The Noise-Sensitivity Phase Transition in Compressed Sensing

David L. Donoho[*] Arian Maleki[†] and Andrea Montanari[*,†]

April 7, 2010

Abstract

Consider the noisy underdetermined system of linear equations: $y = Ax^0 + z^0$, with $n \times N$ measurement matrix A, $n < N$, and Gaussian white noise $z^0 \sim \mathsf{N}(0, \sigma^2 I)$. Both y and A are known, both x^0 and z^0 are unknown, and we seek an approximation to x^0.

When x^0 has few nonzeros, useful approximations are often obtained by ℓ_1-penalized ℓ_2 minimization, in which the reconstruction $\hat{x}^{1,\lambda}$ solves $\min \|y - Ax\|_2^2/2 + \lambda\|x\|_1$.

Evaluate performance by mean-squared error ($\mathrm{MSE} = \mathbb{E}||\hat{x}^{1,\lambda} - x^0||_2^2/N$). Consider matrices A with iid Gaussian entries and a large-system limit in which $n, N \to \infty$ with $n/N \to \delta$ and $k/n \to \rho$. Call the ratio MSE/σ^2 the *noise sensitivity*. We develop formal expressions for the MSE of $\hat{x}^{1,\lambda}$, and evaluate its worst-case formal noise sensitivity over all types of k-sparse signals. The phase space $0 \le \delta, \rho \le 1$ is partitioned by curve $\rho = \rho_{\mathrm{MSE}}(\delta)$ into two regions. Formal noise sensitivity is bounded throughout the region $\rho < \rho_{\mathrm{MSE}}(\delta)$ and is unbounded throughout the region $\rho > \rho_{\mathrm{MSE}}(\delta)$.

The phase boundary $\rho = \rho_{\mathrm{MSE}}(\delta)$ is *identical to* the previously-known phase transition curve for equivalence of $\ell_1 - \ell_0$ minimization in the k-sparse noiseless case. Hence a single phase boundary describes the fundamental phase transitions both for the noiseless and noisy cases.

Extensive computational experiments validate the predictions of this formalism, including the existence of game theoretical structures underlying it (saddlepoints in the payoff, least-favorable signals and maximin penalization).

Underlying our formalism is an approximate message passing soft thresholding algorithm (AMP) introduced earlier by the authors. Other papers by the authors detail expressions for the formal MSE of AMP and its close connection to ℓ_1-penalized reconstruction. Here we derive the minimax formal MSE of AMP and then read out results for ℓ_1-penalized reconstruction.

Key Words. Approximate Message Passing. Lasso. Basis Pursuit. Minimax Risk over Nearly-Black Objects. Minimax Risk of Soft Thresholding.

Acknowledgements. Work partially supported by NSF DMS-0505303, NSF DMS-0806211, NSF CAREER CCF-0743978. Thanks to Iain Johnstone and Jared Tanner for helpful discussions.

[*]Department of Statistics, Stanford University

[†]Department of Electrical Engineering, Stanford University

1 Introduction

Consider the noisy underdetermined system of linear equations:

$$y = Ax^0 + z^0\,, \tag{1.1}$$

where the matrix A is $n \times N$, $n < N$, the N-vector x^0 is k-sparse (i.e. it has at most k non-zero entries), and $z^0 \in \mathbb{R}^n$ is a Gaussian white noise $z^0 \sim \mathsf{N}(0, \sigma^2 I)$. Both y and A are known, both x^0 and z^0 are unknown, and we seek an approximation to x^0.

A very popular approach estimates x^0 via the solution $x^{1,\lambda}$ of the following convex optimization problem

$$(P_{2,\lambda,1}) \qquad \text{minimize} \quad \frac{1}{2}\,\|y - Ax\|_2^2 + \lambda \|x\|_1. \tag{1.2}$$

Thousands of articles use or study this approach, which has variously been called LASSO, Basis Pursuit, or more prosaically, ℓ_1-penalized least-squares [Tib96, CD95, CDS98]. There is a clear need to understand the extent to which $(P_{2,\lambda,1})$ accurately recovers x^0. Dozens of papers present partial results, setting forth often loose bounds on the behavior of $\hat{x}^{1,\lambda}$ (more below).

Even in the noiseless case $z^0 = 0$, understanding the reconstruction problem (1.1) poses a challenge, as the underlying system of equations $y = Ax^0$ is underdetermined. In this case it is informative to consider ℓ_1 minimization,

$$(P_1) \qquad \text{minimize } \|x\|_1\,, \tag{1.3}$$

$$\text{subject to } y = Ax. \tag{1.4}$$

This is the $\lambda = 0$ limit of (1.2): its solution obeys $\hat{x}^{1,0} = \lim_{\lambda \to 0} x^{1,\lambda}$.

The most precise information about behavior of $\hat{x}^{1,0}$ is obtained by large-system analysis; let n, N tend to infinity so that[1] $n \sim \delta N$ and correspondingly let the number of nonzeros $k \sim \rho n$; thus we have a phase space $0 \le \delta, \rho \le 1$, expressing different combinations of undersampling δ and sparsity ρ. When the matrix A has iid Gaussian elements, phase space $0 \le \delta, \rho \le 1$ can be divided into two components, or *phases*, separated by a curve $\rho = \rho_{\ell_1}(\delta)$, which can be explicitly computed. Below this curve, x^0 is sufficiently sparse that $\hat{x}^{1,0} = x^0$ with high probability and therefore ℓ_1 minimization perfectly recovers the sparse vector $\hat{x}^{1,0}$. Above this curve, sparsity is not sufficient: we have $\hat{x}^{1,0} \neq x^0$ with high probability. Hence the curve $\rho = \rho_{\ell_1}(\delta)$, $0 < \delta < 1$, indicates the precise tradeoff between undersampling and sparsity.

Many authors have considered the behavior of $\hat{x}^{1,\lambda}$ in the noisy case but results are somewhat less conclusive. The most well-known analytic approach is the Restricted Isometry Principle (RIP), developed by Candès and Tao [CT05, CT07]. Again in the case where A has iid Gaussian entries, and in the same large-system limit, the RIP implies that, under sufficient sparsity of x^0, with high probability one has stability bounds of the form $\|\hat{x}^{1,\lambda} - x^0\|_2 \le C(\delta, \rho)\|z^0\|_2 \log N$. The region where $C(\delta, \rho) < \infty$ was orginally an implicitly known, but clearly nonempty region of the (δ, ρ) phase space. Blanchard, Cartis and Tanner [BCT09] recently improved the estimates of C in the case of Gaussian matrices A, by careful large deviations analysis, and by developing an asymmetric RIP, obtaining the largest region where $\hat{x}^{1,\lambda}$ is currently known to be stable. Unfortunately as they show, this region is still relatively small compared to the region $\rho < \rho_{\ell_1}(\delta)$, $0 < \delta < 1$.

It may seem that, in the presence of noise, the precise tradeoff between undersampling and sparsity worsens dramatically, compared to the noiseless case. In fact, the opposite is true. In this

[1] Here and below we write $a \sim b$ if $a/b \to 1$ as both quantities tend to infinity.

paper, we show that in the presence of Gaussian white noise, the mean-squared error of the optimally tuned ℓ_1 penalized least squares estimator behaves well over quite a large region of the phase plane, in fact, it is finite over the exact same region of the phase plane as the region of $\ell_1 - \ell_0$ equivalence derived in the noiseless case.

Our main results, stated in Section 3, give explicit evaluations for the the worst-case formal mean square error of $\hat{x}^{1,\lambda}$ under given conditions of noise, sparsity and undersampling. Our results indicate the noise sensitivity of solutions to (1.2), the optimal penalization parameter λ, and the hardest-to-recover sparse vector. As we show, the noise sensitivity exhibits a phase transition in the undersampling-sparsity (δ, ρ) domain along a curve $\rho = \rho_{\mathrm{MSE}}(\delta)$, and this curve is precisely the same as the ℓ_1-ℓ_0 equivalence curve ρ_{ℓ_1}.

Our results might be compared to work of Xu and Hassibi [XH09], who considered a different departure from the noiseless case. In their work, the noise z^0 was still vanishing, but the vector x_0 was allowed to be an ℓ_1-norm bounded perturbation to a k-sparse vector. They considered stable recovery with respect to such small perturbations and showed that the natural boundary for such stable recovery is again the curve $\rho = \rho_{\mathrm{MSE}}(\delta)$.

1.1 Results of our Formalism

We define below a so-called formal MSE (fMSE), and evaluate the (minimax, formal) *noise sensitivity*:

$$M^*(\delta, \rho) = \sup_{\sigma > 0} \max_{\nu} \min_{\lambda} \mathrm{fMSE}(\hat{x}^{1,\lambda}, \nu, \sigma^2)/\sigma^2; \tag{1.5}$$

here ν denotes the marginal distribution of x^0 (which has fraction of nonzeros not larger than $\rho\delta$), and λ denotes the tuning parameter of the ℓ_1-penalized ℓ_2 minimization. Let $M^{\pm}(\varepsilon)$ denote the minimax MSE of scalar thresholding, defined in Section 2 below. Let $\rho_{\mathrm{MSE}}(\delta)$ denote the solution of

$$M^{\pm}(\rho\delta) = \delta\,. \tag{1.6}$$

Our main *theoretical* result is the formula

$$M^*(\delta, \rho) = \begin{cases} \frac{M^{\pm}(\delta\rho)}{1 - M^{\pm}(\delta\rho)/\delta}, & \rho < \rho_{\mathrm{MSE}}(\delta), \\ \infty, & \rho \geq \rho_{\mathrm{MSE}}(\delta). \end{cases} \tag{1.7}$$

Quantity (1.5) is the payoff of a traditional two-person zero sum game, in which the undersampling and sparsity are fixed in advance, the researcher plays against Nature, Nature picks both a noise level and a signal distribution, and the researcher picks a penalization level, in knowledge of Nature's choices. It is traditional in analyzing such games to identify the least-favorable strategy of Nature (who maximizes payout from the researcher), and the optimal strategy for the researcher (who wants to minimize payout). We are able to identify both and give explicit formulas for the so-called saddlepoint strategy, where Nature plays the least-favorable strategy against the researcher and the researcher minimizes the consequent damage. In Proposition 3.1 below we give formulas for this pair of strategies. The phase-transition structure evident in (1.7) is saying that above the curve ρ_{MSE}, Nature has available unboundedly good strategies, to which the researcher has no effective response.

1.2 Structure of the Formalism

Our approach is presented in Section 4, and uses a combination of ideas from decision theory in mathematical statistics, and message passing algorithms in information theory. On the one hand,

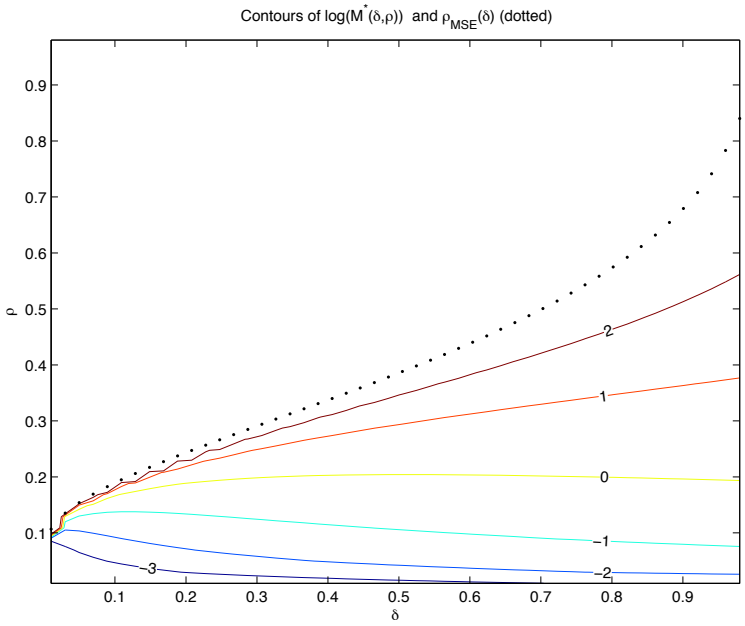

Figure 1: Contour lines of the minimax noise sensitivity $M^*(\delta, \rho)$ in the (ρ, δ) plane. The dotted black curve graphs the phase boundary $(\delta, \rho_{\mathrm{MSE}}(\delta))$. Above this curve, $M^*(\delta, \rho) = \infty$. The colored lines present level sets of $M^*(\delta, \rho) = 1/8,\ 1/4,\ 1/2,\ 1,\ 2,\ 4$ (from bottom to top).

as already evident from formula (1.7), quantities from mathematical statistics play a key role in our formulas. But since these quantities concern a completely different estimator in a completely different problem – the behavior of soft thresholding in estimating a single normal mean, likely to be zero – the superficial appearance of the formulas conceals the type of analysis we are doing. That analysis concerns the properties of an iterative soft thresholding scheme introduced by the authors in [DMM09a], and further developed here. Our formalism neatly describes properties of the formal MSE of AMP as expectations taken in the equilibrium states of a state evolution. As described in [DMM10b], we can calibrate AMP to have the same operating characteristics as ℓ_1-penalized least squares, and by recalibration of the minimax formal MSE for AMP, we get the above results.

1.3 Empirical Validation

We use the word *formalism* for the machinery underlying our derivations because it is not (yet) a rigorously-proven method which is known to give correct results under established regularity conditions. In this sense our method has similarities to the replica and cavity methods of statistical physics, famously useful tools without rigorous general justification.

Our theoretical results are validated here by computational experiments which show that the predictions of our formulas are accurate, and, even more importantly, that the underlying formal structure leading to our predictions – least-favorable objects, game-theoretic saddlepoints of the MSE payoff function, maximin tuning of λ, unboundedness of the noise sensitivity above phase transition– can all be observed experimentally. Because our formalism makes so many different kinds of predictions about quantities with clear operational significance and about their dynamical

evolution in the AMP algorithm, it is quite different than some other formalisms, such as the replica method, in which many fewer checkable predictions are made. In particular, as demonstrated in [DMM09a], the present formalism describes precisely the evolution of an actual low complexity algorithm.

Admittedly, by computational means we can only check individual predictions in specific cases, whereas a full proof could cover all such cases. However, we make available software which checks these features so that interested researchers can check the same phenomena at parameter values that we did not investigate here. *The evidence of our simulations is strong; it is not a realistic possibility that ℓ^1-penalized least squares fails to have the limit behavior discovered here.*

We focused in this paper on measurement matrices A with Gaussian iid entries. It was recently proved that the state evolution formalism at the core of our analysis is indeed asymptotically correct for Gaussian matrices A [BM10]. We believe that similar results hold for matrices A with uniformly bounded iid entries with zero mean and variance $1/n$. However our results should extend to a broader universality class including matrices with iid entries with same mean and variance, under an appropriate light tail condition. It is an outstanding mathematical challenge to prove that such predictions are indeed correct for a broader universality class of estimation problems.

As discussed in Section 7, an alternative route also from statistical physics, using the replica method has been recently used to investigate similar questions. We will argue that the present framework which makes predictions about actual dynamical behavior of algorithms, is computationally verifiable in great detail, whereas the replica method itself applies to no constructive algorithm and makes comparatively many fewer predictions.

2 Minimax MSE of Soft Thresholding

We briefly recall notions from, e.g., [DJHS92, DJ94] and then generalize them. We wish to recover an N vector $x^0 = (x^0(i) : 1 \le i \le N)$ which is observed in Gaussian white noise

$$y(i) = x^0(i) + z^0(i), \qquad 1 \le i \le N,$$

with $z^0(i) \sim \mathsf{N}(0, \sigma^2)$ independent and identically distributed. This can be regarded as special case of the compressed sensing model (1.1), whereby $n = N$ and $A = I$ is the identity matrix – i.e. there is no underdetermined system of equations. We assume that x^0 is sparse. It makes sense to consider soft thresholding

$$\hat{x}^\tau(i) = \eta(y(i); \tau\sigma), \qquad 1 \le i \le N,$$

where the soft threshold function (with threshold level θ) is defined by

$$\eta(x;\theta) = \begin{cases} x - \theta & \text{if } \theta < x, \\ 0 & \text{if } -\theta \le x \le \theta, \\ x + \theta & \text{if } x \le -\theta. \end{cases} \tag{2.1}$$

In words, the estimator (2) 'shrinks' the observations y towards the origin by a multiple τ of the noise level σ.

In place of studying x^0 which are k-sparse, [DJHS92, DJ94] consider random variables X which obey $\mathbb{P}\{X \neq 0\} \le \varepsilon$, where $\varepsilon = k/n$. So let $\mathcal{F}_\varepsilon$ denote the set of probability measures placing all but ε of their mass at the origin:

$$\mathcal{F}_\varepsilon = \{\nu : \nu \text{ is probability measure with } \nu(\{0\}) \ge 1 - \varepsilon\}.$$

We define the soft thresholding mean square error by

$$\mathsf{mse}(\sigma^2;\nu,\tau) \;\equiv\; \mathbb{E}\Big\{\big[\eta\big(X+\sigma\cdot Z;\tau\sigma\big)-X\big]^2\Big\}. \tag{2.2}$$

Here expectation is with respect to independent random variables $Z\sim\mathsf{N}(0,1)$ and $X\sim\nu$.

