Asymptotic MMSE Analysis Under Sparse Representation Modeling*

Wasim Huleihel and Neri Merhav
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL
E-mail: {wh@tx, merhav@ee}.technion.ac.il

Abstract

Compressed sensing is a signal processing technique in which data is acquired directly in a compressed form. There are two modeling approaches that can be considered: the worst-case (Hamming) approach and a statistical mechanism, in which the signals are modeled as random processes rather than as individual sequences. In this paper, the second approach is studied. Accordingly, we consider a model of the form $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}$, where each comportment of $\boldsymbol{X}$ is given by $\boldsymbol{X}_i = \boldsymbol{S}_i\boldsymbol{U}_i$, where $\{\boldsymbol{U}_i\}$ are i.i.d. Gaussian random variables, and $\{\boldsymbol{S}_i\}$ are binary random variables independent of $\{\boldsymbol{U}_i\}$, and not necessarily independent and identically distributed (i.i.d.), $\boldsymbol{H} \in \mathbb{R}^{k\times n}$ is a random matrix with i.i.d. entries, and $\boldsymbol{W}$ is white Gaussian noise. Using a direct relationship between optimum estimation and certain partition functions, and by invoking methods from statistical mechanics and from random matrix theory (RMT), we derive an asymptotic formula for the minimum mean-square error (MMSE) of estimating the input vector $\boldsymbol{X}$ given $\boldsymbol{Y}$ and $\boldsymbol{H}$, as $k, n \to \infty$, keeping the measurement rate, $R = k/n$, fixed. In contrast to previous derivations, which are based on the replica method, the analysis carried in this paper is rigorous.

Index Terms

Compressed Sensing (CS), minimum mean-square error (MMSE), partition function, statistical-mechanics, replica method, conditional mean estimation, phase transitions, threshold effect, random matrix.

*This research was partially supported by The Israeli Science Foundation (ISF), grant no. 412/12.

I. Introduction

Compressed sensing [1, 2] is a signal processing technique that compresses analog vectors by means of a linear transformation. Using some prior knowledge on the signal *sparsity*, and by designing efficient "encoders" and "decoders", the goal is to achieve effective compression in the sense of taking a much smaller number of measurements than the dimension of the original signal.

A general setup of compressed sensing is shown in Fig. 1. The mechanism is as follows: A real vector $\boldsymbol{X} \in \mathbb{R}^n$ is mapped into $\boldsymbol{V} \in \mathbb{R}^k$ by an encoder (or compressor) $f : \mathbb{R}^n \to \mathbb{R}^k$. The decoder (decompressor) $g : \mathbb{R}^k \to \mathbb{R}^n$ receives $\boldsymbol{Y}$, which is a noisy version of $\boldsymbol{V}$, and outputs $\hat{\boldsymbol{X}}$ as the estimation of $\boldsymbol{X}$. The measurement rate, or compression ratio, is defined as

$$R \stackrel{\triangle}{=} \frac{k}{n}. \tag{1}$$

Generally, there are two approaches to the choice of the encoder. The first approach is to constrain the encoder to be a *linear* mapping, denoted by a matrix $\boldsymbol{H} \in \mathbb{R}^{k \times n}$, usually called the *sensing matrix* or *measurement matrix*. Under this encoding linearity constraint, it is reasonable to consider optimal deterministic and random sensing matrices. The other approach is to consider *non-linear* encoders. In this paper, we will focus on random linear encoders; $\boldsymbol{H}$ is assumed to be a random matrix with i.i.d. entries of zero mean and variance $1/n$. On the decoder side, most of the compressed sensing literature focuses on low-complexity decoding algorithms which are robust with respect to observation noise, for example, decoders based on convex optimization, greedy algorithms, etc. (see, for example [3-6]). In this paper, on the other hand, the decoder is assumed to be optimal, namely, it is given by the minimum mean-square error (MMSE) estimator. The input vector $\boldsymbol{X}$ is assumed to be random distributing according some measure that is modeling/capturing sparsity. Note that this statistical assumption (or, Bayesian formulation) is incompatible to "usual" compressive sensing models, in which the underlying signal is assumed to be deterministic and the performance is measured on a worst-case basis with respect to $\boldsymbol{X}$ (Hamming theory). This statistical approach has been previously adopted in the literature (see, for example, [5-12]). Finally, the noise is assumed to additive white and Gaussian.

The main goal of this paper is to analyze rigorously the asymptotic behavior of the MMSE, namely, to find the MMSE for $k, n \to \infty$ with a fixed ratio R. Using the asymptotic MMSE, one can investigate the fundamental tradeoff between optimal reconstruction errors and measurement rates, as a function of the signal and noise statistics. For example, it will be seen that there exists a phase transition threshold of the measurement rate (which only depends on the input statistics). Above the threshold, the noise sensitivity (defined as the ratio between that MMSE and the noise variance) is bounded for all noise

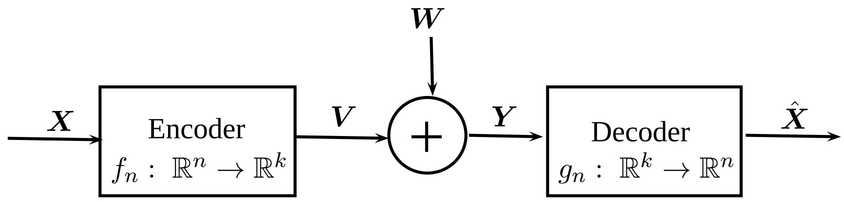

Fig. 1. Noisy compressed sensing setup.

variances. Below the threshold, the noise sensitivity goes to infinity as the noise variance tends to zero.

A. Known results and new contributions

There are several previously reported results that are related (directly or indirectly) to this work. Some of these results were derived rigorously and some of them were not, since they were based on the powerful, but non-rigorous, *replica* method. In the following, we briefly state some of these results. In [12], using the replica method, a decoupling principle of the posterior distribution was claimed, namely, the outcome of inferring about any fixed collection of signal elements becomes independent conditioned on the measurements. Also, it was shown that each signal-element-posterior becomes asymptotically identical to the posterior resulting from inferring the same element in scalar Gaussian noise. Accordingly, this principle allows us to calculate the MMSE of estimating the signal input given the observations. In [11], among other results, it was shown rigorously that for i.i.d. input processes, distributing according to any discrete-continuous mixture measure, the phase transition threshold for optimal encoding is given by the input information dimension. This result serves as a rigorous verification of the replica calculations in [12]. In [10], using the replica method and the decoupling principle, the authors extend the scope of conventional noisy compressive sampling where the sensing matrix is assumed to have i.i.d. entries to allow it to satisfy a certain freeness condition (encompassing Haar matrices and other unitarily invariant matrices). In [13, 14], the authors designed structured sensing matrices (not necessarily i.i.d.), and a corresponding reconstruction procedure, that allows compressed sensing to be performed at acquisition rates approaching to the theoretical optimal limits. A wide variety of previous works are concerning low-complexity decoders, which are robust with respect to the noise, e.g., decoders based on convex optimizations (such as ℓ_1-minimization and ℓ_1-penalized least-squares) [3, 4], graph-based iterative decoders such as linear MMSE estimation and approximate message passing (AMP) [5], etc. For example, in [6], the linear MMSE and LASSO estimators were studied for the case of i.i.d. sensing matrices as special cases of the AMP algorithm, the performance of which was rigorously characterized for Gaussian sensing matrices [15], and generalized for a broad class of sensing matrices in [9]. Another, somewhat related, subject is the recovery of the sparsity pattern with vanishing and non-vanishing error probability, which was studied in a number of recent works, e.g., [6, 16-22].

In this paper, under the previously mentioned model assumptions, we rigorously derive the asymptotic MMSE in a single-letter form. The key idea in our analysis is the fact that by using some direct relationship between optimum estimation and certain partition functions [23], the MMSE can be represented in some mathematically "convenient" form which (due to the previously mentioned input and noise Gaussian statistics assumptions) consists of functions of the *Stieltjes* and *Shannon* transforms. This observation allows us to use some powerful results from random matrix theory (RMT), concerning the asymptotic behavior (a.k.a. deterministic equivalents) of the Stieltjes and Shannon transforms (see e.g., [24, 25] and many references therein). Our asymptotic MMSE formula seems to appear different than the one that is obtained from the replica method [12]. Nevertheless, numerical calculations suggest that the results are equivalent. Thus, similarly to other known cases in statistical mechanics, for which the replica predictions were proved to be correct, our results support the replica method predictions. In the same breath, we believe that our formula is more insightful compared to the replica method results. Also, in contrast to previous works in which only memoryless sources were considered (an indispensable assumption in the analysis), we consider a more general model which allows a certain structured dependency among the various components of the source. Finally, we mention that in a previous related paper [26], the authors have used similar methodologies to obtain the asymptotic mismatched MSE of a codeword (from a randomly selected code), corrupted by a Gaussian vector channel.

B. *Organization*

The remaining part of this paper is organized as follows. In Section II, the model is presented and the problem is formulated. In Section III, the main results are stated and discussed along with a numerical example that demonstrates the theoretical result. In Section IV, the main result is proved, and finally, our conclusions appear in Section V.

II. Notation Conventions and Problem Formulation

A. *Notation Conventions*

Throughout this paper, scalar random variables (RV's) will be denoted by capital letters, their sample values will be denoted by the respective lower case letters and their alphabets will be denoted by the respective calligraphic letters. A similar convention will apply to random vectors and matrices and their sample values, which will be denoted with same symbols in the bold face font. Probability measures will be denoted generically by the letter $\mathbb{P}$. In particular, $\mathbb{P}(\boldsymbol{X}, \boldsymbol{Y})$ is the joint density of the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$. Accordingly, $\mathbb{P}(\boldsymbol{X})$ will denote the marginal of $\boldsymbol{X}$, $\mathbb{P}(\boldsymbol{Y} \mid \boldsymbol{X})$ will denote the conditional density of $\boldsymbol{Y}$ given $\boldsymbol{X}$, and so on.

The expectation operator of a measurable function $f(\boldsymbol{X}, \boldsymbol{Y})$ with respect to (w.r.t.) $\mathbb{P}(\boldsymbol{X}, \boldsymbol{Y})$ will be denoted by $\mathbb{E}\{f(\boldsymbol{X}, \boldsymbol{Y})\}$. The conditional expectation of the same function given a realization $\boldsymbol{y}$ of $\boldsymbol{Y}$, will be denoted by $\mathbb{E}\{f(\boldsymbol{X}, \boldsymbol{Y}) \mid \boldsymbol{Y} = \boldsymbol{y}\}$. When using vectors and matrices in a linear-algebraic format, n-dimensional vectors, like $\boldsymbol{x}$, will be understood as column vectors, the operators $(\cdot)^T$ and $(\cdot)^H$ will denote vector or matrix transposition and vector or matrix conjugate transposition, respectively, and so, $\boldsymbol{X}^T$ would be a row vector. For two positive sequences $\{a_n\}$ and $\{b_n\}$, the notations $a_n \doteq b_n$ and $a_n \approx b_n$ mean equivalence in the exponential order, i.e., $\lim_{n\to\infty} \frac{1}{n} \ln(a_n/b_n) = 0$, and $\lim_{n\to\infty} (a_n/b_n) = 1$, respectively. For two sequences $\{a_n\}$ and $\{b_n\}$, the notation $a_n \asymp b_n$ means that $\lim_{n\to\infty} (a_n - b_n) = 0$. Finally, the indicator function of an event $\mathcal{A}$ will be denoted by $\mathbb{1}_{\mathcal{A}}$.

B. *Model and Problem Formulation*

As was mentioned earlier, we consider sparse signals, supported on a subspace with dimension smaller than n. In the literature, it is often assumed that the input process $\boldsymbol{X}$ has i.i.d. components. In this work, however, we generalize this assumption by considering the following stochastic model: Each component, X_i, $1 \leq i \leq n$, of $\boldsymbol{X}$, is given by $X_i = S_i U_i$ where $\{U_i\}$ are i.i.d. Gaussian random variables with zero mean and variance σ^2, and $\{S_i\}$ are binary random variables taking values in $\{0, 1\}$, independently of $\{U_i\}$. Now, instead of assuming that the *pattern* sequence $\boldsymbol{S} = (S_1, \ldots, S_n)$ is i.i.d., we will assume a more general distribution but we keep certain symmetry properties among the various possible sequences $\{\boldsymbol{S}\}$. In particular, we postulate that all sequences $\{\boldsymbol{S}\}$ with the same number of 1's are equally likely, namely, all configurations with the same "*magnetization*"[1]

$$m_s = \frac{1}{n} \sum_{i=1}^{n} S_i \tag{2}$$

have the same probability. This literally means that the measure $\mathbb{P}(\boldsymbol{S})$ depends on $\boldsymbol{S}$ only via m_s. Consider then the following form

$$\mathbb{P}(\boldsymbol{S}) = C_n \cdot \exp\{n f(m_s)\} \tag{3}$$

[1]The term "magnetization" is borrowed from the field of statistical mechanics of spin array systems, in which S_i is taking values in $\{-1, 1\}$. Nevertheless, for the sake of convince, we will use this term also in our problem.

where $f(\cdot)$ is a certain function independent of n and C_n is a normalization constant. Note that for the popular i.i.d. assumption, f is a linear function. By using the method of types [27], we obtain

$$\begin{aligned} C_n &= \left(\sum_{\boldsymbol{s} \in \{0,1\}^n} \exp\left\{ n f\left(m_s\right) \right\} \right)^{-1} \\ &= \left(\sum_{m \in [0,1]} \Omega\left(m\right) \exp\left\{ n f\left(m\right) \right\} \right)^{-1} \\ &\doteq \exp\left\{ -n \cdot \max_m \left\{ h_2\left(m\right) + f\left(m\right) \right\} \right\} \\ &= \exp\left\{ -n \left[h_2\left(m_a\right) + f\left(m_a\right) \right] \right\} \end{aligned} \tag{4}$$

where $\Omega(m)$ designates the number of binary n-vectors with magnetization m, $h_2(\cdot)$ designates the binary entropy function, and m_a is the maximizer of $h_2(m) + f(m)$ over $[0,1]$. In other words, m_a is the *a-priori* magnetization, namely, the magnetization that *dominates* the measure $\mathbb{P}(\boldsymbol{S})$.

Remark 1 While the Gaussian assumption on U_i's is mandatory in our analysis, the assumption that S_i is taking values in $\{0,1\}$, can be generalized to any discrete probability measure. Such a generalization has some practical motivations [28]. Also, as was reported in [29], statistical dependency in the pattern sequence may lead to the appearance of phase transitions caused by the source, in addition to the phase transition caused by the channel.

Remark 2 In the i.i.d. case, each X_i is distributed according to following mixture distribution (a.k.a. Bernoulli-Gaussian measure)

$$P(x) = (1-p) \cdot \delta(x) + p \cdot P_G(x) \tag{5}$$

where $\delta(x)$ is the Dirac function, $P_G(x)$ is a Gaussian density function and $0 \leq p \leq 1$. Consider a random vector $\boldsymbol{X}$ in which each component is *independently* drawn from $P(x)$. Then, by the law of large numbers (LLN), $\frac{1}{n}\|\boldsymbol{X}\|_0 \xrightarrow{\mathbb{P}} p$, where $\|\boldsymbol{X}\|_0$ designates the number of non-zero elements of a vector $\boldsymbol{X}$. Thus, it is clear that the weight p parametrizes the signal sparsity and P_G is the prior distribution of the non-zero entries.

Finally, we consider the following observation model

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}, \tag{6}$$

where $\boldsymbol{H}$ is a $k \times n$ random matrix, a.k.a. the *sensing matrix*, with i.i.d. entries of zero mean and variance $1/n$. The components of the noise $\boldsymbol{W}$ are i.i.d. Gaussian random variables with zero mean and variance $1/\beta$. We denote by $R \triangleq k/n$ the measurement rate.

The MMSE of $\boldsymbol{X}$ given $\boldsymbol{Y}$ and $\boldsymbol{H}$ is defined as follows

$$\text{mmse}\left(\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{H}\right) \triangleq \mathbb{E}\left\|\boldsymbol{X} - \mathbb{E}\left\{\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{H}\right\}\right\|^2 \tag{7}$$

where $\mathbb{E}\left\{\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{H}\right\}$ is the conditional expectation w.r.t. the measure $\mathbb{P}\left(\cdot \mid \boldsymbol{Y}, \boldsymbol{H}\right)$. Accordingly, we define the *asymptotic MMSE* as

$$D\left(R, \beta\right) \triangleq \limsup_{n \to \infty} \frac{1}{n} \text{mmse}\left(\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{H}\right). \tag{8}$$

As was mentioned earlier, our main goal is to rigorously derive computable, single-letter expression for $D\left(R, \beta\right)$.

III. Main Result

In this section, our main result is first presented and discussed. Then, we provide a numerical example in order to illustrate the obtained theoretical results. The proof of the main theorem is provided in Section IV.

Before we state our main result, we define some auxiliary functions of a generic variable $x \in [0, 1]$:

$$b\left(x\right) \triangleq \frac{-\left[1 + \beta\sigma^2\left(R - x\right)\right] + \sqrt{\left[1 + \beta\sigma^2\left(R - x\right)\right]^2 + 4\beta\sigma^2 x}}{2\beta\sigma^2 x}, \tag{9}$$

$$g\left(x\right) \triangleq 1 + \beta\sigma^2 x b\left(x\right), \tag{10}$$

$$\bar{I}\left(x\right) \triangleq \frac{R}{x}\ln g\left(x\right) - \ln b\left(x\right) - \frac{\beta\sigma^2 R b\left(x\right)}{g\left(x\right)}, \tag{11}$$

$$V\left(x\right) \triangleq \frac{\beta^3\sigma^4 b^2\left(x\right) x^2}{2g^2\left(x\right)}, \tag{12}$$

$$L\left(x\right) \triangleq \frac{\beta^2\sigma^2 b\left(x\right)}{2g^2\left(x\right)}, \tag{13}$$

and

$$t\left(x\right) \triangleq f\left(x\right) - \frac{x}{2}\bar{I}\left(x\right) + V\left(x\right)\left[m_a R\sigma^2 + \frac{R}{\beta}\right]. \tag{14}$$

Next, for $x, y \in [0, 1]$ define the functions

$$\nu_1\left(x, y\right) \triangleq \frac{\beta R}{g\left(x\right)} - \frac{\beta^2 R\sigma^2 b\left(x\right) y}{g^2\left(x\right)} + \frac{1}{\sigma^2}, \tag{15}$$

$$\nu_2\left(x\right) \triangleq \frac{\beta R}{g\left(x\right)} + \frac{1}{\sigma^2}, \tag{16}$$

and

$$\alpha\left(x,y\right)\triangleq\frac{1}{\nu_{1}\left(x,y\right)\nu_{2}\left(x\right)}. \tag{17}$$

The asymptotic MMSE is given in the following theorem.

