Consider the noisy underdetermined system of linear equations: $y = Ax_0 + z$, with $A$ an $n \times N$ measurement matrix, $n < N$, and $z \sim N(0, \sigma^2 I)$ a Gaussian white noise. Both $y$ and $A$ are known, both $x_0$ and $z$ are unknown, and we seek an approximation to $x_0$. When $x_0$ has few nonzeros, useful approximations are often obtained by $\ell_1$-penalized $\ell_2$ minimization, in which the reconstruction $\hat{x}_{1,\lambda}$ solves $\min \{\|y - Ax\|_2^2 / 2 + \lambda \|x\|_1\}$.

Consider the reconstruction mean-squared error $\text{MSE} = E\|\hat{x}_{1,\lambda} - x_0\|_2^2 / N$, and define the ratio $\text{MSE}/\sigma^2$ as the noise sensitivity. Consider matrices $A$ with iid Gaussian entries and a large-system limit in which $n, N \to \infty$ with $n/N \to \delta$ and $k/n \to \rho$. We develop exact expressions for the asymptotic MSE of $\hat{x}_{1,\lambda}$, and evaluate its worst-case noise sensitivity over all types of $k$-sparse signals. The phase space $0 \leq \delta, \rho \leq 1$ is partitioned by the curve $\rho = \rho_{\text{MSE}}(\delta)$ into two regions. Formal noise sensitivity is bounded throughout the region $\rho < \rho_{\text{MSE}}(\delta)$ and is unbounded throughout the region $\rho > \rho_{\text{MSE}}(\delta)$.

The phase boundary $\rho = \rho_{\text{MSE}}(\delta)$ 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 these predictions, 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 $\ell_1$-penalized reconstruction. The focus of the present paper is on computing the minimax formal MSE within the class of sparse signals $x^0$.


**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.
1 Introduction

Consider the noisy underdetermined system of linear equations:

\[ y = Ax_0 + z, \]  

(1.1)

where \( A \) is an \( n \times N \) matrix, \( n < N \), normalized in such a way that \( \sum_{i=1}^{n} A_{ij}^2 \approx 1 \) (see below for more precise definitions). The \( N \)-vector \( x_0 \) is \( k \)-sparse (i.e. it has at most \( k \) non-zero entries), and \( z \in \mathbb{R}^n \) is a Gaussian white noise \( z \sim N(0, \sigma^2 I) \). Both \( y \) and \( A \) are known, both \( x_0 \) and \( z \) are unknown, and we seek an approximation to \( x_0 \).

A very popular approach estimates \( x_0 \) via the solution \( \hat{x}^{1,\lambda} \) of the following convex optimization problem

\[
(P_{2,\lambda,1}) \quad \text{minimize} \quad \frac{1}{2} \|y - Ax\|^2 + \lambda \|x\|_1.
\]

(1.2)

Thousands of articles use or study this approach, which has variously been called LASSO, Basis Pursuit, or more prosaically, \( \ell_1 \)-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 \( z = 0 \), 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 \( \ell_1 \) minimization,

\[
(P_1) \quad \text{minimize} \quad \|x\|_1, \\
\text{subject to} \quad y = Ax.
\]

(1.3)

(1.4)

For the classes of instances considered in the present paper, the minimum of this problem is achieved at a single point to be denoted by \( \hat{x}^{1,0} \). The problem \( (P_1) \) can be regarded as the \( \lambda \to 0 \) limit of \( (P_{2,\lambda,1}) \) above: its solution obeys \( \hat{x}^{1,0} = \lim_{\lambda \to 0} \hat{x}^{1,\lambda} \).

The most precise information about the behavior of \( \hat{x}^{1,0} \) is obtained by large-system analysis. Let \( n, N \) tend to infinity so that \( n \sim \delta N \) and correspondingly let the number of nonzeros \( k \sim \rho n \). We thus have a phase space \( 0 \leq \delta, \rho \leq 1 \), expressing different combinations of undersampling \( \delta \) and sparsity \( \rho \). When the matrix \( A \) has iid Gaussian elements, phase space \( 0 \leq \delta, \rho \leq 1 \) can be divided into two components, or phases, separated by a curve \( \rho = \rho_\delta(\delta) \), 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 \( \ell_1 \) minimization perfectly recovers the sparse vector \( \hat{x}^{1,0} \) (here and below ‘with high probability’ means with probability converging to 1 as \( n, N \to \infty \)). Above this curve, sparsity is not sufficient: we have \( \hat{x}^{1,0} \neq x_0 \) with high probability. Hence the curve \( \rho = \rho_\delta(\delta) \), \( 0 < \delta < 1 \), indicates the precise tradeoff between undersampling and sparsity. Exact expressions for the curve \( (\delta, \rho_\delta(\delta)) \) were first derived in [Don06] using results from random polytope geometry (see also [DT05, DT09]). This is sometimes referred to as the phase boundary for weak \( \ell_0-\ell_1 \) equivalence, whereby ‘\( \ell_0-\ell_1 \)’ refers to the striking phenomenon that \( \ell_1 \) minimization recovers the sparsest solution, and the qualification ‘weak’ corresponds to the fact that \( \hat{x}^{1,0} = x_0 \) for most (but not necessarily for all) signal vectors \( x_0 \) with sufficient 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 based on the Restricted Isometry Property

\(^1\)Here and below we write \( a \sim b \) if \( a/b \to 1 \) as both quantities tend to infinity.
(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, RIP implies that, under sufficient sparsity of $x_0$, with high probability one has stability bounds\(^2\) of the form $\|\hat{x}^{1,\lambda} - x_0\|_2^2 \leq C(\delta, \rho)k^2 \sigma^2 \log N/k$. The region where $C(\delta, \rho) < \infty$ was originally an implicitly known, but clearly nonempty region of the $(\delta, \rho)$ phase space. Blanchard, Cartis and Tanner [BCT11] recently improved the estimates of $C$ in the case of Gaussian matrices $A$, by large deviations analysis, and by developing an asymmetric version of 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 paper, we show that in the presence of Gaussian white noise, the mean-squared error of the optimally tuned $\ell_1$ 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 asymptotic 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 $\lambda$, and the hardest-to-recover sparse vector. As we show, the noise sensitivity exhibits a phase transition in the undersampling-sparsity $(\delta, \rho)$ domain along a curve $\rho = \rho_{\text{MSE}}(\delta)$, and this curve is precisely the same as the $\ell_1-\ell_0$ equivalence curve $\rho_{\ell_1}$.

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 $z$ was still vanishing, but the vector $x_0$ was allowed to be an $\ell_1$-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 $\rho = \rho_{\text{MSE}}(\delta)$.

1.1 Results of our formalism

In the following we shall introduce a formal mean square error formula

$$f_{\text{MSE}} = f_{\text{MSE}}(\lambda; \nu, \delta, \sigma). \quad (1.5)$$

This provides a prediction for the mean square error in terms of only three features of the reconstruction problem $(P_{2, \lambda, 1})$: the regularization parameter $\lambda$, the undersampling ratio $\delta$, the noise level $\sigma$ and the empirical distribution of the entries of $x_0$, $\nu$. This is defined by:

$$\nu = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_{0,i}}, \quad (1.6)$$

where $\delta_a$ denotes the point mass measure at $a \in \mathbb{R}$, i.e., the measure defined as $\delta_a(S) = 1$ if $a \in S$, and $\delta_a(S) = 0$ otherwise. As discussed further below, the formal mean square error $f_{\text{MSE}}(\lambda; \nu, \delta, \sigma)$ coincides with the asymptotic MSE for sequences of random matrices with iid Gaussian entries, and is conjectured to be asymptotically exact for a broader class of sensing matrices. The fact that the predicted MSE depends on $x_0$ only through the empirical distribution $\nu$ is a consequence of the fact that the columns of $A$ are exchangeable. Note that $x_0$ is not assumed to be random.

\(^2\)Alternative bounds of the form $\|\hat{x}^{1,\lambda} - x_0\|_2^2 \leq C_2 N \sigma^2$ have also been proved [CRT06]. These however do not capture the correct dependence on the sparsity of $x_0$. 

---
In the present paper, we take the point of view that the formal MSE, \( f_{\text{MSE}}(\lambda; \nu, \delta, \sigma) \), is a scientific prediction. As such, we can deduce useful consequences and test them. For many practical scenarios, it is unrealistic to assume that the empirical distribution \( \nu \) of the signal, or the noise variance \( \sigma^2 \) are known. Therefore, a useful way of dealing with these unknowns is to consider the worst case \( \nu \) under the assumption that the fraction of nonzero entries in \( x_0 \) is at most \( \rho \delta \). In other words, \( \nu(\{0\}) \geq 1 - \rho \delta \). Under this model, the mean square error is completely characterized by the following minimax (formal) noise sensitivity:

\[
M^*(\delta, \rho) \equiv \sup_{\sigma > 0} \max_{\nu: \nu(\{0\}) \geq 1 - \rho \delta} \min_{\lambda} \frac{1}{\sigma^2} f_{\text{MSE}}(\lambda; \nu, \delta, \sigma). \tag{1.7}
\]

A certain (finite) value of \( M^*(\delta, \rho) \) ensures that, for any choice of \( \nu \) satisfying our sparsity assumption, and for any choice of \( \sigma^2 \), the mean square error incurred by \( \hat{x}^{1,\lambda} \) will not be larger (asymptotically) than \( M^*(\delta, \rho) \sigma^2 \), provided \( \lambda \) is appropriately chosen. Further, \( M^*(\delta, \rho) \) is the smallest constant such that this statement holds.

Our main theoretical result is an explicit expression for \( M^*(\delta, \rho) \) in terms of well known quantities in statistical decision theory. Let \( M^\pm(\epsilon) \) denote the minimax MSE of scalar thresholding, defined in Section 2 below. Roughly speaking this is the noise sensitivity in the scalar case \( N = n = 1 \), and is a well known function that can be easily evaluated numerically. Further, let \( \rho_{\text{MSE}}(\delta) \) denote the solution of

\[
M^\pm(\rho \delta) = \delta. \tag{1.8}
\]

We then have the following formula expressing the noise sensitivity of \((P_{2,\lambda,1})\) in terms of the scalar noise sensitivity \( M^\pm(\cdot) \):

\[
M^*(\delta, \rho) = \begin{cases} 
M^\pm(\delta \rho) \frac{1}{1 - M^\pm(\delta \rho)/\delta}, & \text{if } \rho < \rho_{\text{MSE}}(\delta), \\
\infty, & \text{if } \rho \geq \rho_{\text{MSE}}(\delta).
\end{cases} \tag{1.9}
\]

The quantity (1.7) can be interpreted as 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. In Proposition 3.1 below we give formulas for the minimax strategies achieving the value of the game \( M^*(\delta, \rho) \). The phase-transition structure evident in (1.9) is saying that above the curve \( \rho_{\text{MSE}} \), 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. The formal MSE expression \( f_{\text{MSE}}(\lambda; \nu, \delta, \sigma) \) is obtained from the analysis of an iterative soft thresholding scheme introduced by the authors in [DMM09a], and called there AMP (for approximate message passing). The AMP algorithm is conjectured to converge with high probability to the LASSO solution, i.e., to the solution of the optimization problem (1.2), for various ensembles of the measurement matrix \( A \). Remarkably, and unlike analogous algorithms for compressed sensing reconstruction, AMP admits an exact large system analysis in the limit \( N, n \to \infty \). The analysis describes the algorithm progress.
in terms of a one-dimensional recursion called state evolution. Hence, fixed points of state evolution describe the large system limit of LASSO optimum.

The electronic supplement of [DMM09a] documented these arguments through a series of large scale numerical simulations. Subsequently (and after the submission of the present paper) this analysis was confirmed rigorously in a sequence of two papers [BM11, BM10] in the case in which the matrix \( A \) is Gaussian with iid entries \( A_{ij} \sim N(0, 1/n) \). In particular, these papers prove that, almost surely

\[
\text{fMSE}(\lambda; \nu, \delta, \sigma) = \lim_{N \to \infty} \frac{1}{N} \| \hat{x}^{1, \lambda} - x_0 \|^2,
\]

where the limit is taken along suitably defined sequences of instances of increasing dimensions \( N, n \) (see below for a precise definition). Under the hypotheses of [BM10], the results presented here are rigorous in an appropriate\(^3\) asymptotic sense.

