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

Sundeep Rangan
Qualcomm Technologies
Bedminster, NJ
srangan@qualcomm.com

Alyson K. Fletcher
University of California, Berkeley
Berkeley, CA
alyson@eecs.berkeley.edu

Vivek K Goyal
Mass. Inst. of Tech.
Cambridge, MA
vgoyal@mit.edu

Abstract

The replica method is a non-rigorous but widely-accepted technique from statistical physics used in the asymptotic analysis of large, random, nonlinear problems. This paper applies the replica method to non-Gaussian maximum a posteriori (MAP) estimation. It is shown that with random linear measurements and Gaussian noise, the asymptotic behavior of the MAP estimate of an n-dimensional vector "decouples" as n scalar MAP estimators. The result is a counterpart to Guo and Verdú's replica analysis of minimum mean-squared error estimation.

The replica 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 exactly computing various performance metrics including mean-squared error and sparsity pattern recovery probability.

1 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. One of the most basic estimators for $\mathbf{x}$ is the maximum a posteriori (MAP) estimate

$$\hat{\mathbf{x}}^{\rm map}(\mathbf{y}) = \mathop{\arg\max}_{\mathbf{x} \in \mathbb{R}^n} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}), \tag{2}$$

which is defined assuming some prior on $\mathbf{x}$. For most priors, the MAP estimate is nonlinear and its behavior is not easily characterizable. Even if the priors for $\mathbf{x}$ and $\mathbf{w}$ are separable, the analysis of the MAP estimate may be difficult since the matrix Φ couples the n unknown components of $\mathbf{x}$ with the m measurements in the vector $\mathbf{y}$.

The primary contribution of this paper—an abridged version of [1]—is to show that with certain large random Φ and Gaussian $\mathbf{w}$, there is an *asymptotic decoupling* of (1) into n scalar MAP estimation problems. Each equivalent scalar problem has an appropriate scalar prior and Gaussian noise with an *effective noise level*. The analysis yields the asymptotic joint distribution of each component x_j of $\mathbf{x}$ and its corresponding estimate $\hat{x}_j$ in the MAP estimate vector $\hat{\mathbf{x}}^{\rm map}(\mathbf{y})$. From the joint distribution, various further computations can be made, such as the mean-squared error (MSE) of the MAP estimate or the error probability of a hypothesis test computed from the MAP estimate.

*This work was supported in part by a University of California President's Postdoctoral Fellowship and National Science Foundation CAREER Award 0643836.

Replica Method. Our analysis is based on a powerful but non-rigorous technique from statistical physics known as the replica method. The replica method was originally developed by Edwards and Anderson [2] 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 [3].

The replica method was first applied to the study of nonlinear MAP estimation problems by Tanaka [4] and Müller [5]. These papers studied 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 $\mathbf{w}$. The results were then generalized in a remarkable paper by Guo and Verdú [6] to vectors $\mathbf{x}$ with arbitrary distributions. Guo and Verdú's result was also able to incorporate a large class of minimum postulated MSE estimators, where the estimator may assume a prior that is different from the actual prior. The main result in this paper is the corresponding MAP statement to Guo and Verdú's result. In fact, our result is derived from Guo and Verdú's by taking appropriate limits with large deviations arguments.

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. Some progress has been made in formally proving these assumptions; a survey of this work can be found in [7]. Also, some of the predictions of the replica method have been validated rigorously by other means [8]. To emphasize our dependence on these unproven assumptions, we will refer to Guo and Verdú's result as the Replica MMSE Claim. Our main result, which depends on Guo and Verdú's analysis, will be called the Replica MAP Claim.

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 [9–11]. In *compressed sensing*, one estimates a sparse vector $\mathbf{x}$ from random linear measurements. Generically, optimal estimation of $\mathbf{x}$ with a sparse prior is NP-hard [12]. Thus, most attention has focused on greedy heuristics such as matching pursuit and convex relaxations such as basis pursuit [13] or lasso [14]. While successful in practice, these algorithms are difficult to analyze precisely.