Theorem 1 (Asymptotic MMSE) Let Q be a random variable distributed according to

$$\mathbb{P}_{Q}\left(q\right)=\frac{1-m_{a}}{\sqrt{2\pi P_{y}}}\exp\left(-\frac{q^{2}}{2P_{y}}\right)+\frac{m_{a}}{\sqrt{2\pi\left(P_{y}+R^{2}\sigma^{2}\right)}}\exp\left(-\frac{q^{2}}{2\left(P_{y}+R^{2}\sigma^{2}\right)}\right) \tag{18}$$

where m_a is defined as in (4) and $P_y \triangleq m_a\sigma^2 R + R/\beta$. Let us define

$$K\left(Q,\alpha_{1},\alpha_{2}\right)\triangleq\frac{1}{2}\left[1+\tanh\left(\frac{L\left(\alpha_{1}\right)Q^{2}-\alpha_{2}}{2}\right)\right] \tag{19}$$

where $\alpha_1 \in [0,1]$ and $\alpha_2 \in \mathbb{R}$. Let m° and γ° be solutions of the system of equations

$$\gamma^{\circ}\triangleq-\mathbb{E}\left\{K\left(Q,m^{\circ},\gamma^{\circ}\right)Q^{2}\left.\frac{\mathrm{d}L\left(m\right)}{\mathrm{d}m}\right|_{m=m^{\circ}}\right\}-\left.\frac{\mathrm{d}t\left(m\right)}{\mathrm{d}m}\right|_{m=m^{\circ}}, \tag{20a}$$

$$m^{\circ}\triangleq\mathbb{E}\left\{K\left(Q,m^{\circ},\gamma^{\circ}\right)\right\} \tag{20b}$$

where in case of more than one solution, (m°,γ°) is the pair with the largest value of

$$t\left(m^{\circ}\right)+\left(m^{\circ}-\frac{1}{2}\right)\gamma^{\circ}+\mathbb{E}\left\{\frac{1}{2}L\left(m^{\circ}\right)Q^{2}+\ln2\cosh\left(\frac{L\left(m^{\circ}\right)Q^{2}-\gamma^{\circ}}{2}\right)\right\}. \tag{21}$$

Finally, define

$$\rho_{1}^{\circ}\triangleq\mathbb{E}\left\{K\left(Q,m^{\circ},\gamma^{\circ}\right)Q^{2}\right\}, \tag{22}$$

$$\rho_{2}^{\circ}\triangleq\mathbb{E}\left\{K^{2}\left(Q,m^{\circ},\gamma^{\circ}\right)\right\}, \tag{23}$$

$$\rho_{3}^{\circ}\triangleq\mathbb{E}\left\{K^{2}\left(Q,m^{\circ},\gamma^{\circ}\right)Q^{2}\right\}. \tag{24}$$

Then, the limit supremum in (8) is, in fact, an ordinary limit, and the asymptotic MMSE is given by

$$\begin{aligned}D\left(R,\beta\right)=\sigma^{2}m^{\circ}b\left(m^{\circ}\right)+\frac{2b\left(m^{\circ}\right)}{r^{3}\left(m^{\circ}\right)}\beta^{3}\sigma^{2}\left[P_{y}-\rho_{1}^{\circ}\right]\left[m^{\circ}\alpha\left(m^{\circ},m^{\circ}\right)-\rho_{2}^{\circ}\alpha\left(m^{\circ},\rho_{2}^{\circ}\right)\right]\\+\frac{\beta^{2}}{r^{2}\left(m^{\circ}\right)}\left[\alpha\left(m^{\circ},m^{\circ}\right)\rho_{1}^{\circ}-\alpha\left(m^{\circ},\rho_{2}^{\circ}\right)\rho_{3}^{\circ}\right].\end{aligned} \tag{25}$$

In the following, we explain the above result qualitatively, and in particular, the various quantities that have been defined in Theorem 1. The first important quantity is m°, which is obtained as the solution of the system of equations in (20), and which we will refer to as the *posterior* magnetization. We use the term "posterior" in order to distinguish it from the a-priori magnetization m_a; while m_a is the magnetization that dominates the probability distribution function of the source, before observing $\boldsymbol{Y}$, the posterior magnetization is the one that dominates the posterior distribution, namely, after observing the measurements. It is instructive to look at another representation of m°, which appears in the analysis, and is given as follows

$$m^\circ = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\left[1+\tanh\left(\frac{L\left(m^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right] \tag{26}$$

where $\boldsymbol{h}_i$ is the ith column of $\boldsymbol{H}$. Note that the summand of the above sum is bounded between zero and one, and hence, so is m°, which makes sense. Intuitively speaking, the first term in argument of the hyperbolic tangent can be interpreted as a projection of the measurements on the sensing matrix columns, and γ° serves as a correction/alignment term so that the overall summation gives the "correct" magnetization (depending on the SNR and the measurement rate). The role of the hyperbolic tangent becomes clearer when considering the low noise case. For large SNR, the hyperbolic tangent behaves very sharply; it converges to the sign function. When the sign function value equals one, the summand in (26) also equals one, which means that $S_i = 1$. On the other hand, when the sign function value equals -1, the summand equals zero, which means that $S_i = 0$. So, for large SNR the posterior magnetization simply equals to the a-priori magnetization. Regarding the MMSE itself, it can be seen that in this case $K\left(\cdot\right) = K^2\left(\cdot\right)$, and thus $\rho_1^\circ = \rho_3^\circ$ and $\rho_2^\circ = m^\circ$. Therefore, according to (25), we see that we are only left with the first term on the right hand side, which for large β behaves, for $R > m^\circ$, like

$$\sigma^2 m_a b\left(m_a\right) \approx \frac{\sigma^2 m_a}{\beta\left(R-m_a\right)}. \tag{27}$$

This result was already noticed in [11][2] for i.i.d. sources under which $m_a = p$.

Corollary 1 ("Infinite" SNR) In the low noise regime, $\beta \to \infty$, the asymptotic MMSE is given by

$$\lim_{\beta\to\infty}\left[\beta\cdot D\left(R,\beta\right)\right] = \sigma^2\frac{m_a}{R-m_a}, \tag{28}$$

for $R > m_a$.

The solution to (21) is known as a *critical point*, beyond which the solution to (20) ceases to be the dominant posterior magnetization, and accordingly, it must jump elsewhere. Furthermore, as we vary one of the other parameters of our model (including the source model), it might happen that the dominant magnetization jumps from one value to another.

[2]In [10, 30], it was stated that (27) is proved rigorously in [11] for i.i.d. sources. However, we suspect that this claim is not true, due to the fact that in [11] the authors use the replica symmetry assumption in order to obtain this result.

It is interesting to note that there are essentially two origins for possible phase transitions in our model: The first one is the channel $\boldsymbol{H}$ that induces "long-range interactions"[3]. The second is the source, which may have possible dependency (or interaction) between its various components (see (3)). Accordingly, in [29, Example E] the problem of estimation of sparse signals, assuming that $\boldsymbol{H} = \boldsymbol{I}$, was considered. It was shown that, despite the fact that there are no long-range interactions induced by the channel, still there are phase transitions if the source is not i.i.d. Indeed, in the i.i.d. case, the problem is analogous to a system of non-interacting particles, where of course, no phase transitions can exist.

In the following, we consider the special case where $f(m)$ is quadratic[4], i.e., $f(m) = am + bm^2/2$, and demonstrate that the dominant posterior magnetization might jump from one value to another. Note that this example was also considered in [29, Example E]. For simplicity of the demonstration, assume that σ^2 and β are so small such that the random fluctuation in (19) are negligible. Accordingly, using (20), we may write

$$m^\circ \approx \frac{1}{2}\left[1+\tanh\left(\frac{1}{2}\left.\frac{\mathrm{d}t(m)}{\mathrm{d}m}\right|_{m=m^\circ}\right)\right] \tag{29}$$

$$\approx \frac{1}{2}\left[1+\tanh\left(\frac{bm^\circ + a}{2}\right)\right], \tag{30}$$

which can be regarded as the same equation of the *spin*-magnetization (namely, after transforming S_i's into spins, $\mu_i \in \{-1, 1\}$, using the transformation $\mu_i = 1 - 2S_i$) as in the Curie-Weiss model of spin arrays (see e.g., [31, Sect. 4.2]). For example, for $a = 0$ and $b > 1$, this equation has two symmetric non-zero solutions $\pm m_0$, which both dominate the partition function. If $0 < a \ll 1$, it is evident that the symmetry is broken, and there is only one dominant solution which is about $m_0 \text{sgn}(m_0)$. Further discussion on the behavior of the above saddle point equation, and various interesting approximations of the dominant magnetization can be found in [29, 31, 32].

It is tempting to compare Theorem 1 with the prediction of the replica method [12]. Unfortunately, we were unable to show analytically that the two results are in agreement, despite the fact that there are some similarities. Nevertheless, numerical calculations suggest that this is the case. Fig. 2 shows the asymptotic MMSE obtained using Theorem 1 and using the replica method, as a function of β, assuming an i.i.d. source with sparsity rate $p = 0.1$, and measurement rate $R = 0.3$. It can be seen that both results

[3]In the considered settings, the posterior, is proportional to $\exp\left\{-\beta\left\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{X}\right\|^2/2\right\}$, and after expansion of the norm, the exponent includes an "external-field term", proportional to $\boldsymbol{y}^T\boldsymbol{H}\boldsymbol{x}$, and a "pairwise spin-spin interaction term", proportional to $\left\|\boldsymbol{H}\boldsymbol{X}\right\|^2$. These terms contain linear subset of components (or "particles") of $\boldsymbol{X}$, which are known as long-range interactions.

[4]As was noted in [29], quadratic model (similar to the *random-field Curie-Weiss model* of spin systems (see e.g., [31, Sect. 4.2])) can be thought of as consisting of the first two terms of the Taylor series expansion of a smooth function.

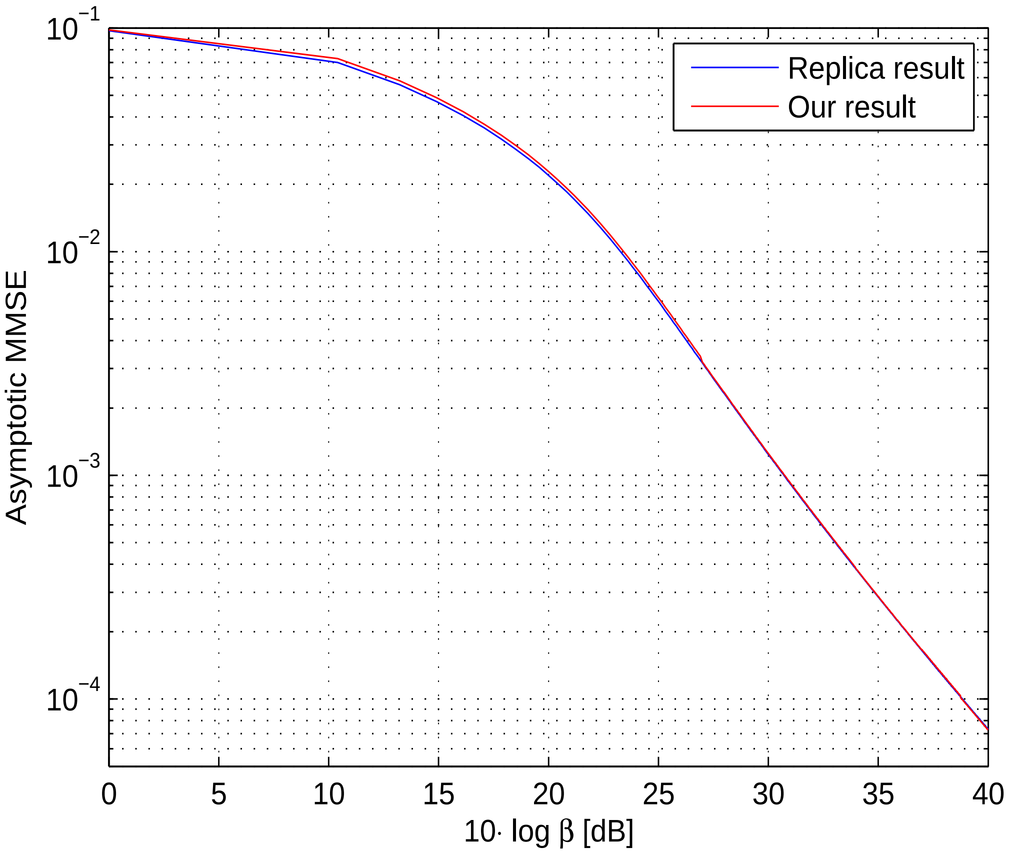

Fig. 2. Comparison of the asymptotic MMSE using Theorem 1 and the replica method as a function of β, for sparsity rate $p = 0.1$, and measurement rate $R = 0.3$.

give the same MMSE. Table I shows the relative error, defined as $|\text{mmse}_{\text{our}} - \text{mmse}_{\text{replica}}| / \text{mmse}_{\text{our}}$, as a function of β. More enlightening numerical examples can be found in [10, 30, 13, 14].

IV. Proof

A. Proof Outline

In this subsection, before getting deep into the proof of Theorem 1, we discuss the techniques and the main steps which will be used in the proof. The analysis is essentially composed of three main steps. The first step is finding a generic expression of the MMSE. This is done by using a direct relationship between the MMSE and some partition function, which can be found in Lemma 1. This expression contains terms that can be asymptotically assessed using the well-known Stieltjes and Shannon transforms. In the second step (appearing in Appendix B), we derive the asymptotic behavior of these functions (which are extremely

TABLE I

COMPARISON BETWEEN THEOREM 1 AND THE REPLICA METHOD

$10 \log \beta$	Relative Error
0	$5.11 \cdot 10^{-3}$
10	$8.09 \cdot 10^{-3}$
15	$6.12 \cdot 10^{-3}$
20	$6.51 \cdot 10^{-3}$
25	$6.03 \cdot 10^{-3}$
30	$4.65 \cdot 10^{-3}$
35	$4.49 \cdot 10^{-3}$
40	$4.59 \cdot 10^{-3}$

complex to analyze for finite n). In other words, we show that these functions converge, with probability tending to one, as $n \to \infty$, to some random functions that are much easier to work with. This is done by invoking recent powerful methods from RMT, such as, the Bai-Silverstein method [33]. The resulting functions are, in general, random, due to the fact that they depend on the observations $\boldsymbol{y}$ and the sensing matrix $\boldsymbol{H}$. Accordingly, we show that for the calculation of the asymptotic MMSE, it is sufficient to take into account "only" combinations of typical vectors $\{\boldsymbol{y}\}$ and matrices $\{\boldsymbol{H}\}$, where "typicality" is defined in accordance to the above-mentioned asymptotic results. Therefore, at the end of the second step, we obtain an approximation (which is exact as $n \to \infty$) for the MMSE. Finally, in the last step, using this approximation and large deviations theory, we obtain the result stated in Theorem 1 (this step can be found in Appendix C).

B. Definitions

An important function, which will be pivotal to our derivation, is the *partition function*, which is defined as follows.

Definition 1 (Partition Function) Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be random vectors with joint density function $\mathbb{P}(\boldsymbol{X}, \boldsymbol{Y})$. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^T$ be a deterministic column vector of n real-valued parameters. The partition function w.r.t. $\mathbb{P}(\boldsymbol{X}, \boldsymbol{Y})$, denoted by $Z(\boldsymbol{Y}; \boldsymbol{\lambda})$, is defined as

$$Z(\boldsymbol{Y}; \boldsymbol{\lambda}) \triangleq \int_{\mathcal{X}^n} \mathrm{d}\boldsymbol{x}\, \mathbb{P}(\boldsymbol{x}, \boldsymbol{Y}) \exp\left\{\boldsymbol{\lambda}^T \boldsymbol{x}\right\}, \tag{31}$$

where it is assumed that the integral converges uniformly at least in some neighborhood of $\boldsymbol{\lambda} = \mathbf{0}$ [5].

The motivation of the above definition is the following simple result [23].

Lemma 1 (MMSE-partition function relation) Let $Z(\boldsymbol{Y};\boldsymbol{\lambda})$ be defined as in (31). Then, the following relation between $Z(\boldsymbol{Y};\boldsymbol{\lambda})$ and the MMSE of $\boldsymbol{X}$ given $\boldsymbol{Y}$, holds true

$$\begin{aligned}\text{mmse}(\boldsymbol{X} \mid \boldsymbol{Y}) &\triangleq \sum_{i=1}^{n} \mathbb{E}\left\{(X_i - \mathbb{E}\{X_i \mid \boldsymbol{Y}\})^2\right\} \\ &= \sum_{i=1}^{n}\left[\mathbb{E}\left\{X_i^2\right\} - \mathbb{E}\left\{\left[\frac{\partial \ln Z(\boldsymbol{Y};\boldsymbol{\lambda})}{\partial \lambda_i}\right]^2\Bigg|_{\boldsymbol{\lambda}=0}\right\}\right] \qquad (32) \\ &= \sum_{i=1}^{n} \mathbb{E}\left\{\frac{\partial^2 \ln Z(\boldsymbol{Y};\boldsymbol{\lambda})}{\partial \lambda_i^2}\Bigg|_{\boldsymbol{\lambda}=0}\right\}. \qquad (33)\end{aligned}$$

Proof: Readily follows by taking the gradient of (31) w.r.t. $\boldsymbol{\lambda}$, and evaluating the results at $\boldsymbol{\lambda} = \mathbf{0}$. ■

Our analysis will rely heavily on methods and results from RMT. Two efficient tools which are commonly being used in RMT are the *Stieltjes* and *Shannon* transforms, which are defined as follows.

Definition 2 (Stieltjes Transform) Let μ be a finite nonnegative measure with support $\text{supp}(\mu) \subset \mathbb{R}$, i.e., $\mu(\mathbb{R}) < \infty$. The Stieltjes transform $S_\mu(z)$ of μ is defined for $z \in \mathbb{C} - \text{supp}(\mu)$ as

$$S_\mu(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu(\lambda)}{\lambda - z}.$$

Let $F_{\boldsymbol{A}}(\cdot)$ be the empirical spectral distribution (ESD) of the eigenvalues of $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, namely,

$$F_{\boldsymbol{A}}(x) \triangleq \frac{1}{N}\{\#\text{ of eigenvalues of } \boldsymbol{A} \leq x\}. \qquad (34)$$

The Stieltjes transform of $F_{\boldsymbol{A}}(x)$ is defined as

$$S_{\boldsymbol{A}}(z) = \int_{\mathbb{R}^+} \frac{dF_{\boldsymbol{A}}(x)}{x - z} = \frac{1}{N}\operatorname{tr}(\boldsymbol{A} - z\boldsymbol{I})^{-1} \qquad (35)$$

for $z \in \mathbb{C} \setminus \mathbb{R}^+$.

The last equality readily follows by using the spectral decomposition of $\boldsymbol{A}$, and the fact that the trace of a matrix equals to the sum of its eigenvalues. For brevity, we will refer to $S_{\boldsymbol{A}}(z)$ as the Stieltjes transform of $\boldsymbol{A}$, rather than the Stieltjes transform of $F_{\boldsymbol{A}}(x)$.

[5] In case that this assumption does not hold, one can instead, parametrize each component λ_i of $\boldsymbol{\lambda}$ as a purely imaginary number $\lambda_i = j\omega_i$ where $i = \sqrt{-1}$, similarly to the definition of the characteristics function.

Definition 3 (Shannon Transform) The Shannon of transform of a non-negative definite matrix $\boldsymbol{A} \in \mathbb{C}^{N \times N}$ is defined as

$$\nu_{\boldsymbol{A}}(z) = \frac{1}{N} \ln \det \left(\frac{1}{z} \boldsymbol{A} + \boldsymbol{I} \right), \tag{36}$$

for $z > 0$.

The relation between our partition function and the Stieltjes and Shannon transforms will become clear in the sequel. Finally, we define the notion of deterministic equivalence.

Definition 4 (Deterministic Equivalence) Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{f_n\}$ be a series of measurable complex-valued functions, $f_n : \Omega \times \mathbb{C} \to \mathbb{C}$. Let $\{g_n\}$ be a series of complex-valued functions, $g_n : \mathbb{C} \to \mathbb{C}$. Then, $\{g_n\}$ is said to be a deterministic equivalent of $\{f_n\}$ on $D \subset \mathbb{C}$, if there exists a set $A \subset \Omega$ with $P(A) = 1$, such that

$$f_n(\omega, z) - g_n(z) \to 0 \tag{37}$$

as $n \to \infty$ for all $\omega \in A$ and for all $z \in D$.

Loosely speaking, $\{g_n\}$ is a deterministic equivalent of a sequence of random variables $\{f_n\}$ if $g_n(z)$ approximates $f_n(\omega, z)$ arbitrarily closely as n grows, for every $z \in D$ and almost every $\omega \in A$.

C. Auxiliary Results

In our derivations, the following asymptotic results will be used.

Lemma 2 Let $(\Omega, \mathcal{F}, P)$ be a probability space, and consider a sequence of random variables $\left\{ X_i^{(n)} \right\}_{i=1}^{n}$. Assume that

$$\max_{1 \le i \le n} \left\{ \mathbb{E} \left| X_i^{(n)} \right|^p \right\} \le \frac{C}{n^{1+\nu}} \tag{38}$$

where C, $\nu > 0$, and $p \ge 1$ are some fixed constants. Then,

$$\lim_{n \to \infty} \mathbb{P} \left(\left\{ \omega \in \Omega : \sup_{m \ge n} \frac{1}{m} \sum_{i=1}^{m} \left| X_i^{(m)}(\omega) \right| \ge \epsilon \right\} \right) = 0 \tag{39}$$

for all $\epsilon > 0$, namely, $\left(\frac{1}{n} \sum_{i=1}^{n} \left| X_i^{(n)} \right| \right)$ converges to zero almost sure (a.s.) as $n \to \infty$.

Proof: Using Chebyshev's inequality and then Jensen's inequality, for a given $\delta > 0$, we have that

$$\mathbb{P} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left| X_i^{(n)} \right| > \delta \right\} \le \frac{1}{\delta^p} \mathbb{E} \left\{ \left(\frac{1}{n} \sum_{i=1}^{n} \left| X_i^{(n)} \right| \right)^p \right\} \tag{40}$$

$$\leq \frac{1}{n\delta^p}\sum_{i=1}^{n}\mathbb{E}\left|X_i^{(n)}\right|^p \tag{41}$$

$$\leq \frac{1}{\delta^p}\max_{1\leq i\leq n}\left\{\mathbb{E}\left|X_i^{(n)}\right|^p\right\} \tag{42}$$

$$\leq \frac{C}{\delta^p n^{1+\nu}} \tag{43}$$

where the last inequality follows by (38). With (43), the desired result follows from common arguments that rely on the Borel-Cantelli lemma. As this argument will be used repeatedly in our analysis, for completeness we explicitly present it here. Indeed, as the right-hand side (r.h.s.) of (43) is summable, by the Borel-Cantelli lemma, we have that

$$\mathbb{P}\left(\left\{\omega\in\Omega:\ \frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(n)}\left(\omega\right)\right|\geq\delta \text{ infinitely often}\right\}\right)=0. \tag{44}$$

But since $\delta > 0$ is arbitrary, the above holds for all rational $\delta > 0$. Since any countable union of sets of zero probability is still a set of zero probability, we conclude that

$$\mathbb{P}\left(\bigcup_{q\in\mathbb{N}}\left\{\omega\in\Omega:\ \frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(n)}\left(\omega\right)\right|\geq\frac{1}{q} \text{ infinitely often}\right\}\right)=0. \tag{45}$$

■

Remark 3 Note that the random variables $\left\{X_i^{(n)}\right\}_{i=1}^{n}$ may depend on each other, and the result will be still true.

The following lemmas deal with the asymptotic behavior of scalar functions of random matrices, in the form of Stieltjes and Shannon transforms, defined earlier. The proofs of the following results are based on a powerful approach by Bai and Silverstein [33], a.k.a. the Stieltjes transform method in the spectral analysis of large-dimensional random matrices.

