasses, statistical mechanics, thresholding**

I. Introduction

Estimating a vector $\mathbf{x} \in \mathbb{R}^n$ from measurements of the form

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\Phi \in \mathbb{R}^{m\times n}$ represents a known *measurement matrix* and $\mathbf{w} \in \mathbb{R}^m$ represents measurement errors or noise, is a generic problem that arises in a range of circumstances. When the noise $\mathbf{w}$ is i.i.d. zero-mean Gaussian with variance σ^2 and $\mathbf{x}$ is i.i.d. with components x_j having a probability distribution function $p(x_j)$, the *maximum a posteriori* (MAP) estimate is given by

$$\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \left[\frac{1}{2\sigma^2}\|\mathbf{y}-\Phi\mathbf{x}\|^2 + \sum_{j=1}^n f(x_j) \right], \tag{2}$$

where $f(x_j) = -\log p(x_j)$. Estimators of the form (2) are also used with the regularization function $f(x_j)$ or noise level parameter σ^2 not matching the true prior or noise level, either since those quantities are not known or since the optimization in (2) using the true values is too difficult to compute. In such cases, the estimator (2) can be interpreted as a MAP estimate for a *postulated* distribution and noise level, and we will thus call estimators of the form (2) *postulated MAP estimators*.

Due to their prevalence, characterizing the behavior of postulated MAP estimators is of interest in a wide range of applications. However, for most regularization functions $f(\cdot)$, the postulated MAP estimator (2) is nonlinear and not easy to analyze. Even if, for the purpose of analysis, one assumes separable priors on $\mathbf{x}$ and $\mathbf{w}$, the analysis of the postulated MAP estimate may be difficult since the matrix Φ couples the n unknown components of $\mathbf{x}$ with the m measurements in the vector $\mathbf{y}$.

This paper provides a general analysis of postulated MAP estimators based on the *replica method*—a non-rigorous but widely-used method from statistical physics for analyzing large random systems. It is shown that, under a key assumption of replica symmetry described below, the replica method predicts that with certain large random Φ and Gaussian $\mathbf{w}$, there is an *asymptotic decoupling* of the vector postulated MAP estimate (2) into n scalar MAP estimators. Specifically, the replica method predicts that the joint distribution of each component x_j of $\mathbf{x}$ and its corresponding component $\hat{x}_j$ in the estimate vector $\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y})$ is asymptotically identical to the outputs of a simple system where $\hat{x}_j$ is a postulated MAP estimate of the scalar random variable x_j observed in Gaussian noise. Using this scalar equivalent model, one can then readily compute the asymptotic joint distribution of $(x_j, \hat{x}_j)$ for any component j.

The replica method's non-rigorous but simple prescription for computing the asymptotic joint componentwise distributions has three key, attractive features:

- *Sharp predictions:* Most importantly, the replica method provides—under the assumption of the replica hypotheses—not just bounds, but sharp predictions of the asymptotic behavior of postulated MAP estimators. From the joint distribution, various further computations

This material is based upon work supported in part by a University of California President's Postdoctoral Fellowship and the National Science Foundation under CAREER Grant No. 0643836.

S. Rangan (email: srangan@poly.edu) is with the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University.

A. K. Fletcher (email: alyson@eecs.berkeley.edu) is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley.

V. K. Goyal (email: vgoyal@mit.edu) is with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology.

can be made, to provide precise predictions of quantities such as the mean-squared error (MSE) and the error probability of any componentwise hypothesis test computed from a postulated MAP estimate.
- *Computational tractability:* Since the scalar equivalent model involves only a scalar random variable x_j, scalar Gaussian noise, and scalar postulated MAP estimate $\hat{x}_j$, any quantity derived from the joint distribution can be computed numerically from one- or two-dimensional integrals.
- *Generality:* The replica analysis can incorporate arbitrary separable distributions on $\mathbf{x}$ and regularization functions $f(\cdot)$. It thus applies to a large class of estimators and test scenarios.

A. Replica Method and Contributions of this Work

The replica method was originally developed by Edwards and Anderson [1] to study the statistical mechanics of spin glasses. Although not fully rigorous from the perspective of probability theory, the technique was able to provide explicit solutions for a range of complex problems where many other methods had previously failed. Indeed, the replica method and related ideas from statistical mechanics have found success in a number of classic NP-hard problems including the traveling salesman problem [2], graph partitioning [3], k-SAT [4] and others [5]. Statistical physics methods have also been applied to the study of error correcting codes [6], [7]. There are now several general texts on the replica method [8]–[11].

The replica method was first applied to the study of non-linear MAP estimation problems by Tanaka [12]. That work applied what is called a replica symmetric analysis to multiuser detection for large CDMA systems with random spreading sequences. Müller [13] considered a mathematically-similar problem for MIMO communication systems. In the context of the estimation problem considered here, Tanaka's and Müller's papers essentially characterized the behavior of the MAP estimator of a vector $\mathbf{x}$ with i.i.d. binary components observed through linear measurements of the form (1) with a large random Φ and Gaussian $\mathbf{w}$.

Tanaka's results were then generalized in a remarkable paper by Guo and Verdú [14] to vectors $\mathbf{x}$ with arbitrary separable distributions. Guo and Verdú's result was also able to incorporate a large class of postulated minimum mean squared error (MMSE) estimators, where the estimator may assume a prior that is different from the actual prior. Replica analyses have also been applied to related communication problems such as lattice precoding for the Gaussian broadcast channel [15]. A brief review of the replica method analysis by Tanaka [12] and Guo and Verdú [14] is provided in Appendix A.

The result in this paper is derived from Guo and Verdú [14] by a standard hardening argument. Specifically, the postulated MAP estimator (2) is first expressed as a limit of the postulated MMSE estimators analyzed in [14]. Then, the behavior of the postulated MAP estimator can be derived by taking appropriate limits of the results in [14] on postulated MMSE estimators. This hardening technique is well-known and is used in Tanaka's original work [12] in the analysis of MAP estimators with binary and Gaussian priors.

Through the limiting analysis via hardening, the postulated MAP results here follow from the postulated MMSE results in [14]. Thus, the central contribution of this work is to work out these limits to provide a set of equations for a general class of postulated MAP estimators. In particular, while Tanaka has derived the equations for replica predictions of MAP estimates for binary and Gaussian priors, the results here provide explicit equations for general priors and regularization functions.

B. Replica Assumptions

The non-rigorous aspect of the replica method involves a set of assumptions that include a self-averaging property, the validity of a "replica trick," and the ability to exchange certain limits. Importantly, this work is based on an additional strong assumption of *replica symmetry*. As described in Appendix A, the replica method reduces the calculation of a certain free energy to an optimization problem over covariance matrices. The replica symmetric (RS) assumption is that the maxima in this optimization satisfy certain symmetry properties. This RS assumption is not always valid, and indeed Appendix A provides several examples from other applications of the replica method where replica symmetry breaking (RSB) solutions are known to be needed to provide correct predictions.

For the analysis of postulated MMSE estimators, [12] and [14] derive analytic conditions for the validity of the RS assumption only in some limited cases. Our analysis of postulated MAP estimators depends on [14], and, unfortunately, we have not provided a general analytic test for the validity of the RS assumption in this work. Following [14], our approach instead is to compare, where possible, the predictions under the RS assumption to numerical simulations of the postulated MAP estimator. As we will see in Section VI, the RS predictions appear to be accurate, at least for many common estimators arising in compressed sensing. That being said, the RS analysis can also provide predictions for optimal MMSE and zero norm-regularized estimators that cannot be simulated tractably. Extra caution must be applied in assuming the validity of the RS predictions for these estimators.

To emphasize our dependence on these unproven assumptions—notably replica symmetry—we will refer to the general MMSE analysis in Guo and Verdú's work [14] as the **replica symmetric postulated MMSE decoupling property**. Our main result will be called the **replica symmetric postulated MAP decoupling property**.

C. Connections to Belief Propagation

Although not explored in this work, it is important to point out that the results of the replica analysis of postulated MMSE and MAP estimation are similar to those derived for belief propagation (BP) estimation. Specifically, there is now a large body of work analyzing BP and approximate BP algorithms for estimation of vectors $\mathbf{x}$ observed through linear measurements of the form (1) with large random Φ. For both certain large sparse random matrices [16]–[22], and more recently for certain large dense random matrices [23]–[26], several results now show that BP estimates exhibit an asymptotic decoupling property similar to RS predictions for postulated MMSE and

MAP estimators. Graphical model arguments have also been used to establish a decoupling property under a very general, random sparse observation model [27].

The effective noise level in the scalar equivalent model for BP and approximate BP methods can be predicted by certain state evolution equations similar to density evolution analysis of BP decoding of LDPC codes [28], [29]. It turns out that in several cases, the fixed point equations for state evolution are identical to the equations for the effective noise level predicted by the RS analysis of postulated MMSE and MAP estimators. In particular, the equations in [23], [24] agree exactly with the RS predictions for LASSO estimation given in this work.

These connections are significant in several regards: Firstly, the state evolution analysis of BP algorithms can be made fully rigorous under suitable assumptions and thus provides an independent, rigorous justification for some of the RS claims.

Secondly, the replica method provides only an *analysis* of estimators, but no method to actually compute those estimators. In contrast, the BP and approximate BP algorithms provide a possible tractable method for achieving the performance predicted by the replica method.

Finally, the BP analysis provides an algorithmic intuition as to why decoupling may occur (and hence when replica symmetry may be valid): As described in [30], BP and approximate BP algorithms can be seen as iterative procedures where the vector estimation problem is reduced to a sequence of "decoupled" scalar estimation problems. This decoupling is based essentially on the principle that, in each iteration, when estimating one component x_j, the uncertainty in the other components $\{x_k,\, k \neq j\}$ can be aggregated as Gaussian noise. Based on the state evolution analysis of BP algorithms, we know that this Central Limit Theorem-based approximation is asymptotically valid when the components of the mixing matrix Φ are sufficiently dense and independent. Thus, the validity of RS is possibly connected to validity of this Gaussian approximation.

D. Applications to Compressed Sensing

As an application of our main result, we will develop a few analyses of estimation problems that arise in compressed sensing [31]–[33]. In *compressed sensing*, one estimates a sparse vector $\mathbf{x}$ from random linear measurements. A vector $\mathbf{x}$ is *sparse* when its number of nonzero entries k is smaller than its length n. Generically, optimal estimation of $\mathbf{x}$ with a sparse prior is NP-hard [34]. Thus, most attention has focused on greedy heuristics such as matching pursuit [35]–[38] and convex relaxations such as basis pursuit [39] or lasso [40]. While successful in practice, these algorithms are difficult to analyze precisely.

Compressed sensing of sparse $\mathbf{x}$ through (1) (using inner products with rows of Φ) is mathematically identical to *sparse approximation* of $\mathbf{y}$ with respect to columns of Φ. An important set of results for both sparse approximation and compressed sensing are the deterministic conditions on the *coherence* of Φ that are sufficient to guarantee good performance of the suboptimal methods mentioned above [41]–[43]. These conditions can be satisfied with high probability for certain large random measurement matrices. Compressed sensing has provided many sufficient conditions that are easier to satisfy than the initial coherence-based conditions. However, despite this progress, the exact performance of most sparse estimators is still not known precisely, even in the asymptotic case of large random measurement matrices. Most results describe the estimation performance via bounds, and the tightness of these bounds is generally not known.

There are, of course, notable exceptions including [44] and [45], which provide matching necessary and sufficient conditions for recovery of strictly sparse vectors with basis pursuit and lasso. However, even these results only consider exact recovery and are limited to measurements that are noise-free or measurements with a signal-to-noise ratio (SNR) that scales to infinity.

Many common sparse estimators can be seen as MAP estimators with certain postulated priors. Most importantly, lasso and basis pursuit are MAP estimators assuming a Laplacian prior. Other commonly-used sparse estimation algorithms, including linear estimation with and without thresholding and zero norm-regularized estimators, can also be seen as postulated MAP-based estimators. For these postulated MAP-based sparse estimation algorithms, the replica method can provide non-rigorous but sharp, easily-computable predictions for the asymptotic behavior. In the context of compressed sensing, this analysis can predict various performance metrics such as MSE or fraction of support recovery. The expressions can apply to arbitrary ratios k/n, n/m, and SNR. Due to the generality of the replica analysis, the methodology can also incorporate arbitrary distributions on $\mathbf{x}$ including several sparsity models, such as Laplacian, generalized Gaussian, and Gaussian mixture priors. Discrete distributions can also be studied.

It should be pointed out that this work is not the first to use ideas from statistical physics for the study of sparse estimation. Guo, Baron and Shamai [46] have provided a replica analysis of compressed sensing that characterizes not just the postulated MAP or postulated MMSE estimate, but the asymptotic posterior marginal distribution. That work also shows an independence property across finite sets of components. Merhav, Guo and Shamai [47] consider, among other applications, the estimation of a sparse vector $\mathbf{x}$ from measurements of the form $\mathbf{y} = \mathbf{x} + \mathbf{w}$. In their model, there is no measurement matrix such as Φ in (1), but the components of $\mathbf{x}$ are possibly correlated. Their work derives explicit expressions for the MMSE as a function of the probability distribution on the number of nonzero components. The analysis does not rely on replica assumptions and is fully rigorous. More recently, Kabashima, Wadayama and Tanaka [48] have used the replica method to derive precise conditions on which sparse signals can be recovered with ℓ_p-based relaxations such as lasso. Their analysis does not consider noise, but can find conditions on recovery on the entire vector $\mathbf{x}$, not just individual components. Also, using free probability theory [49], [50], a recent analysis [51] extends the replica analysis of compressed sensing to larger classes of matrices, including matrices Φ that are possibly not i.i.d.

E. Outline

The remainder of the paper is organized as follows. The precise estimation problem is described in Section II. We review the RS postulated MMSE decoupling property of Guo and Verdú in Section III. We then present our main result, an RS postulated MAP decoupling property, in Section IV. The results are applied to the analysis of compressed sensing algorithms in Section V, which is followed by numerical simulations in Section VI. Conclusions are possible avenues for future work are given in Section VII. The proof of the main result is somewhat long and given in a set of appendices; Appendix B provides an overview of the proof and a guide through the appendices with detailed arguments.

II. Estimation Problem and Assumptions

Consider the estimation of a random vector $\mathbf{x} \in \mathbb{R}^n$ from linear measurements of the form

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{w} = \mathbf{A}\mathbf{S}^{1/2}\mathbf{x} + \mathbf{w}, \tag{3}$$

where $\mathbf{y} \in \mathbb{R}^m$ is a vector of observations; $\Phi = \mathbf{A}\mathbf{S}^{1/2}$, with $\mathbf{A} \in \mathbb{R}^{m\times n}$, is a measurement matrix; $\mathbf{S}$ is a diagonal matrix of positive scale factors,

$$\mathbf{S} = \mathrm{diag}\,(s_1, \ldots, s_n), \quad s_j > 0; \tag{4}$$

and $\mathbf{w} \in \mathbb{R}^m$ is zero-mean, white Gaussian noise. We consider a sequence of such problems indexed by n, with $n \to \infty$. For each n, the problem is to determine an estimate $\widehat{\mathbf{x}}$ of $\mathbf{x}$ from the observations $\mathbf{y}$ knowing the measurement matrix $\mathbf{A}$ and scale factor matrix $\mathbf{S}$.

The components x_j of $\mathbf{x}$ are modeled as zero mean and i.i.d. with some prior probability distribution $p_0(x_j)$. The per-component variance of the Gaussian noise is $\mathbf{E}|w_j|^2 = \sigma_0^2$. We use the subscript "0" on the prior and noise level to differentiate these quantities from certain "postulated" values to be defined later. When we develop applications in Section V, the prior $p_0(x_j)$ will incorporate presumed sparsity of $\mathbf{x}$.

In (3), we have factored $\Phi = \mathbf{A}\mathbf{S}^{1/2}$ so that even with the i.i.d. assumption on $\{x_j\}_{j=1}^n$ above and an i.i.d. assumption on entries of $\mathbf{A}$, the model can capture variations in powers of the components of $\mathbf{x}$ that are known *a priori* at the estimator. Specifically, multiplication by $\mathbf{S}^{1/2}$ scales the variance of the jth component of $\mathbf{x}$ by a factor s_j. Variations in the power of $\mathbf{x}$ that are not known to the estimator should be captured in the distribution of $\mathbf{x}$.