Recent compressed sensing research has provided scaling laws on numbers of measurements that guarantee good performance of these methods [15–17]. However, these scaling laws are in general conservative. There are, of course, notable exceptions including [18] and [19] 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 MAP-based estimators. For these algorithms, the replica method provides—under the assumption of the replica hypotheses—not just bounds, but the exact asymptotic behavior. This in turns permits exact expressions for various performance metrics such as MSE or fraction of support recovery. The expressions apply for arbitrary ratios k/n, n/m and SNR.

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

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

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.

In (3), we have factored $\Phi = \mathbf{A}\mathbf{S}^{1/2}$ so that even with the i.i.d. assumption on x_js 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. 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 $p_0(x_j)$.

(c) The noise $\mathbf{w}$ is Gaussian with $\mathbf{w} \sim \mathcal{N}(0, \sigma_0^2 I_m)$.

(d) The components of the matrix $\mathbf{A}$ are i.i.d. zero mean with variance $1/m$.

(e) The scale factors s_j are i.i.d. and satisfy $s_j > 0$ almost surely.

(f) The scale factor matrix $\mathbf{S}$, measurement matrix $\mathbf{A}$, vector $\mathbf{x}$ and noise $\mathbf{w}$ are independent.

3 Review of the Replica MMSE Claim

We begin by reviewing the Replica MMSE Claim of Guo and Verdú [6]. 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 *minimum postulated MSE (MPMSE)* estimate of $\mathbf{x}$ as

$$\hat{\mathbf{x}}^{\rm mpmse}(\mathbf{y}) = \mathbf{E}\left(\mathbf{x} \mid \mathbf{y}\,;\, p_{\rm post}, \sigma^2_{\rm post}\right) = \int \mathbf{x} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}\,;\, p_{\rm post}, \sigma^2_{\rm post})\, d\mathbf{x},$$

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 and noise variance specified as parameters after the semicolon:

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}\,;\, q, \sigma^2) = C^{-1} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|^2\right) q(\mathbf{x}), \qquad q(\mathbf{x}) = \prod_{j=1}^{n} q(x_j), \quad (5)$$

where C is a normalization constant.

The Replica MMSE Claim describes the asymptotic behavior of the postulated MMSE estimator via 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

$$p_{x|z}(x \mid z\,;\, q, \mu) = \left[\int_{x\in\mathcal{X}} \phi(z - x\,;\, \mu) q(x)\, dx\right]^{-1} \phi(z - x\,;\, \mu) q(x) \quad (6)$$

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

$$\phi(v\,;\, \mu) = \frac{1}{\sqrt{2\pi\mu}} e^{-|v|^2/(2\mu)}. \quad (7)$$

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

$$z = x + \sqrt{\mu} v, \quad (8)$$

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

$$\hat{x}^{\rm mmse}_{\rm scalar}(z\,;\, q, \mu) = \int_{x\in\mathcal{X}} x p_{x|z}(x|z\,;\, \mu)\, dx. \quad (9)$$

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

$$\mathsf{mse}(p_1, p_0, \mu_1, \mu_0, z) = \int_{x \in \mathcal{X}} |x - \hat{x}^{\rm mmse}_{\rm scalar}(z\,;\,p_1, \mu_1)|^2 p_{x|z}(x \mid z\,;\,p_0, \mu_0)\,dx, \qquad (10)$$

which is the mean-squared error in estimating the scalar x from the variable z in (8) 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 MMSE Claim [6]. *Consider the estimation problem in Section 2. Let $\hat{\mathbf{x}}^{\rm mpmse}(\mathbf{y})$ be the MPMSE 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 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 $(x_j, s_j, \hat{x}^{\rm mpmse}_j)$ converge in distribution to the random vector $(x, s, \hat{x})$ where x, s, and v are independent with $x \sim p_0(x)$, $s \sim p_S(s)$, $v \sim \mathcal{N}(0,1)$, and

$$\hat{x} = \hat{x}^{\rm mmse}_{\rm scalar}(z\,;\,p_{\rm post}, \mu_p), \quad z = x + \sqrt{\mu}v. \qquad (11)$$

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

(b) The effective noise levels satisfy the equations

$$\sigma^2_{\rm eff} = \sigma^2_0 + \beta\mathbf{E}\left[s\,\mathsf{mse}(p_{\rm post}, p_0, \mu_p, \mu, z)\right] \qquad (12a)$$

$$\sigma^2_{\rm p-eff} = \sigma^2_{\rm post} + \beta\mathbf{E}\left[s\,\mathsf{mse}(p_{\rm post}, p_{\rm post}, \mu_p, \mu_p, z)\right], \qquad (12b)$$

where the expectations are taken over $s \sim p_S(s)$ and z generated by (11).