It is important to allow general σ in calculations below. However, note to the scale invariance

$$\mathsf{mse}(\sigma^2;\nu,\tau)=\sigma^2\mathsf{mse}(1;\nu^{1/\sigma},\tau)\,, \tag{2.3}$$

where ν^a is the probability distribution obtained by rescaling ν: $\nu^a(S)=\nu(\{x:\,a\,x\,\in S\})$. It follows that all calculations can be made in the $\sigma=1$ setting and results rescaled to obtain final answers. Below, when we deal with $\sigma=1$, we will suppress the σ argument, and simply write $\mathsf{mse}(\nu,\tau)\equiv\mathsf{mse}(1;\nu,\tau)$

The *minimax threshold MSE* was defined in [DJHS92, DJ94] by

$$M^{\pm}(\varepsilon)=\inf_{\tau>0}\sup_{\nu\in\mathcal{F}_\varepsilon}\mathsf{mse}(\nu,\tau)\,. \tag{2.4}$$

(The superscript $\pm$ reminds us that, when the estimand X is nonzero, it may take either sign. In Section 6.1, the superscript $+$ will be used to cover the case where $X\ge 0$). We will denote by $\tau^{\pm}(\varepsilon)$ the threshold level achieving the infimum. Figure 2 depicts the behavior of $M^{\pm}$ and $\tau^{\pm}$ as a function of ε. $M^{\pm}(\varepsilon)$ was studied in [DJ94] where one can find a considerable amount of information about the behavior of the optimal threshold $\tau^{\pm}$ and the least favorable distribution $\nu_\varepsilon^{\pm}$. In particular, the optimal threshold behaves as

$$\tau^{\pm}(\varepsilon)\sim\sqrt{2\log(\varepsilon^{-1})}\,,\qquad\text{as}\quad\varepsilon\to 0,$$

and is explicitly computable at finite ε.

A peculiar aspect of the results in [DJ94] requires us to generalize their results somewhat. For a given, fixed $\tau>0$, the worst case MSE obeys

$$\sup_{\nu\in\mathcal{F}_\varepsilon}\mathsf{mse}(\nu,\tau)=\varepsilon\,(1+\tau^2)+(1-\varepsilon)[2(1+\tau^2)\,\Phi(-\tau)-2\tau\,\phi(\tau)]\,, \tag{2.5}$$

with $\phi(z)=\exp(-z^2/2)/\sqrt{2\pi}$ the standard normal density and $\Phi(z)=\int_{-\infty}^{z}\phi(x)\,\mathrm{d}x$ the Gaussian distribution. This supremum is "achieved" only by a three-point mixture on the *extended* real line $\mathbb{R}\cup\{-\infty,\infty\}$:

$$\nu_\varepsilon^*=(1-\varepsilon)\delta_0+\frac{\varepsilon}{2}\delta_\infty+\frac{\varepsilon}{2}\delta_{-\infty}.$$

We will need approximations which place no mass at ∞. We say distribution $\nu_{\varepsilon,\alpha}$ is *α-least-favorable* for $\eta(\,\cdot\,;\tau)$ if it is the least-dispersed distribution in $\mathcal{F}_\varepsilon$ achieving a fraction $(1-\alpha)$ of the worst case risk for $\eta(\,\cdot\,;\tau)$, i.e. if both (i)

$$\mathsf{mse}(\nu_{\varepsilon,\alpha},\tau^{\pm}(\varepsilon))=(1-\alpha)\cdot\sup_{\nu\in\mathcal{F}_\varepsilon}\mathsf{mse}(\nu,\tau^{\pm}(\varepsilon))\,,$$

and (ii) ν has the smallest second moment for which (i) is true. The least favorable distribution $\nu_{\varepsilon,\alpha}$ has the form of a three-point mixture

$$\nu_{\varepsilon,\alpha}=(1-\varepsilon)\,\delta_0+\frac{\varepsilon}{2}\delta_{\mu^{\pm}(\varepsilon,\alpha)}+\frac{\varepsilon}{2}\delta_{-\mu^{\pm}(\varepsilon,\alpha)}\,.$$

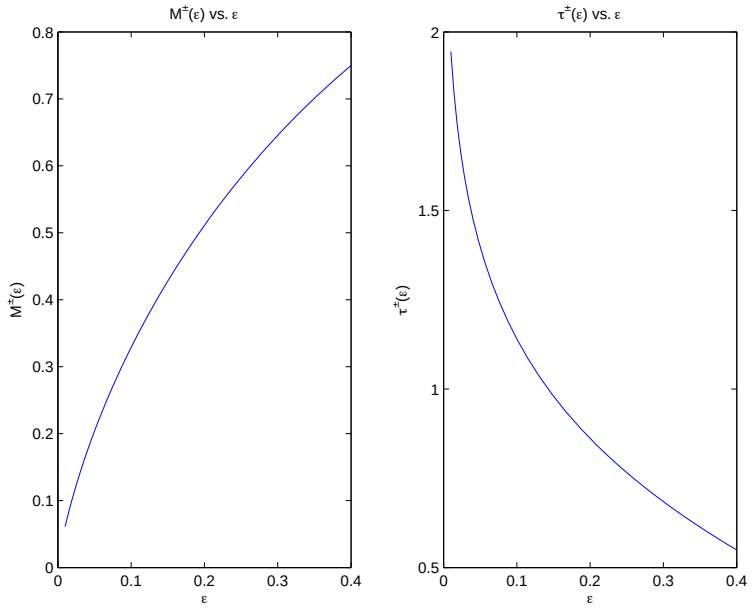

Figure 2: Left: $M^{\pm}(\varepsilon)$ as a function of ε; Right: $\tau^{\pm}(\varepsilon)$ as a function of ε.

Here $\mu^{\pm}(\varepsilon, \alpha)$ is an explicitly computable function, see below, and for $\alpha > 0$ fixed we have

$$\mu^{\pm}(\varepsilon, \alpha) \sim \sqrt{2 \log(\varepsilon^{-1})}, \qquad \text{as} \quad \varepsilon \to 0 .$$

Note in particular the relatively weak role played by α. This shows that although the precise least-favorable situation places mass at infinity, in fact, an approximately least-favorable situation is already achieved much closer to the origin.

3 Main Results

The notation of the last section allows us to state our main results.

3.1 Terminology

Definition 3.1. (Large-System Limit). *A sequence of problem size parameters n, N will be said to* **grow proportionally** *if both $n, N \to \infty$ while $n/N \to \delta \in (0, 1)$.*

Consider a sequence of random variables $(W_{n,N})$, where n, N grow proportionally. Suppose that $W_{n,N}$ converges in probability to a deterministic quantity W_∞, which may depend on $\delta > 0$. Then we say that $W_{n,N}$ has **large-system limit** *W_∞, denoted*

$$W_\infty = \text{ls lim}(W_{n,N}).$$

Definition 3.2. (Large-System Framework). *We denote by* $\text{LSF}(\delta, \rho, \sigma, \nu)$ *a sequence of problem instances $(y, A, x^0)_{n,N}$ as per Eq. (1.1) indexed by problem sizes n, N growing proportionally: $n/N \to \delta$. In each instance, the entries of the $n \times N$ matrix A are Gaussian iid $\mathsf{N}(0, 1/n)$, the entries of z^0 are Gaussian iid $\mathsf{N}(0, \sigma^2)$ and the entries of x^0 are iid ν.*

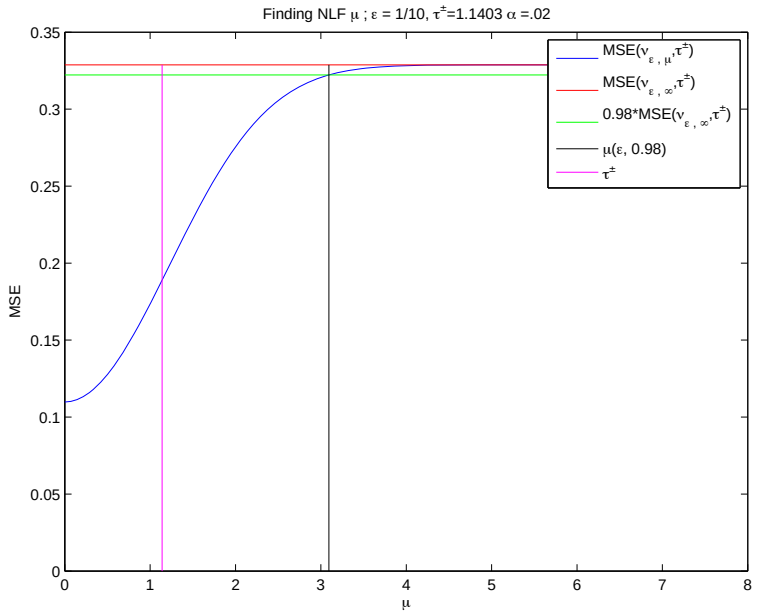

Figure 3: Illustration of α-least-favorable ν. For $\varepsilon = 1/10$, we consider soft thresholding with the minimax parameter $\tau^{\pm}(\varepsilon)$. We identify the smallest μ such that the measure $\nu_{\varepsilon,\mu} = (1-\varepsilon)\delta_0 + \frac{\varepsilon}{2}\delta_\mu + \frac{\varepsilon}{2}\delta_{-\mu}$ has $\mathsf{mse}(\nu_{\varepsilon,\mu}, \tau^*) \geq 0.98\, M^{\pm}(0.1)$ (i.e. the MSE is at least 98 % of the minimax MSE).

For the sake of concreteness we focus here on problem sequences whereby the matrix A has iid Gaussian entries. An obvious generalization of this setting would be to assume that the entries are iid with mean 0 and variance $1/n$. We expect our result to hold for a broad set of distributions in this class.

In order to match the k-sparsity condition underlying (1.1) we consider the standard framework only for $\nu \in \mathcal{F}_{\delta\rho}$.

Definition 3.3. (Observable). *Let $\hat{x}$ denote the output of a reconstruction algorithm on problem instance (y, A, x^0). An* observable J *is a function $J(y, A, x^0, \hat{x})$ of the tuple $(y, A, x^0, \hat{x})$.*

In an abuse of notation, the realized values $J_{n,N} = J(y, A, x^0, \hat{x})$ in this framework will also be called observables. An example is the observed per-coordinate MSE:

$$\mathrm{MSE} \equiv \frac{1}{N}\,\|\hat{x} - x^0\|_2^2\,.$$

The MSE depends explicitly on x^0 and implicitly on y and A (through the reconstruction algorithm). Unless specified, we shall assume that the reconstruction algorithm solves the LASSO problem (1.2), and hence $\hat{x}^{1,\lambda} = \hat{x}$. Further in the following we will drop the dependence of the observable on the arguments $y, A, x^0, \hat{x}$, and the problem dimensions n, N, when clear from context.

Definition 3.4. (Formalism). *A* formalism *is a procedure that assigns a* purported *large-system limit* Formal(J) *to an observable J in the* LSF$(\delta, \rho, \sigma, \nu)$. *This limit in general depends on δ, ρ, σ^2, and $\nu \in \mathcal{F}_{\delta\rho}$:* Formal$(J) =$ Formal$(J; \delta, \rho, \sigma, \nu)$.

Thus, in sections below we will consider $J = \mathrm{MSE}(y, A, x^0, \hat{x}^{1,\lambda})$ and describe a specific formalism yielding Formal(MSE), the formal MSE (also denoted by fMSE). Our formalism has the following character when applied to MSE: for each σ^2, δ, and probability measure ν on $\mathbb{R}$, it calculates a purported limit $\mathrm{fMSE}(\delta, \nu, \sigma)$. For a problem instance with large n, N realized from the standard framework $\mathrm{LSF}(\delta, \rho, \sigma, \nu)$, we claim the MSE will be approximately $\mathrm{fMSE}(\delta, \nu, \sigma)$. In fact we will show how to calculate formal limits for several observables. For clarity, we always attach the modifier *formal* to any result of our formalism: e.g., *formal* MSE, *formal False Alarm Rate*, *formally optimal threshold parameter*, and so on.

Definition 3.5. (Validation). *A formalism is* theoretically *validated by proving that, in the standard asymptotic framework, we have*

$$\mathrm{ls}\lim(J_{n,N}) = \mathrm{Formal}(J)$$

for a class $\mathcal{J}$ of observables to which the formalism applies, and for a range of $\mathrm{LSF}(\delta, \rho, \sigma^2, \nu)$.

A formalism is empirically *validated by showing that, for problem instances (y, A, x^0) realized from* $\mathrm{LSF}(\delta, \rho, \sigma, \nu)$ *with large N we have*

$$J_{n,N} \approx \mathrm{Formal}(J; \delta, \rho, \sigma, \nu),$$

for a collection of observables $J \in \mathcal{J}$ and a range of asymptotic framework parameters $(\delta, \rho, \sigma, \nu)$; here the approximation $\approx$ should be evaluated by usual standards of empirical science.

Obviously, theoretical validation is stronger than empirical validation, but careful empirical validation is still validation. We do not attempt here to theoretically validate this formalism in any generality; see [BM10] results in this direction. Instead we view the formalism as calculating *predictions* of empirical results. We have compared these predictions with empirical results and found a persuasive level of agreement. For example, our formalism has been used to predict the MSE of reconstructions by (1.2), and actual empirical results match the predictions, i.e.:

$$\frac{1}{N}\|\hat{x}^{1,\lambda} - x^0\|_2^2 \approx \mathrm{fMSE}(\delta, \rho, \nu, \sigma).$$

3.2 Results of the Formalism

The behavior of formal mean square error changes dramatically at the following phase boundary.

Definition 3.6 (Phase Boundary)**.** *For each $\delta \in [0, 1]$, let $\rho_{\mathrm{MSE}}(\delta)$ be the value of ρ solving*

$$M^{\pm}(\rho\delta) = \delta\,. \tag{3.1}$$

It is well known that $M^{\pm}(\varepsilon)$ is monotone increasing and concave in ε, with $M^{\pm}(0) = 0$ and $M^{\pm}(1) = 1$. As a consequence, ρ_{MSE} is also a monotone increasing function of δ, in fact $\rho_{\mathrm{MSE}}(\delta) \to 0$ as $\delta \to 0$ and $\rho_{\mathrm{MSE}}(\delta) \to 1$ as $\delta \to 1$. An explicit expression for the curve $(\delta, \rho_{\mathrm{MSE}}(\delta))$ is provided in Appendix A.

Proposition 3.1. Results of Formalism. *The formalism developed below yields the following conclusions.*

1.a In the region $\rho < \rho_{\mathrm{MSE}}(\delta)$, the minimax formal noise sensitivity obeys the formula

$$M^*(\delta,\rho) \equiv \frac{M^{\pm}(\rho\delta)}{1 - M^{\pm}(\rho\delta)/\delta}.$$

In particular, M^* is finite throughout this region.

1.b With σ^2 the noise level in (1.1), define the formal noise-plus interference *level* $\mathrm{fNPI} = \mathrm{fNPI}(\tau;\delta,\rho,\sigma,\nu)$

$$\mathrm{fNPI} = \sigma^2 + \mathrm{fMSE}/\delta,$$

and its minimax value $\mathrm{NPI}^*(\delta,\rho;\sigma) \equiv \sigma^2 \cdot (1 + M^*(\delta,\rho)/\delta)$. *For $\alpha > 0$, define*

$$\mu^*(\delta,\rho;\alpha) \equiv \mu^{\pm}(\delta\rho,\alpha) \cdot \sqrt{\mathrm{NPI}^*(\delta,\rho)}$$

In $\mathrm{LSF}(\delta,\rho,\sigma,\nu)$ *let $\nu \in \mathcal{F}_{\delta\rho}$ place fraction $1-\delta\rho$ of its mass at zero and the remaining mass equally on $\pm\mu^*(\delta,\rho;\alpha)$. This ν is $\tilde{\alpha}$-least-favorable: the formal noise sensitivity of $\hat{x}^{1,\lambda}$ equals $(1-\tilde{\alpha})M^*(\delta,\rho)$, with $(1-\tilde{\alpha}) = (1-\alpha)(1-M^{\pm}(\delta\rho))/(1-(1-\alpha)M^{\pm}(\delta\rho))$.*

1.c The formally maximin penalty *parameter obeys*

$$\lambda^*(\nu;\delta,\rho,\sigma) \equiv \tau^{\pm}(\delta\rho) \cdot \sqrt{\mathrm{fNPI}(\tau^{\pm};\delta,\rho,\sigma,\nu)} \cdot (1 - \mathrm{EqDR}(\nu;\tau^{\pm}(\delta\rho))/\delta)\,,$$

where $\mathrm{EqDR}(\cdots)$ *is the asymptotic* detection rate*, i.e. the asymptotic fraction of coordinates that are estimated to be nonzero. (An explicit expression for this quantity is given in Section 4.5.)*

In particular with this ν-adaptive choice of penalty parameter, the formal MSE *of $\hat{x}^{1,\lambda}$ does not exceed $M^* \cdot \sigma^2$.*

2 In the region $\rho > \rho_{\mathrm{MSE}}(\delta)$, the formal noise sensitivity is infinite. Throughout this phase, for each fixed number $M < \infty$, there exists $\alpha > 0$ such that the probability distribution $\nu \in \mathcal{F}_{\delta\rho}$ placing its nonzeros at $\pm\mu^*(\delta,\rho,\alpha)$, yields formal MSE *larger than M.*

We explain the formalism and derive these results in Section 4 below.