Lemma 3 ([34]) Let $\boldsymbol{X}_m \in \mathbb{C}^{m\times l}$ be a sequence of random matrices with i.i.d. entries, $\mathbb{E}\left|X_{i,j}-\mathbb{E}X_{i,j}\right|^2 = 1/l$, and let $\boldsymbol{G}_l = \operatorname{diag}\left(g_1,\ldots,g_l\right)\in\mathbb{R}^{l\times l}$ be a sequence of deterministic matrices, satisfying $g_j \geq 0$ for all $1\leq j\leq l$ and $\sup_j g_j < \infty$. Denote $\boldsymbol{B}_m = \boldsymbol{X}_m\boldsymbol{G}_l\boldsymbol{X}_m^H$ and let $l, m\rightarrow\infty$ with fixed $0 < c \triangleq m/l < \infty$. Then, for every $\gamma > 0$

$$\frac{1}{m}\ln\det\left(\frac{1}{\gamma}\boldsymbol{B}_m+\boldsymbol{I}_m\right)-\eta\left(\gamma\right)\rightarrow 0, \quad \text{a.s.} \tag{46}$$

where

$$\eta\left(\gamma\right)\triangleq\frac{1}{m}\sum_{j=1}^{l}\ln\left(1+cg_j\bar{S}\left(-\gamma\right)\right)-\ln\left(\gamma\bar{S}\left(-\gamma\right)\right)-\frac{1}{l}\sum_{j=1}^{l}\frac{g_j\bar{S}\left(-\gamma\right)}{1+cg_j\bar{S}\left(-\gamma\right)} \tag{47}$$

and $\bar{S}(z)$ is defined by the unique positive solution of the equation

$$\bar{S}(z) = \left(\frac{1}{l} \sum_{j=1}^{l} \frac{g_j}{1 + c g_j \bar{S}(z)} - z \right)^{-1}. \tag{48}$$

The next lemma deals with the asymptotic behavior of the Stieltjes transform.

Lemma 4 ([35]) Let $\boldsymbol{X}_m$, $\boldsymbol{G}_l$, and $\boldsymbol{B}_m$ be defined as in Lemma 3. Let $\boldsymbol{\Theta}_m \in \mathbb{C}^{m\times m}$ be a deterministic sequence of matrices having uniformly bounded spectral norms (with respect to m)[6]. Then, as $m, l \to \infty$ we a.s. have that

$$\frac{1}{m} \operatorname{tr}\left(\boldsymbol{\Theta}_m \left(\boldsymbol{B}_m - z\boldsymbol{I}_m \right)^{-1} \right) - \frac{1}{m} \operatorname{tr}\left(\boldsymbol{\Theta}_m \right) \bar{S}(z) \to 0, \text{ for all } z \in \mathbb{C} \setminus \mathbb{R}_+. \tag{49}$$

Remark 4 In [36], the authors propose a somewhat more restrictive (but useful) version of Lemma 4. Assuming that $\boldsymbol{\Theta}_m$ has a uniformly bounded Frobenius norm (for all m), they show similarly that

$$\left| \operatorname{tr}\left(\boldsymbol{\Theta}_m \left(\boldsymbol{B}_m - z\boldsymbol{I}_m \right)^{-1} \right) - \operatorname{tr}\left(\boldsymbol{\Theta}_m \right) \bar{S}(z) \right| \to 0, \text{ for } z \in \mathbb{C} \setminus \mathbb{R}_+. \tag{50}$$

a.s. as $m, l \to \infty$.

In order to apply the above results in our analysis, a somewhat more general version will be needed. First, the matrix $\boldsymbol{\Theta}_m$ in the Lemma 4 is assumed to be deterministic and bounded (in the spectral or Frobenius senses). In our case, however, we will need to deal with a random matrix $\boldsymbol{\Theta}_m$ which is independent of the other random variables. The following proposition accounts for this problem. The proof is relegated to Appendix A.

Proposition 1 The assertion of Lemma 4 holds true also for a random $\boldsymbol{\Theta}_m \in \mathbb{C}^{m\times m}$, which is independent of $\boldsymbol{X}_m$, and has a uniformly bounded spectral norm (with respect to m) in the a.s. sense.

Remark 5 In Proposition 1, it is assumed that $\boldsymbol{\Theta}_m$ has uniformly bounded spectral norm (uniformly in m) in the a.s. sense, namely,

$$\limsup_{m\to\infty} \|\boldsymbol{\Theta}_m\| < \infty \tag{51}$$

with probability one. In other words, for every $\epsilon > 0$, there exists some positive M_0 such that for all $m > M_0$ we have that $\|\boldsymbol{\Theta}_m\| < D + \epsilon$ for some finite constant D.

[6]Actually we only need to demand the distribution $F_{\boldsymbol{\Theta}_m}$ to be tight, namely, for all $\epsilon > 0$ there exists $M > 0$ such that $F_{\boldsymbol{\Theta}_m}(M) > 1 - \epsilon$ for all m.

The second issue is regarding the assumption that the ratio $c = m/l$, in the previous lemmas, tends to a strictly positive limit. In our case, however, this limit may be zero. Fortunately, it turns out that the previous results still hold true also in this case, namely, a continuity property w.r.t. c. Technically speaking, this fact can be shown by repeating the original proofs of the above results and noticing that the positivity assumption is superfluous. In case that m is fixed while l goes to infinity (and then c vanishes), using the strong law of large numbers (SLLN), it is easy to see that the previous lemmas indeed hold true. Also, if $m \ll \sqrt{l}$, then a simple approach is to show that the diagonal elements of the matrix $\boldsymbol{X}_m \boldsymbol{X}_m^T$ concentrate around a fixed value, and that the row sum of off-diagonal terms converges to zero. Then using Gershgorin's circle theorem [37] one obtains the deterministic equivalent.

In the following subsection, we prove Theorem 1. The proof contains several tedious calculations and lemmas, which will relegated to appendices for the sake of convenience.

D. Main Steps in the Derivation of Theorem 1

Let $\boldsymbol{s}$ and $\boldsymbol{r}$ be two binary sequences of length n, and let $\mathcal{S} \triangleq \text{spt}(\boldsymbol{s})$ and $\mathcal{R} \triangleq \text{spt}(\boldsymbol{r})$ designate their respective *generalized supports*, defined as $\text{spt}(\boldsymbol{s}) \triangleq \{i \in \mathbb{N} : \; S_i \neq 0\}$, and similarly for $\boldsymbol{r}$. Also, define

$$\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \triangleq \sum_{j\in\mathcal{S}\cap\mathcal{R}} \boldsymbol{e}_{m_j^s} \tilde{\boldsymbol{e}}_{m_j^r}^T \tag{52}$$

where $\boldsymbol{e}_{m_j^s}$ and $\tilde{\boldsymbol{e}}_{m_j^r}$ denote unit vectors of size $|\mathcal{S}| \times 1$[7] and $|\mathcal{R}| \times 1$, having "1" at the indexes $m_j^s \triangleq \sum_{l=1}^j s_l$ and $m_j^r \triangleq \sum_{l=1}^j r_l$, respectively.

Example 1 Let $n = 6$, and consider $\boldsymbol{s} = (1,1,0,0,1,1)$ and $\boldsymbol{r} = (0,1,1,0,0,1)$. Then, $\mathcal{S} = \{1,2,5,6\}$, $\mathcal{R} = \{2,3,6\}$, and thus $\mathcal{S} \cap \mathcal{R} = \{2,6\}$. Whence $m_2^s = 2$, $m_2^r = 1$, $m_1^6 = 4$, and $m_6^r = 3$. Accordingly, the matrix $\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}$ is given by

$$\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T = \left(\boldsymbol{e}_2 \tilde{\boldsymbol{e}}_1^T + \boldsymbol{e}_4 \tilde{\boldsymbol{e}}_3^T\right)^T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

For a vector $\boldsymbol{v}$ and a matrix $\boldsymbol{V}$, we define $\boldsymbol{v}_{\boldsymbol{s}} \triangleq \boldsymbol{v}|_{\mathcal{S}}$ and $\boldsymbol{V}_{\boldsymbol{s}} \triangleq \boldsymbol{V}|_{\mathcal{S}}$, which is the restriction of the entries of $\boldsymbol{v}$ and the columns of $\boldsymbol{V}$ on the support $\mathcal{S}$, respectively. Finally, for brevity, we define the

[7]For a set $\mathcal{A}$, we use $|\mathcal{A}|$ to designate its cardinality.

following quantities

$$\mathcal{H}^{\boldsymbol{s}} \triangleq \left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}, \tag{53}$$

$$\mathcal{H}_{\boldsymbol{i}}^{\boldsymbol{s}} \triangleq \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}, \tag{54}$$

$$\mathcal{H}_{\boldsymbol{i,j}}^{\boldsymbol{s}} \triangleq \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_{i,j} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}, \tag{55}$$

where $\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i \triangleq \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} - \boldsymbol{z}_i \boldsymbol{z}_i^T$, and $\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_{i,j} \triangleq \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i - \boldsymbol{z}_j \boldsymbol{z}_j^T$, in which $\boldsymbol{z}_i$ is the ith row of the $\boldsymbol{H}_{\boldsymbol{s}}$.

In the following, we first derive a generic expression for the MMSE. Under the model described in Section II, one have that

$$\mathbb{P}\left(\boldsymbol{y} \mid \boldsymbol{H}, \boldsymbol{x}\right) = \frac{1}{\left(2\pi/\beta\right)^{k/2}} \exp\left(-\frac{\beta}{2} \left\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\right\|^2\right) \tag{56}$$

and that

$$\mathbb{P}\left(\boldsymbol{x} \mid \boldsymbol{s}\right) = \sum_{\boldsymbol{s} \in \{0,1\}^n} \mathbb{P}\left(\boldsymbol{s}\right) \prod_{i:\, s_i = 0} \delta\left(x_i\right) \prod_{i:\, s_i = 1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} x_i^2}. \tag{57}$$

Therefore, the partition function (31) is given by

$$Z\left(\boldsymbol{y}, \boldsymbol{H}; \boldsymbol{\lambda}\right) = \sum_{\boldsymbol{s} \in \{0,1\}^n} \mathbb{P}\left(\boldsymbol{s}\right) \int_{\mathbb{R}^n} \frac{\exp\left(-\beta \left\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\right\|^2/2 + \boldsymbol{\lambda}^T \boldsymbol{x}\right)}{\left(2\pi/\beta\right)^{k/2}} \prod_{i:\, s_i = 0} \delta\left(x_i\right) \prod_{i:\, s_i = 1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} x_i^2} \mathrm{d}\boldsymbol{x}.$$

Now, note that

$$\left\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\right\|^2 \prod_{s_i = 0} \delta\left(x_i\right) = \left[\left\|\boldsymbol{y}\right\|^2 - 2\sum_{i \in \mathcal{S}} \boldsymbol{h}_i^T \boldsymbol{y} x_i + \sum_{i,j \in \mathcal{S}} x_i x_j \boldsymbol{h}_i^T \boldsymbol{h}_j\right] \prod_{s_i = 0} \delta\left(x_i\right) \tag{58}$$

$$= \left[\left\|\boldsymbol{y}\right\|^2 - 2\boldsymbol{x}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y} + \boldsymbol{x}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{x}_{\boldsymbol{s}}\right] \prod_{s_i = 0} \delta\left(x_i\right) \tag{59}$$

where $\boldsymbol{h}_i$ denotes the ith column of $\boldsymbol{H}$, and similarly,

$$\boldsymbol{\lambda}^T \boldsymbol{x} \prod_{s_i = 0} \delta\left(x_i\right) = \left(\sum_{i \in \mathcal{S}} x_i \lambda_i\right) \prod_{s_i = 0} \delta\left(x_i\right) \tag{60}$$

$$= \boldsymbol{\lambda}_{\boldsymbol{s}}^T \boldsymbol{x}_{\boldsymbol{s}} \prod_{s_i = 0} \delta\left(x_i\right) \tag{61}$$

Using the fact that $\delta\left(\cdot\right)$ is a measure on $\mathbb{R}$, one may conclude that

$$Z\left(\boldsymbol{y}, \boldsymbol{H}; \boldsymbol{\lambda}\right) = \sum_{\boldsymbol{s} \in \{0,1\}^n} \mathbb{P}\left(\boldsymbol{s}\right) \frac{1}{\left(2\pi/\beta\right)^{k/2}} \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^{|\mathcal{S}|}} \exp\left(-\frac{\beta}{2} \left\|\boldsymbol{y}\right\|^2\right)$$

$$\times \int_{\mathbb{R}^{|\mathcal{S}|}} \exp\left(-\boldsymbol{x}_{\boldsymbol{s}}^T\left(\frac{\beta}{2}\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{2\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)\boldsymbol{x}_{\boldsymbol{s}}+\boldsymbol{x}_{\boldsymbol{s}}^T\left(\boldsymbol{\lambda}_{\boldsymbol{s}}+\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\right)\right)\mathrm{d}\boldsymbol{x}_{\boldsymbol{s}} \tag{62}$$

$$= \sum_{\boldsymbol{s}\in\{0,1\}^n} \frac{\mathbb{P}(\boldsymbol{s})\exp\left(-\frac{\beta}{2}\|\boldsymbol{y}\|^2\right)}{(2\pi/\beta)^{k/2}\left(\sqrt{2\pi\sigma^2}\right)^{|\mathcal{S}|}\det^{1/2}\left[\frac{1}{2\pi}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)\right]}$$

$$\times \exp\left\{\frac{1}{2}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)^T\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\right\} \tag{63}$$

$$= C\cdot\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\frac{\exp\left\{\frac{1}{2}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\right\}}{\sqrt{\det\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)}} \tag{64}$$

where C is independent of $\boldsymbol{\lambda}$, but depends on β and $\boldsymbol{y}$. We are now in a position to find a preliminary expression of the MMSE, using Lemma 1. Let

$$\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\triangleq\exp\left\{\frac{1}{2}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)-\frac{1}{2}\ln\det\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)\right\}, \tag{65}$$

and therefore

$$Z\left(\boldsymbol{y},\boldsymbol{H};\boldsymbol{\lambda}\right)=C\cdot\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{\lambda}_{\boldsymbol{s}}\right). \tag{66}$$

Now,

$$\frac{\partial}{\partial\lambda_i}\left\{\frac{1}{2}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\right\}=\boldsymbol{e}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\mathbb{1}_{i\in\mathcal{S}}, \tag{67}$$

and thus

$$\frac{\partial}{\partial\lambda_i}\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{\lambda}_{\boldsymbol{s}}\right)=\boldsymbol{e}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\mathbb{1}_{i\in\mathcal{S}}\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{\lambda}_{\boldsymbol{s}}\right). \tag{68}$$

Recall that for a positive, twice differential function f,

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln f(x)=\frac{1}{f(x)}\left(\frac{\mathrm{d}}{\mathrm{d}x}f(x)\right) \tag{69}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\ln f(x)=\frac{1}{f(x)}\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2}f(x)\right)-\frac{1}{[f(x)]^2}\left(\frac{\mathrm{d}}{\mathrm{d}x}f(x)\right)^2. \tag{70}$$

Thus, using (66) and (68), we have that (for $1\leq i\leq n$),

$$\frac{\partial}{\partial\lambda_i}\ln Z\left(\boldsymbol{y},\boldsymbol{H};\boldsymbol{\lambda}\right)=\frac{\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,\boldsymbol{e}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\mathbb{1}_{i\in\mathcal{S}}\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{\lambda}_{\boldsymbol{s}}\right)}{Z\left(\boldsymbol{y},\boldsymbol{H};\boldsymbol{\lambda}\right)}. \tag{71}$$

Let us calculate the second derivative. First, using (70) we may write

$$\frac{\partial^2}{\partial\lambda_i^2}\ln Z\left(\boldsymbol{y},\boldsymbol{H};\boldsymbol{\lambda}\right)=\frac{\frac{\partial}{\partial\lambda_i}\left(\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,\boldsymbol{e}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\mathbb{1}_{i\in\mathcal{S}}\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\right)}{Z\left(\boldsymbol{y},\boldsymbol{H};\boldsymbol{\lambda}\right)}$$

$$-\frac{\left(\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,\boldsymbol{e}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\mathbb{1}_{i\in\mathcal{S}}\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\right)^2}{\left[Z\left(\boldsymbol{y},\boldsymbol{H};\boldsymbol{\lambda}\right)\right]^2}. \tag{72}$$

We have that

$$\frac{\partial}{\partial \lambda_i}\left\{\boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right) \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}, \boldsymbol{\lambda}_{\boldsymbol{s}}\right)\right\}=\boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}, \boldsymbol{\lambda}_{\boldsymbol{s}}\right) + \boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right) \boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right) \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}, \boldsymbol{\lambda}_{\boldsymbol{s}}\right) \tag{73}$$

$$=\boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}, \boldsymbol{\lambda}_{\boldsymbol{s}}\right)+\boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right)^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}, \boldsymbol{\lambda}_{\boldsymbol{s}}\right). \tag{74}$$

Let $\xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right) \triangleq \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}, \mathbf{0}\right)$. Hence,

$$\left.\frac{\partial}{\partial \lambda_i}\left\{\boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}+\boldsymbol{\lambda}_{\boldsymbol{s}}\right) \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}, \boldsymbol{\lambda}_{\boldsymbol{s}}\right)\right\}\right|_{\boldsymbol{\lambda}=0}=\boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right) + \boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \beta^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y} \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right). \tag{75}$$

Thus, substituting the last result in (72), evaluated at $\boldsymbol{\lambda}=\mathbf{0}$, we obtain

$$\left.\frac{\partial^2}{\partial \lambda_i^2} \ln Z(\boldsymbol{y}, \boldsymbol{H} ; \boldsymbol{\lambda})\right|_{\boldsymbol{\lambda}=0}=\frac{\sum_{\boldsymbol{s} \in\{0,1\}^n} \mathbb{P}(\boldsymbol{s}) \boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right)}{\sum_{\boldsymbol{s} \in\{0,1\}^n} \mathbb{P}(\boldsymbol{s}) \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right)} + \frac{\sum_{\boldsymbol{s} \in\{0,1\}^n} \mathbb{P}(\boldsymbol{s}) \boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \beta^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y} \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right)}{\sum_{\boldsymbol{s} \in\{0,1\}^n} \mathbb{P}(\boldsymbol{s}) \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right)} - \frac{\left[\sum_{\boldsymbol{s} \in\{0,1\}^n} \mathbb{P}(\boldsymbol{s}) \boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y} \mathbb{1}_{i \in \mathcal{S}} \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right)\right]^2}{\left[\sum_{\boldsymbol{s} \in\{0,1\}^n} \mathbb{P}(\boldsymbol{s}) \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right)\right]^2}. \tag{76}$$

By Lemma 1, in order to obtain the MMSE, we need to sum the above equations over $1 \leq i \leq n$. Recall that for an $n \times n$ matrix $\boldsymbol{A}$, the trace operator can be represented as $\operatorname{tr}(\boldsymbol{A})=\sum_{i=1}^n \hat{\boldsymbol{e}}_i^T \boldsymbol{A} \hat{\boldsymbol{e}}_i$ where $\hat{\boldsymbol{e}}_i$ is the ith column of the $n \times n$ identity matrix. Thus, we have that

$$\sum_{i=1}^n \boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}}=\operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \tag{77}$$

$$\sum_{i=1}^n \boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \beta^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y} \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}}=\sum_{i=1}^n \operatorname{tr}\left(\boldsymbol{e}_i^T \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \beta^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y} \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{e}_i \mathbb{1}_{i \in \mathcal{S}}\right) \tag{78}$$

$$=\operatorname{tr}\left(\boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \beta^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y} \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \sum_{i=1}^n \boldsymbol{e}_i \boldsymbol{e}_i^T \mathbb{1}_{i \in \mathcal{S}}\right) \tag{79}$$

$$=\beta^2 \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}. \tag{80}$$

Accordingly, let us define

$$J_1\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right) \triangleq \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}+\frac{\beta^2}{n} \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}. \tag{81}$$

Now, the summation (normalized by n) over the first two terms in the r.h.s. of (76) gives

$$\frac{1}{\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,\xi(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}})}\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,J_1(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}})\,\xi(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}})\,. \tag{82}$$

Regarding the last term in the r.h.s. of (76), we have

$$\begin{aligned}&\left[\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,\boldsymbol{e}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\mathbb{1}_{i\in\mathcal{S}}\xi(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}})\right]^2\\&=\sum_{\boldsymbol{s}\in\{0,1\}^n}\sum_{\boldsymbol{r}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,\mathbb{P}(\boldsymbol{r})\,\boldsymbol{e}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\beta^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{r}}\boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\tilde{\boldsymbol{e}}_i\mathbb{1}_{i\in\mathcal{S}\cap\mathcal{R}}\xi(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}})\,\xi(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{r}})\,.\end{aligned} \tag{83}$$

Note that $\boldsymbol{s}$ and $\boldsymbol{r}$ may not have the same support, and in particular, they may not have even the same support size. This explains the appearance of $\tilde{\boldsymbol{e}}_i$ which is of size $|\mathcal{R}|\times 1$. Now, we have that

$$\sum_{i=1}^n\boldsymbol{e}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\beta^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{r}}\boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\tilde{\boldsymbol{e}}_i\mathbb{1}_{i\in\mathcal{S}\cap\mathcal{R}}=\sum_{i=1}^n\mathrm{tr}\left(\boldsymbol{e}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\beta^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{r}}\boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\tilde{\boldsymbol{e}}_i\mathbb{1}_{i\in\mathcal{S}\cap\mathcal{R}}\right) \tag{84}$$

$$=\mathrm{tr}\left(\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\beta^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{r}}\boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\sum_{i=1}^n\tilde{\boldsymbol{e}}_i\boldsymbol{e}_i^T\mathbb{1}_{i\in\mathcal{S}\cap\mathcal{R}}\right) \tag{85}$$

$$=\beta^2\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{s}}\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{y} \tag{86}$$

where we have used the fact that

$$\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T=\sum_{i=1}^n\tilde{\boldsymbol{e}}_i\boldsymbol{e}_i^T\mathbb{1}_{i\in\mathcal{S}\cap\mathcal{R}}. \tag{87}$$

Let us define

$$J_2(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{H}_{\boldsymbol{r}})\triangleq\frac{\beta^2}{n}\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{s}}\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{y}. \tag{88}$$

Therefore, the summation (normalized by n) of the third term over $1\leq i\leq n$ reads

$$\frac{1}{\left(\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,\xi(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}})\right)^2}\sum_{\boldsymbol{s}\in\{0,1\}^n}\sum_{\boldsymbol{r}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{s})\,\mathbb{P}(\boldsymbol{r})\,J_2(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{H}_{\boldsymbol{r}})\,\xi(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}})\,\xi(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{r}})\,. \tag{89}$$

Finally, the difference between (82) and (89) gives the normalized MMSE, which can be represented as

$$\frac{\mathrm{mmse}(\boldsymbol{X}\mid\boldsymbol{Y},\boldsymbol{H})}{n}=\mathbb{E}\left\{\mathbb{E}_{\mu_s}\left[J_1(\boldsymbol{Y},\boldsymbol{H}_{\boldsymbol{s}})\right]-\mathbb{E}_{\mu_{s\times r}}\left[J_2(\boldsymbol{Y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{H}_{\boldsymbol{r}})\right]\right\} \tag{90}$$

where $\mathbb{E}_{\mu_s}$ denotes the expectation taken w.r.t. the discrete measure

$$\mu(\boldsymbol{s}\mid\boldsymbol{Y},\boldsymbol{H})\triangleq\frac{\mathbb{P}(\boldsymbol{s})\,\xi(\boldsymbol{Y},\boldsymbol{H}_{\boldsymbol{s}})}{\sum_{\boldsymbol{u}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{u})\,\xi(\boldsymbol{Y},\boldsymbol{H}_{\boldsymbol{u}})}, \tag{91}$$

and $\mathbb{E}_{\mu_{s\times r}}$ denotes the expectation taken w.r.t. the discrete product measure

$$\mu(\boldsymbol{s}\mid\boldsymbol{Y},\boldsymbol{H})\cdot\mu(\boldsymbol{r}\mid\boldsymbol{Y},\boldsymbol{H})\triangleq\frac{\mathbb{P}(\boldsymbol{s})\,\mathbb{P}(\boldsymbol{r})\,\xi(\boldsymbol{Y},\boldsymbol{H}_{\boldsymbol{s}})\,\xi(\boldsymbol{Y},\boldsymbol{H}_{\boldsymbol{r}})}{\left[\sum_{\boldsymbol{u}\in\{0,1\}^n}\mathbb{P}(\boldsymbol{u})\,\xi(\boldsymbol{Y},\boldsymbol{H}_{\boldsymbol{u}})\right]^2}. \tag{92}$$

At this stage, the relation to the Stieltjes and Shannon transforms is clear: The structure of the various terms in ξ, J_1, J_2 suggest an application of an extended version of the Stieltjes and Shannon transforms of the matrix $\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}$.