The Replica MMSE Claim 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 $\hat{x}(z\,;\,p_{\rm post}, \mu_p)$.

The effective noise levels $\sigma^2_{\rm eff}$ and $\sigma^2_{\rm p-eff}$ are described by the solutions to fixed-point equations (12). 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. When there are multiple solutions to these equations, the true solution is the minimizer of a certain Gibbs' function [6].

4 Replica MAP Claim

We now turn to MAP estimation. Let $\mathcal{X} \subseteq \mathbb{R}$ be some (measurable) set and consider an estimator of the form

$$\hat{\mathbf{x}}^{\rm map}(\mathbf{y}) = \mathop{\arg\min}_{\mathbf{x} \in \mathcal{X}^n} \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|^2_2 + \sum_{j=1}^{n} f(x_j), \qquad (13)$$

where $\gamma > 0$ is an algorithm parameter and $f : \mathcal{X} \to \mathbb{R}$ is some scalar-valued, non-negative cost function. We will assume that the objective function in (13) has a unique essential minimizer for almost all $\mathbf{y}$.

The estimator (13) can be interpreted as a MAP estimator. Specifically, for any $u > 0$, it can be verified that $\hat{\mathbf{x}}^{\rm map}(\mathbf{y})$ is the MAP estimate

$$\hat{\mathbf{x}}^{\rm map}(\mathbf{y}) = \mathop{\arg\max}_{\mathbf{x} \in \mathcal{X}^n} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}\,;\,p_u, \sigma^2_u),$$

where $p_u(\mathbf{x})$ and σ^2_u are the prior and noise level

$$p_u(\mathbf{x}) = \left[\int_{\mathbf{x} \in \mathcal{X}^n} \exp(-uf(\mathbf{x}))d\mathbf{x}\right]^{-1} \exp(-uf(\mathbf{x})), \quad \sigma^2_u = \gamma/u, \qquad (14)$$

where $f(\mathbf{x}) = \sum_j f(x_j)$. To analyze this MAP estimator, we consider a sequence of MMSE estimators

$$\widehat{\mathbf{x}}^u(\mathbf{y}) = \mathbf{E}\left(\mathbf{x} \mid \mathbf{y}\,;\, p_u, \sigma_u^2\right). \tag{15}$$

The proof of the Replica MAP Claim below (see [1]) uses a standard large deviations argument to show that

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

for all $\mathbf{y}$. Under the assumption that the behaviors of the MMSE estimators are described by the Replica MMSE Claim, we can then extrapolate the behavior of the MAP estimator.

To state the claim, define the scalar MAP estimator

$$\hat{x}^{\rm map}_{\rm scalar}(z\,;\,\lambda) = \mathop{\arg\min}_{x\in\mathcal{X}} F(x,z,\lambda), \quad F(x,z,\lambda) = \frac{1}{2\lambda}|z-x|^2 + f(x). \tag{16}$$

where, again, we assume that (16) has a unique essential minimizer for almost all λ and almost all z. We also assume that the limit

$$\sigma^2(z,\lambda) = \lim_{x\to\hat{x}} \frac{|x-\hat{x}|^2}{2(F(x,z,\lambda)-F(\hat{x},z,\lambda))}, \tag{17}$$

exists where $\hat{x} = \hat{x}^{\rm map}_{\rm scalar}(z;\lambda)$. We make the following additional assumptions:

Assumption 1 *Consider the MAP estimator (13) applied to the estimation problem in Section 2. Assume:*

(a) *For all $u>0$ sufficiently large, assume the postulated prior p_u and noise level σ_u^2 satisfy the Replica MMSE Claim. Also, assume that for the corresponding effective noise levels, $\sigma^2_{\rm eff}(u)$ and $\sigma^2_{\rm p-eff}(u)$, the following limits exists:*

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

(b) *Suppose for each n, $\hat{x}^u_j(n)$ 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 noise level σ_u^2. Then, assume that the following limits can be interchanged:*

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

where the limits are in distribution.