3.3 Interpretation of the Predictions

Figure 1 displays the noise sensitivity; above the phase transition boundary $\rho = \rho_{\mathrm{MSE}}(\delta)$, it is infinite. The different contour lines show positions in the δ,ρ plane where a given noise sensitivity is achieved. As one might expect, the sensitivity blows up rather dramatically as we approach the phase boundary.

Figure 4 displays the least-favorable coefficient amplitude $\mu^*(\delta,\rho,\alpha = 0.02)$. Notice that $\mu^*(\delta,\rho,\alpha)$ diverges as the phase boundary is approached. Indeed beyond the phase boundary arbitrarily large MSE can be produced by choosing μ large enough.

Figure 5 displays the value of the optimal penalization parameter amplitude $\lambda^* = \lambda^*(\nu^*_{\delta,\rho};\delta,\rho,\sigma = 1)$. Note that the parameter tends to zero as we approach phase transition.

For these figures, the region above phase transition is not decorated, because the values there are infinite or not defined.

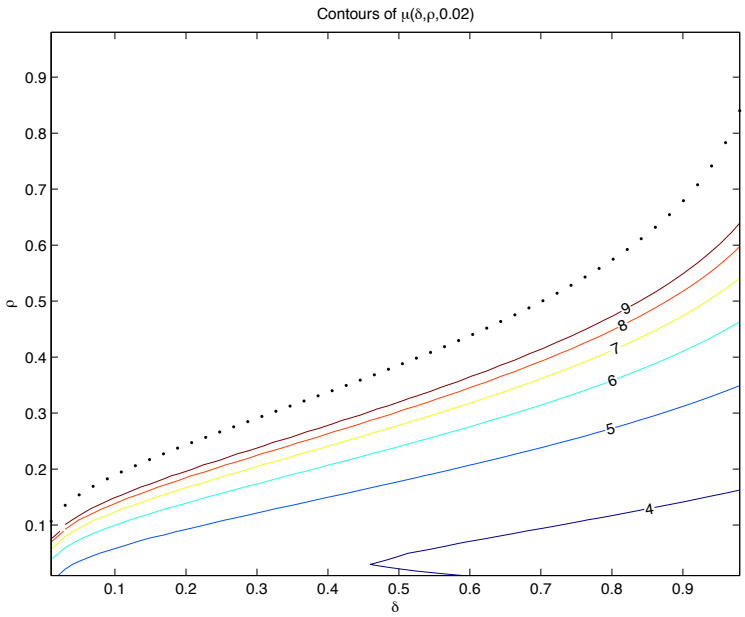

Figure 4: Contour lines of the near-least-favorable signal amplitude $\mu^*(\delta, \rho, \alpha)$ in the (ρ, δ) plane. The dotted line corresponds to the phase transition $(\delta, \rho_{\mathrm{MSE}}(\delta))$, while the colored solid lines portray level sets of $\mu^*(\delta, \rho, \alpha)$. The 3-point mixture distribution $(1-\varepsilon)\delta_0 + \frac{\varepsilon}{2}\delta_\mu + \frac{\varepsilon}{2}\delta_{-\mu}$, $(\varepsilon = \delta\rho)$ will cause 98% of the worst-case MSE. When a k-sparse vector is drawn from this distribution, its nonzeros are all at $\pm\mu$.

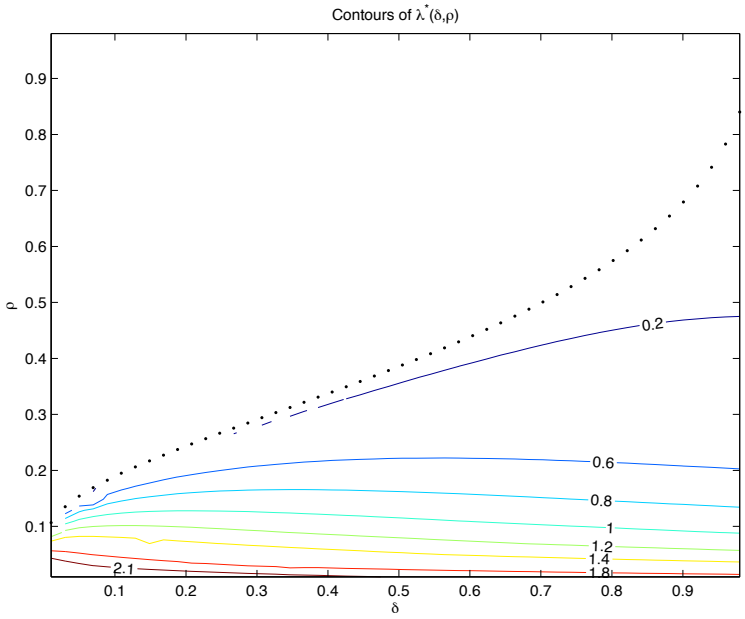

Figure 5: Contour lines of the maximin penalization parameter: $\lambda^*(\delta, \rho)$ in the (ρ, δ) plane. The dotted line corresponds to the phase transition $(\delta, \rho_{\mathrm{MSE}}(\delta))$, while thin lines are contours for $\lambda^*(\delta, \rho, \alpha)$. Close to phase transition, the maximin value approaches 0.

3.4 Comparison to other phase transitions

In view of the importance of the phase boundary for Proposition 3.1, we note the following:

Finding 3.1. Phase Boundary Equivalence. *The phase boundary ρ_{MSE} is identical to the phase boundary ρ_{ℓ_1} below which ℓ_1 minimization and ℓ_0 minimization are equivalent.*

In words, throughout the phase where ℓ_1 minimization is equivalent to ℓ_0 minimization, the solution to (1.2) has bounded formal MSE. When we are outside that phase, the solution has unbounded formal MSE. The verification of Finding 3.1 follows in two steps. First, the formulas for the phase boundary discussed in this paper are identical to the phase boundary formulas given in [DMM09b]; Second, in [DMM09b] it was shown that these formulas agree numerically with the formulas known for ρ_{ℓ_1}.

3.5 Validating the Predictions

Proposition 3.1 makes predictions for the behavior of solutions to (1.2). It will be validated empirically, by showing that such solutions behave as predicted.

In particular, simulation evidence will be presented to show that in the phase where noise sensitivity is *finite*:

1. Running (1.2) for data (y, A) generated from vectors x_0 with coordinates with distribution ν which is nearly least-favorable results in an empirical MSE approximately equal to $M^*(\delta, \rho) \cdot \sigma^2$.
2. Running (1.2) for data (y, A) generated from vectors x_0 with coordinates with distribution ν which is far from least-favorable results in empirical MSE noticeably smaller than $M^*(\delta, \rho) \cdot \sigma^2$.
3. Running (1.2) with a suboptimal penalty parameter λ results in empirical MSE noticeably greater than $M^*(\delta, \rho) \cdot \sigma^2$.

Second, in the phase where formal MSE is *infinite*:

4. Running (1.2) on vectors x_0 generated by formally least-favorable results in an empirical MSE which is very large.

Evidence for all these claims will be given below.

4 The formalism

4.1 The AMPT Algorithm

We now consider a reconstruction approach seemingly very different from $(P_{2,\lambda,1})$. This algorithm, called *first-order approximate message passing* (AMP) algorithm proceeds iteratively, starting at $\hat{x}^0 = 0$ and producing the estimate $\hat{x}^t$ of x^0 at iteration t according to the iteration:

$$z^t = y - A\hat{x}^t + z^{t-1}\frac{\mathrm{df}_t}{n} \tag{4.1}$$

$$\hat{x}^{t+1} = \eta(A^* z^t + \hat{x}^t; \theta_t)\,, \tag{4.2}$$

Here $\hat{x}^t \in \mathbb{R}^p$ is the current estimate of x^0, and $\mathrm{df}_t = \|\hat{x}^t\|_0$ is the number of nonzeros in the current estimate. Again $\eta(\,\cdot\,;\,\cdot\,)$ is the *soft threshold* nonlinearity with threshold parameter θ_t

$$\theta_t = \tau \cdot \sigma_t; \tag{4.3}$$

τ is a tuning constant, fixed throughout iterations and σ_t is an empirical measure of the scale of the residuals. Finally $z^t \in \mathbb{R}^n$ is the current *working residual*. Compare with the usual residual defined by $r^t = y - A\hat{x}^t$ via the identity $z^t = r^t + z^{t-1}\frac{\mathrm{df}_t}{n}$. The extra term in AMP plays a subtle but crucial role. [2]

4.2 Formal MSE, and its evolution

Let $\mathsf{npi}(m;\sigma,\delta) \equiv \sigma^2 + m/\delta$. We define the *MSE map* Ψ through

$$\Psi(m,\delta,\sigma,\tau,\nu) \equiv \mathsf{mse}(\mathsf{npi}(m,\sigma,\delta);\nu,\tau)\,, \tag{4.4}$$

where the function $\mathsf{mse}(\,\cdot\,;\nu,\tau)$ is the soft thresholding mean square error already introduced in Eq. (2.2). It describes the MSE of soft thresholding in a problem where the noise level is $\sqrt{\mathsf{npi}}$. A heuristic explanation of the meaning and origin of npi will be given below.

Definition 4.1. State Evolution. *The* state *is a 5-tuple* $(m;\delta,\sigma,\tau,\nu)$. State evolution *is the evolution of the state by the rule*

$$\begin{aligned}(m_t;\delta,\sigma,\tau,\nu) &\mapsto (\Psi(m_t);\delta,\sigma,\tau,\nu),\\ t &\mapsto t+1.\end{aligned}$$

As the parameters (δ,σ,τ,ν) remain fixed during evolution, we usually omit mention of them and think of state evolution simply as the iterated application of Ψ:

$$\begin{aligned}m_t &\mapsto m_{t+1} \equiv \Psi(m_t),\\ t &\mapsto t+1.\end{aligned}$$

Definition 4.2. Stable Fixed Point. *The Highest Fixed Point of the continuous function Ψ is*

$$\mathrm{HFP}(\Psi) = \sup\{m : \Psi(m) \geq m\}.$$

The stability coefficient of the continuously differentiable function Ψ is

$$\mathrm{SC}(\Psi) = \left.\frac{\mathrm{d}}{\mathrm{d}m}\Psi(m)\right|_{m=\mathrm{HFP}(\Psi)}.$$

We say that $\mathrm{HFP}(\Psi)$ *is a stable fixed point if* $0 \leq \mathrm{SC}(\Psi) < 1$.

To illustrate this, Figure 6 shows the MSE map and fixed points in three cases.

In what follows we denote by $\mu_2(\nu) = \int x^2 \mathrm{d}\nu$ the second-moment of the distribution ν.

Lemma 4.1. *Let $\Psi(\,\cdot\,) = \Psi(\,\cdot\,,\delta,\sigma,\tau,\nu)$, and assume either $\sigma^2 > 0$ or $\mu_2(\nu) > 0$. Then the sequence of iterates m_t defined by $m_{t+1} = \Psi(m_t)$ starting from $m_0 = \mu_2(\nu)$ converges monotonically to* $\mathrm{HFP}(\Psi)$*:*

$$m_t \to \mathrm{HFP}(\Psi), \qquad t \to \infty.$$

[2] A similar-looking algorithm was introduced by the authors in [DMM09a], with identical steps (4.2)-(4.1); it differed only in the choice of threshold; instead of a tuning parameter τ like in (4.3) – one that can be set freely – a fixed choice $\tau(\delta)$ was made for each specific δ. Here we call that algorithm AMPM - M for *minimax*, as explained in [DMM09b]. In contrast, the current algorithm is tunable, allowing choice of τ, we label it $\mathrm{AMPT}(\tau)$, T for tunable.

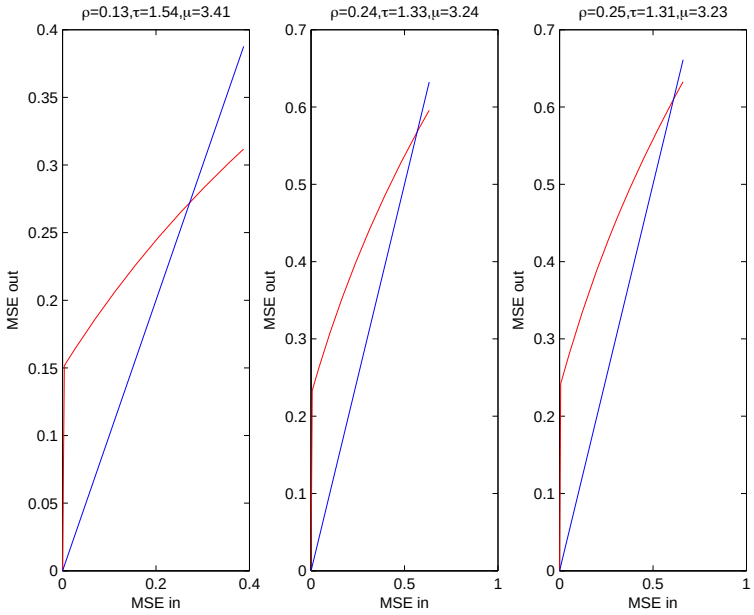

Figure 6: MSE Map Ψ in three cases, and associated fixed points. Left: $\delta = 0.25$, $\rho = \rho_{\mathrm{MSE}}/2$, $\sigma = 1$, $\nu = \nu^*(\delta, \rho, \alpha)$ Center: $\delta = 0.25$, $\rho = \rho_{\mathrm{MSE}} \times 0.95$, $\sigma = 1$, $\nu = \nu^*(\delta, \rho, \alpha)$ Right: $\delta = 0.25$, $\rho = \rho_{\mathrm{MSE}}$, $\sigma = 1$, $\nu = \nu^*(\delta, \rho, \alpha)$

Further, if $\sigma > 0$ *then* $\mathrm{HFP}(\Psi) \in (0, \infty)$ *is the unique fixed point.*

Suppose further that the stability coefficient satisfies $0 < \mathrm{SC}(\Psi) < 1$. *Then there exists a constant* $\mathcal{A}(\nu, \Psi)$ *such that*

$$\left| m_t - \mathrm{HFP}(\Psi) \right| \le \mathcal{A}(\nu, \Psi)\, \mathrm{SC}(\Psi)^t \,.$$

Finally, if $\mu_2(\nu) \ge \mathrm{HFP}(\Psi)$ *then the sequence* $\{m_t\}$ *is monotonically decreasing to* $\mu_2(\nu)$ *with*

$$(m_t - \mathrm{HFP}(\Psi)) \le \mathrm{SC}(\Psi)^t \cdot (\mu_2(\nu) - \mathrm{HFP}(\Psi)).$$

In short, barring the trivial case $x^0 = 0$, $z^0 = 0$ (no signal, no noise), state evolution converges to the highest fixed point. If the stability coefficient is smaller than 1, convergence is exponentially fast.

Proof (Lemma 4.1). This Lemma is an immediate consequence of the fact that $m \mapsto \Psi(m)$ is a concave non-decreasing function, with $\Psi(0) > 0$ as long as $\sigma > 0$ and $\Psi(0) = 0$ for $\sigma = 0$.

Indeed in [DMM09b] the authors showed that at noise level $\sigma = 0$, the MSE map $m \to \Psi(m; \delta, \sigma, \nu, \tau)$ is concave as a function of m. We have the identity

$$\Psi(m; \delta, \sigma, \nu, \tau) = \Psi(m + \sigma^2 \cdot \delta; \delta, \sigma = 0, \nu, \tau),$$

relating the noise-level 0 MSE map to the noise-level σ MSE map. From this it follows that Ψ is concave for $\sigma > 0$ as well. Also, [DMM09b] shows that $\Psi(m = 0; \delta, \sigma = 0, \nu, \tau) = 0$ and $\frac{\mathrm{d}\Psi}{\mathrm{d}m}(m = 0; \delta, \sigma = 0, \nu, \tau) > 0$, whence $\Psi(m = 0; \delta, \sigma, \nu, \tau) > 0$ for any positive noise level σ. □

In the same paper [DMM09b], the authors derived the least-favorable stability coefficient in the noiseless case $\sigma = 0$:

$$\mathrm{SC}^*(\delta, \rho, \sigma = 0) = \sup_{\nu \in \mathcal{F}_{\delta\rho}} \mathrm{SC}(\Psi(\,\cdot\,; \delta, \sigma = 0, \nu, \tau))\,.$$

They showed that, for $M^{\pm}(\delta, \rho) < \delta$ the only fixed point is at $m = 0$ and has stability coefficient

$$\mathrm{SC}^*(\delta, \rho, \sigma = 0) = M^{\pm}(\delta\rho)/\delta\,.$$

Hence, it follows that $\mathrm{SC}^*(\delta, \rho, \sigma = 0) < 1$ throughout the region $\rho < \rho_{\mathrm{MSE}}(\delta)$.

Define

$$\mathrm{SC}^*(\delta, \rho) = \sup_{\sigma > 0} \sup_{\nu \in \mathcal{F}_{\delta\rho}} \mathrm{SC}(\Psi(\,\cdot\,; \delta, \sigma, \nu, \tau)).$$

Concavity of the noise level 0 MSE map implies

$$\mathrm{SC}^*(\delta, \rho) = \mathrm{SC}^*(\delta, \rho, \sigma = 0)).$$

We therefore conclude that throughout the region $\rho < \rho_{\mathrm{MSE}}(\delta)$ For this reason, that region can also be called the *stability phase*, not only the stability coefficient is smaller than 1, $\mathrm{SC}(\Psi) < 1$, but that it can be bounded away from 1 uniformly in the signal distribution ν.