We summarize the situation and make additional assumptions to specify the problem precisely as follows:

(a) The number of measurements $m = m(n)$ is a deterministic quantity that varies with n and satisfies
$$\lim_{n\to\infty} n/m(n) = \beta$$
for some $\beta \geq 0$. (The dependence of m on n is usually omitted for brevity.)
(b) The components x_j of $\mathbf{x}$ are i.i.d. with probability distribution $p_0(x_j)$. All moments of x_j are finite.
(c) The noise $\mathbf{w}$ is Gaussian with $\mathbf{w} \sim \mathcal{N}(0, \sigma_0^2\mathbf{I}_m)$.
(d) The components of the matrix $\mathbf{A}$ are i.i.d. and distributed as $A_{ij} \sim (1/\sqrt{m})A$ for some random variable A with zero mean, unit variance and all other moments of A finite.
(e) The scale factors s_j are i.i.d., satisfy $s_j > 0$ almost surely, and all moments of s_j are finite.
(f) The scale factor matrix $\mathbf{S}$, measurement matrix $\mathbf{A}$, vector $\mathbf{x}$, and noise $\mathbf{w}$ are all independent.

III. Review of the Replica Symmetric Postulated MMSE Decoupling Property

We begin by reviewing the RS postulated MMSE decoupling property of Guo and Verdú [14].

A. Postulated MMSE Estimators

To define the concept of a postulated MMSE estimator, suppose one is given a "postulated" prior distribution $p_{\rm post}$ and a postulated noise level $\sigma^2_{\rm post}$ that may be different from the true values p_0 and σ_0^2. We define the *postulated minimum MSE* (PMMSE) estimate of $\mathbf{x}$ as

$$\begin{aligned}\hat{\mathbf{x}}^{\rm pmmse}(\mathbf{y}) &= \mathbf{E}\left(\mathbf{x} \mid \mathbf{y}\,;\, p_{\rm post}, \sigma^2_{\rm post}\right) \\ &= \int \mathbf{x} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}\,;\, p_{\rm post}, \sigma^2_{\rm post})\, d\mathbf{x}, \end{aligned} \tag{5}$$

where $p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}\,;\, q, \sigma^2)$ is the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ under the $\mathbf{x}$ distribution q and noise variance σ^2 specified as parameters after the semicolon. We will use this sort of notation throughout the rest of the paper, including the use of p without a subscript for the p.d.f. of the scalar or vector quantity understood from context. In this case, due to the Gaussianity of the noise, we have

$$\begin{aligned} &p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}\,;\, q, \sigma^2) \\ &\quad = C^{-1} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|^2\right) q(\mathbf{x}), \end{aligned} \tag{6}$$

where the normalization constant is

$$C = \int \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|^2\right) q(\mathbf{x})\, d\mathbf{x}$$

and $q(\mathbf{x})$ is the joint p.d.f.

$$q(\mathbf{x}) = \prod_{j=1}^n q(x_j).$$

In the case when $p_{\rm post} = p_0$ and $\sigma^2_{\rm post} = \sigma_0^2$, so that the postulated and true values agree, the PMMSE estimate reduces to the true MMSE estimate.

B. Decoupling under Replica Symmetric Assumption

The essence of the RS PMMSE decoupling property is that the asymptotic behavior of the PMMSE estimator is described by an equivalent scalar estimator. Let $q(x)$ be a probability distribution defined on some set $\mathcal{X} \subseteq \mathbb{R}$. Given $\mu > 0$, let $p_{x|z}(x \mid z\,;\, q, \mu)$ be the conditional distribution

$$\begin{aligned} &p_{x|z}(x \mid z\,;\, q, \mu) \\ &\quad = \left[\int_{x\in\mathcal{X}} \phi(z - x\,;\, \mu) q(x)\, dx\right]^{-1} \phi(z - x\,;\, \mu) q(x) \end{aligned} \tag{7}$$

where $\phi(\cdot)$ is the Gaussian distribution

$$\phi(v\,;\,\mu) = \frac{1}{\sqrt{2\pi\mu}} e^{-|v|^2/(2\mu)}. \tag{8}$$

The distribution $p_{x|z}(x|z\,;\,q,\mu)$ is the conditional distribution of the scalar random variable $x \sim q(x)$ given an observation of the form

$$z = x + \sqrt{\mu} v, \tag{9}$$

where $v \sim \mathcal{N}(0,1)$. Using this distribution, we can define the scalar conditional MMSE estimate

$$\hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,q,\mu) = \int_{x\in\mathcal{X}} x\, p_{x|z}(x \mid z\,;\,\mu)\, dx. \tag{10}$$

Also, given two distributions, $p_0(x)$ and $p_1(x)$, and two noise levels, $\mu_0 > 0$ and $\mu_1 > 0$, define

$$\begin{aligned}&\mathsf{mse}(p_1, p_0, \mu_1, \mu_0, z)\\ &= \int_{x\in\mathcal{X}} |x - \hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,p_1,\mu_1)|^2\, p_{x|z}(x \mid z\,;\,p_0,\mu_0)\, dx,\end{aligned} \tag{11}$$

which is the MSE in estimating the scalar x from the variable z in (9) when x has a true distribution $x \sim p_0(x)$ and the noise level is $\mu = \mu_0$, but the estimator assumes a distribution $x \sim p_1(x)$ and noise level $\mu = \mu_1$.

Replica Symmetric Postulated MMSE Decoupling Property [14]: Consider the estimation problem in Section II. Let $\hat{\mathbf{x}}^{\rm pmmse}(\mathbf{y})$ be the PMMSE estimator based on a postulated prior $p_{\rm post}$ and postulated noise level $\sigma^2_{\rm post}$. For each n, let $j = j(n)$ be some deterministic component index with $j(n) \in \{1,\ldots,n\}$. Then under replica symmetry, there exist *effective noise levels* $\sigma^2_{\rm eff}$ and $\sigma^2_{\rm p-eff}$ such that:

(a) As $n \to \infty$, the random vectors $(x_j, s_j, \hat{x}_j^{\rm pmmse})$ converge in distribution to the random vector $(x, s, \hat{x})$ consistent with the block diagram in Fig. 1. Here x, s, and v are independent with $x \sim p_0(x)$, $s \sim p_S(s)$, $v \sim \mathcal{N}(0,1)$, and

$$\begin{aligned}\hat{x} &= \hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,p_{\rm post},\mu_p), &\text{(12a)}\\ z &= x + \sqrt{\mu} v, &\text{(12b)}\end{aligned}$$

where $\mu = \sigma^2_{\rm eff}/s$ and $\mu_p = \sigma^2_{\rm p-eff}/s$.

(b) The effective noise levels satisfy the equations

$$\begin{aligned}\sigma^2_{\rm eff} &= \sigma_0^2 + \beta\, \mathbf{E}\left[s\, \mathsf{mse}(p_{\rm post}, p_0, \mu_p, \mu, z)\right] &\text{(13a)}\\ \sigma^2_{\rm p-eff} &= \sigma^2_{\rm post}\\ &\quad + \beta\, \mathbf{E}\left[s\, \mathsf{mse}(p_{\rm post}, p_{\rm post}, \mu_p, \mu_p, z)\right] &\text{(13b)}\end{aligned}$$

where the expectations are taken over $s \sim p_S(s)$ and z generated by (12b).

This result asserts that the asymptotic behavior of the joint estimation of the n-dimensional vector $\mathbf{x}$ can be described by n equivalent scalar estimators. In the scalar estimation problem, a component $x \sim p_0(x)$ is corrupted by additive Gaussian noise yielding a noisy measurement z. The additive noise variance is $\mu = \sigma^2_{\rm eff}/s$, which is the effective noise divided by the scale factor s. The estimate of that component is then described by the (generally nonlinear) scalar estimator $\hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,p_{\rm post},\mu_p)$.

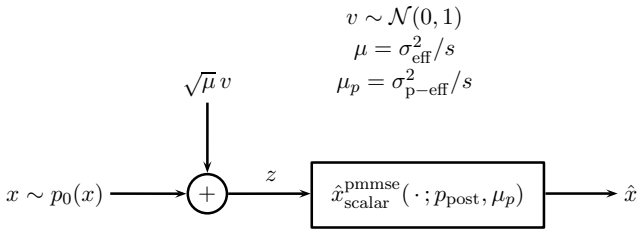

Fig. 1. Equivalent scalar model for the estimator behavior predicted by the replica symmetric postulated MMSE decoupling property.

The effective noise levels $\sigma^2_{\rm eff}$ and $\sigma^2_{\rm p-eff}$ are described by the solutions to fixed-point equations (13). Note that $\sigma^2_{\rm eff}$ and $\sigma^2_{\rm p-eff}$ appear implicitly on the left- and right-hand sides of these equations via the terms μ and μ_p. In general, there is no closed form solution to these equations. However, the expectations can be evaluated via (one-dimensional) numerical integration.

It is important to point out that there may, in general, be multiple solutions to the fixed-point equations (13). In this case, it turns out that the true solution is the minimizer of a certain Gibbs' function described in [14].

C. Effective Noise and Multiuser Efficiency

To understand the significance of the effective noise level $\sigma^2_{\rm eff}$, it is useful to consider the following estimation problem with side information. Suppose that when estimating the component x_j an estimator is given as side information the values of all the other components $\{x_\ell,\, \ell \neq j\}$. Then, this hypothetical estimator with side information can "subtract out" the effect of all the known components and compute

$$z_j = \frac{1}{\|\mathbf{a}_j\|^2\sqrt{s_j}} \mathbf{a}_j' \left(\mathbf{y} - \sum_{\ell\neq j} \sqrt{s_\ell}\, \mathbf{a}_\ell\, x_\ell\right),$$

where $\mathbf{a}_\ell$ is the ℓth column of the measurement matrix $\mathbf{A}$. It is easily checked that

$$\begin{aligned}z_j &= \frac{1}{\|\mathbf{a}_j\|^2\sqrt{s_j}} \mathbf{a}_j' \left(\sqrt{s_j}\, \mathbf{a}_j\, x_j + \mathbf{w}\right)\\ &= x_j + \sqrt{\mu_0}\, v_j,\end{aligned} \tag{14}$$

where

$$v_j = \frac{1}{\sigma_0 \|\mathbf{a}_j\|^2} \mathbf{a}_j' \mathbf{w}, \quad \mu_0 = \frac{\sigma_0^2}{s_j}.$$

Thus, (14) shows that with side information, estimation of x_j reduces to a scalar estimation problem where x_j is corrupted by additive noise $\sqrt{\mu_0}\, v_j$. Since $\mathbf{w}$ is Gaussian with mean zero and per-component variance σ_0^2, v_j is Gaussian with mean zero and variance $1/\|\mathbf{a}_j\|^2$. Also, since $\mathbf{a}_j$ is an m-dimensional vector whose components are i.i.d. with variance $1/m$, $\|\mathbf{a}_j\|^2 \to 1$ as $m \to \infty$. Therefore, for large m, v_j will approach $v_j \sim \mathcal{N}(0,1)$.

Comparing (14) with (12b), we see that the equivalent scalar model predicted by the RS PMMSE decoupling property (12b) is identical to the estimation with perfect side information (14), except that the noise level is increased by a factor

$$1/\eta = \mu/\mu_0 = \sigma^2_{\rm eff}/\sigma_0^2. \tag{15}$$

In multiuser detection, the factor η is called the *multiuser efficiency* [52], [53].

The multiuser efficiency can be interpreted as degradation in the effective signal-to-noise ratio (SNR): With perfect side-information, an estimator using z_j in (14) can estimate x_j with an effective SNR of

$$\mathsf{SNR}_0(s) = \frac{1}{\mu_0}\mathbf{E}|x_j|^2 = \frac{s}{\sigma_0^2}\mathbf{E}|x_j|^2. \tag{16}$$

In CDMA multiuser detection, the factor $\mathsf{SNR}_0(s)$ is called the post-despreading SNR with no multiple access interference. The RS PMMSE decoupling property shows that without side information, the effective SNR is given by

$$\mathsf{SNR}(s) = \frac{1}{\mu}\mathbf{E}|x_j|^2 = \frac{s}{\sigma_{\rm eff}^2}\mathbf{E}|x_j|^2. \tag{17}$$

Therefore, the multiuser efficiency η in (15) is the ratio of the effective SNR with and without perfect side information.

IV. Analysis of Postulated MAP Estimators via Hardening

The main result of the paper is developed in this section.

A. Postulated MAP Estimators

Let $\mathcal{X} \subseteq \mathbb{R}$ be some (measurable) set and consider an estimator of the form

$$\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathcal{X}^n} \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|_2^2 + \sum_{j=1}^n f(x_j), \tag{18}$$

where $\gamma > 0$ is an algorithm parameter and $f : \mathcal{X} \to \mathbb{R}$ is some scalar-valued, nonnegative cost function. We will assume that the objective function in (18) has a unique essential minimizer for almost all $\mathbf{y}$.

The estimator (18) can be interpreted as a MAP estimator. To see this, suppose that for u sufficiently large,

$$\int_{\mathbf{x}\in\mathcal{X}^n} e^{-uf(\mathbf{x})}\,d\mathbf{x} < \infty, \tag{19}$$

where we have extended the notation $f(\,\cdot\,)$ to vector arguments such that

$$f(\mathbf{x}) = \sum_{j=1}^n f(x_j). \tag{20}$$

When (19) is satisfied, we can define a prior probability distribution depending on u:

$$p_u(\mathbf{x}) = \left[\int_{\mathbf{x}\in\mathcal{X}^n} \exp(-uf(\mathbf{x}))\,d\mathbf{x}\right]^{-1} \exp(-uf(\mathbf{x})). \tag{21}$$

Also, let

$$\sigma_u^2 = \gamma/u. \tag{22}$$

Substituting (21) and (22) into (6), we see that

$$\begin{aligned} &p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}\,;\, p_u, \sigma_u^2) \\ &\quad = \; C_u \exp\left[-u\left(\frac{1}{2\gamma}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|^2 + f(\mathbf{x})\right)\right] \end{aligned} \tag{23}$$

for some constant C_u that does not depend on $\mathbf{x}$. (The scaling of the noise variance along with p_u enables the factorization in the exponent of (23).) Comparing to (18), we see that

$$\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y}) = \arg\max_{\mathbf{x}\in\mathcal{X}^n} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}\,;\, p_u, \sigma_u^2).$$

Thus for all sufficiently large u, we indeed have a MAP estimate—assuming the prior p_u and noise level σ_u^2.

B. Decoupling under Replica Symmetric Assumption

To analyze the postulated MAP (PMAP) estimator, we consider a sequence of postulated MMSE estimators indexed by u. For each u, let

$$\widehat{\mathbf{x}}^u(\mathbf{y}) = \mathbf{E}\left(\mathbf{x} \mid \mathbf{y}\,;\, p_u, \sigma_u^2\right), \tag{24}$$

which is the MMSE estimator of $\mathbf{x}$ under the postulated prior p_u in (21) and noise level σ_u^2 in (22). Using a standard large deviations argument, one can show that under suitable conditions

$$\lim_{u\to\infty} \widehat{\mathbf{x}}^u(\mathbf{y}) = \hat{\mathbf{x}}^{\rm pmap}(\mathbf{y})$$

for all $\mathbf{y}$. A formal proof is given in Appendix D (see Lemma 4). Under the assumption that the behaviors of the postulated MMSE estimators are described by the RS PMMSE decoupling property, we can then extrapolate the behavior of the postulated MAP estimator. This will yield our main result.

In statistical physics the parameter u has the interpretation of inverse temperature (see a general discussion in [54]). Thus, the limit as $u \to \infty$ can be interpreted as a cooling or "hardening" of the system.

In preparation for the main result, define the scalar MAP estimator

$$\hat{x}_{\rm scalar}^{\rm pmap}(z\,;\,\lambda) = \arg\min_{x\in\mathcal{X}} F(x, z, \lambda) \tag{25}$$

where

$$F(x, z, \lambda) = \frac{1}{2\lambda}|z - x|^2 + f(x). \tag{26}$$

The estimator (25) plays a similar role as the scalar MMSE estimator (10).

The main result pertains to the estimator (18) applied to the sequence of estimation problems defined in Section II. Our assumptions are as follows:

Assumption 1: For all $u > 0$ sufficiently large, assume that the postulated MMSE estimator (5) with the postulated prior p_u in (21) and postulated noise level σ_u^2 in (22) satisfy the RS PMMSE decoupling property in Section III-B.

Assumption 2: Let $\sigma_{\rm eff}^2(u)$ and $\sigma_{\rm p-eff}^2(u)$ be the effective noise levels when using the postulated prior p_u and noise level σ_u^2. Assume the following limits exist:

$$\begin{aligned} \sigma_{\rm eff,map}^2 &= \lim_{u\to\infty} \sigma_{\rm eff}^2(u), \\ \gamma_p &= \lim_{u\to\infty} u\sigma_{\rm p-eff}^2(u). \end{aligned}$$

Assumption 3: Suppose for each n, $\hat{x}_j^u(n)$ is the MMSE estimate of the component x_j for some index $j \in \{1, \ldots, n\}$ based on the postulated prior p_u and postulated noise level

σ_u^2. Then, assume that limits can be interchanged to give the following equality:

$$\lim_{u\to\infty}\lim_{n\to\infty}\hat{x}_j^u(n) = \lim_{n\to\infty}\lim_{u\to\infty}\hat{x}_j^u(n),$$

where the limits are in distribution.