In the present paper, we start from (1.10) that we interpret as a scientific prediction. We use it to derive useful concrete consequences and test them numerically. There is a number of important reasons to proceed in this way. (1) The proof of (1.10) in [BM11, BM10] only covers the Gaussian case, and it is likely that the exact domain of validity of such formula will remain mathematically unknown for a long time. It is therefore important to put forward a principled approach to numerical validation, which we exemplify here. (2) Even in the Gaussian case, (1.10) refers to specific sequences

\[^3\]A rigorous definition of minimax problem in the \( N \to \infty \) requires some care due to uniformity issues. A framework that settles these issues is put forward in [DJMM11].
of instances. Using the minimax approach in statistical decision theory, we show how to extract uniform stability bounds over classes of sparse signals. This is a crucial step, that can and should be carried out in several other classes of problems for which formal predictions of the type (1.10) are available (e.g. through statistical physics methods). (3) Predictions of the form (1.10) are only asymptotic, and it is an outstanding mathematical challenge to convert them into non-asymptotic estimates. It is therefore important to check their precision at moderate dimensions \( N, n \). (4) Last but not the least, in practice it is the present empirical investigations that eventually inspired the rigorous arguments, and not viceversa.

1.3 Empirical Validation

Our theoretical results are validated here by computational experiments which show that the predictions of our formulas are accurate. Even more importantly, we demonstrated the underlying formal structure leading to our predictions – least-favorable objects, game-theoretic saddlepoints of the MSE payoff function, maximin tuning of \( \lambda \), unboundedness of the noise sensitivity above phase transition—can all be observed experimentally. The usefulness of the present approach –based on interplay of asymptotic analysis and statistical decision theory– is demonstrated by the many predictions about quantities with clear operational significance.

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 \( \ell^1 \)-penalized least squares fails to have the limit behavior discovered here.

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 argue that the ideas developed here and in companion papers give access to a richer set of operational predictions, to efficient reconstruction algorithms (such as AMP) [DMM09a] and opens the way to rigorous work [BM11, BM10].

1.4 Plan of the paper

The rest of the paper is organized as follows. In Section 2 we discuss a special case \( A = I \) in (1.1). This background is crucial for a proper understanding of our main results, presented in Section 3. Indeed the latter describes a characterization of the minimax MSE for the compressed sensing problem, in terms of the minimax MSE of the scalar (or \( A = I \)) problem. The formalism at the core of this result is further explained in Section 4. Section 5 validates these predictions through extensive simulations. Finally, Sections 7 and 6 discuss connections with earlier literature and generalizations of the present work. The appendix A describes a characterization of the phase boundary \( \rho = \rho_{\text{MSE}}(\delta) \) that can be useful for reproducing our results.

2 Minimax MSE of Soft Thresholding

In this section we consider the simpler case in which the matrix \( A \) in model (1.1) is replaced by the identity \( A = I \). In other words

\[
y_i = x_{0,i} + z_i, \quad 1 \leq i \leq N,
\]
with \( z_i \sim N(0, \sigma^2) \) independent and identically distributed, and we assume that \( x_0 \) is sparse. This problem has received considerable interest, e.g., [DJHS92, DJ94]. We will recall some basic facts from this literature, and introduce some straightforward generalizations that are important for the present case.

Under the sparsity assumption, it makes sense to consider the soft thresholding estimator

\[
\hat{x}_i = \eta(y_i; \tau \sigma), \quad 1 \leq i \leq N, \tag{2.1}
\]

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

\[
\eta(x; \theta) = \begin{cases} 
    x - \theta & \text{if } \theta < x, \\
    0 & \text{if } -\theta \leq x \leq \theta, \\
    x + \theta & \text{if } x \leq -\theta.
\end{cases} \tag{2.2}
\]

In words, the estimator (2.1) ‘shrinks’ the observations \( y \) towards the origin by a multiple \( \tau \) of the noise level \( \sigma \).

In place of studying \( x_0 \) which are \( k \)-sparse following [DJHS92, DJ94], we can consider random variables \( X \) which obey \( P\{X \neq 0\} \leq \varepsilon \), where \( \varepsilon = k/n \). So let \( F_\varepsilon \) denote the set of probability measures placing all but \( \varepsilon \) of their mass at the origin:

\[ F_\varepsilon = \{ \nu : \nu \text{ is probability measure with } \nu(\{0\}) \geq 1 - \varepsilon \}. \]

We define the soft thresholding mean square error by

\[
\text{mse}(\sigma^2; \nu, \tau) \equiv E\left\{ \left[ \eta(X + \sigma Z; \tau \sigma) - X \right]^2 \right\}. \tag{2.3}
\]

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

It is important to allow general \( \sigma \) in the calculations below. However, note the scale invariance

\[
\text{mse}(\sigma^2; \nu, \tau) = \sigma^2 \text{mse}(1; \nu^{1/\sigma}, \tau), \tag{2.4}
\]

where \( \nu^a \) is the probability distribution obtained by rescaling \( \nu \): \( \nu^a(S) = \nu(\{ x : a x \in S \}) \). 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 \( \sigma \) argument, and simply write \( \text{mse}(\nu, \tau) \equiv \text{mse}(1; \nu, \tau) \).

The minimax threshold \textit{MSE} was defined in [DJHS92, DJ94] by

\[
M^\pm(\varepsilon) = \inf_{\tau > 0} \sup_{\nu \in F_\varepsilon} \text{mse}(\nu, \tau). \tag{2.5}
\]

(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 \( X \geq 0 \).) This minimax problem was studied in [DJ94] where one can find a considerable amount of information.

With a slight abuse of notation, denote by \( M^\pm(\varepsilon, \tau) \) the worst case MSE \textit{for a fixed} \( \tau > 0 \):

\[
M^\pm(\varepsilon, \tau) \equiv \sup_{\nu \in F_\varepsilon} \text{mse}(\nu, \tau). \tag{2.6}
\]

It follows immediately from [DJ94] that the supremum is “achieved” only by a three-point mixture on the \textit{extended} real line \( \mathbb{R} \cup \{-\infty, \infty\} \), namely

\[
\nu^*_\varepsilon = (1 - \varepsilon)\delta_0 + \frac{\varepsilon}{2}\delta_\infty + \frac{\varepsilon}{2}\delta_{-\infty}. \tag{2.7}
\]
Substituting this form in Eq. (2.3), we obtain the following explicit formula for the worst-case MSE

\[ M^\pm(\epsilon, \tau) = \epsilon (1 + \tau^2) + (1 - \epsilon) [2(1 + \tau^2) \Phi(-\tau) - 2\tau \phi(\tau)] , \tag{2.8} \]

with \( \phi(z) = \exp(-z^2/2)/\sqrt{2\pi} \) the standard normal density and \( \Phi(z) = \int_{-\infty}^{z} \phi(x) \, dx \) the Gaussian distribution.

We let \( \tau^\pm(\epsilon) \) be the minimax threshold level, i.e., the value of \( \tau \) minimizing the worst case MSE \( M^\pm(\epsilon, \tau) \), as given by (2.8). The corresponding value of the MSE is the minimax MSE \( M^\pm(\epsilon) = M^\pm(\epsilon, \tau^\pm(\epsilon)) \). Hence \( \tau^\pm(\epsilon) \) and \( M^\pm(\epsilon) \) can be computed by solving a simple one-dimensional optimization problem. Figure 2 depicts the behavior of \( M^\pm \) and \( \tau^\pm \) as a function of \( \epsilon \). Of particular interest is the sparse limit \( \epsilon \). In this regime, [DJ94] yields the pleasingly simple expressions

\[ M^\pm(\epsilon) \sim 2\epsilon \log(\epsilon^{-1}) , \quad \tau^\pm(\epsilon) \sim \sqrt{2\log(\epsilon^{-1})} , \quad \text{as} \quad \epsilon \to 0. \]

A peculiar aspect of the results in [DJ94] requires us to generalize their results somewhat: the worst case distribution (2.7) puts positive mass at \( \pm\infty \). We will instead need approximations which place no mass at \( \infty \). This constraint can be formalized as follows.

**Definition 2.1.** We say distribution \( \nu_{\epsilon, \alpha} \) is \( \alpha \)-least-favorable for \( \eta(\cdot; \tau) \) if it is the least-dispersed distribution in \( \mathcal{F}_\epsilon \) achieving a fraction \( (1 - \alpha) \) of the worst case risk for \( \eta(\cdot; \tau) \), i.e. if

(i) We have

\[ \text{mse}(\nu_{\epsilon, \alpha}, \tau^\pm(\epsilon)) = (1 - \alpha) \cdot \sup_{\nu \in \mathcal{F}_\epsilon} \text{mse}(\nu, \tau^\pm(\epsilon)) . \]

(ii) The distribution \( \nu \) has the smallest second moment among all those for which (i) holds.

Figure 3 illustrates this definition.
Finding NLF $\mu ; \varepsilon = 1/10, \tau^*=1.1403, \alpha = .02$

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

Adapting the arguments of [DJ94] it is immediate to show that the least favorable distribution $\nu_{\varepsilon, \alpha}$ has the form of a three-point mixture

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

Here $\mu^{\pm}(\varepsilon, \alpha)$ is a function that can be obtained by solving one-dimensional optimization problem given by conditions (i) and (ii) above. Further, asymptotic analysis [DJ94] shows that, for any $\alpha > 0$ fixed, we have

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

Notice the relatively weak role played by $\alpha$. This shows that although the least-favorable distribution places mass at infinity, an approximately least-favorable distribution is already achieved with maximum amplitude $\mu^{\pm}(\varepsilon, \alpha)$.

3 Main Results

The notation of the last section allows us to state our main results. We first pause for introducing some terminology.

3.1 Terminology

An instance of the compressed sensing reconstruction problem is a triple $I_{n,N} = (x_0^{(N)}, z^{(n)}, A^{(n,N)})$. Further, we say that the sensing matrix $A$ is distributed according to the ensemble $\text{GAUSS}(n, N)$ (for
short $A \sim \text{Gauss}(n, N)$ if $A \in \mathbb{R}^{n \times N}$ and its entries $A_{ij}$ are iid $\mathcal{N}(0, 1/n)$.

**Definition 3.1. Convergent sequence of problem instances.** The sequence of problem instances $S = \{(x_0^{(N)}, z^{(N)}, A^{(n,N)})\}_{n,N}$ is said to be a convergent sequence if $n/N \to \delta \in (0, \infty)$, and in addition the following conditions hold:

(a) **Convergence of object marginals.** The empirical distribution of the entries of $x_0^{(N)}$ converges weakly to a probability measure $\nu$ on $\mathbb{R}$ with bounded second moment. Further $N^{-1}\|x_0^{(N)}\|_2^2 \to \mathbb{E}_\nu X^2$.

(b) **Convergence of noise marginals.** The empirical distribution of the entries of $z^{(n)}$ converges weakly to a probability measure $\omega$ on $\mathbb{R}$ with bounded second moment. Further $n^{-1}\|z_i^{(n)}\|_2^2 \to \mathbb{E}_\omega Z^2 \equiv \sigma^2$.

(c) **Normalization of Matrix Columns.** If $\{e_i\}_{1 \leq i \leq N}$, $e_i \in \mathbb{R}^N$ denotes the standard basis, then $\max_{i \in [N]} \|A^{(n,N)} e_i\|_2$, $\min_{i \in [N]} \|A^{(n,N)} e_i\|_2 \to 1$, as $N \to \infty$ where $[N] \equiv \{1, 2, \ldots, N\}$.

We shall say that $S$ is a convergent sequence of problem instances, and will write $S \in CS(\delta, \nu, \omega, \sigma)$ to label the implied limits.

Finally, we say that $S$ is a Gaussian sequence if $A^{(n,N)} \sim \text{Gauss}(n, N)$.

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 3.2. (Observable).** An observable $J = (J_{n,N})$ is a sequence of functions:

$$J_{n,N} : (x_0^{(N)}, z^{(n)}, A^{(n,N)}) \to J_{n,N}(x_0^{(N)}, z^{(n)}, A^{(n,N)}). \quad (3.1)$$

With an abuse of notation we will also denote by $(J_{n,N})$ the sequence of values taken by the observable $J$ on a specific sequence $S$. Namely $J_{n,N} = J_{n,N}(x_0^{(N)}, z^{(n)}, A^{(n,N)})$.

In the following we shall drop subscripts in the above quantities whenever un-necessary, or clear from the context. An example is the observed per-coordinate mean square error of a specific reconstruction algorithm alg:

$$\text{MSE}_{\text{alg}} \equiv \frac{1}{N} \|\hat{x}_{\text{alg}} - x_0\|_2^2.$$ 

The MSE depends explicitly on $x_0$ and implicitly on $y$ and $A$ (through the estimator $\hat{x}_{\text{alg}}$). Unless specified otherwise, we shall assume that the reconstruction algorithm solves the LASSO problem (1.2), and hence $\hat{x}_{\text{alg}} = \hat{x}^{1,\lambda}$.

