IEEE TRANSACTIONS ON
INFORMATION THEORY

IEEE Xplore® Digital Library
http://ieeexplore.ieee.org/Xplore

The Replica-Symmetric Prediction for Random Linear Estimation with Gaussian Matrices is Exact

Galen Reeves, *Member, IEEE* and Henry D. Pfister, *Senior Member, IEEE*

***Abstract*—This paper considers the fundamental limit of random linear estimation for i.i.d. signal distributions and i.i.d. Gaussian measurement matrices. Its main contribution is a rigorous characterization of the asymptotic mutual information (MI) and minimum mean-square error (MMSE) in this setting. Under mild technical conditions, our results show that the limiting MI and MMSE are equal to the values predicted by the replica method from statistical physics. This resolves a well-known problem that has remained open for over a decade.**

***Index Terms*—Compressed Sensing, MMSE Estimation, Phase Transitions, Random Linear Mixing, Replica Method**

I. Introduction

The canonical random linear estimation (or compressed sensing) problem can be formulated as follows. The signal is a random n-dimensional vector $X^n = (X_1, \dots, X_n)$ whose entries are drawn independently from a common distribution P_X with finite variance. The signal is observed using noisy linear measurements of the form

$$Y_k = \langle A_k, X^n \rangle + W_k,$$

where $\{A_k\}$ is a sequence of n-dimensional measurement vectors, $\{W_k\}$ is a sequence of standard Gaussian random variables, and $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product between vectors. The primary goal is to reconstruct X^n from the set of m measurements $\{(Y_k, A_k)\}_{k=1}^m$. Since the reconstruction problem is symmetric under simultaneous scaling of X^n and $\{W_k\}$, the unit-variance assumption on $\{W_k\}$ incurs no loss of generality. In matrix form, the relationship between the signal and a set of m measurements is given by

$$Y^m = A^m X^n + W^m \tag{1}$$

where A^m is an $m \times n$ measurement matrix whose k-th row is A_k.

This paper analyzes the minimum mean-square error (MMSE) reconstruction in the asymptotic setting where the number of measurements m and the signal length n increase to infinity. The focus is on scaling regimes in which the measurement ratio $\delta_n = m/n$ converges to a number $\delta \in (0, \infty)$. The objective is to show that the normalized mutual information (MI) and MMSE converge to limits,

$$\mathcal{I}_n(\delta_n) \triangleq \frac{1}{n} I(X^n; Y^m \,|\, A^m) \to \mathcal{I}(\delta)$$
$$\mathcal{M}_n(\delta_n) \triangleq \frac{1}{n} \mathsf{mmse}(X^n \,|\, Y^m, A^m) \to \mathcal{M}(\delta),$$

almost everywhere and to characterize these limits in terms of the measurement ratio δ and the signal distribution P_X. We note that, throughout this paper, all logarithms are natural and all mutual informations are computed in nats.

Using the replica method from statistical physics, Guo and Verdú [1] provide an elegant characterization of these limits in the setting of i.i.d. measurement matrices. Their result was stated originally as a generalization of Tanaka's replica analysis of code-division multiple-access (CDMA) with binary signaling [2]. The replica method was also applied specifically to compressed sensing in [3]–[8]. The main issue, however, is that the replica method is not rigorous. It requires an exchange of limits that is unjustified, and it requires the assumption of replica symmetry, which is unproven in the context of the random linear estimation problem.

The main result of this paper is that replica prediction is correct for i.i.d. Gaussian measurement matrices provided that the signal distribution, P_X, has bounded fourth moment and satisfies a certain 'single-crossing' property. The proof differs from previous approaches in that we first establish some properties of the finite-length MMSE and MI sequences, and then use these properties to uniquely characterize their limits.

A. The Replica-Symmetric Prediction

For the stated problem, the replica method provides conjectured single-letter expressions for the asymptotic mutual information $\mathcal{I}(\delta)$ and the asymptoic MMSE $\mathcal{M}(\delta)$. The resulting replica formulas are defined in terms of the scalar mutual information function $I_X(s) \triangleq I(X; \sqrt{s}X + N)$ and the scalar MMSE function $\mathsf{mmse}_X(s) \triangleq \mathsf{mmse}(X \mid \sqrt{s}X + N)$. In these expressions, X is a scalar random variable drawn according to the signal distribution P_X, N is independent standard Gaussian noise, and $s \in \mathbb{R}_+$ parameterizes the signal-to-noise ratio of the scalar estimation problem.

There are a number of equivalent ways to express the replica formulas for this problem. Guo and Verdú provide

The work of G. Reeves was supported in part by funding from the Laboratory for Analytic Sciences (LAS) and the NSF under Grant No. 1718494. The work of H. Pfister was supported in part by the NSF under Grants No. 1545143 and No. 1718494. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors. This paper was presented in part at the Institut Henri Poincaré Spring 2016 Thematic Program on the Nexus of Information and Computation Theories [28] and at the 2016 International Symposium on Information Theory [29].

G. Reeves is with the Department of Electrical and Computer Engineering and the Department of Statistical Science, Duke University, Durham, NC 27708 USA (e-mail: galen.reeves@duke.edu). H. Pfister is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA (e-mail: henry.pfister@duke.edu)

Digital Object Identifier: 10.1109/TIT.2019.2891664

one representation in [1, Section II-D]. This paper uses an alternative representation that is more closely related to the arguments in our proof. An explicit connection between these two characterizations is given in Appendix E. Our expression depends on the function $R : \mathbb{R}_+^2 \to \mathbb{R}_+$, which is defined by

$$R(\delta, z) = I_X\left(\frac{\delta}{1+z}\right) + \frac{\delta}{2}\left[\log(1+z) - \frac{z}{1+z}\right]. \qquad (2)$$

Definition 1. The replica-MI function $\mathcal{I}_{\mathrm{RS}} \colon \mathbb{R}_+ \to \mathbb{R}_+$ and the replica-MMSE function $\mathcal{M}_{\mathrm{RS}} \colon \mathbb{R}_+ \to \mathbb{R}_+$ are defined as

$$\mathcal{I}_{\mathrm{RS}}(\delta) = \min_{z \geq 0} R(\delta, z)$$
$$\mathcal{M}_{\mathrm{RS}}(\delta) \in \arg\min_{z \geq 0} R(\delta, z).$$

The function $\mathcal{I}_{\mathrm{RS}}(\delta)$ is increasing because $R(\delta, z)$ is increasing in δ and it is concave because it is the pointwise minimum of concave functions. The concavity implies that $\mathcal{I}'_{\mathrm{RS}}(\delta)$ exists almost everywhere and is decreasing. It can also be shown that $\mathcal{M}_{\mathrm{RS}}(\delta)$ is also decreasing and, thus, continuous almost everywhere. If the minimizer is not unique, then $\mathcal{M}_{\mathrm{RS}}(\delta)$ may have jump discontinuities and may not be uniquely defined at those points; see Figure 1.

B. Statement of Main Result

In order to state our results, we need some further definitions. Let $R_z(\delta, z) = \frac{\partial}{\partial z} R(\delta, z)$ denote the partial derivative of $R(\delta, z)$ with respect to z. The *fixed-point curve* FP is the set of (δ, z) pairs where z is a stationary point of $R(\delta, z)$, i.e.,

$$\mathrm{FP} = \left\{(\delta, z) \in \mathbb{R}_+^2 \,:\, R_z(\delta, z) = 0\right\}.$$

To emphasize the connection with mutual information, we often plot this curve using the change of variables $z \mapsto \frac{1}{2}\log(1+z)$. The resulting curve, $(\delta, \frac{1}{2}\log(1+z))$, is called the *fixed-point information curve*; see Figure 1.

Definition 2 (Single-Crossing Property)**.** Informally, a signal distribution P_X has the single-crossing property if the replica-MMSE crosses the fixed-point curve FP at most once. A formal definition of the single-crossing property is given in Section VI-A.

Assumption 1 (IID Gaussian Measurements)**.** The rows of the measurement matrix $\{A_k\}$ are independent Gaussian vectors with mean zero and covariance $n^{-1} I_n$. Furthermore, the noise $\{W_k\}$ is i.i.d. Gaussian with mean zero and variance one.

Assumption 2 (IID Signal Entries)**.** The signal entries $\{X_i\}$ are independent copies of a random variable $X \sim P_X$ with bounded fourth moment $\mathbb{E}\left[X^4\right] \leq B$.

Assumption 3 (Single-Crossing Property)**.** The signal distribution P_X satisfies the single-crossing property.

Theorem 1. *Under Assumptions 1-3, we have*

(i) The sequence of MI functions $\mathcal{I}_n(\delta)$ converges to the replica prediction. In other words, for all $\delta \in \mathbb{R}_+$,

$$\lim_{n \to \infty} \mathcal{I}_n(\delta) = \mathcal{I}_{\mathrm{RS}}(\delta).$$

(ii) The sequence of MMSE functions $\mathcal{M}_n(\delta)$ converges almost everywhere to the replica prediction. In other words, for all continuity points of $\mathcal{M}_{\mathrm{RS}}(\delta)$,

$$\lim_{n \to \infty} \mathcal{M}_n(\delta) = \mathcal{M}_{\mathrm{RS}}(\delta).$$

Remark 1. The primary contribution of Theorem 1 is for the case where $\mathcal{M}_{\mathrm{RS}}(\delta)$ has a discontinuity. This occurs, for example, in applications such as compressed sensing with sparse priors and CDMA with finite alphabet signaling. For the special case where $\mathcal{M}_{\mathrm{RS}}(\delta)$ is continuous, the validity of the replica prediction can also be established by combining the AMP analysis with the I-MMSE relationship [9]–[13].

Remark 2. For a given signal distribution P_X the single-crossing property can be verified by numerically evaluating the replica-MMSE and checking whether it crosses the fixed-point curve more than once.

C. Context and Related Work

Throughout this paper we focus on the Bayes-optimal setting where the signal distribution P_{X^n} is known. This is an important special case of the more general setting in which reconstruction is based on a postulated distribution Q_{X^n} that is possibly different from true distribution. Within the context of the statistical physics literature, Bayes-optimal inference is directly related to the so-called Nishimori line [14] (see also [15, Section 2.6]).

The replica method was developed originally to study mean-field approximations in spin glasses [16], [17]. It was first applied to linear estimation problems in the context of CDMA wireless communication [1], [2], [18], with subsequent work focusing specifically on compressed sensing [3]–[8]. For a nice overview of the replica method and its application to inference problems, see [19].

Previous work has used the replica method to study the mean-square error (MSE) when the true distribution is i.i.d. P_X and the postulated distribution is i.i.d. Q_X. In this setting, Guo and Verdú [1] derive explicit formulas for the MSE corresponding to the conditional expectation. Building upon this work, Rangan et al. [5] used a 'hardening technique' that is well known in the statistics literature to derive conjectured formulas for MSE associate with larger class of estimators that includes maximum a posteriori (MAP) estimation. Recent work by Bereyhi et al. [20], [21] has derived conjectured formulas for the MSE using replica symmetry breaking. The work of Bereyhi et al. shows that replica-symmetric solution is not stable when there is significant mismatch between the true distortion and the postulated distribution.

Within the context of random linear estimation, the results of the replica method have been proven rigorously in a number of settings. One example is given by message passing on matrices with special structure, such as sparsity [9], [22], [23] or spatial coupling [15], [24], [25]. However, in the case of i.i.d. matrices, the results are limited to signal distributions with a unique fixed point [10], [12] (e.g., Gaussian inputs [26], [27]). For the special case of i.i.d. matrices with binary inputs, it has also been shown that the replica prediction provides an upper bound for the asymptotic mutual information [28]. Bounds on

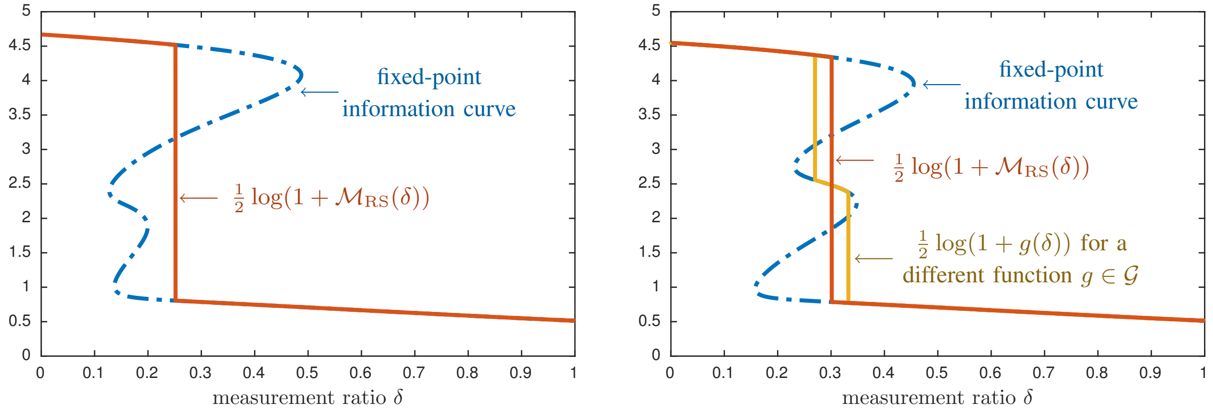

Fig. 1. Plot of the replica-MMSE as a function of the measurement ratio δ. The signal distribution is given by a three-component Gaussian mixture of the form $P_X = 0.4\mathcal{N}(0,5) + \alpha\mathcal{N}(40,5) + (0.6-\alpha)\mathcal{N}(220,5)$. In the left panel, $\alpha = 0.1$ and the distribution satisfies the single-crossing property. In the right panel, $\alpha = 0.3$ and the distribution does not satisfy the single-crossing property. The fixed-point information curve (dashed blue line) is given by $\frac{1}{2}\log(1+z)$ where z satisfies the fixed-point equation $R_z(\delta, z) = 0$.

the locations of discontinuities in the MMSE with sparse priors have also been obtain by analyzing the problem of approximate support recovery [6]–[8].

Using tools from statistical physics and random matrix theory, earlier work by Huleihel and Merhav [29] provides a rigorous characterization of the asymptotic MMSE for signals in a class of Gaussian mixture models [30], [31]. After our initial presentation [32] and publication [33], Barbier et al. [34] obtained similar results using a substantially different method.

D. Notation

We use C to denote an absolute constant and C_θ to denote a number that depends on a parameter θ. In all cases, the numbers C and C_θ are positive and finite, although their values change from place to place. The Euclidean norm is denoted by $\|\cdot\|$. The positive part of a real number x is denoted by $(x)_+ \triangleq \max(x,0)$. The nonnegative real line $[0,\infty)$ is denoted by $\mathbb{R}_+$ and the positive integers $\{1,2,\cdots\}$ are denoted by $\mathbb{N}$. For each $n \in \mathbb{N}$ the set $\{1,2,\cdots,n\}$ is denoted by $[n]$.

The joint distribution of the random variables X, Y is denoted by $P_{X,Y}$ and the marginal distributions are denoted by P_X and P_Y. The conditional distribution of X given $Y = y$ is denoted by $P_{X|Y=y}$ and the conditional distribution of X corresponding to a random realization of Y is denoted by $P_{X|Y}$. The expectation over a single random variable is denoted by $\mathbb{E}_X$. For example, this implies that $\mathbb{E}[f(X,Y)|Y] = \mathbb{E}_X[f(X,Y)]$.

Using this notation, the mutual information between X and Y can be expressed in terms of the expected Kullback-Leibler divergence as follows:

$$\begin{aligned} I(X;Y) &= D_{\mathrm{KL}}(P_{X,Y} \,\|\, P_X \times P_Y) \\ &= \mathbb{E}\big[D_{\mathrm{KL}}\big(P_{X|Y} \,\big\|\, P_X\big)\big] \\ &= \mathbb{E}\big[D_{\mathrm{KL}}\big(P_{Y|X} \,\big\|\, P_Y\big)\big], \end{aligned}$$

where the expectation in the second line is with respect to Y and the expectation in the third line is with respect to X.

The conditional variance of a random variable X given Y is denoted by

$$\mathsf{Var}(X\,|\,Y) = \mathbb{E}\big[(X - \mathbb{E}[X\,|\,Y])^2 \;\big|\; Y\big],$$

and the conditional covariance matrix of a random vector X given Y is denoted by

$$\mathsf{Cov}(X\,|\,Y) = \mathbb{E}\big[(X - \mathbb{E}[X\,|\,Y])(X - \mathbb{E}[X\,|\,Y])^T \;\big|\; Y\big].$$

The conditional variance and conditional covariance matrix are random because they are functions of the random conditional distribution $P_{X|Y}$.

The minimum mean-square error (MMSE) of X given Y is defined to be the expected squared difference between X and its conditional expectation and is denoted by

$$\mathsf{mmse}(X\,|\,Y) = \mathbb{E}\big[\|X - \mathbb{E}[X\,|\,Y]\|^2\big].$$

Since the expectation is taken with respect to both X and Y, the MMSE is a deterministic functions of the joint distribution $P_{X,Y}$. The MMSE can also be expressed in terms of the expected trace of the conditional covariance matrix:

$$\mathsf{mmse}(X\,|\,Y) = \mathbb{E}[\mathrm{tr}(\mathsf{Cov}(X\,|\,Y))].$$

II. Overview of Main Steps in Proof

We begin with some additional definitions. The finite-length MI sequence $I : \mathbb{N}^2 \to \mathbb{R}_+$ and MMSE sequence $M : \mathbb{N}^2 \to \mathbb{R}_+$ are defined according to

$$\begin{aligned} I_{m,n} &\triangleq I(X^n; Y^m \,|\, A^m), \\ M_{m,n} &\triangleq \frac{1}{n}\mathsf{mmse}(X^n \,|\, Y^m, A^m), \end{aligned}$$

where the relationship between the n-dimensional signal X^n, m-dimensional measurements Y^m, and $m \times n$ measurement matrix A^m is given by the measurement model (1). Furthermore, the first and second order MI difference sequences are defined according to

$$\begin{aligned} I'_{m,n} &\triangleq I_{m+1,n} - I_{m,n} \\ I''_{m,n} &\triangleq I'_{m+1,n} - I'_{m,n}. \end{aligned}$$

To simplify notation we will often drop the explicit dependence on the signal length n and simply write I_m, M_m, I'_m, and I''_m.

A. Properties of Finite-Length Sequences

At its foundation, our proof relies on certain relationships between the MI and MMSE sequences. Observe that by the chain rule for mutual information,

$$\underbrace{I(X^n; Y^m \,|\, A^m)}_{I_{m,n}} = \sum_{k=1}^{m-1} \underbrace{I(X^n; Y_{k+1} \,|\, Y^k, A^{k+1})}_{I'_{k,n}}.$$

Here, the conditioning in the mutual information on the right-hand side depends only on A^{k+1} because the measurement vectors are generated independently of the signal.

The above decomposition shows that the MI difference is given by $I'_{m,n} = I(X^n; Y_{m+1} \mid Y^m, A^{m+1})$. In other words, it is the mutual information between the signal and a new measurement Y_{m+1}, conditioned on the previous data (Y^m, A^{m+1}). One of the key steps in our proof is to show that the MI difference and MMSE satisfy the following relationship almost everywhere

$$I'_{m,n} \approx \frac{1}{2}\log(1 + M_{m,n}). \tag{3}$$

Our approach relies on the fact that the gap between the right and left sides of (3) can be viewed as a measure of the non-Gaussianness of the conditional distribution of the new measurement. By relating this non-Gaussianness to certain properties of the posterior distribution, we are able to show that (3) is tight whenever the second order MI difference sequence is small. The details of these steps are given in Section IV.

Another important relationship that is used in our proof is the following fixed-point identity for the MMSE

$$M_{m,n} \approx \mathsf{mmse}_X\left(\frac{m/n}{1 + M_{m,n}}\right). \tag{4}$$

where $\mathsf{mmse}_X(s) = \mathbb{E}\Big[\big(X - \mathbb{E}[X \,|\, \sqrt{s}X + N]\big)^2\Big]$ is the scalar MMSE function of $X \sim P_X$ under independent Gaussian noise $N \sim \mathcal{N}(0,1)$ with signal-to-noise ratio $s \in \mathbb{R}_+$ [1]. In words, this says that the MMSE of the random linear estimation problem is approximately equal to that of a scalar problem whose signal-to-noise ratio is a function of the measurement rate. In Section V it is shown that the tightness in (4) can be bounded in terms of the tightness of (3).

B. Asymptotic Constraints

The previous subsection focused on relationships between the finite-length MI and MMSE sequences. To characterize these relationships in the asymptotic setting of interest, the finite-length sequences are extended to functions of a continuous parameter $\delta \in \mathbb{R}_+$ according to

$$\begin{aligned}\mathcal{I}'_n(\delta) &= I'_{\lfloor \delta n \rfloor, n}\\ \mathcal{I}_n(\delta) &= \int_0^\delta \mathcal{I}'_n(\gamma)\,\mathrm{d}\gamma\\ \mathcal{M}_n(\delta) &= M_{\lfloor \delta n \rfloor, n}.\end{aligned}$$

This choice of interpolation has the convenient property that the MI function $\mathcal{I}_n$ is continuous and differentiable almost everywhere. Furthermore, by construction, $\mathcal{I}_n$ corresponds to the normalized mutual information and satisfies $\mathcal{I}_n\big(\frac{m}{n}\big) = \frac{1}{n} I_{m,n}$ for all integers m and n.

With this notation in hand, we are now ready to state two of the main theorems in this paper. These theorems provide precise bounds on the relationships given in (3) and (4). The proofs are given in Section IV and Section V.

Theorem 2. *Under Assumptions 1 and 2, the MI and MMSE functions satisfy*

$$\int_0^\delta \left| \mathcal{I}'_n(\gamma) - \frac{1}{2}\log(1 + \mathcal{M}_n(\gamma)) \right| \mathrm{d}\gamma \le C_{B,\delta} \cdot n^{-r},$$

for all $n \in \mathbb{N}$ and $\delta \in \mathbb{R}_+$ where $r \in (0,1)$ is a universal constant.

Theorem 3. *Under Assumptions 1 and 2, the MMSE function satisfies*

$$\int_0^\delta \left| \mathcal{M}_n(\gamma) - \mathsf{mmse}_X\left(\frac{\gamma}{1 + \mathcal{M}_n(\gamma)}\right) \right| \mathrm{d}\gamma \le C_{B,\delta} \cdot n^{-r},$$

for all $n \in \mathbb{N}$ and $\delta \in \mathbb{R}_+$ where $r \in (0,1)$ is a universal constant.

The bounds given in Theorems 2 and 3 are with respect the integrals over $\mathcal{I}'_n$ and $\mathcal{M}_n$, and thus prove convergence in L_1 over bounded intervals. This is sufficient to show that the relationships hold almost everywhere in the limit. Importantly, though, these bounds still allow for the possibility that the relationships do not hold at countably many points, and thus allow for the possibility of phase transitions.

For distributions that have a phase transition, our proof technique requires a boundary condition for the mutual information. This boundary condition is used to determine the location of the phase transition. The next result shows that the replica-MI is equal to the MI-function in the limit as the measurement rate increases to infinity, and thus the replica-MI can be used as a boundary condition. The proof is given in Section III-D.

Theorem 4. *Under Assumptions 1 and 2, the MI function satisfies*

$$|\mathcal{I}_n(\delta) - \mathcal{I}_{\mathrm{RS}}(\delta)| \le C \cdot \delta^{-\frac{1}{2}},$$

for all $n \in \mathbb{N}$ and $\delta \ge 4$.

At first glance, it may seem surprising that Theorem 4 holds for all signal lengths. However, this bound is tight in the regime where the number of measurements is much larger than the number of signal entries. From the rotational invariance of the Gaussian noise and monotonicity of the mutual information with respect to the signal-to-noise ratio, one can obtain the sandwiching relation

$$n\,\mathbb{E}\big[I_X\big(\sigma^2_{\min}(A^m)\big)\big] \le I_{m,n} \le n\,\mathbb{E}\big[I_X\big(\sigma^2_{\max}(A^m)\big)\big],$$

where the upper and lower bounds depend only on the minimum and maximum singular values of the random $m \times n$ measurement matrix. For fixed n, it is well known that the ratio of these singular values converges to one almost surely in the limit as m increases to infinity. Our proof of Theorem 4 uses a more refined analysis, based on the QR decomposition, but the basic idea is the same.

C. Uniqueness of Limit

The final step of the proof is to make the connection between the asymptotic constraints on the MI and MMSE described in the previous subsection and the replica-MI and replica-MMSE functions given in Definition 1.

We begin by describing two functional properties of the replica limits. The first property follows from the fact that the MMSE is a global minimizer of the function $R(\delta, z)$ with respect to its second argument. Since $R(\delta, z)$ is differentiable for all $\delta, z \in \mathbb{R}_+$, any minimizer z^* of $R(\delta, z)$ must satisfy the equation $R_z(\delta, z^*) = 0$, where $R_z(\delta, z) = \frac{\partial}{\partial z} R(\delta, z)$. Using the I-MMSE relationship [35], it can be shown that, for each $\delta \in \mathbb{R}_+$, the replica-MMSE $\mathcal{M}_{\mathrm{RS}}(\delta)$ satisfies the fixed-point equation

$$\mathcal{M}_{\mathrm{RS}}(\delta) = \mathsf{mmse}_X\left(\frac{\delta}{1 + \mathcal{M}_{\mathrm{RS}}(\delta)}\right). \tag{5}$$

Note that if this fixed-point equation has a unique solution, then it provides an equivalent definition of the replica-MMSE. However, in the presence of multiple solutions, only the solutions that correspond to global minimizers of $R(\delta, z)$ are valid.

The second property relates the derivative of the replica-MI to the replica-MMSE. Specifically, as a consequence of the envelope theorem [36] and (5), it can be shown that the derivative of the replica-MI satisfies

$$\mathcal{I}'_{\mathrm{RS}}(\delta) = \frac{1}{2}\log(1 + \mathcal{M}_{\mathrm{RS}}(\delta)), \tag{6}$$

almost everywhere.

These properties show that the replica limits $\mathcal{I}_{\mathrm{RS}}$ and $\mathcal{M}_{\mathrm{RS}}$ satisfy the relationships given in Theorems 2 and 3 with equality. In order to complete proof we need to show that, in conjunction with a boundary condition imposed by the large measurement rate limit, the constraints (5) and (6) provide an equivalent characterization of the replica limits.

Definition 3. For a given signal distribution P_X, let $\mathcal{V}$ be the subset of non-increasing functions from $\mathbb{R}_+ \to \mathbb{R}_+$ such that all $g \in \mathcal{V}$ satisfy the fixed-point condition

$$g(\delta) = \mathsf{mmse}_X\left(\frac{\delta}{1 + g(\delta)}\right). \tag{7}$$

Furthermore, let $\mathcal{G} \subseteq \mathcal{V}$ be the subset such that all $g \in \mathcal{G}$ also satisfy the large measurement rate boundary condition

$$\lim_{\delta \to \infty}\left|\int_0^\delta \frac{1}{2}\log(1 + g(\gamma))\mathrm{d}\gamma - \mathcal{I}_{\mathrm{RS}}(\delta)\right| = 0.$$

In Section VI-A, it is shown that if the signal distribution P_X has the single-crossing property, then $\mathcal{M}_{\mathrm{RS}}(\delta)$ has at most one discontinuity and all $g \in \mathcal{G}$ satisfy $g(\delta) = \mathcal{M}_{\mathrm{RS}}(\delta)$ almost everywhere. In other words, the single-crossing property provides a sufficient condition under which the replica limits can be obtained uniquely from (6) and (7). A graphical illustration is provided in Figure 1.

III. Properties of MI and MMSE

A. Single-Letter Functions

The single-letter MI and MMSE functions corresponding to a real valued input distribution P_X under additive Gaussian noise are defined by

$$I_X(s) \triangleq I(X; \sqrt{s}X + N)$$
$$\mathsf{mmse}_X(s) \triangleq \mathsf{mmse}(X \mid \sqrt{s}X + N),$$

where $X \sim P_X$ and $N \sim \mathcal{N}(0,1)$ are independent and $s \in \mathbb{R}_+$ parametrizes the signal-to-noise ratio. Many properties of these functions have been studied in the literature [35], [37]–[39]. The function $I_X(s)$ is concave and non-decreasing with $I_X(0) = 0$. If $I_X(s)$ is finite for some $s > 0$ then it is finite for all $s \in \mathbb{R}_+$ [39, Theorem 6]. The MMSE function is non-increasing with $\mathsf{mmse}_X(0) = \mathsf{Var}(X)$. Both $I_X(s)$ and $\mathsf{mmse}_X(s)$ are infinitely differentiable on $(0, \infty)$ [37, Proposition 7].