Assumption 4: For every n, $\mathbf{A}$, and $\mathbf{S}$, assume that for almost all $\mathbf{y}$, the minimization in (18) achieves a unique essential minimum. Here, essential should be understood in the standard measure-theoretic sense in that the minimum and essential infimum agree.

Assumption 5: Assume that $f(x)$ is nonnegative and satisfies

$$\lim_{|x|\to\infty}\frac{f(x)}{\log|x|} = \infty,$$

where the limit must hold over all sequences in $\mathcal{X}$ with $|x| \to \infty$. If $\mathcal{X}$ is compact, this limit is automatically satisfied (since there are no sequences in $\mathcal{X}$ with $|x| \to \infty$).

Assumption 6: For all $\lambda \in \mathbb{R}$ and almost all z, the minimization in (25) has a unique, essential minimum. Moreover, for all λ and almost all z, there exists a $\sigma^2(z,\lambda)$ such that

$$\lim_{x\to\hat{x}}\frac{|x-\hat{x}|^2}{2(F(x,z,\lambda)-F(\hat{x},z,\lambda))} = \sigma^2(z,\lambda), \qquad (28)$$

where $\hat{x} = \hat{x}_{\rm scalar}^{\rm pmap}(z\,;\,\lambda)$.

Assumption 1 is simply stated to again point out that we are assuming the validity of replica symmetry for the postulated MMSE estimates. We make the additional Assumptions 2 and 3, which are also difficult to verify but similar in spirit. Taken together, Assumptions 1–3 reflect the main limitations of the replica symmetric analysis and precisely state the manner in which the analysis is non-rigorous.

Assumptions 4–6 are technical conditions on the existence and uniqueness of the MAP estimate. Assumption 4 will be true for any strictly convex regularization $f(x_j)$, although it is difficult to verify in the non-convex case. The other two assumptions, Assumptions 5 and 6, will be verified for the problems of interest. In fact, we will explicitly calculate $\sigma^2(z,\lambda)$.

We can now state our extension of the RS PMMSE decoupling property.

Replica Symmetric Postulated MAP Decoupling Property: Consider the estimation problem in Section II. Let $\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y})$ be the postulated MAP estimator (18) defined for some $f(x)$ and $\gamma > 0$ satisfying Assumptions 1–6. For each n, let $j = j(n)$ be some deterministic component index with $j(n) \in \{1,\ldots,n\}$. Then under replica symmetry (as part of Assumption 1):

(a) As $n \to \infty$, the random vectors $(x_j, s_j, \hat{x}_j^{\rm pmap})$ converge in distribution to the random vector $(x, s, \hat{x})$ consistent with the block diagram in Fig. 2 for the limiting *effective noise levels* $\sigma_{\rm eff}^2$ and γ_p in Assumption 2. Here x, s, and v are independent with $x \sim p_0(x)$, $s \sim p_S(s)$, $v \sim \mathcal{N}(0,1)$, and

$$\hat{x} = \hat{x}_{\rm scalar}^{\rm pmap}(z,\lambda_p), \qquad (29a)$$
$$z = x + \sqrt{\mu}v, \qquad (29b)$$

where $\mu = \sigma_{\rm eff,map}^2/s$ and $\lambda_p = \gamma_p/s$.

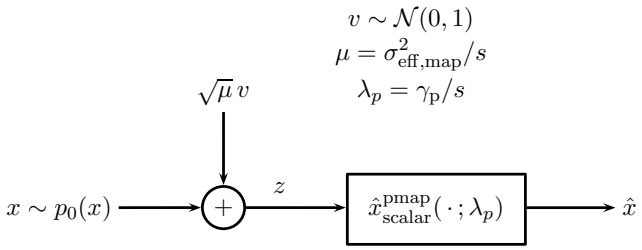

Fig. 2. Equivalent scalar model for the estimator behavior predicted by the replica symmetric postulated MAP decoupling property.

(b) The limiting effective noise levels $\sigma_{\rm eff,map}^2$ and γ_p satisfy the equations

$$\sigma_{\rm eff,map}^2 = \sigma_0^2 + \beta\,\mathbf{E}\left[s|x-\hat{x}|^2\right], \qquad (30a)$$
$$\gamma_p = \gamma + \beta\,\mathbf{E}\left[s\sigma^2(z,\lambda_p)\right], \qquad (30b)$$

where the expectations are taken over $x \sim p_0(x)$, $s \sim p_S(s)$, and $v \sim \mathcal{N}(0,1)$, with $\hat{x}$ and z defined in (29).

Proof: See Appendices B–F. ∎

Analogously to the RS PMMSE decoupling property, the RS PMAP decoupling property asserts that asymptotic behavior of the PMAP estimate of any single component of $\mathbf{x}$ is described by a simple equivalent scalar estimator. In the equivalent scalar model, the component of the true vector $\mathbf{x}$ is corrupted by Gaussian noise and the estimate of that component is given by a scalar PMAP estimate of the component from the noise-corrupted version.

V. Analysis of Compressed Sensing

Our results thus far hold for any separable distribution for $\mathbf{x}$ (see Section II) and under mild conditions on the cost function f (see especially Assumption 5, but other assumptions also implicitly constrain f). In this section, we provide additional details on replica analysis for choices of f that yield PMAP estimators relevant to compressed sensing. Since the role of f is to determine the estimator, this is not the same as choosing sparse priors for $\mathbf{x}$. Numerical evaluations of asymptotic performance with sparse priors for $\mathbf{x}$ are given in Section VI.

A. Linear Estimation

We first apply the RS PMAP decoupling property to the simple case of linear estimation. Linear estimators only use second-order statistics and generally do not directly exploit sparsity or other aspects of the distribution of the unknown vector $\mathbf{x}$. Nonetheless, for sparse estimation problems, linear estimators can be used as a first step in estimation, followed by thresholding or other nonlinear operations [55], [56]. It is therefore worthwhile to analyze the behavior of linear estimators even in the context of sparse priors.

The asymptotic behavior of linear estimators with large random measurement matrices is well known. For example, using the Marčenko-Pastur theorem [57], Verdú and Shamai [58] characterized the behavior of linear estimators with large i.i.d. matrices $\mathbf{A}$ and constant scale factors $\mathbf{S} = I$. Tse and Hanly [59] extended the analysis to general $\mathbf{S}$. Guo and

Verdú [14] showed that both of these results can be recovered as special cases of the general RS PMMSE decoupling property. We show here that the RS PMAP decoupling property can also recover these results. Although the calculations are very similar to [14], and indeed we arrive at precisely the same results, walking through the computations will illustrate how the RS PMAP decoupling property is used.

To simplify the notation, suppose that the true prior on $\mathbf{x}$ is such that each component has zero mean and unit variance. Choose the cost function

$$f(x) = \frac{1}{2}|x|^2,$$

which corresponds to the negative log of a Gaussian prior also with zero mean and unit variance. With this cost function, the PMAP estimator (18) reduces to the linear estimator

$$\hat{\mathbf{x}}^{\text{pmap}}(\mathbf{y}) = \mathbf{S}^{1/2}\mathbf{A}'\left(\mathbf{A}\mathbf{S}\mathbf{A}' + \gamma I\right)^{-1}\mathbf{y}. \tag{31}$$

When $\gamma = \sigma_0^2$, the true noise variance, the estimator (31) is the linear MMSE estimate.

Now, let us compute the effective noise levels from the RS PMAP decoupling property. First note that $F(x, z, \lambda)$ in (26) is given by

$$F(x, z, \lambda) = \frac{1}{2\lambda}|z - x|^2 + \frac{1}{2}|x|^2,$$

and therefore the scalar MAP estimator in (25) is given by

$$\hat{x}^{\text{pmap}}_{\text{scalar}}(z\,;\,\lambda) = \frac{1}{1+\lambda}z. \tag{32}$$

A simple calculation also shows that $\sigma^2(z, \lambda)$ in (28) is given by

$$\sigma^2(z, \lambda) = \frac{\lambda}{1+\lambda}. \tag{33}$$

As part (a) of the RS PMAP decoupling property, let $\mu = \sigma^2_{\text{eff,map}}/s$ and $\lambda_p = \gamma_p/s$. Observe that

$$\begin{aligned}
&\mathbf{E}\left[s\,|x - \hat{x}^{\text{pmap}}_{\text{scalar}}(z\,;\,\lambda_p)|^2\right] \\
&\quad\overset{(a)}{=}\quad \mathbf{E}\left[s\left|x - \frac{1}{1+\lambda_p}z\right|^2\right] \\
&\quad\overset{(b)}{=}\quad \mathbf{E}\left[s\left|\frac{\lambda_p}{1+\lambda_p}x - \frac{\sqrt{\mu}}{1+\lambda_p}v\right|^2\right] \\
&\quad\overset{(c)}{=}\quad \frac{s(\lambda_p^2+\mu)}{(1+\lambda_p)^2},
\end{aligned} \tag{34}$$

where (a) follows from (32); (b) follows from (29b); and (c) follows from the fact that x and v are uncorrelated with zero mean and unit variance. Substituting (33) and (34) into the fixed-point equations (30), we see that the limiting noise levels $\sigma^2_{\text{eff,map}}$ and γ_p must satisfy

$$\begin{aligned}
\sigma^2_{\text{eff,map}} &= \sigma_0^2 + \beta\,\mathbf{E}\left[\frac{s(\lambda_p^2+\mu)}{(1+\lambda_p)^2}\right], \\
\gamma_p &= \gamma + \beta\,\mathbf{E}\left[\frac{s\lambda_p}{1+\lambda_p}\right],
\end{aligned}$$

where the expectation is over $s \sim p_S(s)$. In the case when $\gamma = \sigma_0^2$, it can be verified that a solution to these fixed-point equations is $\sigma^2_{\text{eff,map}} = \gamma_p$, which results in $\mu = \lambda_p$ and

$$\begin{aligned}
\sigma^2_{\text{eff,map}} &= \sigma_0^2 + \beta\,\mathbf{E}\left[\frac{s\lambda_p}{1+\lambda_p}\right] \\
&= \sigma_0^2 + \beta\,\mathbf{E}\left[\frac{s\sigma^2_{\text{eff,map}}}{s+\sigma^2_{\text{eff,map}}}\right].
\end{aligned} \tag{35}$$

The expression (35) is precisely the Tse-Hanly formula [59] for the effective interference. Given a distribution on s, this expression can be solved numerically for $\sigma^2_{\text{eff,map}}$. In the special case of constant s, (35) reduces to Verdú and Shamai's result in [60] and can be solved via a quadratic equation.

The RS PMAP decoupling property now states that for any component index j, the asymptotic joint distribution of $(x_j, s_j, \hat{x}_j)$ is described by x_j corrupted by additive Gaussian noise with variance $\sigma^2_{\text{eff,map}}/s$ followed by a scalar linear estimator.

As described in [14], the above analysis can also be applied to other linear estimators including the matched filter (where $\gamma \to \infty$) or the decorrelating receiver ($\gamma \to 0$).

B. Lasso Estimation

We next consider lasso estimation, which is widely used for estimation of sparse vectors. The lasso estimate [40] (sometimes referred to as basis pursuit denoising [39]) is given by

$$\hat{\mathbf{x}}^{\text{lasso}}(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_1, \tag{36}$$

where $\gamma > 0$ is an algorithm parameter. The estimator is essentially a least-squares estimator with an additional $\|\mathbf{x}\|_1$ regularization term to encourage sparsity in the solution. The parameter γ is selected to trade off the sparsity of the estimate with the prediction error. An appealing feature of lasso estimation is that the minimization in (36) is convex; lasso thus enables computationally-tractable algorithms for finding sparse estimates.

The lasso estimator (36) is identical to the PMAP estimator (18) with the cost function

$$f(x) = |x|.$$

With this cost function, $F(x, z, \lambda)$ in (26) is given by

$$F(x, z, \lambda) = \frac{1}{2\lambda}|z - x|^2 + |x|,$$

and therefore the scalar MAP estimator in (25) is given by

$$\hat{x}^{\text{pmap}}_{\text{scalar}}(z\,;\,\lambda) = T^{\text{soft}}_{\lambda}(z), \tag{37}$$

where $T^{\text{soft}}_{\lambda}(z)$ is the soft thresholding operator

$$T^{\text{soft}}_{\lambda}(z) = \begin{cases} z - \lambda, & \text{if } z > \lambda; \\ 0, & \text{if } |z| \le \lambda; \\ z + \lambda, & \text{if } z < -\lambda. \end{cases} \tag{38}$$

The RS PMAP decoupling property now states that there exists effective noise levels $\sigma^2_{\text{eff,map}}$ and γ_p such that for any component index j, the random vector $(x_j, s_j, \hat{x}_j)$ converges

in distribution to the vector $(x, s, \hat{x})$ where $x \sim p_0(x)$, $s \sim p_S(s)$, and $\hat{x}$ is given by

$$\hat{x} = T^{\rm soft}_{\lambda_p}(z), \qquad z = x + \sqrt{\mu} v, \tag{39}$$

where $v \sim \mathcal{N}(0,1)$, $\lambda_p = \gamma_p/s$, and $\mu = \sigma^2_{\rm eff,map}/s$. Hence, the asymptotic behavior of lasso has a remarkably simple description: the asymptotic distribution of the lasso estimate $\hat{x}_j$ of the component x_j is identical to x_j being corrupted by Gaussian noise and then soft-thresholded to yield the estimate $\hat{x}_j$.

This soft-threshold description has an appealing interpretation. Consider the case when the measurement matrix $\mathbf{A} = I$. In this case, the lasso estimator (36) reduces to n scalar estimates,

$$\hat{x}_j = T^{\rm soft}_{\lambda}\left(x_j + \sqrt{\mu_0} v_j\right), \qquad j = 1, 2, \ldots, n, \tag{40}$$

where $v_i \sim \mathcal{N}(0,1)$, $\lambda = \gamma/s$, and $\mu_0 = \sigma_0^2/s$. Comparing (39) and (40), we see that the asymptotic distribution of $(x_j, s_j, \hat{x}_j)$ with large random $\mathbf{A}$ is identical to the distribution in the trivial case where $\mathbf{A} = I$, except that the noise levels γ and σ_0^2 are replaced by effective noise levels γ_p and $\sigma^2_{\rm eff,map}$.

To calculate the effective noise levels, one can perform a simple calculation to show that $\sigma^2(z,\lambda)$ in (28) is given by

$$\sigma^2(z,\lambda) = \left\{ \begin{array}{ll} \lambda, & \text{if } |z| > \lambda; \\ 0, & \text{if } |z| \leq \lambda. \end{array} \right. \tag{41}$$

Hence,

$$\begin{aligned} \mathbf{E}\left[s\sigma^2(z,\lambda_p)\right] &= \mathbf{E}\left[s\lambda_p \Pr(|z| > \lambda_p)\right] \\ &= \gamma_p \Pr(|z| > \gamma_p/s), \end{aligned} \tag{42}$$

where we have used the fact that $\lambda_p = \gamma_p/s$. Substituting (37) and (42) into (30), we obtain the fixed-point equations

$$\sigma^2_{\rm eff,map} = \sigma_0^2 + \beta \mathbf{E}\left[s|x - T^{\rm soft}_{\lambda_p}(z)|^2\right], \tag{43a}$$

$$\gamma_p = \gamma + \beta\gamma_p \Pr(|z| > \gamma_p/s), \tag{43b}$$

where the expectations are taken with respect to $x \sim p_0(x)$, $s \sim p_S(s)$, and z in (39). Again, while these fixed-point equations do not have a closed-form solution, they can be relatively easily solved numerically given distributions of x and s.

C. Zero Norm-Regularized Estimation

Lasso can be regarded as a convex relaxation of zero norm-regularized estimator

$$\hat{\mathbf{x}}^{\rm zero}(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_0, \tag{44}$$

where $\|\mathbf{x}\|_0$ is the number of nonzero components of $\mathbf{x}$. For certain strictly sparse priors, zero norm-regularized estimation may provide better performance than lasso. While *computing* the zero norm-regularized estimate is generally very difficult, we can use the replica analysis to provide a simple prediction of its *performance*. This analysis can provide a bound on the performance achievable by practical algorithms.