**Definition 3.3. (Formalism).** A formalism is a procedure that assigns a purported large-system limit $\text{Formal}(J)$ to an observable $J$ along the sequence $S \in CS(\delta, \nu, \omega, \sigma)$. This limit in general depends on $\delta$, $\rho$, $\sigma$, and $\nu$: $\text{Formal}(J) = \text{Formal}(J; \delta, \rho, \sigma, \nu)$. 

10
Thus, in sections below we will consider \( J = \text{MSE} \) 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 \( \sigma \), \( \delta \), and probability measure \( \nu \) on \( \mathbb{R} \), it calculates a purported limit \( \text{fMSE}(\delta, \nu, \sigma) \). For a problem instance with large \( n, N \) realized from the standard framework LSF(\( \delta, \rho, \sigma, \nu \)), we claim the MSE will be approximately \( \text{fMSE}(\delta, \nu, \sigma) \). 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 3.4. (Validation). A formalism is theoretically validated by proving that, in the standard asymptotic framework, we have

\[
\lim_{N,n \to \infty} J_{n,N}(x_0^{(N)}, z^{(n)}, A^{(n,N)}) = \text{Formal}(J)
\]

for a class \( J \) of observables to which the formalism applies, and for a range of converging sequences \( S \in CS(\delta, \nu, \omega, \sigma) \).

A formalism is empirically validated by showing that, for problem instances \( I_{n,N} = (x_0^{(N)}, z^{(n)}, A^{(n,N)}) \in S, S \in CS(\delta, \nu, \omega, \sigma) \), with large \( N \), we have

\[
J_{n,N} \approx \text{Formal}(J; \delta, \rho, \sigma, \nu),
\]

for a collection of observables \( J \in J \) and a range of asymptotic framework parameters \( (\delta, \rho, \sigma, \nu) \). Here the approximation error hidden by \( \approx \) should be demonstrated to be empirically consistent with a systematic effect vanishing in the large system limit.

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 Section 4 for a discussion of 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.,

\[
\frac{1}{N} \| \hat{x}^{1,\lambda} - x_0 \|_2^2 \approx \text{fMSE}(\delta, \rho, \nu, \sigma).
\]

3.2 Results of the Formalism

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

Definition 3.5 (Phase Boundary). For each \( \delta \in [0, 1] \), let \( \rho_{\text{MSE}}(\delta) \) be the unique value of \( \rho \) solving

\[
M^\pm(\rho \delta) = \delta.
\]

It is known [DJ94] that \( M^\pm(\varepsilon) \) is monotone increasing and concave in \( \varepsilon \), with \( M^\pm(0) = 0 \) and \( M^\pm(1) = 1 \), see Figure 2. As a consequence, \( \rho_{\text{MSE}} \) is also a monotone increasing function of \( \delta \), with \( \rho_{\text{MSE}}(\delta) \to 0 \) as \( \delta \to 0 \) and \( \rho_{\text{MSE}}(\delta) \to 1 \) as \( \delta \to 1 \). A parametric expression for the curve \( (\delta, \rho_{\text{MSE}}(\delta)) \) is provided in Appendix A.

Proposition 3.1. Results of Formalism. The formalism developed below yields the following conclusions.
1.a In the region \( \rho < \rho_{\text{MSE}}(\delta) \), the minimax formal noise sensitivity defined in (1.7) obeys the formula

\[
M^*(\delta, \rho) \equiv \frac{M^\pm(\rho\delta)}{1 - M^\pm(\rho\delta)/\delta}. \tag{3.3}
\]

In particular, \( M^*(\delta, \rho) \) is finite throughout this region.

1.b With \( \sigma^2 \) the noise level in (1.1), define the formal noise-plus interference level \( f_{\text{NPI}} = f_{\text{NPI}}(\tau; \delta, \rho, \sigma, \nu) \)

\[
f_{\text{NPI}} = \sigma^2 + f_{\text{MSE}}/\delta,
\]

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

\[
\mu^*(\delta, \rho; \alpha) \equiv \mu^\pm(\delta\rho, \alpha) \cdot \sqrt{\text{NPI}^*(\delta, \rho; \sigma)}
\]

Construct a sequence of instances \( \mathbf{S} \in \text{CS}(\delta, \nu, \omega, \sigma) \), in such a way that \( \nu \in \mathcal{F}_{\delta\rho} \) place fraction 1 - \( \delta \rho \) of its mass at zero and the remaining mass equally on \( \pm \mu^*(\delta, \rho; \alpha) \). Such a sequence 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 formal maximin penalty parameter obeys

\[
\lambda^*(\nu; \delta, \rho, \sigma) \equiv \tau^\pm(\delta\rho) \cdot \sqrt{f_{\text{NPI}}} \cdot (1 - \text{EqDR}/\delta),
\]

where \( f_{\text{NPI}} = f_{\text{NPI}}(\tau^\pm; \delta, \rho, \sigma, \nu) \) and \( \text{EqDR} \) is given by

\[
\text{EqDR} \equiv \mathbb{P}\{|X + \sqrt{f_{\text{NPI}}}Z| > \tau^\pm(\delta\rho) \cdot \sqrt{f_{\text{NPI}}}|X \sim \nu \text{ and } Z \sim \mathcal{N}(0,1) \text{ independent} \}
\]

for \( X \sim \nu \) and \( Z \sim \mathcal{N}(0,1) \) independent (\( \text{EqDR} \) has the interpretation of asymptotic detection rate, i.e., the asymptotic fraction of coordinates that are estimated to be nonzero, see below).

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

2 In the region \( \rho > \rho_{\text{MSE}}(\delta) \), 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 \( \rho = \rho_{\text{MSE}}(\delta) \), it is infinite. The different contour lines show positions in the \((\delta, \rho)\) 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 \( \mu \) large enough. Figure 5 displays the value of the optimal penalization parameter amplitude \( \lambda^* = \lambda^*(\nu_{\delta\rho}^\alpha; \delta, \rho, \sigma = 1) \). This is the value of the regularization parameter that achieves the smallest worst case mean square error over the class of signals \( x_0 \) with at most \( N\rho\delta \) non-zero entries, and such that \( \|x_0^2\|^2 \leq Na \) with \( a = \rho\delta(\mu^*)^2 \). 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.
Contours of $\mu^*(\delta, \rho, 0.02)$ in the $(\delta, \rho)$ plane. The dotted line corresponds to the phase transition $(\delta, \rho_{\text{MSE}}(\delta))$, while the colored solid lines portray level sets of $\mu^*(\delta, \rho, \alpha)$. The 3-point mixture distribution $(1 - \varepsilon)\delta_0 + \varepsilon_1 \mu + \varepsilon_2 \mu, (\varepsilon = \delta \rho)$ will cause 98% of the worst-case MSE. Equivalently, 98% of the worst-case MSE is produced by vectors $x_0 \in \mathbb{R}^N$ with $N\varepsilon/2$ entries equal to $+\mu$, $N\varepsilon/2$ entries equal to $-\mu$, and $N(1 - \varepsilon)$ entries equal to 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 $\rho_{\text{MSE}}$ is identical to the phase boundary $\rho_{\ell_1}$ below which $\ell_1$ minimization and $\ell_0$ minimization are weakly equivalent.

The notion of weak equivalence is defined following [Don06, DT05]: Given a vector $x_0$ with at most $N\rho \delta$ non-zero entries, choose $A$ at random (in our case $A \sim \text{Gauss}(n, N)$). Weak equivalence requires that, with high probability with respect to such random choice, the solution of $(P_1)$ recovers the unknown signal, $\hat{x}_{1,0} = x_0$. The random polytope analysis of [Don06, DT05] implies that, under the Gaussian model, weak equivalence holds for all $(\delta, \rho)$ such that $\rho < \rho_{\ell_1}(\delta)$.

Finding 3.1 implies that the same phase boundary $\rho_{\ell_1}(\delta) = \rho_{\text{MSE}}(\delta)$ also describes the behavior in presence of noise. In words, throughout the phase where $\ell_1$ minimization is equivalent to $\ell_0$ minimization for the noiseless reconstruction problem, the solution to (1.2) has bounded formal MSE for noisy reconstruction. 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 $\rho_{\ell_1}$.

3.5 Validating the Predictions

Proposition 3.1 makes predictions for the behavior of solutions to (1.2). Since it is based on the formal MSE expression, that describes the $N, n \to \infty$ behavior of converging sequences, we will validate it empirically. The interest and importance of such a validation was discussed in the introduction.
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 \(\nu\) 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 \(\nu\) 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 \(\lambda\) 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

In this section, we present our formalism, and the formal mean square error expression \(\text{fMSE}(\lambda; \nu, \delta, \sigma)\) which is at the basis of Proposition 3.1. This will require a detour describing the AMP algorithm, and its analysis.
4.1 The AMP Algorithm

We now consider a reconstruction approach seemingly very different from the convex optimization method \((P_{2\lambda,1})\). 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:

\[
    z^t = y - A\hat{x}^t + b_t z^{t-1}, \quad b_t \equiv \frac{\|\hat{x}^t\|_0}{n}
\]

\[
    \hat{x}^{t+1} = \eta(A^*z^t + \hat{x}^t; \theta_t)
\]

Here \(\hat{x}^t \in \mathbb{R}^p\) is the current estimate of \(x_0\), and \(\|\hat{x}^t\|_0\) is the number of nonzeros in the current estimate. Again \(\eta(\cdot; \cdot)\) is the soft threshold nonlinearity with threshold parameter \(\theta_t\). We choose the latter following the rule \(\theta_t = \tau \cdot \sigma_t\),

\[
    \theta_t = \tau \cdot \sigma_t,
\]

where \(\tau\) is a tuning constant, fixed throughout iterations and \(\sigma_t\) is an empirical measure of the scale of the residuals. In practice, we can estimate asymptotically \(\sigma_t\) by \(\|z^t\|^2_2/n\). Finally \(z^t \in \mathbb{R}^n\) is the current working residual. This can be compared with the usual residual given by \(r^t = y - A\hat{x}^t\) via the identity \(z^t = r^t + b_t z^{t-1}\) with the scalar coefficient \(b_t = (\|\hat{x}^t\|_0/n)\).

The extra term \(b_t z^{t-1}\) in AMP plays a subtle but crucial role. Note that, similar to the step size, the coefficient \(b_t\) has to follow a specific prescription and cannot be scaled arbitrarily. This fails the standard convergence proofs [BT09]. However, it facilitates an exact characterization of the high-dimensional limit \(N,n \to \infty\). The AMP algorithm was introduced by the authors in [DMM09a]. We refer to this paper as well as to the following work [BM11] for further justification of the specific update, and of the term \(b_t z^{t-1}\).

In the present paper we introduce one important difference with respect to [DMM09a], namely we let the parameter \(\tau\) in (4.3) be set freely. In contrast, [DMM09a] proposed a fixed choice \(\tau(\delta)\) for each specific \(\delta\). As explained in [DMM09b] this choice is minimax optimal in the sense of maximizing the number of non-zero elements, \(\rho \delta N\), for which the reconstruction is successful in the noiseless case. The current algorithm is instead tunable and we therefore call it AMPT(\(\tau\)), where \(T\) stands for tunable.