The following proposition is essentially the core of our analysis; it provides approximations (which are asymptotically exact in the a.s. sense) of ξ, J_1, J_2. Before stating the proposition, we define the following terms. Let $m_s \triangleq \frac{1}{n}\sum_{i=1}^n s_i$, $m_r \triangleq \frac{1}{n}\sum_{i=1}^n r_i$, and $m_{s,r} \triangleq \frac{1}{n}\sum_{i=1}^n s_i r_i$, and recall the auxiliary variables defined in (9)-(17). The following results are proved in Appendix B.

Proposition 2 (Asymptotic approximations) Under the assumptions and definition presented earlier, the following relations hold in the a.s. sense:

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} = \sigma^2 m_s b\left(m_s\right), \tag{93}$$

$$\lim_{n\to\infty} \frac{1}{n} \ln\det\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right) = m_s \bar{I}\left(m_s\right), \tag{94}$$

$$\lim_{n\to\infty} \frac{1}{n} \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y} - f_n = 0, \tag{95}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}} \boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{y} - q_n = 0, \tag{96}$$

where

$$f_n \triangleq \beta \frac{\sigma^4 b^2\left(m_s\right) m_s^2}{g^2\left(m_s\right)} \frac{\left\|\boldsymbol{y}\right\|^2}{n} + \frac{\sigma^2 b\left(m_s\right)}{g^2\left(m_s\right)} \frac{\left\|\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}\right\|^2}{n}, \tag{97}$$

and (with some abuse of notations $\alpha = \alpha\left(m_s, m_r, m_{s,r}\right)$)

$$\begin{aligned} q_n \triangleq & \frac{\alpha}{g\left(m_s\right) g\left(m_r\right)} \frac{\boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{y}}{n} \\ & - \frac{\alpha}{g\left(m_s\right) g\left(m_r\right)} \beta\sigma^2 m_{s,r} \left(\frac{b\left(m_r\right)}{g\left(m_r\right)} \frac{\left\|\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{y}\right\|^2}{n} + \frac{b\left(m_s\right)}{g\left(m_s\right)} \frac{\left\|\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y}\right\|^2}{n}\right) \\ & + \frac{\alpha}{g\left(m_s\right) g\left(m_r\right)} \beta\sigma^2 m_{s,r} \left(\frac{b\left(m_r\right)}{g\left(m_r\right)} m_r + \frac{b\left(m_s\right)}{g\left(m_s\right)} m_s\right) \frac{\left\|\boldsymbol{y}\right\|^2}{n}. \end{aligned} \tag{98}$$

The next step is to apply Proposition 2 to the obtained MMSE. The main observation here is as follows: Let $\epsilon > 0$ and define

$$\begin{aligned} \mathcal{T}_\epsilon^{\boldsymbol{s},\boldsymbol{r}} \triangleq \Bigg\{ & \boldsymbol{y} \in \mathbb{R}^{k\times 1}, \boldsymbol{H} \in \mathbb{R}^{k\times n} : \ \left|\frac{1}{n}\operatorname{tr}\boldsymbol{\mathcal{H}}^{\boldsymbol{s}} - \sigma^2 m_s b\left(m_s\right)\right| < \epsilon, \ \left|\frac{1}{n}\boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{y} - f_n\right| < \epsilon, \\ & \left|\frac{1}{n}\boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}} \boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{y} - q_n\right| < \epsilon, \ \left|\frac{1}{n}\ln\det\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right) - m_s \bar{I}\left(m_s\right)\right| < \epsilon \Bigg\} \end{aligned} \tag{99}$$

and

$$\mathcal{T}_\epsilon \triangleq \bigcup_{\boldsymbol{s},\boldsymbol{r}\in\{0,1\}^n\times\{0,1\}^n} \mathcal{T}_\epsilon^{\boldsymbol{s},\boldsymbol{r}}. \tag{100}$$

By Proposition 2, this set has probability tending to one as $k, n \to \infty$. Accordingly, $\mathcal{T}_\epsilon$ is the set of "typical" $\{\boldsymbol{y}, \boldsymbol{H}\}$-pairs of observation vectors and sensing matrices. The main purpose is to calculate the following quantity

$$\begin{aligned}\mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n\left\{X_i^2 - \left(\mathbb{E}\left\{X_i \mid \boldsymbol{y},\boldsymbol{H}\right\}\right)^2\right\}\right\} =& \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n\left\{X_i^2 - \left(\mathbb{E}\left\{X_i \mid \boldsymbol{y},\boldsymbol{H}\right\}\right)^2\right\}\mathbb{1}_{\mathcal{T}_\epsilon}\right\} \\ &+ \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n\left\{X_i^2 - \left(\mathbb{E}\left\{X_i \mid \boldsymbol{y},\boldsymbol{H}\right\}\right)^2\right\}\mathbb{1}_{\mathcal{T}_\epsilon^c}\right\}\end{aligned} \tag{101}$$

where $\mathcal{T}_\epsilon^c$ is the complementary (w.r.t. $\mathbb{R}^k \times \mathbb{R}^{k\times n}$) of $\mathcal{T}_\epsilon$. However, by using the Cauchy-Schwartz inequality we have that

$$\left|\mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n\left\{X_i^2 - \left(\mathbb{E}\left\{X_i \mid \boldsymbol{y},\boldsymbol{H}\right\}\right)^2\right\}\mathbb{1}_{\mathcal{T}_\epsilon^c}\right\}\right|^2 \leq \left|\mathbb{E}\left\{\frac{1}{n}\left\|\boldsymbol{X}\right\|^2\mathbb{1}_{\mathcal{T}_\epsilon^c}\right\}\right|^2 \tag{102}$$

$$\leq \mathbb{P}\left\{\mathcal{T}_\epsilon^c\right\}\mathbb{E}\left\{\frac{1}{n^2}\left\|\boldsymbol{X}\right\|^4\right\}, \tag{103}$$

but, since $\mathbb{E}\left\{\frac{1}{n^2}\left\|\boldsymbol{X}\right\|^4\right\}$ is bounded (for any n), and $\mathbb{P}\left\{\mathcal{T}_\epsilon^c\right\} \to 0$ as $n \to \infty$, it follows that the last expectation asymptotically vanishes. Thus, for the asymptotic calculation of the MMSE, only the first term at the r.h.s. of (101) prevails.

Note that

$$\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{s}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{y} = \sum_{i=1}^n\left|\boldsymbol{h}_i^T\boldsymbol{y}\right|^2 s_i r_i, \tag{104}$$

$$\left\|\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\right\|^2 = \sum_{i=1}^n\left|\boldsymbol{h}_i^T\boldsymbol{y}\right|^2 s_i, \tag{105}$$

and

$$\left\|\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{y}\right\|^2 = \sum_{i=1}^n\left|\boldsymbol{h}_i^T\boldsymbol{y}\right|^2 r_i. \tag{106}$$

Using Proposition 2 (along with the previous typicality considerations), and large deviations theory, the asymptotic MMSE given in Theorem 1 is derived in Appendix C.

V. Conclusion

In this paper, we considered the calculation of the asymptotic MMSE calculation under sparse representation modeling. As opposed to the popular worst-case approach, we adopt a statistical framework for compressed sensing by modeling the input signal as a random process rather than as an individual sequence. In contrast to previous derivations, which were based on the (non-rigorous) replica method, the analysis carried out in this paper is rigorous. The derivation builds upon a simple relation between the MMSE and a certain function, which can be viewed as a partition function, and hence can be analyzed using methods of statistical mechanics. It was shown that the MMSE can be represented in a special form that contains functions of the Stieltjes and Shannon transforms. This observation allowed us to invoke some powerful results from RMT concerning the asymptotic behavior of these transforms. Although our asymptotic MMSE formula seems to be different from the one that is obtained by the replica method, numerical calculations suggest that they are actually the same. This supports the results of the replica method.

Finally, we believe that the tools developed in this paper, for handling the MMSE, can be used in order to obtain the MMSE estimator itself. An example for such calculation can be found in a recent paper [26], where the MMSE (or, more generally, the mismatched MSE), along with the estimator itself, were derived for a model of a codeword (from a randomly selected code), corrupted by a Gaussian vector channel. Also, we believe that our results, can be generalized to the case of mismatch, namely, mismatched compressed sensing. An example for an interesting mismatch model could be a channel mismatch, namely, the receiver has a wrong assumption on the channel $\boldsymbol{H}$, which can be modeled as $\hat{\boldsymbol{H}} = \tau\boldsymbol{H} + \sqrt{1-\tau^2}\boldsymbol{Q}$, where $\boldsymbol{Q}$ is some random matrix, independent of $\boldsymbol{H}$, and $0 \leq \tau \leq 1$ quantifies the proximity between $\hat{\boldsymbol{H}}$ and $\boldsymbol{H}$. Another mismatch configuration could be noise-variance mismatch, namely, the receiver has wrong knowledge about the noise variance. It is then interesting to investigate the resulted MSE in these cases, and in particular, to check whether there are new phase transitions caused by the mismatch.

Appendix A

Proof of Proposition 1: Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be a sequence of i.i.d. random matrices (the subscript index designates the “left” matrix dimension k) defined over the probability space $(\mathcal{X}, \mathcal{F}_X, \mu_{\mathcal{X}})$, and let $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \ldots$ be a sequence of random matrices defined over the probability space $(\mathcal{D}, \mathcal{F}_D, \mu_{\mathcal{D}})$. Now, let $(\mathcal{X}\times\mathcal{D}, \mathcal{F}_X \times \mathcal{F}_D, \mu_{\mathcal{X}\times\mathcal{D}})$ be the respective product space. Obviously, since $\boldsymbol{Q}_m \triangleq \boldsymbol{\Theta}_m \left(\boldsymbol{B}_m - z\boldsymbol{I}_m\right)^{-1}$ is determined by $\boldsymbol{X}_m$ and $\boldsymbol{\Theta}_m$, we can write every possible sequence $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \ldots = \boldsymbol{Q}_1(x,d), \boldsymbol{Q}_2(x,d), \ldots$ for some $(x,d) \in \mathcal{X}\times\mathcal{D}$. Accordingly, we need to prove that the set

$$\mathcal{A} \triangleq \left\{(x,d) \in \mathcal{X}\times\mathcal{D} : \text{ Lemma 4 holds true}\right\}$$

has probability one in our product space. From Tonelli's theorem [38],

$$\mathbb{P}\{\mathcal{A}\} = \int_{\mathcal{A}} \mathrm{d}\mu_{\mathcal{X}\times\mathcal{D}}(x,d) \tag{A.1}$$

$$= \int_{\mathcal{X}\times\mathcal{D}} \mathbb{1}_{\mathcal{A}}(x,d)\,\mathrm{d}\mu_{\mathcal{X}\times\mathcal{D}}(x,d) \tag{A.2}$$

$$= \int_{\mathcal{D}} \left[\int_{\mathcal{X}} \mathbb{1}_{\mathcal{A}}(x,d)\,\mathrm{d}\mu_{\mathcal{X}}(x)\right] \mathrm{d}\mu_{\mathcal{D}}(d). \tag{A.3}$$

Now, let $d_0 \in \mathcal{D}$ be a realization of $\boldsymbol{\Theta}_m$, namely, a sequence of matrices $\boldsymbol{\Theta}_1(d_0), \boldsymbol{\Theta}_2(d_0), \ldots$ such that $\boldsymbol{\Theta}_m$ maintains the boundedness condition (or distribution tightness). Accordingly, for this d_0 we can apply Lemma 4. Namely, the set of realizations x such that $(x, d_0) \in \mathcal{A}$ has probability one, and therefore, for this d_0 we have that

$$\int_{\mathcal{X}} \mathbb{1}_{\mathcal{A}}(x,d)\,\mathrm{d}\mu_{\mathcal{X}}(x) = 1.$$

Let $\mathcal{B} \subset \mathcal{D}$ be the set of all realizations d such that $\boldsymbol{\Theta}_m$ maintains the boundedness condition. Then,

$$\mathbb{P}\{\mathcal{A}\} = \int_{\mathcal{B}} \mathrm{d}\mu_{\mathcal{D}}(d) + \int_{\mathcal{D}\setminus\mathcal{B}} \left[\int_{\mathcal{X}} \mathbb{1}_{\mathcal{A}}(x,d)\,\mathrm{d}\mu_{\mathcal{X}}(x)\right] \mathrm{d}\mu_{\mathcal{D}}(d) \tag{A.4}$$

$$\geq 1 \tag{A.5}$$

where the last equality follows from the fact that the boundedness condition happens w.p. 1. ■

Appendix B

Proof of Proposition 2: As can be seen from Proposition 2, we will deal with terms which consist of scalar functions (e.g. Stieltjes and Shannon) of the following matrix

$$\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} = \sigma^2\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}.$$

In the following analysis, we need to use Lemmas 3 and 4, where the central quantity to be calculated is $S(z)$ given in (48). Indeed, given $S(z)$, using (47) and (49), we will obtain the limit of the Shannon and Stieltjes transforms. Accordingly, we substitute in these lemmas: $\boldsymbol{X} = \boldsymbol{H}_{\boldsymbol{s}}^T$, $\boldsymbol{G} = \beta\sigma^2 R \boldsymbol{I}_{\boldsymbol{s}}$, $c = |\mathcal{S}|/k = m_s/R$. Note that by using these substitutions, we obtain $\boldsymbol{B} = \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^T = \beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}$. Then, using (48) for $z = -1$, we obtain that $S(-1)$ is given by the solution of

$$S(-1) = \left(\frac{1}{|\mathcal{S}|}\sum_{l=1}^{|\mathcal{S}|} \frac{g_l}{1 + c g_l S(-1)} + 1\right)^{-1}.$$

Thus, substituting $g_l = \beta\sigma^2 R$ (independently of the index l) and $c = m_s/R$, we obtain

$$S(-1) = \left(\frac{\beta\sigma^2 R}{1 + \beta\sigma^2 R \frac{m_s}{R} S(-1)} + 1\right)^{-1} \tag{B.1}$$

$$= \frac{1+\sigma^2\beta m_s S(-1)}{1+\beta\sigma^2 R+\beta\sigma^2 m_s S(-1)}, \tag{B.2}$$

and thus

$$S(-1) = \frac{-\left[1+\beta\sigma^2(R-m_s)\right]+\sqrt{\left[1+\beta\sigma^2(R-m_s)\right]^2+4\beta\sigma^2 m_s}}{2\beta\sigma^2 m_s}. \tag{B.3}$$

Note that $S(-1)$ is recognized as $b(m_s)$ defined in (9), and which will be used from now on.

A. Derivation of (93) *and* (94)

The results given in (93) and (94) follow directly from Lemmas 3 and 4. Indeed, using Lemma 4 (in particular (49)) with $\boldsymbol{\Theta} = \boldsymbol{I}_{\boldsymbol{s}}$, one obtains that[8]

$$\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} = \sigma^2\frac{1}{n}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \to \sigma^2 m_s b(m_s), \tag{B.4}$$

a.s. as $n\to\infty$.

The second item follows directly from Lemma 3. Recall that

$$\eta(\gamma) \triangleq \frac{1}{k}\sum_{l=1}^{|\mathcal{S}|}\ln\left(1+cg_l S(-\gamma)\right)-\ln\left(\gamma^2 S(-\gamma)\right)-\frac{1}{|\mathcal{S}|}\sum_{l=1}^{|\mathcal{S}|}\frac{g_l S(-\gamma)}{1+cg_l S(-\gamma)}. \tag{B.5}$$

Thus, under our model, and by choosing $\gamma=1$, we obtain

$$\eta(1) = \frac{R}{m_s}\ln\left[1+\beta\sigma^2 b(m_s)m_s\right]-\ln b(m_s)-\frac{\beta\sigma^2 R b(m_s)}{1+\beta\sigma^2 b(m_s)m_s}, \tag{B.6}$$

which is recognized as $\bar{I}(m_s)$ defined in (11), and which will be used from now on. Thus, by Lemma 3, we conclude that

$$\frac{1}{n}\ln\det\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right) \to m_s\bar{I}(m_s) \tag{B.7}$$

a.s. as $n\to\infty$.

B. Derivation of (95)

Equation (95) is closely related to the terms appearing in Lemma 4. However, we cannot directly apply it on our terms, unless we choose $\boldsymbol{\Theta}$ to be dependent on $\boldsymbol{H}$, which is not supported by Proposition 1. Instead, we use the following idea: Let $\boldsymbol{z}_i$ denote the ith row of the matrix $\boldsymbol{H}_{\boldsymbol{s}}$, and hence

$$\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y} = \sum_{i=1}^{k} y_i\boldsymbol{z}_i.$$

[8]Note that the fact that Lemma 4 holds true also for matrices $\boldsymbol{X}$ with a vanishing ratio c is in use here (see discussion after Proposition 1). Indeed, as the summation over the pattern sequences $\boldsymbol{s}$ is over the whole space $\{0,1\}^n$, the ratio m can, in general, vanish.