To apply the RS PMAP decoupling property to the zero norm-regularized estimator (44), we observe that the zero norm-regularized estimator is identical to the PMAP estimator (18) with the cost function

$$f(x) = \left\{ \begin{array}{ll} 0, & \text{if } x = 0; \\ 1, & \text{if } x \neq 0. \end{array} \right. \tag{45}$$

Technically, this cost function does not satisfy the conditions of the RS PMAP decoupling property. For one thing, without bounding the range of x, the bound (19) is not satisfied. Also, the minimum of (25) does not agree with the essential infimum. To avoid these problems, we can consider an approximation of (45),

$$f_{\delta,M}(x) = \left\{ \begin{array}{ll} 0, & \text{if } |x| < \delta; \\ 1, & \text{if } |x| \in [\delta, M], \end{array} \right.$$

which is defined on the set $\mathcal{X} = \{x : |x| \leq M\}$. We can then take the limits $\delta \to 0$ and $M \to \infty$. For space considerations and to simplify the presentation, we will just apply the decoupling property with $f(x)$ in (45) and omit the details of taking the appropriate limits.

With $f(x)$ given by (45), the scalar MAP estimator in (25) is given by

$$\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda) = T^{\rm hard}_t(z), \qquad t = \sqrt{2\lambda}, \tag{46}$$

where $T^{\rm hard}_t$ is the hard thresholding operator,

$$T^{\rm hard}_t(z) = \left\{ \begin{array}{ll} z, & \text{if } |z| > t; \\ 0, & \text{if } |z| \leq t. \end{array} \right. \tag{47}$$

Now, similar to the case of lasso estimation, the RS PMAP decoupling property states that there exists effective noise levels $\sigma^2_{\rm eff,map}$ and γ_p such that for any component index j, the random vector $(x_j, s_j, \hat{x}_j)$ converges in distribution to the vector $(x, s, \hat{x})$ where $x \sim p_0(x)$, $s \sim p_S(s)$, and $\hat{x}$ is given by

$$\hat{x} = T^{\rm hard}_t(z), \qquad z = x + \sqrt{\mu} v, \tag{48}$$

where $v \sim \mathcal{N}(0,1)$, $\lambda_p = \gamma_p/s$, $\mu = \sigma^2_{\rm eff,map}/s$, and

$$t = \sqrt{2\lambda_p} = \sqrt{2\gamma_p/s}. \tag{49}$$

Thus, the zero norm-regularized estimation of a vector $\mathbf{x}$ is equivalent to n scalar components corrupted by some effective noise level $\sigma^2_{\rm eff,map}$ and hard-thresholded based on an effective noise level γ_p.

The fixed-point equations for the effective noise levels $\sigma^2_{\rm eff,map}$ and γ_p can be computed similarly to the case of lasso. Specifically, one can verify that (41) and (42) are both satisfied for the hard thresholding operator as well. Substituting (42) and (46) into (30), we obtain the fixed-point equations

$$\sigma^2_{\rm eff,map} = \sigma_0^2 + \beta \mathbf{E}\left[s|x - T^{\rm hard}_t(z)|^2\right], \tag{50a}$$

$$\gamma_p = \gamma + \beta\gamma_p \Pr(|z| > t), \tag{50b}$$

where the expectations are taken with respect to $x \sim p_0(x)$, $s \sim p_S(s)$, z in (48), and t given by (49). These fixed-point equations can be solved numerically.

D. Optimal Regularization

The lasso estimator (36) and zero norm-regularized estimator (44) require the setting of a regularization parameter γ. Qualitatively, the parameter provides a mechanism to trade off the sparsity level of the estimate with the fitting error. One of the benefits of the replica analysis is that it provides a simple mechanism for optimizing the parameter level given the problem statistics.

Consider first the lasso estimator (36) with some $\beta > 0$ and distributions $x \sim p_0(x)$ and $s \sim p_S(s)$. Observe that there exists a solution to (43b) with $\gamma > 0$ if and only if

$$\Pr\left(|z| > \gamma_p/s\right) < 1/\beta. \tag{51}$$

This leads to a natural optimization: we consider an optimization over two variables $\sigma^2_{\text{eff,map}}$ and γ_p, where we minimize $\sigma^2_{\text{eff,map}}$ subject to (43a) and (51).

One simple procedure for performing this minimization is as follows: Start with $t = 0$ and some initial value of $\sigma^2_{\text{eff,map}}(0)$. For any iteration $t \geq 0$, we update $\sigma^2_{\text{eff,map}}(t)$ with the minimization

$$\sigma^2_{\text{eff,map}}(t+1) = \sigma_0^2 + \beta \min_{\gamma_p} \mathbf{E}\left[s|x - T^{\text{soft}}_{\lambda_p}(z)|^2\right], \tag{52}$$

where, on the right-hand side, the expectation is taken over $x \sim p_0(x)$, $s \sim p_S(s)$, z in (39), $\mu = \sigma^2_{\text{eff,map}}(t)/s$, and $\lambda_p = \gamma_p/s$. The minimization in (52) is over $\gamma_p > 0$ subject to (51). One can show that with a sufficiently high initial condition, the sequence $\sigma^2_{\text{eff,map}}(t)$ monotonically decreases to a local minimum of the objective function. Given the final value for γ_p, one can then recover γ from (43b). A similar procedure can be used for the zero norm-regularized estimator.

VI. Numerical Simulations

A. Bernoulli–Gaussian Mixture Distribution

As discussed above, the replica method is based on certain unproven assumptions and even then the decoupling results under replica symmetry are only asymptotic for the large dimension limit. To validate the predictive power of the RS PMAP decoupling property for finite dimensions, we first performed numerical simulations where the components of $\mathbf{x}$ are a zero-mean Bernoulli–Gaussian process, or equivalently a two-component, zero-mean Gaussian mixture where one component has zero variance. Specifically,

$$x_j \sim \begin{cases} \mathcal{N}(0,1), & \text{with prob. } \rho; \\ 0, & \text{with prob. } 1-\rho, \end{cases}$$

where ρ represents a sparsity ratio. In the experiments, $\rho = 0.1$. This is one of many possible sparse priors.

We took the vector $\mathbf{x}$ to have $n = 100$ i.i.d. components with this prior, and we varied m for 10 different values of $\beta = n/m$ from 0.5 to 3. For the measurements (3), we took a measurement matrix $\mathbf{A}$ with i.i.d. Gaussian components and a constant scale factor matrix $\mathbf{S} = I$. The noise level σ_0^2 was set so that SNR_0 = 10 dB, where SNR_0 is the signal-to-noise ratio with perfect side information defined in (16).

We simulated various estimators and compared their performances against the asymptotic values predicted by the replica analysis. For each value of β, we performed 1000 Monte Carlo trials of each estimator. For each trial, we measured the normalized squared error (SE) in dB

$$10\log_{10}\left(\frac{\|\widehat{\mathbf{x}} - \mathbf{x}\|^2}{\|\mathbf{x}\|^2}\right),$$

where $\widehat{\mathbf{x}}$ is the estimate of $\mathbf{x}$. The results are shown in Fig. 3, with each set of 1000 trials represented by the median normalized SE in dB.

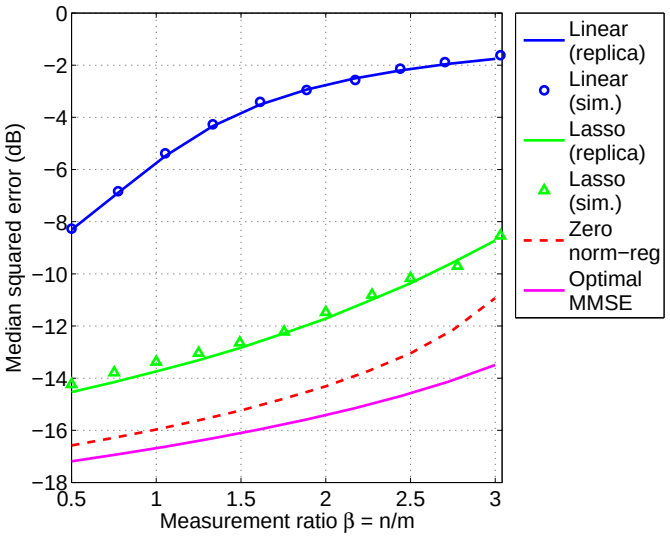

Fig. 3. MSE performance prediction with the RS PMAP decoupling property. Plotted is the median normalized SE for various sparse recovery algorithms: linear MMSE estimation, lasso, zero norm-regularized estimation, and optimal MMSE estimation. Solid lines show the asymptotic predicted MSE from the replica method. For the linear and lasso estimators, the circles and triangles show the actual median SE over 1000 Monte Carlo simulations. The unknown vector has i.i.d. Bernoulli–Gaussian components with a 90% probability of being zero. The noise level is set so that SNR_0 = 10 dB. See text for details.

The top curve shows the performance of the linear MMSE estimator (31). As discussed in Section V-A, the RS PMAP decoupling property applied to the case of a constant scale matrix $\mathbf{S} = I$ reduces to Verdú and Shamai's result in [60]. As can be seen in Fig. 3, the result predicts the simulated performance of the linear estimator extremely well.

The next curve shows the lasso estimator (36) with the factor γ selected to minimize the MSE as described in Section V-D. To compute the predicted value of the MSE from the RS PMAP decoupling property, we numerically solve the fixed-point equations (43) to obtain the effective noise levels $\sigma^2_{\text{eff,map}}$ and γ_p. We then use the scalar MAP model with the estimator (37) to predict the MSE. We see from Fig. 3 that the predicted MSE matches the median SE within 0.3 dB over a range of β values. At the time of initial dissemination of this work [61], precise prediction of lasso's performance given a specific noise variance and prior was not achievable with any other method. Now, as discussed in Section I-C, such asymptotic performance predictions can also be proven rigorously through connections with approximate belief propagation.

Fig. 3 also shows the theoretical minimum MSE (as computed with the RS PMMSE decoupling property) and the

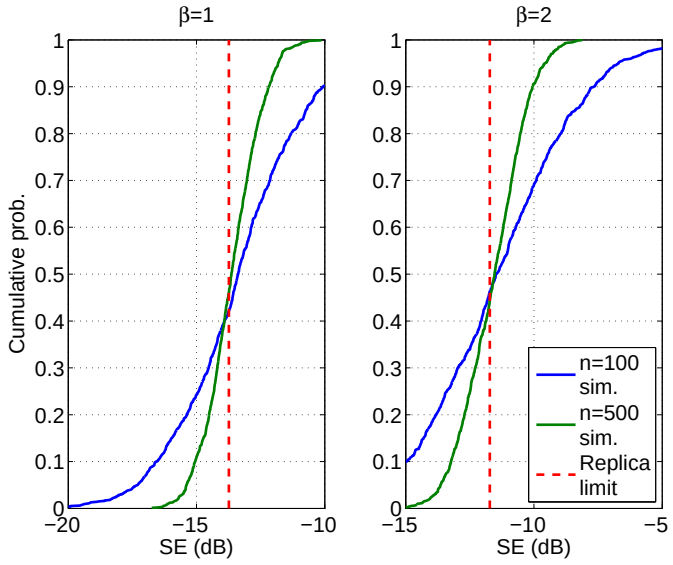

Fig. 4. Convergence to the asymptotic limit from the RS PMAP decoupling property. Plotted are the CDFs of the SE over 1000 Monte Carlo trials of the lasso method for the Gaussian mixture distribution. Details are in the text. The CDF is shown for dimensions $n = 100$ and $n = 500$ and $\beta = 1$ and 2. As vector dimension increases, the performance begins to concentrate around the limit predicted by the RS PMAP decoupling property.

theoretical MSE from the zero norm-regularized estimator as computed in Section V-C. For these two cases, the estimators cannot be simulated since they involve NP-hard computations. But we have depicted the curves to show that the replica method can be used to calculate the gap between practical and impractical algorithms. Interestingly, we see that there is about a 2.0 to 2.5 dB gap between lasso and zero norm-regularized estimation, and another 1 to 2 dB gap between zero norm-regularized estimation and optimal MMSE.

It is, of course, not surprising that zero norm-regularized estimation performs better than lasso for the strictly sparse prior considered in this simulation, and that optimal MMSE performs better yet. However, what is valuable is that replica analysis can quantify the precise performance differences.

In Fig. 3, we plotted the median SE since there is actually considerable variation in the SE over the random realizations of the problem parameters. To illustrate the degree of variability, Fig. 4 shows the CDF of the SE values over the 1000 Monte Carlo trials. Each trial has different noise and measurement matrix realizations, and both contribute to SE variations. We see that the variation of the SE is especially large at the smaller dimension $n = 100$. While the median value agrees well with the theoretical replica limit, any particular instance of the problem can vary considerably from that limit. This is a significant drawback of the replica method: at lower dimensions, the replica method may provide accurate predictions of the median behavior, but it does not bound the variations from the median.

As one might expect, at the higher dimension of $n = 500$, the level of variability is reduced and the observed SE begins to concentrate around the replica limit. In his original paper [12], Tanaka assumes that concentration of the SE will occur; he calls this the *self-averaging* assumption. Fig. 4 provides some empirical evidence that self-averaging does indeed occur. However, even at $n = 500$, the variation is not insignificant. As a result, caution should be exercised in using the replica predictions on particular low-dimensional instances.

B. Discrete Distribution with Dynamic Range

The RS PMAP decoupling property can also be used to study the effects of dynamic range in power levels. To validate the replica analysis with power variations, we ran the following experiment: the vector $\mathbf{x}$ was generated with i.i.d. components

$$x_j = \sqrt{s_j}\, u_j, \tag{53}$$

where s_j is a random power level and u_j is a discrete three-valued random variable with probability mass function

$$u_j \sim \begin{cases} 1/\sqrt{\rho}, & \text{with prob } = \rho/2; \\ -1/\sqrt{\rho}, & \text{with prob } = \rho/2; \\ 0, & \text{with prob } = 1-\rho. \end{cases} \tag{54}$$

As before, the parameter ρ represents the sparsity ratio and we chose a value of $\rho = 0.1$. The measurements were generated by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} = \mathbf{A}\mathbf{S}^{1/2}\mathbf{u} + \mathbf{w},$$

where $\mathbf{A}$ is an i.i.d. Gaussian measurement matrix and $\mathbf{w}$ is Gaussian noise. As in the previous section, the post-despreading SNR with side-information was normalized to 10 dB.

The factor s_j in (53) accounts for power variations in x_j. We considered two random distributions for s_j: (a) $s_j = 1$, so that the power level is constant; and (b) s_j is uniform (in dB scale) over a 10 dB range with unit average power.

In case (b), when there is variation in the power levels, we can analyze two different scenarios for the lasso estimator:

- *Power variations unknown:* If the power level s_j in (53) is unknown to the estimator, then we can apply the standard lasso estimator:

$$\widehat{\mathbf{x}}(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_1, \tag{55}$$

  which does not need knowledge of the power levels s_j. To analyze the behavior of this estimator with the replica method, we simply incorporate variations of both u_j and s_j into the prior of x_j and assume a constant scale factor s in the replica equations.
- *Power variations known:* If the power levels s_j are known, the estimator can compute

$$\widehat{\mathbf{u}}(\mathbf{y}) = \arg\min_{\mathbf{u}\in\mathbb{R}^n} \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{u}\|_2^2 + \|\mathbf{u}\|_1 \tag{56}$$

  and then take $\widehat{\mathbf{x}} = \mathbf{S}^{1/2}\widehat{\mathbf{u}}$. This can be analyzed with the replica method by incorporating the distribution of s_j into the scale factors.

Fig. 5 shows the performance of the lasso estimator for the different power range scenarios. As before, for each β, the figure plots the median SE over 1000 Monte Carlo simulation trials. Fig. 5 also shows the theoretical asymptotic performance

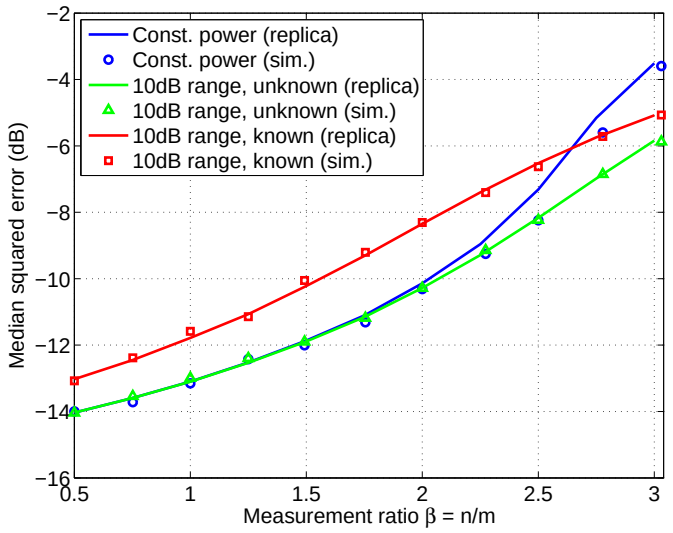

Fig. 5. MSE performance prediction by the replica method of the lasso estimator with power variations in the components. Plotted is the median SE of the lasso method in estimating a discrete-valued distribution. Three scenarios are considered: (a) all components have the same power; (b) the components have a 10 dB range in power that is unknown to the estimator; and (c) the power range is known to the estimator and incorporated into the measurement matrix. Solid lines represent the asymptotic prediction from the RS PMAP decoupling property, and the circles, triangles, and squares show the median SE over 1000 Monte Carlo simulation. See text for details.

as predicted with the RS PMAP decoupling property. Simulated values are based on a vector dimension of $n = 100$ and optimal selection of γ as described in Section V-D.