4.2 Formal MSE, and its evolution

It is convenient to introduce the noise plus interference function \(\text{npi}(m;\sigma,\delta) \equiv \sigma^2 + m/\delta\). We then define the MSE map \(\Psi\) through

\[
    \Psi(m, \delta, \sigma, \tau, \nu) \equiv \text{mse}(\text{npi}(m, \sigma, \delta); \nu, \tau),
\]

where the function \(\text{mse}(\cdot; \nu, \tau)\) is the soft thresholding mean square error introduced in (2.3). It describes the MSE of soft thresholding in a problem where the noise level is \(\sqrt{\text{npi}}\). A heuristic explanation of the meaning and origin of \(\text{npi}\) 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

\[
    (m_t; \delta, \sigma, \tau, \nu) \mapsto (\Psi(m_t); \delta, \sigma, \tau, \nu),
\]

\[
    t \mapsto t + 1.
\]
As the parameters \((\delta, \sigma, \tau, \nu)\) remain fixed during evolution, we usually omit mention of them and think of state evolution simply as the iterated application of \(\Psi\):

\[
m_t \mapsto m_{t+1} \equiv \Psi(m_t),
\]

\[
t \mapsto t + 1.
\]

As mentioned in [DMM09a] and [BM11], in the asymptotic settings in which \(n, N \to \infty\) with \(n/N \to \delta\) and \(k/n \to \rho\), \(m_t\) characterizes the MSE of the AMPT estimates at every iteration. The fact that state evolution accurately describes the large system limit of the AMP algorithm (4.1), (4.2) is a non-trivial mathematical phenomenon. Some intuition can be gained by considering a modified evolution whereby (4.1), (4.2) are replaced by

\[
z_t = y(t) - A(t)\hat{x}_t, \quad y(t) = A(t)x_0 + w
\]

and

\[
\hat{x}_{t+1} = \eta(A(t)^*z_t + \hat{x}_t; \theta_t), \quad \text{with } A(1), A(2), \ldots, A(t) \sim A \text{ a sequence of random matrices.}
\]

In other words, the measurement matrix is sampled independently at each iteration, and the term \(b_t z_t\) is dropped. While this recursion does not correspond to any actual algorithm, it is fairly easy to analyze, showing that—for instance— \(z_t\) is asymptotically normal with iid entries of variance \(m_t\). Of course the use of the same matrix \(A\) introduces complex dependencies across iterations, but the term \(b_t z_t\) surprisingly ‘cancels’ such dependencies asymptotically [BM11].

**Definition 4.2. Stable Fixed Point.** The Highest Fixed Point of the continuous function \(\Psi\) is

\[
\text{HFP}(\Psi) = \sup\{m : \Psi(m) \geq m\}.
\]

The stability coefficient of the continuously differentiable function \(\Psi\) is

\[
\text{SC}(\Psi) = \left. \frac{d}{dm} \Psi(m) \right|_{m=\text{HFP}(\Psi)}.
\]

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

To illustrate this, Figure 6 shows the MSE map and fixed points in three cases. The rightmost point of these curves corresponds to \(m_0 = \mu_2(\nu)\), where \(\mu_2(\nu) = \int x^2 d\nu\) is the second-moment of the distribution \(\nu\). We will see that the LASSO mean square error is described by the fixed point \(\text{HFP}(\Psi)\). Notice that the curves in Figure 6 intersect to the left of the point \(m_0 = \mu_2(\nu)\). This implies that, in these examples, the LASSO reduces the mean square error with respect to the trivial estimator \(\hat{x} = 0\). This happens because \((\delta, \rho)\) are chosen within the stability region.

**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 \(\text{HFP}(\Psi)\):

\[
m_t \to \text{HFP}(\Psi), \quad t \to \infty.
\]

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

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

\[
|m_t - \text{HFP}(\Psi)| \leq A(\nu, \Psi) \text{SC}(\Psi)^t.
\]

Finally, if \(\mu_2(\nu) \geq \text{HFP}(\Psi)\) then the sequence \(\{m_t\}\) is monotonically decreasing to \(\mu_2(\nu)\) with

\[
(m_t - \text{HFP}(\Psi)) \leq \text{SC}(\Psi)^t \cdot (\mu_2(\nu) - \text{HFP}(\Psi)).
\]
Figure 6: MSE map $\Psi : m \mapsto \Psi(m)$ in three cases, and associated fixed points. In each case we plot $m$ (MSE input) on the horizontal axis, and $\Phi(m)$ (MSE output) on the vertical axis as red curves. Blue lines give $m$ versus $m$. Left: $\delta = 0.25$, $\rho = \rho_{\text{MSE}}/2$, $\sigma = 1$, $\nu = \nu^\ast(\delta, \rho, \alpha)$ Center: $\delta = 0.25$, $\rho = \rho_{\text{MSE}} \times 0.95$, $\sigma = 1$, $\nu = \nu^\ast(\delta, \rho, \alpha)$ Right: $\delta = 0.25$, $\rho = \rho_{\text{MSE}}$, $\sigma = 1$, $\nu = \nu^\ast(\delta, \rho, \alpha)$ (this case is marginally stable in the worst case, but since we use $\nu^\ast(\delta, \rho, \alpha)$ with bounded second moment, we still see a stable solution).
In short, barring for technical reasons the trivial case \( x_0 = 0, z = 0 \) (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 \( m \rightarrow \Psi(m; \delta, \sigma = 0, \nu, \tau) \) is concave as a function of \( m \), for any \( \delta, \nu, \tau \). 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 \( \sigma \) MSE map. From this it follows that \( \Psi \) is concave for \( \sigma > 0 \) as well. Also, [DMM09b] shows that \( \Psi(m = 0; \delta, \sigma = 0, \nu, \tau) = 0 \) and \( \frac{d \Psi}{d m}(m = 0; \delta, \sigma = 0, \nu, \tau) > 0 \), whence \( \Psi(m = 0; \delta, \sigma, \nu, \tau) > 0 \) for any positive noise level \( \sigma \).

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

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

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

\[
SC^*(\delta, \rho, \sigma = 0) = M^\pm(\delta \rho) / \delta.
\]

Hence, it follows that \( SC^*(\delta, \rho, \sigma = 0) < 1 \) throughout the region \( \rho < \rho_{\text{MSE}}(\delta) \), with \( \rho_{\text{MSE}}(\delta) \) given as per Definition 3.5.

Here we are interested in the noisy case and hence define the worst-case stability coefficient

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

Concavity of the MSE map \( m \mapsto \Psi(m) \) for \( \sigma = 0 \) implies

\[
SC^*(\delta, \rho) = SC^*(\delta, \rho, \sigma = 0)).
\]

For this reason, we will refer to the region \( \rho < \rho_{\text{MSE}}(\delta) \) as to the stability phase: not only the stability coefficient is smaller than 1, \( SC(\Psi) < 1 \), but it is bounded away from 1 uniformly in the signal distribution \( \nu \). We thus have proved the following:

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

Outside the stability region, for any arbitrarily large value \( m \in \mathbb{R} \), we can find measures \( \nu \) obeying the sparsity constraint \( \nu \in \mathcal{F}_\delta \) for which \( HFP(\Psi) > \mu_2(\nu) > m \) (see the construction in Section 4.5). This corresponds to a MSE larger than \( m \).

Figure 7 shows the MSE map and the state evolution in three cases. As for the analogous plots in Figure 6, the rightmost point (and vertical lines) correspond to \( m = \mu_2(\nu) \). In the first case, \( \rho \) is well below \( \rho_{\text{MSE}} \) and the fixed point is well below \( \mu_2(\nu) \). In the second case, \( \rho \) is slightly below \( \rho_{\text{MSE}} \) and the fixed point is close to \( \mu_2(\nu) \). In the third case, \( \rho \) is above \( \rho_{\text{MSE}} \) and the fixed point lies above \( \mu_2(\nu) \).
Figure 7: Crossing the phase transition: effects on MSE map $m \mapsto \Psi(m)$, and associated state evolution. In each case we plot $m$ (MSE input) on the horizontal axis, and $\Phi(m)$ (MSE output) on the vertical axis as red curves. Blue lines give $m$ versus $m$. Left: $\delta = 0.25, \rho = \rho_{\text{MSE}}/2, \sigma = 1, \nu = \nu(\delta, \rho, 0.01)$ Middle: $\delta = 0.25, \rho = 0.9 \cdot \rho_{\text{MSE}}, \sigma = 1, \nu = \nu(\delta, \rho, 0.01)$ Right: $\delta = 0.25, \rho = 1.5 \cdot \rho_{\text{MSE}}, \sigma = 1, \nu = \nu(\delta, \rho, 0.01)$. In each case $\tau = \tau(\delta \rho)$. As explained in the text, the fact that the curves do not intersect (for $m \leq \mu_2(\nu)$ in the rightmost frame is a consequence of the instability for $\rho < \rho_{\text{MSE}}(\delta)$.

The quantity $\mu_2(\nu)$ can be interpreted as the MSE one suffers by ‘doing nothing’: setting threshold $\lambda = \infty$ and taking $\hat{x} = 0$. When $HFP(\Psi) > \mu_2(\nu)$, 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 $\hat{x}^{t=1}$ 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

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

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

Suppose we are given $(\delta, \sigma, \nu, \tau)$, and a fixed point $m^*, m^* = HFP(\Psi)$ 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 $E(\zeta|S^*)$.

**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
Table 1: Some observables and their names.

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbrev.</th>
<th>( \zeta = \zeta(u,v,w) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Square Error</td>
<td>MSE</td>
<td>( \zeta = (u - w)^2 )</td>
</tr>
<tr>
<td>False Alarm Rate</td>
<td>FAR</td>
<td>( \zeta = 1_{{w \neq 0 &amp; u = 0}/(1 - \rho \delta)} )</td>
</tr>
<tr>
<td>Detection Rate</td>
<td>DR</td>
<td>( \zeta = 1_{{w \neq 0}} )</td>
</tr>
<tr>
<td>Missed Detection Rate</td>
<td>MDR</td>
<td>( \zeta = 1_{{w = 0 &amp; u \neq 0}/(\rho \delta)} )</td>
</tr>
<tr>
<td>False Detection Rate</td>
<td>FDeR</td>
<td>( \zeta = 1_{{w \neq 0 &amp; u = 0}/(\rho \delta)} )</td>
</tr>
</tbody>
</table>

\( \zeta : \mathbb{R}^3 \to \mathbb{R} \) associate the observable

\[
J^\zeta(y, A, x_0, \hat{x}) = \frac{1}{N} \sum_{i=1}^{N} \zeta(x_{0,i}, (A^T z)_i + \hat{x}_i - x_{0,i}, \hat{x}_i).
\]

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

\[
\text{Formal}(J^\zeta) = \mathcal{E}(\zeta|S^*).
\]

Validity of the state evolution formalism for AMPT entails that, for a convergent sequence of problem instances \( S = \{I_{n,N}\} = \{(x_0^{(N)}, z^{(n)}, A^{(n,N)})\}_{n,N} \) in \( CS(\delta, \nu, \omega, \sigma) \) the large-system limit for observable \( J^\zeta_{n,N} \) is simply the expectation in the equilibrium state:

\[
\lim_{n,N \to \infty} J^\zeta(y^{(n)}, A^{(n,N)}, x_0^{(N)}, \hat{x}^{(N)}) = \mathcal{E}(\zeta|S^*). \tag{4.5}
\]

The class \( \mathcal{J} \) of observables representable by the form \( J^\zeta \) is quite rich, by choosing \( \zeta(u,v,w) \) appropriately. Table 1 gives examples of well-known observables and the \( \zeta \) 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

\[
\text{Formal}(\text{FDR}) = \frac{\text{Formal}(\text{FDeR})}{\text{Formal}(\text{DR})}.
\]

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 AMPT(\( \tau \)) by State Evolution**

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

- Equilibrium MSE
  
  \[
  \text{EqMSE} = m_{\infty} = \text{HFP}(\Psi(\cdot; \nu, \tau); \delta, \sigma).
  \]

- Equilibrium Noise Plus Interference Level
  
  \[
  \text{npi}_{\infty} = \frac{1}{\delta} m_{\infty} + \sigma^2
  \]
– Equilibrium Threshold (absolute units)

$$\theta_\infty = \tau \cdot \sqrt{n\pi_\infty}.$$  

– Equilibrium Mean Squared Residual. Let $Y_\infty = X + \sqrt{n\pi_\infty} Z$ for $X \sim \nu$ and $Z \sim \mathcal{N}(0,1)$ independent$^4$. Then

$$\text{EqMSR} = \mathbb{E}\{ [Y_\infty - \eta(Y_\infty; \theta_\infty)]^2 \}.$$  

– Equilibrium Mean Absolute Estimate

$$\text{EqMAE} = \mathbb{E}\{ |\eta(Y_\infty; \theta_\infty)| \}.$$  

– Equilibrium Detection Rate

$$\text{EqDR} = \mathbb{P}\{ |Y_\infty| > \theta_\infty \}.$$ \hspace{1cm} (4.6)  

– Equilibrium Penalized MSR

$$\text{EqPMSR} = \text{EqMSR}/2 + \theta_\infty \cdot (1 - \text{EqDR}/\delta) \cdot \text{EqMAE}.$$  

\textbf{Remark 4.1.} Building on the present paper \cite{BM11, BM10} proved that the limit (4.5) indeed holds for convergent Gaussian sequences, i.e., under the assumption $A^{(n,N)} \sim \text{GAUSS}(n, N)$, and for sufficiently regular functions $\zeta$.

Notice that this type of result does not immediately implies minimax risk estimates, because of uniformity issues. A framework to deal with such difficulties is proposed in \cite{DJMM11}.