(c) *Assume that $f(x)$ is non-negative and satisfies $f(x)/\log|x| \to \infty$ as $|x|\to\infty$.*

Item (a) is stated to reiterate that we are assuming the Replica MMSE Claim is valid. See [1, Sect. IV] for additional discussion of technical assumptions.

Replica MAP Claim [1]. *Consider the estimation problem in Section 2. Let $\hat{\mathbf{x}}^{\rm map}(\mathbf{y})$ be the MAP estimator (13) defined for some $f(x)$ and $\gamma>0$ satisfying Assumption 1. For each n, let $j=j(n)$ be some deterministic component index with $j(n)\in\{1,\ldots,n\}$. Then:*

(a) *As $n\to\infty$, the random vectors $(x_j, s_j, \hat{x}^{\rm map}_j)$ converge in distribution to the random vector $(x,s,\hat{x})$ where x, s, and v are independent with $x\sim p_0(x)$, $s\sim p_S(s)$, $v\sim\mathcal{N}(0,1)$, and*

$$\hat{x} = \hat{x}^{\rm map}_{\rm scalar}(z,\lambda_p), \quad z = x+\sqrt{\mu}v, \tag{18}$$

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

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

$$\sigma^2_{\rm eff,map} = \sigma_0^2 + \beta\mathbf{E}\left[s|x-\hat{x}|^2\right] \tag{19a}$$

$$\gamma_p = \gamma + \beta\mathbf{E}\left[s\sigma^2(z,\lambda_p)\right], \tag{19b}$$

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 (18).

Analogously to the Replica MMSE Claim, the Replica MAP Claim asserts that asymptotic behavior of the MAP 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 MAP estimate of the component from the noise-corrupted version.

5 Analysis of Compressed Sensing

Our results thus far hold for any separable distribution for $\mathbf{x}$ and under mild conditions on the cost function f. The role of f is to determine the estimator. In this section, we first consider choices of f that yield MAP estimators relevant to compressed sensing. We then additionally impose a sparse prior for $\mathbf{x}$ for numerical evaluations of asymptotic performance.

Lasso Estimation. We first consider the lasso or basis pursuit estimate [13, 14] given by

$$\hat{\mathbf{x}}^{\rm lasso}(\mathbf{y}) = \mathop{\arg\min}_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_1, \tag{20}$$

where $\gamma > 0$ is an algorithm parameter. This estimator is identical to the MAP estimator (13) with the cost function

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

With this cost function, the scalar MAP estimator in (16) is given by

$$\hat{x}^{\rm map}_{\rm scalar}(z\,;\,\lambda) = T^{\rm soft}_\lambda(z), \tag{21}$$

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

$$T^{\rm soft}_\lambda(z) = \left\{ \begin{array}{rl} z-\lambda, & \mbox{if } z > \lambda; \\ 0, & \mbox{if } |z| \leq \lambda; \\ z+\lambda, & \mbox{if } z < -\lambda. \end{array} \right. \tag{22}$$

The Replica MAP Claim now 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

$$\hat{x} = T^{\rm soft}_{\lambda_p}(z), \qquad z = x + \sqrt{\mu}v, \tag{23}$$

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 $\hat{x}_j$.

To calculate the effective noise levels, one can perform a simple calculation to show that $\sigma^2(z,\lambda)$ in (17) is given by

$$\sigma^2(z,\lambda) = \left\{ \begin{array}{ll} \lambda, & \mbox{if } |z| > \lambda; \\ 0, & \mbox{if } |z| \leq \lambda. \end{array} \right. \tag{24}$$

Hence,

$$\mathbf{E}\left[s\sigma^2(z,\lambda_p)\right] = \mathbf{E}\left[s\lambda_p \Pr(|z| > \lambda_p)\right] = \gamma_p \Pr(|z| > \gamma_p/s) \tag{25}$$

where we have use the fact that $\lambda_p = \gamma_p/s$. Substituting (21) and (25) into (19), we obtain the fixed-point equations

$$\sigma^2_{\rm eff,map} = \sigma_0^2 + \beta\mathbf{E}\left[s|x - T^{\rm soft}_{\lambda_p}(z)|^2\right] \tag{26a}$$

$$\gamma_p = \gamma + \beta\gamma_p \Pr(|z| > \gamma_p/s), \tag{26b}$$

where the expectations are taken with respect to $x \sim p_0(x)$, $s \sim p_S(s)$, and z in (23). 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.