We see that in all three cases (constant power and power variations unknown and known to the estimator), the replica prediction is in excellent agreement with the simulated performance. With one exception, the replica method matches the simulated performance within 0.2 dB. The one exception is for $\beta = 2.5$ with constant power, where the replica method underpredicts the median SE by about 1 dB. A simulation at a higher dimension of $n = 500$ (not shown here) reduced this discrepancy to 0.2 dB, suggesting that the replica method is still asymptotically correct.

We can also observe two interesting phenomena in Fig. 5. First, the lasso method's performance with constant power is almost identical to the performance with unknown power variations for values of $\beta < 2$. However, at higher values of β, the power variations actually *improve* the performance of the lasso method, even though the average power is the same in both cases. Wainwright's analysis [44] demonstrated the significance of the minimum component power in dictating lasso's performance. The above simulation and the corresponding replica predictions suggest that dynamic range may also play a role in the performance of lasso. That increased dynamic range can improve the performance of sparse estimation has been observed for other estimators [62], [63].

A second phenomena we see in Fig. 5 is that knowing the power variations and incorporating them into the measurement matrix can actually degrade the performance of lasso. Indeed, knowing the power variations appears to result in a 1 to 2 dB loss in MSE performance.

Of course, one cannot conclude from this one simulation that these effects of dynamic range hold more generally. The study of the effect of dynamic range is interesting and beyond the scope of this work. The point is that the replica method provides a simple analytic method for quantifying the effect of dynamic range that appears to match actual performance well.

C. Support Recovery with Thresholding

In estimating vectors with strictly sparse priors, one important problem is to detect the *locations* of the nonzero components in the vector $\mathbf{x}$. This problem, sometimes called *support recovery*, arises for example in subset selection in linear regression [64], where finding the support of the vector $\mathbf{x}$ corresponds to determining a subset of features with strong linear influence on some observed data $\mathbf{y}$. Several works have attempted to find conditions under which the support of a sparse vector $\mathbf{x}$ can be fully detected [44], [56], [65] or partially detected [66]–[68]. Unfortunately, with the exception of [44], the only available results are bounds that are not tight.

One of the uses of RS PMAP decoupling property is to *exactly* predict the fraction of support that can be detected correctly. To see how to predict the support recovery performance, observe that the decoupling property provides the asymptotic joint distribution for the vector $(x_j, s_j, \hat{x}_j)$, where x_j is the component of the unknown vector, s_j is the corresponding scale factor and $\hat{x}_j$ is the component estimate. Now, in support recovery, we want to estimate θ_j, the indicator function that x_j is nonzero

$$\theta_j = \begin{cases} 1, & \text{if } x_j \neq 0; \\ 0, & \text{if } x_j \neq 0. \end{cases}$$

One natural estimate for θ_j is to compare the magnitude of the component estimate $\hat{x}_j$ to some scale-dependent threshold $t(s_j)$,

$$\widehat{\theta}_j = \begin{cases} 1, & \text{if } |\hat{x}_j| > t(s_j); \\ 0, & \text{if } |\hat{x}_j| \leq t(s_j). \end{cases}$$

This idea of using thresholding for sparsity detection has been proposed in [55] and [69]. Using the joint distribution $(x_j, s_j, \hat{x}_j)$, one can then compute the probability of sparsity misdetection

$$p_{\text{err}} = \Pr(\widehat{\theta}_j \neq \theta_j).$$

The probability of error can be minimized over the threshold levels $t(s)$.

To verify this calculation, we generated random vectors $\mathbf{x}$ with $n = 100$ i.i.d. components given by (53) and (54). We used a constant power ($s_j = 1$) and a sparsity fraction of $\rho = 0.2$. As before, the observations $\mathbf{y}$ were generated with an i.i.d. Gaussian matrix with SNR_0 = 10 dB.

Fig. 6 compares the theoretical probability of sparsity misdetection predicted by the replica method against the actual probability of misdetection based on the average of 1000 Monte Carlo trials. We tested two algorithms: linear MMSE estimation and lasso estimation. For lasso, the regularization parameter was selected for minimum MMSE as described in Section V-D. The results show a good match.

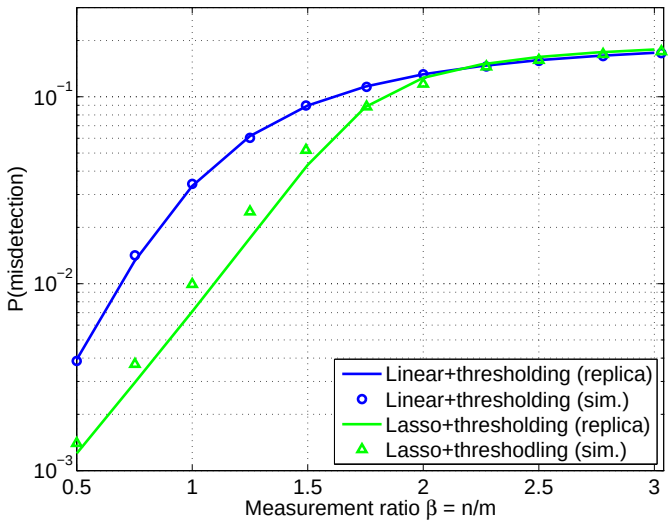

Fig. 6. Support recovery performance prediction with the replica method. The solid lines show the theoretical probability of error in sparsity misdetection using linear and lasso estimation followed by optimal thresholding. The circles and triangles are the corresponding mean probabilities of misdetection over 1000 Monte Carlo trials.

VII. Conclusions and Future Work

We have applied the replica method from statistical physics for computing the asymptotic performance of postulated MAP estimators of non-Gaussian vectors with large random linear measurements, under a replica symmetric assumption. The method can be readily applied to problems in compressed sensing. While the method is not theoretically rigorous, simulations show an excellent ability to predict the performance for a range of algorithms, performance metrics, and input distributions. Indeed, we believe that the replica method provides the only method to date for asymptotically-exact prediction of performance of compressed sensing algorithms that can apply in a large range of circumstances.

Moreover, we believe that the availability of a simple scalar model that exactly characterizes certain sparse estimators opens up numerous avenues for analysis. For one thing, it would be useful to see if the replica analysis of lasso can be used to recover the scaling laws of Wainwright [44] and Donoho and Tanner [45] for support recovery and to extend the latter to the noisy setting. Also, the best known bounds for MSE performance in sparse estimation are given by Haupt and Nowak [70] and Candès and Tao [71]. Since the replica analysis is asymptotically exact (subject to various assumptions), we may be able to obtain much tighter analytic expressions. In a similar vein, several researchers have attempted to find information-theoretic lower bounds with optimal estimation [56], [65], [72]. Using the replica analysis of optimal estimators, one may be able to improve these scaling laws as well.

Finally, there is a well-understood connection between statistical mechanics and belief propagation-based decoding of error correcting codes [6], [7]. These connections may suggest improved iterative algorithms for sparse estimation as well.

Appendix A
Review of the Replica Method

We provide a brief summary of the replica method, with a focus on some of the details of the replica symmetric analysis of postulated MMSE estimation in [12], [14]. This review will elucidate some of the key assumptions, notably the assumption of replica symmetry. General descriptions of the replica method can be found in texts such as [8]–[11].

The replica method is based on evaluating variants of the so-called *asymptotic free energy*

$$\mathcal{F} = -\lim_{n\to\infty} \frac{1}{n}\mathbf{E}\left[\log Z(\mathbf{y},\Phi)\right], \tag{57}$$

where $Z(\mathbf{y},\Phi)$ is the postulated partition function

$$Z(\mathbf{y},\Phi) = \mathbf{E}\left[\log p_{\mathbf{y}}(\mathbf{y} \mid \Phi\,;\, p_{\rm post}, \sigma^2_{\rm post})\right]$$

and the expectation in (57) is with respect to the true distribution on $\mathbf{y}$. For the replica PMMSE and PMAP analyses in [12], [14], various joint moments of the variables x_j and $\hat{x}_j$ are computed from certain variants of the free energy, and the convergence of the joint distribution of $(x_j, \hat{x}_j)$ is then analyzed based on these moments.

To evaluate the asymptotic free energy, the replica method uses the identity that, for any random variable Z,

$$\mathbf{E}[\log Z] = \lim_{\nu\to 0} \frac{\partial}{\partial \nu} \log \mathbf{E}\left[Z^\nu\right].$$

Therefore, the asymptotic free energy (57) can be rewritten as

$$\mathcal{F} = -\lim_{n\to\infty} \frac{1}{n} \lim_{\nu\to 0} \frac{\partial}{\partial \nu} \log \mathbf{E}\left[Z^\nu(\mathbf{y},\Phi)\right]. \tag{58}$$

The "replica trick" involves evaluating the expectation $\mathbf{E}[Z^\nu(\mathbf{y},\Phi)]$ for positive integer values of ν and then assuming an analytic continuation so that the resulting expression is valid for real ν in the vicinity of zero. For positive integer values of ν, the quantity $Z^\nu(\mathbf{y},\Phi)$ can be written as

$$Z^\nu(\mathbf{y},\Phi) = \mathbf{E}\left[\prod_{a=1}^{\nu} p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x}_a, \Phi\,;\, p_{\rm post}, \sigma^2_{\rm post})\right], \tag{59}$$

where the expectation is over independent copies of the vectors $\mathbf{x}_a$, $a = 1,\ldots,\nu$, with i.i.d. components $x_{aj} \sim p_{\rm post}(x_{aj})$. The motivation for the replica trick is that the quantity $Z^\nu(\mathbf{y},\Phi)$ in (59) can be thought of as a partition function of a new system with ν "replicated" copies of the variables $\mathbf{x}_a$, $a = 1,\ldots,\nu$. The parameter ν is called the replica number.

The replicated system is relatively easy to analyze. Specifically, to evaluate $\mathbf{E}[Z^\nu(\mathbf{y},\Phi)]$, the replica analysis in [12], [14] first assumes a *self-averaging* property that essentially assumes that the variations in $Z^\nu(\mathbf{y},\Phi)$ due to randomness of the measurement matrix Φ vanish in the limit as $n \to \infty$. Although a large number of statistical physics quantities exhibit such self-averaging, the self-averaging of the relevant quantities for the general PMMSE and PMAP analyses has not been rigorously established. Following [12], [14], self-averaging in this work is thus simply assumed.

Under the self-averaging assumption, the expectation in (59) is evaluated in [14] by first conditioning on the $(\nu+1)$-by-$(\nu+1)$ correlation matrix $\mathbf{Q} = (1/n)\mathbf{X}^T\mathbf{X}$, where $\mathbf{X}$ is the

n-by-$(\nu+1)$ matrix

$$\mathbf{X} = [\mathbf{x} \;\; \mathbf{x}_1 \;\; \ldots \;\; \mathbf{x}_\nu],$$

with $\mathbf{x}$ having i.i.d. components according to the true distribution $x_j \sim p_0(x_j)$ and the vectors $\mathbf{x}_a$ being independent with i.i.d. components following the postulated distribution $x_{aj} \sim p_{\rm post}(x_{aj})$. The conditioning on $\mathbf{Q}$ reduces the expectation in (59) to an integral of the form

$$\begin{aligned}\frac{1}{n}&\mathbf{E}[Z^\nu(\mathbf{y},\Phi)] \\ &= \frac{1}{n}\log\int \exp\left(\frac{n}{\beta}G^{(\nu)}(\mathbf{Q})\right)\mu_n^{(\nu)}(d\mathbf{Q}) + O\left(\frac{1}{n}\right) \qquad (60)\end{aligned}$$

where $G^{(\nu)}(\mathbf{Q})$ is some function of the correlation matrix $\mathbf{Q}$ and $\mu_n^{(\nu)}(\mathbf{Q})$ is a probability measure on $\mathbf{Q}$. It is then argued that the measures $\mu_n^\nu(\mathbf{Q})$ satisfy a large deviations property with some rate function $I^\nu(\mathbf{Q})$. Then, using standard large deviations arguments as in [73], the asymptotic value of the expectation in (60) reduces to a maximization of the form

$$\lim_{n\to\infty}\mathbf{E}[Z^\nu(\mathbf{y},\Phi)] = \sup_{\mathbf{Q}}\left[\frac{1}{\beta}G^{(\nu)}(\mathbf{Q}) - I^\nu(\mathbf{Q})\right], \qquad (61)$$

where the supremum is over the set of covariance matrices $\mathbf{Q}$. The correlation matrix $\mathbf{Q}$ plays a similar role as the so-called overlap matrix in replica analyses of systems with discrete energy states [10].

The maximization in (61) over all covariance matrices is, in general, difficult to perform. The key replica symmetry (RS) assumption used in [12] and [14], and hence implicitly used in this paper, is that the maxima are achieved with matrices $\mathbf{Q}$ that are symmetric with respect to permutations of the ν replica indices. Under this symmetry assumption, the space of covariance matrices is greatly reduced and the maxima (61) can be explicitly evaluated.

The RS assumption is not always valid, even though the system itself is symmetric across the replica indices. For example, it is well-known that even in the simple random energy model, the corresponding maximization may not satisfy the RS assumption, particularly at low temperatures [10]; see, also [74]. More recently, it has been shown that replica symmetry may also be broken when analyzing lattice precoding for the Gaussian broadcast channel [15].

In absence of replica symmetry, one must search through a larger class of overlap or covariance matrices $\mathbf{Q}$. One such hierarchy of classes of matrices that is often used is described by the so-called k-step replica symmetry breaking (RSB) matrices, a description of which can be found in various texts [8]–[11]. In this regard, the analysis in this paper, which assumes replica symmetry, is thus only a 0-step RSB analysis or 0th-level prediction.

It is difficult to derive general tests for whether the RS assumption is rigorously valid. Tanaka's original work [12] derived an explicit condition for the validity of the RS assumption based on the Almeida–Thouless (AT) test [75] that considers asymmetric perturbations around the RS saddle points of the maximization (61). For the case of binary signals, the condition has a simple formula with the SNR and measurement ratio β. In [48], an AT condition was also derived for RS analysis of ℓ_p reconstruction with Bernoulli–Gaussian priors. Unfortunately, no equivalent condition has been derived for the general scenario considered in Guo and Verdú's extension in [14].

In this work, we simply assume replica symmetry for the all values of the scale factor $u > 0$. Since u is analogous to inverse temperature [54] and validity of the RS assumption is more problematic at low temperatures, one must be cautious in interpreting our results. As stated in Section I, where possible we have confirmed the replica predictions by comparison to numerical experiments. However, such experiments are limited to computable estimators such as LASSO and linear estimators. For other estimators, such as the true MMSE or zero norm-regularized estimator, the RS assumption may very well not hold.

Appendix B
Proof Overview

Fix a deterministic sequence of indices $j = j(n)$ with $j(n) \in \{1,\ldots,n\}$. For each n, define the random vector triples

$$\theta^u(n) = (x_j(n), s_j(n), \hat{x}_j^u(n)), \qquad (62a)$$
$$\theta^{\rm map}(n) = (x_j(n), s_j(n), \hat{x}_j^{\rm pmap}(n)), \qquad (62b)$$

where $x_j(n)$, $\hat{x}_j^u(n)$, and $\hat{x}_j^{\rm pmap}(n)$ are the jth components of the random vectors $\mathbf{x}$, $\widehat{\mathbf{x}}^u(\mathbf{y})$, and $\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y})$, and $s_j(n)$ is the jth diagonal entry of the matrix $\mathbf{S}$.

For each u, we will use the notation

$$\hat{x}^u_{\rm scalar}(z\,;\,\lambda) = \hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,p_u, \lambda/u), \qquad (63)$$

where p_u is defined in (21) and $\hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,\cdot,\cdot)$ is defined in (10). Also, for every σ and $\gamma > 0$ define the random vectors

$$\theta^u_{\rm scalar}(\sigma^2,\gamma) = (x, s, \hat{x}^u_{\rm scalar}(z\,;\,\gamma/s)), \qquad (64a)$$
$$\theta^{\rm map}_{\rm scalar}(\sigma^2,\gamma) = (x, s, \hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\gamma/s)), \qquad (64b)$$

where x and s are independent with $x \sim p_0(x)$, $s \sim p_S(s)$, and

$$z = x + \frac{\sigma}{\sqrt{s}}v \qquad (65)$$

with $v \sim \mathcal{N}(0,1)$.