Thus,

$$\frac{1}{n}\boldsymbol{y}^T\boldsymbol{H_s}\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{H_s}^T\boldsymbol{y}=\frac{1}{n}\sum_{i=1}^{k}y_i^2\boldsymbol{z}_i^T\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i \tag{B.8}$$

$$+\frac{1}{n}\sum_{i\neq j}^{k}y_iy_j\boldsymbol{z}_i^T\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_j. \tag{B.9}$$

Let us start with the first term at the r.h.s. of (B.9). Recall that

$$\boldsymbol{H_s}^T\boldsymbol{H_s}=\sum_{i=1}^{k}\boldsymbol{z}_i\boldsymbol{z}_i^T.$$

In the sequel, we will repeatedly use Lemmas 7-15, which all appear in Appendix D. Using the matrix inversion lemma (Lemma 7), we have that

$$\frac{1}{n}\sum_{i=1}^{k}y_i^2\boldsymbol{z}_i^T\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i=\frac{1}{n}\sum_{i=1}^{k}y_i^2\frac{\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i}{1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i}. \tag{B.10}$$

Since the matrix $\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}$ is statistically independent on $\boldsymbol{z}_i$, we can write

$$\frac{1}{n}\sum_{i=1}^{k}\frac{y_i^2\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i}{1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i}\asymp\frac{1}{n}\sum_{i=1}^{k}\frac{y_i^2\frac{1}{n}\operatorname{tr}\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}}{1+\beta\frac{1}{n}\operatorname{tr}\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}} \tag{B.11}$$

$$\asymp\frac{1}{n}\sum_{i=1}^{k}\frac{y_i^2\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}}{1+\beta\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}} \tag{B.12}$$

$$\asymp\frac{1}{n}\sum_{i=1}^{k}\frac{y_i^2m_s\sigma^2b\left(m_s\right)}{1+\beta\sigma^2m_sb\left(m_s\right)} \tag{B.13}$$

$$=\frac{m_s\sigma^2b\left(m_s\right)}{1+\beta\sigma^2m_sb\left(m_s\right)}\frac{\left\|\boldsymbol{y}\right\|^2}{n} \tag{B.14}$$

where in the first passage, we applied the trace lemma (Lemma 11) and Lemma 12, in the second passage we have used the rank-1 perturbation lemma (Lemma 13), and the third passage is due to Lemma 4 (actually the first item of Proposition 2). In the following, we provide a rigorous justification to the above derivation. We first show that the first passage is true, namely, that we have a.s.,

$$\frac{1}{n}\sum_{i=1}^{k}y_i^2\left(\frac{\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i}{1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i}-\frac{\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}}{1+\beta\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}}\right)\to 0. \tag{B.15}$$

There are at least two approaches to prove the last statement: using a graph-combinatorial method (very powerful but tedious), or the following approach. By Lemma 2, it is enough to prove that

$$\max_{1\leq i\leq k}\mathbb{E}\left\{\left|\frac{\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i}{1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H_s}^T\boldsymbol{H_s}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}\boldsymbol{z}_i}-\frac{\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}}{1+\beta\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H_s}^T\boldsymbol{H_s}+\frac{1}{\sigma^2}\boldsymbol{I_s}\right)^{-1}}\right|^p\right\}\leq\mathcal{O}\left(\frac{1}{n^{1+\delta}}\right), \tag{B.16}$$

for $\delta > 0$. Instead of showing (B.16), we will equivalently show that[9]

$$\max_{1\le i\le k} \mathbb{E}\left\{\left|\frac{\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i-\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i}\right|^p\right\} \le \mathcal{O}\left(\frac{1}{n^{1+\delta}}\right), \tag{B.17}$$

and that

$$\max_{1\le i\le k} \mathbb{E}\left\{\left|\frac{\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i}-\frac{\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1+\beta\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}\right|^p\right\} \le \mathcal{O}\left(\frac{1}{n^{1+\delta}}\right). \tag{B.18}$$

We now show (B.17). First, note that

$$|e_i| \triangleq \left|\frac{\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i-\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i}\right| \tag{B.19}$$

$$\overset{(a)}{\le} \left|\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i-\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right| \tag{B.20}$$

$$\overset{(b)}{\le} \left|\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i-\frac{1}{n}\operatorname{tr}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right|$$
$$+\left|\frac{1}{n}\operatorname{tr}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right| \tag{B.21}$$

where (a) follows from the fact that $\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i$ is non-negative, and thus,

$$\frac{1}{1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i} \le 1, \tag{B.22}$$

and (b) follows by adding and subtracting the term $\frac{1}{n}\operatorname{tr}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}$, and then using the triangle inequality. Applying Lemma 14 to the second term in the r.h.s. of (B.21), one readily obtains that

$$\left|\frac{1}{n}\operatorname{tr}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-\frac{1}{n}\operatorname{tr}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right| \le \frac{\sigma^2\left\|\boldsymbol{I}_{\boldsymbol{s}}\right\|}{n} = \frac{\sigma^2}{n}, \tag{B.23}$$

uniformly in $\boldsymbol{s}$. Applying Lemma 10 to the first term at the r.h.s. of (B.21), we obtain

$$\mathbb{E}\left\{\left|\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i-\frac{1}{n}\operatorname{tr}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right|^p\right\} \le \frac{\tilde{C}}{n^{p/2}}, \tag{B.24}$$

where according to Lemma 10, the constant $\tilde{C}$ is given by

$$\tilde{C} = C_p\cdot\mathbb{E}\left(\frac{1}{|\mathcal{S}|}\operatorname{tr}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-2}\right)^{p/2} \tag{B.25}$$

[9]The equivalence readily follows by adding and subtracting a common term and then using the triangle inequality.

$$\leq C_p \sigma^{2p}, \tag{B.26}$$

where in the last inequality, we have used the fact that $\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i$ is non-negative, and thus

$$\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \preceq \left(\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} = \sigma^2 \boldsymbol{I}_{\boldsymbol{s}} \tag{B.27}$$

where for two matrices $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{B} \in \mathbb{R}^{N\times N}$ the notation $\boldsymbol{A} \preceq \boldsymbol{B}$ means that the difference $\boldsymbol{B} - \boldsymbol{A}$ is non-negative definite. Thus, the bound in (B.24) is uniform in $\boldsymbol{s}$. Therefore,

$$\mathbb{E}\left\{|e_i|^p\right\} \leq \mathcal{O}\left(\frac{1}{n^{p/2}}\right). \tag{B.28}$$

Thus, taking any $p > 2$, we obtain (B.17). Similarly, for (B.18), we see that

$$\begin{aligned}
|\tilde{e}_i| &\triangleq \left|\frac{\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1 + \beta \boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i} - \frac{\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1 + \beta\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}\right| \\
&= \frac{\left|\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i - \frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right|}{\left(1 + \beta \boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i\right)\left(1 + \beta\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right)}
\end{aligned} \tag{B.29}$$

$$\begin{aligned}
&\overset{(a)}{\leq} \frac{\beta}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\left|\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i - \frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right| \\
&\overset{(b)}{\leq} \beta\sigma^2\left|\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i - \frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right|
\end{aligned} \tag{B.30}$$

where (a) follows from (B.22), and the fact that

$$\frac{1}{1 + \beta\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} \leq 1, \tag{B.31}$$

and (b) follows from

$$\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \leq \frac{1}{n}\operatorname{tr}\left(\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} = \sigma^2 \tag{B.32}$$

Therefore, as before, by applying Lemma 10, we obtain that $\mathbb{E}\left|\tilde{e}_i\right|^p \leq \mathcal{O}\left(n^{-p/2}\right)$, as required. Finally, we show that the error due to the passage from (B.12) to (B.13) can be bounded uniformly in $\boldsymbol{s}$. Indeed, let the error be denoted by

$$\hat{e} \triangleq \frac{\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1 + \beta\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} - \frac{m_s\sigma^2 b\left(m_s\right)}{1 + \beta\sigma^2 m_s b\left(m_s\right)}. \tag{B.33}$$

First, we see that

$$|\hat{e}| = \frac{\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1 + \beta\frac{1}{n}\operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} - \frac{m_s\sigma^2 b\left(m_s\right)}{1 + \beta\sigma^2 m_s b\left(m_s\right)} \tag{B.34}$$

$$= \sigma^2 \frac{\left| \frac{1}{n} \operatorname{tr} \left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} - m_s b\left(m_s\right) \right|}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right) \left(1 + \beta \frac{1}{n} \operatorname{tr}\left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right)} \tag{B.35}$$

$$\leq \sigma^2 \left| \frac{1}{n} \operatorname{tr} \left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} - m_s b\left(m_s\right) \right|, \tag{B.36}$$

where the last inequality follows from (B.31) and the fact that $1+\beta\sigma^2 m_s b\left(m_s\right) \geq 1$. Recall that $b\left(m_s\right)$ is the solution of the following equation (given in (B.2))

$$b\left(m_s\right) = \left(\frac{\beta\sigma^2 R}{1+\beta\sigma^2 m_s b\left(m_s\right)} + 1 \right)^{-1}. \tag{B.37}$$

Let us define

$$w \triangleq \frac{1}{n} \operatorname{tr} \left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} - \frac{1}{n} \operatorname{tr} \left(\frac{R\beta\sigma^2}{1+\beta\sigma^2 \frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} + 1 \right)^{-1} \boldsymbol{I}_{\boldsymbol{s}}. \tag{B.38}$$

Then, note that

$$\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} - \left(\frac{R\beta\sigma^2}{1+\beta\sigma^2 \frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} + 1 \right)^{-1} \boldsymbol{I}_{\boldsymbol{s}}$$

$$\stackrel{(a)}{=} \left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \left[\frac{R\beta\sigma^2}{1+\beta\sigma^2 \frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} \boldsymbol{I}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} - \beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} - \boldsymbol{I}_{\boldsymbol{s}} \right]$$

$$\left(\frac{R\beta\sigma^2}{1+\beta\sigma^2 \frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} + 1 \right)^{-1} \boldsymbol{I}_{\boldsymbol{s}} \tag{B.39}$$

$$= \left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \left[\frac{R\beta\sigma^2}{1+\beta\sigma^2 \frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} \boldsymbol{I}_{\boldsymbol{s}} - \beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \right]$$

$$\left(\frac{R\beta\sigma^2}{1+\beta\sigma^2 \frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} + 1 \right)^{-1} \boldsymbol{I}_{\boldsymbol{s}} \tag{B.40}$$

$$= -\vartheta \left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}$$

$$+ \vartheta \left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \frac{R\beta\sigma^2}{1+\beta\sigma^2 \frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} \tag{B.41}$$

where (a) is due to Lemma 9, and in the last equalities we canceled out and rearranged the various terms, and

$$\vartheta \triangleq \left(\frac{R\beta\sigma^2}{1+\beta\sigma^2 \frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} + 1 \right)^{-1}. \tag{B.42}$$

Therefore, using (B.41) we obtain

$$w = \frac{1}{n} \operatorname{tr} \left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} - \frac{1}{n} \operatorname{tr} \left(\frac{R\beta\sigma^2}{1+\beta\sigma^2 \frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} + 1 \right)^{-1} \boldsymbol{I}_{\boldsymbol{s}}$$

$$\stackrel{\text{(B.41)}}{=} -\varpi\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right)+\varpi\frac{R\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} \tag{B.43}$$

$$= -\varpi\frac{1}{n}\sum_{i=1}^{k}\beta\sigma^2\boldsymbol{z}_i^T\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i+\varpi\frac{1}{n}\sum_{i=1}^{k}\frac{\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}} \tag{B.44}$$

where in the last equality we have used the fact that $R = k/n$, and that

$$\operatorname{tr}\left(\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right)=\operatorname{tr}\left(\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\sum_{i=1}^{k}\boldsymbol{z}_i\boldsymbol{z}_i^T\right) \tag{B.45}$$

$$=\sum_{i=1}^{k}\boldsymbol{z}_i^T\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i. \tag{B.46}$$

Therefore

$$|w| = \left|\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-\frac{1}{n}\operatorname{tr}\left(R\frac{\beta\sigma^2}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}+1\right)^{-1}\boldsymbol{I}_{\boldsymbol{s}}\right|$$

$$\stackrel{\text{(B.44)}}{=}\left|\varpi\frac{1}{n}\sum_{i=1}^{k}\beta\sigma^2\boldsymbol{z}_i^T\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i-\varpi\frac{1}{n}\sum_{i=1}^{k}\frac{\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}\right|$$

$$=\left|\varpi\frac{1}{n}\sum_{i=1}^{k}\left[\beta\sigma^2\boldsymbol{z}_i^T\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i-\frac{\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}\right]\right|$$

$$\stackrel{(a)}{=}\varpi\frac{1}{n}\left|\sum_{i=1}^{k}\left[\frac{\beta\sigma^2\boldsymbol{z}_i^T\operatorname{tr}\left(\beta\sigma^2\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i}{1+\beta\sigma^2\boldsymbol{z}_i^T\operatorname{tr}\left(\beta\sigma^2\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i}-\frac{\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}\right]\right| \tag{B.47}$$

where (a) follows by the matrix inversion lemma (Lemma 7). Now, note that

$$\left|\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-m_sb\left(m_s\right)\right|$$

$$\stackrel{(a)}{=}\left|\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-m_s\left(\frac{R\beta\sigma^2}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}+1\right)^{-1}\right.$$

$$\left.+m_s\left(\frac{R\beta\sigma^2}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}+1\right)^{-1}-m_sb\left(m_s\right)\right|$$

$$\stackrel{(b)}{\leq}\left|\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-m_s\left(\frac{R\beta\sigma^2}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}+1\right)^{-1}\right|$$

$$+m_s\left|\left(\frac{R\beta\sigma^2}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}+1\right)^{-1}-b\left(m_s\right)\right| \tag{B.48}$$

$$\stackrel{(c)}{=}|w|+m_s\left|\left(\frac{R\beta\sigma^2}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}+1\right)^{-1}-b\left(m_s\right)\right| \tag{B.49}$$

where in (a) we added and subtracted a common term, in (b) we have used the triangle inequality, and in (c) we noticed that the first term is w given in (B.38). But using (B.37), we notice that

$$\left|\left(\frac{R\beta\sigma^2}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}+1\right)^{-1}-b\left(m_s\right)\right|$$

$$=\left|\left(\frac{R\beta\sigma^2}{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}+1\right)^{-1}-\left(\frac{\beta\sigma^2R}{1+\beta\sigma^2m_sb\left(m_s\right)}+1\right)^{-1}\right| \tag{B.50}$$

$$=\left|\frac{1+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)}{1+\beta\sigma^2R+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}}-\frac{1+\beta\sigma^2m_sb\left(m_s\right)}{1+\beta\sigma^2R+\beta\sigma^2m_sb\left(m_s\right)}\right| \tag{B.51}$$

$$=\frac{\beta^2\sigma^4R\left|\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-m_sb\left(m_s\right)\right|}{\left(1+\beta\sigma^2R+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right)\left(1+\beta\sigma^2R+\beta\sigma^2m_sb\left(m_s\right)\right)} \tag{B.52}$$

$$\triangleq\kappa\left|\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-m_sb\left(m_s\right)\right| \tag{B.53}$$

where

$$\kappa\triangleq\frac{\beta^2\sigma^4R}{\left(1+\beta\sigma^2R+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right)\left(1+\beta\sigma^2R+\beta\sigma^2m_sb\left(m_s\right)\right)}. \tag{B.54}$$

Thus, using (B.49) and (B.53), we obtain

$$\left|\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-m_sb\left(m_s\right)\right|\leq|w|+\kappa m_s\left|\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}-m_sb\left(m_s\right)\right|. \tag{B.55}$$

In the following, we show that $0<\kappa m_s<1$. First, for $m_s\leq R$ we see that

$$\kappa m_s=\frac{\beta^2\sigma^4Rm_s}{\left(1+\beta\sigma^2R+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right)\left(1+\beta\sigma^2R+\beta\sigma^2m_sb\left(m_s\right)\right)} \tag{B.56}$$

$$\stackrel{(a)}{\leq}\frac{\beta^2\sigma^4R^2}{\left(1+\beta\sigma^2R\right)^2} \tag{B.57}$$

$$\leq 1. \tag{B.58}$$

where (a) follows from the facts that $\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\geq 0$ and that $b\left(m_s\right)\geq 0$. For $m_s>R$, we first note that $b\left(m_s\right)\geq\left(m_s-R\right)/m_s$, which follows from the facts that $b\left(m_s\right)$ is monotonically decreasing in β (by definition), and that

$$\lim_{\beta\to\infty}b\left(m_s\right)=\frac{m_s-R}{m_s}. \tag{B.59}$$

Whence,

$$\kappa m_s=\frac{\beta^2\sigma^4Rm_s}{\left(1+\beta\sigma^2R+\beta\sigma^2\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right)\left(1+\beta\sigma^2R+\beta\sigma^2m_sb\left(m_s\right)\right)} \tag{B.60}$$

$$\leq \frac{\beta^2\sigma^4 R m_s}{(1+\beta\sigma^2 R)\left(1+\beta\sigma^2 R+\beta\sigma^2 m_s \frac{m_s-R}{m_s}\right)} \tag{B.61}$$

$$= \frac{\beta^2\sigma^4 R m_s}{(1+\beta\sigma^2 R)(1+\beta\sigma^2 m_s)} \tag{B.62}$$

$$\leq \frac{\beta^2\sigma^4 R}{(1+\beta\sigma^2 R)(1+\beta\sigma^2)} \leq 1. \tag{B.63}$$

Thus, using (B.55), we obtain

$$\left|\frac{1}{n}\operatorname{tr}\left(\beta\sigma^2 \boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} - m_s b\left(m_s\right)\right| \leq \frac{1}{1-m_s\kappa}\left|w\right| \tag{B.64}$$

$$\leq \tilde{\kappa}\left|w\right| \tag{B.65}$$

where $\tilde{\kappa} > 0$ depends only on β, σ^2, R. But, comparing (B.47) with (B.15), we readily conclude that $|w|$ converges to zero a.s., and uniformly in $\boldsymbol{s}$. Accordingly, based on (B.36) and (B.65), we conclude that the error $|\hat{e}|$ in (B.36) converges to zero a.s., and can be bounded by a vanishing term that is uniform in $\boldsymbol{s}$

Recalling (B.9), in order to finish the proof of (95), it remains to handle the second term on the r.h.s. of (B.9). Essentially, there is nothing different in this term compared to the first term on the r.h.s. of (B.9). Therefore, and for the sake of brevity, in the following, we use the same reasoning as in the passage from (B.11) to (B.14). Nevertheless, the same arguments we have used to show (B.16), can be readily applied also here. First, note that by applying Lemma 7 twice (first, we remove from $\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}$ the ith term, namely, $\boldsymbol{z}_i\boldsymbol{z}_i^T$, and then we remove the jth term), we obtain

$$\frac{1}{n}\sum_{i\neq j}^{k} y_i y_j \boldsymbol{z}_i^T\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_j$$

$$= \frac{1}{n}\sum_{i\neq j}^{k} \frac{y_i y_j \boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_{i,j}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_j}{\left(1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_{i}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i\right)\left(1+\beta\boldsymbol{z}_j^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_{i,j}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_j\right)}, \tag{B.66}$$

and thus the matrix inverse terms are statistically independent of $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$. Now, we use the same arguments as before. Indeed, we may write that

$$\frac{1}{n}\sum_{i\neq j}^{k} \frac{y_i y_j \boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_{i,j}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_j}{\left(1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_{i}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i\right)\left(1+\beta\boldsymbol{z}_j^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_{i,j}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_j\right)} \tag{B.67}$$

$$\overset{(a)}{\asymp} \frac{1}{n}\sum_{i\neq j}^{k} \frac{y_i y_j \boldsymbol{z}_i^T\boldsymbol{z}_j\sigma^2 b\left(m_s\right)}{\left(1+\beta\frac{1}{n}\operatorname{tr}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_{i}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right)\left(1+\beta\frac{1}{n}\operatorname{tr}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_{i,j}+\frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\right)} \tag{B.68}$$

$$\underset{\smile}{\overset{(b)}{}} \frac{1}{n}\sum_{i\neq j}^{k}\frac{y_i y_j \boldsymbol{z}_i^T\boldsymbol{z}_j\sigma^2 b\left(m_s\right)}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)^2} \tag{B.69}$$

$$\overset{(c)}{=} \frac{\sigma^2 b\left(m_s\right)}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)^2}\left[\frac{1}{n}\sum_{i,j=1}^{k} y_i y_j \boldsymbol{z}_i^T\boldsymbol{z}_j - m_s\frac{\left\|\boldsymbol{y}\right\|^2}{n}\right] \tag{B.70}$$

$$= \frac{\sigma^2 b\left(m_s\right)}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)^2}\left[\frac{\left\|\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\right\|^2}{n} - m_s\frac{\left\|\boldsymbol{y}\right\|^2}{n}\right] \tag{B.71}$$

where in (a) we applied Lemma 4 (or more precisely, (50)) to the numerator, and Lemma 10 to the denominator, in (b) we applied Lemma 13 and then Lemma 4 to the denominator, and in (c) we have used (B.14). Therefore, based on (B.14), (B.71), and (B.9), we may conclude that

$$f_n = \frac{m_s\sigma^2 b\left(m_s\right)}{1+\beta\sigma^2 m_s b\left(m_s\right)}\frac{\left\|\boldsymbol{y}\right\|^2}{n} + \frac{\sigma^2 b\left(m_s\right)}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)^2}\left[\frac{\left\|\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\right\|^2}{n} - m_s\frac{\left\|\boldsymbol{y}\right\|^2}{n}\right] \tag{B.72}$$

$$= \frac{\beta\sigma^4 m_s^2 b^2\left(m_s\right)}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)^2}\frac{\left\|\boldsymbol{y}\right\|^2}{n} + \frac{\sigma^2 b\left(m_s\right)}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)^2}\frac{\left\|\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\right\|^2}{n}, \tag{B.73}$$

where in the last equality we have just rearranged terms. Therefore, we obtained (95), as claimed.

Remark 6 Finally, before we turn into the proof of (96), we emphasize that the above derivation shows that the magnitude of the errors that result from the above approximation (e.g., (B.21) and (B.65)), can be upper and lower bounded by a vanishing term of $\mathcal{O}\left(n^{-1-\delta}\right)$ for $\delta > 0$, that is uniform in $\boldsymbol{s}$. These bounds, however, are random variables in general, depending on $\boldsymbol{y}$ and $\boldsymbol{H}$. We will use this fact in the asymptotic evaluation of the MMSE.

C. *Derivation of* (96)

Showing (96) is much more challenging due to the fact that in contrast to (95), we will need to develop new deterministic equivalent results (in the form of Lemma 4), so that we will be able to obtain its asymptotic behavior. It will be seen that the main idea in our derivation is actually based on "guessing" the form of the limit. This idea of guessing the limit is similar to a popular approach in RMT known as Bai-Silverstein method [33].