The so-called I-MMSE relationship [35] states that

$$\frac{\mathrm{d}}{\mathrm{d}s} I_X(s) = \frac{1}{2}\mathsf{mmse}_X(s). \tag{8}$$

This relationship was originally stated for input distributions with finite second moments [35, Theorem 1], and was later shown to hold for any input distributions with finite mutual information [39, Theorem 6]. This relationship can also be viewed as a consequence of De-Bruijn's identity [40].

Furthermore, it is well known that under a second moment constraint, the MI and MMSE are maximized when the input distribution is Gaussian. This yields the following inequalities

$$I_X(s) \le \frac{1}{2}\log(1 + s\,\mathsf{Var}(X)) \tag{9}$$

$$\mathsf{mmse}_X(s) \le \frac{\mathsf{Var}(X)}{1 + s\,\mathsf{Var}(X)}. \tag{10}$$

More generally, the MMSE function satisfies the upper bound $\mathsf{mmse}_X(s) \le 1/s$, for every input distribution P_X and $s > 0$ [37, Proposition 4]. Combining this inequality with (8) leads to

$$I_X(t) - I_X(s) \le \frac{1}{2}\log\left(\frac{t}{s}\right), \quad 0 < s \le t, \tag{11}$$

which holds for any input distribution with finite mutual information.

Finally, the derivative of the MMSE with respect to s is given by the second moment of the conditional variance [37, Proposition 9],

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathsf{mmse}_X(s) = -\mathbb{E}\Big[\big(\mathsf{Var}(X \,|\, Y)\big)^2\Big]. \tag{12}$$

This expression can be used to show that $\mathsf{mmse}_X(s)$ is Lipschitz continuous for any P_X with finite fourth moment. The following result is proved in Appendix B-A.

Lemma 5. *The single-letter MMSE function satisfies the following bounds:*

(i) For any input distribution P_X with finite fourth moment and $s, t \in \mathbb{R}_+$,

$$|\mathsf{mmse}_X(s) - \mathsf{mmse}_X(t)| \le 4\mathbb{E}\big[X^4\big]|s - t|. \tag{13}$$

(ii) For every input distribution P_X and $s, t \in (0, \infty)$,

$$|\mathsf{mmse}_X(s) - \mathsf{mmse}_X(t)| \le 12\left|\frac{1}{s} - \frac{1}{t}\right|. \tag{14}$$

B. Multivariate MI and MMSE

From the chain rule for mutual information, we see that the MI difference sequence is given by

$$I'_{m,n} = I(X^n; Y_{m+1} \,|\, Y^m, A^{m+1}). \tag{15}$$

By the non-negativity of mutual information, this establishes that the MI sequence is non-decreasing in m. Alternatively, by the data-processing inequality for MMSE [41, Proposition 4], we also see that $M_{m,n}$ is non-increasing in m, and also

$$M_{m,n} \le \mathsf{Var}(X).$$

The next result shows that the second order MI difference can also be expressed in terms of mutual information. The proof is given in Appendix B-B.

Lemma 6. *Under Assumption 1, the second order MI difference sequence satisfies*

$$I''_{m,n} = -I(Y_{m+1}; Y_{m+2} \mid Y^m, A^{m+2}). \tag{16}$$

One consequence of Lemma 6 is that the first order MI difference sequence is non-increasing in m, and thus

$$I'_{m,n} \le I'_{1,n} = I_{1,n}. \tag{17}$$

This inequality plays an important role later on in our proof, when we show that certain terms of interest are bounded by the magnitude of the second order MI difference.

The next result provides non-asymptotic bounds in terms of the single-letter MI and MMSE functions corresponding to the signal distribution P_X. The proof is given in Appendix B-C

Lemma 7. *Under Assumptions 1 and 2, the MI and MMSE sequences satisfy*

$$\sum_{k=1}^{\min(n,m)} \mathbb{E}\big[I_X\big(\tfrac{1}{n}\chi^2_{m-k+1}\big)\big] \le I_{m,n} \le n\,\mathbb{E}\big[I_X\big(\tfrac{1}{n}\chi^2_m\big)\big] \tag{18}$$

$$\mathbb{E}\big[\mathsf{mmse}_X\big(\tfrac{1}{n}\chi^2_m\big)\big] \le M_{m,n} \le \mathbb{E}\big[\mathsf{mmse}_X\big(\tfrac{1}{n}\chi^2_{m-n+1}\big)\big], \tag{19}$$

where χ^2_k denotes a chi-squared random variable with k degrees of freedom and the upper bound on $M_{m,n}$ is valid for all $m \ge n$.

Remark 3. The proof of Lemma 7 does not require the assumption that the signal distribution has bounded fourth moment. In fact, (18) holds for any signal distribution with finite mutual information and (19) holds for any signal distribution.

Remark 4. The upper bound in (18) and lower bound in (19) are not new and are special cases of results given in [8].

Combining Lemma 7 with Inequalities (9) and (10), leads to upper bounds on the MI and MMSE that depend only on the variance of the signal distribution. Alternatively, combining Lemma 7 with the smoothness of the single-letter functions given in (11) and (14) leads to the following characterization, which is tight whenever m is much larger than n. The proof is given in Appendix B-D

Lemma 8. *Under Assumptions 1 and 2, the MI and MMSE sequences satisfy, for all $m \ge n+2$,*

$$\left|\tfrac{1}{n}I_{m,n} - I_X\left(\tfrac{m}{n}\right)\right| \le \tfrac{1}{2}\left[\tfrac{n+1}{m-n-1} + \sqrt{\tfrac{2}{m-2}}\right] \tag{20}$$

$$\left|M_{m,n} - \mathsf{mmse}_X\left(\tfrac{m}{n}\right)\right| \le \tfrac{12\,n}{m}\left[\tfrac{n+1}{m-n-1} + \sqrt{\tfrac{2}{m-2}}\right]. \tag{21}$$

For any fixed n, the right-hand sides of (20) and (21) converge to zero as m increases to infinity. Consequently, the large m behavior of the MI sequence is given by

$$\lim_{m\to\infty} I_{m,n} = \begin{cases} H(X^n), & \text{if } P_X \text{ has finite entropy} \\ +\infty, & \text{otherwise.} \end{cases}$$

C. Properties of Replica Prediction

Using the I-MMSE relationship, the partial derivative of $R(\delta, z)$ with respect to its second argument is given by

$$R_z(\delta, z) = \frac{\delta}{2(1+z)^2}\left[z - \mathsf{mmse}_X\left(\frac{\delta}{1+z}\right)\right]. \tag{22}$$

From this expression, we see that the $R_z(\delta, z) = 0$ is equivalent to the fixed-point condition

$$z = \mathsf{mmse}\left(\frac{\delta}{1+z}\right).$$

Furthermore, since $\mathcal{I}_{\mathrm{RS}}(\delta)$ is concave, it is differentiable almost everywhere. For all δ where $\mathcal{I}'_{\mathrm{RS}}(\delta)$ exists, it follows from the envelope theorem [36] that $\mathcal{I}'_{\mathrm{RS}}(\delta) = R_\delta(\delta, \mathcal{M}_{\mathrm{RS}}(\delta))$, where $R_\delta(z, \delta)$ is the partial derivative of $R(\delta, z)$ with respect to its first argument. Direct computation yields

$$R_\delta(\delta, z) = \frac{1}{2}\log(1+z) + \frac{1}{2(1+z)}\left[\mathsf{mmse}_X\left(\frac{\delta}{1+z}\right) - z\right].$$

Finally, noting that the second term on the right-hand side is equal to zero whenever $z = \mathcal{M}_{\mathrm{RS}}(\delta)$, leads to

$$\mathcal{I}'_{\mathrm{RS}}(\delta) = \frac{1}{2}\log(1 + \mathcal{M}_{\mathrm{RS}}(\delta)).$$

The proof of the next result is given in Appendix B-E.

Lemma 9. *The Replica-MI and Replica-MMSE functions satisfy, for all $\delta \ge 1$,*

$$I_X(\delta - 1) \le \mathcal{I}_{\mathrm{RS}}(\delta) \le I_X(\delta), \tag{23}$$

$$\mathsf{mmse}_X(\delta) \le \mathcal{M}_{\mathrm{RS}}(\delta) \le \mathsf{mmse}_X(\delta - 1). \tag{24}$$

It is interesting to note the parallels between the bounds on the MI and MMSE sequences in Lemma 7 and the bounds on the replica functions in Lemma 9. Combining Lemma 9 with the smoothness of the single-letter functions given in (11) and (14) leads to

$$|\mathcal{I}_{\mathrm{RS}}(\delta) - I_X(\delta)| \le \frac{1}{2(\delta - 1)}$$

$$|\mathcal{M}_{\mathrm{RS}}(\delta) - \mathsf{mmse}_X(\delta)| \le \frac{1}{\delta(\delta - 1)}.$$

D. Proof of Theorem 4

This proof follows from combining Lemmas 8 and 9. Fix any $n \in \mathbb{R}_+$ and $\delta > 4$ and let $m = \lfloor \delta n \rfloor$ and $\lambda = m+1-\delta n$. The MI function obeys the upper bound

$$\begin{aligned}\mathcal{I}_n(\delta) &= \frac{1}{n}[\lambda I_{m,n} + (1-\lambda) I_{m+1,n}] \\ &\overset{(a)}{\le} \lambda \mathbb{E}\big[I_X\big(\tfrac{1}{n}\chi_m^2\big)\big] + (1-\lambda)\mathbb{E}\big[I_X\big(\tfrac{1}{n}\chi_{m+1}^2\big)\big] \\ &\overset{(b)}{\le} I_X(\delta), \end{aligned} \tag{25}$$

where: (a) follows from (18); and (b) follows from Jensen's inequality and the concavity of $I_X(s)$. The MI function also obeys the lower bound

$$\begin{aligned}\mathcal{I}_n(\delta) &\overset{(a)}{\ge} \frac{1}{n} I_{m,n} \\ &\overset{(b)}{\ge} I_X(\delta) - \tfrac{1}{2(\delta-1)} - \tfrac{1}{2}\Big[\tfrac{n+1}{(\delta-1)n-2} - \sqrt{\tfrac{2}{\delta-3}}\Big], \end{aligned} \tag{26}$$

where (a) follows from the fact that $I_{m,n}$ is non-decreasing in m and (b) follows from (20), (11), and the fact that $m \ge \delta n - 1 \ge \delta - 1$. Finally, we have

$$\begin{aligned}|\mathcal{I}_n(\delta) - \mathcal{I}_{\mathrm{RS}}(\delta)| &\overset{(a)}{\le} |\mathcal{I}_n(\delta) - I_X(\delta)| + |I_X(\delta) - \mathcal{I}_{\mathrm{RS}}(\delta)| \\ &\overset{(b)}{\le} \tfrac{1}{(\delta-1)} + \tfrac{1}{2}\tfrac{n+1}{(\delta-1)n-2} + \tfrac{1}{2}\sqrt{\tfrac{2}{\delta-3}} \\ &\overset{(c)}{\le} \Big(4+\sqrt{2}\Big)\delta^{-\frac{1}{2}}, \end{aligned}$$

where: (a) follows from the triangle inequality; (b) follows from (23), (25), and (26); and (c) follows from the assumption $\delta \ge 4$. This completes the proof of Theorem 4.

E. Concentration of MI Density

In order to establish the MI and MMSE relationships used in our proof, we need to show that certain functions of the random tuple (X^n, W^n, A^n) concentrate about their expectations.

Our first result bounds the variation in the mutual information corresponding to the measurement matrix. Let $I_{m,n}(A^m)$ denote the MI sequence as a function of the random matrix A^m. The following result follows from the Gaussian Poincaré inequality and the multivariate I-MMSE relationship. The proof is given in Appendix B-F.

Lemma 10. *Under Assumption 1, the variance of the MI with respect to the measurement matrix satisfies*

$$\mathsf{Var}(I_{m,n}(A^m)) \le M_{m,n}. \tag{27}$$

It is interesting to note that Lemma 10 does not require any assumptions about the signal distribution. An important consequence of this result is that the variance of the normalized mutual information $\frac{1}{n} I_{m,n}(A^m)$ converges to zero as $\max(m,n)$ increases to infinity.

Next, we focus on the concentration of the mutual information density, which is a random variable whose expectation is equal to the mutual information.

Definition 4. Given a distribution $P_{X,Y,Z}$, the *conditional mutual information density* between X and Y given Z is defined as

$$\imath(X;Y\,|\,Z) \triangleq \log\left(\frac{\mathrm{d}P_{X,Y|Z}(X,Y\,|\,Z)}{\mathrm{d}\big(P_{X|Z}(X\,|\,Z) \times P_{Y|Z}(Y\,|\,Z)\big)}\right),$$

where $(X,Y,Z) \sim P_{X,Y,Z}$. This is well-defined because a joint distribution is absolutely continuous with respect to the product of its marginals.

The mutual information density satisfies many of the same properties as mutual information, such as the chain rule and invariance to one-to-one transformations; see [43, Chapter 5.5]. For this random linear estimation problem, the mutual information density can be expressed in terms of the density functions $f_{Y^m|X^n,A^m}$ and $f_{Y^m|A^m}$, which are guaranteed to exist because of the additive Gaussian noise:

$$\imath(X^n;Y^m\,|\,A^m) = \log\left(\frac{f_{Y^m|X^n,A^m}(Y^m\,|\,X^n,A^m)}{f_{Y^m|A^m}(Y^m\,|\,A^m)}\right).$$

The next result bounds the variance of the mutual information density in terms of the fourth moment of the signal distribution and the problem dimensions. The proof is given in Appendix B-G.

Lemma 11. *Under Assumptions 1 and 2, the variance of the MI density satisfies*

$$\mathsf{Var}(\imath(X^n;Y^m \mid A^m)) \le C_B \cdot \left(1+\frac{m}{n}\right)^2 n.$$

IV. Proof of Theorem 2

This section describes the proof of Theorem 2. An outline of the dependences between various steps is provided in Figure 2.

A. Further Definitions

The conditional distribution induced by the data (Y^m, A^m) plays an important role and is referred to throughout as the *posterior distribution*. The optimal signal estimate with respect to squared error is given by the mean of the posterior distribution, and the squared error associated with this estimate is denoted by

$$\mathcal{E}_{m,n} = \frac{1}{n}\|X^n - \mathbb{E}[X^n\,|\,Y^m, A^m]\|^2.$$

The conditional expectation of the squared error with respect to the posterior distribution is referred to as the *posterior variance* and is denoted by

$$V_{m,n} = \mathbb{E}[\mathcal{E}_{m,n}\,|\,Y^m, A^m].$$

Both $\mathcal{E}_{m,n}$ and $V_{m,n}$ are random variables. By construction, their expectations are equal to the MMSE, that is

$$M_{m,n} = \mathbb{E}[V_{m,n}] = \mathbb{E}[\mathcal{E}_{m,n}].$$

Next, recall that the MI difference sequence can be expressed in terms of the mutual information between the signal and a new measurement:

$$I'_{m,n} = I(X^n; Y_{m+1}\,|\,Y^m, A^{m+1}).$$

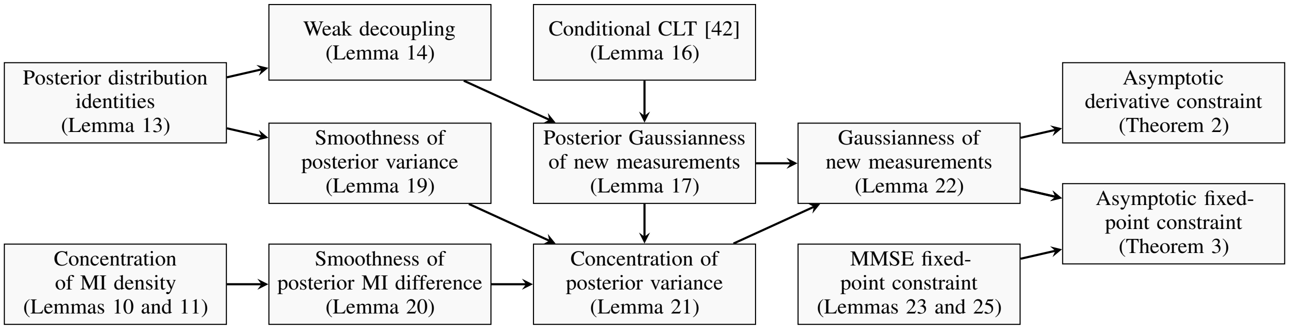
Posterior distribution identities (Lemma 13)
Weak decoupling (Lemma 14)
Conditional CLT [42] (Lemma 16)
Smoothness of posterior variance (Lemma 19)
Posterior Gaussianness of new measurements (Lemma 17)
Gaussianness of new measurements (Lemma 22)
Asymptotic derivative constraint (Theorem 2)
Asymptotic fixed-point constraint (Theorem 3)
Concentration of MI density (Lemmas 10 and 11)
Smoothness of posterior MI difference (Lemma 20)
Concentration of posterior variance (Lemma 21)
MMSE fixed-point constraint (Lemmas 23 and 25)

Fig. 2. Outline of the main steps in the proofs of Theorem 2 and Theorem 3 .

The *MI difference density* is defined to be the random variable

$$\mathcal{J}_{m,n} = \imath(X^n; Y_{m+1} \,|\, Y^m, A^{m+1}),$$

where the mutual information density is defined in Definition 4. The conditional expectation of the MI difference density with respect to the posterior distribution is referred to as the *posterior MI difference* and is denoted by

$$J_{m,n} = \mathbb{E}\big[\imath(X^n; Y_{m+1} \,|\, Y^m, A^{m+1}) \;\big|\; Y^m, A^m\big].$$

Both $\mathcal{J}_{m,n}$ and $J_{m,n}$ are random variables. By construction, their expectations are equal to the MI difference, that is

$$\mathbb{E}[\mathcal{J}_{m,n}] = \mathbb{E}[J_{m,n}] = I'_{m,n}.$$

A summary of this notation is provided in Table I.

The next result shows that the moments of the square error and MI difference density can be bounded in terms of the fourth moment of the signal distribution. The proof is given in Appendix C-A.

Lemma 12. *Under Assumptions 1 and 2,*

$$\mathbb{E}\left[\left|\mathcal{E}_{m,n}\right|^2\right] \le C \cdot B \tag{28}$$

$$\mathbb{E}\left[\left|\mathcal{J}_{m,n}\right|^2\right] \le C \cdot (1 + B) \tag{29}$$

$$\mathbb{E}\left[\left|Y_{m+1}\right|^4\right] \le C \cdot (1 + B) \tag{30}$$

$$\mathbb{E}\left[\left|Y_{m+1} - \widehat{Y}_{m+1}\right|^4\right] \le C \cdot (1 + B), \tag{31}$$

where $\widehat{Y}_{m+1} = \mathbb{E}[Y_{m+1} \,|\, Y^m, A^m]$.

B. Weak Decoupling

The posterior distribution of the signal cannot, in general, be expressed as the product of its marginals since the measurements introduce dependence between the signal entries. Nevertheless, it has been observed that in some cases, the posterior distribution satisfies a *decoupling principle* [1], in which the posterior distribution on a small subset of the signal entries is well approximated by the product of the marginals on that subset. One way to express this decoupling is say that, for any fixed index set $\{i_1, \cdots i_L\}$ with $L \ll n$, the random posterior distribution satisfies

$$P_{X_{i_1}, \cdots, X_{i_L} | Y^m, A^m} \approx \prod_{\ell=1}^{L} P_{X_{i_\ell} | Y^m, A^m},$$

with high probability with respect to the data (Y^m, A^m).

One of the main ideas in our proof is to use decoupling to show that the MI and MMSE sequences satisfy certain relationships. For the purposes of our proof, it is sufficient to work with a weaker notion of decoupling that depends only on the statistics of the pairwise marginals of the posterior distribution.

Definition 5. The posterior signal distribution $P_{X^n|Y^m,A^m}$ satisfies *weak decoupling* if

$$\mathbb{E}[|\mathcal{E}_{m,n} - V_{m,n}|] \to 0$$

$$\frac{1}{n}\mathbb{E}[\|\mathsf{Cov}(X^n \,|\, Y^m, A^m)\|_F] \to 0.$$

as m and n increase to infinity.

The first condition in the definition of weak decoupling says that the magnitude of the squared error must concentrate about its conditional expectation. The second condition says that the average correlation between the signal entries under the posterior distribution is converging to zero. Note that both of these conditions are satisfied a priori in the case of $m = 0$ measurements, because that prior signal distribution is i.i.d. with finite fourth moments.

The next result provides several key identities, which show that certain properties of the posterior distribution can be expressed in terms of the second order statistics of new measurements. This result does not require Assumption 2, and thus holds generally for any prior distribution on X^n. The proof is given in Appendix C-B.

Lemma 13. *Under Assumption 1, the following identities hold for all integers* $m < i < j$*:*

(i) The posterior variance satisfies

$$V_{m,n} = \mathbb{E}_{A_i}[\mathsf{Var}(Y_i \,|\, Y^m, A^m, A_i)] - 1. \tag{32}$$

(ii) The posterior covariance matrix satisfies

$$\frac{1}{n^2}\|\mathsf{Cov}(X^n \,|\, Y^m, A^m)\|_F^2 = \mathbb{E}_{A_i, A_j}\left[\left|\mathsf{Cov}(Y_i, Y_j \,|\, Y^m, A^m, A_i, A_j)\right|^2\right]. \tag{33}$$

(iii) The conditional variance of $\sqrt{1 + \mathcal{E}_{m,n}}$ *satisfies*

$$\mathsf{Var}\Big(\sqrt{1 + \mathcal{E}_{m,n}} \;\Big|\; Y^m, A^m\Big) = \frac{\pi}{2}\,\mathsf{Cov}\Big(\big|Y_i - \widehat{Y}_i\big|, \big|Y_j - \widehat{Y}_j\big| \;\Big|\; Y^m, A^m\Big), \tag{34}$$

TABLE I
SUMMARY OF NOTATION USED IN THE PROOFS OF THEOREM 2 AND THEOREM 3

	Random Variable	Posterior Expectation	Expectation
Squared Error	$\mathcal{E}_m = \frac{1}{n}\|X^n - \mathbb{E}[X^n \,|\, Y^m, A^m]\|^2$	$V_m = \mathbb{E}[\mathcal{E}_m \,|\, Y^m, A^m]$	$M_m = \mathbb{E}[\mathcal{E}_m]$
MI Difference	$\mathcal{J}_m = \imath(X^n; Y_{m+1} \,|\, Y^m, A^{m+1})$	$J_m = \mathbb{E}[\mathcal{J}_m \,|\, Y^m, A^m]$	$I'_m = \mathbb{E}[\mathcal{J}_m]$

where $\widehat{Y}_i = \mathbb{E}[Y_i \mid Y^m, A^m, A_i]$.

Identity (33) relates the correlation of the signal entries under the posterior distribution to the correlation of new measurements. Identity (34) relates the deviation of the squared error under the posterior distribution to the correlation between new measurements. Combining these identities with the bounds on the relationship between covariance and mutual information given in Appendix A-C leads to the following result. The proof is given in Appendix C-C.

Lemma 14. *Under Assumptions 1 and 2, the posterior variance and the posterior covariance matrix satisfy*

$$\mathbb{E}[|\mathcal{E}_{m,n} - V_{m,n}|] \le C_B \cdot \left|I''_{m,n}\right|^{\frac{1}{4}} \tag{35}$$

$$\frac{1}{n}\mathbb{E}[\|\mathsf{Cov}(X^n \mid Y^m, A^m)\|_F] \le C_B \cdot \left|I''_{m,n}\right|^{\frac{1}{4}}. \tag{36}$$

C. Gaussiannness of New Measurements

The *centered measurement* $\bar{Y}_{m+1}$ is defined to be the difference between a new measurement and its conditional expectation given the previous data:

$$\bar{Y}_{m+1} \triangleq Y_{m+1} - \mathbb{E}\big[Y_{m+1} \mid Y^m, A^{m+1}\big].$$

Conditioned on the data (Y^m, A^{m+1}), the centered measurement provides the same information as Y_{m+1}, and thus the posterior MI difference and the MI difference can be expressed equivalently as

$$\begin{aligned} J_m &= \mathbb{E}\big[\imath(X^n; \bar{Y}_{m+1} \mid Y^m, A^{m+1}) \bigm| Y^m, A^m\big] \\ I'_m &= I(X^n; \bar{Y}_{m+1} \mid Y^m, A^{m+1}). \end{aligned}$$

Furthermore, by the linearity of expectation, the centered measurement can be viewed as a noisy linear projection of the signal error:

$$\bar{Y}_{m+1} = \langle A_{m+1}, \bar{X}^n \rangle + W_{m+1}, \tag{37}$$

where $\bar{X}^n = X^n - \mathbb{E}[X^n \,|\, Y^m, A^m]$. Since the measurement vector A_{m+1} and noise term W_{m+1} are independent of everything else, the variance of the centered measurement can be related directly to the posterior variance V_m and the MMSE M_m via the following identities:

$$\mathsf{Var}(\bar{Y}_{m+1} \,|\, Y^m, A^m) = 1 + V_{m,n} \tag{38}$$

$$\mathsf{Var}(\bar{Y}_{m+1}) = 1 + M_{m,n}. \tag{39}$$

Identity (38) follows immediately from Lemma 13. Identity (39) follows from the fact that the centered measurement has zero mean, by construction, and thus its variance is equal to the expectation of (38).

At this point, the key question for our analysis is the extent to which the conditional distribution of the centered measurement can be approximated by a zero-mean Gaussian distribution. We focus on two different measures of non-Gaussianness. The first measure, which is referred to as the *posterior non-Gaussianness*, is defined by the random variable

$$\Delta^P_{m,n} \triangleq \mathbb{E}_{A_{m+1}}\Big[D_{\mathrm{KL}}\Big(P_{\bar{Y}_{m+1}|Y^m, A^{m+1}} \Big\| \mathcal{N}(0, 1 + V_m)\Big)\Big].$$

This is the Kullback–Leibler divergence with respect to the Gaussian distribution whose variance is matched to the conditional variance of $\bar{Y}_{m+1}$ given the data (Y^m, A^m).

The second measure, which is referred to simply as the *non-Gaussianness*, is defined by

$$\Delta_{m,n} \triangleq \mathbb{E}\Big[D_{\mathrm{KL}}\Big(P_{\bar{Y}_{m+1}|Y^m, A^{m+1}} \Big\| \mathcal{N}(0, 1 + M_m)\Big)\Big].$$

Here, the expectation is taken with respect to the tuple (Y^m, A^{m+1}) and the comparison is with respect to the Gaussian distribution whose variance is matched to the marginal variance of $\bar{Y}_{m+1}$.

The connection between the non-Gaussianness of the centered measurement and the relationship between the mutual information and MMSE sequences is given by the following result. The proof is given in Appendix C-D.

Lemma 15. *Under Assumption 1, the posterior non-Gaussianness and the non-Gaussianness satisfy the following identities:*

$$\Delta^P_{m,n} = \frac{1}{2}\log(1 + V_{m,n}) - J_{m,n} \tag{40}$$

$$\Delta_{m,n} = \frac{1}{2}\log(1 + M_{m,n}) - I'_{m,n}. \tag{41}$$

Identity (41) shows the integral relationship between mutual information and MMSE in Theorem 2 can be stated equivalently in terms of the non-Gaussianness of the centered measurements. Furthermore, by combining (41) with (40), we see that the non-Gaussianness can be related to the expected posterior non-Gaussianness using the following decomposition:

$$\Delta_{m,n} = \mathbb{E}\big[\Delta^P_{m,n}\big] + \frac{1}{2}\mathbb{E}\left[\log\left(\frac{1 + M_{m,n}}{1 + V_{m,n}}\right)\right]. \tag{42}$$

The rest of this subsection is focused on bounding the expected posterior non-Gaussianness. The second term on the right-hand side of (42) corresponds to the deviation of the posterior variance and is considered in the next subsection.

The key step in bounding the posterior non-Gaussianness is provided by the following result, which bounds the expected Kullback–Leibler divergence between the conditional distribution of a random projection and a Gaussian approximation [42].

Lemma 16 ([42]). *Let U be an n-dimensional random vector with mean zero and $\mathbb{E}\big[\|U\|^4\big] < \infty$, and let $Y = \langle A, U \rangle + W$, where $A \sim \mathcal{N}(0, \frac{1}{n} I_n)$ and $W \sim \mathcal{N}(0, 1)$ are independent.*

Then, the expected KL divergence between $P_{Y|A}$ and the Gaussian distribution with the same mean and variance as P_Y satisfies

$$\begin{aligned}&\mathbb{E}\big[D_{\mathrm{KL}}\big(P_{Y|A}\,\big\|\,\mathcal{N}(0,\mathsf{Var}(Y))\big)\big]\\&\quad\le\frac{1}{2}\mathbb{E}\big[\big|\tfrac{1}{n}\|U\|-\tfrac{1}{n}\mathbb{E}\big[\|U\|^2\big]\big|\big]\\&\qquad+C\cdot\left|\tfrac{1}{n}\|\mathsf{Cov}(U)\|_F\left(1+\tfrac{1}{n}\sqrt{\mathbb{E}[\|U\|^4]}\right)\right|^{\frac{2}{5}}.\end{aligned}$$

Combining Lemma 16 with Lemma 14 leads to the following result, which bounds the expected posterior non-Gaussianness in terms of the second order MI difference. The proof is given in Appendix C-E.