\section*{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 \cite{DMM10b}:

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

By \textit{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 \cite{DMM10b}. The correct mapping can be guessed from the following remarks:

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

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

$^4$Notice that $Y_\infty$ does not capture the distribution of the actual measurement vector $y$, but is rather plays the role of an 'effective' measurement in our formulas. The measurements would have the distribution of $Y_\infty$ if $A$ was the identity matrix, and the noise has variance $n\pi_\infty$.  

\hspace{1cm} 21
• AMPT(τ): At a fixed point \( \hat{x}^\infty, z^\infty \), no working residual exceeds the equilibrium threshold \( \theta^\infty \): \( \| A^T z^\infty \| \leq \theta^\infty \). Further

\[
\begin{align*}
\hat{x}^\infty_i > 0 & \iff (A^T z^\infty)_i = \theta^\infty, \\
\hat{x}^\infty_i = 0 & \iff |(A^T z^\infty)_i| < \theta^\infty, \\
\hat{x}^\infty_i < 0 & \iff (A^T z^\infty)_i = -\theta^\infty.
\end{align*}
\]

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

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

\[ \lambda = \theta^\infty \cdot (1 - df/n). \quad (4.7) \]

In other words, if we have a fixed point of AMPT(\( \tau \)) we can choose \( \lambda \) in such a way that this is also an optimum of LASSO(\( \lambda \)). Viceversa, any optimum of LASSO(\( \lambda \)) can be realized as a fixed point of AMPT(\( \tau \)): notice in fact that the relation (4.7) is invertible whenever \( df < n \).

This simple rule gives a calibration relationship between \( \tau \) and \( \lambda \), i.e. a one-one correspondence between \( \tau \) and \( \lambda \) that renders the two apparently different reconstruction procedures equivalent, provided the iteration AMPT(\( \tau \)) converges rapidly to its fixed point. Our empirical results confirm that this is indeed what happens for typical large system frameworks CS(\( \delta, \nu, \omega, \sigma \)). As mentioned above, this is a non-trivial finding, since AMP is not a priori guaranteed to converge.

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

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

\[ \lambda = \theta^\infty(\tau) \cdot (1 - \text{EqDR}(\tau)/\delta). \]

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

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

4.4 Derivation of Proposition 3.1

Consider the following Minimax Problem for AMPT(\( \tau \)). With \( f\text{MSE}(\tau; \delta, \rho, \sigma, \nu) \) denoting the equilibrium formal MSE for AMPT(\( \tau \)) for the framework CS(\( \delta, \nu, \omega, \sigma \)), fix \( \sigma = 1 \) and define

\[ M^0(\delta, \rho) = \inf_{\tau} \sup_{\nu \in F_0} f\text{MSE}(\tau; \delta, \rho, \sigma = 1, \nu). \quad (4.8) \]

We will first show that this definition obeys the formula just like the one in Proposition 3.1, given for \( M^*(\delta, \rho) \). Later we show that \( M^0(\delta, \rho) = M^*(\delta, \rho) \).
Figure 8: Contour lines of $\tau^*(\delta,\rho)$ in the $(\delta,\rho)$ plane. The dotted line corresponds to the phase transition $(\delta,\rho_{\text{MSE}}(\delta))$, while thin lines are contours for $\tau^*(\delta,\rho)$.

**Proposition 4.1.** For $M^\flat$ defined by (4.8),

$$M^\flat(\delta,\rho) = \frac{M^\pm(\delta\rho)}{1 - M^\pm(\delta\rho)/\delta}$$  \hspace{1cm} (4.9)

The AMPT threshold rule

$$\tau^*(\delta,\rho) = \tau^\pm(\delta\rho), \quad 0 < \rho < \rho_{\text{MSE}}(\delta),$$  \hspace{1cm} (4.10)

minimizes the formal MSE:

$$\sup_{\nu \in \mathcal{F}_{\delta\rho}} \text{fMSE}(\tau^*;\delta,\rho,1,\nu) = \inf_{\tau} \sup_{\nu \in \mathcal{F}_{\delta\rho}} \text{fMSE}(\tau;\delta,\rho,1,\nu) = M^\flat(\delta,\rho).$$  \hspace{1cm} (4.11)

Figure 8 depicts the behavior of $\tau^*$ in the $(\delta,\rho)$ plane.

**Proof (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 $\Psi(m;\nu) = \Psi(m,\delta,\sigma = 1,\tau^*,\nu) = \text{mse}(\text{npi}(m,1,\delta);\nu,\tau^*)$, cf. Eq. (4.4). Then $m^*$ obeys, by definition of fixed point,

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

We can use the scale invariance $\text{mse}(\sigma^2;\nu,\tau^*) = \text{mse}(1;\tilde{\nu},\tau^*)$, 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

$$m^* = \text{mse}(\text{npi}(m^*,1,\delta);\nu,\tau^*) = \text{mse}(1;\tilde{\nu},\tau^*) \cdot \text{npi}(m^*,1,\delta) \leq M^\pm(\delta\rho) \cdot \text{npi}(m^*,1,\delta),$$
where we used the fact that \( \nu^*(\delta, \rho) = \tau^\pm(\delta \rho) \). Hence

\[
\frac{m^*}{\text{npi}(m^*; 1, \delta)} \leq M^\pm(\delta \rho).
\]

The function \( m \mapsto m/\text{npi}(m; 1, \delta) = m/(1 + m/\delta) \) is one-to-one strictly increasing from the interval \([0, \infty)\) to \([0, \delta)\). Thus, provided that \( 1 - M^\pm(\delta \rho)/\delta > 0 \), i.e. \( \rho < \rho_{\text{MSR}} \), we have

\[
m^* \leq \frac{M^\pm(\delta \rho)}{1 - M^\pm(\delta \rho)/\delta}.
\]

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

\[
\sup_{\nu \in \mathcal{F}_\delta} \text{fMSE}(\nu^*; \delta, \rho, 1, \nu) \leq \frac{M^\pm(\delta \rho)}{1 - M^\pm(\delta \rho)/\delta}.
\]

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

\[
m^* = \text{mse}(\text{npi}(m^*, 1, \delta); \nu^*, \nu^*) = \text{mse}(1; \nu^*, \nu^*) \cdot \text{npi}(m^*, 1, \delta), \tag{4.12}
\]

where \( \nu^* \) is again a rescaled probability measure, this time with \( \nu^* = x \cdot \sqrt{\text{npi}(m^*, 1, \delta)} \in B = \nu\{x \in B\} \). Now since \( m^* \leq M^\rho \), we have \( \text{npi}(m^*, 1, \delta) \leq \text{NPI}' \), and hence

\[
\frac{\xi}{\sqrt{\text{npi}(m^*, 1, \delta)}} = \mu^\pm(\varepsilon, \alpha) \cdot \sqrt{\frac{\text{NPI}'}{\text{npi}(m^*, 1, \delta)}} > \mu^\pm(\varepsilon, \alpha).
\]

Note that \( \text{mse}(m; (1 - \varepsilon)\delta_0 + (\varepsilon/2)\delta_{-\xi} + (\varepsilon/2)\delta_{\xi}, \tau) \) 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 \( \alpha \)-least favorable for the minimax problem (2.5). Consequently,

\[
\text{mse}(1; \nu^*, \nu^*) \geq \text{mse}(1; \nu_{\delta, \alpha}, \nu^*) = (1 - \alpha) \cdot M^\pm(\delta, \rho).
\]

Using the scale-invariance relation, Eq. (4.12), we conclude that

\[
\frac{m^*}{\text{npi}(m^*; 1, \delta)} \geq (1 - \alpha) \cdot M^\pm(\delta \rho).
\]

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

\[
\text{fMSE}(\nu^*; \delta, \rho, 1, \nu) \geq \frac{(1 - \alpha) \cdot M^\pm(\delta \rho)}{1 - (1 - \alpha) \cdot M^\pm(\delta \rho)/\delta}.
\]

As \( \alpha > 0 \) is arbitrary, we conclude

\[
\sup_{\nu \in \mathcal{F}_\delta} \text{fMSE}(\nu^*; \delta, \rho, 1, \nu) \geq \frac{M^\pm(\delta \rho)}{1 - M^\pm(\delta \rho)/\delta}.
\]
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 (1.7) can be rewritten as a fixed-scale \( \sigma = 1 \) property:

\[
M^*(\delta, \rho) = \sup_{\nu \in F_{\delta \rho}} \inf_{\lambda} \text{fMSE}(\nu, \lambda | \text{LASSO}),
\]

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

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

\[
M^*(\delta, \rho) = M^\flat(\delta, \rho).
\]

Also, for \( \lambda^* \) as defined in Proposition 3.1,

\[
M^*(\delta, \rho) = \sup_{\nu \in F_{\delta \rho}} \text{fMSE}(\nu, \lambda^*(\nu; \delta, \rho, \sigma) | \text{LASSO}).
\]

In words, \( \lambda^* \) is the maximin penalization and the maximin MSE of LASSO is precisely given by the formula (4.9).

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

\[
\text{fMSE}(\nu, \lambda(\tau) | \text{LASSO}) = \text{fMSE}(\nu, \tau | \text{AMPT}).
\]

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

\[
\sup_{\nu \in F_{\delta \rho}} \text{fMSE}(\nu, \lambda^*(\nu; \delta, \rho, \sigma) | \text{LASSO}) = \sup_{\nu \in F_{\delta \rho}} \text{fMSE}(\nu, \lambda(\tau^*) | \text{LASSO})
\]

\[
= \sup_{\nu \in F_{\delta \rho}} \text{fMSE}(\nu, \tau^* | \text{AMPT})
\]

\[
= \sup_{\nu \in F_{\delta \rho}} \inf_{\tau} \text{fMSE}(\nu, \tau | \text{AMPT})
\]

\[
(4.13)
\]

\[
= \sup_{\nu \in F_{\delta \rho}} \inf_{\tau} \text{fMSE}(\nu, \lambda(\tau) | \text{LASSO})
\]

\[
= \sup_{\nu \in F_{\delta \rho}} \inf_{\lambda} \text{fMSE}(\nu, \lambda | \text{LASSO}).
\]

Display (4.13) shows that all these equalities are equal to \( M^\flat(\delta, \rho) \). \( \square \)

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 \( \delta, \rho \) with \( \rho > \rho_{\text{MSE}}(\delta) \) 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 \( \text{mse}(\mu, \tau) \) denote the MSE of scalar soft thresholding for amplitude of the non-zeros equal to
Consider values of the AMPT threshold $\tau$ such that $\text{mse}(0, \tau) < \delta$; this will be possible for all $\tau$ sufficiently large. Pick a number $\gamma \in \mathbb{R}$ obeying

$$0 \leq \gamma \text{mse}(0, \tau)/\delta. \quad (4.14)$$

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

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

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

$$\text{fMSE}(\nu, \tau | \text{AMPT}) = \frac{\delta \gamma}{1 - \gamma},$$

$$\text{fNPI}(\nu, \tau | \text{AMPT}) = \frac{1}{1 - \gamma}.$$

Assumption (4.14) permits us to choose $\gamma$ 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 $\tau$’s which can be associated to $\lambda$ obey

$$0 < \text{EqDR}(\nu, \tau) \leq \delta,$$

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

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

$$\text{mse}(0, \tau) < \delta \quad \text{and} \quad \text{EqDR}(\nu, \tau) < \delta,$$

the parameter $\lambda \geq 0$ defined by the calibration relation

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

has the formal MSE

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

One can check that, for each $\lambda \geq 0$, for each phase space point above phase transition, the above construction allows to construct a measure $\mu$ with $\varepsilon = \delta \rho$ mass on nonzeros and with arbitrarily high formal MSE. This completes the derivation of part 2 of Proposition 3.1.
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 $f_{\text{MSE}}(\lambda; \delta, \rho, \sigma, \nu)$ denote the formal MSE we assign to $x_1^\lambda$ for problem instances from $CS(\delta, \nu, \omega, \sigma)$. Let $e_{\text{MSE}}(\lambda)$ be the empirical MSE of the LASSO estimator $x_1^\lambda$ in a problem instance drawn from $CS(\delta, \nu, \omega, \sigma)$ at a given problem size $n, N$. In claiming that the noise sensitivity of $x_1^\lambda$ is bounded above by $M^*(\delta, \rho)$, we are saying that in empirical trials, the ratio $e_{\text{MSE}}/\sigma^2$ will not be larger than $M^*(\delta, \rho)$ 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 1.b of Proposition 3.1. If the empirical ratio $e_{\text{MSE}}/\sigma^2$ 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 $\rho \in \{\frac{1}{2} \cdot \rho_{\text{MSE}}, \frac{3}{4} \cdot \rho_{\text{MSE}}, \frac{9}{10} \cdot \rho_{\text{MSE}}, \frac{19}{20} \cdot \rho_{\text{MSE}}\}$. The predictions of the SE formalism, $\mu^*(\delta, \rho, 0.02)$, $\lambda^*(\delta, \rho)$, and $f_{\text{MSE}}$, are detailed in Tables 3 and 4. These are computed using the formulae in Proposition 3.1, part 1.c.