Let $\boldsymbol{z}_i$ and $\tilde{\boldsymbol{z}}_i$ denote the ith rows of the matrices $\boldsymbol{H}_{\boldsymbol{s}}$ and $\boldsymbol{H}_{\boldsymbol{r}}$, respectively. Then, using the following facts

$$\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{s}} = \sum_{i=1}^{k} y_i\boldsymbol{z}_i^T, \tag{B.74}$$

and

$$\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{y} = \sum_{i=1}^{k} y_i \tilde{\boldsymbol{z}}_i, \tag{B.75}$$

we have that

$$\begin{aligned}
&\frac{1}{n} \boldsymbol{y}^T \boldsymbol{H}_{\boldsymbol{s}} \left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{y} \\
&= \frac{1}{n} \sum_{i=1}^{k} y_i^2 \boldsymbol{z}_i^T \left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \tilde{\boldsymbol{z}}_i \\
&\quad + \frac{1}{n} \sum_{i \neq j}^{k} y_i y_j \boldsymbol{z}_i^T \left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \tilde{\boldsymbol{z}}_j.
\end{aligned} \tag{B.76}$$

Let us start with the first term at the r.h.s. of (B.76). Applying the matrix inversion lemma (Lemma 7) we obtain

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{k} y_i^2 \boldsymbol{z}_i^T \left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \tilde{\boldsymbol{z}}_i \\
&= \frac{1}{n} \sum_{i=1}^{k} \frac{y_i^2 \boldsymbol{z}_i^T \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \tilde{\boldsymbol{z}}_i}{\left(1 + \beta \boldsymbol{z}_i^T \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{z}_i\right) \left(1 + \beta \tilde{\boldsymbol{z}}_i^T \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \tilde{\boldsymbol{z}}_i\right)}.
\end{aligned} \tag{B.77}$$

Note that contrary to the previous case (95), where already at this stage, we were able to continue the asymptotic analysis (e.g. see the passages used to obtain (B.14)), in this case we cannot, because currently, we do not know how the numerator behaves. Thus, in order to continue, we wish to find a real function h_n for which

$$\boldsymbol{z}_i^T \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \tilde{\boldsymbol{z}}_i - h_n \to 0 \tag{B.78}$$

a.s. as $n \to \infty$. First of all, using Lemma 11, we readily obtain that[10]

$$\begin{aligned}
&\boldsymbol{z}_i^T \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \tilde{\boldsymbol{z}}_i \\
&- \frac{1}{n} \operatorname{tr} \left[\left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T\right] \to 0
\end{aligned} \tag{B.79}$$

a.s. as $n \to \infty$. Accordingly, h_n is to be chosen such that

$$\frac{1}{n} \operatorname{tr} \left[\left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T\right] - h_n \to 0. \tag{B.80}$$

[10]Note that this passage is not essential, and can be avoided (for the second term at the r.h.s. of (B.76) this passage will not be used).

To this end, let us choose h_n as follows

$$h_n = \frac{1}{n}\operatorname{tr}\left(\boldsymbol{D}_s^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{D}_r^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) \tag{B.81}$$

where $\boldsymbol{D}_s$ and $\boldsymbol{D}_r$ are two matrices to be determined such that (B.80) holds true. First, note that

$$\begin{aligned}
&\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T - \boldsymbol{D}_s^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{D}_r^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\\
&= \left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\\
&\quad - \boldsymbol{D}_s^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\\
&\quad + \boldsymbol{D}_s^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T - \boldsymbol{D}_s^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{D}_r^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T \qquad \text{(B.82)}\\
&= \left[\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} - \boldsymbol{D}_s^{-1}\right]\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\\
&\quad + \boldsymbol{D}_s^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left[\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} - \boldsymbol{D}_r^{-1}\right]\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T, \qquad \text{(B.83)}
\end{aligned}$$

where in the first equality we added and subtracted a common term, and in the second passage we took out the common factors. Thus, according to (B.80), we wish to show that the trace of the above two terms, when normalized by n, will converge to zero a.s. as $n \to \infty$. Let us start with the first term in (B.83). First, by Lemma 9,

$$\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} - \boldsymbol{D}_s^{-1} = \boldsymbol{D}_s^{-1}\left[\boldsymbol{D}_s - \beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i - \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right]\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}. \tag{B.84}$$

Let us choose $\boldsymbol{D}_s$ as follows

$$\boldsymbol{D}_s = \left(\psi_n + \frac{1}{\sigma^2}\right)\boldsymbol{I}_{\boldsymbol{s}} \tag{B.85}$$

where ψ_n is to be determined such that (B.80) holds true. Thus, substituting the above choice of $\boldsymbol{D}_s$ in (B.84), we obtain

$$\begin{aligned}
\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} - \boldsymbol{D}_s^{-1} &= \psi_n\boldsymbol{D}_s^{-1}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\\
&\quad - \beta\boldsymbol{D}_s^{-1}\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}.
\end{aligned} \tag{B.86}$$

Therefore, the first term of (B.83) reads

$$\left[\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} - \boldsymbol{D}_s^{-1}\right]\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T$$

$$= \psi_n \boldsymbol{D}_s^{-1} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T$$

$$- \beta \boldsymbol{D}_s^{-1} \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \tag{B.87}$$

$$= \psi_n \boldsymbol{D}_s^{-1} \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T - \beta \boldsymbol{D}_s^{-1} \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T. \tag{B.88}$$

For simplicity of notation, we define $C \triangleq 1/\left(\psi_n + 1/\sigma^2\right)$, and recall the notation $\boldsymbol{\mathcal{H}}^{\boldsymbol{s}} \triangleq \left(\beta \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}$. The trace of the second term on the r.h.s. of the above equality can be written as (note that $\boldsymbol{D}_s$ is diagonal)

$$\frac{1}{n} \operatorname{tr} \left(\beta \boldsymbol{D}_s^{-1} \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \right)$$

$$\stackrel{(a)}{=} \frac{1}{n} \operatorname{tr} \left(\beta \boldsymbol{D}_s^{-1} \sum_{j=1}^{k} \boldsymbol{z}_i \boldsymbol{z}_i^T \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \right)$$

$$\stackrel{(b)}{=} \frac{C\beta}{n} \sum_{j=1}^{k} \boldsymbol{z}_j^T \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \boldsymbol{z}_j \tag{B.89}$$

$$\stackrel{(c)}{=} \frac{C\beta}{n} \sum_{j=1}^{k} \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \boldsymbol{z}_j}{1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j} \tag{B.90}$$

where in (a) we have used the fact that $\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} = \sum_{i=1}^{k} \boldsymbol{z}_i \boldsymbol{z}_i^T$, in (b) we have used the cyclic property of the trace operator, and in (c) we have used the matrix inversion lemma. Then applying Lemma 8 on $\left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}$ we obtain

$$\frac{1}{n} \operatorname{tr} \left(\beta \boldsymbol{D}_s^{-1} \boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \left(\beta \left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}} \right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}} \right)^{-1} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \right)$$

$$= \frac{C\beta}{n} \sum_{j=1}^{k} \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \boldsymbol{z}_j}{1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j} - \frac{C\beta}{n} \sum_{j=1}^{k} \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \beta \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \boldsymbol{z}_j}{\left(1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j\right) \left(1 + \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j\right)}. \tag{B.91}$$

Thus, using the last equality, the normalized trace of (B.88) is given by

$$\psi_n C \frac{1}{n} \operatorname{tr} \left(\boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \right) - \beta C \frac{1}{n} \operatorname{tr} \left(\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \right)$$

$$= \psi_n C \frac{1}{n} \operatorname{tr} \left(\boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \right) - \frac{C\beta}{n} \sum_{j=1}^{k} \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \boldsymbol{z}_j}{1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j}$$

$$+ \frac{C\beta}{n} \sum_{j=1}^{k} \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \beta \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \boldsymbol{z}_j}{\left(1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j\right) \left(1 + \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j\right)}. \tag{B.92}$$

Now we are in a position to choose ψ_n. Recalling the trace lemma, by setting

$$\psi_n = \frac{\beta R}{1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}}} - \frac{\beta^2 R \frac{1}{n} \operatorname{tr} \left(\boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s} \cap \boldsymbol{r}}^T \right)}{\left(1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\right) \left(1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\right)}, \tag{B.93}$$

the term on the r.h.s. of (B.92) will converge to zero as n, k grow large. Let us show that this is indeed the right choice. Choosing (B.93), (B.92) can be explicitly written as

$$\psi_n C \frac{1}{n} \operatorname{tr}\left(\boldsymbol{\mathcal{H}}_i^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_i^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) - \frac{C\beta}{n} \sum_{j=1}^{k} \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T \boldsymbol{z}_j}{1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j}$$
$$+ \frac{C\beta}{n} \sum_{j=1}^{k} \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \beta \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T \boldsymbol{z}_j}{\left(1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j\right)\left(1 + \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j\right)} \tag{B.94}$$

$$= \frac{C\beta}{n} \sum_{j=1}^{k} \left[\frac{\frac{1}{n} \operatorname{tr}\left(\boldsymbol{\mathcal{H}}_i^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_i^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right)}{1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}}} - \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T \boldsymbol{z}_j}{1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j} \right]$$
$$+ \frac{C\beta}{n} \sum_{j=1}^{k} \left[\frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \beta \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T \boldsymbol{z}_j}{\left(1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j\right)\left(1 + \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j\right)} - \frac{\frac{1}{n} \operatorname{tr}\left(\boldsymbol{\mathcal{H}}_i^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_i^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) \beta \frac{1}{n} \operatorname{tr}\left(\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right)}{\left(1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\right)\left(1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\right)} \right] \tag{B.95}$$

where the above equality follows by substituting ψ_n, and rearranging the two sums. Now, proving that the first term convergences a.s. to zero as $n \to \infty$, can be shown exactly as was already done in (B.16). The convergence of the second term is essentially very similar to the first term, but with more terms involved (actually the second term can be seen as an extension of the first term). Indeed, by Lemma 2, it is enough to prove that

$$\mathbb{E}\left\{ \left| \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \beta \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T \boldsymbol{z}_j}{\left(1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j\right)\left(1 + \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j\right)} - \frac{\frac{1}{n} \operatorname{tr}\left(\boldsymbol{\mathcal{H}}_i^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_i^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) \beta \frac{1}{n} \operatorname{tr}\left(\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right)}{\left(1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\right)\left(1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\right)} \right|^p \right\}$$
$$\leq \mathcal{O}\left(\frac{1}{n^{1+\delta}}\right), \tag{B.96}$$

or equivalently that (again, we add and subtract a common term and then we use the triangle inequality)

$$\mathbb{E}\left\{ \left| \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \beta \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T \boldsymbol{z}_j - \frac{1}{n} \operatorname{tr}\left(\boldsymbol{\mathcal{H}}_i^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_i^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) \beta \frac{1}{n} \operatorname{tr}\left(\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right)}{\left(1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j\right)\left(1 + \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j\right)} \right|^p \right\}$$
$$\leq \mathcal{O}\left(\frac{1}{n^{1+\delta}}\right), \tag{B.97}$$

and that

$$\mathbb{E}\left\{ \left| \frac{\frac{1}{n} \operatorname{tr}\left(\boldsymbol{\mathcal{H}}_i^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_i^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) \beta \frac{1}{n} \operatorname{tr}\left(\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right)}{\left(1 + \beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{z}_j\right)\left(1 + \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j\right)} - \frac{\frac{1}{n} \operatorname{tr}\left(\boldsymbol{\mathcal{H}}_i^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_i^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) \beta \frac{1}{n} \operatorname{tr}\left(\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right)}{\left(1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\right)\left(1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\right)} \right|^p \right\} \leq \mathcal{O}\left(\frac{1}{n^{1+\delta}}\right). \tag{B.98}$$

Let us show (B.97). First, note that

$$|d_i| \triangleq \left| \frac{\boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i,j}^{r} \beta \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j - \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \beta \frac{1}{n}\operatorname{tr}\left(\boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}^{r} \boldsymbol{Q}_{s\cap r}^T\right)}{\left(1+\beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{s} \boldsymbol{z}_j\right)\left(1+\beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \tilde{\boldsymbol{z}}_j\right)} \right| \tag{B.99}$$

$$\overset{(a)}{\leq} \left| \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i,j}^{r} \beta \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j - \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \beta \frac{1}{n}\operatorname{tr}\left(\boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \right| \tag{B.100}$$

$$\begin{aligned}\overset{(b)}{=} & \left| \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i,j}^{r} \beta \tilde{\boldsymbol{z}}_j \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j - \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j \right. \\ & \left. + \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j - \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \beta \frac{1}{n}\operatorname{tr}\left(\boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \right|\end{aligned} \tag{B.101}$$

$$\begin{aligned}\overset{(c)}{\leq} & \left| \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i,j}^{r} \tilde{\boldsymbol{z}}_j - \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \right| \left| \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j \right| \\ & + \beta \left| \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \right| \left| \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j - \frac{1}{n}\operatorname{tr}\left(\boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \right|\end{aligned} \tag{B.102}$$

where (a) follows from the two obvious facts (that we already used)

$$\frac{1}{1+\beta \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{s} \boldsymbol{z}_j} \leq 1 \tag{B.103}$$

$$\frac{1}{1+\beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \tilde{\boldsymbol{z}}_j} \leq 1, \tag{B.104}$$

(b) follows by adding and subtracting the term

$$\frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j,$$

and (c) follows from the triangle inequality and pulling out the common factor. But using the Cauchy-Schwartz inequality, we may write

$$\begin{aligned}&\mathbb{E}\left\{ \left| \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i,j}^{r} \tilde{\boldsymbol{z}}_j - \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \right|^p \left| \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j \right|^p \right\} \\ &\leq \left(\mathbb{E}\left| \boldsymbol{z}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i,j}^{r} \tilde{\boldsymbol{z}}_j - \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{i}^{s} \boldsymbol{Q}_{s\cap r} \boldsymbol{\mathcal{H}}_{i}^{r} \boldsymbol{Q}_{s\cap r}^T\right) \right|^{2p} \right)^{1/2} \left(\mathbb{E}\left| \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j \right|^{2p} \right)^{1/2}\end{aligned} \tag{B.105}$$

$$\leq \mathcal{O}\left(\frac{1}{n^{p/2}}\right) \tag{B.106}$$

where the last inequality follows from the fact that $\mathbb{E}\left| \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{i,j}^{r} \boldsymbol{Q}_{s\cap r}^T \boldsymbol{z}_j \right|^{2p}$ is bounded (Lemma 15) and by using Lemma 10. The second term in (B.102) is handled similarly. Thus, taking any $p > 2$, we obtain (B.97). Similar arguments can be applied to show that (B.98) holds true.

So, hitherto we show that

$$\psi_n C \frac{1}{n} \operatorname{tr}\left(\boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) - \beta C \frac{1}{n} \operatorname{tr}\left(\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{s}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) \to 0 \tag{B.107}$$

a.s. as $n \to \infty$ where ψ_n is given by (B.93). Next, we consider the second term in (B.83). Using Lemma 9 we may write

$$\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} - \boldsymbol{D}_r^{-1} = \boldsymbol{D}_r^{-1}\left[\boldsymbol{D}_r - \beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i - \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right]\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}. \tag{B.108}$$

Let us choose

$$\boldsymbol{D}_r = \left(\eta_n + \frac{1}{\sigma^2}\right) \boldsymbol{I}_{\boldsymbol{r}}, \tag{B.109}$$

and thus

$$\begin{aligned} \boldsymbol{D}_s^{-1} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left[\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} - \boldsymbol{D}_r^{-1}\right] &= \eta_n \boldsymbol{D}_s^{-1} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{D}_r^{-1}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \\ &\quad - \boldsymbol{D}_s^{-1} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{D}_r^{-1} \beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}. \end{aligned} \tag{B.110}$$

Let $\tilde{C} = 1/\left(\eta_n + 1/\sigma^2\right)$. Then, we have that

$$\frac{1}{n} \operatorname{tr}\left(\boldsymbol{D}_s^{-1} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{D}_r^{-1} \beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i \boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) = \frac{\tilde{C} C \beta}{n} \sum_{j=1}^{k} \frac{\tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \tilde{\boldsymbol{z}}_j}{1 + \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j} \tag{B.111}$$

where, as before, we have used the fact that $\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}} = \sum_{i=1}^k \boldsymbol{z}_i \boldsymbol{z}_i^T$ along with the cyclic property of the trace operator, and the matrix inversion lemma. Therefore, using the same reasoning as before, by setting

$$\eta_n = \frac{\beta R}{1 + \beta \frac{1}{n} \operatorname{tr} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}}}, \tag{B.112}$$

the second term in (B.83) is given by

$$\frac{1}{n} \operatorname{tr}\left(\boldsymbol{D}_s^{-1} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left[\boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{r}} - \boldsymbol{D}_r^{-1}\right] \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) = \frac{\tilde{C} C \beta}{n} \sum_{j=1}^{k}\left[\frac{\tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \tilde{\boldsymbol{z}}_j}{1 + \beta \tilde{\boldsymbol{z}}_j^T \boldsymbol{\mathcal{H}}_{\boldsymbol{i},\boldsymbol{j}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j} - \frac{\frac{1}{n} \operatorname{tr}\left(\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \mathcal{H}_i^r \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right)}{1 + \beta \frac{1}{n} \operatorname{tr} \mathcal{H}^r}\right]. \tag{B.113}$$

This term converges a.s. to zero as n, k grow large exactly due to the same reasons as before. Whence, with the choice of $\boldsymbol{D}_s$ and $\boldsymbol{D}_r$ in (B.85) and (B.109), respectively, h_n given in (B.81) reads

$$h_n = \frac{1}{\left(\eta_n + \frac{1}{\sigma^2}\right)\left(\psi_n + \frac{1}{\sigma^2}\right)} \frac{\operatorname{tr}\left(\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right)}{n}, \tag{B.114}$$

and we overall show that

$$\frac{1}{n} \operatorname{tr}\left[\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T \boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{s}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T \boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2} \boldsymbol{I}_{\boldsymbol{r}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right] - h_n \to 0. \tag{B.115}$$

Thus far, we found a random function h_n which approximates the term of interest. This function involves the Stieltjes transforms $\frac{1}{n}\operatorname{tr}\boldsymbol{\mathcal{H}}^{\boldsymbol{r}}$, $\frac{1}{n}\operatorname{tr}\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}$ and $\frac{1}{n}\operatorname{tr}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T$. However, fortunately, Lemma 4 exactly provides the behavior of these transforms. So, based on this observation, we conclude that

$$h_n - \alpha m_{s,r} \to 0 \tag{B.116}$$

where we have used the fact that $\operatorname{tr}\left(\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}^T\right) = \sum_{i=1}^n s_i r_i = n m_{s,r}$, and we have defined

$$\alpha \triangleq \frac{1}{\left(\eta_0 + \frac{1}{\sigma^2}\right)\left(\psi_0 + \frac{1}{\sigma^2}\right)}, \tag{B.117}$$

in which

$$\eta_0 \triangleq \frac{\beta R}{1+\beta\sigma^2 m_r b\left(m_r\right)}, \tag{B.118}$$

and

$$\psi_0 \triangleq \frac{\beta R}{1+\beta\sigma^2 m_r b\left(m_r\right)} - \frac{\beta^2\sigma^2 R b\left(m_s\right) m_{s,r}}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)\left(1+\beta\sigma^2 m_r b\left(m_r\right)\right)}. \tag{B.119}$$

Returning back to (B.77), we now may write

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^k \boldsymbol{z}_i^T\left(y_i^2\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\tilde{\boldsymbol{z}}_i\\
&= \frac{1}{n}\sum_{i=1}^k \frac{y_i^2\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\tilde{\boldsymbol{z}}_i}{\left(1+\beta\boldsymbol{z}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{z}_i\right)\left(1+\beta\tilde{\boldsymbol{z}}_i^T\left(\beta\left[\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}}\right]_i + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\tilde{\boldsymbol{z}}_i\right)}\\
&\asymp \frac{1}{n}\sum_{i=1}^k \frac{y_i^2\alpha m_{s,r}}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)\left(1+\beta\sigma^2 m_r b\left(m_r\right)\right)}\\
&= \frac{\alpha m_{s,r}}{\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)\left(1+\beta\sigma^2 m_r b\left(m_r\right)\right)}\frac{\left\|\boldsymbol{y}\right\|^2}{n}.
\end{aligned} \tag{B.120}$$