Lemma 4.2. *Throughout the region $\rho < \rho_{\mathrm{MSE}}(\delta)$, $0 < \delta < 1$, for every $\nu \in \mathcal{F}_{\delta\rho}$, we have* $\mathrm{SC}(\Psi) \le \mathrm{SC}^*(\delta, \rho) < 1$.

Outside the stability region, for each large m, we can find measures ν obeying the sparsity constraint $\nu \in \mathcal{F}_{\delta\rho}$ for which state evolution converges to a fixed point suffering equilibrium MSE $> m$. The construction in section 4.5 shows that $\mathrm{HFP}(\Psi) > \mu_2(\nu) > m$. Figure 7 shows the MSE map and the state evolution in three cases which may be compared to 6. In the first case, ρ is well below ρ_{MSE} and the fixed point is well below $\mu_2(\nu)$. In the second case, ρ is slightly below ρ_{MSE} and the fixed point is close to $\mu_2(\nu)$. In the third case, ρ is above ρ_{MSE} and the fixed point, lies above $\mu_2(\nu)$.

$\mu_2(\nu)$ is the MSE one suffers by 'doing nothing': setting threshold $\lambda = \infty$ and taking $\hat{x} = 0$. When $\mathrm{HFP}(\Psi) > \mu_2(\nu)$, one iteration of thresholding makes things *worse*, not better. In words, the phase boundary is exactly the place below which we are sure that, if $\mu_2(\nu)$ is large, a single iteration of thresholding gives an estimate $\hat{x}^1$ that is better than the starting point $\hat{x}^0$. Above the phase boundary, even a single iteration of thresholding may be a catastrophically bad thing to do.

Definition 4.3. (Equilibrium States and State-Conditional Expectations)
Consider a real-valued function $\zeta : \mathbb{R}^3 \mapsto \mathbb{R}$, its expectation in state $S = (m; \delta, \sigma, \nu)$ *is*

$$\mathcal{E}(\zeta|S) = \mathbb{E}\left\{\zeta(X, Z, \eta(X + \sqrt{\mathsf{npi}}\, Z; \tau\sqrt{\mathsf{npi}}))\right\},$$

where $\mathsf{npi} = \mathsf{npi}(m; \sigma, \delta)$ *and* $X \sim \nu$, $Z \sim \mathsf{N}(0, 1)$ *are independent random variables.*

Suppose we are given $(\delta, \sigma, \nu, \tau)$, and a fixed point m^*, $m^* = \mathrm{HFP}(\Psi)$ with $\Psi = \Psi(\,\cdot\,; \delta, \sigma, \nu, \tau)$. The tuple $S^* = (m^*; \delta, \sigma, \nu)$ is called the equilibrium state *of state evolution. The* expectation in the equilibrium state *is $\mathcal{E}(\zeta|S^*)$.*

Definition 4.4. (State Evolution Formalism for AMPT) *. Run the AMPT algorithm and assume that the sequence of estimates $(\hat{x}^t, z^t)$ converges to the fixed point $(\hat{x}^\infty, z^\infty)$. To each function $\zeta : \mathbb{R}^3 \mapsto \mathbb{R}$ associate the observable*

$$J^\zeta(y, A, x^0, \hat{x}) = \frac{1}{N}\sum_{i=1}^{N} \zeta\big(x^0(i), A^T z(i) + \hat{x}(i) - x^0(i), \hat{x}(i)\big)\,.$$

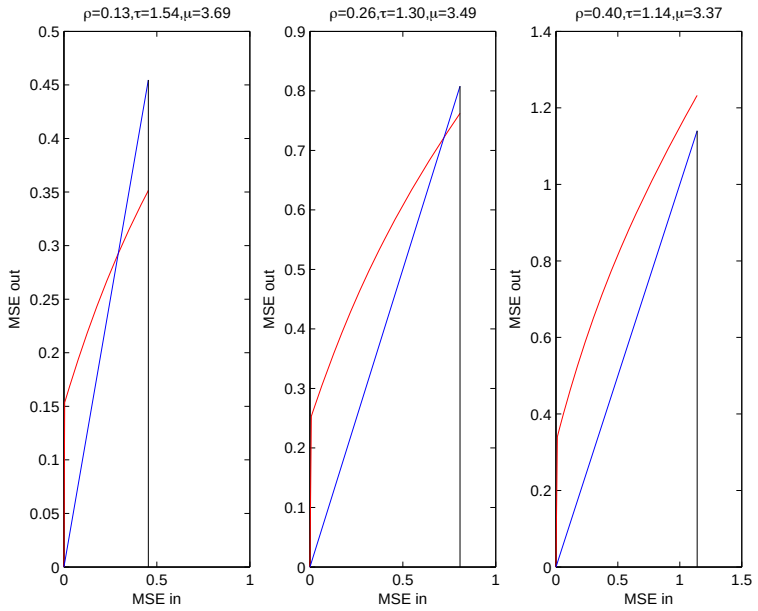

Figure 7: Crossing the phase transition: effects on MSE Map Ψ, and associated state evolution. Left: $\delta = 0.25$, $\rho = \rho_{\mathrm{MSE}}/2$, $\sigma = 1$, $\nu = \nu(\delta, \rho, 0.01)$ Middle: $\delta = 0.25$, $\rho = 0.9 \cdot \rho_{\mathrm{MSE}}$, $\sigma = 1$, $\nu = \nu(\delta, \rho, 0.01)$ Right: $\delta = 0.25$, $\rho = 1.5 \cdot \rho_{\mathrm{MSE}}$, $\sigma = 1$, $\nu = \nu(\delta, \rho, 0.01)$. In each case $\tau = \tau^{\pm}(\delta\rho)$.

Let S^* denote the equilibrium state reached by state evolution in a given situation $(\delta, \sigma, \nu, \tau)$. The state evolution formalism assigns the purported limit value

$$\mathrm{Formal}(J^{\zeta}) = \mathcal{E}(\zeta|S^*).$$

Validity of the state evolution formalism for AMPT entails that, for a sequence of problem instances (y, A, x^0) drawn from $\mathrm{LSF}(\delta, \rho, \sigma, \nu)$, the large-system limit for observable $J^{\zeta}_{n,N}$ is simply the expectation in the equilibrium state:

$$\mathrm{ls}\lim J^{\zeta}_{n,N} = \mathcal{E}(\zeta|S^*).$$

The class $\mathcal{J}$ of observables representable by the form J^{ζ} is quite rich, by choosing $\zeta(u, v, w)$ appropriately. Table 1 gives examples of well-known observables and the ζ which will generate them. Formal values for other interesting observables can in principle be obtained by combining such simple ones. For example, the False Discovery rate FDR is the ratio FDeR/DR and so the ratio of two elementary observables of the kind for which the formalism is defined. We assign it the purported limit value

$$\mathrm{Formal}(\mathrm{FDR}) = \frac{\mathrm{Formal}(\mathrm{FDeR})}{\mathrm{Formal}(\mathrm{DR})}.$$

Below we list a certain number of observables for which the formalism was checked empirically and that play an important role in characterizing the fixed point estimates.

Calculation of Formal Operating Characteristics of $\mathrm{AMPT}(\tau)$ by State Evolution

Name	Abbrev.	$\zeta = \zeta(u, v, w)$
Mean Square Error	MSE	$\zeta = (u - w)^2$
False Alarm Rate	FAR	$\zeta = 1_{\{w \neq 0 \& u = 0\}}/(1 - \rho\delta)$
Detection Rate	DR	$\zeta = 1_{\{w \neq 0\}}$
Missed Detection Rate	MDR	$\zeta = 1_{\{w = 0 \& u \neq 0\}}/(\rho\delta)$
False Detection Rate	FDeR	$\zeta = 1_{\{w \neq 0 \& u = 0\}}/(\rho\delta)$

Table 1: Some observables and their names.

Given $\delta, \sigma, \nu, \tau$, identify the fixed point $\mathrm{HFP}(\Psi(\,\cdot\,; \delta, \sigma, \nu, \tau)$. Calculate the following quantities

- Equilibrium MSE
$$\mathrm{EqMSE} = m_\infty = \mathrm{HFP}(\Psi(\,\cdot\,; \nu, \tau); \delta, \sigma).$$
- Equilibrium Noise Plus Interference Level
$$\mathsf{npi}_\infty = \frac{1}{\delta} m_\infty + \sigma^2$$
- Equilibrium Threshold (absolute units)
$$\theta_\infty = \tau \cdot \sqrt{\mathsf{npi}_\infty}.$$
- Equilibrium Mean Squared Residual. Let $Y_\infty = X + \sqrt{\mathsf{npi}_\infty}\, Z$ for $X \sim \nu$ and $Z \sim \mathsf{N}(0, 1)$ are independent. Then
$$\mathrm{EqMSR} = \mathbb{E}\big\{ \left[Y_\infty - \eta(Y_\infty; \theta_\infty) \right]^2 \big\}.$$
- Equilibrium Mean Absolute Estimate
$$\mathrm{EqMAE} = \mathbb{E}\{ |\eta(Y_\infty; \theta_\infty)| \}.$$
- Equilibrium Detection Rate
$$\mathrm{EqDR} = \mathbb{P}\{ \eta(Y_\infty; \theta_\infty) \neq 0 \}. \tag{4.5}$$
- Equilibrium Penalized MSR
$$\mathrm{EqPMSR} = \mathrm{EqMSR}/2 + \theta_\infty \cdot (1 - \mathrm{EqDR}/\delta) \cdot \mathrm{EqMAE}.$$

4.3 AMPT - LASSO Calibration

Of course at this point the reader is entitled to feel that the introduction of AMPT is a massive digression. The relevance of AMPT is indicated by the following conclusion from [DMM10b]:

Finding 4.1. *In the large system limit, the operating characteristics of* $\mathrm{AMPT}(\tau)$ *are equivalent to those of* $\mathrm{LASSO}(\lambda)$ *under an appropriate calibration* $\tau \leftrightarrow \lambda$.

By *calibration*, we mean a rescaling that maps results on one problem into results on the other problem. The notion is explained at greater length in [DMM10b]. The correct mapping can be guessed from the following remarks:

LASSO(λ): no *residual* exceeds λ: $\|A^T(y - A\hat{x}^{1,\lambda})\|_\infty \le \lambda$. Further

$$\begin{aligned} \hat{x}_i^{1,\lambda} > 0 &\Leftrightarrow (A^T(y - A\hat{x}^{1,\lambda}))_i = \lambda\,, \\ \hat{x}_i^{1,\lambda} = 0 &\Leftrightarrow |(A^T(y - A\hat{x}^{1,\lambda}))_i| < \lambda\,, \\ \hat{x}_i^{1,\lambda} < 0 &\Leftrightarrow (A^T(y - A\hat{x}^{1,\lambda}))_i = -\lambda\,. \end{aligned}$$

- AMPT(τ): At a fixed point $\hat{x}^\infty$, z^∞, no *working residual* exceeds the equilibrium threshold θ_∞: $\|A^T z^\infty\|_\infty \le \theta_\infty$. Further

$$\begin{aligned} \hat{x}_i^{\infty} > 0 &\Leftrightarrow (A^T z^{\infty})_i = \theta_\infty\,, \\ \hat{x}_i^{\infty} = 0 &\Leftrightarrow |(A^T z^{\infty})_i| < \theta_\infty\,, \\ \hat{x}_i^{\infty} < 0 &\Leftrightarrow (A^T z^{\infty})_i = -\theta_\infty\,. \end{aligned}$$

Define $df = \#\{i : \hat{x}_i^\infty \neq 0\}$. Further notice that at the AMPT fixed point $(1 - \mathrm{df}/n)z^\infty = y - A^T\hat{x}^\infty$. We can summarize these remarks in the following statement

Lemma 4.3. *Solutions $\hat{x}^{1,\lambda}$ of* LASSO(λ) *(i.e. optima of the problem (1.2)) are in correspondence with fixed points $(\hat{x}^\infty, z^\infty)$ of the* AMPT(τ) *under the bijection $\hat{x}^\infty = \hat{x}^{1,\lambda}$, $z^\infty = (y - A^T\hat{x}^{1,\lambda})/(1 - \mathrm{df}/n)$, provided the threshold parameters are in the following relation*

$$\lambda = \theta_\infty \cdot (1 - \mathrm{df}/n)\,. \tag{4.6}$$

In other words, if we have a fixed point of AMPT(τ) we can choose λ in such a way that this is also an optimum of LASSO(λ). Viceversa, any optimum of LASSO(λ) can be realized as a fixed point of AMPT(τ): notice in fact that the relation (4.6) is invertible whenever $\mathrm{df} < n$.

This simple rule gives a calibration relationship between τ and λ, i.e. a one-one correspondence between τ and λ that renders the two apparently different reconstruction procedures equivalent, provided the iteration AMPT(τ) converges rapidly to its fixed point. Our empirical results confirm that this is indeed what happens for typical large system frameworks LSF($\delta, \rho, \sigma, \nu$).

The next lemma characterizes the equilibrium calibration relation between AMP and LASSO.

Lemma 4.4. *Let* $\mathrm{EqDR}(\tau) = \mathrm{EqDR}(\tau; \delta, \rho, \nu, \sigma)$ *denote the equilibrium detection rate obtained from state evolution when the tuning parameter of AMPT is τ. Define $\tau^0(\delta, \rho, \nu, \sigma) > 0$, so that* $\mathrm{EqDR}(\tau) \le \delta$ *when $\tau > \tau^0$. For each $\lambda \ge 0$, there is a unique value $\tau(\lambda) \in [\tau_0, \infty)$ such that*

$$\lambda = \theta_\infty(\tau) \cdot (1 - \mathrm{EqDR}(\tau)/\delta).$$

We can restate Finding 4.1 in the following more convenient form.

Finding 4.2. *For each $\lambda \in [0, \infty)$ we find that* AMPT($\tau(\lambda)$) *and* LASSO(λ) *have statistically equivalent observables. In particular the* MSE*,* MAE*,* MSR*,* DR*, have the same distributions.*

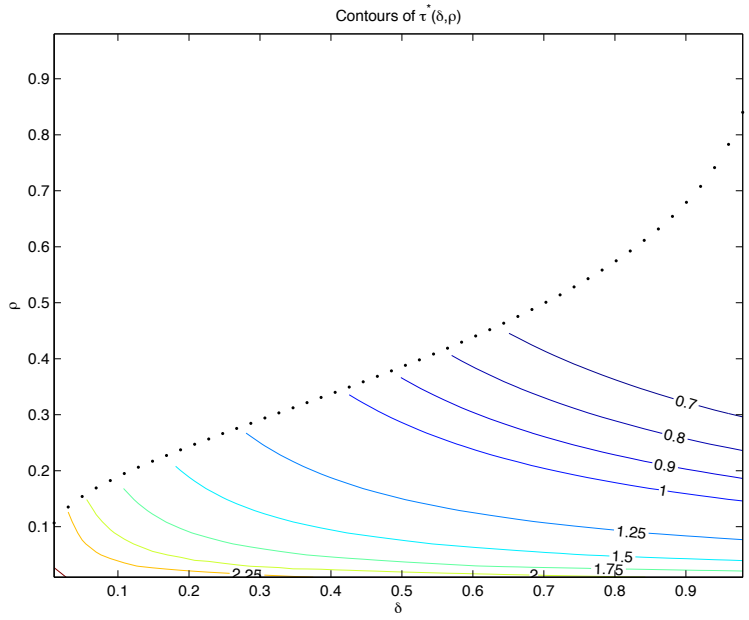

Figure 8: Contour lines of $\tau^*(\delta, \rho)$ in the (ρ, δ) plane. The dotted line corresponds to the phase transition $(\delta, \rho_{\mathrm{MSE}}(\delta))$, while thin lines are contours for $\tau^*(\delta, \rho)$

4.4 Derivation of Proposition 3.1

Consider the following Minimax Problem for AMPT(τ). With fMSE$(\tau; \delta, \rho, \sigma, \nu)$ denoting the equilibrium formal MSE for $AMPT(\tau)$ for the framework LSF$(\delta, \rho, \sigma, \nu)$, fix $\sigma = 1$ and *define*

$$M^\flat(\delta, \rho) = \inf_{\tau} \sup_{\nu \in \mathcal{F}_{\delta\rho}} \mathrm{fMSE}(\tau; \delta, \rho, \sigma = 1, \nu). \tag{4.7}$$

We will first show that this definition obeys the formula just like the one in Proposition 3.1, given for M^*. Later we show that $M^\flat = M^*$.

Proposition 4.1. *For $M^\flat$ defined by (4.7),*

$$M^\flat(\delta, \rho) = \frac{M^\pm(\delta\rho)}{1 - M^\pm(\delta\rho)/\delta} \tag{4.8}$$

The AMPT threshold rule

$$\tau^*(\delta, \rho) = \tau^\pm(\delta\rho), \quad 0 < \rho < \rho_{\mathrm{MSE}}(\delta)\,, \tag{4.9}$$

minimaxes the formal MSE:

$$\sup_{\nu \in \mathcal{F}_{\delta\rho}} \mathrm{fMSE}(\tau^*; \delta, \rho, 1, \nu) = \inf_{\tau} \sup_{\nu \in \mathcal{F}_{\delta\rho}} \mathrm{fMSE}(\tau; \delta, \rho, 1, \nu) = M^\flat(\delta, \rho). \tag{4.10}$$

Figure 8 depicts the behavior of τ^* in the (δ, ρ) plane.