Now, to prove the RS PMAP decoupling property, we need to show that (under the stated assumptions)

$$\lim_{n\to\infty}\theta^{\rm map}(n) = \theta^{\rm map}_{\rm scalar}(\sigma^2_{\rm eff,map}, \gamma_p), \qquad (66)$$

where the limit is in distribution and the noise levels $\sigma^2_{\rm eff,map}$ and γ_p satisfy part (b) of the claim. This desired equivalence is depicted in the right column of Fig. 7.

To show this limit we first observe that under Assumption 1, for u sufficiently large, the postulated prior distribution $p_u(x)$ in (21) and noise level σ_u^2 in (22) are assumed to satisfy the RS PMMSE decoupling property. This implies that

$$\begin{aligned}\lim_{n\to\infty}&(x_j(n), s_j(n), \hat{x}_j^u(n)) \\ &= (x, s, \hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,p_u, \sigma^2_{\rm p-eff}(u)/s)), \qquad (67)\end{aligned}$$

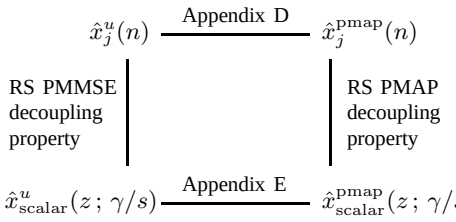

Fig. 7. The RS PMAP decoupling property of this paper relates $\hat{x}_j^{\rm pmap}(n)$ to $\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\gamma/s)$ through an $n\to\infty$ limit. We establish the equivalence of its validity to the validity of the RS PMMSE decoupling property [14] through two $u\to\infty$ limits: Appendix D relates $\hat{x}_j^u(n)$ and $\hat{x}_j^{\rm pmap}(n)$; Appendix E relates $\hat{x}^u_{\rm scalar}(z\,;\,\gamma/s)$ and $\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\gamma/s)$.

where the limit is in distribution, $x\sim p_0(x)$, $s\sim p_S(s)$, and

$$z = x + \frac{\sigma_{\rm eff}(u)}{\sqrt{s}}v, \qquad v\sim\mathcal{N}(0,1).$$

Using the notation above, we can rewrite this limit as

$$\begin{aligned}\lim_{n\to\infty}\theta^u(n) &\overset{(a)}{=} \lim_{n\to\infty}(x_j(n),s_j(n),\hat{x}_j^u(n))\\ &\overset{(b)}{=} (x,s,\hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,p_u,\sigma^2_{\rm p-eff}(u)/s))\\ &\overset{(c)}{=} (x,s,\hat{x}^u_{\rm scalar}(z\,;\,u\sigma^2_{\rm p-eff}(u)/s))\\ &\overset{(d)}{=} \theta^u_{\rm scalar}(\sigma^2_{\rm eff}(u),u\sigma^2_{\rm p-eff}(u)),\end{aligned} \tag{68}$$

where all the limits are in distribution and (a) follows from the definition of $\theta^u(n)$ in (62a); (b) follows from (67); (c) follows from (63); and (d) follows from (64a). This equivalence is depicted in the left column of Fig. 7.

The key part of the proof is to use a large deviations argument to show that for almost all $\mathbf{y}$,

$$\lim_{u\to\infty}\widehat{\mathbf{x}}^u(\mathbf{y}) = \hat{\mathbf{x}}^{\rm pmap}(\mathbf{y}).$$

This limit in turn shows (see Lemma 5 of Appendix D) that for every n,

$$\lim_{u\to\infty}\theta^u(n) = \theta^{\rm map}(n) \tag{69}$$

almost surely and in distribution. A large deviation argument is also used to show that for every λ and almost all z,

$$\lim_{u\to\infty}\hat{x}^u_{\rm scalar}(z\,;\,\lambda) = \hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda).$$

Combining this with the limits in Assumption 2, we will see (see Lemma 7 of Appendix E) that

$$\begin{aligned}&\lim_{u\to\infty}\theta^u_{\rm scalar}(\sigma^2_{\rm eff}(u),u\sigma^2_{\rm p-eff}(u))\\ &\quad= \theta^{\rm map}_{\rm scalar}(\sigma^2_{\rm eff,map},\gamma_p)\end{aligned} \tag{70}$$

almost surely and in distribution.

The equivalences (69) and (70) are shown as rows in Fig. 7. As shown, they combine with the RS PMMSE decoupling property to prove the RS PMAP decoupling property. In equations instead of diagrammatic form, the combination of limits is

$$\begin{aligned}\lim_{n\to\infty}\theta^{\rm map}(n) &\overset{(a)}{=} \lim_{n\to\infty}\lim_{u\to\infty}\theta^u(n)\\ &\overset{(b)}{=} \lim_{u\to\infty}\lim_{n\to\infty}\theta^u(n)\\ &\overset{(c)}{=} \lim_{u\to\infty}\theta^u_{\rm scalar}(\sigma^2_{\rm eff}(u),u\sigma^2_{\rm p-eff}(u))\\ &\overset{(d)}{=} \theta^{\rm map}_{\rm scalar}(\sigma^2_{\rm eff,map},\gamma_p)\end{aligned}$$

where all the limits are in distribution and (a) follows from (69); (b) follows from Assumption 3; (c) follows from (68); and (d) follows from (70). This proves (66) and part (a) of the claim.

Therefore, to prove the claim we prove the limit (69) in Appendix D and the limit (70) in Appendix E and show that the limiting noise levels $\sigma^2_{\rm eff,map}$ and γ_p satisfy the fixed-point equations in part (b) of the claim in Appendix F. Before these results are given, we review in Appendix C some requisite results from large deviations theory.

Appendix C
Large Deviations Results

The above proof overview shows that the RS predictions for the postulated MAP estimate are calculated by taking the limit as $u\to\infty$ of the RS predictions of the postulated MMSE estimates. These limits are evaluated with large deviations theory and we begin, in this appendix, by deriving some simple modifications of standard large deviations results. The main result we need is Laplace's principle as described in [73]:

Lemma 1 (Laplace's Principle): Let $\varphi(\mathbf{x})$ be any measurable function defined on some measurable subset $\mathcal{D}\subseteq\mathbb{R}^n$ such that

$$\int_{\mathbf{x}\in\mathcal{D}}\exp(-\varphi(\mathbf{x}))\,d\mathbf{x} < \infty. \tag{71}$$

Then

$$\lim_{u\to\infty}\frac{1}{u}\log\int_{\mathbf{x}\in\mathcal{D}}\exp(-u\varphi(\mathbf{x}))\,d\mathbf{x} = -\operatorname*{ess\,inf}_{\mathbf{x}\in\mathcal{D}}\varphi(\mathbf{x}).$$

Given $\varphi(\mathbf{x})$ as in Lemma 1, define the probability distribution

$$q_u(\mathbf{x}) = \left[\int_{\mathbf{x}\in\mathcal{D}}\exp(-u\varphi(\mathbf{x}))\,d\mathbf{x}\right]^{-1}\exp(-u\varphi(\mathbf{x})). \tag{72}$$

We want to evaluate expectations of the form

$$\lim_{u\to\infty}\int_{\mathbf{x}\in\mathcal{D}}g(u,\mathbf{x})q_u(\mathbf{x})\,d\mathbf{x}$$

for some real-valued measurable function $g(u,\mathbf{x})$. The following lemma shows that this integral is described by the behavior of $g(u,\mathbf{x})$ in a neighborhood of the minimizer of $\varphi(\mathbf{x})$.

Lemma 2: Suppose that $\varphi(\mathbf{x})$ and $g(u,\mathbf{x})$ are real-valued measurable functions such that:

(a) The function $\varphi(\mathbf{x})$ satisfies (71) and has a unique essential minimizer $\widehat{\mathbf{x}}\in\mathbb{R}^n$ such that for every open neighborhood U of $\widehat{\mathbf{x}}$,

$$\inf_{\mathbf{x}\notin U}\varphi(\mathbf{x}) > \varphi(\widehat{\mathbf{x}}).$$

(b) The function $g(u,\mathbf{x})$ is positive and satisfies

$$\limsup_{u\to\infty}\sup_{\mathbf{x}\notin U}\frac{\log g(u,\mathbf{x})}{u(\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}}))}\leq 0$$

for every open neighborhood U of $\widehat{\mathbf{x}}$.

(c) There exists a constant g_∞ such that for every $\epsilon>0$, there exists a neighborhood U of $\hat{x}$ such that

$$\limsup_{u\to\infty}\left|\int_U g(u,\mathbf{x})q_u(\mathbf{x})\,dx-g_\infty\right|\leq\epsilon.$$

Then,

$$\lim_{u\to\infty}\int g(u,\mathbf{x})q_u(\mathbf{x})\,d\mathbf{x}=g_\infty.$$

Proof: Due to item (c), we simply have to show that for any open neighborhood U of $\widehat{\mathbf{x}}$,

$$\limsup_{u\to\infty}\int_{\mathbf{x}\in U^c}g(u,\mathbf{x})q_u(\mathbf{x})\,d\mathbf{x}=0.$$

To this end, let

$$Z(u)=\log\int_{\mathbf{x}\in U^c}g(u,\mathbf{x})q_u(\mathbf{x})\,d\mathbf{x}.$$

It suffices to show that $Z(u)\to-\infty$ as $u\to\infty$. Using the definition of $q_u(\mathbf{x})$ in (72), it is easy to check that

$$Z(u)=Z_1(u)-Z_2(u), \tag{73}$$

where

$$\begin{aligned}Z_1(u) &= \log\int_{\mathbf{x}\in U^c}g(u,\mathbf{x})\exp\left(-u(\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}}))\right)\,d\mathbf{x},\\ Z_2(u) &= \log\int_{\mathbf{x}\in\mathcal{D}}\exp\left(-u(\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}}))\right)\,d\mathbf{x}.\end{aligned}$$

Now, let

$$M=\operatorname*{ess\,inf}_{\mathbf{x}\in U^c}\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}}).$$

By item (a), $M>0$. Therefore, we can find a $\delta>0$ such that

$$-M(1-\delta)+3\delta<0. \tag{74}$$

Now, from item (b), there exists a u_0 such that for all $u>u_0$,

$$Z_1(u) \leq \log\int_{\mathbf{x}\in U^c}\exp(-u(1-\delta)(\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}})))\,d\mathbf{x}.$$

By Laplace's principle, we can find a u_1 such that for all $u>u_1$,

$$\begin{aligned}Z_1(u) &\leq u\left[\delta-\inf_{\mathbf{x}\in U^c}(1-\delta)(\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}}))\right]\\ &= u(-M(1-\delta)+\delta).\end{aligned} \tag{75}$$

Also, since $\widehat{\mathbf{x}}$ is an essential minimizer of $\varphi(\mathbf{x})$,

$$\operatorname*{ess\,inf}_{\mathbf{x}\in\mathcal{D}}\varphi(\mathbf{x})=\varphi(\widehat{\mathbf{x}}).$$

Therefore, by Laplace's principle, there exists a u_2 such that for $u>u_2$,

$$Z_2(u)\geq u\left[-\delta-\operatorname*{ess\,inf}_{\mathbf{x}\in\mathcal{D}}(\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}}))\right]=-u\delta. \tag{76}$$

Substituting (75) and (76) into (73) we see that for u sufficiently large,

$$Z(u)\leq u(-M(1-\delta)+\delta)+u\delta<-u\delta,$$

where the last inequality follows from (74). This shows $Z(u)\to-\infty$ as $u\to\infty$ and the proof is complete. ■

One simple application of this lemma is as follows:

Lemma 3: Let $\varphi(\mathbf{x})$ and $h(\mathbf{x})$ be real-valued measurable functions satisfying the following:

(a) The function $\varphi(\mathbf{x})$ has a unique essential minimizer $\widehat{\mathbf{x}}$ such that for every open neighborhood U of $\widehat{\mathbf{x}}$,

$$\inf_{\mathbf{x}\notin U}\varphi(\mathbf{x})>\varphi(\widehat{\mathbf{x}}).$$

(b) The function $h(\mathbf{x})$ is continuous at $\widehat{\mathbf{x}}$.

(c) There exists a $c>0$ and compact set K such that for all $\mathbf{x}\notin K$,

$$\varphi(\mathbf{x})\geq c\log|h(\mathbf{x})|. \tag{77}$$

Then,

$$\lim_{u\to\infty}\int h(\mathbf{x})q_u(\mathbf{x})\,d\mathbf{x}=h(\widehat{\mathbf{x}}).$$

Proof: We will apply Lemma 2 with $g(u,\mathbf{x})=|h(\mathbf{x})-h(\widehat{\mathbf{x}})|$ and $g_\infty=0$. Item (a) of this lemma shows that $\varphi(\mathbf{x})$ satisfies item (a) in Lemma 2.

To verify that item (b) of Lemma 2 holds, we first claim there exists a constant $M>0$ such that for all $\mathbf{x}$,

$$\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}})\geq M\log|h(\mathbf{x})-h(\widehat{\mathbf{x}})|. \tag{78}$$

We find a valid constant M for three regions. First, let U be the set of $\mathbf{x}$ such that $|h(\mathbf{x})-h(\widehat{\mathbf{x}})|<1$. Since $h(\mathbf{x})$ is continuous in $\mathbf{x}$, U is an open neighborhood of $\widehat{\mathbf{x}}$. Also, for $\mathbf{x}\in U$, the right hand side of (78) is negative. Since the left-hand side of (78) is positive, the inequality will be satisfied in U for any $M>0$.

Next, consider the set $K_1=K\backslash U$ where K is the compact set in item (c) of this lemma. Since K is compact and $h(\mathbf{x})$ is continuous, there exists a $c_1>0$ such that $\log|h(\mathbf{x})-h(\widehat{\mathbf{x}})|<c_1$ for all $\mathbf{x}\in K$. Also, since U is an open neighborhood of $\widehat{\mathbf{x}}$, by item (a), there exists a $c_2>0$ such that $\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}})\geq c_2$ for all $\mathbf{x}\notin U$. Hence, the inequality (78) is satisfied with $M=c_2/c_1$ in the set K_1.

Finally, consider the set K^c. In this set, (77) is satisfied for some $c>0$. Combining this inequality with the fact that $\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}})\geq c_2$ for some $c_2>0$, one can show that (78) also holds for some $M>0$. Hence, for each of the regions U, $K\backslash U$ and K^c, (78) is satisfied for some $M>0$. Taking the maximum of the three values of M, one can assume (78) for all $\mathbf{x}$.

Applying (78), we obtain

$$\frac{\log g(u,\mathbf{x})}{\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}})}=\frac{\log|h(\mathbf{x})-h(\widehat{\mathbf{x}})|}{\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}})}\leq\frac{1}{M}.$$

Hence, for any open neighborhood U of $\widehat{\mathbf{x}}$,

$$\limsup_{u\to\infty}\sup_{\mathbf{x}\notin U}\frac{\log g(u,\mathbf{x})}{u(\varphi(\mathbf{x})-\varphi(\widehat{\mathbf{x}}))}\leq\lim_{u\to\infty}\frac{1}{uM}=0.$$

Now let us verify that item (c) of Lemma 2 holds. Let $\epsilon>0$. Since $h(\mathbf{x})$ is continuous at $\mathbf{x}$, there exists an open neighborhood U of $\mathbf{x}$ such that $g(u,\mathbf{x})<\epsilon$ for all $\mathbf{x}\in U$ and u. This implies that for all u,

$$\int_U g(u,\mathbf{x})q_u(\mathbf{x})\,d\mathbf{x}<\epsilon\int_U q_u(\mathbf{x})\,d\mathbf{x}\leq\epsilon,$$

which shows that $g(u,\mathbf{x})$ satisfies item (c) of Lemma 2. Thus

$$\begin{aligned}\left|\int h(\mathbf{x})q_u(\mathbf{x})\,d\mathbf{x} - h(\widehat{\mathbf{x}})\right| &= \left|\int (h(\mathbf{x})-h(\widehat{\mathbf{x}}))q_u(\mathbf{x})\,d\mathbf{x}\right| \\ &\leq \int |h(\mathbf{x})-h(\widehat{\mathbf{x}})|q_u(\mathbf{x})\,d\mathbf{x} \\ &\leq \int g(u,\mathbf{x})q_u(\mathbf{x})\,d\mathbf{x} \to 0,\end{aligned}$$

where the last limit is as $u\to\infty$ and follows from Lemma 2. ■

APPENDIX D
EVALUATION OF $\lim_{u\to\infty}\widehat{\mathbf{x}}^u(\mathbf{y})$

We can now apply Laplace's principle in the previous section to prove (69). We begin by examining the pointwise convergence of the PMMSE estimator $\widehat{\mathbf{x}}^u(\mathbf{y})$.