Lemma 17. *Under Assumptions 1 and 2, the expected posterior non-Gaussianness satisfies*

$$\mathbb{E}\big[\Delta^P_{m,n}\big]\le C_B\cdot\big|I''_{m,n}\big|^{\frac{1}{10}}.$$

D. Concentration of Posterior Variance

We now turn our attention to the second term on the right-hand side of (42). By the concavity of the logarithm, this term is nonnegative and measures the deviation of the posterior variance about its expectation.

We begin with the following result, which provides useful bounds on the deviation of the posterior variance. The proof is given in Appendix C-F.

Lemma 18. *Under Assumption 2, the posterior variance satisfies the following inequalities:*

$$\begin{aligned}\mathbb{E}\left[\log\left(\frac{1+M_{m,n}}{1+V_{m,n}}\right)\right]&\le\mathbb{E}[|V_{m,n}-M_{m,n}|]\\&\le C_B\cdot\sqrt{\inf_{t\in\mathbb{R}}\mathbb{E}\big[\tfrac{1}{2}\log(1+V_{m,n})-t\big]}.\end{aligned}\tag{43}$$

The next step is to bound the right-hand side of (43). Observe that by Lemma 15, the term $\frac{1}{2}\log(1+V_{m,n})$ can be expressed in terms of the posterior non-Gaussianness and the posterior MI difference. Accordingly, the main idea behind our approach is to show that the deviation of this term can be upper bounded in terms of deviation of the posterior MI difference.

Rather than working with the sequences $V_{m,n}$ and $J_{m,n}$ directly, however, we bound the averages of these terms corresponding to a sequence of ℓ measurements, where ℓ is an integer that that is chosen at the end of the proof to yield the tightest bounds. The main reason that we introduce this averaging is so that we can take advantage of the bound on the variance of the mutual information density given in Lemma 11.

The key technical results that we need are given below. Their proofs are given in Appendices C-G and C-H.

Lemma 19. *Under Assumptions 1 and 2, the posterior variance satisfies*

$$\frac{1}{\ell}\sum_{k=m}^{m+\ell-1}\mathbb{E}[|V_m-V_k|]\le C_B\cdot\big|I'_{m,n}-I'_{m+\ell-1,n}\big|^{\frac{1}{2}},$$

for all $(\ell,m,n)\in\mathbb{N}^3$.

Lemma 20. *Under Assumptions 1 and 2, the posterior MI difference satisfies*

$$\inf_{t\in\mathbb{R}}\mathbb{E}\left[\left|\frac{1}{\ell}\sum_{k=m}^{m+\ell-1}J_k-t\right|\right]\le C_B\cdot\left[\left(1+\frac{m}{n}\right)\frac{\sqrt{n}}{\ell}+\frac{1}{\sqrt{n}}\right],$$

for all $(\ell,m,n)\in\mathbb{N}^3$.

Finally, combining Lemmas 17, 19, and 20 leads to the following result, which bounds the deviation of the posterior variance in terms of the MI difference difference sequence. The proof is given in Appendix C-I.

Lemma 21. *Under Assumptions 1 and 2, the posterior variance satisfies,*

$$\begin{aligned}\mathbb{E}[|V_{m,n}-M_{m,n}|]\le C_B\cdot\Big[&\big|I'_{m,n}-I'_{m+\ell-1,n}\big|^{\frac{1}{4}}+\ell^{-\frac{1}{20}}\\&+\big(1+\tfrac{m}{n}\big)^{\frac{1}{2}}n^{\frac{1}{4}}\ell^{-\frac{1}{2}}+n^{-\frac{1}{4}}\Big].\end{aligned}$$

for all $(\ell,m,n)\in\mathbb{N}^3$.

E. Final Steps in Proof of Theorem 2

The following result is a straightforward consequence of Identity (42) and Lemmas 17 and 21. The proof is given in Appendix C-I.

Lemma 22. *Under Assumptions 1 and 2, the non-Gaussianness of new measurements satisfies the upper bound*

$$\begin{aligned}\Delta_{m,n}\le C_B\cdot\Big[&\big|I''_{m,n}\big|^{\frac{1}{10}}+\big|I'_{m,n}-I'_{m+\ell_n,n}\big|^{\frac{1}{4}}\\&+\big(1+\tfrac{m}{n}\big)^{\frac{1}{2}}n^{-\frac{1}{24}}\Big],\end{aligned}\tag{44}$$

where $\ell_n=\lceil n^{\frac{5}{6}}\rceil$.

We now show how the proof of Theorem 2 follows as a consequence of Identity (41) and Lemma 22. Fix any $n\in\mathbb{N}$ and $\delta\in\mathbb{R}_+$ and let $m=\lceil\delta n\rceil$ and $\ell=\lceil n^{\frac{5}{6}}\rceil$. Then, we can write

$$\begin{aligned}&\int_0^\delta\left|\mathcal{I}'_n(\gamma)-\frac{1}{2}\log(1+\mathcal{M}_n(\gamma))\right|\mathrm{d}\gamma\\&\quad\le\sum_{k=0}^{m-1}\int_{\frac{k}{n}}^{\frac{k+1}{n}}\left|\mathcal{I}'_n(\gamma)-\frac{1}{2}\log(1+\mathcal{M}_n(\gamma))\right|\mathrm{d}\gamma\\&\quad\overset{(a)}{=}\frac{1}{n}\sum_{k=0}^{m-1}\left|I'_{k,n}-\frac{1}{2}\log(1+M_{k,n})\right|\\&\quad\overset{(b)}{=}\frac{1}{n}\sum_{k=0}^{m-1}\Delta_{k,n}\\&\quad\overset{(c)}{\le}\frac{C_B}{n}\sum_{k=0}^{m-1}\Big[\big|I''_{k,n}\big|^{\frac{1}{10}}+\big|I'_{k+\ell,n}-I'_{k,n}\big|^{\frac{1}{4}}\\&\qquad\qquad+\big(1+\tfrac{k}{n}\big)^{\frac{1}{2}}n^{-\frac{1}{24}}\Big],\end{aligned}\tag{45}$$

where: (a) follows from the definitions of $\mathcal{I}'_n(\delta)$ and $\mathcal{M}_n(\delta)$; (b) follows from Identity (41); and (c) follows from Lemma 22. To further bound the right-hand side of (45), observe that

$$\frac{1}{n}\sum_{k=0}^{m-1}\big|I''_{k,n}\big|^{\frac{1}{10}}\overset{(a)}{\le}\frac{m^{\frac{9}{10}}}{n}\left(\sum_{k=0}^{m-1}\big|I''_{k,n}\big|\right)^{\frac{1}{10}}$$

$$\begin{aligned} &\overset{(b)}{=} \frac{m^{\frac{9}{10}}}{n} \left| I'_{1,n} - I'_{m,n} \right|^{\frac{1}{10}} \\ &\overset{(c)}{\le} \frac{m^{\frac{9}{10}}}{n} (I_{1,n})^{\frac{1}{10}} \\ &\overset{(d)}{\le} C_B \cdot (1+\delta)^{\frac{9}{10}} \, n^{-\frac{1}{10}}, \end{aligned} \tag{46}$$

where: (a) follows from Hölders inequality; (b) follows from the fact that $I''_{m,n}$ is non-positive; (c) follows from the fact that $I'_{m,n}$ is non-increasing in m; and (d) follows from the fact that $I_{1,n}$ is upper bounded by a constant that depends only on B. Along similar lines,

$$\begin{aligned} \frac{1}{n}\sum_{k=0}^{m-1} \left| I'_{k,n} - I'_{k+\ell,n} \right|^{\frac{1}{4}} &\overset{(a)}{\le} \frac{m^{\frac{3}{4}}}{n} \left(\sum_{k=0}^{m-1} \left| I'_{k,n} - I'_{k+\ell,n} \right| \right)^{\frac{1}{4}} \\ &\overset{(b)}{=} \frac{m^{\frac{3}{4}}}{n} \left(\sum_{k=0}^{\ell-1} (I'_{k,n} - I'_{k+m,n}) \right)^{\frac{1}{4}} \\ &\overset{(c)}{\le} \frac{m^{\frac{3}{4}}}{n} (\ell \cdot I_{1,n})^{\frac{1}{4}} \\ &\overset{(d)}{\le} C'_B \cdot (1+\delta)^{\frac{3}{4}} \, n^{-\frac{1}{24}}, \end{aligned} \tag{47}$$

where: (a) follows from Hölders inequality; (b) and (c) follow from the fact that $I'_{m,n}$ is non-increasing in m; and (d) follows from the fact that $I_{1,n}$ is upper bounded by a constant that depends only on B. Finally,

$$\begin{aligned} \frac{1}{n}\sum_{k=0}^{m-1} \left(1 + \tfrac{k}{n}\right)^{\frac{1}{2}} n^{-\frac{1}{24}} &\le \tfrac{m}{n}\left(1 + \tfrac{m-1}{n}\right)^{\frac{1}{2}} n^{-\frac{1}{24}} \\ &\le (1+\delta)^{\frac{3}{2}} \, n^{-\frac{1}{24}}. \end{aligned} \tag{48}$$

Plugging (46), (47), and (48) back into (45) and retaining only the dominant terms yields

$$\int_0^\delta \left| \mathcal{I}'_n(\gamma) - \frac{1}{2}\log(1+\mathcal{M}_n(\gamma)) \right| \mathrm{d}\gamma \le C_B \cdot (1+\delta)^{\frac{3}{2}} \, n^{-\frac{1}{24}}.$$

This completes the proof of Theorem 2.

V. Proof of Theorem 3

A. MMSE Fixed-Point Relationship

This section shows how the MMSE can be bounded in terms of a fixed-point equation defined by the single-letter MMSE function of the signal distribution. At a high level, our approach focuses on the MMSE of an augmented measurement model, which contains an extra measurement and extra signal entry, and shows that this augmented MMSE can be related to M_n in two different ways.

For a fixed signal length n and measurement number m, the augmented measurement model consists of the measurements (Y^m, A^m) plus an additional measurement given by

$$Z_{m+1} = Y_{m+1} + \sqrt{G_{m+1}}\, X_{n+1},$$

where $G_m \sim \frac{1}{n}\chi^2_m$ has a scaled chi-square distribution and is independent of everything else. The observed data is given by the tuple $(Y^m, A^m, \mathcal{D}_{m+1})$ where

$$\mathcal{D}_{m+1} = (Z_{m+1}, A_{m+1}, G_{m+1}).$$

The augmented MMSE $\widetilde{M}_{m,n}$ is defined to be the average MMSE of the first n signal entries given this data:

$$\widetilde{M}_{m,n} \triangleq \frac{1}{n}\mathsf{mmse}(X^n \mid Y^m, A^m, \mathcal{D}_{m+1}).$$

The augmented measurement Z_{m+1} is a noisy version of the measurement Y_{m+1}. Therefore, as far as estimation of the signal X^n is concerned, the augmented measurements are more informative than (Y^m, A^m), but less informative than (Y^{m+1}, A^{m+1}). An immediate consequence of the data processing inequality for MMSE [41, Proposition 4], is that the augmented MMSE is sandwiched between the MMSE sequence:

$$M_{m+1,n} \le \widetilde{M}_{m,n} \le M_{m,n}. \tag{49}$$

The following result follows immediately from (49) and the smoothness of the posterior variance given in Lemma 19. The proof is given in Appendix D-A.

Lemma 23. *Under Assumptions 1 and 2, the augmented MMSE $\widetilde{M}_{m,n}$ and the MMSE $M_{m,n}$ satisfy*

$$\left| \widetilde{M}_{m,n} - M_{m,n} \right| \le C_B \cdot \left| I''_{m,n} \right|^{\frac{1}{2}}. \tag{50}$$

The next step in the proof is to show that $\widetilde{M}_m$ can also be expressed in terms of the single-letter MMSE function $\mathsf{mmse}_X(s)$. The key property of the augmented measurement model that allows us to make this connection is given by the following result. The proof is given in Appendix D-B.

Lemma 24. *Under Assumptions 1 and 2, the augmented MMSE can be expressed equivalently in terms of the last signal entry:*

$$\widetilde{M}_{m,n} = \mathsf{mmse}(X_{n+1} \mid Y^m, A^m, \mathcal{D}_{m+1}). \tag{51}$$

To see why the characterization in (51) is useful note that the first m measurements (Y^m, A^m) are independent of X_{n+1}. Thus, as far as estimation of X_{n+1} is concerned, the relevant information provided by these measurements is summarized by the conditional distribution of Y_{m+1} given (Y^m, A^{m+1}). This observation allows us to leverage results from Section IV-C, which focused on the non-Gaussianness of this distribution. The proof of the following result is given in Section D-C.

Lemma 25. *Under Assumptions 1 and 2, the augmented MMSE and the MMSE satisfy*

$$\left| \widetilde{M}_{m,n} - \mathsf{mmse}_X\!\left(\frac{m/n}{1 + M_{m,n}} \right) \right| \le C_B \cdot \left(\sqrt{\Delta_{m,n}} + \frac{\sqrt{m}}{n} \right),$$

where $\Delta_{m,n}$ is the non-Gaussianness of new measurements.

B. Final Steps in Proof of Theorem 3

Fix any $n \in \mathbb{N}$ and $\delta \in \mathbb{R}_+$ and let $m = \lceil \delta n \rceil$. Then, we can write

$$\begin{aligned} &\int_0^\delta \left| \mathcal{M}_n(\gamma) - \mathsf{mmse}_X\!\left(\frac{\gamma}{1+\mathcal{M}_n(\gamma)} \right) \right| \mathrm{d}\gamma \\ &\quad \le \sum_{k=0}^{m-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} \left| \mathcal{M}_n(\gamma) - \mathsf{mmse}_X\!\left(\frac{\gamma}{1+\mathcal{M}_n(\gamma)} \right) \right| \mathrm{d}\gamma \end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{=} \sum_{k=0}^{m-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} \left| M_{k,n} - \mathsf{mmse}_X\left(\frac{\gamma}{1+M_{k,n}}\right)\right| \mathrm{d}\gamma \\
&\overset{(b)}{\le} \frac{1}{n} \sum_{k=0}^{m-1} \left(\left| M_{k,n} - \mathsf{mmse}_X\left(\frac{k/n}{1+M_{k,n}}\right)\right| + \frac{4B}{n}\right) \\
&\overset{(c)}{\le} \frac{C_B}{n} \sum_{k=0}^{m-1} \left(\left| I''_{k,n}\right|^{\frac{1}{2}} + |\Delta_{k,n}|^{\frac{1}{2}} + \frac{\sqrt{k}}{n} + \frac{1}{n} \right), \qquad (52)
\end{aligned}
$$

where: (a) follows from definition of $\mathcal{M}_n(\delta)$; (b) follows from the triangle inequality and Lemma 5; and (c) follows from Lemmas 23 and 25. To further bound the right-hand side of (52), observe that, by the same steps that let to Inequality (46),

$$\frac{1}{n} \sum_{k=0}^{m-1} \left| I''_{k,n}\right|^{\frac{1}{2}} \le C_B \cdot (1+\delta)^{\frac{1}{2}} n^{-\frac{1}{2}}. \qquad (53)$$

Furthermore,

$$
\begin{aligned}
\frac{1}{n} \sum_{k=0}^{m-1} |\Delta_{k,n}|^{\frac{1}{2}} &\overset{(a)}{\le} \sqrt{\frac{m}{n}} \sqrt{\frac{1}{n} \sum_{k=0}^{m-1} \Delta_{k,n}} \\
&\overset{(b)}{\le} C_B \cdot (1+\delta)^{\frac{1}{2}} \sqrt{(1+\delta)^{\frac{3}{2}}\, n^{-\frac{1}{24}}} \\
&= C_B \cdot (1+\delta)^{\frac{7}{4}}\, n^{-\frac{1}{48}} \qquad (54)
\end{aligned}
$$

where (a) follows from the Cauchy-Schwarz inequality, and (b) follows from the proof of Theorem 2. Finally,

$$\frac{1}{n} \sum_{k=0}^{m-1} \left(\frac{\sqrt{k}}{n} + \frac{1}{n}\right) \le \frac{m(1+\sqrt{m-1})}{n^2} \le (1+\delta)^{\frac{3}{2}} n^{-\frac{1}{2}}. \qquad (55)$$

Plugging (53), (54), and (55), back into (52) and keeping only the dominant terms leads to

$$
\begin{aligned}
&\int_0^\delta \left| \mathcal{M}_n(\gamma) - \mathsf{mmse}_X\left(\frac{\gamma}{1+\mathcal{M}_n(\gamma)}\right)\right| \mathrm{d}\gamma \\
&\qquad \le C_B \cdot (1+\delta)^{\frac{7}{4}}\, n^{-\frac{1}{48}}.
\end{aligned}
$$

This completes the proof of Theorem 3.

VI. Proof of Theorem 1

The proof of Theorem 1 is established by combining implications of the single-crossing property, the constraints on the MI and MMSE given in Theorems 2, 3, and 4, and standard results from functional analysis.

A. The Single-Crossing Property

The fixed point curve is given by the graph of the function

$$\delta_{\mathrm{FP}}(z) = (1+z)\mathsf{mmse}_X^{-1}(z),$$

where $\mathsf{mmse}_X^{-1}(z)$ is the functional inverse of $\mathsf{mmse}_X(s)$. The function $\delta_{\mathrm{FP}}(z)$ is continuously differentiable over its domain because $\mathsf{mmse}_X(s)$ is smooth on $(0,\infty)$ [37, Proposition 7].

The function $\delta_{\mathrm{RS}}(z)$ is defined to be the functional inverse of the replica-MMSE function

$$\delta_{\mathrm{RS}}(z) = \mathcal{M}_{\mathrm{RS}}^{-1}(z).$$

The function δ_{RS} is continuous and non-increasing because $\mathcal{M}_{\mathrm{RS}}$ is strictly decreasing. Note that jump discontinuities in $\mathcal{M}_{\mathrm{RS}}(\delta)$ correspond to flat sections in $\delta_{\mathrm{RS}}(z)$.

Using these definitions we can now provide a formal definition of the single-crossing property.

Definition 6 (Single-Crossing Property)**.** A signal distribution P_X has the single-crossing property if $\delta_{\mathrm{RS}} - \delta_{\mathrm{FP}}$ has at most one zero-crossing. In other words, there exists $z_* \in \mathbb{R}_+$ such that $\delta_{\mathrm{RS}} - \delta_{\mathrm{FP}}$ is nonpositive or nonnegative on $[0, z_*]$ and nonpositive or nonnegative on $[z_*, \infty)$.

Lemma 26. *If the signal distribution P_X has the single-crossing property there exists a point $(z_*, \delta_*) \in \mathrm{FP}$ such that*

$$\delta_{\mathrm{RS}}(z) = \begin{cases} \max(\delta_{\mathrm{FP}}(z), \delta_*), & \textit{if } z \in [0, z_*] \\ \min(\delta_{\mathrm{FP}}(z), \delta_*), & \textit{if } z \in [z_*, \infty). \end{cases}$$

Proof. If $\delta_{\mathrm{RS}}(z) = \delta_{\mathrm{FP}}(z)$ for all $z \in \mathbb{R}_+$ then this representation holds for every point in FP because δ_{RS} is non-increasing. Alternatively, if there exists $(u_*, \delta_*) \in \mathbb{R}_+^2$ such that $\delta_* = \delta_{\mathrm{RS}}(u_*) \neq \delta_{\mathrm{FP}}(u_*)$, then it must be the case that the global minimum of $Q_*(z) \triangleq R(\delta_*, z)$ is attained at more than one point. More precisely, there exists $z_1 < u_* < z_2$ such that

$$Q_*(z_1) = Q_*(z_2) = \min_z Q_*(z) \qquad (56)$$

$$Q_*(u) > \min_z Q_*(z) \quad \text{for some } u \in (z_1, z_2). \qquad (57)$$

To see why the second constraint follows from the assumption $\delta_{\mathrm{RS}}(u_*) \neq \delta_{\mathrm{FP}}(u_*)$, note that if $Q_*(z)$ were constant over the interval $[z_1, z_2]$, that would mean that $Q'_*(z) = R_z(\delta_*, z) = 0$ for all $z \in [z_1, z_2]$. This is equivalent to saying that every point on the line from (δ_*, z_1) to (δ_*, z_2) is on the fixed-point curve, which is a contradiction.

Now, since $\delta_{\mathrm{RS}}(z)$ is non-increasing and equal to δ_* at both z_1 and z_2 we know that $\delta_{\mathrm{RS}}(z) = \delta_*$ for all $z \in [z_1, z_2]$. Furthermore, since z_1 and z_2 are minimizers of $R(\delta_*, z)$, we also know that $\delta_{\mathrm{FP}}(z_1) = \delta_{\mathrm{FP}}(z_2) = \delta_*$.

Next, we will show that the the function $\delta_{\mathrm{FP}}(z) - \delta_*$ must have at least one negative-to-positive zero-crossing on (z_1, z_2). Recall that the function $Q_*(z)$ is continuous, has global minima at z_1 and z_2, and it not constant over $[z_1, z_2]$. Therefore, it must attain a local maximum on the open interval (z_1, z_2). Since it is continuously differentiable, this means that there exists $u_1, u_2 \in (z_1, z_2)$ with $u_1 < u_2$ such that $Q'_*(u_1) > 0$ and $Q'_*(u_2) < 0$. The sign changes in $Q_*(z)$ can be related to the sign changes in $\delta_{\mathrm{FP}}(z) - \delta_*$ be noting that

$$
\begin{aligned}
\mathrm{sgn}(Q'_*(z)) &\overset{(a)}{=} \mathrm{sgn}\left(z - \mathsf{mmse}_X\left(\frac{\delta_*}{1+z}\right)\right) \\
&\overset{(b)}{=} -\,\mathrm{sgn}\left(\mathsf{mmse}_X^{-1}(z) - \frac{\delta_*}{1+z}\right) \\
&= -\,\mathrm{sgn}(\delta_{\mathrm{FP}}(z) - \delta_*),
\end{aligned}
$$

where (a) follows from (22) the fact that δ_* can be taken to be strictly positive and (b) follows from the fact that $\mathsf{mmse}_X(s)$ is strictly decreasing. As a consequence, we see that $\delta_{\mathrm{FP}}(u_1) < \delta_*$ and $\delta_{\mathrm{FP}}(u_2) > \delta_*$, and thus $\delta_{\mathrm{FP}}(z) - \delta_*$ has at least one negative-to-positive zero-crossing on the interval (z_1, z_2).

At this point, we have shown that every tuple (δ_*, z_1, z_2) satisfying (56) and (57) leads to at least one negative-to-positive zero-crossing of $\delta_{\mathrm{FP}} - \delta_{\mathrm{RS}}$. Therefore, if the signal distribution has the single-crossing property, there can be at most one such tuple. This implies that $\delta_{\mathrm{FP}}(z) = \delta_{\mathrm{RS}}(z)$ for all $z \in [0, z_1] \cup [z_2, \infty)$. Furthermore, by the continuity of δ_{FP}, there exists a point $z_* \in (z_1, z_2)$ such that $\delta_{\mathrm{FP}}(z_*) = \delta_*$ and

$$\begin{aligned} z \le z_* &\implies \delta_{\mathrm{RS}}(z) \ge \delta_{\mathrm{FP}}(z) \\ z \ge z_* &\implies \delta_{\mathrm{RS}}(z) \le \delta_{\mathrm{FP}}(z). \end{aligned}$$

Combining these observations leads to the stated result. □

Next, for each $g \in \mathcal{V}$, we use $\delta_g(z) = g^{-1}(z)$ to denote the functional inverse. The function δ_g is continuous and non-increasing because g is strictly decreasing.

Lemma 27. *If the signal distribution P_X has the single-crossing property then, for every $g \in \mathcal{V}$, the function δ_g is either an upper bound or lower bound on δ_{RS}.*

Proof. Let (δ_*, z_*) be the point described in Lemma 26. Since δ_g is non-increasing, we have $\delta_g(z) \ge \delta_g(z_*)$ for all $z \in [0, z_*]$ and $\delta_g(z) \le \delta_g(z_*)$ for all $z \in [z_*, \infty)$. Combining these inequalities with the fact that δ_g is lower bounded by the lower envelope of the fixed-point curve leads to

$$\delta_g(z) \ge \begin{cases} \min\limits_{u \in [0,z]} \max(\delta_{\mathrm{FP}}(u), \delta_g(z_*)), & \text{if } z \in [0, z_*] \\ \min\limits_{u \in [z_*,z]} \min(\delta_{\mathrm{FP}}(u), \delta_g(z_*)), & \text{if } z \in [z_*, \infty). \end{cases}.$$

Therefore, if $\delta_g(z_*) \ge \delta_*$, we see that

$$\begin{aligned} \delta_g(z) &\ge \begin{cases} \min\limits_{u \in [0,z]} \max(\delta_{\mathrm{FP}}(u), \delta_*), & \text{if } z \in [0, z_*] \\ \min\limits_{u \in [z_*,z]} \min(\delta_{\mathrm{FP}}(u), \delta_*), & \text{if } z \in [z_*, \infty). \end{cases} \\ &\overset{(a)}{=} \begin{cases} \min\limits_{u \in [0,z]} \delta_{\mathrm{RS}}(u), & \text{if } z \in [0, z_*] \\ \min\limits_{u \in [z_*,z]} \delta_{\mathrm{RS}}(u), & \text{if } z \in [z_*, \infty). \end{cases} \\ &\overset{(b)}{=} \delta_{\mathrm{RS}}(z), \end{aligned}$$

where (a) follows from Lemma 26 and (b) follows from the fact that δ_{RS} is non-increasing. Alternatively, if $\delta_g(z_*) \le \delta_*$ then a similar argument can be used to show that $\delta_g(z) \le \delta_{\mathrm{RS}}(z)$ for all $z \in \mathbb{R}_+$. □

Lemma 28. *If the signal distribution P_X has the single-crossing property, then $\mathcal{G}$ is equal to the equivalence class of functions in $\mathcal{V}$ that are equal to $\mathcal{M}_{\mathrm{RS}}$ almost everywhere.*

Proof. Recall that $\mathcal{G}$ is the set of all functions $g \in \mathcal{V}$ that satisfy the boundary condition

$$\lim_{\delta \to \infty} \left| \int_0^\delta \frac{1}{2} \log(1 + g(\gamma)) \mathrm{d}\gamma - \mathcal{I}_{\mathrm{RS}}(\delta) \right| = 0. \qquad (58)$$

Furthermore, for each $g \in \mathcal{V}$ and $\delta \in \mathbb{R}_+$, we can write

$$\begin{aligned} &\left| \int_0^\delta \tfrac{1}{2} \log(1 + g(\gamma)) \mathrm{d}\gamma - \mathcal{I}_{\mathrm{RS}}(\delta) \right| \\ &\overset{(a)}{=} \left| \int_0^\delta \tfrac{1}{2} \log(1 + \mathcal{M}_{\mathrm{RS}}(\gamma)) \mathrm{d}\gamma - \int_0^\delta \tfrac{1}{2} \log(1 + g(\gamma)) \mathrm{d}\gamma \right| \\ &\overset{(b)}{=} \int_0^\delta \left| \tfrac{1}{2} \log(1 + \mathcal{M}_{\mathrm{RS}}(\gamma)) - \tfrac{1}{2} \log(1 + g(\gamma)) \right| \mathrm{d}\gamma, \qquad (59) \end{aligned}$$

where (a) follows from (6) and (b) follows from the monotonicity of g and $\mathcal{M}_{\mathrm{RS}}$ and Lemma 27. Combining (58) and (59), we see that, for all $g \in \mathcal{G}$,

$$\int_0^\infty \left| \frac{1}{2} \log(1 + \mathcal{M}_{\mathrm{RS}}(\gamma)) - \frac{1}{2} \log(1 + g(\gamma)) \right| \mathrm{d}\gamma = 0,$$

and thus $\mathcal{M}_{\mathrm{RS}}$ and g are equal almost everywhere. □

B. Convergence of Subsequences

For each $n \in \mathbb{N}$, the function $\mathcal{M}_n$ is a non-increasing function from $\mathbb{R}_+$ to $\mathbb{R}_+$. Convergence of the sequence $\mathcal{M}_n$ can be treated in a few different ways. In our original approach [33], we focused on the Lévy metric [44, Ch. 2]. Here, we present a more direct argument based on the Helly Selection Theorem [45, Thm. 12].