Proposition 4.1. Consider $\nu \in \mathcal{F}_{\delta\rho}$ and $\sigma^2 = 1$ and set $\tau^*(\delta,\rho) = \tau^\pm(\delta\rho)$ as in the statement. Let for short $\Psi(m;\nu) = \Psi(m,\delta,\sigma=1,\tau^*,\nu) = \mathsf{mse}(\mathsf{npi}(m,1,\delta);\nu,\tau^*)$, cf. Eq. (4.4). Then $m^* = \mathrm{HFP}(\Psi)$ obeys, by definition of fixed point,

$$m^* = \Psi(m^*;\nu)\,.$$

We can use the scale invariance $\mathsf{mse}(\sigma^2;\nu,\tau^*) = \mathsf{mse}(1;\tilde\nu,\tau^*)$, where $\tilde\nu$ is a rescaled probability measure, $\tilde\nu\{x\cdot\sigma\in B\} = \nu\{x\in B\}$. For $\nu\in\mathcal{F}_{\delta\rho}$, we have $\tilde\nu\in\mathcal{F}_{\delta\rho}$ as well and we therefore obtain

$$m^* = \mathsf{mse}(\mathsf{npi}(m^*,1,\delta);\nu,\tau^*) = \mathsf{mse}(1;\tilde\nu,\tau^*)\cdot\mathsf{npi}(m^*,1,\delta) \le M^\pm(\delta\rho)\cdot\mathsf{npi}(m^*;1,\delta)\ ,$$

where we used the fact that $\tau^*(\delta,\rho) = \tau^\pm(\delta\rho)$. Hence

$$\frac{m^*}{\mathsf{npi}(m^*;1,\delta)} \le M^\pm(\delta\rho)\,.$$

The function $m\mapsto \frac{m}{\mathsf{npi}(m;\delta,1)}$ is one-to-one strictly increasing on the interval $[0,\delta)$. Thus, provided that $1-M^\pm(\delta\rho)/\delta > 0$, i.e. $\rho<\rho_{\mathrm{MSE}}$, we have

$$m^* \le \frac{M^\pm(\delta\rho)}{1-M^\pm(\delta\rho)/\delta}.$$

As this inequality applies to *any* HFP produced by our formalism, in particular the largest one consistent with $\nu\in\mathcal{F}_{\delta\rho}$, we have

$$\sup_{\nu\in\mathcal{F}_{\delta\rho}} \mathrm{fMSE}(\tau^*;\delta,\rho,1,\nu) \le \frac{M^\pm(\delta\rho)}{1-M^\pm(\delta\rho)/\delta}.$$

We now develop the reverse inequality. To do so, we make a specific choice $\overline{\nu}$ of ν. Fix $\alpha>0$ small. Now for $\varepsilon=\delta\rho$, define $\xi = \mu^\pm(\varepsilon,\alpha)\cdot\sqrt{\mathrm{NPI}^*}$, where $\mathrm{NPI}^* = 1+M^\flat/\delta$ (with $M^\flat = M^\pm(\delta\rho)/(1-M^\pm(\delta\rho)/\delta)$ as in the thesis). Let $\overline{\nu} = (1-\varepsilon)\delta_0 + (\varepsilon/2)\,\delta_{-\xi} + (\varepsilon/2)\delta_\xi$. Denote by $m^* = m^*(\overline{\nu})$ the highest fixed point corresponding to the signal distribution $\overline{\nu}$. Using once again scale invariance, we have

$$m^* = \mathsf{mse}(\mathsf{npi}(m^*,1,\delta);\overline{\nu},\tau^*) = \mathsf{mse}(1;\tilde\nu,\tau^*)\cdot\mathsf{npi}(m^*,1,\delta)\,, \tag{4.11}$$

where $\tilde\nu$ is again a rescaled probability measure, this time with $\tilde\nu\{x\cdot\sqrt{\mathsf{npi}(m^*,1,\delta)}\in B\} = \overline{\nu}\{x\in B\}$. Now since $m^*\le M^\flat$, we have $\mathsf{npi}(m^*,1,\delta)\le\mathrm{NPI}^*$, and hence

$$\frac{\xi}{\sqrt{\mathsf{npi}(m^*,1,\delta)}} = \mu^\pm(\varepsilon,\alpha)\cdot\sqrt{\frac{\mathrm{NPI}^*}{\mathsf{npi}(m^*,1,\delta)}} > \mu^\pm(\varepsilon,\alpha)\,.$$

Note that $\mathsf{mse}(m;(1-\varepsilon)\delta_0+(\varepsilon/2)\delta_{-x}+(\varepsilon/2)\delta_x,\tau)$ is monotone increasing in $|x|$. Recall that $\nu_{\varepsilon,\alpha} = (1-\varepsilon)\delta_0 + (\varepsilon/2)\delta_{-\mu^\pm(\varepsilon,\alpha)} + (\varepsilon/2)\delta_{\mu^\pm(\varepsilon,\alpha)}$ is α-least favorable for the minimax problem (2.4). Consequently,

$$\mathsf{mse}(1;\tilde\nu,\tau^*) \ge \mathsf{mse}(1;\nu_{\delta\rho,\alpha},\tau^*) = (1-\alpha)\cdot M^\pm(\delta,\rho)\,.$$

Using the scale-invariance relation, Eq. (4.11), we conclude that

$$\frac{m^*}{\mathsf{npi}(m^*;\delta,1)} \ge (1-\alpha)\cdot M^\pm(\delta\rho)\,.$$

Again, in the region $\rho < \rho_{\mathrm{MSE}}(\delta)$, the function $m \mapsto \frac{m}{\mathsf{npi}(m;\delta,1)}$ is one-to-one and monotone and therefore so

$$\mathrm{fMSE}(\tau^*;\delta,\rho,1,\overline{\nu}) \geq \frac{(1-\alpha)\cdot M^{\pm}(\delta\rho)}{1-(1-\alpha)\cdot M^{\pm}(\delta\rho)/\delta}.$$

As $\alpha > 0$ is arbitrary, we conclude

$$\sup_{\nu\in\mathcal{F}_{\delta\rho}} \mathrm{fMSE}(\tau^*;\delta,\rho,1,\nu) \geq \frac{M^{\pm}(\delta\rho)}{1-M^{\pm}(\delta\rho)/\delta}.$$

□

We now explain how this result about AMPT leads to our claim for the behavior of the LASSO estimator $\hat{x}^{1,\lambda}$. By a scale invariance the quantity (1.5) can be rewritten as a fixed-scale $\sigma = 1$ property:

$$M^*(\delta,\rho) = \sup_{\nu\in\mathcal{F}_{\delta\rho}} \inf_{\lambda} \mathrm{fMSE}(\nu,\lambda|\mathrm{LASSO}),$$

where we introduced explicit reference to the algorithm used, and dropped the irrelevant arguments. We will analogously write $\mathrm{fMSE}(\nu,\tau|\mathrm{AMPT})$ for the $\mathrm{AMPT}(\tau)$ MSE.

Proposition 4.2. *Assume the validity of our calibration relation i.e. the equivalence of formal operating characteristics of* $\mathrm{AMPT}(\tau)$ *and* $\mathrm{LASSO}(\lambda(\tau))$*. Then*

$$M^*(\delta,\rho) = M^{\flat}(\delta,\rho).$$

Also, for λ^* *as defined in Proposition 3.1,*

$$M^*(\delta,\rho) = \sup_{\nu\in\mathcal{F}_{\delta\rho}} \mathrm{fMSE}(\nu,\lambda^*(\nu;\delta,\rho,\sigma)|\mathrm{LASSO}).$$

In words, λ^* is the maximin penalization and the maximin MSE of LASSOis precisely given by the formula (4.8).

Proof. Taking the validity of our calibration relationship $\tau \leftrightarrow \lambda(\tau)$ as given, we must have

$$\mathrm{fMSE}(\nu,\lambda(\tau)|\mathrm{LASSO}) = \mathrm{fMSE}(\nu,\tau|\mathrm{AMPT}).$$

Our definition of λ^* in Proposition 3.1 is simply the calibration relation applied to the minimax AMPT threshold τ^*, i.e. $\lambda^* = \lambda(\tau^*)$. Hence assuming the validity of our calibration relation, we have:

$$\begin{aligned}
\sup_{\nu\in\mathcal{F}_{\delta\rho}} \mathrm{fMSE}(\nu,\lambda^*(\nu;\delta,\rho,\sigma)|\mathrm{LASSO}) &= \sup_{\nu\in\mathcal{F}_{\delta\rho}} \mathrm{fMSE}(\nu,\lambda(\tau^*)|\mathrm{LASSO}) \\
&= \sup_{\nu\in\mathcal{F}_{\delta\rho}} \mathrm{fMSE}(\nu,\tau^*|\mathrm{AMPT}) \\
&= \sup_{\nu\in\mathcal{F}_{\delta\rho}} \inf_{\tau} \mathrm{fMSE}(\nu,\tau|\mathrm{AMPT}) \qquad (4.12) \\
&= \sup_{\nu\in\mathcal{F}_{\delta\rho}} \inf_{\tau} \mathrm{fMSE}(\nu,\lambda(\tau)|\mathrm{LASSO}) \\
&= \sup_{\nu\in\mathcal{F}_{\delta\rho}} \inf_{\lambda} \mathrm{fMSE}(\nu,\lambda|\mathrm{LASSO}).
\end{aligned}$$

Display (4.12) shows that all these equalities are equal to $M^{\flat}(\delta,\rho)$. □

The proof of Proposition 3.1, points $1a$, $1b$, $1c$ follows immediately from the above.

4.5 Formal MSE above Phase Transition

We now make an explicit construction showing that noise sensitivity is unbounded above PT.

We first consider the AMPT algorithm above PT. Fix δ, ρ with $\rho > \rho_{\mathrm{MSE}}(\delta)$ and set $\varepsilon = \delta\rho$.

In this section we focus on 3 point distributions with mass at 0 equal to $1-\varepsilon$. With an abuse of notation we let $\mathsf{mse}(\mu,\tau)$ denote the MSE of scalar soft thresholding for amplitude of the non-zeros equal to μ, and noise variance equal to 1. In formulas, $\mathsf{mse}(\mu,\tau) \equiv \mathsf{mse}(1;(1-\varepsilon)\delta_0 + (\varepsilon/2)\delta_\mu + (\varepsilon/2)\delta_{-\mu},\tau)$, and

$$\mathsf{mse}(\mu,\tau) = (1-\varepsilon)\mathbb{E}\eta(Z;\tau)^2 + \varepsilon\mathbb{E}\left(\mu - \eta(\mu+Z;\tau)\right)^2 .$$

Consider values of the AMPT threshold τ such that $\mathsf{mse}(0,\tau) < \delta$; this will be possible for all τ sufficiently large. Pick a number $\gamma \in (0,1)$ obeying

$$1 < \gamma < \mathsf{mse}(0,\tau)/\delta. \tag{4.13}$$

Let $M^{\pm}(\varepsilon,\tau) = \sup_\mu \mathsf{mse}(\mu,\tau)$ denote the worst case risk of $\eta(\,\cdot\,;\tau)$ over the class $\mathcal{F}_\varepsilon$. Let $\mu^{\pm}(\varepsilon,\alpha,\tau)$ denote the α-least-favorable μ for threshold τ:

$$\mathsf{mse}(\mu^{\pm},\tau) = (1-\alpha)M^{\pm}(\varepsilon,\tau).$$

Define $\alpha^* = 1-\gamma\delta/M^{\pm}(\varepsilon,\tau)$, and note that $\alpha^* \in (0,1)$ by earlier assumptions. Let $\mu^* = \mu^{\pm}(\alpha^*,\tau,\varepsilon)$. A straightforward calculation along the lines of the previous section yields.

Lemma 4.5. *For the measure* $\nu = (1-\varepsilon)\delta_0 + (\varepsilon/2)\delta_{\mu^*} + (\varepsilon/2)\delta_{-\mu^*}$, *the formal* MSE *and formal* NPI *are given by*

$$\begin{aligned} \mathrm{fMSE}(\nu,\tau|\mathrm{AMPT}) &= \frac{\delta\gamma}{1-\gamma}, \\ \mathrm{fNPI}(\nu,\tau|\mathrm{AMPT}) &= \frac{1}{1-\gamma}. \end{aligned}$$

Assumption (4.13) permits us to choose γ very close to 1. Hence the above formulas show explicitly that MSE is unbounded above phase transition.

What do the formulas say about $\hat{x}^{1,\lambda}$ above PT? The τ's which can be associated to λ obey

$$0 < \mathrm{EqDR}(\nu,\tau) \le \delta,$$

where $\mathrm{EqDR}(\nu,\tau) = \mathrm{EqDR}(\tau;\delta,\rho,\nu,\sigma)$ is the equilibrium detection rate for a signal with distribution ν. Equivalently, they are those τ where the equilibrium discovery number is n or smaller.

Lemma 4.6. *For each* $\tau > 0$, *obeying both*

$$\mathsf{mse}(0,\tau) < \delta \qquad \textit{and} \qquad \mathrm{EqDR}(\nu,\tau) < \delta,$$

the parameter $\lambda \ge 0$ *defined by the calibration relation*

$$\lambda(\tau) = \frac{\tau}{\sqrt{1-\gamma}} \cdot (1 - \mathrm{EqDR}(\nu,\tau)/\delta),$$

has the formal MSE

$$\mathrm{fMSE}(\nu,\tau|\mathrm{LASSO}) = \frac{\delta\gamma}{1-\gamma}.$$

One can check that, for each $\lambda \ge 0$, for each phase space point above phase transition, the above construction allows to construct a measure μ with $\varepsilon = \delta\rho$ mass on nonzeros and with arbitrarily high formal MSE. This completes the derivation of part 2 of Proposition 3.1.

δ	ρ	ε	$M^{\pm}(\varepsilon)$	$\tau^{\pm}(\varepsilon)$	$\mu^{\pm}(\varepsilon, 0.02)$	$M^*(\delta, \rho)$	$\mu^*(\delta, \rho, 0.02)$	$\tau^*(\delta, \rho)$	λ^*
0.10	0.09	0.01	0.06	1.96	3.74	0.14	5.79	1.96	1.28
0.10	0.14	0.01	0.08	1.83	3.63	0.41	8.24	1.83	0.83
0.10	0.17	0.02	0.09	1.77	3.58	1.20	12.90	1.77	0.51
0.10	0.18	0.02	0.10	1.75	3.57	2.53	18.28	1.75	0.41
0.25	0.13	0.03	0.15	1.54	3.41	0.39	5.46	1.54	0.98
0.25	0.20	0.05	0.20	1.40	3.29	1.12	7.68	1.40	0.62
0.25	0.24	0.06	0.23	1.33	3.24	3.28	12.22	1.33	0.39
0.25	0.25	0.06	0.24	1.31	3.23	6.89	17.31	1.31	0.30
0.50	0.19	0.10	0.32	1.15	3.11	0.90	5.19	1.15	0.70
0.50	0.29	0.14	0.42	1.00	2.99	2.55	7.35	1.00	0.42
0.50	0.35	0.17	0.47	0.92	2.93	7.51	11.75	0.92	0.26
0.50	0.37	0.18	0.48	0.90	2.91	15.75	16.67	0.90	0.20

Table 2: Parameters of quasi-Least-Favorable Settings studied in the empirical results presented here.

5 Empirical Validation

So far our discussion explains how state evolution calculations are carried out so others might reproduce them. The actual 'science contribution' of our paper comes in showing that these calculations describe the actual behavior of solutions to (1.2). We check these calculations in two ways: first, to show that individual MSE predictions are accurate, and second, to show that the mathematical structures (least-favorable, minimax saddlepoint, maximin threshold) that lead to our predictions are visible in empirical results.

5.1 Below phase transition

Let $\mathrm{fMSE}(\lambda; \delta, \rho, \sigma, \nu)$ denote the formal MSE we assign to $\hat{x}^{1,\lambda}$ for problem instances from $\mathrm{LSF}(\delta, \rho, \sigma, \nu)$. Let $\mathrm{eMSE}(\lambda)_{n,N}$ denote the empirical MSE of the LASSO estimator $\hat{x}^{1,\lambda}$ in a problem instance drawn from $\mathrm{LSF}(\delta, \rho, \sigma, \nu)$ at a given problem size n, N. In claiming that the noise sensitivity of $\hat{x}^{1,\lambda}$ is bounded above by $M^*(\delta, \rho)$, we are saying that in empirical trials, the ratio eMSE/σ^2 will not be larger than M^* with statistical significance. We now present empirical evidence for this claim.

5.1.1 Accuracy of MSE at the LF signal

We first consider the accuracy of theoretical predictions at the nearly-least-favorable signals generated by $\nu_{\delta,\rho,\alpha} = (1-\varepsilon)\delta_0 + (\varepsilon/2)\delta_{-\mu^*(\delta,\rho,\alpha)} + (\varepsilon/2)\delta_{\mu^*(\delta,\rho,\alpha)}$ defined by Part 2.*b* of Proposition 3.1. If the empirical ratio eMSE/σ^2 is substantially above the theoretical bound $M^*(\delta, \rho)$, according to standards of statistical significance, we have falsified the proposition.