Lemma 4: For every n, $\mathbf{A}$, and $\mathbf{S}$ and almost all $\mathbf{y}$,

$$\lim_{u\to\infty}\widehat{\mathbf{x}}^u(\mathbf{y}) = \hat{\mathbf{x}}^{\rm pmap}(\mathbf{y}),$$

where $\widehat{\mathbf{x}}^u(\mathbf{y})$ is the PMMSE estimator in (24) and $\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y})$ is the PMAP estimator in (18).

Proof: The lemma is a direct application of Lemma 3. Fix n, $\mathbf{y}$, $\mathbf{A}$, and $\mathbf{S}$ and let

$$\varphi(\mathbf{x}) = \frac{1}{2\lambda}\|\mathbf{y}-\mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|^2 + f(\mathbf{x}). \tag{79}$$

The definition of $\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y})$ in (18) shows that

$$\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathcal{X}^n}\varphi(\mathbf{x}).$$

Assumption 4 shows that this minimizer is unique for almost all $\mathbf{y}$. Also (23) shows that

$$\begin{aligned}&p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\mid\mathbf{y}\,;\,p_u,\sigma_u^2) \\ &\quad= \left[\int_{\mathbf{x}\in\mathcal{X}^n}\exp\left(-u\varphi(\mathbf{x})\right)\,d\mathbf{x}\right]^{-1}\exp(-u\varphi(\mathbf{x})) \\ &\quad= q_u(\mathbf{x}),\end{aligned}$$

where $q_u(\mathbf{x})$ is given in (72) with $\mathcal{D}=\mathcal{X}^n$. Therefore, using (24),

$$\widehat{\mathbf{x}}^u(\mathbf{y}) = \mathbf{E}\left(\mathbf{x}\mid\mathbf{y}\,;\,p_u,\sigma_u^2\right) = \int_{\mathbf{x}\in\mathcal{X}^n}\mathbf{x}\,q_u(\mathbf{x})\,d\mathbf{x}. \tag{80}$$

Now, to prove the lemma, we need to show that

$$\lim_{n\to\infty}\hat{x}_j^u(\mathbf{y}) = \hat{x}_j^{\rm pmap}(\mathbf{y})$$

for every component $j=1,\ldots,n$. To this end, fix a component index j. Using (80), we can write the jth component of $\widehat{\mathbf{x}}^u(\mathbf{y})$ as

$$\hat{x}_j^u(\mathbf{y}) = \int_{\mathbf{x}\in\mathcal{X}^n}h(\mathbf{x})q_u(\mathbf{x})\,d\mathbf{x},$$

where $h(\mathbf{x})=x_j$. The function $h(\mathbf{x})$ is continuous. To verify item (c) of Lemma 3, using Assumption 5, we first find a compact set K such that $|\mathbf{x}|\notin K$ implies that

$$f(x_j) > c\log|x_j|. \tag{81}$$

Then, for $\mathbf{x}\notin K$,

$$\varphi(\mathbf{x}) \overset{(a)}{\geq} f(\mathbf{x}) \overset{(b)}{\geq} f(x_j) \overset{(c)}{\geq} c\log|x_j|,$$

where (a) follows from the definition of $\varphi(\mathbf{x})$ in (79); (b) follows from (20) and the assumption that the cost functions $f(x_i)$ are non-negative; and (c) follows from (81). Therefore, item (c) of Lemma 3 follows since $h(x_j)=x_j$. Thus, all the hypotheses of Lemma 3 are satisfied and we have the limit

$$\lim_{u\to\infty}\hat{x}_j^u(\mathbf{y}) = h(\hat{\mathbf{x}}^{\rm pmap}(\mathbf{y})) = \hat{x}_j^{\rm pmap}(\mathbf{y}).$$

This proves the lemma. ■

Lemma 5: Consider the random vectors $\theta^u(n)$ and $\theta^{\rm map}(n)$ defined in (62a) and (62b), respectively. Then, for all n,

$$\lim_{u\to\infty}\theta^u(n) = \theta^{\rm map}(n) \tag{82}$$

almost surely and in distribution.

Proof: The vectors $\theta^u(n)$ and $\theta^{\rm map}(n)$ are deterministic functions of $\mathbf{x}(n)$, $\mathbf{A}(n)$, $\mathbf{S}(n)$, and $\mathbf{y}$. Lemma 4 shows that the limit (82) holds for any values of $\mathbf{x}(n)$, $\mathbf{A}(n)$, and $\mathbf{S}(n)$, and almost all $\mathbf{y}$. Since $\mathbf{y}$ has a continuous probability distribution (due to the additive noise $\mathbf{w}$ in (3)), the set of values where this limit does not hold must have probability zero. Thus, the limit (82) holds almost surely, and therefore, also in distribution. ■

APPENDIX E
EVALUATION OF $\lim_{u\to\infty}\hat{x}^u_{\rm scalar}(z\,;\,\lambda)$

We first show the pointwise convergence of the scalar MMSE estimator $\hat{x}^u_{\rm scalar}(z\,;\,\lambda)$.

Lemma 6: Consider the scalar estimators $\hat{x}^u_{\rm scalar}(z\,;\,\lambda)$ defined in (63) and $\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)$ defined in (25). For all $\lambda>0$ and almost all z, we have the deterministic limit

$$\lim_{u\to\infty}\hat{x}^u_{\rm scalar}(z\,;\,\lambda) = \hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda).$$

Proof: The proof is similar to that of Lemma 4. Fix z and λ and consider the conditional distribution $p_{x|z}(x\mid z\,;\,p_u,\lambda/u)$. Using (7) along with the definition of $p_u(x)$ in (21) and an argument similar to the proof of Lemma 4, it is easily checked that

$$p_{x|z}(x\mid z\,;\,p_u,\lambda/u) = q_u(x), \tag{83}$$

where $q_u(x)$ is given by (72) with $\mathcal{D}=\mathcal{X}$ and

$$\varphi(x) = F(x,z,\lambda), \tag{84}$$

where $F(x,z,\lambda)$ is defined in (26). Using (63) and (10),

$$\begin{aligned}\hat{x}^u_{\rm scalar}(z\,;\,\lambda) &= \hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,p_u,\lambda/u) \\ &= \int_{x\in\mathcal{X}} x\,p_{x|z}(x\mid z\,;\,p_u,\lambda/u)\,dx \\ &= \int_{x\in\mathcal{X}} h(x)q_u(x)\,dx,\end{aligned}$$

with $h(x)=x$.

We can now apply Lemma 3. The definition of $\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)$ in (25) shows that

$$\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda) = \arg\min_{x\in\mathcal{X}}\varphi(x). \tag{85}$$

Assumption 6 shows that for all $\lambda > 0$ and almost all z, this minimization is unique so

$$\varphi(x) > \varphi(\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda))$$

for all $x \neq \hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)$. Also, using (26),

$$\begin{aligned}\lim_{|x|\to\infty} \varphi(x) &\overset{(a)}{=} \lim_{|x|\to\infty} F(x,z,\lambda) \\ &\overset{(b)}{\geq} \lim_{|x|\to\infty} f(x) \overset{(c)}{=} \infty \end{aligned} \tag{86}$$

where (a) follows from (84); (b) follows from (26); and (c) follows from Assumption 5. Equations (85) and (86) show that item (a) of Lemma 3 is satisfied. Item (b) of Lemma 3 is also clearly satisfied since $h(x) = x$ is continuous.

Also, similar to the proof of Lemma 4, one can show using Assumption 5 that item (c) of Lemma 3 is satisfied for some $c > 0$. Thus, all the hypotheses of Lemma 3 are satisfied and we have the limit

$$\lim_{u\to\infty} \hat{x}^u_{\rm scalar}(z\,;\,\lambda) = h(\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)) = \hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda).$$

This proves the lemma. ∎

We now turn to convergence of the random variable $\theta^u_{\rm scalar}(\sigma^2_{\rm eff}(u), u\sigma^2_{\rm p-eff}(u))$.

Lemma 7: Consider the random vectors $\theta^u_{\rm scalar}(\sigma^2,\gamma)$ defined in (64a) and $\theta^{\rm map}_{\rm scalar}(\sigma^2,\gamma)$ in (64b). Let $\sigma^2_{\rm eff}(u)$, $\sigma^2_{\rm p-eff}(u)$, $\sigma^2_{\rm eff,map}$ and γ_p be as defined in Assumption 2. Then the following limit holds:

$$\lim_{u\to\infty} \theta^u_{\rm scalar}(\sigma^2_{\rm eff}(u), u\sigma^2_{\rm p-eff}(u)) = \theta^{\rm map}_{\rm scalar}(\sigma^2_{\rm eff,map},\gamma_p) \tag{87}$$

almost surely and in distribution.

Proof: The proof is similar to that of Lemma 5. For any σ^2 and $\gamma > 0$, the vectors $\theta^u_{\rm scalar}(\sigma^2,\gamma)$ and $\theta^{\rm map}_{\rm scalar}(\sigma^2,\gamma)$ are deterministic functions of the random variables $x \sim p_0(x)$, $s \sim p_S(s)$, and z given (65) with $v \sim \mathcal{N}(0,1)$. Lemma 6 shows that the limit

$$\lim_{u\to\infty} \theta^u_{\rm scalar}(\sigma^2,\gamma) = \theta^{\rm map}_{\rm scalar}(\sigma^2,\gamma) \tag{88}$$

holds for any values of σ^2, γ, x, and s and almost all z. Also, if we fix x, s, and v, by Assumption 6, the function

$$\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\gamma/s) = \hat{x}^{\rm pmap}_{\rm scalar}(x + \frac{\sigma}{\sqrt{s}}v\,;\,\gamma/s)$$

is continuous in γ and σ^2 for almost all values of v. Therefore, we can combine (88) with the limits in Assumption 2 to show that

$$\lim_{u\to\infty} \theta^u_{\rm scalar}(\sigma^2_{\rm eff}(u), u\sigma^2_{\rm p-eff}(u)) = \theta^{\rm map}_{\rm scalar}(\sigma^2_{\rm eff,map},\gamma_p)$$

for almost all x and s and almost all z. Since z has a continuous probability distribution (due to the additive noise v in (65)), the set of values where this limit does not hold must have probability zero. Thus, the limit (87) holds almost surely, and therefore, also in distribution. ∎

APPENDIX F
PROOF OF THE FIXED-POINT EQUATIONS

For the final part of the proof, we need to show that the limits $\sigma^2_{\rm eff,map}$ and γ_p in Assumption 2 satisfy the fixed-point equations (30). The proof is straightforward, but we just need to keep track of the notation properly. We begin with the following limit.

Lemma 8: The following limit holds:

$$\begin{aligned}\lim_{u\to\infty} &\mathbf{E}\left[s\,\mathsf{mse}(p_u,p_0,\mu^u_p,\mu^u,z^u)\right] \\ &= \mathbf{E}\left[s|x-\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)|^2\right],\end{aligned}$$

where the expectations are taken over $x \sim p_0(x)$ and $s \sim p_S(s)$, and z and z^u are the random variables

$$z^u = x + \sqrt{\mu^u}v, \tag{89a}$$

$$z = x + \sqrt{\mu}v, \tag{89b}$$

with $v \sim \mathcal{N}(0,1)$ and $\mu^u = \sigma^2_{\rm eff}(u)/s$, $\mu^u_p = \sigma^2_{\rm p-eff}(u)/s$, $\mu = \sigma^2_{\rm eff,map}/s$, and $\lambda = \gamma_p/s$.

Proof: Using the definitions of mse in (11) and $\hat{x}^u_{\rm scalar}(z\,;\,\cdot)$ in (63),

$$\begin{aligned}&\mathsf{mse}(p_u,p_0,\mu^u_p,\mu^u,z^u) \\ &= \int_{x\in\mathcal{X}} |x-\hat{x}^{\rm pmmse}_{\rm scalar}(z^u\,;\,p_u,\mu^u_p)|^2 p_{x|z}(x\mid z^u\,;\,p_0,\mu^u)\,dx \\ &= \int_{x\in\mathcal{X}} |x-\hat{x}^u_{\rm scalar}(z^u\,;\,\mu^u_p/u)|^2 p_{x|z}(x\mid z^u\,;\,p_0,\mu^u)\,dx.\end{aligned}$$

Therefore, fixing s (and hence μ^u_p and μ^u), we obtain the conditional expectation

$$\begin{aligned}&\mathbf{E}\left[\mathsf{mse}(p_u,p_0,\mu^u_p,\mu^u,z^u)\mid s\right] \\ &\quad= \mathbf{E}\left[|x-\hat{x}^u_{\rm scalar}(z^u\,;\,\mu^u_p/u)|^2\mid s\right],\end{aligned} \tag{90}$$

where the expectation on the right is over $x \sim p_0(x)$ and z^u given by (89a).

Also, observe that the definitions $\mu^u = \sigma^2_{\rm eff}(u)/s$ and $\mu = \sigma^2_{\rm eff,map}/s$ and along with the limit in Assumption 2 show that

$$\lim_{u\to\infty} \mu^u = \mu. \tag{91}$$

Similarly, since $\mu^u_p = \sigma^2_{\rm p-eff}(u)/s$ and $\lambda = \gamma_p/s$, Assumption 2 shows that

$$\lim_{u\to\infty} \frac{\mu^u_p}{u} = \lambda. \tag{92}$$

Taking the limit as $u \to \infty$,

$$\begin{aligned}\lim_{u\to\infty} &\mathbf{E}\left[s\,\mathsf{mse}(p_u,p_0,\mu^u_p,\mu^u,z^u)\right] \\ &\overset{(a)}{=} \lim_{u\to\infty} \mathbf{E}\left[s|x-\hat{x}^u_{\rm scalar}(z^u\,;\,\mu^u_p/u)|^2\right], \\ &\overset{(b)}{=} \lim_{u\to\infty} \mathbf{E}\left[s|x-\hat{x}^u_{\rm scalar}(z^u\,;\,\lambda)|^2\right], \\ &\overset{(c)}{=} \lim_{u\to\infty} \mathbf{E}\left[s|x-\hat{x}^u_{\rm scalar}(z\,;\,\lambda)|^2\right], \\ &\overset{(d)}{=} \lim_{u\to\infty} \mathbf{E}\left[s|x-\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)|^2\right],\end{aligned}$$

where (a) follows from (90); (b) follows from (92); (c) follows from (91), which implies that $z^u \to z$; and (d) follows from Lemma 6. ∎

The previous lemma enables us to evaluate the limit of the MSE in (30a). To evaluate the limit of the MSE in (30b), we need the following lemma.

Lemma 9: Fix z and λ, and let

$$g(u,x) = u|x-\hat{x}|^2, \qquad \hat{x} = \hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda). \tag{93}$$

Also, let $\varphi(x)$ be given by (84) and $q_u(x)$ be given by (72) with $\mathcal{D} = \mathcal{X}$. Then, for any $\epsilon > 0$, there exists an open neighborhood $U \subseteq \mathcal{X}$ of $\hat{x}$ such that

$$\limsup_{u\to\infty}\left|\int_{x\in U} g(u,x)q_u(x)\,dx - \sigma^2(z,\lambda)\right| < \epsilon,$$

where $\sigma^2(z,\lambda)$ is given in Assumption 6.

Proof: The proof is straightforward but somewhat tedious. We will just outline the main steps. Let $\delta > 0$. Using Assumption 5, one can find an open neighborhood $U \subseteq \mathcal{X}$ of $\hat{x}$ such that for all $x \in U$ and $u > 0$,

$$\phi\left(x,\sigma^2_-(u)\right) \le \exp(-u(\varphi(x)-\varphi(\hat{x}))) \le \phi\left(x,\sigma^2_+(u)\right), \tag{94}$$

where $\phi(x,\sigma^2)$ is the unnormalized Gaussian distribution

$$\phi(x,\sigma^2) = \exp\left(-\frac{1}{2\sigma^2}|x-\hat{x}|^2\right)$$

and

$$\begin{aligned}\sigma^2_+(u) &= (1+\delta)\sigma^2(z,\lambda)/u,\\ \sigma^2_-(u) &= (1-\delta)\sigma^2(z,\lambda)/u.\end{aligned}$$

Combining the bounds in (94) with the definition of $q_u(x)$ in (72) and the fact that $U \subseteq \mathcal{X}$ shows that for all $x \in U$ and $u > 0$,

$$\begin{aligned}q_u(x) &= \left[\int_{x\in\mathcal{X}} e^{-u\varphi(x)}\,dx\right]^{-1} e^{-u\varphi(x)}\\ &\le \left[\int_{x\in U}\phi(x,\sigma^2_-(u))\,dx\right]^{-1}\phi(x,\sigma^2_+(u)).\end{aligned}$$

Therefore,

$$\begin{aligned}\int_{x\in U} g(u,x)q_u(x)\,dx &= \int_{x\in U} u|x-\hat{x}|^2 q_u(x)\,dx\\ &\le \left[\int_{x\in U}\phi(x,\sigma^2_-(u))\,dx\right]^{-1}\\ &\quad \int_{x\in U} u|x-\hat{x}|^2\phi(x,\sigma^2_+(u))\,dx.\end{aligned} \tag{95}$$

Now, it can be verified that

$$\lim_{u\to\infty}\int_{x\in U} u^{1/2}\phi(x,\sigma^2_-(u))\,dx = \sqrt{2\pi(1-\delta)}\sigma(z,\lambda) \tag{96}$$

and

$$\begin{aligned}&\lim_{u\to\infty}\int_{x\in U} u^{3/2}|x-\hat{x}|^2\phi(x,\sigma^2_+(u))\,dx\\ &\quad = \sqrt{2\pi(1+\delta)^3}\sigma(z,\lambda)^3.\end{aligned} \tag{97}$$

Substituting (96) and (97) into (95) shows that

$$\limsup_{u\to\infty}\int_{x\in U} g(u,x)q_u(x)\,dx \le \frac{(1+\delta)^{3/2}}{1-\delta}\sigma^2(z,\lambda).$$

A similar calculation shows that

$$\liminf_{u\to\infty}\int_{x\in U} g(u,x)q_u(x)\,dx \ge \frac{(1-\delta)^{3/2}}{1+\delta}\sigma^2(z,\lambda).$$

Therefore, with appropriate selection of δ, one can find a neighborhood U of $\hat{x}$ such that

$$\limsup_{u\to\infty}\left|\int_{x\in U} g(u,x)q_u(x)\,dx - \sigma^2(z,\lambda)\right| < \epsilon,$$

and this proves the lemma. ∎

Using the above result, we can evaluate the scalar MSE.