Next, we take care of the second term in the r.h.s. of (B.76), which by using Lemma 7 and Lemma 8 can be rewritten as

$$\begin{aligned}
\frac{1}{n}\sum_{i\neq j}^k y_i y_j \boldsymbol{z}_i^T\boldsymbol{\mathcal{H}}^{\boldsymbol{s}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{\mathcal{H}}^{\boldsymbol{r}}\tilde{\boldsymbol{z}}_j &= \frac{1}{n}\sum_{i\neq j}^k \frac{y_i y_j \boldsymbol{z}_i^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i,j}}^{\boldsymbol{s}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{\mathcal{H}}_{\boldsymbol{i,j}}^{\boldsymbol{r}}\tilde{\boldsymbol{z}}_j}{\left(1+\beta\boldsymbol{z}_i^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{s}}\boldsymbol{z}_i\right)\left(1+\beta\tilde{\boldsymbol{z}}_j^T\boldsymbol{\mathcal{H}}_{\boldsymbol{j}}^{\boldsymbol{r}}\tilde{\boldsymbol{z}}_j\right)}\\
&- \frac{1}{n}\sum_{i\neq j}^k \frac{y_i y_j \beta \boldsymbol{z}_i^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i,j}}^{\boldsymbol{s}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{\mathcal{H}}_{\boldsymbol{i,j}}^{\boldsymbol{r}}\tilde{\boldsymbol{z}}_i\left(\tilde{\boldsymbol{z}}_i^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i,j}}^{\boldsymbol{r}}\tilde{\boldsymbol{z}}_j\right)}{\left(1+\beta\boldsymbol{z}_i^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{s}}\boldsymbol{z}_i\right)\left(1+\beta\tilde{\boldsymbol{z}}_j^T\boldsymbol{\mathcal{H}}_{\boldsymbol{j}}^{\boldsymbol{r}}\tilde{\boldsymbol{z}}_j\right)\left(1+\beta\tilde{\boldsymbol{z}}_i^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i,j}}^{\boldsymbol{r}}\tilde{\boldsymbol{z}}_i\right)}\\
&- \frac{1}{n}\sum_{i\neq j}^k \frac{y_i y_j \beta\left(\boldsymbol{z}_i^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i,j}}^{\boldsymbol{s}}\boldsymbol{z}_j\right)\boldsymbol{z}_j^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i,j}}^{\boldsymbol{s}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{\mathcal{H}}_{\boldsymbol{j}}^{\boldsymbol{r}}\tilde{\boldsymbol{z}}_j}{\left(1+\beta\boldsymbol{z}_i^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i}}^{\boldsymbol{s}}\boldsymbol{z}_i\right)\left(1+\beta\tilde{\boldsymbol{z}}_j^T\boldsymbol{\mathcal{H}}_{\boldsymbol{j}}^{\boldsymbol{r}}\tilde{\boldsymbol{z}}_j\right)\left(1+\beta\boldsymbol{z}_j^T\boldsymbol{\mathcal{H}}_{\boldsymbol{i,j}}^{\boldsymbol{s}}\boldsymbol{z}_j\right)}.
\end{aligned} \tag{B.121}$$

However, it can be seen that there is nothing essentially different in the above terms, when compared to the behavior of the first term in the r.h.s. of (B.76), we analyzed earlier. Indeed, based on the previous results and analysis, we can infer that

$$\frac{1}{n}\sum_{i\neq j}^{k} y_i y_j \boldsymbol{z}_i^T \mathcal{H}^s \boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}} \boldsymbol{\mathcal{H}}^{\boldsymbol{r}} \tilde{\boldsymbol{z}}_j \asymp \frac{\alpha\left[\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{s}}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{y} - m_{s,r}\left\|\boldsymbol{y}\right\|^2\right]}{n\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)\left(1+\beta\sigma^2 m_r b\left(m_r\right)\right)}$$
$$-\frac{\alpha m_{s,r}\beta\sigma^2 b\left(m_r\right)\left[\left\|\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{r}}\right\|^2 - m_r\left\|\boldsymbol{y}\right\|^2\right]}{n\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)\left(1+\beta\sigma^2 m_r b\left(m_r\right)\right)^2} - \frac{\alpha m_{s,r}\beta\sigma^2 b\left(m_s\right)\left[\left\|\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{s}}\right\|^2 - m_s\left\|\boldsymbol{y}\right\|^2\right]}{n\left(1+\beta\sigma^2 m_s b\left(m_s\right)\right)^2\left(1+\beta\sigma^2 m_r b\left(m_r\right)\right)}. \quad \text{(B.122)}$$

Therefore, using the last result, (B.120) and (B.76), we conclude that

$$\frac{1}{n}\boldsymbol{y}^T\boldsymbol{H}_{\boldsymbol{s}}\left(\beta\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{H}_{\boldsymbol{s}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{s}}\right)^{-1}\boldsymbol{Q}_{\boldsymbol{s}\cap\boldsymbol{r}}\left(\beta\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{H}_{\boldsymbol{r}} + \frac{1}{\sigma^2}\boldsymbol{I}_{\boldsymbol{r}}\right)^{-1}\boldsymbol{H}_{\boldsymbol{r}}^T\boldsymbol{y} - q_n \asymp 0 \quad \text{(B.123)}$$

where q_n is given by (98). ■

APPENDIX C

Derivation of (25)*:* In this appendix, using the previous asymptotic results, we derive the asymptotic MMSE. Recall that the MMSE is given by (90)

$$\frac{\text{mmse}\left(\boldsymbol{X}\mid\boldsymbol{Y},\boldsymbol{H}\right)}{n} = \mathbb{E}\left\{\mathbb{E}_{\mu_s}\left[J_1\left(\boldsymbol{Y},\boldsymbol{H}_{\boldsymbol{s}}\right)\right] - \mathbb{E}_{\mu_{s\times r}}\left[J_2\left(\boldsymbol{Y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{H}_{\boldsymbol{r}}\right)\right]\right\}. \quad \text{(C.1)}$$

In the following, we asymptotically estimate each of the various terms in the outer expectation. We start with the analysis of the second term, and accordingly define

$$\mathscr{X}\left(\boldsymbol{y},\boldsymbol{H}\right) \triangleq \sum_{\boldsymbol{s}\in\{0,1\}^n}\sum_{\boldsymbol{r}\in\{0,1\}^n}\mathbb{P}\left(\boldsymbol{s}\right)\mathbb{P}\left(\boldsymbol{r}\right)J_2\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{H}_{\boldsymbol{r}}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{r}}\right).$$

Over the typical set T_ϵ, using the definitions of ξ, J_1, J_2 in (65), (81), and (88), respectively, we know that

$$\left|J_2\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{H}_{\boldsymbol{r}}\right) - \beta^2 q_n\right| < \epsilon, \quad \text{(C.2)}$$

and that

$$\left|\frac{1}{n}\ln\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right) - \frac{\beta^2}{2}f_n - \frac{1}{2}m_s\bar{I}\left(m_s\right)\right| < \epsilon. \quad \text{(C.3)}$$

For brevity, we will henceforth use the following notations

$$\frac{\beta^2}{2}f_n = \frac{\beta^3\sigma^4 b^2\left(m_s\right)m_s^2}{2g^2\left(m_s\right)}\frac{\left\|\boldsymbol{y}\right\|^2}{n} + \frac{\beta^2\sigma^2 b\left(m_s\right)}{2g^2\left(m_s\right)}\frac{\left\|\boldsymbol{H}_{\boldsymbol{s}}^T\boldsymbol{y}\right\|^2}{n}$$
$$\triangleq V\left(m_s\right)\frac{\left\|\boldsymbol{y}\right\|^2}{n} + L\left(m_s\right)\frac{\sum_{i=1}^n\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i}{n}, \quad \text{(C.4)}$$

and

$$\beta^2 q_n \triangleq q\left(m_s, m_r, m_{s,r}, \frac{\sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 s_i}{n}, \frac{\sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 r_i}{n}, \frac{\sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 s_i r_i}{n}\right). \tag{C.5}$$

The last notation emphasizes the important fact that q_n depends on the pattern sequences $\boldsymbol{s}$ and $\boldsymbol{r}$ only through the quantities $m_s, m_r, m_{s,r}, \sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 s_i, \sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 r_i, \sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 s_i r_i$. In the following, we omit the notation of this dependency and use $g\left(\boldsymbol{s}, \boldsymbol{r}\right)$ instead. With these definitions, we now use the fact that for the calculation of the MMSE we only care about $(\boldsymbol{y}, \boldsymbol{H})$-typical sequences. Let φ denote a random fluctuation term that results from the approximation we use in (C.3), namely, for large n

$$\frac{1}{n}\ln \xi\left(\boldsymbol{y}, \boldsymbol{H}_{\boldsymbol{s}}\right) \approx \frac{\beta^2}{2} f_n + \frac{1}{2} m_s \bar{I}\left(m_s\right) + \varphi. \tag{C.6}$$

As was shown in Appendix B (see remark at the end of the appendix), this fluctuation term is typically lower and upper bounded by a vanishing term that is uniform in $\boldsymbol{s}$ (and $\boldsymbol{r}$), namely, $|\varphi| \leq \mathcal{O}\left(1/n\right)$[11].

Therefore, over $\mathcal{T}_\epsilon$ and for large n and k, the function $\mathscr{Z}\left(\boldsymbol{y}, \boldsymbol{H}\right)$ is lower and upper bounded as follows

$$\mathscr{Z}_-\left(\boldsymbol{y}, \boldsymbol{H}\right) \leq \mathscr{Z}\left(\boldsymbol{y}, \boldsymbol{H}\right) \leq \mathscr{Z}_+\left(\boldsymbol{y}, \boldsymbol{H}\right) \tag{C.7}$$

where

$$\begin{aligned} \mathscr{Z}_\pm\left(\boldsymbol{y}, \boldsymbol{H}\right) \triangleq \sum_{\boldsymbol{s}\in\{0,1\}^n} \sum_{\boldsymbol{r}\in\{0,1\}^n} q\left(\boldsymbol{s}, \boldsymbol{r}\right) \exp\Bigg\{ n \Bigg(\tilde{t}\left(m_s\right) + \tilde{t}\left(m_r\right) + L\left(m_s\right) \frac{1}{n}\sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 s_i \\ + L\left(m_r\right)\frac{1}{n}\sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 r_i \pm \varphi \Bigg)\Bigg\} \end{aligned} \tag{C.8}$$

where

$$\tilde{t}\left(m\right) \triangleq f\left(m\right) - \frac{m}{2}\bar{I}\left(m\right) + V\left(m\right)\frac{\left\|\boldsymbol{y}\right\|^2}{n}. \tag{C.9}$$

Based on (C.8), we need to handle a double summation (over $\boldsymbol{s}$ and $\boldsymbol{r}$). In the following, we first assess the exponential order of the summation over $\boldsymbol{r}$. First, we rewrite $\mathscr{Z}_\pm\left(\boldsymbol{y}, \boldsymbol{H}\right)$ as follows

$$\begin{aligned} \mathscr{Z}_\pm\left(\boldsymbol{y}, \boldsymbol{H}\right) = \sum_{\boldsymbol{s}\in\{0,1\}^n} \exp\left\{ n \left(\tilde{t}\left(m_s\right) + L\left(m_s\right)\frac{1}{n}\sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 s_i \right)\right\} \sum_{\boldsymbol{r}\in\{0,1\}^n} q\left(\boldsymbol{s}, \boldsymbol{r}\right) \exp\Bigg\{ n \Bigg(\tilde{t}\left(m_r\right) \\ + L\left(m_r\right)\frac{1}{n}\sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 r_i \pm \varphi \Bigg)\Bigg\} \end{aligned} \tag{C.10}$$

[11]Physically, over the typical set, this fluctuation will not affect the asymptotic behavior of any *intensive* quantity, namely, a quantity that does not depend on n (e.g., the dominant magnetization).

$$\triangleq \sum_{\boldsymbol{s}\in\{0,1\}^n} \exp\left\{n\left(\tilde{t}\left(m_s\right)+L\left(m_s\right)\frac{1}{n}\sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i\right)\right\}\tilde{\mathscr{Z}}_{\pm}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s}\right) \tag{C.11}$$

where for any $\boldsymbol{s}\in\{0,1\}^n$, we define

$$\tilde{\mathscr{Z}}_{\pm}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s}\right) \triangleq \sum_{\boldsymbol{r}\in\{0,1\}^n} q\left(\boldsymbol{s},\boldsymbol{r}\right)\exp\left\{n\left(\tilde{t}\left(m_r\right)+L\left(m_r\right)\frac{1}{n}\sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 r_i \pm\varphi\right)\right\}. \tag{C.12}$$

Now, note that $\tilde{\mathscr{Z}}_{\pm}$ can be equivalently rewritten as

$$\tilde{\mathscr{Z}}_{\pm}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s}\right) = \sum_{m_r}\exp\left\{n\left(\tilde{t}\left(m_r\right)\pm\varphi\right)\right\}\hat{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right) \tag{C.13}$$

where the summation is over $m_r\in[0/n,1/n,\ldots,n/n]$, and

$$\hat{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right) \triangleq \sum_{\boldsymbol{r}:\, m(\boldsymbol{r})=m_r} q\left(\boldsymbol{s},\boldsymbol{r}\right)\exp\left(L\left(m_r\right)\sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 r_i\right) \tag{C.14}$$

where with slight abuse of notations, the summation is performed over sequences $\boldsymbol{r}$ with magnetization, $m\left(\boldsymbol{r}\right)\triangleq\frac{1}{n}\sum_{i=1}^{n}r_i$, fixed to m_r. For the sake of brevity, we will omit the $\pm$ sign.

In the following, we will find the asymptotic behavior of $\hat{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right)$, and then the asymptotic behavior of $\tilde{\mathscr{Z}}_{\pm}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s}\right)$. For $\hat{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right)$, we will need to count the number of sequences $\{\boldsymbol{r}\}$, having a given magnetization m_r, and also admit some linear constraint. Accordingly, consider the following set

$$\mathcal{F}_{\delta}\left(\left\{\rho_l\right\}_{l=1}^{L},m\right) \triangleq \left\{\boldsymbol{v}\in\{0,1\}^n:\ \left|\sum_{i=1}^{n}v_i-nm\right|\leq\delta,\ \left|\sum_{i=1}^{n}v_iu_{i,l}-n\rho_l\right|\leq\delta,\ l=1,\ldots,L\right\} \tag{C.15}$$

where $L\in\mathbb{N}$ is fixed, and $\{u_{i,l}\}_{i=1}^{n}$ for $l=1,\ldots,L$ are given sequences of real numbers. Thus, the above set contains binary sequences that admit a set of linear constraints. We will upper and lower bound the cardinality of $\mathcal{F}_{\delta}\left(\left\{\rho_l\right\}_{l=1}^{L},m\right)$ for a given $\delta>0$, m, and $\left\{\rho_l\right\}_{l=1}^{L}$. Then, we will use the result in order to approximate $\hat{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right)$.

Define

$$\mathbb{P}\left(v_i;\left\{\alpha_l\right\}_{l=1}^{L},\gamma\mid\left\{u_{i,l}\right\}_{l=1}^{L}\right) \triangleq \frac{\exp\left\{\sum_{l=1}^{L}\alpha_l v_i u_{i,l}-\gamma v_i\right\}}{2\exp\left\{\frac{1}{2}\left(\sum_{l=1}^{L}\alpha_l u_{i,l}-\gamma\right)\right\}\cosh\left(\frac{\sum_{l=1}^{L}\alpha_l u_{i,l}-\gamma}{2}\right)} \tag{C.16}$$

where $\left\{\alpha_l\right\}_{l=1}^{L}$ and γ are auxiliary parameters. Now, for $\boldsymbol{v}=(v_1,\ldots,v_n)$, let

$$\mathbb{P}\left(\boldsymbol{v};\left\{\alpha_l\right\}_{l=1}^{L},\gamma\mid\left\{\boldsymbol{u}_l\right\}_{l=1}^{L}\right) \triangleq \frac{\exp\left\{\sum_{l=1}^{L}\alpha_l\sum_{i=1}^{n}v_iu_{i,l}-\gamma\sum_{i=1}^{n}v_i\right\}}{2^n\exp\left\{\frac{1}{2}\left(\sum_{l=1}^{L}\alpha_l\sum_{i=1}^{n}u_{i,l}-n\gamma\right)\right\}\prod_{i=1}^{n}\cosh\left(\frac{\sum_{l=1}^{L}\alpha_l u_{i,l}-\gamma}{2}\right)}. \tag{C.17}$$

Then, we have that

$$1 \geq \mathbb{P}\left(\boldsymbol{v} \in \mathcal{F}_{\delta}\left(\rho, m\right); \left\{\alpha_l\right\}_{l=1}^{L}, \gamma \mid \left\{\boldsymbol{u}_l\right\}_{l=1}^{L}\right) \tag{C.18}$$

$$= \sum_{\boldsymbol{v} \in \mathcal{F}_{\delta}} \frac{\exp\left\{\sum_{l=1}^{L} \alpha_l \sum_{i=1}^{n} v_i u_{i,l} - \gamma \sum_{i=1}^{n} v_i\right\}}{2^n \exp\left\{\frac{1}{2}\left(\sum_{l=1}^{L} \alpha_l \sum_{i=1}^{n} u_{i,l} - n\gamma\right)\right\} \prod_{i=1}^{n} \cosh\left(\frac{\sum_{l=1}^{L} \alpha_l u_{i,l} - \gamma}{2}\right)} \tag{C.19}$$

$$\geq \sum_{\boldsymbol{v} \in \mathcal{F}_{\delta}} \frac{\exp\left\{\sum_{l=1}^{L} \alpha_l \left(n\rho_l - \delta\right) - \gamma\left(nm - \delta\right)\right\}}{2^n \exp\left\{\frac{1}{2}\left(\sum_{l=1}^{L} \alpha_l \sum_{i=1}^{n} u_{i,l} - n\gamma\right)\right\} \prod_{i=1}^{n} \cosh\left(\frac{\sum_{l=1}^{L} \alpha_l u_{i,l} - \gamma}{2}\right)} \tag{C.20}$$

$$= \left|\mathcal{F}_{\delta}\left\{\left(\rho_l\right)_{l=1}^{L}, m\right\}\right| \frac{\exp\left\{\sum_{l=1}^{L} \alpha_l \left(n\rho_l - \delta\right) - \gamma\left(nm - \delta\right)\right\}}{2^n \exp\left\{\frac{1}{2}\left(\sum_{l=1}^{L} \alpha_l \sum_{i=1}^{n} u_{i,l} - n\gamma\right)\right\} \prod_{i=1}^{n} \cosh\left(\frac{\sum_{l=1}^{L} \alpha_l u_{i,l} - \gamma}{2}\right)}. \tag{C.21}$$

It is easy to verify that $\left(\left\{\alpha_l^{\circ}\right\}_{l=1}^{L}, \gamma^{\circ}\right)$ given by the solution of the following set of equations

$$\rho_l = \frac{\delta}{n} + \frac{1}{2n} \sum_{i=1}^{n} u_{i,l} + \frac{1}{2n} \sum_{i=1}^{n} \tanh\left(\frac{\sum_{l=1}^{L} \alpha_l^{\circ} u_{i,l} - \gamma^{\circ}}{2}\right) u_{i,l}, \quad l = 1, \ldots, L, \tag{C.22}$$

and

$$m = \frac{\delta}{n} + \frac{1}{2} + \frac{1}{2n} \sum_{i=1}^{n} \tanh\left(\frac{\sum_{l=1}^{L} \alpha_l u_{i,l} - \gamma^{\circ}}{2}\right), \tag{C.23}$$

maximize the right hand side of (C.21) (w.r.t. $(\alpha)_{l=1}^{L}$ and γ). Thus, using the last results, we have the following upper bound

$$\left|\mathcal{F}_{\delta}\left(\left\{\rho_l\right\}_{l=1}^{L}, m\right)\right| \leq \frac{\exp\left\{\frac{1}{2}\left(\sum_{l=1}^{L} \alpha_l^{\circ} \sum_{i=1}^{n} u_{i,l} - n\gamma^{\circ}\right)\right\} \prod_{i=1}^{n} 2\cosh\left(\frac{\sum_{l=1}^{L} \alpha_l^{\circ} u_{i,l} - \gamma^{\circ}}{2}\right)}{\exp\left\{\sum_{l=1}^{L} \alpha_l^{\circ} \left(n\rho_l - \delta\right) - \gamma^{\circ}\left(nm - \delta\right)\right\}}$$

$$= \exp\left\{\frac{1}{2}\left(\sum_{l=1}^{L} \alpha_l^{\circ} \sum_{i=1}^{n} u_{i,l} - n\gamma^{\circ}\right) - \left(\sum_{l=1}^{L} \alpha_l^{\circ}\left(n\rho_l - \delta\right) - \gamma^{\circ}\left(nm - \delta\right)\right) + \sum_{i=1}^{n} \ln\left[2\cosh\left(\frac{\sum_{l=1}^{L} \alpha_l^{\circ} u_{i,l} - \gamma^{\circ}}{2}\right)\right]\right\} \tag{C.24}$$

$$\triangleq R_{\delta}. \tag{C.25}$$

For a lower bound, we first note that

$$1 = \mathbb{P}\left(\boldsymbol{v} \in \mathcal{F}_{\delta}\left(\left\{\rho_l\right\}_{l=1}^{L}, m\right); \left\{\alpha_l\right\}_{l=1}^{L}, \gamma \mid \left\{\boldsymbol{u}_l\right\}_{l=1}^{L}\right) + \mathbb{P}\left(\boldsymbol{v} \in \mathcal{F}_{\delta}^{c}\left(\left\{\rho_l\right\}_{l=1}^{L}, m\right); \left\{\alpha_l\right\}_{l=1}^{L}, \gamma \mid \left\{\boldsymbol{u}_l\right\}_{l=1}^{L}\right) \tag{C.26}$$

$$\leq \left|\mathcal{F}_{\delta}\left(\left\{\rho_l\right\}_{l=1}^{L}, m\right)\right| \frac{1}{R_{-\delta}} + \mathbb{P}\left(\boldsymbol{v} \in \mathcal{F}_{\delta}^{c}\left(\left\{\rho_l\right\}_{l=1}^{L}, m\right); \left\{\alpha_l\right\}_{l=1}^{L}, \gamma \mid \left\{\boldsymbol{u}_l\right\}_{l=1}^{L}\right) \tag{C.27}$$

where the last inequality follows by the same considerations we have used for obtaining (C.21) (but now with δ instead of $-\delta$). Using Boole's inequality,

$$\mathbb{P}\left(\boldsymbol{v} \in \mathcal{F}_\delta^c\left(\{\rho_l\}_{l=1}^L, m\right); \{\alpha_l\}_{l=1}^L, \gamma \mid \{\boldsymbol{u}_l\}_{l=1}^L\right) \leq \mathbb{P}\left(\boldsymbol{v}: \left|\sum_{i=1}^n v_i - nm\right| > \delta; \{\alpha_l\}_{l=1}^L, \gamma \mid \{\boldsymbol{u}_l\}_{l=1}^L\right)$$
$$+ \mathbb{P}\left(\boldsymbol{v}: \left|\sum_{i=1}^n v_i u_{i,l} - n\rho_l\right| > \delta,\ l = 1, \ldots, L; \{\alpha_l\}_{l=1}^L, \gamma \mid \{\boldsymbol{u}_l\}_{l=1}^L\right). \tag{C.28}$$

It is easy to verify that the parameters $\{\alpha_l\}_{l=1}^L$ and γ that are solving the following the following equations

$$\mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n v_i u_{i,l} \,\middle|\, \{\boldsymbol{u}_l\}_{l=1}^L\right\} = \rho_l,\ l = 1, \ldots, L, \tag{C.29}$$

and

$$\mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n v_i \,\middle|\, \{\boldsymbol{u}_l\}_{l=1}^L\right\} = m \tag{C.30}$$

where the expectation is taken w.r.t. the conditional distribution (C.17), are also maximizing the conditional distribution (maximum-likelihood)[12]. Therefore, using the SLLN, the two terms on the right hand side of (C.28) are negligible as $n \to \infty$, namely,

$$\mathbb{P}\left(\boldsymbol{v} \in \mathcal{F}_\delta^c\left(\{\rho_l\}_{l=1}^L, m\right); \alpha, \gamma \mid \{\boldsymbol{u}_l\}_{l=1}^L\right) \leq \tau \tag{C.31}$$

for any $\tau > 0$. Thus,

$$\left|\mathcal{F}_\delta\left(\{\rho_l\}_{l=1}^L, m\right)\right| \geq (1-\tau)\, R_{-\delta}. \tag{C.32}$$

Whence, (C.25) and (C.32) provide tight (as $\delta \to 0$) upper and lower bounds on cardinality of $\mathcal{F}_\delta\left(\{\rho_l\}_{l=1}^L, m\right)$.