We consider parameter points $\delta \in \{0.10, 0.25, 0.50\}$ and $\rho \in \{\frac{1}{2} \cdot \rho_{\mathrm{MSE}}, \frac{3}{4} \cdot \rho_{\mathrm{MSE}}, \frac{9}{10} \cdot \rho_{\mathrm{MSE}}, \frac{19}{20} \cdot \rho_{\mathrm{MSE}}\}$. The predictions of the SE formalism are detailed in Table 2.

δ	ρ	μ	λ^*	fMSE	eMSE	SE
0.100	0.095	5.791	1.258	0.136	0.126	0.0029
0.100	0.142	8.242	0.804	0.380	0.329	0.0106
0.100	0.170	12.901	0.465	1.045	0.755	0.0328
0.100	0.180	18.278	0.338	2.063	1.263	0.0860
0.250	0.134	5.459	0.961	0.374	0.373	0.0046
0.250	0.201	7.683	0.592	1.028	1.002	0.0170
0.250	0.241	12.219	0.351	2.830	2.927	0.0733
0.250	0.254	17.314	0.244	5.576	5.169	0.1978
0.500	0.193	5.194	0.689	0.853	0.836	0.0078
0.500	0.289	7.354	0.400	2.329	2.251	0.0254
0.500	0.347	11.746	0.231	6.365	6.403	0.1157
0.500	0.366	16.667	0.159	12.427	11.580	0.2999

Table 3: Results at $N = 1500$. MSE of LASSO(λ^*) at nearly-least-favorable situations, together with standard errors (SE)

Results at $N = 1500$

To test these predictions, we generate in each situation $R = 200$ random realizations of size $N = 1500$ from LSF($\delta, \rho, \sigma, \nu$) with the parameters shown in Table 2 and run the LARS/LASSO solver to find the solution $\hat{x}^{1,\lambda}$. Table 3 shows the empirical average MSE in 200 trials at each tested situation.

Except at $\delta = 0.10$ the mismatch between empirical and theoretical a few to several percent - reasonable given the sample size $R = 200$. At $\delta = 0.10$, $\rho = 0.180$ – close to phase transition – there is a mismatch needing attention. (In fact, at each level of δ the most serious mismatch is at the value of ρ closest to phase transition. This can be attributed partially to the blowup of the quantity being measured as we approach phase transition.) We will pursue this mismatch below.

We also ran trials at $\delta \in \{0.15, 0.20, 0.30, 0.35, 0.40, 0.45\}$. These cases exhibited the same patterns seen above, with adequate fit except at small δ, especially near phase transition. We omit the data here.

In all our trials, we measured numerous observables – not only the MSE. The trend in mismatch between theory and observation in such observables was comparable to that seen for MSE. In [DMM09b, DMM10b], the reader can find discussion and presentation of evidence for other observables.

Results at $N = 4000$

Statistics of random sampling dictate that there always be some measure of disagreement between empirical averages and expectations. When the expectations are taken in the large-system limit, as ours are, there are additional small-N effects that appear separate from random sampling effects. However, both sorts of effects should visibly decline with increasing N.

Table 4 presents results for $N = 4000$; we expect the discrepancies to shrink when the experiments are run at larger value of N. We study the same ρ and δ that were studied for $N = 1500$, and see that the mismatches in our MSE's have grown smaller with N.

δ	ρ	μ	λ^*	fMSE	eMSE	SE
0.100	0.095	5.791	1.258	0.136	0.128	0.0016
0.100	0.142	8.242	0.804	0.380	0.348	0.0064
0.100	0.170	12.901	0.465	1.045	0.950	0.0228
0.100	0.180	18.278	0.338	2.063	1.588	0.0619
0.250	0.134	5.459	0.961	0.374	.371	0.0028
0.250	0.201	7.683	0.592	1.028	1.023	0.0106
0.250	0.241	12.219	0.351	2.830	2.703	0.0448
0.250	0.254	17.314	0.244	5.576	5.619	0.0428
0.500	0.193	5.194	0.689	0.853	0.849	0.0047
0.500	0.289	7.354	0.400	2.329	2.296	0.016
0.500	0.347	11.746	0.231	6.365	6.237	0.0677
0.500	0.366	16.667	0.159	12.427	12.394	0.171

Table 4: Results at $N = 4000$. Theoretical and empirical MSE's of LASSO(λ^*) at nearly-least-favorable situations, together with standard errors (SE).

δ	ρ	μ	λ^*	fMSE	eMSE	SE
0.100	0.095	5.791	1.258	0.136	0.131	0.0012
0.100	0.142	8.242	0.804	0.380	0.378	0.0046
0.100	0.170	12.901	0.465	1.045	1.024	0.0186
0.100	0.180	18.278	0.338	2.063	1.883	0.0458

Table 5: Results at $N = 8000$. Theoretical and empirical MSE's of LASSO(λ^*) at nearly-least-favorable situations with $\delta = 0.10$, together with standard errors (SE) of the empirical MSE's

Results at $N = 8000$

Small values of δ have the largest discrepancy specially when ρ is chosen very close to the phase transition curve. To show that this discrepancy shrinks as we increase the value of N, we do a similar experiment for $\delta = 0.10$ but this time with $N = 8000$. Table 5 summarizes the results of this simulation and shows better agreement between the formal predictions and empirical results.

The alert reader will no doubt have noticed that the discrepancy between theoretical predictions and empirical results is in many cases quite a bit larger in magnitude than the size of the the formal standard errors reported in the above tables. We emphasize that the theoretical predictions are formal limits for the $N \to \infty$ case, while empirical results take place at finite N. In both statistics and statistical physics it is quite common for mismatches between finite-N results and N-large to occur as either $O(N^{-1/2})$ (eg Normal approximation to the Poisson) or $O(N^{-1})$ effects (eg Normal approximation to fair coin tossing). Analogously, we might anticipate that mismatches in this setting of order $N^{-\alpha}$ with α either 1/2 or 1. Figure 9 presents empirical and theoretical results taken from the cases $N = 1500$, 4000, and 8000 and displays them on a common graph, with y-axis a mean-squared error (empirical or theoretical) and on the x axis the inverse system size $1/N$. The case

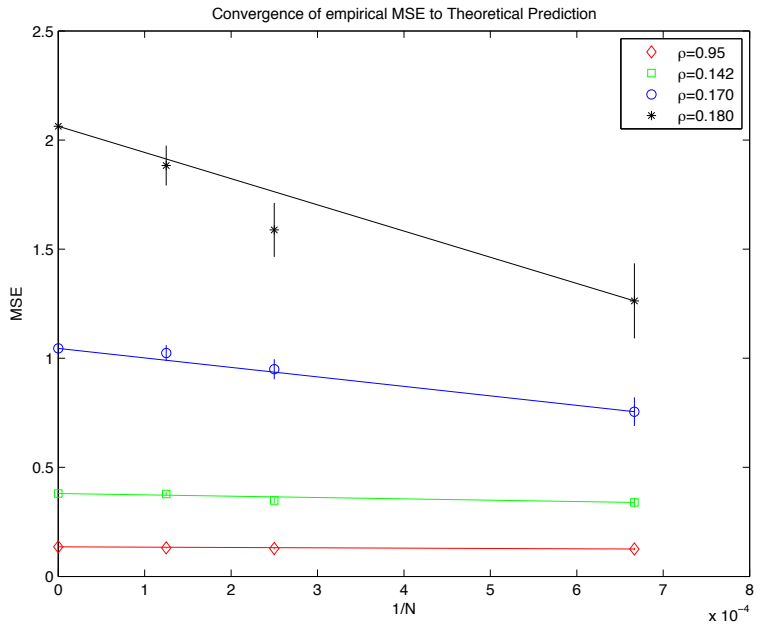

Figure 9: Finite-N scaling of empirical MSE. Empirical MSE results from the cases $N = 1500$, $N = 4000$ and $N = 8000$ and $\delta = 0.1$. Vertical axis: empirical MSE. Horizontal axis: $1/N$. Different colors/symbols indicate different values of the sparsity control parameter δ. Vertical bars denote $\pm 2SE$ limits. Theoretical predictions for the $N = \infty$ case appear at $1/N = 0$. Lines connect the cases $N = 1500$ and $N = \infty$.

$1/N = 0$ presents the formal large-system limit predicted by our calculations and the other cases $1/N > 0$ present empirical results described in the tables above. As can be seen, the discrepancy between formal MSE and empirical MSE tends to zero linearly with $1/N$. (A similar plot with $1/\sqrt{N}$ on the x-axis would not be so convincing.)

Finding 5.1. *The formal and empirical* MSE*'s at the quasi saddlepoint* (ν^*, λ^*) *show statistical agreement at the cases studied, in the sense that either the* MSE*'s are consistent with standard statistical sampling formulas, or, where they were not consistent at* $N = 1500$*, fresh data at* $N = 4000$ *and* $N = 8000$ *showed marked reductions in the anomalies confirming that the anomalies decline with increasing* N*.*

5.1.2 Existence of Game-Theoretic Saddlepoint in eMSE

Underlying our derivations of minimax formal MSE is a game-theoretic saddlepoint structure, illustrated in Figure 10. The loss function MSE has the following structure around the quasi saddlepoint (ν^*, λ^*): any variation of μ to lower values, will cause a reduction in loss, while a variation of λ to other values will cause an increase in loss.

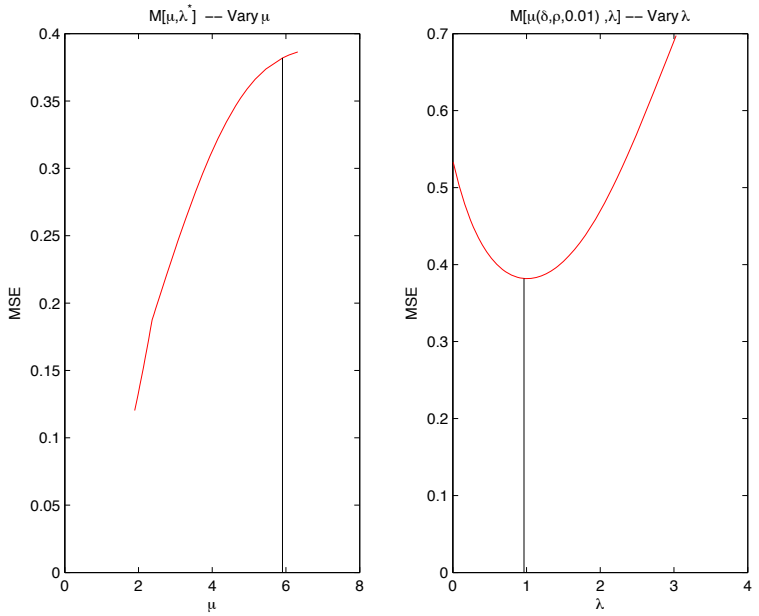

Figure 10: Saddlepoint in formal MSE. Right panel: Behavior of formal MSE as λ is varied away from λ^*. Left panel: Behavior of formal MSE as μ is varied away from μ^* in the direction of smaller values. Black lines indicate locations of μ^* and λ^*. $\delta = 0.25$, $\rho = \rho_{\mathrm{MSE}}(\delta)/2$.

5.1.3 Other penalization gives larger MSE

If our formalism is correct in deriving optimal penalization for $\hat{x}^{1,\lambda}$, we will see that changes of the penalization away from λ^* will cause MSE to increase. We consider the same situations as earlier, but now vary λ away from the minimax value, while holding the other aspects of the problem fixed. In the Appendix, Tables 7 and 8 presents numerical values of the empirical MSE obtained. Note the agreement of formal MSE, in which a saddlepoint is rigorously proven, and empirical MSE, which represents actual LARS/LASSO reconstructions. Also in this case we used $R = 200$ Monte Carlo replications.

To visualize the information in those tables, we refer to Figure 11.

5.1.4 MSE with more favorable measures is smaller

In our formalism, fixing $\lambda = \lambda^*$, and varying μ to smaller values will cause a reduction in formal MSE. Namely, if instead of $\mu^*(\delta, \rho, 0.01)$ we used $\mu^*(\delta, \rho, \alpha)$ for α significantly larger than 0.01, we would see a significant reduction in MSE, by an amount matching the predicted amount.

Recall that $\mathsf{mse}(\nu, \tau)$ denotes the 'risk' (MSE) of scalar soft thresholding as in Section 2, with input distribution ν, noise variance 1, and threshold τ. Now suppose that $\mathsf{mse}(\nu_0, \tau) > \mathsf{mse}(\nu_1, \tau)$. Then also the resulting formal noise-plus-interference obeys $\mathrm{fNPI}(\nu_0, \tau) > \mathrm{fNPI}(\nu_1, \tau)$. As noticed several times in Section 4.4, the formal MSE of AMPT obeys $\mathrm{fMSE}(\nu, \tau) = \mathsf{mse}(\tilde{\nu}, \tau) \cdot \mathrm{fNPI}(\nu, \tau)$, where $\tilde{\nu}$ denotes a rescaled probability measure (as in the proof of Proposition 4.1). Hence

$$\mathrm{fMSE}(\nu_1, \tau) \leq \mathsf{mse}(\tilde{\nu_1}, \tau) \cdot \mathrm{fNPI}(\nu_0, \tau)\,,$$

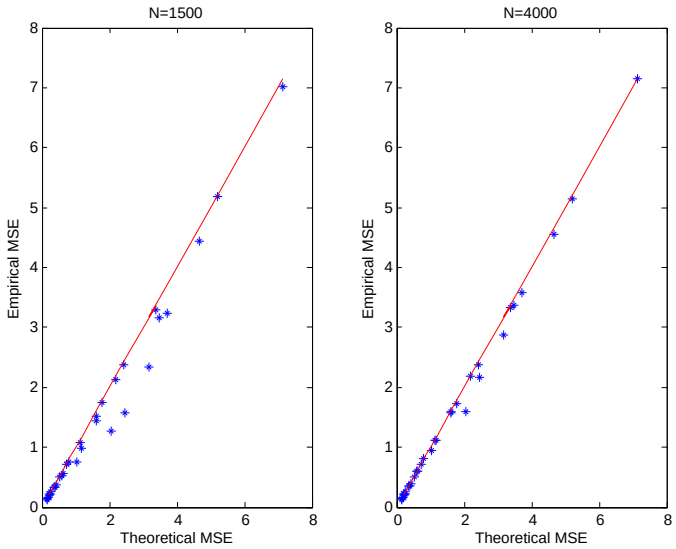

Figure 11: Scatterplots comparing Theoretical and Empirical MSE's found in Tables 7 and 8. Left Panel: results at $N = 1500$. Right Panel: results at $N = 4000$. Note visible tightening of the scatter around the identity line as N increases.

where the scaling uses $\mathrm{fNPI}(\nu_0)$. In particular, for $\mu = \mu^*(\delta, \rho, \alpha) = \mu^{\pm}(\delta \cdot \rho, \alpha)\sqrt{\mathrm{NPI}^*(\delta, \rho)}$, the three point mixture: $\nu_{\delta,\rho,\alpha}$ has

$$\mathrm{fMSE}(\nu_{\delta,\rho,\alpha}, \tau^*) \le (1 - \alpha)M^*(\delta, \rho),$$

and we ought to be able to see this. Table 9 shows results of simulations at $N = 1500$. The theoretical MSE drops as we move away from the nearly least favorable μ in the direction of smaller μ, and the empirical MSE responds similarly.

Finding 5.2. *The empirical data exhibit the saddlepoint structures predicted by the SE formalism.*

5.1.5 MSE of Mixtures

The SE formalism contains a basic mathematical structure which allows one to infer that behavior at one saddlepoint determines the global minimax value: behavior under taking convex combinations (mixtures) of measures ν.