Results at $N = 1500$

To test these predictions, we generate in each situation $R = 200$ random realizations of size $N = 1500$ from $CS(\delta, \nu, \omega, \sigma)$ with the parameters shown in Tables 3 and 4, and run the LARS/LASSO solver to find the solution $x_1^\lambda$. Table 2 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 $\delta$ the most serious mismatch is at the value of $\rho$ 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 $\delta$, 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.
Table 2: Results at $N = 1500$. MSE of LASSO($\lambda^*$) at nearly-least-favorable situations, together with standard errors (SE)

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\rho$</th>
<th>$\mu^*(\delta, \rho, .02)$</th>
<th>$\lambda^*(\delta, \rho)$</th>
<th>fMSE</th>
<th>eMSE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>1.258</td>
<td>0.136</td>
<td>0.126</td>
<td>0.0029</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>0.804</td>
<td>0.380</td>
<td>0.329</td>
<td>0.0106</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12.901</td>
<td>0.465</td>
<td>1.045</td>
<td>0.755</td>
<td>0.0328</td>
</tr>
<tr>
<td>0.100</td>
<td>0.180</td>
<td>18.278</td>
<td>0.338</td>
<td>2.063</td>
<td>1.263</td>
<td>0.0860</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.459</td>
<td>0.961</td>
<td>0.374</td>
<td>0.373</td>
<td>0.0046</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.683</td>
<td>0.592</td>
<td>1.028</td>
<td>1.002</td>
<td>0.0170</td>
</tr>
<tr>
<td>0.250</td>
<td>0.241</td>
<td>12.219</td>
<td>0.351</td>
<td>2.830</td>
<td>2.927</td>
<td>0.0733</td>
</tr>
<tr>
<td>0.250</td>
<td>0.254</td>
<td>17.314</td>
<td>0.244</td>
<td>5.576</td>
<td>5.169</td>
<td>0.1978</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.194</td>
<td>0.689</td>
<td>0.853</td>
<td>0.836</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.354</td>
<td>0.400</td>
<td>2.329</td>
<td>2.251</td>
<td>0.0254</td>
</tr>
<tr>
<td>0.500</td>
<td>0.347</td>
<td>11.746</td>
<td>0.231</td>
<td>6.365</td>
<td>6.403</td>
<td>0.1157</td>
</tr>
<tr>
<td>0.500</td>
<td>0.366</td>
<td>16.667</td>
<td>0.159</td>
<td>12.427</td>
<td>11.580</td>
<td>0.2999</td>
</tr>
</tbody>
</table>

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 3 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 $\rho$ and $\delta$ that were studied for $N = 1500$, and see that the mismatches in our MSE's have grown smaller with $N$.

Results at $N = 8000$

Small values of $\delta$ correspond to the largest discrepancy, in particular when $\rho$ is 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 4 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 $\alpha$ either 1/2 or 1. Figure 9 presents empirical and theoretical results taken from the cases $N = 1500, 4000, \text{and} 8000$ 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 $1/N = 0$ presents the formal large-system limit predicted by our calculations and the other cases
<table>
<thead>
<tr>
<th>(\delta)</th>
<th>(\rho)</th>
<th>(\mu^*(\delta, \rho, .02))</th>
<th>(\lambda^*(\delta, \rho))</th>
<th>(\text{fMSE})</th>
<th>(\text{eMSE})</th>
<th>(\text{SE})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>1.258</td>
<td>0.136</td>
<td>0.128</td>
<td>0.0016</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>0.804</td>
<td>0.380</td>
<td>0.348</td>
<td>0.0064</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12.901</td>
<td>0.465</td>
<td>1.045</td>
<td>0.950</td>
<td>0.0228</td>
</tr>
<tr>
<td>0.100</td>
<td>0.180</td>
<td>18.278</td>
<td>0.338</td>
<td>2.063</td>
<td>1.588</td>
<td>0.0619</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.459</td>
<td>0.961</td>
<td>0.374</td>
<td>.371</td>
<td>0.0028</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.683</td>
<td>0.592</td>
<td>1.028</td>
<td>1.023</td>
<td>0.0106</td>
</tr>
<tr>
<td>0.250</td>
<td>0.241</td>
<td>12.219</td>
<td>0.351</td>
<td>2.830</td>
<td>2.703</td>
<td>0.0448</td>
</tr>
<tr>
<td>0.250</td>
<td>0.254</td>
<td>17.314</td>
<td>0.244</td>
<td>5.576</td>
<td>5.619</td>
<td>0.0428</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.194</td>
<td>0.689</td>
<td>0.853</td>
<td>0.849</td>
<td>0.0047</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.354</td>
<td>0.400</td>
<td>2.329</td>
<td>2.296</td>
<td>0.016</td>
</tr>
<tr>
<td>0.500</td>
<td>0.347</td>
<td>11.746</td>
<td>0.231</td>
<td>6.365</td>
<td>6.237</td>
<td>0.0677</td>
</tr>
<tr>
<td>0.500</td>
<td>0.366</td>
<td>16.667</td>
<td>0.159</td>
<td>12.427</td>
<td>12.394</td>
<td>0.171</td>
</tr>
</tbody>
</table>

Table 3: Results at \(N = 4000\). Theoretical and empirical MSE’s of LASSO\((\lambda^*)\) at nearly-least-favorable situations, together with standard errors (SE).

<table>
<thead>
<tr>
<th>(\delta)</th>
<th>(\rho)</th>
<th>(\mu^*(\delta, \rho, .02))</th>
<th>(\lambda^*(\delta, \rho))</th>
<th>(\text{fMSE})</th>
<th>(\text{eMSE})</th>
<th>(\text{SE})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>1.258</td>
<td>0.136</td>
<td>0.131</td>
<td>0.0012</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>0.804</td>
<td>0.380</td>
<td>0.378</td>
<td>0.0046</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12.901</td>
<td>0.465</td>
<td>1.045</td>
<td>1.024</td>
<td>0.0186</td>
</tr>
<tr>
<td>0.100</td>
<td>0.180</td>
<td>18.278</td>
<td>0.338</td>
<td>2.063</td>
<td>1.883</td>
<td>0.0458</td>
</tr>
</tbody>
</table>

Table 4: Results at \(N = 8000\). Theoretical and empirical MSE’s of LASSO\((\lambda^*)\) at nearly-least-favorable situations with \(\delta = 0.10\), together with standard errors (SE) of the empirical MSE’s. 

29
Convergence of empirical MSE to theoretical prediction

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 $\delta$. 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 $N = \infty$.

$1/N > 0$ 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 $(\nu^*, \lambda^*)$ 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 $(\nu^*, \lambda^*)$: any variation of $\mu$ to lower values, will cause a reduction in loss, while a variation of $\lambda$ to other values will cause an increase in loss.

### 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 $\lambda^*$ will cause MSE to increase. We consider the same situations as earlier, but now vary $\lambda$ away from the minimax value, while holding the other aspects of the problem fixed. In the appendix, Tables 6 and 7 present 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 $\mu$ to smaller values will cause a reduction in formal MSE. This can be checked as follows. If instead of $\mu^*(\delta, \rho, 0.01)$ we used $\mu^*(\delta, \rho, \alpha)$ for $\alpha$ significantly larger than 0.01, we should see a significant reduction in MSE, by an amount matching the predicted amount.

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

$$\text{fMSE}(\nu_1, \tau) \leq \text{mse}(\tilde{\nu}_1, \tau) \cdot \text{fNPI}(\nu_0, \tau),$$

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

$$\text{fMSE}(\nu_{\delta, \rho, \alpha}, \tau^*) \leq (1 - \alpha) M^*(\hat{\delta}, \rho),$$

and we ought to be able to see this. Table 8 shows results of simulations at $N = 1500$. The theoretical MSE drops as we move away from the nearly least favorable $\mu$ in the direction of smaller $\mu$, and the empirical MSE responds similarly.
Figure 11: Scatterplots comparing Theoretical and Empirical MSE’s found in Tables 6 and 7. 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.

**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 $\nu$.

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

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

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

$$f\text{MSE}((1 - \gamma)\nu_0 + \gamma\nu_1, \tau) \leq (1 - \gamma) \cdot \text{mse}(\tilde{\nu}_0, \tau) \cdot \text{NPI}(\nu_0, \tau) + \gamma \cdot \text{mse}(\tilde{\nu}_1, \tau) \cdot \text{NPI}(\nu_0, \tau), \quad (5.1)$$

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, $\nu_0 = \nu_{b,\rho,0.02}$ and $\nu_1 = \nu_{b,\rho,0.50}$, 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
Figure 12: Convexity structures in formal MSE. Behavior of formal MSE of 5 point mixture combining nearly least-favorable $\mu$ 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. $\delta = 0.25$, $\rho = \rho_{\text{MSE}}(\delta)/2$.

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 which takes pace at $\rho_{\text{MSE}}(0.25) \approx 0.2675$ – and selected a range of $\tau$ and $\gamma$ values allowed by our formalism. For each pair $\gamma, \tau$, we generated $R = 200$ Monte Carlo realizations and obtained LASSO solutions with the given penalization parameter $\lambda$. The results are described in Table 5. The match between formal MSE and empirical MSE is acceptable.

A similar behavior can be found for any $\rho > \rho_{\text{MSE}}(\delta)$. However, very close to the phase transition the asymptotic predictions are approached more slowly as $n, N \to \infty$.

**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

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

(6.1)

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

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

(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 \( \ell_1 \)-penalized least-squares estimator \( x^{1,\lambda^+,\ell} \), the solution to

\[
(P^+_{2,\lambda,1}) \quad \text{minimize} \quad \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \\
\text{subjectto} \quad x \geq 0.
\]

(6.3)

(6.4)

We define the minimax, formal noise sensitivity:

\[
M^{+\ast}(\delta, \rho) \equiv \sup_{\sigma > 0} \max_{\nu \in \mathcal{F}_{\rho\delta}^+} \min_{\lambda} \frac{1}{\sigma^2} f\text{MSE}_+(\lambda; \nu, \delta, \sigma),
\]

(6.5)

here \( \nu \in \mathcal{F}_{\rho\delta}^+ \) is the empirical distribution of \( x_0 \), and and \( f\text{MSE}_+ = f\text{MSE}_+(\lambda; \nu, \delta, \sigma) \) is the formal mean square error in the present case. Let \( \rho^+_{\text{MSE}}(\delta) \) denote the solution of

\[
M^+(\rho\delta) = \delta.
\]

(6.6)

In complete analogy to (1.9) we have the formula:

\[
M^{+\ast}(\delta, \rho) = \begin{cases} 
  M^+(\delta\rho) & \text{if } \rho < \rho^+_{\text{MSE}}(\delta), \\
  \frac{M^+(\delta\rho)}{1 - M^+(\delta\rho)/\delta} & \text{if } \rho \geq \rho^+_{\text{MSE}}(\delta).
\end{cases}
\]  

(6.7)

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 \( M^{+\ast} \). Again in analogy to the sign-unconstrained case, the phase boundary \( \rho^+_{\text{MSE}} \) occurs at precisely the same location at the phase boundary for the weak \( \ell_1-\ell_0 \) equivalence with non-negative signals; 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 $\ell_1$-minimization has its $\ell_1$-$\ell_0$ equivalence phase transition at the boundary $\rho_{\text{MSE}}$ not only for Gaussian matrices but for a wide collection of ensembles, including partial Fourier, partial Hadamard, expander graphs, iid $\pm 1$. 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.

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. This is often referred to as a conditional random field (CRF). In the present case this is easily done, provided the a priori probability distribution of the signal \( x = (x_1, \ldots, x_N) \) takes a \( \nu = \nu_1 \times \nu_2 \cdots \times \nu_N \).

The posterior is then

\[
\mu(dx) = \frac{1}{Z} \prod_{a=1}^{n} \exp \left\{ -\frac{\beta}{2} (y_a - (Ax)_a)^2 \right\} \prod_{i=1}^{N} \nu_i(dx_i).
\]

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 problems. From the information theory perspective, the term \(+b_t z^{t-1}\) 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 \( \nu_i \) 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, this is directly related to the \( \ell_1 \) regularization term 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.