First, we let $L^1([0, S])$ represent the standard Banach space of Lebesgue integrable functions from $[0, S]$ to $\mathbb{R}$ with norm

$$\int_0^S |f(\delta)| \mathrm{d}\delta.$$

In this space, two functions f, g are called *equivalent* if they are equal almost-everywhere (i.e., $\int_0^S |f(\delta) - g(\delta)| \mathrm{d}\delta = 0$). Next, we recall that monotone functions are continuous almost everywhere (e.g., except for a countable set of jump discontinuities) [46]. Thus, f, g are equivalent if and only if they are equal at all points of continuity.

The following lemmas outline our approach to convergence.

Lemma 29. *Under Assumptions 1 and 2, for any $S > 0$ and any subsequence of $(\mathcal{M}_n, \mathcal{I}_n)$, there is a further subsequence (whose index is denoted by n') and some $g \in \mathcal{G}$ such that*

$$\begin{aligned} &\lim_{n'} \int_0^S \left| \mathcal{M}_{n'}(\delta) - g(\delta) \right| \mathrm{d}\delta = 0 \\ &\lim_{n'} \int_0^S \left| \mathcal{I}'_{n'}(\delta) - \frac{1}{2} \log(1 + g(\delta)) \right| \mathrm{d}\delta = 0. \end{aligned}$$

Proof. For any $S > 0$ and each $n \in \mathbb{N}$, the restriction of $\mathcal{M}_n(\delta)$ to $\delta \in [0, S]$ is non-increasing and uniformly bounded by $\mathcal{M}_n(0) = \mathsf{Var}(X)$. Since $\mathcal{M}_n(\delta)$ is nonnegative and non-increasing, its total variation on $[0, T]$ equals $\mathcal{M}_n(0) - \mathcal{M}_n(S) \le \mathsf{Var}(X)$ [47, Section 6.3].

Based on this, the Helly Selection Theorem [45, Thm. 12] shows that any subsequence of $\mathcal{M}_n$ contains a further subsequence that converges in $L^1([0, S])$. Let $\mathcal{M}_{n'}$ denote this further subsequence and $\mathcal{M}_*$ denote its limit so that

$$\lim_{n'} \int_0^S |\mathcal{M}_{n'}(\delta) - \mathcal{M}_*(\delta)| \mathrm{d}\delta = 0.$$

To simplify notation, we define the operator $T \colon L^1([0, S]) \to L^1([0, S])$ via $(Tf)(\delta) \mapsto \mathsf{mmse}_X(\delta/(1 + f(\delta)))$. To analyze $\mathcal{M}_*(\delta)$, we observe that, for all n, one has

$$\begin{aligned} \int_0^S \left| \mathcal{M}_*(\delta) - T\mathcal{M}_*(\delta) \right| \mathrm{d}\delta &\le \int_0^S \left| \mathcal{M}_*(\delta) - \mathcal{M}_n(\delta) \right| \mathrm{d}\delta \\ &+ \int_0^S \left| \mathcal{M}_n(\delta) - T\mathcal{M}_n(\delta) + T\mathcal{M}_n(\delta) - T\mathcal{M}_*(\delta) \right| \mathrm{d}\delta \end{aligned}$$

$$\leq (1+L_T)\int_0^S \big|\mathcal{M}_*(\delta)-\mathcal{M}_n(\delta)\big|\mathrm{d}\delta$$
$$+\int_0^S \big|\mathcal{M}_n(\delta)-T\mathcal{M}_n(\delta)\big|\mathrm{d}\delta,$$

where L_T is the Lipschitz constant of T. Under Assumption 2, one can use Lemma 5, to show that $L_T \leq 4BS$. Since $\int_0^S \big|\mathcal{M}_*(\delta)-\mathcal{M}_{n'}(\delta)\big|\mathrm{d}\delta \to 0$ by construction and $\int_0^S \big|\mathcal{M}_{n'}(\delta)-T\mathcal{M}_{n'}(\delta)\big|\mathrm{d}\delta \to 0$ by Theorem 3, taking the limit along this subsequnce shows that $\mathcal{M}_*$ equals $T\mathcal{M}_*$ almost everywhere on $[0,S]$. As S was arbitrary, we see that $\mathcal{M}_*$ satisfies the first condition of Definition 3.

To establish the second condition of Definition 3, we focus on the sequence $\mathcal{I}_n$. Recall that each $\mathcal{I}_n$ is concave and differentiable, with derivative $\mathcal{I}_n'$. Also, the set $\{\mathcal{I}_n\}$ is uniformly bounded on $[0,S]$ and uniformly Lipschitz by (9), (17), and (18). By the Arzelà-Ascoli theorem [47, Section 10.1], this implies that any subsequence of $\mathcal{I}_n$ contains a further subsequence that converges uniformly on $[0,S]$. Moreover, the limiting function is concave and the further subsequence of derivatives also converges to the derivative of the limit function at each point where it is differentiable [48, Corollary 1.3.8].

Thus, from any subsequence of $(\mathcal{M}_n,\mathcal{I}_n)$, we can choose a further subsequence (whose index is denoted by n') such that $\int_0^S \big|\mathcal{M}_{n'}(\delta)-\mathcal{M}_*(\delta)\big|\mathrm{d}\delta \to 0$ and $\mathcal{I}_{n'}$ converges uniformly on $[0,S]$ to a concave limit function $\mathcal{I}_*$. Moreover, the sequence of derivatives $\mathcal{I}_{n'}'$ also converges to $\mathcal{I}_*'$ at each point where $\mathcal{I}_*$ is differentiable. Since $\mathcal{I}$ is concave, it is differentiable almost everywhere and we have

$$\lim_{n'} \mathcal{I}_{n'}'(\delta) = \mathcal{I}_*'(\delta)$$

almost everywhere on $[0,S]$. Since $|\mathcal{I}_{n'}'(\delta)-\mathcal{I}_*'(\delta)|$ is bounded and converges to zero almost everywhere on $[0,S]$, we can apply the dominated convergence theorem to see also that $\int_0^S \big|\mathcal{I}_{n'}'(\delta)-\mathcal{I}_*'(\delta)\big|\mathrm{d}\delta \to 0$. Next, we can apply Theorem 2 to see that

$$\lim_{n'}\int_0^S\left|\mathcal{I}_{n'}'(\delta)-\frac{1}{2}\log(1+\mathcal{M}_{n'}(\delta))\right|\mathrm{d}\delta = 0.$$

Since $\frac{1}{2}\log(1+z)$ is Lipschitz in z, one finds that

$$\int_0^S\left|\frac{1}{2}\log(1+\mathcal{M}_{n'}(\delta))-\frac{1}{2}\log(1+\mathcal{M}_*(\delta))\right|\mathrm{d}\delta \to 0$$

follows from the fact that $\int_0^S \big|\mathcal{M}_{n'}(\delta)-\mathcal{M}_*(\delta)\big|\mathrm{d}\delta \to 0$. Along with the triangle inequality, this shows that $\mathcal{I}_{n'}'(\delta)$ converges to $\frac{1}{2}\log(1+\mathcal{M}_*(\delta))$ almost everywhere on $[0,S]$. Since $\big|\mathcal{I}_{n'}'(\delta)-\frac{1}{2}\log(1+\mathcal{M}_*(\delta))\big|$ is bounded and converges to zero almost everywhere on $[0,S]$, we can apply the dominated convergence theorem to see that

$$\mathcal{I}_*(S) = \lim_{n'}\int_0^S \mathcal{I}_{n'}'(\delta)\mathrm{d}\delta = \int_0^S \frac{1}{2}\log(1+\mathcal{M}_*(\delta))\mathrm{d}\delta. \quad (60)$$

Next, we observe that Theorem 4 implies

$$\lim_{n'}|\mathcal{I}_{\mathrm{RS}}(S)-\mathcal{I}_{n'}(S)| \leq \frac{C}{\sqrt{S}},$$

for all $S \geq 4$. With (60), this implies that

$$\left|\mathcal{I}_{\mathrm{RS}}(S)-\int_0^S\frac{1}{2}\log\big(1+\mathcal{M}_*(\delta)\big)\mathrm{d}\delta\right|$$
$$\leq |\mathcal{I}_{\mathrm{RS}}(S)-\mathcal{I}_{n'}(S)| + \left|\mathcal{I}_{n'}(S)-\int_0^S\frac{1}{2}\log\big(1+\mathcal{M}_*(\delta)\big)\mathrm{d}\delta\right|$$
$$\leq \frac{C}{\sqrt{S}} + \epsilon_{n'},$$

where $\lim_{n'}\epsilon_{n'}=0$. Taking the limit $n'\to\infty$ followed by the limit $S\to\infty$, we see

$$\lim_{S\to\infty}\left|\mathcal{I}_{\mathrm{RS}}(S)-\int_0^S\frac{1}{2}\log\big(1+\mathcal{M}_*(\delta)\big)\mathrm{d}\delta\right| = 0$$

and, thus, that $\mathcal{M}_*\in\mathcal{G}$. Notice that we focus first on finite $S\in\mathbb{R}_+$ and then take the limit $S\to\infty$. This is valid because the functions $\mathcal{I}(\delta)$ and $\mathcal{M}_n(\delta)$ are defined for all $\delta\in\mathbb{R}_+$ but restricted to $[0,S]$ for the convergence proof. □

Now, we can complete the proof of Theorem 1. The key idea is to combine Lemma 28 with Lemma 29. From these two results, it follows that, for any $S>0$, every subsequence of $\mathcal{M}_n(\delta)$ has a further subsequence that converges to $\mathcal{M}_{\mathrm{RS}}(\delta)$. This holds because the further subsequence must converge to some function in $\mathcal{G}$ (by Lemma 29) but there is only one function up to almost everywhere equivalence (by Lemma 28).

The final step is to realize that this is sufficient to prove that, for all $S>0$, we have

$$\lim_{n\to\infty}\int_0^S|\mathcal{M}_{\mathrm{RS}}(\delta)-\mathcal{M}_n(\delta)|\mathrm{d}\delta = 0.$$

To see this, suppose that $\mathcal{M}_n(\delta)$ does not converge to $\mathcal{M}_{\mathrm{RS}}(\delta)$ in $L^1([0,S])$. In this case, there is an $\epsilon>0$ and an infinite subsequence $n(i)$ such that

$$\int_0^S\big|\mathcal{M}_{\mathrm{RS}}(\delta)-\mathcal{M}_{n(i)}(\delta)\big|\mathrm{d}\delta > \epsilon$$

for all $i\in\mathbb{N}$. But, applying Lemma 29 shows that $n(i)$ has a further subsequence (denoted by n') such that

$$\lim_{n'}\int_0^S|\mathcal{M}_{\mathrm{RS}}(\delta)-\mathcal{M}_{n'}(\delta)|\mathrm{d}\delta = 0.$$

From this contradiction, one must conclude that $\mathcal{M}_n(\delta)$ converges to $\mathcal{M}_{\mathrm{RS}}(\delta)$ in $L^1([0,S])$ for any $S>0$.

VII. Conclusion and Future Directions

In this paper, we resolve a well-known open problem by presenting a rigorous derivation of the fundamental limits of random linear estimation for i.i.d. signal distributions and i.i.d. Gaussian measurement matrices. We show that the limiting MI and MMSE are equal to the values predicted by the replica method from statistical physics.

The assumption in our proof that the signal entries are i.i.d. is made to simplify the statement of the results and the analysis. With minor modifications, the proof technique can be extended to the setting where the entries in X^n can be partitioned into blocks of size L such that each block is drawn i.i.d. according to some distribution on $\mathbb{R}^L$ with finite-fourth moment. For $L = 2$, this includes the case where the odd/even entries of the signal are drawn independently from different distributions and also the case where they are drawn jointly from a single distribution on $\mathbb{R}^2$. The primary change to the replica formulas is that the scalar mutual information and MMSE functions are replaced by L-block mutual information and MMSE functions, which again map $\mathbb{R}_+$ to $\mathbb{R}_+$. In particular, the modified MMSE fixed-point equation is defined with respect to the MMSE of a length-L block observed in isotropic Gaussian noise.

The assumption that the entries of the matrix A^m are i.i.d. Gaussian plays an important role in our proof technique. In particular, this allows one to establish concentration of the mutual information density. It is possible that our results can be extended to a larger class of random matrix distributions, such as those with i.i.d. sub-Gaussian entries following along the same lines as Korada and Macris [28, Theorem 4]. More generally, it would be interesting to see if the ideas used in the paper can also be used to prove the conjectured formulas obtained by Tulino et al. [8] for the setting of orthogonally invariant random matrix distributions.

An interesting direction for future work is to see if one can obtain rigorous proofs for the setting of mismatched estimation. Such a result could then be used characterize the behavior of a large class of estimators that can be viewed as conditional expectations with respect to some prior. Although elements of our approach can be applied in this setting, the fact that replica-symmetry breaking may be needed for low-temperature limits (see Bereyhi et al. [20], [21]) suggests that a full characterization of the problem will require some new ideas. The setting of mismatched estimation can also be approached using ideas from statistical decision theory [49], where the goal is characterize the minimax performance over a class of distributions [6], [50]–[53].

Appendix A
Useful Results

A. Basic Inequalities

We begin by reviewing a number of basic inequalities. For numbers $x_1, \cdots x_n$ and $p \geq 1$, Jensen's inequality combined with the convexity of $|\cdot|^p$ yields,

$$\left|\sum_{i=1}^n x_i\right|^p \leq n^{p-1}\sum_{i=1}^n |x_i|^p, \qquad p \geq 1. \tag{61}$$

In the special case $n = 2$ and $p \in \{2, 4\}$, we obtain

$$(a+b)^2 \leq 2(a^2+b^2) \tag{62}$$
$$(a+b)^4 \leq 8(a^4+b^4). \tag{63}$$

For random variables $X_1, \cdots, X_n$ and $p \geq 1$, a consequence of Minkowski's inequality [54, Theorem 2.16] is that

$$\mathbb{E}\left[\left|\sum_{i=1}^n X_i\right|^p\right] \leq \left(\sum_{i=1}^n (\mathbb{E}[|X_i|^p])^{\frac{1}{p}}\right)^p, \qquad p \geq 1. \tag{64}$$

Also, for random variables X and Y, an immediate consequence of Jensen's inequality is that expectation of the absolute difference $|X - Y|$ can be upper bounded in terms of higher moments, i.e.,

$$\mathbb{E}[|X-Y|] \leq |\mathbb{E}[|X-Y|^p]|^{\frac{1}{p}}, \qquad p \geq 1.$$

Sometimes, we need to bound the difference in terms of weaker measures of deviation between X and Y. The following Lemma provides two such bounds that also depend on the moments of X and Y.

Lemma 30. *For nonnegative random variables X and Y, the expectation of the absolute difference $|X - Y|$ obeys the following upper bounds:*

$$\mathbb{E}[|X-Y|] \leq \sqrt{\frac{1}{2}(\mathbb{E}[X^2]+\mathbb{E}[Y^2])\mathbb{E}[|\log(X/Y)|]} \tag{65}$$
$$\mathbb{E}[|X-Y|] \leq \sqrt{2(\mathbb{E}[X]+\mathbb{E}[Y])\mathbb{E}\left[\left|\sqrt{X}-\sqrt{Y}\right|\right]}. \tag{66}$$

Proof. We begin with (65). For any numbers $0 < x < y$, the difference $y - x$ can be upper bounded as follows:

$$\begin{aligned} y - x &= \int_x^y \sqrt{u}\frac{1}{\sqrt{u}}du \\ &\leq \sqrt{\int_x^y u\, du}\sqrt{\int_x^y \frac{1}{u}du} \\ &= \sqrt{\frac{1}{2}(y^2-x^2)}\sqrt{\log(y/x)} \\ &\leq \sqrt{\frac{1}{2}(y^2+x^2)}\sqrt{\log(y/x)}, \end{aligned}$$

where the first inequality is due to the Cauchy-Schwarz inequality. Thus, the absolute difference between X and Y obeys

$$|X-Y| \leq \sqrt{\frac{1}{2}(X^2+Y^2)}\sqrt{|\log(X/Y)|}.$$

Taking the expectation of both sides and using the Cauchy-Schwarz inequality leads to (65).

To prove (66), observe that the difference between between X and Y can be decomposed as

$$X - Y = (\sqrt{X}+\sqrt{Y})(\sqrt{X}-\sqrt{Y}).$$

Thus, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}[|X-Y|] &= \sqrt{\mathbb{E}\left[(\sqrt{X}+\sqrt{Y})^2\right]}\sqrt{\mathbb{E}\left[(\sqrt{X}-\sqrt{Y})^2\right]} \\ &\leq \sqrt{2(\mathbb{E}[X]+\mathbb{E}[Y])}\sqrt{\mathbb{E}\left[(\sqrt{X}-\sqrt{Y})^2\right]}, \end{aligned}$$

where the last step is due to (62). □

B. Variance Decompositions

This section reviews some useful decompositions and bounds on the variance. As a starting point, observe that the variance of a random variable X can be expressed in terms of an independent copy X' according to

$$\mathsf{Var}(X) = \frac{1}{2}\mathbb{E}\left[(X-X')^2\right].$$

This representation can extended to conditional variance, by letting X_y and X'_y denote independent draws from the conditional distribution $P_{X|Y=y}$, so that

$$\mathsf{Var}(X \,|\, Y = y) = \frac{1}{2}\mathbb{E}\big[(X_y - X'_y)^2\big].$$

For a random draw of Y, it then follows that the random conditional variance of X given Y can be expressed as

$$\mathsf{Var}(X \,|\, Y) = \frac{1}{2}\mathbb{E}\big[(X_Y - X'_Y)^2 \,|\, Y\big], \tag{67}$$

where X_Y and X'_Y are conditionally independent draws from the random conditional distribution $P_{X|Y}(\cdot|Y)$.

Using this representation, the moments of the conditional variance can be bounded straightforwardly. For all $p \geq 1$,

$$\begin{aligned}\mathbb{E}[|\mathsf{Var}(X|Y)|^p] &= \frac{1}{2^p}\mathbb{E}\Big[\big|\mathbb{E}\big[(X_Y - X'_Y)^2 \mid Y\big]\big|^p\Big] \\ &\overset{(a)}{\leq} \frac{1}{2^p}\mathbb{E}\Big[|X_Y - X'_Y|^{2p}\Big] \\ &\overset{(b)}{\leq} 2^{p-1}\Big(\mathbb{E}\Big[|X_Y|^{2p}\Big] + \mathbb{E}\Big[|X'_Y|^{2p}\Big]\Big) \\ &\overset{(c)}{=} 2^p\mathbb{E}\Big[|X|^{2p}\Big],\end{aligned}$$

where (a) follows from Jensen's inequality and the convexity of $|\cdot|^p$, (b) follows from (61), and (c) follows from the fact that X_Y and X'_Y both have the same distribution as X.

The law of total variance gives

$$\mathsf{Var}(X) = \mathbb{E}[\mathsf{Var}(X \,|\, Y)] + \mathsf{Var}(\mathbb{E}[X \,|\, Y]). \tag{68}$$

As an immediate consequence, we obtain the data processing inequality for MMSE (see e.g. [41, Proposition 4]) , which states that conditioning cannot increase the MMSE on average. In particular, if $X \to Y \to Z$ form a Markov chain, then,

$$\mathsf{mmse}(X \,|\, Y) \leq \mathsf{mmse}(X \,|\, Z).$$

C. Bounds using KL Divergence

This section provides a number results that that allow us to bound differences in expectations in terms of Kullback–Leibler divergence. One of the consequences of Lemma 31 (given below) is that random variables $X \sim P_X$ and $Y \sim P_Y$ with positive and finite second moments satisfy

$$\frac{|\mathbb{E}[X] - \mathbb{E}[Y]|}{\sqrt{2\mathbb{E}[X^2] + 2\mathbb{E}[Y^2]}} \leq \sqrt{D_{\mathrm{KL}}(P_X \,\|\, P_Y)}.$$

We begin by reviewing some basic definitions (see e.g., [55, Section 3.3]). Let P and Q be probability measures with densities p and q with respect to a dominating measure λ. The Hellinger distance $d_H(P, Q)$ is defined as the $\mathcal{L}_2$ distance between the square roots of the densities $\sqrt{p}$ and $\sqrt{q}$, and the squared Helliger distance is given by

$$d_H^2(P, Q) = \int (\sqrt{p} - \sqrt{q})^2 \mathrm{d}\lambda.$$

The Kullback–Leibler divergence (also known as relative entropy) is defined as

$$D_{\mathrm{KL}}(P \,\|\, Q) = \int p \log\Big(\frac{p}{q}\Big)\mathrm{d}\lambda.$$

The squared Hellinger distance is upper bounded by the KL divergence [55, pg. 62],

$$d_H^2(P, Q) \leq D_{\mathrm{KL}}(P \,\|\, Q). \tag{69}$$

Lemma 31. *Let f be a function that is measurable with respect to P and Q. Then*

$$\left|\int f(\mathrm{d}P - \mathrm{d}Q)\right| \leq \sqrt{2\int f^2(\mathrm{d}P + \mathrm{d}Q)}\,\min\Big\{\sqrt{D_{\mathrm{KL}}(P \,\|\, Q)}, 1\Big\}.$$

Proof. Let p and q be the densities of P and Q with respect to a dominating measures λ. Then, we can write

$$\begin{aligned}\left|\int f(\mathrm{d}P - \mathrm{d}Q)\right| &= \left|\int f(p - q)\mathrm{d}\lambda\right| \\ &\overset{(a)}{=} \left|\int f(\sqrt{p} + \sqrt{q})(\sqrt{p} - \sqrt{q})\mathrm{d}\lambda\right| \\ &\overset{(b)}{\leq} \sqrt{\int f^2(\sqrt{p} + \sqrt{q})^2\mathrm{d}\lambda\, d_H^2(P, Q)} \\ &\overset{(c)}{\leq} \sqrt{\int f^2(\sqrt{p} + \sqrt{q})^2\mathrm{d}\lambda\, D_{\mathrm{KL}}(P \,\|\, Q)} \\ &\overset{(d)}{\leq} \sqrt{2\int f^2(p + q)\mathrm{d}\lambda\, D_{\mathrm{KL}}(P \,\|\, Q)},\end{aligned} \tag{70}$$

where (a) is justified by the non-negativity of the densities, (b) follows from the Cauchy-Schwarz inequality, (c) follows from (69), and (d) follows from (62).

Alternatively, we also have the upper bound

$$\begin{aligned}\left|\int f(\mathrm{d}P - \mathrm{d}Q)\right| &\overset{(a)}{\leq} \left|\int f\mathrm{d}P\right| + \left|\int f\mathrm{d}Q\right| \\ &\overset{(b)}{\leq} \sqrt{\int f^2\mathrm{d}P} + \sqrt{\int f^2\mathrm{d}Q} \\ &\overset{(c)}{\leq} \sqrt{2\int f^2(\mathrm{d}P + \mathrm{d}Q)},\end{aligned} \tag{71}$$

where (a) follows from the triangle inequality, (b) follows from Jensen's inequality, and (c) follows from (62). Taking the minimum of (70) and (71) leads to the stated result. □

Lemma 32. *For any distribution $P_{X,Y,Z}$ and $p \geq 1$,*

$$\begin{aligned}&\mathbb{E}[|\mathsf{Var}(X \,|\, Y) - \mathsf{Var}(X \,|\, Z)|^p] \\ &\qquad \leq 2^{2p+\frac{1}{2}}\sqrt{\mathbb{E}[|X|^{4p}]\,\mathbb{E}\big[D_{\mathrm{KL}}\big(P_{X|Y} \,\big\|\, P_{X|Z}\big)\big]}.\end{aligned}$$

Proof. Let P and Q be the random probability measures on $\mathbb{R}^2$ defined by

$$\begin{aligned}P &= P_{X|Y} \times P_{X|Y} \\ Q &= P_{X|Z} \times P_{X|Z},\end{aligned}$$

and let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$. Then, by the variance decomposition (67), we can write

$$\mathsf{Var}(X \,|\, Y) = \int f\mathrm{d}P, \quad \mathsf{Var}(X \,|\, Z) = \int f\mathrm{d}Q.$$

Furthermore, by the upper bound $f^2(x_1,x_2) \le 2(x_1^4 + x_2^4)$, the expectation of the f^2 satisfies

$$\int f^2(\mathrm{d}P + \mathrm{d}Q) \le 4\mathbb{E}\big[X^4 \,|\, Y\big] + 4\mathbb{E}\big[X^4 \,|\, Z\big]. \tag{72}$$

Therefore, by Lemma 31, the difference between the conditional variances satisfies

$$\begin{aligned} &|\mathsf{Var}(X \,|\, Y) - \mathsf{Var}(X \,|\, Z)| \\ &\quad \le \sqrt{2\int f^2(\mathrm{d}P + \mathrm{d}Q)}\sqrt{\min\{D_{\mathrm{KL}}(P \,\|\, Q), 1\}} \\ &\quad \le \sqrt{8\mathbb{E}[X^4 \,|\, Y] + 8\mathbb{E}[X^4 \,|\, Z]}\sqrt{\min\{D_{\mathrm{KL}}(P \,\|\, Q), 1\}}, \end{aligned} \tag{73}$$

where the second inequality follows from (72).

The next step is to bound the expected p-th power of the right-hand side of (73). Starting with the Cauchy-Schwarz inequality, we have

$$\begin{aligned} &\mathbb{E}[|\mathsf{Var}(X \mid Y) - \mathsf{Var}(X \mid Z)|^p] \\ &\quad \le \sqrt{\mathbb{E}[|8\mathbb{E}[X^4 \,|\, Y] + 8\mathbb{E}[X^4 \,|\, Z]|^p]} \\ &\quad\quad \times \sqrt{\mathbb{E}[|\min\{D_{\mathrm{KL}}(P \,\|\, Q), 1\}|^p]}. \end{aligned} \tag{74}$$

For the first term on the right-hand side of (74), observe that, by Jensen's inequality,

$$\begin{aligned} \sqrt{\mathbb{E}[|8\mathbb{E}[X^4 \,|\, Y] + 8\mathbb{E}[X^4 \,|\, Z]|^p]} &\le \sqrt{8^p\mathbb{E}[X^{4p}] + 8^p\mathbb{E}[X^{4p}]} \\ &= 4^p\sqrt{\mathbb{E}[X^{4p}]}. \end{aligned} \tag{75}$$

Meanwhile, the expectation in the second term on the right-hand side of (74) satisfies

$$\begin{aligned} \mathbb{E}[|\min\{D_{\mathrm{KL}}(P \,\|\, Q), 1\}|^p] &\le \mathbb{E}[D_{\mathrm{KL}}(P \,\|\, Q)] \\ &= 2\mathbb{E}\big[D_{\mathrm{KL}}\big(P_{X|Y} \,\big\|\, P_{X|Z}\big)\big], \end{aligned} \tag{76}$$

where the second step follows from the definition of P and Q. Plugging (75) and (76) back into (74) leads to the stated result. □

Lemma 33. *For any distribution $P_{X,Y,Z}$ and $p \ge 1$,*

$$\begin{aligned} &\mathbb{E}[|\mathsf{Cov}(X, Y \,|\, Z)|^p] \\ &\quad \le 2^p\sqrt{\mathbb{E}\Big[|\mathbb{E}[X^4 \,|\, Z]\mathbb{E}[Y^4 \,|\, Z]|^{\frac{p}{2}}\Big] I(X; Y \,|\, Z)}. \end{aligned}$$

Proof. Let P and Q be the random probability measures on $\mathbb{R}^2$ defined by

$$\begin{aligned} P &= P_{X,Y|Z} \\ Q &= P_{X|Z} \times P_{Y|Z}, \end{aligned}$$

and let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = xy$. Then, the conditional covariance between X and Y can be expressed as

$$\mathsf{Cov}(X, Y \,|\, Z) = \int f(\mathrm{d}P - \mathrm{d}Q).$$

Furthermore

$$\begin{aligned} \int f^2(\mathrm{d}P + \mathrm{d}Q) &= \mathbb{E}\big[|XY|^2 \mid Z\big] + \mathbb{E}\big[X^2 \,|\, Z\big]\mathbb{E}\big[Y^2 \,|\, Z\big] \\ &\le 2\sqrt{\mathbb{E}[X^4 \,|\, Z]\mathbb{E}[Y^4 \,|\, Z]}, \end{aligned}$$

where the second step follows from the Cauchy-Schwarz inequality and Jensen's inequality. Therefore, by Lemma 31, the magnitude of the covariance satisfies

$$\begin{aligned} &|\mathsf{Cov}(X, Y \,|\, Z)| \\ &\quad \le \sqrt{2\int f^2(\mathrm{d}P + \mathrm{d}Q)}\sqrt{\min\{D_{\mathrm{KL}}(P \,\|\, Q), 1\}} \\ &\quad \le 2\big|\mathbb{E}\big[X^4 \,|\, Z\big]\mathbb{E}\big[Y^4 \,|\, Z\big]\big|^{\frac{1}{4}}|\min\{D_{\mathrm{KL}}(P \,\|\, Q), 1\}|^{\frac{1}{2}}. \end{aligned} \tag{77}$$

The next step is to bound the expected p-th power of the right-hand side of (77). Starting with the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \mathbb{E}[|\mathsf{Cov}(X, Y \mid Z)|^p] &\le 2^p\sqrt{\mathbb{E}\Big[|\mathbb{E}[X^4 \,|\, Z]\mathbb{E}[Y^4 \,|\, Z]|^{\frac{p}{2}}\Big]} \\ &\quad \times \sqrt{\mathbb{E}[|\min\{D_{\mathrm{KL}}(P \,\|\, Q), 1\}|^p]}. \end{aligned} \tag{78}$$

Note that the expectation in the second term on the right-hand side of (78) satisfies

$$\begin{aligned} \mathbb{E}[|\min\{D_{\mathrm{KL}}(P \,\|\, Q), 1\}|^p] &\le \mathbb{E}[D_{\mathrm{KL}}(P \,\|\, Q)] \\ &= I(X; Y \,|\, Z), \end{aligned} \tag{79}$$

where the second step follows from the definition of P and Q. Plugging (79) back into (78) leads to the stated result. □

Lemma 34. *For any distributions $P_{X,Y}$ and $P_{X,Z}$,*

$$\begin{aligned} &|\mathsf{mmse}(X \,|\, Y) - \mathsf{mmse}(X \,|\, Z)| \\ &\quad \le 2^{\frac{5}{2}}\sqrt{\mathbb{E}[|X|^4]\, D_{\mathrm{KL}}(P_{X,Y} \,\|\, P_{X,Z})}. \end{aligned}$$

Proof. Let P and Q be the distributions given by

$$\begin{aligned} P(x_1, x_2, y) &= P_{X|Y}(x_1 \mid y)P_{X|Y}(x_2 \mid y)P_Y(y) \\ Q(x_1, x_2, y) &= P_{X|Z}(x_1 \mid y)P_{X|Z}(x_2 \mid y)P_Z(y) \end{aligned}$$

Then,

$$\begin{aligned} \mathsf{mmse}(X \,|\, Y) &= \frac{1}{2}\int (x_1 - x_2)^2 \mathrm{d}P(x_1, x_2, y) \\ \mathsf{mmse}(X \,|\, Z) &= \frac{1}{2}\int (x_1 - x_2)^2 \mathrm{d}Q(x_1, x_2, y) \end{aligned}$$

and so, by Lemma 31,

$$\begin{aligned} &|\mathsf{mmse}(X \,|\, Y) - \mathsf{mmse}(X \,|\, Z)| \\ &\quad \le \sqrt{\frac{1}{2}\int (x_1 - x_2)^4(\mathrm{d}P(x_1, x_2, y) + \mathrm{d}Q(x_1, x_2, y))} \\ &\quad\quad \times \sqrt{D_{\mathrm{KL}}(P \,\|\, Q)}. \end{aligned} \tag{80}$$

For the first term on the right-hand side of (80), observe that

$$\begin{aligned} &\int (x_1 - x_2)^4(\mathrm{d}P(x_1, x_2, y) + \mathrm{d}Q(x_1, x_2, y)) \\ &\quad \overset{(a)}{\le} 8\int (x_1^4 + x_2^4)(\mathrm{d}P(x_1, x_2, y) + \mathrm{d}Q(x_1, x_2, y)) \\ &\quad \overset{(b)}{=} 32\,\mathbb{E}\big[|X|^4\big], \end{aligned} \tag{81}$$

where (a) follows from (63) and (b) holds because the marginal distributions of X_1 and X_2 are identical under P and Q.