Zero Norm-Regularized Estimation. Lasso can be regarded as a convex relaxation of zero norm-regularized estimation

$$\hat{\mathbf{x}}^{\text{zero}}(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_0, \tag{27}$$

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 characterization of its *performance*. This analysis can provide a bound on the performance achievable by practical algorithms.

The zero norm-regularized estimator is identical to the MAP estimator (13) with the cost function

$$f(x) = \begin{cases} 0, & \text{if } x = 0; \\ 1, & \text{if } x \neq 0. \end{cases} \tag{28}$$

Technically, this cost function does not satisfy the conditions of the Replica MAP Claim. To avoid this problem, we can consider an approximation of (28),

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

which is defined on the set $\mathcal{X} = \{x : |x| \leq M\}$. We can then take the limits $\delta \to 0$ and $M \to \infty$. To simplify the presentation, we will just apply the Replica MAP Claim with $f(x)$ in (28) and omit the details in taking the appropriate limits.

With $f(x)$ given by (28), the scalar MAP estimator in (16) is given by

$$\hat{x}^{\text{map}}_{\text{scalar}}(z\,;\,\lambda) = T^{\text{hard}}_t(z), \qquad t = \sqrt{2\lambda}, \tag{29}$$

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

$$T^{\text{hard}}_t(z) = \begin{cases} z, & \text{if } |z| > t; \\ 0, & \text{if } |z| \leq t. \end{cases} \tag{30}$$

Now, similar to the case of lasso estimation, the Replica MAP Claim states 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

$$\hat{x} = T^{\text{hard}}_t(z), \qquad z = x + \sqrt{\mu}v, \tag{31}$$

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

$$t = \sqrt{2\lambda_p} = \sqrt{2\gamma_p/s}. \tag{32}$$

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

The fixed-point equations for the effective noise levels $\sigma^2_{\text{eff,map}}$ and γ_p can be computed similarly to the case of lasso. Specifically, one can verify that (24) and (25) are both satisfied for the hard thresholding operator as well. Substituting (25) and (29) into (19), we obtain the fixed-point equations

$$\sigma^2_{\text{eff,map}} = \sigma_0^2 + \beta\mathbf{E}\left[s|x - T^{\text{hard}}_t(z)|^2\right], \tag{33a}$$

$$\gamma_p = \gamma + \beta\gamma_p \Pr(|z| > t), \tag{33b}$$

where the expectations are taken with respect to $x \sim p_0(x)$, $s \sim p_S(s)$, z in (31), and t given by (32). These fixed-point equations can be solved numerically.