Let $\mathsf{mse}(\nu, \lambda)$ denote the 'risk' (MSE) of scalar soft thresholding as in Section 2. For such scalar thresholding, we have the affine relation

$$\mathsf{mse}((1 - \gamma)\nu_0 + \gamma\nu_1, \tau) = (1 - \gamma)\mathsf{mse}(\nu_0, \tau) + \gamma \cdot \mathsf{mse}(\nu_1, \tau)\,.$$

Now suppose that $\mathsf{mse}(\nu_0, \tau) > \mathsf{mse}(\nu_1, \tau)$. Then also $\mathrm{NPI}(\nu_0, \tau) > \mathrm{NPI}(\nu_1, \tau)$. The formal MSE of AMPT obeys the scaling relation $\mathrm{fMSE}(\nu, \tau) = \mathsf{mse}(\tilde{\nu}, \tau) \cdot \mathrm{NPI}(\nu, \tau)$, where $\tilde{\nu}$ denotes the rescaled probability measure, argument rescaled by $1/\sqrt{NPI}$. We conclude that

$$\mathrm{fMSE}((1 - \gamma)\nu_0 + \gamma\nu_1, \tau) \le (1 - \gamma) \cdot \mathsf{mse}(\tilde{\nu}_0, \tau) \cdot \mathrm{NPI}(\nu_0, \tau) + \gamma \cdot \mathsf{mse}(\tilde{\nu}_1, \tau) \cdot \mathrm{NPI}(\nu_0, \tau), \qquad (5.1)$$

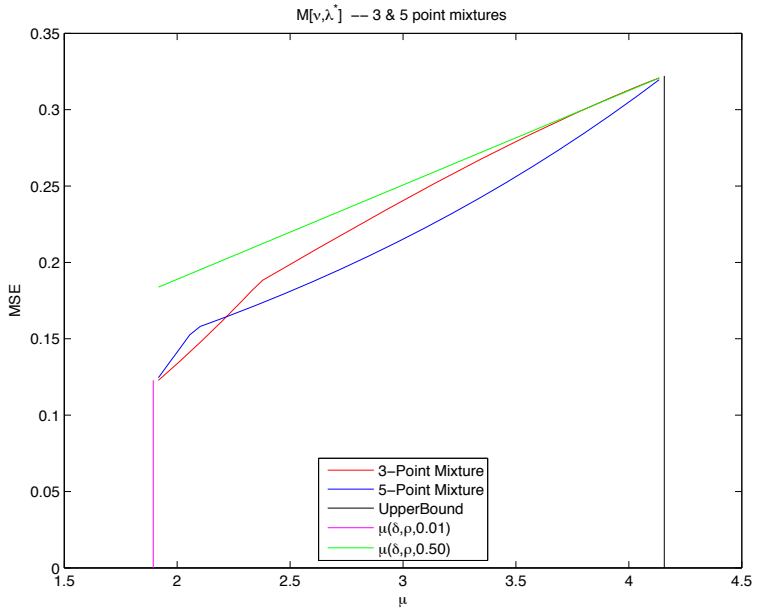

Figure 12: Convexity structures in formal MSE. Behavior of formal MSE of 5 point mixture combining nearly least-favorable μ with discount of 1% and one with discount of 50%. Also, the convexity bound (5.1) and the formal MSE of associated 3-point mixtures is displayed. $\delta = 0.25$, $\rho = \rho_{\text{MSE}}(\delta)/2$.

This 'quasi-affinity' relation allows to extend the saddlepoint structure from 3 point mixtures to more general measures.

To check this, we consider two near-least-favorable measures, $\nu_0 = \nu_{\delta,\rho,0.02}$ and $\nu_1 = \nu_{\delta,\rho,0.50}$. and generate a range of cases $\nu^{(\alpha)} = (1-\alpha)\nu_0 + \alpha\nu_1$ by varying alpha. When $\alpha \notin \{0,1\}$ this is a 5 point mixture rather than one of the 3-point mixtures we have been studying. Figure 12 displays the convexity bound (5.1), and the behavior of the formal MSE of this 5 point mixture. For comparison it also presents the formal MSE of the 3 point mixture having its mass at the weighted mean $(1-\alpha)\mu(\delta,\rho,0.02) + \alpha\mu(\delta,\rho,0.50)$. Evidently, the 5 point mixture typically has smaller MSE than the comparable 3-point mixture, and it always is below the convexity bound.

Finding 5.3. *The empirical MSE obeys the mixture inequalities predicted by the SE formalism.*

5.2 Above Phase Transition

We conducted an empirical study of the formulas derived in Section 4.5. At $\delta = 0.25$ we chose $\rho = 0.401$ - well above phase transition - and selected a range of τ and γ values allowed by our formalism. For each pair γ, τ, we generated $R = 200$ Monte Carlo realizations and obtained LASSO solutions with the given penalization parameter λ. The results are described in Table 6. The match between formal MSE and empirical MSE is acceptable.

Finding 5.4. *Running $\hat{x}^{1,\lambda}$ at the 3-point mixtures defined for the regime above phase transition in Lemma 4.6 yields empirical MSE consistent with the formulas of that Lemma.*

This validates the unboundedness of MSE of LASSO above phase transition.

6 Extensions

6.1 Positivity Constraints

A completely parallel treatment can be given for the case where $x^0 \geq 0$. In that setting, we use the positivity-constrained soft-threshold

$$\eta^+(x;\theta) = \begin{cases} x-\theta & \text{if } \theta < x, \\ 0 & \text{if } x \leq \theta, \end{cases} \tag{6.1}$$

and consider the corresponding positive-constrained thresholding minimax MSE [DJHS92]

$$M^+(\varepsilon) = \inf_{\tau>0} \sup_{\nu \in \mathcal{F}_\varepsilon^+} \mathbb{E}\Big\{ \big[\eta^+\big(X+\sigma\cdot Z;\tau\sigma\big) - X\big]^2 \Big\}, \tag{6.2}$$

where

$$\mathcal{F}_\varepsilon^+ = \{\nu \,:\, \nu \text{ is probability measure with } \nu[0,\infty) = 1, \nu(\{0\}) \geq 1-\varepsilon\}.$$

We consider the positive-constrained ℓ_1-penalized least-squares estimator $x^{1,\lambda,+}$, the solution to

$$(P^+_{2,\lambda,1}) \qquad \text{minimize}_{x\geq 0} \quad \frac{1}{2}\,\|y-Ax\|_2^2 + \lambda\|x\|_1. \tag{6.3}$$

We define the minimax, formal *noise sensitivity*:

$$M^{+,*}(\delta,\rho) = \sup_{\sigma>0} \max_{\nu} \min_{\lambda} \text{fMSE}(x^{1,\lambda,+},\nu,\sigma^2)/\sigma^2; \tag{6.4}$$

here $\nu \in \mathcal{F}^+_{\rho\delta}$ is the marginal distribution of x_0. Let $\rho^+_{\text{MSE}}(\delta)$ denote the solution of

$$M^+(\rho\delta) = \delta\,. \tag{6.5}$$

In complete analogy to (1.7) we have the formula:

$$M^{+,*}(\delta,\rho) = \begin{cases} \frac{M^+(\delta\rho)}{1-M^+(\delta\rho)/\delta}, & \rho < \rho^+_{\text{MSE}}(\delta), \\ \infty, & \rho \geq \rho^+_{\text{MSE}}(\delta). \end{cases} \tag{6.6}$$

The argument is the same as above, using the AMP formalism, with obvious modifications. The papers [DMM09a, DMM09b] show in more detail how to make arguments for AMP that apply simultaneously to the sign-constrained and unconstrained case. All other features of Proposition 3.1 carry over, with obvious substitutions. Figure 13 shows the phase transition for the positivity constrained case, as well as the contour lines of $M^{+,*}$. Again in analogy to the sign-unconstrained case, the phase boundary ρ^+_{MSE} occurs at precisely the same location at the phase boundary for ℓ_1-ℓ_0 equivalence; as earlier this can be inferred from formulas in this paper and in [DMM09a].

6.2 Other Classes of Matrices

We focused here on matrices A with Gaussian iid entries.

Previously, extensive empirical evidence was presented by Donoho and Tanner [DT09], that pure ℓ_1-minimization has its ℓ_1-ℓ_0 equivalence phase transition at the boundary $\rho^{\pm}_{\text{MSE}}$ not only for Gaussian matrices but for a wide collection of ensembles, including partial Fourier, partial Hadamard, expander graphs, iid ± 1. This is the noiseless, $\lambda = 0$ case of the general noisy, $\lambda \geq 0$ case studied here.

We believe that similar results to those obtained here hold for matrices A with uniformly bounded iid entries with zero mean and variance $1/n$. In fact, we believe our results should extend to a broader universality class including matrices with iid entries with same mean and variance, under an appropriate light tail condition.

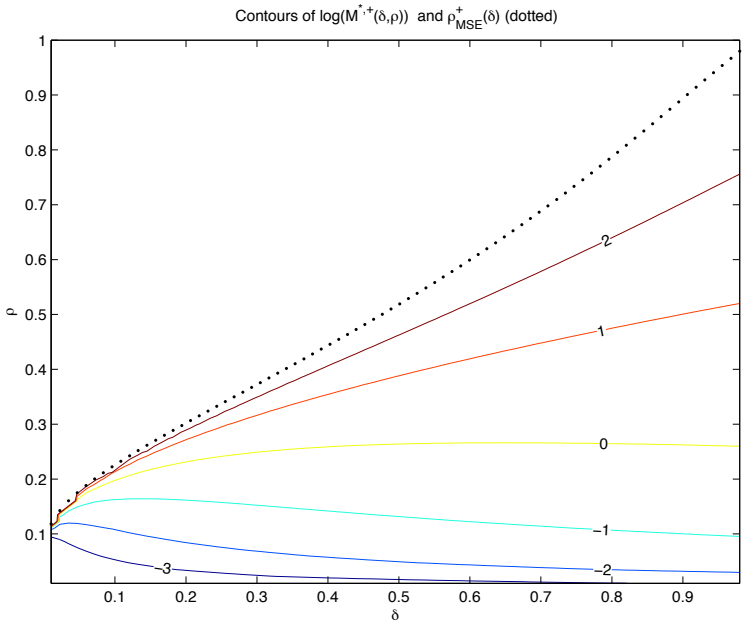

Figure 13: Contour lines of the positivity-constrained minimax noise sensitivity $M^{*,+}(\delta,\rho)$ in the (ρ,δ) plane. The dotted black curve graphs the phase boundary $(\delta,\rho^+_{\rm MSE}(\delta))$. Above this curve, $M^{*,+}(\delta,\rho)=\infty$. The colored lines present level sets of $M^{*,+}(\delta,\rho)=1/8,\,1/4,\,1/2,\,1,\,2,\,4$ (from bottom to top).

7 Relations with Statistical Physics and Information Theory

This section outlines the relations of the approach advocated here with ideas in information theory (in particular, with the theory of sparse graph codes), graphical models and statistical physics (more precisely spin glass theory). We will not discuss such relations in full mathematical detail, but only stress some important points that might be useful for researchers in each of those fields.

7.1 Information theory and message passing algorithms

Message passing algorithms, and most notably belief propagation, have been intensively investigated in coding theory and communications, in particular because of their success in decoding sparse graph codes [RU08]. Belief propagation is defined whenever the *a posteriori* joint distribution of the variables to be inferred x conditional on the observations y can be written as a graphical model. In the present case this is easily done, provided the a priori probability distribution of the signal $x=(x_1,\dots,x_N)$ takes a $\nu=\nu_1\times\nu_2\cdots\times\nu_N$. The posterior is then

$$\mu(\mathrm{d}x)=\frac{1}{Z}\prod_{a=1}^{n}\exp\Big\{-\frac{\beta}{2}(y_a-(Ax)_a)^2\Big\}\prod_{i=1}^{N}\nu_i(\mathrm{d}x_i)\,.$$

Graphical models of this type were (implicitly or explicitly) considered in the context of multiuser detection [Kab03, NS05, MPT06, MT06]. The underlying factor graph [KFL01] is the complete bipartite graph over N variable nodes and n factor nodes.

Applying belief propagation to such a model incurs two obvious difficulties: the graph is dense (and hence the complexity per iteration scales at least like n^3, and in fact worse), and the alphabet is continuous (and hence messages are not finitely representable). As discussed in [DMM10a], AMP solves these problem. From the information theoretical perspective, the term $+z^{t-1}\mathrm{df}_t/n$ in Eq. (4.1) corresponds to 'subtracting intrinsic information'.

An important difference between the message passing algorithms in coding theory and what is presented here is that no precise information is available on the priors ν_i in Eq. (7.1). Therefore the AMP rules should not be sensitive to the prior. The use of the soft threshold function $\eta(\,\cdot\,;\theta)$ makes the AMP robust within the class of sparse priors. Also, it is directly related to the ℓ_1 regularization in the LASSO.

In coding theory, message passing algorithms are analyzed through density evolution [RU08]. The common justification for density evolution is that the underlying graph is random and sparse, and hence converges locally to a tree in the large system limit. In the case of trees density evolution is exact, hence it is asymptotically exact for sparse random graphs. Such an easy justification is not available in the cases of dense graphs treated here and a deeper mathematical analysis is required. In [BM10], this analysis was carried out in the case of Gaussian matrices A. It remains a challenge to generalize such analysis beyond the case of Gaussian matrices A.

Having outlined the relation with belief propagation and coding, it is important to clarify a key point. In the context of sparse graph coding, belief propagation performances and MAP (maximum a posteriori probability) performances do not generally coincide even asymptotically (although they are intimately related [MMU04, MMRU09]). In the present paper we instead conjecture that AMP and LASSO have asymptotically equal MSE under appropriate calibration. This is due to the fact that that the state evolution recursion $m_t \mapsto m_{t+1} = \Psi(m_t)$ has only one stable fixed point.

7.2 Statistical physics

There is a well studied connection between statistical physics techniques and message passing algorithms [MM09]. In particular, the sum-product algorithm corresponds to the Bethe-Peierls approximation in statistical physics, and its fixed points are stationary points of the Bethe free energy. In the context of spin glass theory, the Bethe-Peierls approximation is also referred to as the 'replica symmetric cavity[3] method'.

The Bethe-Peierls approximation postulates a set of non-linear equations on quantities that correspond to the belief propagation messages, and allow to compute posterior marginals under the distribution (7.1). In the special cases of spin glasses on the complete graph (the celebrated Sherrington-Kirkpatrick model), these equations reduce to the so-called TAP equations, named after Thouless, Anderson and Palmer who first used them [TAP77].

The original TAP equations where a set of non-linear equations for local magnetizations (i.e. expectations of a single variable). Thouless, Anderson and Palmer first recognized that naive mean field is not accurate enough in the spin glass model, and corrected it by adding the so called Onsager reaction term that is analogous to the term $+z^{t-1}\mathrm{df}_t/n$ in Eq. (4.1). More than 30 years after the original paper, a complete mathematical justification of the TAP equations remains an open problem in spin glass theory, although important partial results exist [Tal03]. While the connection between belief propagation and Bethe-Peierls approximation stimulated a considerable amount of research [YFW05], the algorithmic uses of TAP equations have received only sparse attention. Remarkable

[3]When this terminology is used in statistical physics, the emphasis is rather on properties of random instances.

exceptions include [OW01, Kab03, NS05].

7.3 State evolution and replica calculations

Within statistical mechanics, the typical properties of probability measures of the form (7.1) are studied using the replica method or the cavity method [MM09]. These can be described as non-rigorous but mathematically sophisticated techniques. Despite intense efforts and some spectacular progresses [Tal03], even a precise statement of the assumptions implicit in such techniques is missing, in a general setting.

The fixed points of state evolution describe the output of the corresponding AMP, when the latter is run for a sufficiently large number of iterations (independent of the dimensions n, N). It is well known, within statistical mechanics [MM09], that the fixed point equations do indeed coincide with the equations obtained form the replica method (in its replica-symmetric form).

During the last few months, several papers investigated compressed sensing problems using the replica method [RFG09, KWT09, GBS09]. In view of the discussion above, it is not surprising that these results can be recovered from the state evolution formalism put forward in [DMM09a]. Let us mention that the latter has several advantages over the replica method:

(1) It is more concrete, and its assumptions can be checked quantitatively through simulations;

(2) It is intimately related to efficient message passing algorithms;

(3) It actually allows to predict the performances of these algorithms (including for instance precise convergence time estimates);

(4) It actually leads to rigorous statements, at least in the case of Gaussian sensing matrices.

A Some explicit formulae

This appendix contain some formulae and analytical derivation omitted from the main text.

The phase boundary curve admits the parametric expression

$$\delta = \frac{2\phi(\tau)}{\tau + 2(\phi(\tau) - \tau\Phi(-\tau))}\,, \tag{A.1}$$

$$\rho = 1 - \frac{\tau\Phi(-\tau)}{\phi(\tau)}\,, \tag{A.2}$$

$$\tag{A.3}$$

This is simply obtained from Eq. (2.5). If we call $G_\varepsilon(\tau)$ the function on the right hand side, then the parametric expression given here follows from $\delta = G_\varepsilon(\tau)$ and $G'_\varepsilon(\tau) = 0$ (which are equivalent to $\delta = M^{\pm}(\varepsilon)$).

B Tables

This appendix contains table of empirical results supporting our claims.

References

[BCT09] J. D. Blanchard, C. Cartis, and J. Tanner, *The restricted isometry property and ℓ_q-regularization: Phase transitions for sparse approximation*, submitted, 2009.

[BM10] M. Bayati and A. Montanari, *The dynamics of message passing on dense graphs, with applications to compressed sensing*, arXiv:1001.3448, 2010.

[CD95] S. S. Chen and D. L. Donoho, *Examples of basis pursuit*, Proceedings of Wavelet Applications in Signal and Image Processing III (San Diego, CA), 1995.

[CDS98] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Journal on Scientific Computing **20** (1998), 33–61.

[CT05] E. J. Candes and T. Tao, *Decoding by linear programming*, IEEE Trans. on Inform. Theory **51** (2005), 4203–4215.

[CT07] ———, *The Dantzig selector: Statistical estimation when p is much larger than n*, Ann. Statist. **35** (2007), 2313–2351.

[DJ94] D. L. Donoho and I. M. Johnstone, *Minimax risk over l_p balls*, Prob. Th. and Rel. Fields **99** (1994), 277–303.

[DJHS92] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, *Maximum entropy and the nearly black object*, Journal of the Royal Statistical Society, Series B (Methodological) **54** (1992), no. 1, 41–81.

[DMM09a] D. L. Donoho, A. Maleki, and A. Montanari, *Message passing algorithms for compressed sensing*, Proceedings of the National Academy of Sciences **106** (2009), no. 45, 18914–18915.