For the second term on the right-hand side of (80), observe that P and Q can be expressed as

$$\begin{aligned} P(x_1, x_2, y) &= \frac{P_{X,Y}(x_1, y) P_{X,Y}(x_2, y)}{P_Y(y)} \\ Q(x_1, x_2, y) &= \frac{P_{X,Z}(x_1, y) P_{X,Z}(x_2, y)}{P_Z(y)}. \end{aligned}$$

Letting $(X_1, X_2, Y) \sim P$, we see that the Kullback-Leibler divergence satisfies

$$\begin{aligned} D_{\mathrm{KL}}(P \| Q) &= \mathbb{E}\left[\log\left(\frac{P_{X,Y}(X_1, Y) P_{X,Y}(X_2, Y) P_Z(Y)}{P_{X,Z}(X_1, Y) P_{X,Z}(X_2, Y) P_Y(Y)} \right) \right] \\ &= 2 D_{\mathrm{KL}}(P_{X,Y} \| P_{X,Z}) - D_{\mathrm{KL}}(P_Y \| P_Z) \\ &\leq 2 D_{\mathrm{KL}}(P_{X,Y} \| P_{X,Z}). \qquad (82) \end{aligned}$$

Plugging (81) and (82) back into (80) completes the proof of Lemma 34. □

Appendix B
Proofs of Results in Section III

A. Proof of Lemma 5

Let $Y = \sqrt{s} X + W$ where $X \sim P_X$ and $W \sim \mathcal{N}(0,1)$ are independent. Letting X_Y and X_Y' denote conditionally independent draws from $P_{X|Y}$ the conditional variance can be expressed as $\mathsf{Var}(X \,|\, Y) = \frac{1}{2}\mathbb{E}\big[(X_Y - X_Y')^2 \,|\, Y\big]$. Therefore,

$$\begin{aligned} \left| \frac{\mathrm{d}}{\mathrm{d}s} \mathsf{mmse}_X(s) \right| &\overset{(a)}{=} \mathbb{E}\Big[\big(\mathsf{Var}(X \,|\, Y)\big)^2 \Big] \\ &= \frac{1}{4} \mathbb{E}\Big[\big(\mathbb{E}\big[(X_Y - X_Y')^2 \,|\, Y\big] \big)^2 \Big] \\ &\overset{(b)}{\leq} \frac{1}{4} \mathbb{E}\Big[(X_Y - X_Y')^4 \Big] \\ &\overset{(c)}{\leq} 4 \mathbb{E}\big[X^4\big], \end{aligned}$$

where (a) follows from (12), (b) follows from Jensen's inequality and (c) follows from (63) and the fact that X_Y and X_Y' have the same distribution as X. This completes the proof of (13).

Next, since $\mathbb{E}[W \,|\, Y] = Y - \sqrt{s}\mathbb{E}[X \,|\, Y]$, the conditional variance can also be expressed as

$$\mathsf{Var}(X \,|\, Y) = \frac{1}{s} \mathsf{Var}(W \,|\, Y).$$

Using the same argument as above leads to

$$\left| \frac{\mathrm{d}}{\mathrm{d}s} \mathsf{mmse}_X(s) \right| \leq \frac{4 \mathbb{E}\big[W^4\big]}{s^2} = \frac{12}{s^2}.$$

This completes the proof of (14).

B. Proof of Lemma 6

The first order MI difference can be decomposed as

$$I'_{m,n} = h(Y_{m+1} \,|\, Y^m, A^{m+1}) - h(Y_{m+1} \,|\, Y^m, A^{m+1}, X^n),$$

where the differential entropies are guaranteed to exist because of the additive Gaussian noise. The second term is given by the entropy of the noise,

$$h(Y_{m+1} \,|\, Y^m, A^{m+1}, X^n) = h(W_{m+1}) = \frac{1}{2}\log(2\pi e),$$

and thus does not depend on m. Using this decomposition, we can now write

$$\begin{aligned} I''_{m,n} &= h(Y_{m+2} \,|\, Y^{m+1}, A^{m+2}) - h(Y_{m+1} \,|\, Y^m, A^{m+1}) \\ &= h(Y_{m+2} \,|\, Y^{m+1}, A^{m+2}) - h(Y_{m+2} \,|\, Y^m, A^{m+2}) \\ &= -I(Y_{m+1}; Y_{m+2} \,|\, Y^m, A^{m+2}), \end{aligned}$$

where the second step follows from the fact that, conditioned on (Y^m, A^m), the new measurements pairs (Y_{m+1}, A_{m+1}) and (Y_{m+2}, A_{m+2}) are identically distributed.

C. Proof of Lemma 7

We first consider the upper bound in (18). Starting with the chain rule for mutual information, we have

$$I(X^n; Y^m \mid A^m) = \sum_{i=1}^{n} I(X_i; Y^m \mid A^m, X^{i-1}). \qquad (83)$$

Next, observe that each summand satisfies

$$\begin{aligned} & I(X_i; Y^m \mid A^m, X^{i-1}) \\ &\quad \overset{(a)}{\leq} I(X_i; X_{i+1}^n, Y^m \mid A^m, X^{i-1}) \\ &\quad \overset{(b)}{=} I(X_i; Y^m \mid A^m, X^{i-1}, X_{i+1}^n), \qquad (84) \end{aligned}$$

where (a) follows from the data processing inequality and (b) follows from expanding the mutual information using the chain rule and noting that $I(X_i; X_{i+1}^n \mid A^m, X^{i-1})$ is equal to zero because the signal entries are independent.

Conditioned on data $(A^m, X^{i-1}, X_{i+1}^n)$, the mutual information provided by Y^m is equivalent to the mutual information provided by the measurement vector

$$Y^m - \mathbb{E}\big[Y^m \mid A^m, X^{i-1}, X_{i+1}^n\big] = A^m(i) X_i + W^m,$$

where $A^m(i)$ is the i-th column of A^m. Moreover, by the rotational invariance of the Gaussian distribution of the noise, the linear projection of this vector in the direction of $A^m(i)$ contains all of the information about X_i. This projection can be expressed as

$$\frac{\langle A^m(i), A^m(i) X_i + W^m \rangle}{\|A^m(i)\|} = \|A^m(i)\| X_i + W,$$

where $W = \langle A^m(i), W^m \rangle / \|A^m(i)\|$ is Gaussian $\mathcal{N}(0,1)$ and independent of X_i and $A^m(i)$, by the Gaussian distribution of W^m. Therefore, the mutual information obeys

$$\begin{aligned} & I(X_i; Y^m \,|\, A^m, X^{i-1}, X_{i+1}^n) \\ &\quad = I(X_i; \|A^m(i)\| X_i + W \mid \|A^m(i)\|) \\ &\quad = \mathbb{E}\big[I_X\big(\|A^m(i)\|^2\big) \big] \\ &\quad = \mathbb{E}\big[I_X\big(\tfrac{1}{n}\chi_m^2\big) \big], \qquad (85) \end{aligned}$$

where the last step follows from the fact that the entries of $A^m(i)$ are i.i.d. Gaussian $\mathcal{N}(0, 1/n)$. Combining (83), (84), and (85) gives the upper bound in (18).

Along similar lines, the lower bound in (19) follows from

$$\begin{aligned} M_{m,n} &\triangleq \frac{1}{n}\mathsf{mmse}(X^n \,|\, Y^m, A^m) \\ &\overset{(a)}{=} \mathsf{mmse}(X_n \,|\, Y^m, A^m) \\ &\overset{(b)}{\ge} \mathsf{mmse}(X_n \,|\, Y^m, A^m, X^{n-1}) \\ &\overset{(c)}{=} \mathsf{mmse}(X_n \,|\, A^m(n) X_n + W^m, A^m(n)) \\ &\overset{(d)}{=} \mathsf{mmse}(X_n \mid \|A^m(n)\| X_i + W, \|A^m(n)\|) \\ &\overset{(e)}{=} \mathbb{E}\big[\mathsf{mmse}_X\big(\tfrac{1}{n}\chi^2_m\big)\big], \end{aligned}$$

where (a) follows from the fact that the distributions of the columns of A^m and entries of X^n are permutation invariant, (b) follows from the data processing inequality for MMSE, (c) follows from the independence of the signal entries, (d) follows from the Gaussian distribution of the noise W^m, and (e) follows from the distribution of $A^m(n)$.

We now turn our attention to the lower bound in (18). Let the QR decomposition of A^m be given by

$$A^m = QR,$$

where Q is an $m \times n$ orthogonal matrix and R is an $m \times n$ upper triangular matrix. Under the assumed Gaussian distribution on A^m, the nonzero entries of R are independent random variables with [56, Theorem 2.3.18]:

$$R_{i,j} \sim \begin{cases} \frac{1}{n}\chi^2_{m-i+1}, & \text{if } i = j \\ \mathcal{N}\big(0, \frac{1}{n}\big), & \text{if } i < j. \end{cases} \tag{86}$$

Let the rotated measurements and noise be defined by

$$\widetilde{Y}^m = Q^T Y^m, \qquad \widetilde{W}^m = Q^T W^m$$

and observe that

$$\widetilde{Y}^m = R X^n + \widetilde{W}^m.$$

By the rotational invariance of the Gaussian distribution of the noise W^m, the rotated noise $\widetilde{W}^m$ is Gaussian $\mathcal{N}(0, I_{m\times m})$ and independent of everything else. Therefore, only the first $d \triangleq \min(m,n)$ measurements provide any information about the signal. Using this notation, the mutual information can be expressed equivalently as

$$\begin{aligned} I(X^n; Y^m \mid A^m) &= I(X^n; \widetilde{Y}^d \mid R) \\ &= \sum_{i=1}^{d} I(X^n; \widetilde{Y}_k \mid \widetilde{Y}^d_{k+1}, R), \end{aligned} \tag{87}$$

where the second step follows from the chain rule for mutual information.

To proceed, note that the measurements $\widetilde{Y}^d_k$ are independent of the first part of the signal X^{k-1}, because R is upper triangular. Therefore, for all $1 \le k \le d$,

$$\begin{aligned} I(X^n; \widetilde{Y}_k \mid \widetilde{Y}^d_{k+1}, R) &= I(X^n_k; \widetilde{Y}_k \mid \widetilde{Y}^d_{k+1}, R) \\ &\overset{(a)}{\ge} I(X_k; \widetilde{Y}_k \mid \widetilde{Y}^d_{k+1}, X^n_{k+1}, R) \\ &\overset{(b)}{=} I(X_k; R_{k,k} X_k + \widetilde{W}_k \mid R) \\ &\overset{(c)}{=} \mathbb{E}\big[I_X\big(\tfrac{1}{n}\chi^2_{m-k+1}\big)\big], \end{aligned} \tag{88}$$

where (a) follows from the data processing inequality, (b) follows from the fact that R is upper triangular and the independence of the signal entries, and (c) follows from (86). Combining (87) and (88) gives the upper bound in (18)

For the upper bound in (19), suppose that $m \ge n$ and let $\widetilde{Y}^n$ be the rotated measurements defined above. Then we have

$$\begin{aligned} M_{m,n} &\triangleq \frac{1}{n}\mathsf{mmse}(X_n \mid Y^m, A^m) \\ &\overset{(a)}{=} \mathsf{mmse}(X_n \,|\, Y^m, A^m) \\ &\overset{(b)}{=} \mathsf{mmse}(X_n \,|\, \widetilde{Y}^m, R) \\ &\overset{(c)}{\le} \mathsf{mmse}(X_n \,|\, \widetilde{Y}_n, R_{n,n}) \\ &\overset{(d)}{=} \mathsf{mmse}(X_n \,|\, R_{n,n} X_n + \widetilde{W}_n, R_{n,n}) \\ &\overset{(e)}{=} \mathbb{E}\big[\mathsf{mmse}_X\big(\tfrac{1}{n}\chi^2_{m-n+1}\big)\big], \end{aligned}$$

where (a) follows from the fact that the distributions of the columns of A^m and entries of X^n are permutation invariant, (b) follows from the fact that multiplication by Q is a one-to-one transformation, (c) follows from the data processing inequality for MMSE, (d) follows from the fact that R is upper triangular with $m \ge n$, and (e) follows from (86).

D. Proof of Lemma 8

Let $U = \frac{1}{n}\chi^2_{m-n+1}$ and $V = \frac{1}{n}\chi^2_{n+1}$ be independent scaled chi-square random variables and let $Z = U + V$. Using this notation, the lower bound in (18) satisfies

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\big[I_X\big(\tfrac{1}{n}\chi^2_{m-k+1}\big)\big] \ge \mathbb{E}[I_X(U)], \tag{89}$$

where we have used the fact that the mutual information function is non-decreasing. Moreover, by (11), we have

$$\begin{aligned} \mathbb{E}[I_X(Z)] - \mathbb{E}[I_X(U)] &\le \frac{1}{2}\mathbb{E}\big[(Z/U - 1)_+\big] \\ &= \frac{1}{2}\mathbb{E}[V/U] \\ &= \frac{1}{2}\frac{n+1}{m-n-1}. \end{aligned} \tag{90}$$

Next, observe that Z has a scaled chi-squared distribution $Z \sim \frac{1}{n}\chi^2_{m+2}$, whose inverse moments are given by

$$\mathbb{E}\big[Z^{-1}\big] = \frac{n}{m}, \qquad \mathsf{Var}(Z^{-1}) = \frac{2n^2}{m^2(m-2)}.$$

Therefore, by (11), we have

$$\begin{aligned} I_X\big(\tfrac{m}{n}\big) - \mathbb{E}[I_X(Z)] &\le \frac{1}{2}\mathbb{E}\left[\left(\frac{m/n}{Z} - 1\right)_+\right] \\ &\le \frac{m}{2n}\mathbb{E}\left[\left|\frac{1}{Z} - \frac{n}{m}\right|\right] \\ &\le \frac{m}{2n}\sqrt{\mathsf{Var}(Z^{-1})} \\ &= \frac{1}{2}\sqrt{\frac{2}{m-2}}, \end{aligned} \tag{91}$$

where the second and third steps follow from Jensen's inequality. Combining (89), (90) and (91) completes the proof of Inequality (20).

We use a similar approach for the MMSE. Note that

$$\begin{aligned}\mathbb{E}[\mathsf{mmse}_X(U)] &= \mathbb{E}\big[\mathsf{mmse}_X\big(\tfrac{1}{n}\chi^2_{m-n+1}\big)\big] \\ \mathbb{E}[\mathsf{mmse}_X(Z)] &\le \mathbb{E}\big[\mathsf{mmse}_X\big(\tfrac{1}{n}\chi^2_{m}\big)\big], \qquad (92)\end{aligned}$$

where the second inequality follows from the monotonicity of the MMSE function. By Lemma 5, the MMSE obeys

$$\begin{aligned}\mathbb{E}[\mathsf{mmse}_X(U)] - \mathbb{E}[\mathsf{mmse}_X(Z)] &\le 12\,\mathbb{E}\left[\left|\frac{1}{U} - \frac{1}{Z}\right|\right] \\ &= 12\big(\mathbb{E}\big[U^{-1}\big] - \mathbb{E}\big[Z^{-1}\big]\big) \\ &= 12\frac{n+1}{m-n-1}. \qquad (93)\end{aligned}$$

Moreover,

$$\begin{aligned}\big|\mathbb{E}[\mathsf{mmse}_X(Z)] - \mathsf{mmse}_X\big(\tfrac{m}{n}\big)\big| &\le 12\,\mathbb{E}\left[\left|\frac{1}{Z} - \frac{n}{m}\right|\right] \\ &\le 12\sqrt{\mathsf{Var}(Z^{-1})} \\ &= \frac{1}{2}\sqrt{\frac{2}{m-2}} \qquad (94)\end{aligned}$$

where the second and third steps follow from Jensen's inequality. Combining (92), (93) and (94) completes the proof of Inequality (21).

E. Proof of Lemma 9

The upper bound in (23) follows from noting that

$$\mathcal{I}_{\mathrm{RS}}(\delta) \triangleq \min_{z\ge 0} R(\delta, z) \le R(\delta, 0) = I_X(\delta).$$

For the lower bound, observe that the replica-MI function can also be expressed in terms of the replica-MMSE function as

$$\mathcal{I}_{\mathrm{RS}}(\delta) = R(\delta, \mathcal{M}_{\mathrm{RS}}(\delta)).$$

Since the term $[\log(1+z) - \frac{z}{1+z}]$ in the definition of $R(\delta, z)$ is non-negative, we have the lower bound

$$\mathcal{I}_{\mathrm{RS}}(\delta) \ge I_X\left(\frac{\delta}{1+\mathcal{M}_{\mathrm{RS}}(\delta)}\right). \qquad (95)$$

Next, we recall that the replica-MMSE function satisfies the fixed-point equation

$$\mathcal{M}_{\mathrm{RS}}(\delta) = \mathsf{mmse}_X\left(\frac{\delta}{1+\mathcal{M}_{\mathrm{RS}}(\delta)}\right). \qquad (96)$$

Also, for any signal distribution P_X with, the MMSE function satisfies the upper bound $\mathsf{mmse}_X(s) \le 1/s$ [37, Proposition 4]. Therefore,

$$\mathcal{M}_{\mathrm{RS}}(\delta) \le \frac{1+\mathcal{M}_{\mathrm{RS}}(\delta)}{\delta}.$$

For $\delta > 1$, rearranging the terms leads to the upper bound

$$\mathcal{M}_{\mathrm{RS}}(\delta) \le \frac{1}{\delta - 1}. \qquad (97)$$

Combining (95) and (97) with the fact that the mutual information is non-decreasing yields

$$\mathcal{I}_{\mathrm{RS}}(\delta) \ge I_X\left(\frac{\delta}{1+\frac{1}{1-\delta}}\right) = I_X(\delta - 1).$$

Lastly, we consider the bounds in (24). Combining (96) and (97) with the fact that the MMSE is non-increasing yields

$$\mathcal{M}_{\mathrm{RS}}(\delta) \le \mathsf{mmse}_X\left(\frac{\delta}{1+\frac{1}{1-\delta}}\right) = \mathsf{mmse}_X(\delta - 1).$$

Alternatively, starting with (96) and using the non-negativity of $\mathcal{M}_{\mathrm{RS}}$ leads to the lower bound. This completes the proof of Lemma 9.

F. Proof of Lemma 10

To lighten notation, we will write $I(A) = I(n^{-1/2}H)$ in place of $I_{m,n}(A^m)$, where $H = \sqrt{n}A$ is an $m \times n$ matrix with i.i.d. standard Gaussian entries. By the Gaussian Poincaré inequality [54, Theorem 3.20], the variance satisfies

$$\begin{aligned}\mathsf{Var}(I(A)) &= \mathsf{Var}(I(n^{-1/2}H)) \qquad (98)\\ &\le \mathbb{E}\left[\left\|\nabla_H I(n^{-1/2}H)\right\|_F^2\right] \qquad (99)\\ &= \frac{1}{n}\mathbb{E}\Big[\|\nabla_A I(A)\|_F^2\Big], \qquad (100)\end{aligned}$$

where ∇_A is the gradient operator with respect to A. Furthermore, by the multivariate I-MMSE relationship [57, Theorem 1], the gradient of the mutual information with respect to A is given by

$$\nabla_A I(A) = AK, \qquad (101)$$

where $K = \mathbb{E}[\mathsf{Cov}(X^n|Y^m, A) \mid A]$ is the expected conditional covariance matrix as a function of the matrix A. Becuase K is a covariance matrix, it is positive semi-definite and the square root $K^{\frac{1}{2}}$ is well-defined. Using these properties, we can write

$$\begin{aligned}\|\nabla_A I(A)\|_F^2 &= \mathrm{tr}\big(AK^2A^T\big) \\ &= \mathrm{tr}\Big(K^{\frac{1}{2}}A^TAK^{\frac{1}{2}}K\Big) \\ &\le \lambda_{\max}(K^{\frac{1}{2}}A^TAK^{\frac{1}{2}})\,\mathrm{tr}(K) \\ &= \lambda_{\max}(AKA^T)\,\mathrm{tr}(K). \qquad (102)\end{aligned}$$

Next, we recognize that $AKA^T = \mathbb{E}[\mathsf{Cov}(AX^n|Y^m, A) \mid A] = \mathbb{E}[\mathsf{Cov}(W^m|Y^m, A) \mid A]$ is the expected conditional covariance matrix of the noise W^n. Consequently, the maximum eigenvalue satisfies

$$\begin{aligned}\lambda_{\max}(AKA^T) &= \sup_{u\in\mathbb{R}^m\,:\,\|u\|=1} u^TAKA^Tu \\ &= \sup_{u\in\mathbb{R}^m\,:\,\|u\|=1} \mathbb{E}\big[\mathsf{Var}(u^TW^m \mid Y^m, A) \mid A\big] \\ &\le \sup_{u\in\mathbb{R}^m\,:\,\|u\|=1} \mathbb{E}\big[\mathsf{Var}(u^TW^m)\big] \\ &= 1, \qquad (103)\end{aligned}$$

where the inequality follows from the law of total variance (68). Plugging (103) back into (102) and then taking the expectation with respect to A leads to

$$\begin{aligned}\frac{1}{n}\mathbb{E}\Big[\|\nabla_A I(A)\|_F^2\Big] &\le \frac{1}{n}\mathbb{E}[\mathrm{tr}(\mathsf{Cov}(X^n \mid Y^m, A)] \\ &= M_{m,n}. \qquad (104)\end{aligned}$$

Combining this inequality with (100) completes the proof of Lemma 10.

G. Proof of Lemma 11

To simplify notation, the mutual information density is denoted by $Z = \imath(X^n; Y^m \,|\, A^m)$. Note that Z can be expressed as a function of the random tuple (X^n, W^m, A^m) according to

$$Z = \log\left(\frac{f_{Y^m|X^n,A^m}(A^m X^n + W^m \,|\, X^n, A^m)}{f_{Y^m|A^m}(A^m X^n + W^m \,|\, A^m)} \right). \tag{105}$$

Starting with the law of total variance (68), we see that

$$\mathsf{Var}(Z) = \mathbb{E}[\mathsf{Var}(Z \,|\, A^m)] + \mathsf{Var}(\mathbb{E}[Z \,|\, A^m]). \tag{106}$$

The second term on the right-hand side of (106) is variance with respect to the matrix, which is bounded by Lemma 10.

The first term on the right-hand side of (106) is the variance with respect to the signal and the noise. Since the entries of X^n and W^m are independent, the variance can be bounded using the the Efron-Stein inequality [54, Theorem 3.1], which yields

$$\begin{aligned}\mathbb{E}[\mathsf{Var}(Z \,|\, A^m)] &\le \sum_{i=1}^n \mathbb{E}\Big[(Z - \mathbb{E}_{X_i}[Z])^2\Big] \\ &\quad + \sum_{i=1}^m \mathbb{E}\Big[(Z - \mathbb{E}_{W_i}[Z])^2\Big]. \end{aligned} \tag{107}$$

At this point, Lemma 11 follows from combining (106), Lemma 10, and (107) with the following inequalities

$$\mathbb{E}\Big[(Z - \mathbb{E}_{X_i}[Z])^2\Big] \le 12\Big(1 + \Big(1 + \sqrt{\tfrac{m}{n}}\Big) B^{\frac{1}{4}}\Big)^4 \tag{108}$$

$$\mathbb{E}\Big[(Z - \mathbb{E}_{W_i}[Z])^2\Big] \le \sqrt{B}. \tag{109}$$

Inequalities (108) and (109) are proved in the following subsections.

1) Proof of Inequality (108)*:* Observe that the expectation $\mathbb{E}\Big[(Z - \mathbb{E}_{X_i}[Z])^2\Big]$ is identical for all $i \in [n]$ because the distribution on the entries in X^n and the columns of A^m are permutation invariant. Throughout this proof, we focus on the variance with respect to the last signal entry X_n.

Starting with the chain rule for mutual information density, we obtain the decomposition

$$Z = \imath(X_n; Y^m \,|\, A^m) + \imath(X^{n-1}; Y^m \,|\, X_n, A^m). \tag{110}$$

The second term is independent of X_n, and thus does not contribute to the conditional variance. To characterize the first term, we introduce a transformation of the data that isolates the effect of the nth signal entry. Let Q be drawn uniformly from the set of $m \times m$ orthogonal matrices whose last row is the unit vector in the direction of the last column of A^m, and let the *rotated* data be defined according to

$$\widetilde{Y}^m = Q Y^m, \qquad \widetilde{A}^m = Q A^m, \qquad \widetilde{W}^m = Q W^m.$$

By construction, the last column of $\widetilde{A}^m$ is zero everywhere except for the last entry with

$$\widetilde{A}_{i,n} = \begin{cases} 0, & \text{if } 1 \le i \le m-1 \\ \|A^m(n)\|, & \text{if } i = m, \end{cases}$$

where $A^m(n)$ denotes the n-th column of A^m. Furthermore, by the rotational invariance of the Gaussian distribution of the noise, the rotated noise $\widetilde{W}^m$ has the same distribution as W^m and is independent of everything else.

Expressing the mutual information density in terms of the rotated data leads to the further decomposition

$$\begin{aligned} \imath(X_n; Y^m \,|\, A^m) &\overset{(a)}{=} \imath(X_n; \widetilde{Y}^m \,|\, \widetilde{A}^m) \\ &\overset{(b)}{=} \imath(X_n; \widetilde{Y}_m \,|\, \widetilde{Y}^{m-1}, \widetilde{A}^m) \\ &\quad + \imath(X_n; \widetilde{Y}^{m-1} \,|\, \widetilde{A}^m) \\ &\overset{(c)}{=} \imath(X_n; \widetilde{Y}_m \,|\, \widetilde{Y}^{m-1}, \widetilde{A}^m), \end{aligned} \tag{111}$$

where (a) follows from fact that multiplication by Q is a one-to-one transformation, (b) follows from the chain rule for mutual information density, and (c) follows from fact that the first $m-1$ entries of $\widetilde{Y}^m$ are independent of X_n.