Lemma 10: Using the notation of Lemma 8,

$$\lim_{u\to\infty}\mathbf{E}\left[us\,\mathsf{mse}(p_u,p_u,\mu^u_p,\mu^u_p,z)\right] = \mathbf{E}\left[s\sigma^2(z,\gamma_p/s)\right].$$

Proof: This is an application of Lemma 2. Fix z and λ and define $g(u,x)$ as in (93). As in the proof of Lemma 6, the conditional distribution $p_{x|z}(x \mid z\,;\,p_u,\lambda/u)$ is given by (83) with $\varphi(x)$ given by (84). The definition of $\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)$ in (25) shows that $\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)$ minimizes $\varphi(x)$. Similar to the proof of Lemma 6, one can show that items (a) and (b) of Lemma 2 are satisfied. Also, Lemma 9 shows that item (c) of Lemma 2 holds with $g_\infty = \sigma^2(z,\lambda)$. Therefore, all the hypotheses of Lemma 2 are satisfied and

$$\lim_{u\to\infty}\int_{x\in\mathcal{X}} u|x-\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)|^2 q_u(x)\,dx = \sigma^2(z,\lambda), \tag{98}$$

for all λ and almost all z.

Now

$$\begin{aligned}&\mathsf{mse}(p_u,p_u,\lambda/u,\lambda/u,z)\\ &\overset{(a)}{=} \int_{x\in\mathcal{X}}|x-\hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,p_u,\lambda/u)|^2 p_{x|z}(x\mid z\,;\,p_u,\lambda/u)\,dx\\ &\overset{(b)}{=} \int_{x\in\mathcal{X}}|x-\hat{x}^{\rm pmmse}_{\rm scalar}(z\,;\,p_u,\lambda/u)|^2 q_u(x)\,dx\\ &\overset{(c)}{=} \int_{x\in\mathcal{X}}|x-\hat{x}^{u}_{\rm scalar}(z\,;\,\lambda)|^2 q_u(x)\,dx,\end{aligned} \tag{99}$$

where (a) is the definition of mse in (11); (b) follows from (83); and (c) follows from (63). Taking the limit of this expression

$$\begin{aligned}&\lim_{u\to\infty} u\,\mathsf{mse}(p_u,p_u,\lambda/u,\lambda/u,z)\\ &\overset{(a)}{=} \lim_{u\to\infty}\int_{x\in\mathcal{X}} u|x-\hat{x}^{u}_{\rm scalar}(z\,;\,\lambda)|^2 q_u(x)\,dx\\ &\overset{(b)}{=} \lim_{u\to\infty}\int_{x\in\mathcal{X}} u|x-\hat{x}^{\rm pmap}_{\rm scalar}(z\,;\,\lambda)|^2 q_u(x)\,dx\\ &\overset{(c)}{=} \sigma^2(z,\lambda),\end{aligned} \tag{100}$$

where (a) follows from (99); (b) follows from Lemma 6; and (c) follows from (98).

The variables z^u and z in (89a) and (89b) as well as μ^u and μ^u_p are deterministic functions of x, v, s, and u. Fixing x, v, and s and taking the limit with respect to u we obtain

the deterministic limit

$$\begin{aligned}
&\lim_{u\to\infty} u\,\mathsf{mse}(p_u, p_u, \mu_p^u, \mu_p^u, z^u) \\
&\overset{(a)}{=} \lim_{u\to\infty} u\,\mathsf{mse}(p_u, p_u, \sigma^2_{\mathrm{p-eff}}(u)/s, \sigma^2_{\mathrm{p-eff}}(u)/s, z^u) \\
&\overset{(b)}{=} \lim_{u\to\infty} \sigma^2(z^u, u\sigma^2_{\mathrm{p-eff}}(u)/s) \\
&\overset{(c)}{=} \lim_{u\to\infty} \sigma^2(z, u\sigma^2_{\mathrm{p-eff}}(u)/s) \\
&\overset{(d)}{=} \sigma^2(z, \gamma_p/s), \qquad (101)
\end{aligned}$$

where (a) follows from the definitions of μ^u and μ_p^u in Lemma 8; (b) follows from (100); (c) follows from the limit (proved in Lemma 8) that $z^u \to z$ as $u \to \infty$; and (d) follows from the limit in Assumption 2.

Finally, observe that for any prior p and noise level μ,

$$\mathsf{mse}(p, p, \mu, \mu, z) \le \mu,$$

since the MSE error must be smaller than the additive noise level μ. Therefore, for any u and s,

$$us\,\mathsf{mse}(p_u, p_u, \mu_p^u, \mu_p^u, z^u) \le us\mu_p^u = u\sigma^2_{\mathrm{eff}}(u),$$

where we have used the definition $\mu_p^u = \sigma^2_{\mathrm{eff}}(u)/s$. Since $u\sigma^2_{\mathrm{eff}}(u)$ converges, there must exists a constant $M > 0$ such that

$$us\,\mathsf{mse}(p_u, p_u, \mu_p^u, \mu_p^u, z^u) \le us\mu_p^u \le M,$$

for all u, s and z^u. The lemma now follows from applying the Dominated Convergence Theorem and taking expectations of both sides of (101). ∎

We can now show that the limiting noise values satisfy the fixed-point equations.

Lemma 11: The limiting effective noise levels $\sigma^2_{\mathrm{eff,map}}$ and γ_p in Assumption 2 satisfy the fixed-point equations (30a) and (30b).

Proof: The noise levels $\sigma^2_{\mathrm{eff}}(u)$ and $\sigma^2_{\mathrm{p-eff}}(u)$ satisfy the fixed-point equations (13a) and (13b) of the RS PMMSE decoupling property with the postulated prior $p_{\mathrm{post}} = p_u$ and noise level $\sigma^2_{\mathrm{post}} = \gamma/u$. Therefore, using the notation in Lemma 8,

$$\begin{aligned}
\sigma^2_{\mathrm{eff}}(u) &= \sigma_0^2 + \beta\,\mathbf{E}\left[s\,\mathsf{mse}(p_u, p_0, \mu_p^u, \mu^u, z^u)\right], \qquad (102a)\\
u\sigma^2_{\mathrm{p-eff}}(u) &= \gamma + \beta\,\mathbf{E}\left[us\,\mathsf{mse}(p_u, p_u, \mu_p^u, \mu_p^u, z^u)\right], \qquad (102b)
\end{aligned}$$

where (as defined in Lemma 8), $\mu^u = \sigma^2_{\mathrm{eff}}(u)/s$ and $\mu_p^u = \sigma^2_{\mathrm{p-eff}}(u)/s$ and the expectations are taken over $s \sim p_S(s)$, $x \sim p_0(x)$, and z^u in (89a).

Therefore,

$$\begin{aligned}
\sigma^2_{\mathrm{eff,map}} &\overset{(a)}{=} \lim_{u\to\infty} \sigma^2_{\mathrm{eff}}(u) \\
&\overset{(b)}{=} \sigma_0^2 + \beta\,\mathbf{E}\left[s\,\mathsf{mse}(p_u, p_0, \mu_p^u, \mu^u, z^u)\right] \\
&\overset{(c)}{=} \sigma_0^2 + \beta\,\mathbf{E}\left[s|x - \hat{x}^{\mathrm{pmap}}_{\mathrm{scalar}}(z\,;\,\lambda)|^2\right],
\end{aligned}$$

where (a) follows from the limit in Assumption 2; (b) follows from (102a); and (c) follows from Lemma 8. This shows that (30a) is satisfied.

Similarly,

$$\begin{aligned}
\gamma_p &\overset{(a)}{=} \lim_{u\to\infty} u\sigma^2_{\mathrm{p-eff}}(u) \\
&\overset{(b)}{=} \gamma + \beta\,\mathbf{E}\left[s\,\mathsf{mse}(p_u, p_u, \mu_p^u, \mu_p^u, z^u)\right] \\
&\overset{(c)}{=} \gamma + \beta\,\mathbf{E}\left[s\sigma^2(z, \lambda_p)\right],
\end{aligned}$$

where (a) follows from the limit in Assumption 2; (b) follows from (102b); and (c) follows from Lemma 10. This shows that (30b) is satisfied. ∎

ACKNOWLEDGMENT

The authors thank reviewers for helpful comments. They also thank Martin Vetterli and Amin Shokrollahi for discussions and support.

REFERENCES

[1] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *J. Phys. F: Metal Physics*, vol. 5, pp. 965–974, 1975.
[2] M. Mézard and G. Parisi, "A replica analysis of the travelling salesman problem," *J. Phys.*, pp. 1285–1296, 1986.
[3] Y. Fu and P. W. Anderson, "Application of statistical mechanics to NP-complete problems in combinatorial optimisation," *J. Phys. A: Math. Gen.*, vol. 19, pp. 1605–1620, 1986.
[4] R. Monasson and R. Zecchina, "Statistical mechanics of the random K-satisfiability model," *Phys. Rev. E*, vol. 56, no. 2, pp. 1357–1370, 1997.
[5] H. Nishimori, *Statistical physics of spin glasses and information processing: An introduction*, ser. International Series of Monographs on Physics. Oxford, UK: Oxford Univ. Press, 2001.
[6] N. Sourlas, "Spin-glass models as error-correcting codes," *Nature*, vol. 339, pp. 693–695, Jun. 1989.
[7] A. Montanari, "Turbo codes: The phase transition," *Europ. Phys. J. B*, vol. 18, pp. 121–136, 2000.
[8] V. Dotsenko, *An Introduction to the Theory of Spin Glasses and Neural Networks*. World Scientific Press, 1995.
[9] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction*. World Scientific Press, 2001.
[10] M. Mézard and A. Montanari, *Information, Physics and Computation*. Oxford University Press, 2009.
[11] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians*. New York: Springer, 2003.
[12] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
[13] 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.
[14] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.
[15] B. Zaidel, R. Müller, A. Moustakas, and R. de Miguel, "Vector precoding for Gaussian MIMO broadcast channels: Impact of replica symmetry breaking," arXiv:1001.3790 [cs.IT]., Jan. 2010.
[16] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1772–1793, Jul. 2002.
[17] T. Tanaka and M. Okada, "Approximate belief propagation, density evolution, and neurodynamics for CDMA multiuser detection," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 700–706, Feb. 2005.
[18] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," arXiv:cs/0602028v1 [cs.IT]., Feb. 2006.
[19] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *Proc. IEEE Inform. Theory Workshop*, Chengdu, China, Oct. 2006, pp. 194–198.
[20] ——, "Random sparse linear systems observed via arbitrary channels: A decoupling principle," in *Proc. IEEE Int. Symp. Inform. Theory*, Nice, France, Jun. 2007, pp. 946–950.

[21] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," in *Proc. Conf. on Inform. Sci. & Sys.*, Princeton, NJ, Mar. 2010, pp. 1–6.

[22] ——, "Estimation with random linear mixing, belief propagation and compressed sensing," arXiv:1001.2228v1 [cs.IT]., Jan. 2010.

[23] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.

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

[25] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," arXiv:1010.5141 [cs.IT]., Oct. 2010.

[26] ——, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inform. Theory*, Saint Petersburg, Russia, Jul.–Aug. 2011, pp. 2174–2178.

[27] A. Montanari, "Estimating random variables from random sparse observations," arXiv:0709.0145v1 [cs.IT]., Feb. 2008.

[28] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[29] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel properties," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[30] A. Montanari, "Graphical models concepts in compressed sensing," arXiv:1011.4328v3 [cs.IT], Mar. 2011.

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

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

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

[34] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Computing*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[35] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.

[36] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[37] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conf. Rec. 27th Asilomar Conf. Sig., Sys., & Comput.*, vol. 1, Pacific Grove, CA, Nov. 1993, pp. 40–44.

[38] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decomposition," *Optical Eng.*, vol. 33, no. 7, pp. 2183–2191, Jul. 1994.

[39] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1999.

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

[41] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[42] ——, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[43] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[44] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso)," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.

[45] D. L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 1–53, Jan. 2009.

[46] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. 47th Ann. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Sep.–Oct. 2009.

[47] N. Merhav, D. Guo, and S. Shamai, "Statistical physics of signal estimation in Gaussian noise: Theory and examples of phase transitions," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1400–1416, 2010.

[48] Y. Kabashima, T. Wadayama, and T. Tanaka, "Typical reconstruction limit of compressed sensing based on l_p-norm minimization," arXiv:0907.0914 [cs.IT]., Jun. 2009.

[49] D. Voiculescu, "Limit laws for random matrices and free products," *Inventiones Mathematicae*, vol. 104, pp. 201–220, 1991.

[50] R. R. Müller, "Random matrices, free probability and the replica method," in *EUSIPCO*, 2004.

[51] G. Caire, S. Shamai, A. Tulino, and S. Verdú, "Support recovery in compressed sensing: Information-theoretic bounds," in *Proc. UCSD Workshop Inform. Theory & Its Applications*, La Jolla, CA, Jan. 2011.

[52] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 32, no. 1, pp. 85–96, Jan. 1986.

[53] S. Verdu, *Multiuser Detection*. New York: Cambridge University Press, 1998.

[54] N. Merhav, "Physics of the Shannon limits," arXiv:0903.1484 [cs.IT]., Mar. 2009.

[55] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.

[56] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.

[57] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR–Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[58] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 622–640, Mar. 1999.

[59] D. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and capacity," *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 641–675, Mar. 1999.

[60] S. Verdú and S. Shamai, "Multiuser detection with random spreading and error-correction codes: Fundamental limits," in *Proc. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Sep. 1997.

[61] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," arXiv:0906.3234v1 [cs.IT]., Jun. 2009.

[62] D. Wipf and B. Rao, "Comparing the effects of different weight distributions on finding sparse representations," in *Proc. Neural Information Process. Syst.*, Vancouver, Canada, Dec. 2006.

[63] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On–off random access channels: A compressed sensing framework," arXiv:0903.1022v1 [cs.IT]., Mar. 2009.

[64] A. Miller, *Subset Selection in Regression*, 2nd ed., ser. Monographs on Statistics and Applied Probability. New York: Chapman & Hall/CRC, 2002, no. 95.

[65] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.

[66] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.

[67] G. Reeves, "Sparse signal sampling using noisy linear projections," Univ. of California, Berkeley, Dept. of Elec. Eng. and Comp. Sci., Tech. Rep. UCB/EECS-2008-3, Jan. 2008.

[68] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 10, pp. 5111–5130, Oct. 2010.

[69] V. Saligrama and M. Zhao, "Thresholded basis pursuit: LP algorithm for oder-wise optimal support recovery for sparse and approximately sparse signals from noisy measurements," *IEEE Trans. Inform. Theory*, vol. 57, no. 3, pp. 1567–1586, Mar. 2011.

[70] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[71] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Stat.*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.

[72] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *Proc. 44th Ann. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Sep. 2006.

[73] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York: Springer, 1998.

[74] H. Nishimori and D. Sherrington, "Absence of replica symmetry breaking in a region of the phase diagram of the Ising spin glass," in *Amer. Inst. Phys. Conf.: Disordered and Complex Sys.*, 2001.

[75] J. R. L. de Almeida and D. J. Thouless, “Stability of the Sherrington–Kirkpatrick solution of a spin glass model,” *J. Phys. A: Math. Gen.*, vol. 11, no. 5, pp. 983–990, 1978.