[DMM09b] ———, *Online supplement to message passing algorithms for compressed sensing*, Proceedings of the National Academy of Sciences **106** (2009), no. 45, 18914–18915.

[DMM10a] ———, *Message Passing Algorithms for Compressed Sensing: I. Motivation and Construction*, IEEE Information Theory Workshop (Cairo, Egypt), January 2010.

[DMM10b] ———, *Theoretical prediction of lasso operatng characteristics*, manuscript, 2010.

[DT09] D.L. Donoho and J. Tanner, *Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing*, Phil. Trans. Roy. Soc. A **367** (2009), 4273–4293.

[GBS09] D. Guo, D. Baron, and S. Shamai, *A single-letter characterization of optimal noisy compressed sensing*, 47th Annual Allerton Conference (Monticello, IL), September 2009.

[Kab03] Y. Kabashima, *A CDMA multiuser detection algorithm on the basis of belief propagation*, J. Phys. A **36** (2003), 11111–11121.

[KFL01] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, *Factor Graphs and the Sum-Product Algorithm*, IEEE Trans. on Inform. Theory **47** (2001), no. 2, 498–519.

[KWT09] Y. Kabashima, T. Wadayama, and T. Tanaka, *A typical reconstruction limit for compressed sensing based on lp-norm minimization*, J. Stat. Mech. (2009), L09003.

[MM09] M. Mézard and A. Montanari, *Information, Physics and Computation*, Oxford University Press, Oxford, 2009.

[MMRU09] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke, *The Generalized Area Theorem and Some of its Consequences*, IEEE Trans. on Inform. Theory **55** (2009), no. 11, 4793–4821.

[MMU04] C. Méasson, A. Montanari, and R. Urbanke, *Maxwell Construction: The Hidden Bridge between Iterative and Maximum a Posteriori Decoding*, IEEE Trans. on Inform. Theory **54** (2004), no. 12, 5277–5307.

[MPT06] A. Montanari, B. Prabhakar, and D. Tse, *Belief Propagation Based Multi–User Detection*, IEEE Information Theory Workshop (Punta del Este, Uruguay), March 2006.

[MT06] A. Montanari and D. Tse, *Analysis of Belief Propagation for Non-Linear Problems: The Example of CDMA (or: How to Prove Tanaka's Formula)*, IEEE Information Theory Workshop (Punta del Este, Uruguay), March 2006.

[NS05] J. P. Neirotti and D. Saad, *Improved message passing for inference in densely connected systems*, Europhys. Lett. **71** (2005), 866–872.

[OW01] M. Opper and O. Winther, *From Naive Mean Field Theory to the TAP Equations*, Advanced mean field methods: theory and practice (M. Opper and D. Saad, eds.), MIT Press, 2001, pp. 7–20.

[RFG09] S. Rangan, A. K. Fletcher, and V. K. Goyal, *Asymptotic analysis of map estimation via the replica method and applications to compressed sensing*, arXiv:0906.3234, 2009.

[RU08] T. J. Richardson and R. Urbanke, *Modern Coding Theory*, Cambridge University Press, Cambridge, 2008, Available online at `http://lthcwww.epfl.ch/mct/index.php`.

[Tal03] M. Talagrand, *Spin glasses: A challenge for mathematicians*, Springer-Verlag, Berlin, 2003.

[TAP77] D. J. Thouless, P. W. Anderson, and R. G. Palmer, *Solution of 'Solvable model of a spin glass'*, Phil. Mag. **35** (1977), 593–601.

[Tib96] R. Tibshirani, *Regression shrinkage and selection with the lasso*, J. Royal. Statist. Soc B **58** (1996), 267–288.

[XH09] W. Xu and B. Hassibi, *On sharp performance bounds for robust sparse signal recovery*, Proc. of the IEEE Int. Symp. on Inform. Theory (Seoul, Korea), July 2009, pp. 493–497.

[YFW05] J. S. Yedidia, W. T. Freeman, and Y. Weiss, *Constructing free energy approximations and generalized belief propagation algorithms*, IEEE Trans. on Inform. Theory **51** (2005), 2282–2313.

δ	ρ	γ	μ	τ	λ	fMSE	eMSE
0.250	0.401	0.75	2.8740	1.500	0.9840	0.750	0.746
0.250	0.401	0.85	4.142	1.500	1.168	1.417	1.425
0.250	0.401	0.90	5.345	1.500	1.366	2.250	2.239
0.250	0.401	0.95	7.954	1.500	1.841	4.750	4.724
0.250	0.401	0.97	10.4781	1.500	2.328	8.083	8.126
0.250	0.401	0.98	12.9628	1.500	2.822	12.250	12.327
0.250	0.401	0.99	18.5172	1.500	3.949	24.750	24.601
0.250	0.401	0.995	26.3191	1.500	5.5558	49.750	49.837
0.250	0.401	0.75	2.9031	2.000	2.8766	1.417	1.409
0.250	0.401	0.85	4.058	2.000	3.626	2.250	2.238
0.250	0.401	0.90	5.158	2.000	4.385	2.250	2.238
0.250	0.401	0.95	7.560	2.000	6.122	4.750	4.742
0.250	0.401	0.97	9.897	2.000	7.861	8.083	8.054
0.250	0.401	0.98	12.205	2.000	9.6019	12.250	12.215
0.250	0.401	0.99	17.380	2.000	13.5425	24.750	24.634
0.250	0.401	0.995	24.662	2.000	19.1260	49.750	49.424
0.250	0.401	0.75	2.817	2.500	4.501	1.417	1.409
0.250	0.401	0.85	3.896	2.500	5.750	2.250	2.241
0.250	0.401	0.90	4.926	2.500	7.004	2.250	2.241
0.250	0.401	0.95	7.181	2.500	9.848	4.750	4.712
0.250	0.401	0.97	9.380	2.500	12.6846	8.083	8.050
0.250	0.401	0.98	11.555	2.500	15.5170	12.250	12.215
0.250	0.401	0.99	16.436	2.500	21.9183	24.750	24.619
0.250	0.401	0.995	23.311	2.500	30.9786	49.750	49.442
0.250	0.401	0.75	2.7649	3.000	5.8144	1.417	1.408
0.250	0.401	0.85	3.809	3.000	7.4730	2.250	2.241
0.250	0.401	0.90	4.806	3.000	9.131	2.250	2.241
0.250	0.401	0.95	6.991	3.000	12.880	4.750	4.735
0.250	0.401	0.97	9.125	3.000	16.6113	8.083	8.053
0.250	0.401	0.98	11.236	3.000	20.3339	12.250	12.218
0.250	0.401	0.99	15.975	3.000	28.7413	24.750	24.621
0.250	0.401	0.995	22.652	3.000	40.6356	49.750	49.419

Table 6: Results above Phase transition. Parameters of the construction as well as theoretical predictions and resulting empirical MSE figures

Table 7: $N = 1500$, λ dependence of the MSE at fixed μ

δ	ρ	μ	λ	fMSE	eMSE	SE
0.100	0.095	5.791	0.402	0.152	0.140	0.0029
0.100	0.095	5.791	1.258	0.136	0.126	0.0029
0.100	0.095	5.791	2.037	0.142	0.133	0.0030
0.100	0.095	5.791	3.169	0.174	0.164	0.0028
0.100	0.095	5.791	4.948	0.239	0.228	0.0025
0.100	0.142	8.242	0.804	0.380	0.329	0.0106
0.100	0.142	8.242	1.960	0.408	0.374	0.0087
0.100	0.142	8.242	3.824	0.534	0.504	0.0084
0.100	0.142	8.242	6.865	0.737	0.716	0.0059
0.100	0.170	12.906	0.465	1.045	0.755	0.0328
0.100	0.170	12.906	2.298	1.178	0.992	0.0326
0.100	0.170	12.906	5.461	1.619	1.520	0.0273
0.100	0.170	12.906	10.607	2.197	2.138	0.0139
0.100	0.180	18.278	0.338	2.063	1.263	0.0860
0.100	0.180	18.278	2.934	2.467	1.573	0.0741
0.100	0.180	18.278	7.545	3.474	3.167	0.0569
0.100	0.180	18.278	14.997	4.677	4.438	0.0321
0.250	0.134	5.459	0.518	0.403	0.390	0.0044
0.250	0.134	5.459	0.961	0.374	0.373	0.0046
0.250	0.134	5.459	1.419	0.385	0.386	0.0046
0.250	0.134	5.459	2.165	0.452	0.455	0.0053
0.250	0.134	5.459	3.555	0.623	0.612	0.0042
0.250	0.201	7.683	0.036	1.151	1.155	0.0174
0.250	0.201	7.683	0.592	1.028	1.002	0.0170
0.250	0.201	7.683	1.183	1.073	1.069	0.0169
0.250	0.201	7.683	2.243	1.324	1.293	0.0158
0.250	0.201	7.683	4.392	1.861	1.837	0.0114
0.250	0.241	12.219	0.351	2.830	2.927	0.0733
0.250	0.241	12.219	1.219	3.065	2.998	0.0661
0.250	0.241	12.219	2.917	4.055	4.020	0.0485
0.250	0.241	12.219	6.444	5.709	5.625	0.0330
0.250	0.254	17.314	0.244	5.576	5.169	0.1978
0.250	0.254	17.314	1.433	6.291	5.992	0.1712
0.250	0.254	17.314	3.855	8.667	8.492	0.1148
0.250	0.254	17.314	8.886	12.154	11.978	0.0697
0.500	0.193	5.194	0.176	1.121	1.108	0.0080
0.500	0.193	5.194	0.470	0.894	0.879	0.0070
0.500	0.193	5.194	0.689	0.853	0.836	0.0078
0.500	0.193	5.194	0.933	0.866	0.862	0.008
0.500	0.193	5.194	1.355	0.965	0.960	0.0078
0.500	0.193	5.194	2.237	1.273	1.263	0.0075
0.500	0.289	7.354	0.179	2.489	2.438	0.0262
0.500	0.289	7.354	0.400	2.329	2.251	0.0254
0.500	0.289	7.354	0.655	2.377	2.329	0.0268
0.500	0.289	7.354	1.137	2.728	2.718	0.0256
0.500	0.289	7.354	2.258	3.704	3.672	0.0212
0.500	0.347	11.746	0.231	6.365	6.403	0.1157
0.500	0.347	11.746	0.558	6.624	6.349	0.1121
0.500	0.347	11.746	1.227	8.089	7.813	0.0819
0.500	0.347	11.746	2.882	11.288	11.189	0.0692
0.500	0.366	16.666	0.159	12.427	11.580	0.2998
0.500	0.366	16.666	0.582	13.300	13.565	0.2851
0.500	0.366	16.666	1.491	17.028	17.194	0.2082
0.500	0.366	16.666	3.769	23.994	23.571	0.1409

Table 8: $N = 4000$, λ dependence of the MSE at fixed μ

δ	ρ	μ	λ	fMSE	eMSE	SE
0.100	0.095	5.791	0.402	0.152	0.144	0.0017
0.100	0.095	5.791	1.258	0.136	0.128	0.0016
0.100	0.095	5.791	2.037	0.142	0.133	0.0016
0.100	0.095	5.791	3.169	0.174	0.168	0.0016
0.100	0.095	5.791	4.948	0.239	0.228	0.0012
0.100	0.142	8.242	0.804	0.380	0.348	0.0064
0.100	0.142	8.242	1.960	0.408	0.389	0.0058
0.100	0.142	8.242	3.824	0.534	0.510	0.0051
0.100	0.142	8.242	6.865	0.737	0.716	0.0034
0.100	0.170	12.906	0.465	1.045	0.950	0.0228
0.100	0.170	12.906	2.298	1.178	1.111	0.0232
0.100	0.170	12.906	5.461	1.619	1.591	0.0159
0.100	0.170	12.906	10.607	2.197	2.182	0.008
0.100	0.180	18.278	0.338	2.063	1.588	0.0619
0.100	0.180	18.278	2.934	2.467	2.171	0.0532
0.100	0.180	18.278	7.545	3.474	3.367	0.0312
0.100	0.180	18.278	14.997	4.677	4.551	0.0169
0.150	0.109	5.631	0.420	0.236	0.228	0.0022
0.150	0.109	5.631	1.073	0.212	0.209	0.0023
0.150	0.109	5.631	1.700	0.218	0.213	0.0021
0.150	0.109	5.631	2.657	0.260	0.251	0.0024
0.150	0.109	5.631	4.284	0.359	0.353	0.0017
0.150	0.163	8.030	0.720	0.588	0.595	0.0072
0.150	0.163	8.030	1.614	0.626	0.610	0.0078
0.150	0.163	8.030	3.135	0.804	0.807	0.0058
0.150	0.163	8.030	5.868	1.125	1.118	0.0047
0.150	0.196	12.577	0.434	1.612	1.572	0.0341
0.150	0.196	12.577	1.814	1.792	1.720	0.0281
0.150	0.196	12.577	4.339	2.433	2.383	0.0205
0.150	0.196	12.577	8.903	3.359	3.333	0.0126
0.150	0.207	17.814	0.305	3.185	2.864	0.0861
0.150	0.207	17.814	2.231	3.715	3.582	0.0722
0.150	0.207	17.814	5.879	5.202	5.141	0.0439
0.150	0.207	17.814	12.455	7.142	7.154	0.0269

Table 9: $N = 1500$, μ dependence of the MSE at fixed λ

δ	ρ	μ	λ	fMSE	eMSE	SE
0.100	0.095	5.291	1.253	0.131	0.125	0.0022
0.100	0.095	5.541	1.256	0.134	0.132	0.0025
0.100	0.095	5.691	1.257	0.135	0.126	0.0027
0.100	0.095	5.791	1.258	0.136	0.129	0.0024
0.100	0.095	5.891	1.259	0.137	0.125	0.0027
0.100	0.095	6.041	1.260	0.138	0.126	0.0030
0.100	0.095	6.291	1.262	0.139	0.127	0.0028
0.100	0.095	6.791	1.264	0.141	0.125	0.0031
0.100	0.142	7.242	0.794	0.349	0.317	0.0074
0.100	0.142	7.742	0.800	0.366	0.335	0.0084
0.100	0.142	7.992	0.802	0.373	0.351	0.0089
0.100	0.142	8.000	0.802	0.373	0.362	0.0094
0.250	0.134	4.459	0.952	0.338	0.336	0.0036
0.250	0.134	4.959	0.957	0.359	0.346	0.0040
0.250	0.134	5.209	0.959	0.367	0.356	0.0044
0.250	0.134	5.359	0.960	0.371	0.373	0.0049
0.250	0.134	5.459	0.961	0.374	0.362	0.0047
0.250	0.134	5.559	0.962	0.376	0.367	0.0045
0.250	0.134	5.709	0.962	0.379	0.372	0.0048
0.250	0.134	5.959	0.963	0.383	0.362	0.0052
0.250	0.134	6.459	0.964	0.387	0.387	0.0058
0.250	0.201	6.683	0.587	0.939	0.899	0.0126
0.250	0.201	7.183	0.590	0.988	0.965	0.0147
0.250	0.201	7.433	0.591	1.009	0.956	0.0147
0.250	0.201	7.583	0.592	1.021	1.027	0.0155
0.500	0.193	4.194	0.684	0.769	0.770	0.0052
0.500	0.193	4.694	0.687	0.818	0.823	0.0066
0.500	0.193	4.944	0.688	0.837	0.838	0.0073
0.500	0.193	5.094	0.689	0.847	0.835	0.0068
0.500	0.193	5.194	0.689	0.853	0.834	0.0073
0.500	0.193	5.294	0.689	0.858	0.845	0.0079
0.500	0.193	5.444	0.690	0.865	0.863	0.0079
0.500	0.193	5.694	0.690	0.874	0.887	0.0085
0.500	0.193	6.194	0.691	0.886	0.868	0.0085
0.500	0.289	6.354	0.398	2.119	2.071	0.0195
0.500	0.289	6.854	0.399	2.234	2.214	0.0235
0.500	0.289	7.104	0.399	2.284	2.157	0.0252
0.500	0.289	7.254	0.400	2.313	2.271	0.0244
0.500	0.289	7.354	0.400	2.329	2.316	0.0275
0.500	0.289	7.454	0.400	2.346	2.287	0.0287
0.500	0.289	7.604	0.400	2.370	2.327	0.0306
0.500	0.289	7.854	0.401	2.404	2.339	0.0284
0.500	0.289	8.000	0.401	2.422	2.409	0.0300

Table 10: $N = 1500$, MSE for 5-point prior

δ	ρ	μ	λ	Theoretical MSE	Empirical MSE	α
0.250	0.134	1.894	0.857	0.120	0.151	0
0.250	0.134	2.171	0.897	0.162	0.163	0.122
0.250	0.134	2.447	0.901	0.178	0.177	0.244
0.250	0.134	2.724	0.906	0.196	0.195	0.366
0.250	0.134	3.001	0.912	0.215	0.210	0.488
0.250	0.134	3.277	0.918	0.237	0.236	0.611
0.250	0.134	3.554	0.926	0.261	0.257	0.7333
0.250	0.134	3.830	0.935	0.287	0.280	0.8556
0.250	0.134	4.107	0.945	0.317	0.307	0.9778
0.250	0.134	4.383	0.957	0.348	0.359	1.1000