To proceed, we introduce the notation $U = (\widetilde{Y}^{m-1}, \widetilde{A}^{m-1})$. Since the noise is independent of the measurements, the conditional density of $\widetilde{Y}_m$ given $(X^n, U, \widetilde{A}_m)$ obeys

$$\begin{aligned} &f_{\widetilde{Y}_m|X^n,U,\widetilde{A}_m}(\widetilde{y}_m \mid x^n, u, \widetilde{a}_m) \\ &\quad = f_{\widetilde{W}_m}(\widetilde{y}_m - \langle \widetilde{a}_m, x^n \rangle) \\ &\quad = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2} (\widetilde{y}_m - \langle \widetilde{a}_m, x^n \rangle)^2 \Big). \end{aligned}$$

Starting with (111), the mutual information density can now be expressed as

$$\begin{aligned} \imath(X_n; Y^m \mid A^m) &= \imath\Big(X_n; \widetilde{Y}_m \mid U, \widetilde{A}_m\Big) \\ &= \log\left(\frac{f_{\widetilde{Y}_m|X^n,U,\widetilde{A}_m}(\widetilde{Y}_m \mid X^n, U, \widetilde{A}_m)}{f_{\widetilde{Y}_m|U,\widetilde{A}_m}(\widetilde{Y}_m \mid U, \widetilde{A}_m)} \right), \\ &= g(U, \widetilde{Y}_m, \widetilde{A}_m) - \frac{1}{2} \widetilde{W}_m^2, \end{aligned} \tag{112}$$

where

$$g(U, \widetilde{Y}_m, A_m) = \log\left(\frac{(2\pi)^{-\frac{1}{2}}}{f_{\widetilde{Y}_m|U,\widetilde{A}_m}(\widetilde{Y}_m \mid U, \widetilde{A}_m)} \right).$$

Furthermore, by (110) and (112), the difference between Z and its conditional expectation with respect to X_n is given by

$$Z - \mathbb{E}_{X_n}[Z] = g(U, \widetilde{Y}_m, \widetilde{A}_m) - \mathbb{E}_{X_n}\Big[g(U, \widetilde{Y}_m, \widetilde{A}_m)\Big].$$

Squaring both sides and taking the expectation yields

$$\begin{aligned} &\mathbb{E}\Big[(Z - \mathbb{E}_{X_n}[Z])^2\Big] \\ &\quad = \mathbb{E}\bigg[\Big(g(U, \widetilde{Y}_m, \widetilde{A}_m) - \mathbb{E}_{X_n}\Big[g(U, \widetilde{Y}_m, \widetilde{A}_m)\Big]\Big)^2\bigg] \\ &\quad \overset{(b)}{\le} \mathbb{E}\Big[g^2(U, \widetilde{Y}_m, \widetilde{A}_m)\Big], \end{aligned} \tag{113}$$

where (b) follows from the law of total variance (68).

Next, we bound the function $g(u, \widetilde{y}_m, \widetilde{a}_m)$. Let X_u^n be drawn according to the conditional distribution $P_{X^n|U=u}$. Then, the conditional density of $\widetilde{Y}_m$ given $(U, \widetilde{A}_m)$ is given by

$$\begin{aligned} &f_{\widetilde{Y}_m|U,\widetilde{A}_m}(\widetilde{y}_m \mid u, \widetilde{a}_m) \\ &\quad = \mathbb{E}\Big[f_{\widetilde{Y}_m|X^n,U,\widetilde{A}_m}(\widetilde{y}_m \mid X_u^n, u, \widetilde{a}_m)\Big] \end{aligned}$$

$$= \mathbb{E}\left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\widetilde{y}_m - \langle a_m, X_u^n\rangle)^2\right)\right].$$

Since the conditional density obeys the upper bound

$$f_{\widetilde{Y}_m|U,\widetilde{A}_m}(\widetilde{y}_m \mid u, \widetilde{a}_m) \le (2\pi)^{-\frac{1}{2}},$$

we see that $g(u, \widetilde{y}_m, \widetilde{a}_m)$ is nonnegative. Alternatively, by Jensen's inequality,

$$\begin{aligned} &f_{\widetilde{Y}_m|U,\widetilde{A}_m}(\widetilde{y}_m \mid u, \widetilde{a}_m) \\ &\quad \ge \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\mathbb{E}\left[(\widetilde{y}_m - \langle a_m, X_u^n\rangle)^2\right]\right), \end{aligned}$$

and thus

$$g(u, \widetilde{y}_m, \widetilde{a}_m) \le \frac{1}{2}\mathbb{E}\left[(\widetilde{y}_m - \langle a_m, X_u^n\rangle)^2\right].$$

Using these facts leads to the following inequality:

$$\begin{aligned} g^2(u, \widetilde{y}_m, \widetilde{a}_m) &\le \frac{1}{4}\left(\mathbb{E}_{X_u^n}\left[(y_m - \langle \widetilde{a}_m, X_u^n\rangle)^2\right]\right)^2 \\ &\overset{(a)}{\le} \frac{1}{4}\mathbb{E}_{X_u^n}\left[(\widetilde{y}_m - \langle \widetilde{a}_m, X_u^n\rangle)^4\right] \\ &\overset{(b)}{\le} 2(\widetilde{y}_m)^4 + 2\mathbb{E}_{X_u^n}\left[(\langle \widetilde{a}_m, X_u^n\rangle)^4\right], \end{aligned}$$

where (a) follows from Jensen's inequality and (b) follows from (63). Taking the expectation with respect to the random tuple $(U, \widetilde{Y}_m, \widetilde{A}_m)$, we can now write

$$\begin{aligned} \mathbb{E}\left[g^2(U, \widetilde{Y}_m, \widetilde{A}_m)\right] &\le 2\mathbb{E}\left[\left(\widetilde{Y}_m\right)^4\right] + 2\mathbb{E}\left[\left(\langle \widetilde{A}_m, X_U^n\rangle\right)^4\right] \\ &\overset{(a)}{=} 2\mathbb{E}\left[\left(\widetilde{Y}_m\right)^4\right] + 2\mathbb{E}\left[\left(\langle \widetilde{A}_m, X^n\rangle\right)^4\right], \end{aligned} \tag{114}$$

where (a) follows from the fact that X_U^n has the same distribution as X^n and is independent of $\widetilde{A}_m$. To upper bound the first term, observe that

$$\begin{aligned} \mathbb{E}\left[\left(\widetilde{Y}_m\right)^4\right] &= \mathbb{E}\left[\left(\widetilde{W}_m + \sum_{i=1}^n \widetilde{A}_{m,i} X_i\right)^4\right] \\ &\overset{(a)}{\le} \left(\left(\mathbb{E}\left[\widetilde{W}_m^4\right]\right)^{\frac{1}{4}} + \left(\mathbb{E}\left[\left(\langle \widetilde{A}_m, X^n\rangle\right)^4\right]\right)^{\frac{1}{4}}\right)^4 \\ &= \left(3^{\frac{1}{4}} + \left(\mathbb{E}\left[\left(\langle \widetilde{A}_m, X^n\rangle\right)^4\right]\right)^{\frac{1}{4}}\right)^4 \end{aligned} \tag{115}$$

where (a) follows from Minkowski's inequality (64). To bound the fourth moment of $\langle \widetilde{A}_m, X^n\rangle$, we use the fact that the entries of the rotated measurement vector $\widetilde{A}_m$ are independent with

$$\widetilde{A}_{m,i} \sim \begin{cases} \mathcal{N}(0, \frac{1}{n}), & \text{if } 1 \le i \le n-1 \\ \frac{1}{\sqrt{n}}\chi_m, & \text{if } i = n, \end{cases}$$

where χ_m denotes a chi random variable with m degrees of freedom. Thus, we have

$$\begin{aligned} &\left(\mathbb{E}\left[\left(\langle \widetilde{A}_m, X^n\rangle\right)^4\right]\right)^{\frac{1}{4}} \\ &\overset{(a)}{\le} \left(\mathbb{E}\left[\left(\sum_{i=1}^{n-1} \widetilde{A}_{m,i} X_i\right)^4\right]\right)^{\frac{1}{4}} + \left(\mathbb{E}\left[\left(\widetilde{A}_{m,n} X_n\right)^4\right]\right)^{\frac{1}{4}} \\ &\overset{(b)}{=} \left(\frac{3}{n^2}\mathbb{E}\left[\|X^{n-1}\|^4\right]\right)^{\frac{1}{4}} + \left(\frac{m(m+2)}{n^2}\mathbb{E}\left[X_n^4\right]\right)^{\frac{1}{4}} \\ &\overset{(c)}{\le} \left(\frac{3(n-1)^2}{n^2}B\right)^{\frac{1}{4}} + \left(\frac{m(m+2)}{n^2}B\right)^{\frac{1}{4}} \\ &\le \left(1 + \sqrt{\frac{m}{n}}\right)(3B)^{\frac{1}{4}}, \end{aligned} \tag{116}$$

where (a) follows from Minkowski's inequality (64), (b) follows from the distribution on $\widetilde{A}^m$ and (c) follows from Assumption 2 and inequality (61) applied to $\|X^{n-1}\|^4 = \left(\sum_{i=1}^{n-1} X_i^2\right)^2$.

Finally, combining (113), (114), (115), and (116) leads to

$$\begin{aligned} \mathbb{E}\left[(Z - \mathbb{E}_{X_n}[Z])^2\right] &\le 2\left(3^{\frac{1}{4}} + \left(1 + \sqrt{\frac{m}{n}}\right)(3B)^{\frac{1}{4}}\right)^4 \\ &\quad + 2\left(1 + \sqrt{\frac{m}{n}}\right)^4 (3B) \\ &\le 12\left(1 + \left(1 + \sqrt{\frac{m}{n}}\right)B^{\frac{1}{4}}\right)^4. \end{aligned}$$

This completes the proof of Inequality (108).

2) Proof of Inequality (109)*:* Observe that the expectation $\mathbb{E}\left[(Z_k - \mathbb{E}_{W_m}[Z_k])^2\right]$ is identical for all $k \in [m]$ because the distributions on the entries in W^n and the rows of A^m are permutation invariant. Throughout this proof, we focus on the variance with respect to the last noise entry W_m.

Recall from (105), Z can be expressed as a function of (X^n, W^m, A^m). By the Gaussian Poincaré inequality [54, Theorem 3.20], the variance with respect to W_m obeys

$$\mathbb{E}\left[(Z - \mathbb{E}_{W_m}[Z])^2\right] \le \mathbb{E}\left[\left(\frac{\partial}{\partial W_m} Z\right)^2\right], \tag{117}$$

where $\frac{\partial}{\partial W_m} Z$ denotes the partial derivative of the right-hand side of (105) evaluated at the point (X^n, W^m, A^m).

To compute the partial derivative, observe that by the chain rule for mutual information density,

$$Z = \imath(X^n; Y^{m-1} \,|\, A^m) + \imath(X^n; Y_m \,|\, Y^{m-1}, A^m).$$

In this decomposition, the first term on the right-hand side is independent of W_m. The second term can be decomposed as

$$\begin{aligned} &\imath(X^n; Y_m \,|\, Y^{m-1}, A^m) \\ &\quad = \log\left(\frac{f_{Y_m|X^n, Y^{m-1}, A^m}(Y_m \mid X^n, Y^{m-1}, A^m)}{f_{Y_m|Y^{m-1}, A^m}(Y_m \mid Y^{m-1}, A^m)}\right) \\ &\quad = \log\left(\frac{f_{W_m}(Y_m - \langle A_m, X^n\rangle)}{f_{Y_m|Y^{m-1}, A^m}(Y_m \mid Y^{m-1}, A^m)}\right) \\ &\quad = \log\left(\frac{f_{W_m}(Y_m)}{f_{Y_m|Y^{m-1}, A^m}(Y_m \mid Y^{m-1}, A^m)}\right) \\ &\qquad + Y_m\langle A_m, X^n\rangle - \frac{1}{2}(\langle A_m, X^n\rangle)^2. \end{aligned}$$

Note that the first term on the right-hand side is the negative of the log likelihood ratio. The partial derivative of this term with respect to Y_m can be expressed in terms of the conditional expectation [58]:

$$\frac{\partial}{\partial Y_m} \log\left(\frac{f_{W_m}(Y_m)}{f_{Y_m|Y^{m-1}, A^m}(Y_m \mid Y^{m-1}, A^m)}\right)$$

$$= -\mathbb{E}[\langle A^m, X^n\rangle \mid Y^m, A^m].$$

Recalling that $Y_m = \langle A_m, X^n\rangle + W_m$, the partial derivative with respect to W_m can now be computed directly as

$$\begin{aligned}\frac{\partial}{\partial W_m} Z &= \frac{\partial}{\partial Y_m} \imath(X^n; Y_m \mid Y^{m-1}, A^m)\\ &= \langle A_m, X^n\rangle - \mathbb{E}[\langle A^m, X^n\rangle \mid Y^m, A^m]\\ &= \langle A_m, X^n - \mathbb{E}[X^n \mid Y^m, A^m]\rangle.\end{aligned}$$

Thus, the expected squared magnitude obeys

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\partial}{\partial W_i} Z_m\right)^2\right] &= \mathbb{E}\Big[(\langle A_m, X^n - \mathbb{E}[X^n \mid Y^m, A^m]\rangle)^2\Big]\\ &= \mathbb{E}[\mathsf{Var}(\langle A_m, X^n\rangle \mid Y^m, A^m)]\\ &\overset{(a)}{\le} \mathbb{E}[\mathsf{Var}(\langle A_m, X^n\rangle \mid A_m)]\\ &= \mathbb{E}\big[A_m^T \,\mathsf{Cov}(X^n) A_m\big]\\ &\overset{(b)}{=} \mathbb{E}\big[X^2\big]\\ &\overset{(c)}{\le} \sqrt{B}, \end{aligned} \tag{118}$$

where (a) follows from the law of total variance (68) and (b) follows from the fact that $\mathbb{E}\big[A_m A_m^T\big] = \frac{1}{n} I_{n\times n}$, and (c) follows from Jensen's inequality and Assumption 2. Combining (117) and (118) completes the proof of Inequality (109).

Appendix C
Proofs of Results in Section IV

A. Proof of Lemma 12.

The squared error obeys the upper bound

$$\begin{aligned}|\mathcal{E}_{m,n}|^2 &= \left(\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}[X_i \,|\, Y^m, A^m])^2\right)^2\\ &\overset{(a)}{\le} \frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}[X_i \,|\, Y^m, A^m])^4\\ &\overset{(b)}{\le} \frac{8}{n}\sum_{i=1}^n \Big[|X_i|^4 + |\mathbb{E}[X_i \,|\, Y^m, A^m]|^4\Big]\\ &\overset{(c)}{\le} \frac{8}{n}\sum_{i=1}^n \Big[|X_i|^4 + \mathbb{E}\Big[|X_i|^4 \,\Big|\, Y^m, A^m\Big]\Big],\end{aligned}$$

where (a) follows from Jensen's inequality (61), (b) follows from (63), and (c) follows from Jensen's inequality. Taking the expectation of both sides leads to

$$\mathbb{E}\Big[|\mathcal{E}_{m,n}|^2\Big] \le \frac{16}{n}\sum_{i=1}^n \mathbb{E}\Big[|X_i|^4\Big] \le 16B,$$

where the second inequality follows from Assumption 2.

Next, observe that the conditional distribution Y_{m+1} given X^n is zero-mean Gaussian with variance $1 + \frac{1}{n}\|X^n\|^2$. Thus,

$$\begin{aligned}\mathbb{E}\big[Y_{m+1}^4\big] &= \mathbb{E}\big[\mathbb{E}\big[Y^4 \,|\, X^n\big]\big]\\ &= 3\mathbb{E}\Big[\big(1 + \tfrac{1}{n}\|X^n\|^2\big)^2\Big]\\ &\overset{(a)}{\le} 6\big(1 + \tfrac{1}{n^2}\mathbb{E}\big[\|X^n\|^4\big]\big)\\ &\overset{(b)}{\le} 6(1+B),\end{aligned}$$

where (a) follows from (62) and (b) follows from (61) and Assumption 2. Along similar lines,

$$\begin{aligned}&\mathbb{E}\Big[\big|Y_{m+1} - \mathbb{E}\big[Y_{m+1} \,|\, Y^m, A^{m+1}\big]\big|^4\Big]\\ &\quad\overset{(a)}{\le} 8\mathbb{E}\Big[|Y_{m+1}|^4\Big] + 8\mathbb{E}\Big[\big|\mathbb{E}\big[Y_{m+1} \,|\, Y^m, A^{m+1}\big]\big|^4\Big]\\ &\quad\overset{(b)}{\le} 16\mathbb{E}\Big[|Y_{m+1}|^4\Big]\\ &\quad\le 96(1+B),\end{aligned}$$

where (a) follows from (63) and (b) follows from Jensen's inequality.

Finally, we note that the proof of (29) can be found in the proof of Lemma 20.

B. Proof of Lemma 13

The starting point for these identities is to observe that the differences between the measurements Y_i and Y_j and their conditional expectations given the data (Y^m, A^m, A_i, A_j) can be expressed as

$$\begin{bmatrix} Y_i - \widehat{Y}_i \\ Y_j - \widehat{Y}_j\end{bmatrix} = \begin{bmatrix} A_i^T \\ A_j^T\end{bmatrix}\Big(X^n - \widehat{X}^n\Big) + \begin{bmatrix} W_i \\ W_j\end{bmatrix}, \tag{119}$$

where $\widehat{X}^n = \mathbb{E}[X^n \mid Y^m, A^m]$ is the signal estimate after the first m measurements.

1) Proof of Identity (32)*:* Starting with (119), we see that the conditional variance of Y_i can be expressed in terms of the posterior covariance matrix of the signal:

$$\mathsf{Var}(Y_i \,|\, Y^m, A^m, A_i) = A_i^T \,\mathsf{Cov}(X^n \,|\, Y^m, A^m) A_i + 1.$$

Taking the expectation of both sides with respect to A_i yields

$$\begin{aligned}&\mathbb{E}_{A_i}[\mathsf{Var}(Y_i \,|\, Y^m, A^m, A_i)]\\ &\qquad= \mathbb{E}_{A_i}\big[A_i^T \,\mathsf{Cov}(X^n \,|\, Y^m, A^m) A_i\big] + 1\\ &\qquad= \mathbb{E}_{A_i}\big[\mathrm{tr}\big(A_i A_i^T \,\mathsf{Cov}(X^n \,|\, Y^m, A^m)\big)\big] + 1\\ &\qquad= \mathrm{tr}\big(\mathbb{E}_{A_i}\big[A_i A_i^T\big] \,\mathsf{Cov}(X^n \,|\, Y^m, A^m)\big) + 1\\ &\qquad= \frac{1}{n}\,\mathrm{tr}(\mathsf{Cov}(X^n \mid Y^m, A^m)) + 1,\end{aligned}$$

where we have used the fact that $\mathbb{E}\big[A_i A_i^T\big] = \frac{1}{n} I_{m\times m}$.

2) Proof of Identity (33)*:* Along the same lines as the conditional variance, we see that the conditional covariance between of Y_i and Y_j is given by

$$\mathsf{Cov}(Y_i, Y_j \mid Y^m, A^m, A_i, A_j) = A_i^T \,\mathsf{Cov}(X^n \mid Y^m, A^m) A_j.$$

Letting $K = \mathsf{Cov}(X^n \mid Y^m, A^m)$, the expectation of the squared covariance with respect to A_i and A_j can be computed as follows:

$$\begin{aligned}&\mathbb{E}_{A_i,A_j}\Big[(\mathsf{Cov}(Y_i, Y_j \mid Y^m, A^m, A_i, A_j))^2\Big]\\ &\qquad= \mathbb{E}_{A_i,A_j}\big[\big(A_i^T K A_j\big)\big(A_j^T K A_i^T\big)\big]\\ &\qquad= \frac{1}{n}\mathbb{E}_{A_i}\big[A_i^T K^2 A_i\big]\\ &\qquad= \frac{1}{n}\mathbb{E}_{A_i}\big[\mathrm{tr}\big(A_i A_i^T K^2\big)\big]\end{aligned}$$

$$= \frac{1}{n} \operatorname{tr}\big(\mathbb{E}_{A_i}\big[A_i A_i^T\big] K^2\big)$$
$$= \frac{1}{n^2} \operatorname{tr}\big(K^2\big),$$

where we have used the fact that A_i and A_j are independent with $\mathbb{E}\big[A_i A_i^T\big] = \frac{1}{n} I_{m\times m}$. Noting that $\operatorname{tr}(K^2) = \|K\|_F^2$ completes the proof of Identity (33).

3) Proof of Identity (34)*:* For this identity, observe that the measurement vectors (A_i, A_j) and the noise terms (W_i, W_j) in (119) are Gaussian and independent of the signal error. Therefore, the conditional distribution of $(Y_i - \widehat{Y}_i, Y_j - \widehat{Y}_j)$ given (X^n, Y^m, A^m) is i.i.d. Gaussian with mean zero and covariance

$$\mathsf{Cov}\left(\begin{bmatrix} Y_i - \widehat{Y}_i \\ Y_j - \widehat{Y}_j \end{bmatrix} \,\middle|\, X^n, Y^m, A^m \right) = \begin{bmatrix} 1 + \mathcal{E}_m & 0 \\ 0 & 1 + \mathcal{E}_m \end{bmatrix}.$$

Using the fact that the expected absolute value of standard Gaussian variable is equal to $\sqrt{2/\pi}$, we see that the conditional absolute moments are given by

$$\mathbb{E}\Big[\big|Y_i - \widehat{Y}_i\big| \,\Big|\, X^n, Y^m, A^m\Big] = \sqrt{\frac{2}{\pi}}\sqrt{1 + \mathcal{E}_m}$$
$$\mathbb{E}\Big[\big|Y_i - \widehat{Y}_i\big|\big|Y_j - \widehat{Y}_j\big| \,\Big|\, X^n, Y^m, A^m\Big] = \frac{2}{\pi}(1 + \mathcal{E}_m).$$

Taking the expectation of both sides with respect to the posterior distribution of X^n given (Y^m, A^m) leads to

$$\mathbb{E}\Big[\big|Y_i - \widehat{Y}_i\big| \,\Big|\, Y^m, A^m\Big] = \sqrt{\frac{2}{\pi}}\mathbb{E}\Big[\sqrt{1 + \mathcal{E}_m} \mid Y^m, A^m\Big]$$
$$\mathbb{E}\Big[\big|Y_i - \widehat{Y}_i\big|\big|Y_j - \widehat{Y}_j\big| \,\Big|\, Y^m, A^m\Big] = \frac{2}{\pi}(1 + V_m).$$

Finally, we see that the conditional covariance is given by

$$\begin{aligned} &\mathsf{Cov}\Big(\big|Y_i - \widehat{Y}_i\big|, \big|Y_j - \widehat{Y}_j\big| \,\Big|\, Y^m, A^m\Big) \\ &\quad = \frac{2}{\pi}\mathbb{E}\bigg[\Big(\sqrt{1 + \mathcal{E}_m}\Big)^2 \,\bigg|\, Y^m, A^m\bigg] \\ &\qquad - \frac{2}{\pi}\Big(\mathbb{E}\Big[\sqrt{1 + \mathcal{E}_m} \,\Big|\, Y^m, A^m\Big]\Big)^2 \\ &\quad = \frac{2}{\pi}\,\mathsf{Var}\Big(\sqrt{1 + \mathcal{E}_m} \mid Y^m, A^m\Big). \end{aligned}$$

This completes the proof of identity (34).

C. Proof of Lemma 14

To simplify notation, we drop the explicit dependence on the problem dimensions and write $\mathcal{E}$ and V instead of $\mathcal{E}_{m,n}$ and $V_{m,n}$. Also, we use $U = (Y^m, A^m)$ to denote the first m measurements. Using this notation, the posterior distribution is given by $P_{X^n|U}$ and the posterior variance is $V = \mathbb{E}[\mathcal{E} \,|\, U]$.

1) Proof of Inequality (35)*:* To begin, let $\mathcal{E}'_m$ be a conditionally independent copy of $\mathcal{E}$ that is drawn according to the posterior distribution $P_{\mathcal{E}|U}$. Starting with the fact that the posterior variance can expressed as $V = \mathbb{E}[\mathcal{E}' \mid U]$, the absolute deviation between $\mathcal{E}$ and V can be upper bounded using the following series of inequalities:

$$\begin{aligned} \mathbb{E}[|\mathcal{E} - V|] &= \mathbb{E}_U\Big[\mathbb{E}\big[|\mathcal{E} - \mathbb{E}[\mathcal{E}' \,|\, U]| \,\big|\, U\big]\Big] \\ &\overset{(a)}{\leq} \mathbb{E}_U\Big[\mathbb{E}\big[|\mathcal{E} - \mathcal{E}'| \,\big|\, U\big]\Big] \\ &= \mathbb{E}[|\mathcal{E} - \mathcal{E}'|] \\ &= \mathbb{E}\bigg[\bigg|\Big(\sqrt{1 + \mathcal{E}} + \sqrt{1 + \mathcal{E}'}\Big) \\ &\qquad \times \Big(\sqrt{1 + \mathcal{E}} - \sqrt{1 + \mathcal{E}'}\Big)\bigg|\bigg] \\ &\overset{(b)}{\leq} \sqrt{\mathbb{E}\bigg[\Big|\sqrt{1 + \mathcal{E}} + \sqrt{1 + \mathcal{E}'}\Big|^2\bigg]} \\ &\quad \times \sqrt{\mathbb{E}\bigg[\Big|\sqrt{1 + \mathcal{E}} - \sqrt{1 + \mathcal{E}'}\Big|^2\bigg]}, \end{aligned} \tag{120}$$

where (a) follows from Jensen's inequality and the fact that $\mathcal{E}$ and $\mathcal{E}'$ are conditionally independent given U and (b) follows from the Cauchy-Schwarz inequality.

For the first term on the right-hand side of (120), observe that

$$\begin{aligned} \mathbb{E}\bigg[\Big|\sqrt{1 + \mathcal{E}} + \sqrt{1 + \mathcal{E}'}\Big|^2\bigg] &\overset{(a)}{\leq} 2\mathbb{E}[(1 + \mathcal{E})] + 2\mathbb{E}[(1 + \mathcal{E}')] \\ &\overset{(b)}{=} 4(1 + M_{m,n}) \\ &\leq C_B, \end{aligned} \tag{121}$$

where (a) follows from (62) and (b) follows from the fact that $\mathcal{E}$ and $\mathcal{E}'$ are identically distributed.

For the second term on the right-hand side of (120), observe that

$$\begin{aligned} &\mathbb{E}\bigg[\Big|\sqrt{1 + \mathcal{E}} - \sqrt{1 + \mathcal{E}'}\Big|^2\bigg] \\ &\quad \overset{(a)}{=} 2\mathbb{E}\Big[\mathsf{Var}(\sqrt{1 + \mathcal{E}} \,|\, U)\Big] \\ &\quad \overset{(b)}{=} \pi\mathbb{E}\big[\mathsf{Cov}\big(|Z_{m+1}|, |Z_{m+2}| \,\big|\, U\big)\big], \end{aligned} \tag{122}$$

with $Z_i = Y_i - \mathbb{E}[Y_i \mid Y^m, A^m, A_i]$. Here, (a) follows from the conditional variance decomposition (67) and (b) follows from Identity (34).

To bound the expected covariance, we apply Lemma 33 with $p = 1$ to obtain

$$\begin{aligned} &\mathbb{E}[\mathsf{Cov}(|Z_{m+1}|, |Z_{m+2}| \mid U)] \\ &\quad \leq 2\sqrt{\mathbb{E}\Big[\sqrt{\mathbb{E}\big[Z_{m+1}^4 \mid U\big]\mathbb{E}\big[Z_{m+2}^4 \mid U\big]}\Big]} \\ &\qquad \times \sqrt{I(|Z_{m+1}|; |Z_{m+2}| \,|\, U)}. \end{aligned} \tag{123}$$

For the first term on the right-hand side, observe that

$$\begin{aligned} \mathbb{E}\Big[\sqrt{\mathbb{E}\big[Z_{m+1}^4 \,|\, U\big]\mathbb{E}\big[Z_{m+2}^4 \,|\, U\big]}\Big] &\overset{(a)}{\leq} \sqrt{\mathbb{E}\big[Z_{m+1}^4\big]\mathbb{E}\big[Z_{m+2}^4\big]} \\ &\overset{(b)}{\leq} C_B, \end{aligned} \tag{124}$$

where (a) follows from the Cauchy-Schwarz inequality and (b) follows from (31).