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 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’ method.

\(^{5}\text{When this terminology is used in statistical physics, the emphasis is rather on the study of distributional properties of random instances.}\)
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 $+b\varepsilon^{t-1}$ 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 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. State evolution predicts the performances of these algorithms (including for instance precise convergence time estimates);
4. Finally, the state evolution approach is more easily amenable to rigorization, as was demonstrated in [BM11, BM10].
A  Formula for $\rho_{MSE}(\delta)$

The phase boundary curve $(\delta, \rho_{MSE}(\delta))$ admits a very simple expression. This is a parametric expression in terms of the parameter $\tau \in (0, \infty)$:

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

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

As $\tau$ goes from 0 to $\infty$, the pair $(\delta, \rho)$ given by these equations traces the phase boundary moving from the corner $(1, 1)$ to $(0, 0)$.

In order to obtain this characterization, recall that the curve $(\delta, \rho_{MSE}(\delta))$ is given implicitly by

$$\delta = M^{\pm}(\rho\delta), \quad (A.3)$$

where $M^{\pm}(\cdot)$ is the soft-thresholding minimax MSE. As discussed in section (2), $M^{\pm}(\varepsilon) = \min_{\tau} M^{\pm}(\varepsilon, \tau)$ with the latter function given explicitly by (2.8). By simple calculus, the minimum is achieved at the unique stationary point of the function $\tau \mapsto M^{\pm}(\varepsilon, \tau)$. Therefore (A.3) is equivalent to

$$\delta = M^{\pm}(\rho\delta, \tau),$$

$$0 = \frac{\partial}{\partial \tau} M^{\pm}(\rho\delta, \tau).$$

Substituting the explicit expression (2.8) in these equations, we recover (A.1), (A.2).

B  Tables

This appendix contains table of empirical results supporting our claims.

References


<table>
<thead>
<tr>
<th>δ</th>
<th>ρ</th>
<th>γ</th>
<th>μ</th>
<th>τ</th>
<th>λ</th>
<th>fMSE</th>
<th>eMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.75</td>
<td>2.8740</td>
<td>1.500</td>
<td>0.9840</td>
<td>0.750</td>
<td>0.746</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.85</td>
<td>4.142</td>
<td>1.500</td>
<td>1.168</td>
<td>1.417</td>
<td>1.425</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.90</td>
<td>5.345</td>
<td>1.500</td>
<td>1.366</td>
<td>2.250</td>
<td>2.239</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.95</td>
<td>7.954</td>
<td>1.500</td>
<td>1.841</td>
<td>4.750</td>
<td>4.724</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.97</td>
<td>10.4781</td>
<td>1.500</td>
<td>2.328</td>
<td>8.083</td>
<td>8.126</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.98</td>
<td>12.962</td>
<td>1.500</td>
<td>2.822</td>
<td>12.250</td>
<td>12.327</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.99</td>
<td>18.5172</td>
<td>1.500</td>
<td>3.949</td>
<td>24.750</td>
<td>24.601</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.995</td>
<td>26.3191</td>
<td>1.500</td>
<td>5.5558</td>
<td>49.750</td>
<td>49.837</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.75</td>
<td>2.9031</td>
<td>2.000</td>
<td>2.8766</td>
<td>1.417</td>
<td>1.409</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.85</td>
<td>4.058</td>
<td>2.000</td>
<td>3.626</td>
<td>2.250</td>
<td>2.238</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.90</td>
<td>5.158</td>
<td>2.000</td>
<td>4.385</td>
<td>2.250</td>
<td>2.238</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.95</td>
<td>7.560</td>
<td>2.000</td>
<td>6.122</td>
<td>4.750</td>
<td>4.742</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.97</td>
<td>9.897</td>
<td>2.000</td>
<td>7.861</td>
<td>8.083</td>
<td>8.054</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.98</td>
<td>12.205</td>
<td>2.000</td>
<td>9.6019</td>
<td>12.250</td>
<td>12.215</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.99</td>
<td>17.380</td>
<td>2.000</td>
<td>13.5425</td>
<td>24.750</td>
<td>24.634</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.995</td>
<td>24.662</td>
<td>2.000</td>
<td>19.126</td>
<td>49.750</td>
<td>49.424</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.75</td>
<td>2.817</td>
<td>2.500</td>
<td>4.501</td>
<td>1.417</td>
<td>1.409</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.85</td>
<td>3.896</td>
<td>2.500</td>
<td>5.750</td>
<td>2.250</td>
<td>2.241</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.90</td>
<td>4.926</td>
<td>2.500</td>
<td>7.004</td>
<td>2.250</td>
<td>2.241</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.95</td>
<td>7.181</td>
<td>2.500</td>
<td>9.848</td>
<td>4.750</td>
<td>4.712</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.97</td>
<td>9.380</td>
<td>2.500</td>
<td>12.6846</td>
<td>8.083</td>
<td>8.050</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.98</td>
<td>11.555</td>
<td>2.500</td>
<td>15.5170</td>
<td>12.250</td>
<td>12.215</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.99</td>
<td>16.436</td>
<td>2.500</td>
<td>21.9183</td>
<td>24.750</td>
<td>24.619</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.995</td>
<td>23.311</td>
<td>2.500</td>
<td>30.9786</td>
<td>49.750</td>
<td>49.442</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.75</td>
<td>2.7649</td>
<td>3.000</td>
<td>5.8144</td>
<td>1.417</td>
<td>1.408</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.85</td>
<td>3.809</td>
<td>3.000</td>
<td>7.4430</td>
<td>2.250</td>
<td>2.241</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.90</td>
<td>4.806</td>
<td>3.000</td>
<td>9.131</td>
<td>2.250</td>
<td>2.241</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.95</td>
<td>6.991</td>
<td>3.000</td>
<td>12.880</td>
<td>4.750</td>
<td>4.735</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.97</td>
<td>9.125</td>
<td>3.000</td>
<td>16.6113</td>
<td>8.083</td>
<td>8.053</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.98</td>
<td>11.236</td>
<td>3.000</td>
<td>20.3339</td>
<td>12.250</td>
<td>12.218</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.99</td>
<td>15.975</td>
<td>3.000</td>
<td>28.7413</td>
<td>24.750</td>
<td>24.621</td>
</tr>
<tr>
<td>0.250</td>
<td>0.401</td>
<td>0.995</td>
<td>22.652</td>
<td>3.000</td>
<td>40.6356</td>
<td>49.750</td>
<td>49.419</td>
</tr>
</tbody>
</table>

Table 5: Results above Phase transition. Parameters of the construction as well as theoretical predictions and resulting empirical MSE figures.
Table 6: \( N = 1500 \), \( \lambda \) dependence of the MSE at fixed \( \mu \)

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( \rho )</th>
<th>( \mu )</th>
<th>( \lambda )</th>
<th>( fMSE )</th>
<th>( eMSE )</th>
<th>( SE )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>0.402</td>
<td>0.152</td>
<td>0.140</td>
<td>0.0029</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>1.258</td>
<td>0.136</td>
<td>0.126</td>
<td>0.0029</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>2.037</td>
<td>0.142</td>
<td>0.133</td>
<td>0.0030</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>3.169</td>
<td>0.174</td>
<td>0.164</td>
<td>0.0028</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>4.948</td>
<td>0.239</td>
<td>0.228</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>0.375</td>
<td>0.417</td>
<td>0.391</td>
<td>0.110</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>0.804</td>
<td>0.380</td>
<td>0.329</td>
<td>0.0106</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>1.960</td>
<td>0.408</td>
<td>0.374</td>
<td>0.0087</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>3.824</td>
<td>0.534</td>
<td>0.504</td>
<td>0.0084</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12.906</td>
<td>0.301</td>
<td>1.106</td>
<td>0.911</td>
<td>0.0410</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12.906</td>
<td>0.465</td>
<td>1.045</td>
<td>0.755</td>
<td>0.0328</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12.906</td>
<td>2.298</td>
<td>1.178</td>
<td>0.992</td>
<td>0.0326</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12.906</td>
<td>5.461</td>
<td>1.619</td>
<td>1.520</td>
<td>0.0273</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.459</td>
<td>0.518</td>
<td>0.403</td>
<td>0.390</td>
<td>0.0044</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.459</td>
<td>0.961</td>
<td>0.374</td>
<td>0.373</td>
<td>0.0046</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.459</td>
<td>1.419</td>
<td>0.385</td>
<td>0.386</td>
<td>0.0046</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.459</td>
<td>2.165</td>
<td>0.452</td>
<td>0.455</td>
<td>0.0053</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.459</td>
<td>3.555</td>
<td>0.623</td>
<td>0.612</td>
<td>0.0042</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.683</td>
<td>0.036</td>
<td>1.151</td>
<td>1.155</td>
<td>0.0174</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.683</td>
<td>0.592</td>
<td>1.028</td>
<td>1.002</td>
<td>0.0170</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.683</td>
<td>1.183</td>
<td>1.073</td>
<td>1.069</td>
<td>0.0169</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.683</td>
<td>2.243</td>
<td>1.234</td>
<td>1.293</td>
<td>0.0158</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.683</td>
<td>4.392</td>
<td>1.861</td>
<td>1.837</td>
<td>0.0114</td>
</tr>
<tr>
<td>0.250</td>
<td>0.241</td>
<td>12.219</td>
<td>0.310</td>
<td>3.111</td>
<td>2.951</td>
<td>0.0901</td>
</tr>
<tr>
<td>0.250</td>
<td>0.241</td>
<td>12.219</td>
<td>0.351</td>
<td>2.830</td>
<td>2.927</td>
<td>0.0733</td>
</tr>
<tr>
<td>0.250</td>
<td>0.241</td>
<td>12.219</td>
<td>1.219</td>
<td>3.065</td>
<td>2.998</td>
<td>0.0661</td>
</tr>
<tr>
<td>0.250</td>
<td>0.241</td>
<td>12.219</td>
<td>2.917</td>
<td>4.055</td>
<td>4.020</td>
<td>0.0485</td>
</tr>
<tr>
<td>0.250</td>
<td>0.241</td>
<td>12.219</td>
<td>4.444</td>
<td>5.709</td>
<td>5.625</td>
<td>0.0330</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.194</td>
<td>0.176</td>
<td>1.121</td>
<td>1.108</td>
<td>0.0080</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.194</td>
<td>0.470</td>
<td>0.894</td>
<td>0.879</td>
<td>0.0070</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.194</td>
<td>0.689</td>
<td>0.853</td>
<td>0.836</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.194</td>
<td>0.933</td>
<td>0.866</td>
<td>0.862</td>
<td>0.008</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.194</td>
<td>1.355</td>
<td>0.965</td>
<td>0.960</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.194</td>
<td>2.237</td>
<td>1.273</td>
<td>1.263</td>
<td>0.0075</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.354</td>
<td>0.179</td>
<td>2.489</td>
<td>2.438</td>
<td>0.0262</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.354</td>
<td>0.400</td>
<td>2.329</td>
<td>2.251</td>
<td>0.0254</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.354</td>
<td>0.655</td>
<td>2.377</td>
<td>2.329</td>
<td>0.0268</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.354</td>
<td>1.137</td>
<td>2.728</td>
<td>2.718</td>
<td>0.0256</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.354</td>
<td>2.258</td>
<td>3.704</td>
<td>3.672</td>
<td>0.0212</td>
</tr>
<tr>
<td>0.500</td>
<td>0.347</td>
<td>11.746</td>
<td>0.19</td>
<td>6.451</td>
<td>6.411</td>
<td>0.1210</td>
</tr>
<tr>
<td>0.500</td>
<td>0.347</td>
<td>11.746</td>
<td>0.231</td>
<td>6.365</td>
<td>6.403</td>
<td>0.1157</td>
</tr>
<tr>
<td>0.500</td>
<td>0.347</td>
<td>11.746</td>
<td>0.558</td>
<td>6.624</td>
<td>6.349</td>
<td>0.1121</td>
</tr>
<tr>
<td>0.500</td>
<td>0.347</td>
<td>11.746</td>
<td>1.227</td>
<td>8.089</td>
<td>7.813</td>
<td>0.0819</td>
</tr>
<tr>
<td>0.500</td>
<td>0.347</td>
<td>11.746</td>
<td>2.882</td>
<td>11.288</td>
<td>11.189</td>
<td>0.0692</td>
</tr>
</tbody>
</table>
Table 7: \( N = 4000 \), \( \lambda \) dependence of the MSE at fixed \( \mu \)