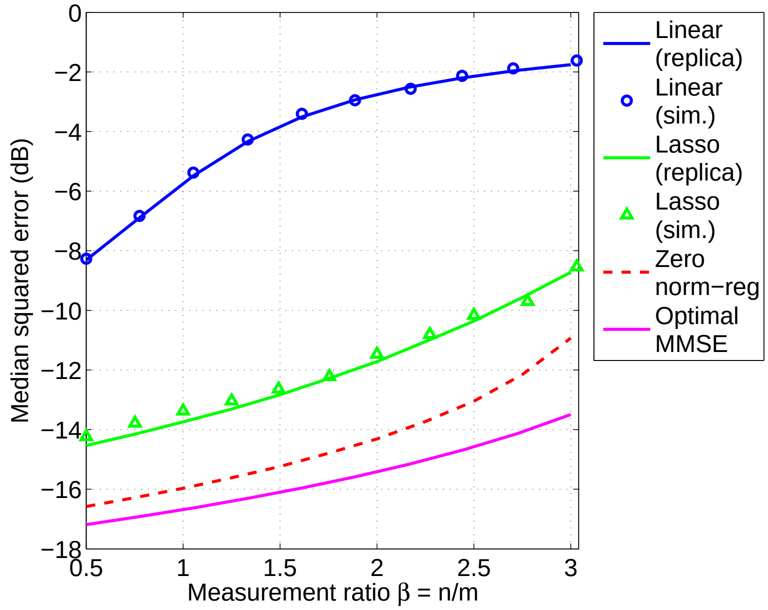

Figure 1: MSE performance prediction with the Replica MAP Claim. Plotted is the median normalized SE for various sparse recovery algorithms: linear MMSE estimation, lasso, zero norm-regularized estimation, and optimal MMSE estimation. Solid lines show the asymptotic predicted MSE from the Replica MAP Claim. For the linear and lasso estimators, the circles and triangles show the actual median SE over 1000 Monte Carlo simulations.

Numerical Simulation. To validate the predictive power of the Replica MAP Claim for finite dimensions, we performed numerical simulations where the components of $\mathbf{x}$ are a zero-mean Bernoulli–Gaussian process. Specifically,

$$x_j \sim \begin{cases} \mathcal{N}(0,1), & \text{with prob. } 0.1; \\ 0, & \text{with prob. } 0.9. \end{cases}$$

We took the vector $\mathbf{x}$ to have $n = 100$ i.i.d. components, and we used ten values of m to vary $\beta = n/m$ from 0.5 to 3. For each problem size, we simulated the lasso and linear MMSE estimators over 1000 independent instances with noise levels chosen such that the SNR with perfect side information is 10 dB. Each set of trials is represented by its median squared error in Fig. 1.

The simulated performance is matched very closely by the asymptotic values predicted by the replica analysis. (Analysis of the linear MMSE estimator using the Replica MAP Claim is detailed in [1]; the Replica MMSE Claim is also applicable to this estimator.) In addition, the replica analysis can be applied to zero norm-regularized and optimal MMSE estimators that are computationally infeasible for large problems. These results are also shown in Fig. 1, illustrating the potential of the replica method to quantify the precise performance losses of practical algorithms.

Additional numerical simulations in [1] illustrate convergence to the replica MAP limit, applicability to discrete distributions for $\mathbf{x}$, effects of power variations in the components, and accurate prediction of the probability of sparsity pattern recovery.

6 Conclusions

We have shown that the behavior of vector MAP estimators with large random measurement matrices and Gaussian noise asymptotically matches that of a set of decoupled scalar estimation problems. We believe that this equivalence to a simple scalar model will open up numerous doors for analysis, particularly in problems of interest in compressed sensing. One can use the model to dramatically improve upon existing performance analyses for sparsity pattern recovery and MSE. Also, the technique is sufficiently general to study effects of dynamic range.

References

[1] 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]., June 2009.

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

[3] H. Nishimori. *Statistical physics of spin glasses and information processing: An introduction*. International Series of Monographs on Physics. Oxford Univ. Press, Oxford, UK, 2001.

[4] T. Tanaka. A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors. *IEEE Trans. Inform. Theory*, 48(11):2888–2910, November 2002.

[5] R. R. Müller. Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation. *IEEE Trans. Signal Process.*, 51(11):2821–2828, November 2003.

[6] D. Guo and S. Verdú. Randomly spread CDMA: Asymptotics via statistical physics. *IEEE Trans. Inform. Theory*, 51(6):1983–2010, June 2005.

[7] M. Talagrand. *Spin Glasses: A Challenge for Mathematicians*. Springer, New York, 2003.

[8] 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]., February 2006.

[9] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, February 2006.

[10] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, April 2006.

[11] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, December 2006.

[12] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Computing*, 24(2):227–234, April 1995.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comp.*, 20(1):33–61, 1999.

[14] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc., Ser. B*, 58(1):267–288, 1996.

[15] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory*, 52(1):6–18, January 2006.

[16] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, October 2004.

[17] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inform. Theory*, 52(3):1030–1051, March 2006.

[18] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso). *IEEE Trans. Inform. Theory*, 55(5):2183–2202, May 2009.

[19] D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.*, 22(1):1–53, January 2009.