Combining (120), (121), (122), (123), and (124) yields

$$\mathbb{E}[|\mathcal{E} - V|] \leq C_B \cdot |I(|Z_{m+1}|; |Z_{m+2}| \,|\, U)|^{\frac{1}{4}}.$$

Thus, in order to complete the proof, we need to show that the mutual information term can be upper bounded in terms of the second order MI difference sequence. To this end, observe that

$$I\big(|Z_{m+1}|; |Z_{m+2}| \,\big|\, U\big)$$

$$
\begin{aligned}
&\overset{(a)}{\le} I(Z_{m+1}; Z_{m+2} \,|\, U)\\
&\overset{(b)}{\le} I(Z_{m+1}, A_{m+1}; Z_{m+2}, A_{m+2} \,|\, U)\\
&\overset{(c)}{=} I(Y_{m+1}, A_{m+1}; Y_{m+2}, A_{m+2} \,|\, U)\\
&\overset{(d)}{=} I(Y_{m+1}; Y_{m+2} \,|\, U, A_{m+1}, A_{m+2})\\
&\overset{(e)}{=} -I''_m,
\end{aligned}
$$

where (a) and (b) both follow from the data processing inequality for mutual information, (c) follows from the fact that, given (U, A_{m+i}), there is a one-to-one mapping between Z_{m+i} and Y_{m+i}, (d) follows from the fact that measurements are generated independently of everything else (Assumption 1), and (e) follows from (16). This completes the Proof of Inequality (35).

2) Proof of Inequality (36)*:* The main idea behind this proof is to combine identity (33) with the covariance bound in Lemma 33. The only tricky part is that the expectation with respect to (A_{m+1}, A_{m+2}) is taken with respect to the squared Frobenius norm whereas the expectation with respect to U is taken with respect to the square root of this quantity.

To begin, observe that for each realization $U = u$, we have

$$
\begin{aligned}
&\frac{1}{n^2}\|\mathsf{Cov}(X^n \mid U = u)\|_F^2\\
&\quad\overset{(a)}{=} \mathbb{E}\Big[\big|\mathsf{Cov}\big(Y_{m+1}, Y_{m+2} \mid U = u, A_{m+1}^{m+2}\big)\big|^2\Big]\\
&\quad\overset{(b)}{\le} 4\sqrt{\mathbb{E}\big[Y_{m+1}^4 \mid U = u\big]\mathbb{E}\big[Y_{m+2}^4 \mid U = u\big]}\\
&\qquad\times \sqrt{I(Y_{m+1}; Y_{m+2} \mid U = u, A_{m+1}^{m+2})},
\end{aligned}
$$

where (a) follows from (33) and (b) follows from Lemma 33 with $p = 2$ and the fact that

$$
\begin{aligned}
&\mathbb{E}\big[\mathbb{E}\big[Y_{m+1}^4 \mid U = u, A_{m+1}^{m+2}\big]\mathbb{E}\big[Y_{m+2}^4 \mid U = u, A_{m+1}^{m+2}\big]\big]\\
&\quad= \mathbb{E}\big[Y_{m+1}^4 \mid U = u\big]\mathbb{E}\big[Y_{m+2}^4 \mid U = u\big].
\end{aligned}
$$

Taking the expectation of the square root of both sides with respect to U leads to

$$
\begin{aligned}
&\frac{1}{n}\mathbb{E}[\|\mathsf{Cov}(X^n \mid U)\|_F]\\
&\quad\overset{(a)}{\le} 4\big|\mathbb{E}\big[Y_{m+1}^4\big]\mathbb{E}\big[Y_{m+2}^4\big]I(Y_{m+1}; Y_{m+2} \mid U, A_{m+1}^{m+2})\big|^{\frac{1}{4}}\\
&\quad\overset{(b)}{\le} C_B \cdot \big|I(Y_{m+1}; Y_{m+2} \mid U, A_{m+1}^{m+2})\big|^{\frac{1}{4}}\\
&\quad\overset{(c)}{=} C_B \cdot |I''_m|^{\frac{1}{4}},
\end{aligned}
$$

where (a) follows from the Cauchy-Schwarz inequality and Jensen's Inequality, (b) follows from (30), and (c) follows from (16). This completes the Proof of Inequality (36).

D. *Proof of Lemma 15*

The mutual information difference density can be decomposed as

$$
\begin{aligned}
\mathcal{J}_{m,n} &= \imath(X^n; \bar{Y}_{m+1} \mid Y^m, A^{m+1})\\
&= -\log\Big(f_{\bar{Y}_{m+1}|Y^m, A^{m+1}}\big(\bar{Y}_{m+1}\big)\Big)\\
&\quad - \frac{1}{2}\log(2\pi) - \frac{1}{2}W_{m+1}^2,
\end{aligned}
$$

where $f_{\bar{Y}_{m+1}|Y^m, A^{m+1}}(y)$ denotes the conditional density function of the centered measurement evaluated with the random data (Y^m, A^{m+1}). Therefore, for every $\sigma^2 > 0$, the Kullback–Leibler divergence between $P_{\bar{Y}_{m+1}|Y^m, A^{m+1}}$ and the Gaussian distribution $\mathcal{N}(0, \sigma^2)$ can be expressed as

$$
\begin{aligned}
&D_{\mathrm{KL}}\Big(P_{\bar{Y}_{m+1}|Y^m, A^{m+1}} \,\Big\|\, \mathcal{N}(0, \sigma^2)\Big)\\
&\quad= \int\bigg(\frac{1}{2\sigma^2}y^2 + \frac{1}{2}\log\big(2\pi\sigma^2\big)\bigg)f_{\bar{Y}_{m+1}|Y^m, A^{m+1}}(y)\mathrm{d}y\\
&\qquad+ \int \log\Big(f_{\bar{Y}_{m+1}|Y^m, A^{m+1}}(y)\Big)f_{\bar{Y}_{m+1}|Y^m, A^{m+1}}(y)\mathrm{d}y\\
&\quad= \frac{1}{2\sigma^2}\mathbb{E}\big[\bar{Y}_{m+1}^2 \,|\, Y^m, A^{m+1}\big] + \frac{1}{2}\log(2\pi\sigma^2)\\
&\qquad- \mathbb{E}\big[\mathcal{J}_{m,n} \,|\, Y^m, A^{m+1}\big] - \frac{1}{2}\log(2\pi) - \frac{1}{2}.
\end{aligned}
$$

Taking the expectation with respect to A_{m+1} and rearranging terms leads to

$$
\begin{aligned}
&\mathbb{E}_{A_{m+1}}\Big[D_{\mathrm{KL}}\Big(P_{\overline{Y}_{m+1}|Y^m, A^{m+1}} \,\Big\|\, \mathcal{N}(0, \sigma^2)\Big)\Big]\\
&\quad= \frac{1}{2}\log(\sigma^2) - J_{m,n} + \frac{1}{2}\bigg(\frac{1 + V_{m,n}}{\sigma^2} - 1\bigg), \qquad (125)
\end{aligned}
$$

where we have used the fact that the conditional variance is given by (38).

At this point, Identity (40) follows immediately by letting $\sigma^2 = 1 + V_{m,n}$. For Identity (41), let $\sigma^2 = 1 + M_{m,n}$ and note that the expectation of the last term in (125) is equal to zero.

E. *Proof of Lemma 17*

The error vector $\bar{X}^n = X^n - \mathbb{E}[X^n \,|\, Y^m, A^m]$ has mean zero by construction. Therefore, by (37) and Lemma 16, the posterior non-Gaussianness satisfies

$$
\begin{aligned}
\Delta_{m,n}^P &\le \frac{1}{2}\mathbb{E}\big[|\mathcal{E}_{m,n} - V_{m,n}| \,\big|\, Y^m, A^m\big]\\
&\quad+ C \cdot \Big|\tfrac{1}{n}\|\,\mathsf{Cov}(X^n \,|\, Y^m, A^m)\|_F\Big(1 + \widetilde{V}_{m,n}^2\Big)\Big|^{\frac{2}{5}},
\end{aligned}
$$

where $\widetilde{V}_{m,n} = \sqrt{\mathbb{E}\big[\mathcal{E}_{m,n}^2 \,|\, Y^m, A^m\big]}$. Taking the expectation of both sides and using the Cauchy-Schwarz inequality and Jensen's inequality leads to

$$
\begin{aligned}
\mathbb{E}\big[\Delta_{m,n}^P\big] &\le \frac{1}{2}\mathbb{E}[|\mathcal{E}_{m,n} - V_{m,n}|]\\
&\quad+ C \cdot \bigg|\tfrac{1}{n}\mathbb{E}[\|\mathsf{Cov}(X^n \,|\, Y^m, A^m)\|_F]\big(1 + \sqrt{\mathbb{E}\big[\mathcal{E}_{m,n}^2\big]}\big)\bigg|^{\frac{2}{5}}.
\end{aligned}
$$

Furthermore, combining this inequality with Lemma 14 and (28) gives

$$
\mathbb{E}\big[\Delta_{m,n}^P\big] \le C_B \cdot \Big[\big|I''_{m,n}\big|^{\frac{1}{4}} + \big|I''_{m,n}\big|^{\frac{1}{10}}\Big].
$$

Finally, since $|I''_{m,n}|$ can be bounded uniformly by a constant that depends only on B, we see that the dominant term on the right-hand side is the one with the smaller exponent. This completes the proof.

F. Proof of Lemma 18

To simplify notation we drop the explicit dependence on the problem parameters and write M and V instead of $M_{m,n}$ and $V_{m,n}$. The first inequality in (43) follows immediately from the fact that the mapping $x \mapsto \log(1+x)$ is one-Lipschitz on $\mathbb{R}_+$.

Next, letting V' be an independent copy of V, the absolute deviation of the posterior variance can be upper bounded as follows:

$$\begin{aligned}\mathbb{E}[|V - M|] &= \mathbb{E}[|V - \mathbb{E}[V']|] \\ &\overset{(a)}{\le} \mathbb{E}[|V - V'|] \\ &\overset{(b)}{\le} \sqrt{\mathbb{E}\Big[(1+V)^2\Big]\mathbb{E}\bigg[\bigg|\log\bigg(\frac{1+V}{1+V'}\bigg)\bigg|\bigg]}, \qquad (126)\end{aligned}$$

where (a) follows from Jensen's inequality and (b) follows from applying Lemma 30 with $X = 1 + V_m$ and $Y = 1 + V'_m$.

The first expectation on right-hand side of (126) obeys:

$$\mathbb{E}\Big[(1+V_m)^2\Big] \overset{(a)}{\le} 2(1 + \mathbb{E}\big[V^2\big]) \overset{(b)}{\le} C_B, \qquad (127)$$

where (a) follows from (62) and (b) follows form Jensen's inequality and (28).

For the second expectation on the right-hand side of (126), observe that by the triangle inequality, we have, for every $t \in \mathbb{R}$,

$$\bigg|\log\bigg(\frac{1+V}{1+V'}\bigg)\bigg| \le |\log(1+V) - t| + |\log(1+V') - t|.$$

Taking the expectation of both sides and minimizing over t yields

$$\mathbb{E}\bigg[\bigg|\log\bigg(\frac{1+V}{1+V'}\bigg)\bigg|\bigg] \le \min_{t \in \mathbb{R}} 2\mathbb{E}[|\log(1+V) - t|]. \qquad (128)$$

Plugging (126) and (127) back into (128) completes the proof of Lemma 18.

G. Proof of Lemma 19

Let $U = (Y^m, A^m)$ and $U_k = (Y_{m+1}^{m+k}, A_{m+1}^{m+k})$. We use the fact that the posterior variance can be expressed in terms of the expected variance of a future measurement. Specifically, by (32), it follows that for any integer $i > m + k$, we have

$$\begin{aligned}V_m &= \mathbb{E}_{A_i}[\mathsf{Var}(Y_i \,|\, U, A_i)] - 1 \\ V_{m+k} &= \mathbb{E}_{A_i}[\mathsf{Var}(Y_i \,|\, U, U_k, A_i)] - 1.\end{aligned}$$

Accordingly, the expectation of the absolute difference can be upper bounded as

$$\begin{aligned}&\mathbb{E}[|V_{m+k} - V_m|] \\ &\quad = \mathbb{E}[|\mathbb{E}_{A_i}[\mathsf{Var}(Y_i \,|\, U, U_k, A_i) - \mathsf{Var}(Y_i \,|\, U, A_i)]|] \\ &\quad \overset{(a)}{\le} \mathbb{E}[|\mathsf{Var}(Y_i \,|\, U, U_k, A_i) - \mathsf{Var}(Y_i \,|\, U, A_i)|] \\ &\quad \overset{(b)}{\le} C \cdot \sqrt{\mathbb{E}[Y_i^4]\mathbb{E}\big[D_{\mathrm{KL}}(P_{Y_i|U,U_k,A_i} \,\|\, P_{Y_i|U,A_i})\big]} \\ &\quad \overset{(c)}{\le} C_B \cdot \sqrt{\mathbb{E}\big[D_{\mathrm{KL}}(P_{Y_i|U,U_k,A_i} \,\|\, P_{Y_i|U,A_i})\big]}, \qquad (129)\end{aligned}$$

where (a) follows from Jensen's inequality, (b) follows from Lemma 32 with $p = 1$, and (c) follows from (30).

Next, the expected Kullback–Leibler divergence can be expressed in terms of a conditional mutual information,

$$\begin{aligned}&\mathbb{E}\big[D_{\mathrm{KL}}(P_{Y_i|U,U_k,A_i} \,\|\, P_{Y_i|U,A_i})\big] \\ &\quad = I(Y_i; U_k \mid U, A_i) \\ &\quad \overset{(a)}{=} I(Y_i; Y_{m+1}^{m+k} \,|\, Y^m, A^{m+k}, A_i) \\ &\quad \overset{(b)}{=} h(Y_i \,|\, Y^m, A^m, A_i) - h(W_i) \\ &\qquad - h(Y_i \,|\, Y^{m+k}, A^{m+k}, A_i) + h(W_i) \\ &\quad \overset{(c)}{=} h(Y_{m+1} \mid Y^m, A^{m+1}) - h(W_{m+1}) \\ &\qquad - h(Y_{m+k+1} \,|\, Y^{m+k}, A^{m+k+1}) + h(W_{m+k}) \\ &\quad \overset{(d)}{=} I(X^n; Y_{m+1} \,|\, Y^m, A^{m+1}) \\ &\qquad - I(X^n; Y_{m+k+1} \,|\, Y^{m+k}, A^{m+k+1}) \\ &\quad \overset{(e)}{=} I'_{m,n} - I'_{m+k,n} \qquad (130)\end{aligned}$$

where (a) follows from the definitions of U and U_k and the fact that the measurements are independent of everything else, (b) follows from expanding the mutual information in terms of the differential entropy of Y_i, (c) follows from the fact that future measurements are identically distributed given the past, (d) follows from the fact that $h(W_m) = h(Y_m \,|\, X^n, A^m)$, and (e) follows from (15).

Combining (129) and (130), we see that the following inequality holds for all integers m and k,

$$\mathbb{E}[|V_{m,n} - V_{k,n}|] \le C_B \cdot \big|I'_{m,n} - I'_{k,n}\big|^{\frac{1}{2}}.$$

Moreover, we can now bound the deviation over ℓ measurements using

$$\begin{aligned}\frac{1}{\ell}\sum_{i=m}^{m+\ell-1} \mathbb{E}[|V_m - V_k|] &\le C_B \cdot \frac{1}{\ell}\sum_{k=m}^{m+\ell-1} |I'_m - I'_k|^{\frac{1}{2}} \\ &\le C_B \cdot \big|I'_m - I'_{m+\ell-1}\big|^{\frac{1}{2}},\end{aligned}$$

where the second inequality follows from the fact that $I'_{m,n}$ is non-increasing in m (see Section III-B). This completes the proof of Lemma 19.

H. Proof of Lemma 20

Starting with the triangle inequality, the sum of the posterior MI difference satisfies, for all $t \in \mathbb{R}$,

$$\bigg|\sum_{i=m}^{m+\ell-1} J_i - t\bigg| \le \bigg|\sum_{i=m}^{m+\ell-1} J_i - \mathcal{J}_i\bigg| + \bigg|\sum_{i=m}^{m+\ell-1} \mathcal{J}_i - t\bigg|.$$

Taking the expectation of both sides and minimizing over t leads to

$$\begin{aligned}\inf_{t \in \mathbb{R}} \mathbb{E}\bigg[\bigg|\sum_{i=m}^{m+\ell-1} J_i - t\bigg|\bigg] &\le \inf_{t \in \mathbb{R}} \mathbb{E}\bigg[\bigg|\sum_{i=m}^{m+\ell-1} \mathcal{J}_i - t\bigg|\bigg] \\ &\quad + \mathbb{E}\bigg[\bigg|\sum_{i=m}^{m+\ell-1} \mathcal{J}_i - J_i\bigg|\bigg]. \qquad (131)\end{aligned}$$

For the first term in (131), observe that

$$\begin{aligned}\inf_{t\in\mathbb{R}} \mathbb{E}\left[\left|\sum_{i=m}^{m+\ell-1} \mathcal{J}_i - t\right|\right] &\overset{(a)}{\le} \inf_{t\in\mathbb{R}} \sqrt{\mathbb{E}\left[\left|\sum_{i=m}^{m+\ell-1} \mathcal{J}_i - t\right|^2\right]} \\ &= \sqrt{\mathsf{Var}\left(\sum_{i=m}^{m+\ell-1} \mathcal{J}_i\right)}, \qquad (132)\end{aligned}$$

where (a) follows from Jensen's inequality. Furthermore, the variance obeys the upper bound

$$\begin{aligned}\mathsf{Var}\left(\sum_{i=m}^{m+\ell-1} \mathcal{J}_i\right) &= \mathsf{Var}\left(\sum_{i=0}^{m+\ell-1} \mathcal{J}_i - \sum_{i=0}^{m-1} \mathcal{J}_i\right) \\ &\overset{(a)}{\le} 2\,\mathsf{Var}\left(\sum_{i=0}^{m+\ell-1} \mathcal{J}_i\right) + 2\,\mathsf{Var}\left(\sum_{i=0}^{m-1} \mathcal{J}_i\right) \\ &\overset{(b)}{=} 2\,\mathsf{Var}\big(\imath\big(X^n; Y^{m+\ell} \mid A^{m+\ell}\big)\big) \\ &\quad + 2\,\mathsf{Var}(\imath(X^n; Y^m \mid A^m)) \\ &\overset{(c)}{\le} C_B \cdot \big(1 + \tfrac{m+\ell}{n}\big)^2 n + C_B \cdot \big(1 + \tfrac{m}{n}\big)^2 n \\ &\le C_B' \cdot \big(1 + \tfrac{m+\ell}{n}\big)^2 n\end{aligned}$$

where (a) follows from (62), (b) follows from the definition of $\mathcal{J}_m$, and (c) follows from Lemma 11. Plugging this bound back into (132) gives

$$\inf_{t\in\mathbb{R}} \mathbb{E}\left[\left|\sum_{i=m}^{m+\ell-1} \mathcal{J}_i - t\right|\right] \le C_B \cdot \left(1 + \frac{m+\ell}{n}\right)\sqrt{n}. \qquad (133)$$

Next, we consider the second term in (131). Note that $\mathcal{J}_m$ can be expressed explicitly as follows:

$$\begin{aligned}\mathcal{J}_m &= \log\left(\frac{f_{Y_{m+1}\mid X^n, Y^m, A^{m+1}}(Y_{m+1} \mid X^n, Y^m, A^{m+1})}{f_{Y_{m+1}\mid Y^m, A^{m+1}}(Y_{m+1} \mid Y^m, A^{m+1})}\right) \\ &= -\log\big(f_{Y_{m+1}\mid Y^m, A^{m+1}}(Y_{m+1} \mid Y^m, A^{m+1})\big) \\ &\quad - \frac{1}{2} W_{m+1}^2 - \frac{1}{2}\log(2\pi).\end{aligned}$$

To proceed, we define the random variables

$$\begin{aligned}\mathcal{H}_m &\triangleq -\log\big(f_{Y_{m+1}\mid Y^m, A^{m+1}}(Y_{m+1} \mid Y^m, A^{m+1})\big) \\ \widehat{\mathcal{H}}_m &\triangleq \mathbb{E}[\mathcal{H}_m \mid Y^m, A^m],\end{aligned}$$

and observe that

$$\begin{aligned}\mathcal{J}_m &= \mathcal{H}_m - \frac{1}{2} W_{m+1}^2 - \frac{1}{2}\log(2\pi) \\ J_m &= \widehat{\mathcal{H}}_m - \frac{1}{2} - \frac{1}{2}\log(2\pi).\end{aligned}$$

Using this notation, we can now write

$$\begin{aligned}&\mathbb{E}\left[\left|\sum_{i=m}^{m+\ell-1} \mathcal{J}_i - J_i\right|\right] \\ &\quad = \mathbb{E}\left[\left|\sum_{i=m}^{m+\ell-1} \mathcal{H}_i - \widehat{\mathcal{H}}_i + \frac{1}{2}(W_i^2 - 1)\right|\right] \\ &\quad \overset{(a)}{\le} \mathbb{E}\left[\left|\sum_{i=m}^{m+\ell-1} \mathcal{H}_i - \widehat{\mathcal{H}}_i\right|\right] + \mathbb{E}\left[\left|\sum_{i=m}^{m+\ell-1} \frac{1}{2}(W_i^2 - 1)\right|\right] \\ &\quad \overset{(b)}{\le} \sqrt{\mathbb{E}\left[\left(\sum_{i=m}^{m+\ell-1} \mathcal{H}_i - \widehat{\mathcal{H}}_i\right)^2\right]} \\ &\qquad + \sqrt{\mathbb{E}\left[\left(\sum_{i=m}^{m+\ell-1} \frac{1}{2}(W_i^2 - 1)\right)^2\right]}, \qquad (134)\end{aligned}$$

where (a) follows from the triangle inequality and (b) follows from Jensen's inequality. For the first term on the right-hand side, observe that square of the sum can be expanded as follows:

$$\begin{aligned}&\mathbb{E}\left[\left(\sum_{i=m}^{m+\ell-1} \mathcal{H}_i - \widehat{\mathcal{H}}_i\right)^2\right] \\ &\quad = \sum_{i=m}^{m+\ell-1} \mathbb{E}\left[\left(\mathcal{H}_i - \widehat{\mathcal{H}}_i\right)^2\right] \\ &\qquad + 2\sum_{i=m}^{m+\ell-1} \sum_{j=i+1}^{m+\ell-1} \mathbb{E}\Big[\Big(\mathcal{H}_i - \widehat{\mathcal{H}}_i\Big)\Big(\mathcal{H}_j - \widehat{\mathcal{H}}_j\Big)\Big]. \\ &\qquad\qquad (135)\end{aligned}$$

To deal with the first term on the right-hand side of (135), observe that

$$\begin{aligned}\mathbb{E}\left[\left(\mathcal{H}_i - \widehat{\mathcal{H}}_i\right)^2\right] &\overset{(a)}{=} \mathbb{E}\big[\mathsf{Var}(\mathcal{H}_i \mid Y^i, A^i)\big] \\ &\overset{(b)}{\le} \mathbb{E}[\mathsf{Var}(\mathcal{H}_i)] \\ &\le \mathbb{E}\left[\left(\mathcal{H}_i - \frac{1}{2}\log(2\pi)\right)^2\right],\end{aligned}$$

where (a) follows from the definition of $\widehat{\mathcal{H}}_i$ and (b) follows from the law of total variance (68). To bound the remaining term, let $U = (Y^m, A^m)$ and let $\widetilde{X}_u^n$ be drawn according to the posterior distribution of X^n given $U = u$. Then, the density of Y_{m+1} given (U, A_{m+1}) can be bounded as follows:

$$\begin{aligned}\frac{1}{\sqrt{2\pi}} &\ge f_{Y_{m+1}\mid U, A_{m+1}}(y_{m+1} \mid u, a_{m+1}) \\ &= \mathbb{E}_{\tilde{X}_u^n}\left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_{m+1} - \langle a_{m+1}, \widetilde{X}_u^n\rangle)^2\right)\right] \\ &\overset{(a)}{\ge} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\mathbb{E}_{\tilde{X}_u^n}\left[(y_{m+1} - \langle a_{m+1}, \widetilde{X}_u^n\rangle)^2\right]\right) \\ &\overset{(b)}{\ge} \frac{1}{\sqrt{2\pi}} \exp\Big(-y_{m+1}^2 - \mathbb{E}_{\tilde{X}_u^n}\Big[(\langle a_{m+1}, \widetilde{X}_u^n\rangle)^2\Big]\Big),\end{aligned}$$

where (a) follows from Jensen's inequality and the convexity of the exponential and (b) follows from (62). Using these bounds, we obtain

$$\begin{aligned}&\mathbb{E}\left[\left(\mathcal{H}_i - \frac{1}{2}\log(2\pi)\right)^2\right] \\ &\quad \le \mathbb{E}\left[\left(Y_{m+1}^2 + \mathbb{E}_{\tilde{X}_U^n}\left[\left(\langle A_{m+1}, \widetilde{X}_U^n\rangle\right)^2\right]\right)^2\right] \\ &\quad \overset{(a)}{\le} 2\mathbb{E}\big[Y_{m+1}^4\big] + 2\mathbb{E}\left[\left(\mathbb{E}_{\tilde{X}_U^n}\left[\left(\langle A_{m+1}, \widetilde{X}_U^n\rangle\right)^2\right]\right)^2\right] \\ &\quad \overset{(b)}{\le} 2\mathbb{E}\big[Y_{m+1}^4\big] + 2\mathbb{E}\Big[(\langle A_{m+1}, X^n\rangle)^4\Big]\end{aligned}$$

$$\leq 4\mathbb{E}\big[Y_{m+1}^4\big] \overset{(c)}{\leq} C_B,$$

where (a) follows from (62), (b) follows from Jensen's inequality and the fact that $\widetilde{X}_U^n$ as the same distribution as X^n and is independent of A_{m+1}, and (c) follows from (30).

To deal with the second term on the right-hand side of (135), note that $\mathcal{H}_i$ and $\widehat{\mathcal{H}}_i$ are determined by (Y^{i+1}, A^{i+1}), and thus, for all $j > i$,

$$\mathbb{E}\Big[\Big(\mathcal{H}_i - \widehat{\mathcal{H}}_i\Big)\Big(\mathcal{H}_j - \widehat{\mathcal{H}}_j\Big) \mid Y^{i+1}, A^{i+1}\Big] = \Big(\mathcal{H}_i - \widehat{\mathcal{H}}_i\Big)\mathbb{E}\Big[\Big(\mathcal{H}_j - \widehat{\mathcal{H}}_j\Big) \mid Y^{i+1}, A^{i+1}\Big] = 0.$$

Consequently, the cross terms in the expansion of (135) are equal to zero, and the first term on the right-hand side of (134) obeys the upper bound

$$\sqrt{\mathbb{E}\left[\left(\sum_{i=m}^{m+\ell-1} \mathcal{H}_i - \widehat{\mathcal{H}}_i\right)^2\right]} \leq C_B \cdot \sqrt{\ell}. \tag{136}$$

As for the second term on the right-hand side of (134), note that $\sum_{i=m}^{m+\ell-1} W_i^2$ is chi-squared with ℓ degrees of freedom, and thus

$$\sqrt{\mathbb{E}\left[\left(\sum_{i=m}^{m+\ell-1} \frac{1}{2}(W_i^2 - 1)\right)^2\right]} = \sqrt{\frac{\ell}{2}}. \tag{137}$$

Plugging (136) and (137) back in to (134) leads to

$$\mathbb{E}\left[\left|\sum_{i=m}^{m+\ell-1} \mathcal{J}_i - J_i\right|\right] \leq C_B \cdot \sqrt{\ell}.$$

Finally, combining this inequality with (131) and (133) gives

$$\inf_{t\in\mathbb{R}} \mathbb{E}\left[\left|\sum_{i=m}^{m+\ell-1} J_i - t\right|\right] \leq C_B \cdot \left(\left(1 + \frac{m+\ell}{n}\right)\sqrt{n} + \sqrt{\ell}\right) \leq C_B' \cdot \left(\left(1 + \frac{m}{n}\right)\sqrt{n} + \frac{\ell}{\sqrt{n}}\right),$$

where the last step follows from keeping only the dominant terms. This completes the proof of Lemma 20.