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( \rho )</th>
<th>( \mu )</th>
<th>( \lambda )</th>
<th>( fMSE )</th>
<th>( eMSE )</th>
<th>( SE )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>0.402</td>
<td>0.152</td>
<td>0.144</td>
<td>0.0017</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>1.258</td>
<td>0.136</td>
<td>0.128</td>
<td>0.0016</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>2.037</td>
<td>0.142</td>
<td>0.133</td>
<td>0.0016</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>3.169</td>
<td>0.174</td>
<td>0.168</td>
<td>0.0016</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>0.375</td>
<td>0.4173</td>
<td>0.398</td>
<td>0.059</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>0.804</td>
<td>0.380</td>
<td>0.348</td>
<td>0.0064</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>1.960</td>
<td>0.408</td>
<td>0.389</td>
<td>0.0058</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.242</td>
<td>3.824</td>
<td>0.534</td>
<td>0.510</td>
<td>0.0051</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12906</td>
<td>0.301</td>
<td>1.106</td>
<td>1.01</td>
<td>0.0231</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12906</td>
<td>0.465</td>
<td>1.045</td>
<td>0.950</td>
<td>0.0228</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12906</td>
<td>2.298</td>
<td>1.178</td>
<td>1.111</td>
<td>0.0232</td>
</tr>
<tr>
<td>0.100</td>
<td>0.170</td>
<td>12906</td>
<td>5.461</td>
<td>1.619</td>
<td>1.591</td>
<td>0.0159</td>
</tr>
<tr>
<td>0.100</td>
<td>0.18</td>
<td>18.278</td>
<td>0.115</td>
<td>2.207</td>
<td>1.911</td>
<td>0.0541</td>
</tr>
<tr>
<td>0.100</td>
<td>0.180</td>
<td>18.278</td>
<td>0.338</td>
<td>2.063</td>
<td>1.588</td>
<td>0.0619</td>
</tr>
<tr>
<td>0.100</td>
<td>0.180</td>
<td>18.278</td>
<td>2.934</td>
<td>2.467</td>
<td>2.171</td>
<td>0.0532</td>
</tr>
<tr>
<td>0.100</td>
<td>0.180</td>
<td>18.278</td>
<td>7.545</td>
<td>3.474</td>
<td>3.367</td>
<td>0.0312</td>
</tr>
<tr>
<td>0.150</td>
<td>0.109</td>
<td>5.631</td>
<td>0.420</td>
<td>0.236</td>
<td>0.228</td>
<td>0.0022</td>
</tr>
<tr>
<td>0.150</td>
<td>0.109</td>
<td>5.631</td>
<td>1.073</td>
<td>0.212</td>
<td>0.209</td>
<td>0.0023</td>
</tr>
<tr>
<td>0.150</td>
<td>0.109</td>
<td>5.631</td>
<td>1.700</td>
<td>0.218</td>
<td>0.213</td>
<td>0.0021</td>
</tr>
<tr>
<td>0.150</td>
<td>0.109</td>
<td>5.631</td>
<td>2.657</td>
<td>0.260</td>
<td>0.251</td>
<td>0.0024</td>
</tr>
<tr>
<td>0.150</td>
<td>0.163</td>
<td>8.030</td>
<td>0.720</td>
<td>0.601</td>
<td>0.594</td>
<td>0.0071</td>
</tr>
<tr>
<td>0.150</td>
<td>0.163</td>
<td>8.030</td>
<td>0.720</td>
<td>0.588</td>
<td>0.595</td>
<td>0.0072</td>
</tr>
<tr>
<td>0.150</td>
<td>0.163</td>
<td>8.030</td>
<td>1.614</td>
<td>0.626</td>
<td>0.610</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.150</td>
<td>0.163</td>
<td>8.030</td>
<td>3.135</td>
<td>0.804</td>
<td>0.807</td>
<td>0.0058</td>
</tr>
<tr>
<td>0.150</td>
<td>0.196</td>
<td>12.577</td>
<td>0.411</td>
<td>1.631</td>
<td>1.603</td>
<td>0.0301</td>
</tr>
<tr>
<td>0.150</td>
<td>0.196</td>
<td>12.577</td>
<td>0.434</td>
<td>1.612</td>
<td>1.572</td>
<td>0.0341</td>
</tr>
<tr>
<td>0.150</td>
<td>0.196</td>
<td>12.577</td>
<td>1.814</td>
<td>1.792</td>
<td>1.720</td>
<td>0.0281</td>
</tr>
<tr>
<td>0.150</td>
<td>0.196</td>
<td>12.577</td>
<td>4.339</td>
<td>2.433</td>
<td>2.383</td>
<td>0.0205</td>
</tr>
<tr>
<td>0.150</td>
<td>0.207</td>
<td>17.814</td>
<td>0.281</td>
<td>3.211</td>
<td>3.005</td>
<td>0.0730</td>
</tr>
<tr>
<td>0.150</td>
<td>0.207</td>
<td>17.814</td>
<td>0.305</td>
<td>3.185</td>
<td>2.864</td>
<td>0.0861</td>
</tr>
<tr>
<td>0.150</td>
<td>0.207</td>
<td>17.814</td>
<td>2.231</td>
<td>3.715</td>
<td>3.582</td>
<td>0.0722</td>
</tr>
<tr>
<td>0.150</td>
<td>0.207</td>
<td>17.814</td>
<td>5.879</td>
<td>5.202</td>
<td>5.141</td>
<td>0.0439</td>
</tr>
<tr>
<td>0.150</td>
<td>0.207</td>
<td>17.814</td>
<td>12.455</td>
<td>7.142</td>
<td>7.154</td>
<td>0.0269</td>
</tr>
</tbody>
</table>
Table 8: $N = 1500$, $\mu$ dependence of the MSE at fixed $\lambda$

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\rho$</th>
<th>$\mu$</th>
<th>$\lambda$</th>
<th>fMSE</th>
<th>eMSE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.291</td>
<td>1.253</td>
<td>0.131</td>
<td>0.125</td>
<td>0.0022</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.541</td>
<td>1.256</td>
<td>0.134</td>
<td>0.132</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.691</td>
<td>1.257</td>
<td>0.135</td>
<td>0.126</td>
<td>0.0027</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.791</td>
<td>1.258</td>
<td>0.136</td>
<td>0.129</td>
<td>0.0024</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>5.891</td>
<td>1.259</td>
<td>0.137</td>
<td>0.125</td>
<td>0.0027</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>6.041</td>
<td>1.260</td>
<td>0.138</td>
<td>0.126</td>
<td>0.0030</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>6.291</td>
<td>1.262</td>
<td>0.139</td>
<td>0.127</td>
<td>0.0028</td>
</tr>
<tr>
<td>0.100</td>
<td>0.095</td>
<td>6.791</td>
<td>1.264</td>
<td>0.141</td>
<td>0.125</td>
<td>0.0031</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>7.242</td>
<td>0.794</td>
<td>0.349</td>
<td>0.317</td>
<td>0.0074</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>7.742</td>
<td>0.800</td>
<td>0.366</td>
<td>0.335</td>
<td>0.0084</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>7.992</td>
<td>0.802</td>
<td>0.373</td>
<td>0.351</td>
<td>0.0089</td>
</tr>
<tr>
<td>0.100</td>
<td>0.142</td>
<td>8.000</td>
<td>0.802</td>
<td>0.373</td>
<td>0.362</td>
<td>0.0094</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>4.459</td>
<td>0.952</td>
<td>0.338</td>
<td>0.336</td>
<td>0.0036</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>4.959</td>
<td>0.957</td>
<td>0.359</td>
<td>0.346</td>
<td>0.0040</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.209</td>
<td>0.959</td>
<td>0.367</td>
<td>0.356</td>
<td>0.0044</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.359</td>
<td>0.960</td>
<td>0.371</td>
<td>0.373</td>
<td>0.0049</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.459</td>
<td>0.961</td>
<td>0.374</td>
<td>0.362</td>
<td>0.0047</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.559</td>
<td>0.962</td>
<td>0.376</td>
<td>0.367</td>
<td>0.0045</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.709</td>
<td>0.962</td>
<td>0.379</td>
<td>0.372</td>
<td>0.0048</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>5.959</td>
<td>0.963</td>
<td>0.383</td>
<td>0.362</td>
<td>0.0052</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>6.459</td>
<td>0.964</td>
<td>0.387</td>
<td>0.387</td>
<td>0.0058</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>6.683</td>
<td>0.587</td>
<td>0.939</td>
<td>0.899</td>
<td>0.0126</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.183</td>
<td>0.590</td>
<td>0.988</td>
<td>0.965</td>
<td>0.0147</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.433</td>
<td>0.591</td>
<td>1.009</td>
<td>0.956</td>
<td>0.0147</td>
</tr>
<tr>
<td>0.250</td>
<td>0.201</td>
<td>7.583</td>
<td>0.592</td>
<td>1.021</td>
<td>1.027</td>
<td>0.0155</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>4.194</td>
<td>0.684</td>
<td>0.769</td>
<td>0.770</td>
<td>0.0052</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>4.694</td>
<td>0.687</td>
<td>0.818</td>
<td>0.823</td>
<td>0.0066</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>4.944</td>
<td>0.688</td>
<td>0.837</td>
<td>0.838</td>
<td>0.0073</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.094</td>
<td>0.689</td>
<td>0.847</td>
<td>0.835</td>
<td>0.0068</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.194</td>
<td>0.689</td>
<td>0.853</td>
<td>0.834</td>
<td>0.0073</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.294</td>
<td>0.689</td>
<td>0.858</td>
<td>0.845</td>
<td>0.0079</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.444</td>
<td>0.690</td>
<td>0.865</td>
<td>0.863</td>
<td>0.0079</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>5.694</td>
<td>0.690</td>
<td>0.874</td>
<td>0.887</td>
<td>0.0085</td>
</tr>
<tr>
<td>0.500</td>
<td>0.193</td>
<td>6.194</td>
<td>0.691</td>
<td>0.886</td>
<td>0.868</td>
<td>0.0085</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>6.354</td>
<td>0.398</td>
<td>2.119</td>
<td>2.071</td>
<td>0.0195</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>6.854</td>
<td>0.399</td>
<td>2.234</td>
<td>2.214</td>
<td>0.0235</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.104</td>
<td>0.399</td>
<td>2.284</td>
<td>2.157</td>
<td>0.0252</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.254</td>
<td>0.400</td>
<td>2.313</td>
<td>2.271</td>
<td>0.0244</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.354</td>
<td>0.400</td>
<td>2.329</td>
<td>2.316</td>
<td>0.0275</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.454</td>
<td>0.400</td>
<td>2.346</td>
<td>2.287</td>
<td>0.0287</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.604</td>
<td>0.400</td>
<td>2.370</td>
<td>2.327</td>
<td>0.0306</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>7.854</td>
<td>0.401</td>
<td>2.404</td>
<td>2.339</td>
<td>0.0284</td>
</tr>
<tr>
<td>0.500</td>
<td>0.289</td>
<td>8.000</td>
<td>0.401</td>
<td>2.422</td>
<td>2.409</td>
<td>0.0300</td>
</tr>
</tbody>
</table>
Table 9: $N = 1500$, MSE for 5-point prior

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\rho$</th>
<th>$\mu$</th>
<th>$\lambda$</th>
<th>Theoretical MSE</th>
<th>Empirical MSE</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>1.894</td>
<td>0.857</td>
<td>0.120</td>
<td>0.151</td>
<td>0.0</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>2.171</td>
<td>0.897</td>
<td>0.162</td>
<td>0.163</td>
<td>0.122</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>2.447</td>
<td>0.901</td>
<td>0.178</td>
<td>0.177</td>
<td>0.244</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>2.724</td>
<td>0.906</td>
<td>0.196</td>
<td>0.195</td>
<td>0.366</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>3.001</td>
<td>0.912</td>
<td>0.215</td>
<td>0.210</td>
<td>0.488</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>3.277</td>
<td>0.918</td>
<td>0.237</td>
<td>0.236</td>
<td>0.611</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>3.554</td>
<td>0.926</td>
<td>0.261</td>
<td>0.257</td>
<td>0.733</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>3.830</td>
<td>0.935</td>
<td>0.287</td>
<td>0.280</td>
<td>0.855</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>4.107</td>
<td>0.945</td>
<td>0.317</td>
<td>0.307</td>
<td>0.977</td>
</tr>
<tr>
<td>0.250</td>
<td>0.134</td>
<td>4.383</td>
<td>0.957</td>
<td>0.348</td>
<td>0.359</td>
<td>1.100</td>
</tr>
</tbody>
</table>