I. *Proof of Lemma 21*

Fix any $(m, n, \ell) \in \mathbb{N}^3$. We begin with the following decomposition, which follows from the triangle inequality:

$$\begin{aligned}
&\inf_{t\in\mathbb{R}} \mathbb{E}\left[\frac{1}{2}\log(1 + V_m) - t\right] \\
&\quad\leq \mathbb{E}\left[\left|\frac{1}{2}\log(1 + V_m) - \frac{1}{\ell}\sum_{k=m}^{m+\ell+1} \frac{1}{2}\log(1 + V_k)\right|\right] \\
&\quad\quad + \mathbb{E}\left[\left|\frac{1}{\ell}\sum_{k=m}^{m+k+1} \frac{1}{2}\log(1 + V_k) - \frac{1}{k}\sum_{i=m}^{m+\ell+1} J_k\right|\right] \\
&\quad\quad + \inf_{t\in\mathbb{R}} \mathbb{E}\left[\left|\frac{1}{\ell}\sum_{i=m}^{m+\ell+1} J_k - t\right|\right].
\end{aligned} \tag{138}$$

The first term on the right-hand side of (138) can be bounded in terms of the smoothness of the mutual information given in Lemma 19. We use the following chain of inequalities:

$$\begin{aligned}
&\mathbb{E}\left[\left|\log(1 + V_m) - \frac{1}{\ell}\sum_{k=m}^{m+\ell-1} \log(1 + V_k)\right|\right] \\
&\quad\overset{(a)}{\leq} \frac{1}{\ell}\sum_{k=m}^{m+\ell-1} \mathbb{E}[|\log(1 + V_m) - \log(1 + V_k)|] \\
&\quad\overset{(b)}{\leq} \frac{1}{\ell}\sum_{k=m}^{m+\ell-1} \mathbb{E}[|V_m - V_k|] \\
&\quad\overset{(c)}{\leq} C_B \cdot \big|I_m' - I_{m+\ell-1}'\big|^{\frac{1}{2}}
\end{aligned} \tag{139}$$

where (a) follows from Jensen's inequality, (b) follows from the fact that the mapping $\log(1 + x) \to x$ is one-Lipschitz on $\mathbb{R}_+$, and (c) follows from Lemma 19.

The second term on the right-hand side of (138) is bounded by the relationship between the posterior variance and posterior mutual information difference:

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{\ell}\sum_{k=m}^{m+\ell-1} \frac{1}{2}\log(1 + V_k) - \frac{1}{\ell}\sum_{k=m}^{m+\ell-1} J_k\right|\right] \\
&\quad\overset{(a)}{=} \frac{1}{\ell}\sum_{k=m}^{m+\ell-1} \mathbb{E}\big[\Delta_k^P\big] \\
&\quad\overset{(b)}{\leq} C_B \cdot \frac{1}{\ell}\sum_{k=m}^{m+\ell-1} |I_k''|^{\frac{1}{10}} \\
&\quad\overset{(c)}{\leq} C_B \cdot \left|\frac{1}{\ell}\sum_{k=m}^{m+\ell-1} |I_k''|\right|^{\frac{1}{10}} \\
&\quad= C_B \cdot \left|\frac{1}{k}(I_{m+k}' - I_m')\right|^{\frac{1}{10}} \\
&\quad\overset{(d)}{\leq} C_B' \cdot \ell^{-\frac{1}{10}},
\end{aligned} \tag{140}$$

where (a) follows from Identity (40), (b) follows from Lemma 17, (c) follows from Jensen's inequality and the non-positiviity of I_m'' and (d) follows from the fact that I_m' is bounded by a constant that depends only on B (see Section III-B).

Finally, the third term on the right-hand side of (138) is bounded by Lemma 20. Plugging (139) and (140) back into (138) leads to

$$\begin{aligned}
\inf_{t\in\mathbb{R}} \mathbb{E}\left[\frac{1}{2}\log(1 + V_m) - t\right] &\leq C_B \cdot \big|I_m' - I_{m+\ell-1}'\big|^{\frac{1}{2}} \\
&\quad + C_B \cdot \left[\ell^{-\frac{1}{10}} + \left(1 + \frac{m}{n}\right)\frac{\sqrt{n}}{\ell} + \frac{1}{\sqrt{n}}\right].
\end{aligned}$$

Combining this inequality with Lemma 18 completes the proof of Lemma 21.

J. *Proof of Lemma 22*

Fix any $(m, n, \ell) \in \mathbb{N}^3$. Combining Identity (42), with Lemmas 17 and 21 yields

$$\Delta_{m,n} = \mathbb{E}\big[\Delta_{m,n}^P\big] + \frac{1}{2}\mathbb{E}\left[\log\left(\frac{1 + M_{m,n}}{1 + V_{m,n}}\right)\right]$$

$$\le C_B \cdot \Big[\big| I''_{m,n} \big|^{\frac{1}{10}} + \big| I'_{m,n} - I'_{m+\ell+1,n} \big|^{\frac{1}{4}} \\ + \ell^{-\frac{1}{20}} + \big(1 + \tfrac{m}{n}\big)^{\frac{1}{2}} n^{\frac{1}{4}} \ell^{-\frac{1}{2}} + n^{-\frac{1}{4}} \Big]. \quad (141)$$

For the specific choice of $\ell = \lceil n^{\frac{5}{6}} \rceil$, we have

$$\ell^{-\frac{1}{20}} + \big(1 + \tfrac{m}{n}\big)^{\frac{1}{2}} n^{\frac{1}{4}} \ell^{-\frac{1}{2}} \le n^{-\frac{1}{24}} + \big(1 + \tfrac{m}{n}\big)^{\frac{1}{2}} n^{-\frac{1}{6}} \\ \le 2\big(1 + \tfrac{m}{n}\big)^{\frac{1}{2}} n^{-\frac{1}{24}}.$$

Plugging this inequality back into (141) completes the proof of Lemma 22.

Appendix D
Proofs of Results in Section V

A. *Proof of Lemma 23*

Recall that M_m is the expectation of the posterior variance V_m. Therefore, the difference between M_{m+1} and M_m can bounded as follows:

$$\begin{aligned} |M_{m+1} - M_m| &= |\mathbb{E}[V_{m+1} - V_m]| \\ &\overset{(a)}{\le} \mathbb{E}[|V_{m+1} - V_m|] \\ &\overset{(b)}{\le} C_B \cdot \sqrt{I''_m}, \end{aligned} \quad (142)$$

where (a) follows from Jensen's inequality and (b) follows from Lemma 19. Combining (142) with the sandwiching relation (49) leads to (50).

B. *Proof of Lemma 24*

Let Q be a random matrix distributed uniformly on the set of $(m+1)\times(m+1)$ orthogonal matrices and define the rotated augmented measurements:

$$\widetilde{Y}^{m+1} = Q \begin{bmatrix} Y^m \\ Z_{m+1} \end{bmatrix}, \qquad \widetilde{A}^{m+1} = Q \begin{bmatrix} A^m & \mathbf{0}_{m\times 1} \\ A_{m+1} & \sqrt{G_{m+1}} \end{bmatrix}.$$

Since multiplication by Q is a one-to-one transformation, the augmented MMSE can be expressed equivalently in terms of the rotated measurements:

$$\begin{aligned} \widetilde{M}_m &\triangleq \frac{1}{n} \mathsf{mmse}(X^n \mid Y^m, A^m, \mathcal{D}_{m+1}) \\ &= \frac{1}{n} \mathsf{mmse}(X^n \mid \widetilde{Y}^{m+1}, \widetilde{A}^{m+1}). \end{aligned} \quad (143)$$

Lemma 35. *The entries of the $(m+1)\times(n+1)$ random matrix $\widetilde{A}^{m+1}$ are i.i.d. Gaussian $\mathcal{N}(0, 1/n)$.*

Proof. The first n columns are i.i.d. Gaussian $\mathcal{N}(0, \frac{1}{n} I_{m+1})$ and independent of Q because of the rotational invariance of the i.i.d. Gaussian distribution on A^{m+1}. The last column of $\widetilde{A}^{m+1}$ is equal to the product of $\sqrt{G_{m+1}}$ and the last column of Q. Since G_{m+1} is proportional to a chi random variable with $m+1$ degrees of freedom and Q is distributed uniformly on the Euclidean sphere of radius one, the last column is also Gaussian $\mathcal{N}(0, \frac{1}{n} I_{m+1})$; see e.g. [56, Theorem 2.3.18]. □

The key takeaway from Lemma 35, is that the distribution on the columns of $\widetilde{A}^{m+1}$ is permutation invariant. Since the distribution on the entries of X^{n+1} is also permutation invariant, this means that the MMSEs of the signal entries are identical, i.e.,

$$\mathsf{mmse}(X_i \mid \widetilde{Y}^{m+1}, \widetilde{A}^{m+1}) = \mathsf{mmse}(X_j \mid \widetilde{Y}^{m+1}, \widetilde{A}^{m+1}),$$

for all $i, j \in [n+1]$. Combining this fact with (143), we see that the augmented MMSE can be expressed equivalently as

$$\begin{aligned} \widetilde{M}_m &= \mathsf{mmse}(X_{n+1} \mid \widetilde{Y}^{m+1}, \widetilde{A}^{m+1}) \\ &= \mathsf{mmse}(X_{n+1} \mid Y^m, A^m, \mathcal{D}_{m+1}), \end{aligned}$$

where the last step follows, again, from the fact that multiplication by Q is a one-to-one transformation of the data. This completes the proof of Lemma 24.

C. *Proof of Lemma 25*

This proof is broken into two steps. First, we show that the augmented MMSE satisfies the inequality,

$$\left| \widetilde{M}_m - \mathbb{E}\left[\mathsf{mmse}_X\left(\frac{G_{m+1}}{1 + M_m} \right) \right] \right| \le C_B \cdot \sqrt{\Delta_m}. \quad (144)$$

Then, we use the smoothness of the of the single-letter MMSE function $\mathsf{mmse}_X(s)$ to show that

$$\left| \mathbb{E}\left[\mathsf{mmse}_X\left(\frac{G_{m+1}}{1 + M_m} \right) \right] - \mathsf{mmse}_X\left(\frac{m/n}{1 + M_m} \right) \right| \\ \le C_B \frac{1 + \sqrt{m}}{n}. \quad (145)$$

The proofs of Inequalities (144) and (145) are given in the following subsections.

1) Proof of Inequality (144)*:* The centered augmented measurement $\bar{Z}_{m+1}$ is defined by

$$\bar{Z}_{m+1} = Z_{m+1} - \mathbb{E}\big[Z_{m+1} \mid Y^m, A^{m+1}\big].$$

Since G_{m+1} and X_{n+1} are independent of the first m measurements, $\bar{Z}_{m+1}$ can also be expressed as

$$\bar{Z}_{m+1} = \sqrt{G_{m+1}} X_{n+1} + \bar{Y}_{m+1},$$

where $\bar{Y}_{m+1} = Y_{m+1} - \mathbb{E}\big[Y_{m+1} \mid Y^m, A^{m+1}\big]$ is the centered measurement introduced in Section IV-C. Starting with (51), we see that the augmented MMSE can expressed as

$$\widetilde{M}_m = \mathsf{mmse}(X_{n+1} \,|\, Y^m, A^m, \bar{Z}_{m+1}, A_{m+1}, G_{m+1}), \quad (146)$$

where we have used the fact that there is a one-to-one mapping between $\bar{Z}_{m+1}$ and Z_{m+1}.

The next step of the proof is to address the extent to which the MMSE in (146) would differ if the 'noise' term $\bar{Y}_{m+1}$ were replaced by an independent Gaussian random variable with the same mean and variance. To make this comparison precise, recall that $\mathbb{E}\big[\bar{Y}_{m+1}\big] = 0$ and $\mathsf{Var}(\bar{Y}_{m+1}) = 1 + M_m$, and let Z^*_{m+1} be defined by

$$Z^*_{m+1} = \sqrt{G_{m+1}} X_{n+1} + Y^*_{m+1},$$

where $Y^*_{m+1} \sim \mathcal{N}(0, 1 + M_m)$ is independent of everything else. Note that the MMSE of X_{n+1} with $\bar{Z}_{m+1}$ replaced by Z^*_{m+1} can be characterized explicitly in terms of the single-letter MMSE function:

$$\mathsf{mmse}(X_{n+1} \mid Y^m, A^m, Z^*_{m+1}, A_{m+1}, G_{m+1})$$

$$\begin{aligned}&\overset{(a)}{=} \mathsf{mmse}(X_{n+1} \mid Z^*_{m+1}, G_{m+1})\\ &= \mathbb{E}\left[\mathsf{mmse}_X\left(\frac{G_{m+1}}{1+M_m}\right)\right] \end{aligned} \tag{147}$$

where (a) follows from the fact that (Y^m, A^{m+1}) is independent of $(X_{n+1}, Z^*_{m+1}, G_{m+1})$.

The next step is to bound the difference between (146) and (147). To proceed, we introduce the notation

$$\begin{aligned}\mathcal{F} &= (Y^m, A^m, \bar{Z}_{m+1}, A_{m+1}, G_{m+1})\\ \mathcal{F}^* &= (Y^m, A^m, Z^*_{m+1}, A_{m+1}, G_{m+1}).\end{aligned}$$

Then, using Lemma 34 yields

$$\begin{aligned}&|\mathsf{mmse}(X_{n+1} \mid \mathcal{F}) - \mathsf{mmse}(X_{n+1} \mid \mathcal{F}^*)|\\ &\quad \le 2^{\frac{5}{2}}\sqrt{\mathbb{E}\left[X^4_{n+1}\right] D_{\mathrm{KL}}\left(P_{\mathcal{F}, X_{n+1}} \,\middle\|\, P_{\mathcal{F}^*, X_{n+1}}\right)}.\end{aligned}$$

By Assumption 2, the fourth moment of X_{n+1} is upper bounded by B. The last step is to show that the Kullback–Leibler divergence is equal to the non-Gaussianenesss Δ_m. To this end, observe that

$$\begin{aligned}&D_{\mathrm{KL}}\left(P_{\mathcal{F}, X_{n+1}} \,\middle\|\, P_{\mathcal{F}^*, X_{n+1}}\right)\\ &\quad\overset{(a)}{=} \mathbb{E}_{X_{n+1}}\left[D_{\mathrm{KL}}\left(P_{\mathcal{F} \mid X_{n+1}} \,\middle\|\, P_{\mathcal{F}^* \mid X_{n+1}}\right)\right]\\ &\quad\overset{(b)}{=} D_{\mathrm{KL}}\left(P_{\bar{Y}_{m+1}, Y^m, A^{m+1}} \,\middle\|\, P_{Y^*_{m+1}, Y^m, A^{m+1}}\right)\\ &\quad\overset{(c)}{=} \mathbb{E}_{Y^m, A^{m+1}}\left[D_{\mathrm{KL}}\left(P_{\bar{Y}_{m+1} \mid Y^m, A^{m+1}} \,\middle\|\, P_{Y^*_{m+1}}\right)\right]\\ &\quad= \Delta_m,\end{aligned}$$

where (a) follows from the chain rule for Kullback–Leibler divergence, (b) follows from the fact that both $\bar{Y}_{m+1}$ and Y^*_{m+1} are independent of (X_{n+1}, G_{m+1}), and (c) follows from the chain rule for Kullback–Leibler divergence and the fact that Y^*_{m+1} is independent of (Y^m, A^{m+1}). This completes the proof of Inequality (144).

2) Proof of Inequality (145)*:* Observe that

$$\begin{aligned}&\left|\mathbb{E}\left[\mathsf{mmse}_X\left(\frac{G_{m+1}}{1+M_m}\right)\right] - \mathsf{mmse}_X\left(\frac{m/n}{1+M_m}\right)\right|\\ &\quad\overset{(a)}{\le} \mathbb{E}\left[\left|\mathsf{mmse}_X\left(\frac{G_{m+1}}{1+M_m}\right) - \mathsf{mmse}_X\left(\frac{m/n}{1+M_m}\right)\right|\right]\\ &\quad\overset{(b)}{\le} 4B\mathbb{E}\left[\left|G_{m+1} - \frac{m}{n}\right|\right]\\ &\quad\overset{(c)}{\le} 4B\left(\mathbb{E}[|G_{m+1} - \mathbb{E}[G_{m+1}]|] + \frac{1}{n}\right)\\ &\quad\overset{(d)}{\le} 4B\left(\sqrt{\mathsf{Var}(G_{m+1})} + \frac{1}{n}\right)\\ &\quad= 4B\left(\frac{\sqrt{2(m+1)}}{n} + \frac{1}{n}\right),\end{aligned}$$

where (a) follows from Jensen's inequality, (b) follows from Lemma 5 and Assumption 2, (c) follows from the triangle inequality and the fact that $\mathbb{E}[G_{m+1}] = \frac{m+1}{n}$, and (d) follows from Jensen's inequality.

Appendix E
Derivation of Replica-Symmetric Formulas

Guo and Verdú [1] studied the random linear estimation problem in the context of randomly spread code-division multiple access (CDMA) and multiuser detection. Similar to the problem formulation in this paper, their analysis focusses on the reconstruction of a random vector X^n with i.i.d. entries from a model of the form (1) where A^m is a random matrix with independent entries and W^n is i.i.d. Gaussian noise. Their assumptions on the matrix A^m are more general than those made in this paper in the sense that they do not require A^m to be Gaussian and they allow for the possibility that different signal entries have different signal-to-noise ratios. Specifically, the assume that

$$A^m = \frac{1}{\sqrt{m}} S^m \operatorname{diag}(\sqrt{\mathsf{snr}_1}, \dots, \sqrt{\mathsf{snr}_n})$$

where S^m has entries that are i.i.d. with mean zero and variance one and snr_i is the signal-to-noise ratio associated with the i-th signal entry. The setting in this paper corresponds to the special case where the the entries of S^m are Gaussian and $\mathsf{snr}_i = m/n$ for all i.

Using the replica method, Guo and Verdú derived conjectured formulas for the asymptotic mean-squared error (MSE) of an an estimation procedure that computes the conditional expectation of X^n with respect to a postulated i.i.d. distribution Q_X that is possibility different from the true distribution P_X. The formulas corresponding to an arbitrary pair (P_X, Q_X) are given in [1, Claim 1].

The Bayes-optimal setting occurs when the postulated prior is equal to the true signal prior. In this setting, the corresponding corresponding MSE is referred to as the minimum mean-square error (MMSE), following from the well known fact that the conditional expectation (associated with the true prior) is the estimator the minimizes the MSE. In [1, Claims 2–3], Guo and Verdú specialize their results to the Bayes optimal setting and provide explicit formulas for the mutual information and the MMSE. These formulas a parametrized in terms of a an effective signal-to-noise ratio $\eta \in [0,1]$, which is referred to as the multiuser efficiency. The conjectured limits for the multiuser efficiency η and the normalized mutual information $\frac{1}{m} I(X^n; Y^n, A^n)$, which is referred to as to the spectral efficiency under joint decoding, are given by the minimizer and minimum, respectively, of the right-hand side of [1, Equation (38)]. The relationship between the multiuser efficiency and the conjectured limit of the MMSE is described in Claim 2 of the their paper.

The replica-MI and replica-MMSE functions introduced in Definition 1 are obtained by specializing the results of Guo and Verdú to the setting studied in this paper. In particular, letting $\mathsf{snr}_i = \delta$ for all $i = 1, 2, \dots$ and normalizing the mutual information by number of signal entries instead of the number of measurements yields

$$\mathcal{I}_{\mathrm{RS}}(\delta) = \min_{\eta \in (0,1]}\left\{I_X(\eta\,\delta) + \frac{\delta}{2}\left[\log\left(\frac{1}{\eta}\right) + \eta - 1\right]\right\}, \tag{148}$$

$$\mathcal{M}_{\mathrm{RS}}(\delta) = \mathsf{mmse}_X(\eta^*\,\delta), \tag{149}$$

where η^* is the minimizer in (148). For the purposes of this paper, we find it convenient to consider slightly different parameterization of these formulas. Making the change of variables $z = 1/\eta - 1$ and recalling the definition of $R(\delta, z)$ in (2), leads to

$$\mathcal{I}_{\mathrm{RS}}(\delta) = \min_{z \in [0,\infty)} R(\delta, z), \qquad (150)$$

$$\mathcal{M}_{\mathrm{RS}}(\delta) = \mathsf{mmse}_X(\delta/(1+z^*)), \qquad (151)$$

where z^* is the minimizer of $R(\delta, z)$. Finally, by the I-MMSE relationship, one sees that every stationary point of the mapping $z \mapsto R(\delta, z)$ satisfies the fixed-point equation $z = \mathsf{mmse}_X(\delta/(1+z))$. Because the minimizer of $R(\delta, z)$ with respect to z is a stationary point, it follows that the the right-hand side of (151) is equal to z^*. Therefore, the expression for $\mathcal{M}_{\mathrm{RS}}(\delta)$ given in Definition 1 is equivalent to (151).

Acknowledgment

G. Reeves is grateful to D. Donoho for many helpful conversations as well as the inspiration behind the sandwiching argument that lead to the fixed-point relationships in Section V-A.

References

[1] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Transactions on Information Theory*, vol. 51, pp. 1983–2010, June 2005.

[2] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Transactions on Information Theory*, vol. 48, pp. 2888–2910, Nov. 2002.

[3] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on l_p-norm minimization," *Journal of Statistical Mechanics: Theory and Experiment*, 2009.

[4] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, (Monticello, IL), Oct. 2009.

[5] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of map estimation via the replica method and applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, pp. 1902 – 1923, Mar. 2012.

[6] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, pp. 3065–3092, May 2012.

[7] G. Reeves and M. Gastpar, "Compressed sensing phase transitions: Rigorous bounds versus replica predictions," in *Proceedings of the Conference on Information Sciences and Systems (CISS)*, (Princeton, NJ), Mar. 2012.

[8] A. Tulino, G. Caire, S. Verdú, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Transactions on Information Theory*, vol. 59, pp. 4243–4271, July 2013.

[9] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *Proceedings of the IEEE Information Theory Workshop (ITW)*, (Chengdu, China), pp. 194–198, Oct. 2006.

[10] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 18914–18919, Nov. 2009.

[11] D. L. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, pp. 6920–6941, Oct. 2011.

[12] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, pp. 764–785, Feb. 2011.

[13] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and iterative algorithms," in *IEEE International Symposium on Information Theory*, (Boston, MA), July 2012.

[14] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Oxford Science Publications, 2001.

[15] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Physical Review X*, vol. 2, May 2012.

[16] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, pp. 965–974, 1975.

[17] M. Mézard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[18] R. Muller, "Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2821–2828, Nov. 2003.

[19] L. Zdeborová and F. Krzakala, "Statistical physics of inference: Thresholds and algorithms," *Advances in Physics*, vol. 65, no. 5, pp. 453–552, 2016.

[20] A. Bereyhi, R. R. Müller, and H. Schulz-Baldes, "RSB decoupling property of MAP estimators," in *Proceedings of the IEEE Information Theory Workshop (ITW)*, (Cambridge, UK), Sept. 2016.

[21] A. Bereyhi, R. R. Müller, and H. Schulz-Baldes, "Statistical mechanics of MAP estimation: General replica ansatz," Oct. 2017. [Online]. Available https://arxiv.org/pdf/1612.01980.pdf.

[22] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proceedings of the IEEE Information Theory Workshop (ITW)*, (Punta del Este, Uruguay), pp. 160–164, 2006.

[23] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 269–280, 2010.

[24] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, (Monticello, IL), Sept. 2010.

[25] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Transactions on Information Theory*, vol. 59, pp. 7434–7464, July 2013.

[26] S. Verdú and S. Shamai, "Spectral efficiency of cdma with random spreading," *IEEE Transactions on Information Theory*, vol. 45, pp. 622–640, Mar. 1999.

[27] D. N. C. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwith and user capacity," *IEEE Transactions on Information Theory*, vol. 45, pp. 641–657, Mar. 1999.

[28] S. B. Korada and N. Macris, "Tight bounds on the capicty of binary input random CDMA systems," *IEEE Transactions on Information Theory*, vol. 56, pp. 5590–5613, Nov. 2010.

[29] W. Huleihel and N. Merhav, "Asymptotic MMSE analysis under sparse representation modeling," *Signal Processing*, vol. 131, pp. 320–332, 2017.

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

[31] N. Merhav, "Optimum estimation via gradients of partition functions and information measures: A statistical-mechanical perspective," *IEEE Transactions on Information Theory*, vol. 57, pp. 3887–3898, June 2011.

[32] G. Reeves, "Understanding the MMSE of compressed sensing one measurement at a time." Presented at the Institut Henri Poincaré Spring 2016 Thematic Program on the Nexus of Information and Computation Theories, Mar. 2016. [Online]. Available: https://youtu.be/vmd8-CMv04I.

[33] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, (Barcelona, Spain), pp. 665 – 669, July 2016.

[34] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The mutual information in random linear estimation," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, (Monticello, IL), 2016.

[35] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 1261–1282, Apr. 2005.

[36] P. Milgrom and I. Segal, "Envelope theorems for arbitrary choice sets," *Econometrica*, vol. 70, pp. 583–601, Mar. 2002.

[37] D. Guo, Y. Wu, S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Transactions on Information Theory*, vol. 57, pp. 2371–2385, Apr. 2011.

[38] Y. Wu and S. Verdú, "MMSE dimension," *IEEE Transactions on Information Theory*, vol. 57, pp. 4857 – 4879, Aug. 2011.

[39] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Transactions on Information Theory*, vol. 58, pp. 1289–1301, Mar. 2012.

[40] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, pp. 101–112, June 1959.

[41] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Transactions on Information Theory*, vol. 57, pp. 33–55, Jan. 2011.

[42] G. Reeves, "Conditional central limit theorems for Gaussian projections," Dec. 2016. [Online]. Available: https://arxiv.org/abs/1612.09252.

[43] R. M. Gray, *Entropy and Information Theory*. Springer-Verlag, 2013.

[44] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, 2nd ed., 1968.

[45] H. Hanche-Olsen and H. Holden, "The Kolmogorov–Riesz compactness theorem," *Expositiones Mathematicae*, vol. 28, no. 4, pp. 385–394, 2010.

[46] W. Rudin, *Principles of Mathematical Analysis (International Series in Pure & Applied Mathematics)*. New York, NY: McGraw-Hill, 1976.

[47] H. Royden and P. Fitzpatrick, *Real Analysis*. Boston, MA: Prentice Hall, 4th ed., 2010.

[48] C. Niculescu and L.-E. Persson, *Convex functions and their applications: A contemporary approach*. Springer Science & Business Media, 2009.

[49] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. New York: Wiley, 2009.

[50] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, pp. 6241 – 6263, Oct. 2012.

[51] G. Reeves and M. Gastpar, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," *IEEE Transactions on Information Theory*, vol. 59, pp. 3451–3465, June 2013.

[52] G. Reeves and D. L. Donoho, "The minimax noise sensitivity in compressed sensing," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, (Istanbul, Turkey), pp. 116–120, July 2013.

[53] G. Reeves, "The fundamental limits of stable recovery in compressed sensing," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, (Honolulu, HI), pp. 994 – 998, July 2014.

[54] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[55] D. Pollard, *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, 2002.

[56] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Monographs and Surveys in Pure and Applied Mathematics, Chapman and Hall/CRC, 1999.

[57] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, pp. 141–154, Jan. 2006.

[58] R. Esposito, "On a relation between detection and estimation in decision theory," *Information and Control*, vol. 12, no. 2, pp. 116–120, 1968.

Galen Reeves (S'07–M'11) received the B.S. degree in Electrical and Computer Engineering from Cornell University in 2005 and the M.S. and Ph.D. degrees in Electrical Engineering and Computer Sciences from the University of California at Berkeley in 2007 and 2011 respectively. From 2011 to 2013 he was a postdoctoral associate in the Departments of Statistics at Stanford University. He is currently an assistant professor at Duke University with a joint appointment in the Department of Electrical and Computer Engineering and the Department of Statistical Science. His research interests include information theory, high-dimensional statistics, and machine learning.

Dr. Reeves received the NSF CAREER award in 2017. He was the General Chair of the 2015 Duke Workshop on Sensing and Analysis of High-Dimensional Data and the Co-Chair of the 2016 North American School of Information Theory.

Henry D. Pfister (S'99–M'03–SM'09) received his Ph.D. in electrical engineering in 2003 from the University of California, San Diego and is currently an associate professor in the Electrical and Computer Engineering Department of Duke University with a secondary appointment in Mathematics. Prior to that, he was a professor at Texas A&M University (2006-2014), a post-doctoral fellow at the École Polytechnique Fédérale de Lausanne (2005-2006), and a senior engineer at Qualcomm Corporate R&D in San Diego (2003-2004). His current research interests include information theory, error-correcting codes, quantum computing, and machine learning.

He received the NSF Career Award in 2008 and a Texas A&M ECE Department Outstanding Professor Award in 2010. He is a coauthor of the 2007 IEEE COMSOC best paper in Signal Processing and Coding for Data Storage and a coauthor of a 2016 Symposium on the Theory of Computing (STOC) best paper. He has served the IEEE Information Theory Society as a member of the Board of Governors (2019-2022), an Associate Editor for the IEEE Transactions on Information Theory (2013-2016), and a Distinguished Lecturer (2015-2016). He was also the General Chair of the 2016 North American School of Information Theory.