Returning back to our problem, we will use the above result in order to find an asymptotic estimate of $\hat{\mathscr{Z}}\left(\boldsymbol{y}, \boldsymbol{H}, \boldsymbol{s}, m_r\right)$:

$$\hat{\mathscr{Z}}\left(\boldsymbol{y}, \boldsymbol{H}, \boldsymbol{s}, m_r\right) \triangleq \sum_{\boldsymbol{r}:\, m(\boldsymbol{r}) = m_r} q\left(\boldsymbol{s}, \boldsymbol{r}\right) \exp\left(L\left(m_r\right)\sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 r_i\right), \tag{C.33}$$

and recall that $q\left(\boldsymbol{s}, \boldsymbol{r}\right)$ depends on $\boldsymbol{s}, \boldsymbol{r}$ as follows

$$q\left(\boldsymbol{s}, \boldsymbol{r}\right) = q\left(m_s, m_r, \sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 r_i, m_{s,r}, \sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 s_i r_i, \sum_{i=1}^n \left|\boldsymbol{y}^T \boldsymbol{h}_i\right|^2 s_i\right). \tag{C.34}$$

[12]Essentially, this follows from the fact that (C.17) maintains all the sufficient statistics induced by $\mathcal{F}_\delta\left(\{\rho_l\}_{l=1}^L, m\right)$.

In accordance to the previous notations used in the calculation of $\left|\mathcal{F}_{\delta}\left(\left\{\rho_l\right\}_{l=1}^{L}, m\right)\right|$, let us define $u_{i,1} \triangleq \left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2$, $u_{i,2} \triangleq s_i$, and $u_{i,3} \triangleq \left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i$, namely, the coefficients of the terms which depend on $\boldsymbol{r}$ (recall (C.34)). Now, the main observation here is that $\hat{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right)$ can be represented as

$$\hat{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right) = 2^n \int_{\mathcal{D}\subset\mathbb{R}^3} g\left(m_s, m_r, \rho_1, \rho_2, \rho_3, \sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i\right)\exp\left(nL\left(m_r\right)\rho_1\right)\mathscr{C}_n\left(\mathrm{d}\rho_1,\mathrm{d}\rho_2,\mathrm{d}\rho_3\right) \tag{C.35}$$

where $\mathcal{D}$ is the codomain[13] of (ρ_1,ρ_2,ρ_3), and $\{\mathscr{C}_n\}$ is a sequence of probability measures that are proportional to the number of sequences $\boldsymbol{r}$ with $\sum_{i=1}^n r_i u_{i,j} \approx n\rho_j$ for $j=1,2,3$, and $\sum_{i=1}^n r_i \approx nm_r$. These probability measures satisfy the large deviations principle (LDP) [39, 40], with the following respective lower semi-continuous rate function

$$I\left(\rho_1,\rho_2,\rho_3\right) = \begin{cases} \ln 2 - \frac{1}{n}\ln R_0, & \text{if } \left\{\rho_l\right\}_{l=1}^3 \in \mathcal{D} \\ \infty, & \text{else} \end{cases} \tag{C.36}$$

where $R_0 \triangleq \lim_{\delta\to 0} R_\delta$ given in (C.25). Indeed, by definition, the probability measure $\mathscr{C}_n$ is the ratio between $\left|\mathcal{F}_{\delta}\left(\left\{\rho_l\right\}_{l=1}^{3}, m_r\right)\right|$ and 2^n (the number of possible sequences). Thus, for any Borel set $\mathcal{B}\subset\mathcal{D}$, we have that $\lim_{n\to\infty}\frac{1}{n}\ln\mathscr{C}_n\left(\mathcal{B}\right) = -I\left(\rho_1,\rho_2,\rho_3\right)$. Accordingly, due to it large deviations properties, applying Varadhan's theorem [39, 40] on (C.35), one obtains

$$\hat{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right) \to q\left(m_s, m_r, \rho_1^\circ, \rho_2^\circ, \rho_3^\circ, \sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i\right)\exp\left\{n\left(\ln_2 + L\left(m_r\right)\rho_1^\circ - I\left(\left\{\rho_l^\circ\right\}_{l=1}^3\right)\right)\right\} \tag{C.37}$$

where $\left\{\rho_l^\circ\right\}_{l=1}^3$ are given by (using the fact that the exponential term is convex)

$$\begin{aligned}\left(\rho_1^\circ,\rho_2^\circ,\rho_3^\circ\right) &= \arg\max_{\rho_1,\rho_2,\rho_3\in\mathbb{R}}\left\{\ln_2 + L\left(m_r\right)\rho_1 - I\left(\left\{\rho_l\right\}_{l=1}^3\right)\right\} \\ &= \arg\max_{\rho_1,\rho_2,\rho_3\in\mathbb{R}}\left\{L\left(m_r\right)\rho_1 + \frac{1}{n}\ln R_0\right\}.\end{aligned} \tag{C.38}$$

Whence, the maximizers are the solutions of the following equations: ρ_1° is the solution of

$$L\left(m_r\right) + \frac{1}{n}\frac{\partial}{\partial\rho_1}\ln R_0 = 0, \tag{C.39}$$

and ρ_j° for $j=2,3$, are the solutions of

$$\frac{\partial}{\partial\rho_j}\ln R_0 = 0. \tag{C.40}$$

[13]Note that we do not need to explicitly define $\mathcal{D}$ simply due to the fact that the exponential term in (C.35) is concave (see (C.38)), and thus the dominating ρ_1,ρ_2,ρ_3 are the same over $\mathcal{D}$ or over $\mathbb{R}^3$.

We have that (for $i=1,2,3$)

$$\frac{1}{n}\frac{\partial}{\partial\rho_i}\ln R_0=\frac{1}{2n}\sum_{l=1}^{3}\frac{\partial\alpha_l^\circ}{\partial\rho_i}\sum_{i=1}^{n}u_{i,l}-\frac{1}{2}\frac{\partial\gamma^\circ}{\partial\rho_i}-\sum_{l=1}^{3}\rho_l\frac{\partial\alpha_l^\circ}{\partial\rho_i}-\alpha_i^\circ+m\frac{\partial\gamma^\circ}{\partial\rho_i}$$
$$+\frac{1}{2n}\sum_{i=1}^{n}\tanh\left(\frac{\sum_{l=1}^{3}\alpha_l^\circ u_{i,l}-\gamma^\circ}{2}\right)\left[\sum_{l=1}^{3}u_{i,l}\frac{\partial\alpha_l^\circ}{\partial\rho_i}-\frac{\partial\gamma^\circ}{\partial\rho_i}\right] \tag{C.41}$$
$$=-\alpha_i^\circ+\sum_{l=1}^{3}\frac{\partial\alpha_l^\circ}{\partial\rho_i}\left[\frac{1}{2n}\sum_{i=1}^{n}u_{i,l}+\frac{1}{2n}\sum_{i=1}^{n}\tanh\left(\frac{\sum_{l=1}^{3}\alpha_l^\circ u_{i,l}-\gamma^\circ}{2}\right)u_{i,l}-\rho_l\right]$$
$$+\frac{\partial\gamma^\circ}{\partial\rho_i}\left[m-\frac{1}{2}-\frac{1}{2n}\sum_{i=1}^{n}\tanh\left(\frac{\sum_{l=1}^{3}\alpha_l^\circ u_{i,l}-\gamma^\circ}{2}\right)\right], \tag{C.42}$$

and by using the saddle point equations (C.22) and (C.23), the last two terms in the above equations vanish, and we remain with

$$\frac{1}{n}\frac{\partial}{\partial\rho_i}\ln R_0=-\alpha_i^\circ. \tag{C.43}$$

Thus, combined with (C.39) and (C.40), we conclude that $\alpha_1^\circ=L\left(m_r\right)$, and that $\alpha_2^\circ=\alpha_3^\circ=0$. Accordingly, the exponential term boils down to

$$L\left(m_r\right)\rho_1^\circ+\frac{1}{n}\ln R_0\Big|_{\rho^\circ}=L\left(m_r\right)\rho_1^\circ+\frac{1}{2n}\left(L\left(m_r\right)\sum_{i=1}^{n}u_{i,1}-n\gamma^\circ\right)-L\left(m_r\right)\rho_1^\circ+m_r\gamma^\circ$$
$$+\frac{1}{n}\sum_{i=1}^{n}\ln\left[2\cosh\left(\frac{L\left(m_r\right)u_{i,1}-\gamma^\circ}{2}\right)\right]$$
$$=m_r\gamma^\circ+\frac{1}{n}\sum_{i=1}^{n}\frac{L\left(m_r\right)u_{i,1}-\gamma^\circ}{2}+\frac{1}{n}\sum_{i=1}^{n}\ln\left[2\cosh\left(\frac{L\left(m_r\right)u_{i,1}-\gamma^\circ}{2}\right)\right]$$
$$\triangleq h\left(\delta^\circ,m_r\right). \tag{C.44}$$

Hence, we obtained that (with the substitution of $u_{i,1}=\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2$)

$$\hat{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right)\to q\left(m_s,m_r,\rho_1^\circ,\rho_2^\circ,\rho_3^\circ,\sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i\right)\exp\left(nh\left(\gamma^\circ,m_r\right)\right) \tag{C.45}$$

where $\gamma^\circ,\{\rho_l^\circ\}_{l=1}^3$ solve the following set of equations (based on (C.22) and (C.23))

$$m_r=\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right], \tag{C.46a}$$

$$\rho_1^\circ=\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2, \tag{C.46b}$$

$$\rho_2^\circ=\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]s_i, \tag{C.46c}$$

$$\rho_3^\circ = \frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L(m_r)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i. \tag{C.46d}$$

Thus far, we approximated $\hat{\mathscr{Z}}(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r)$. Recalling (C.13), the next step in our analysis is to approximate $\tilde{\mathscr{Z}}_\pm(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s})$. Using the last approximation, and applying once again Varadhan's theorem (or simply, the Laplace method [41, 42]) on (C.13), one obtains that

$$\begin{aligned}
\tilde{\mathscr{Z}}_\pm(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s}) &= \sum_{m_r}\exp\left(n\left(\tilde{t}(m_r)\pm\varphi\right)\right)\hat{\mathscr{Z}}(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r) \\
&\asymp q\left(m_s,m_r^\circ,\rho_1^\circ(m_r^\circ),\rho_2^\circ(m_r^\circ,\boldsymbol{s}),\rho_3^\circ(m_r^\circ,\boldsymbol{s}),\sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i\right)\exp\left\{n\left(h(\gamma^\circ,m_r^\circ)+\tilde{t}(m_r^\circ)\right)\right\}
\end{aligned} \tag{C.47}$$

where the dominating m_r° is the saddle point, i.e., one of the solutions to the equation

$$\frac{\partial}{\partial m}f(m)-\frac{1}{2}\bar{I}(m)-\frac{m}{2}\frac{\partial}{\partial m}\bar{I}(m)+\frac{1}{n}\frac{\partial}{\partial m}V(m)\frac{\|\boldsymbol{y}\|^2}{n}+\frac{\partial}{\partial m}h(\gamma^\circ,m)=0 \tag{C.48}$$

where we have used the fact that $\tilde{t}(m)=f(m)-\frac{m}{2}\bar{I}(m)+V(m)\|\boldsymbol{y}\|^2/n$. Simple calculations reveal that the derivative of $h(\gamma^\circ,m)$ w.r.t. m is given by (note that γ° also depends on m_r)

$$\begin{aligned}
\frac{\partial}{\partial m}h(\gamma^\circ,m) &= \gamma^\circ+m\frac{\partial}{\partial m}\gamma^\circ+\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\left[\frac{\partial}{\partial m}L(m)u_{i,1}-\frac{\partial}{\partial m}\gamma^\circ\right] \\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\tanh\left(\frac{L(m)u_{i,1}-\gamma^\circ}{2}\right)\frac{1}{2}\left[\frac{\partial}{\partial m}L(m)u_{i,1}-\frac{\partial}{\partial m}\gamma^\circ\right] \qquad \text{(C.49)} \\
&= \gamma^\circ+\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L(m)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\frac{\partial L(m)}{\partial m}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 \\
&\quad+\frac{\partial}{\partial m}\gamma^\circ\left[m-\frac{1}{2}-\frac{1}{2n}\sum_{i=1}^{n}\tanh\left(\frac{L(m)u_{i,1}-\gamma^\circ}{2}\right)\right], \qquad \text{(C.50)}
\end{aligned}$$

but the last term in r.h.s. of the above equation is zero (due to (C.23)), and thus

$$\frac{\partial}{\partial m}h(\gamma^\circ,m)=\gamma^\circ+\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L(m)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\frac{\partial L(m)}{\partial m}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2. \tag{C.51}$$

Thus, substituting the last result in (C.48), we have that

$$\begin{aligned}
\gamma^\circ(m_r^\circ) &= -\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L(m_r^\circ)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\frac{\partial L(m_r^\circ)}{\partial m_r^\circ}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\frac{\partial}{\partial m_r^\circ}f(m_r^\circ)+\frac{1}{2}\bar{I}(m_r^\circ) \\
&\quad+\frac{m_r^\circ}{2}\frac{\partial}{\partial m_r^\circ}\bar{I}(m_r^\circ)-\frac{\partial}{\partial m_r^\circ}V(m_r^\circ)\frac{\|\boldsymbol{y}\|^2}{n}.
\end{aligned} \tag{C.52}$$

So, hitherto, we obtained that the asymptotic behavior of $\tilde{Z}_\pm(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s})$ is given by (C.47), and the various dominating terms are given by

$$\gamma^\circ(m_r^\circ) = -\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L(m_r^\circ)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\frac{\partial L(m_r^\circ)}{\partial m_r^\circ}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\frac{\partial}{\partial m_r^\circ}f(m_r^\circ)+\frac{1}{2}\bar{I}(m_r^\circ)$$

$$+\frac{m_r^\circ}{2}\frac{\partial}{\partial m_r^\circ}\bar{I}\left(m_r^\circ\right)-\frac{\partial}{\partial m_r^\circ}V\left(m_r^\circ\right)\frac{\left\|\boldsymbol{y}\right\|^2}{n}, \tag{C.53a}$$

$$m_r^\circ=\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right], \tag{C.53b}$$

$$\rho_1^\circ=\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2, \tag{C.53c}$$

$$\rho_2^\circ=\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]s_i, \tag{C.53d}$$

$$\rho_3^\circ=\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i. \tag{C.53e}$$

This concludes the asymptotic analysis of the summation over $\boldsymbol{r}$ in (C.10). We now take care of the summation over $\boldsymbol{s}$ in (C.11). Let

$$q\left(\boldsymbol{s}\right)\stackrel{\triangle}{=}q\left(m_s,m_r^\circ,\rho_1^\circ\left(m_r^\circ\right),\rho_2^\circ\left(m_r^\circ,\boldsymbol{s}\right),\rho_3^\circ\left(m_r^\circ,\boldsymbol{s}\right),\sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i\right). \tag{C.54}$$

Applying (C.47) on (C.11), we have that

$$\begin{aligned}\mathscr{Z}_{\pm}\left(\boldsymbol{y},\boldsymbol{H}\right)&\asymp e^{\left\{n\left(h\left(\gamma^\circ,m_r^\circ\right)+\tilde{t}\left(m_r^\circ\right)\right)\right\}}\sum_{\boldsymbol{s}\in\{0,1\}^n}q\left(\boldsymbol{s}\right)\exp\left\{n\left(\tilde{t}\left(m_s\right)+L\left(m_s\right)\frac{1}{n}\sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i\pm\varphi\right)\right\}\\&\stackrel{\triangle}{=}e^{\left\{n\left(h\left(\gamma^\circ,m_r^\circ\right)+\tilde{t}\left(m_r^\circ\right)\right)\right\}}\sum_{m_s}\exp\left(n\left(\tilde{t}\left(m_s\right)\pm\varphi\right)\right)\bar{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},m_s\right)\end{aligned} \tag{C.55}$$

where as before

$$\bar{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},m_s\right)\stackrel{\triangle}{=}\sum_{\boldsymbol{s}:\,m(\boldsymbol{s})=m_s}q\left(\boldsymbol{s}\right)\exp\left(L\left(m_s\right)\sum_{i=1}^{n}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 s_i\right). \tag{C.56}$$

However, $\bar{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},m_s\right)$ has essentially the same form of $\tilde{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},\boldsymbol{s},m_r\right)$, which we have analyzed earlier. So, using the same technique, we readily obtain that

$$\bar{\mathscr{Z}}\left(\boldsymbol{y},\boldsymbol{H},m_s\right)\asymp\bar{q}\left(m_s\right)\exp\left(nh\left(\tilde{\gamma}^\circ,m_s\right)\right) \tag{C.57}$$

where $h\left(\tilde{\gamma}^\circ,m_s\right)$ is defined as in (C.44) (note that the exponential term is similar to the previous one), and

$$\bar{q}\left(m_s\right)\stackrel{\triangle}{=}q\left(m_s,m_r^\circ,\rho_1^\circ\left(m_r^\circ\right),\rho_2^\circ\left(m_r^\circ,m_s\right),\rho_3^\circ\left(m_r^\circ,m_s\right),\rho_4^\circ\left(m_s\right)\right) \tag{C.58}$$

in which $\tilde{\gamma}^\circ,\{\rho_l^\circ\}_{l=2}^4$ solve the following set of equations

$$m_s=\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_s\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\tilde{\gamma}^\circ}{2}\right)\right], \tag{C.59a}$$

$$\rho_2^{\circ} = \frac{1}{4n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^{\circ}}{2}\right)\right]\left[1+\tanh\left(\frac{L\left(m_s\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\tilde{\gamma}^{\circ}}{2}\right)\right], \tag{C.59b}$$

$$\rho_3^{\circ} = \frac{1}{4n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^{\circ}}{2}\right)\right]\left[1+\tanh\left(\frac{L\left(m_s\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\tilde{\gamma}^{\circ}}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2 \tag{C.59c}$$

$$\rho_4^{\circ} = \frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_s\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\tilde{\gamma}^{\circ}}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2. \tag{C.59d}$$

Finally, the summation over m_s in (C.55) is again estimated by using the Laplace method, and we obtain

$$\begin{aligned}\mathscr{Z}_{\pm}\left(\boldsymbol{y},\boldsymbol{H}\right) \asymp\; & q\left(m_s^{\circ},m_r^{\circ},\rho_1^{\circ}\left(m_r^{\circ}\right),\rho_2^{\circ}\left(m_r^{\circ},m_s^{\circ}\right),\rho_3^{\circ}\left(m_r^{\circ},m_s^{\circ}\right),\rho_4^{\circ}\left(m_s^{\circ}\right)\right)\\ & \times\exp\left\{n\left(h\left(\gamma^{\circ},m_r^{\circ}\right)+h\left(\tilde{\gamma}^{\circ},m_s^{\circ}\right)+\tilde{t}\left(m_r^{\circ}\right)+\tilde{t}\left(m_s^{\circ}\right)\pm\varphi\right)\right\}\end{aligned} \tag{C.60}$$

where

$$\begin{aligned}\gamma^{\circ}\left(m_r^{\circ}\right) = & -\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^{\circ}}{2}\right)\right]\frac{\partial L\left(m_r^{\circ}\right)}{\partial m_r^{\circ}}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\frac{\partial}{\partial m_r^{\circ}}f\left(m_r^{\circ}\right)+\frac{1}{2}\bar{I}\left(m_r^{\circ}\right)\\ & +\frac{m_r^{\circ}}{2}\frac{\partial}{\partial m_r^{\circ}}\bar{I}\left(m_r^{\circ}\right)-\frac{\partial}{\partial m_r^{\circ}}V\left(m_r^{\circ}\right)\frac{\|\boldsymbol{y}\|^2}{n},\\ \tilde{\gamma}^{\circ}\left(m_s^{\circ}\right) = & -\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_s^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\tilde{\gamma}^{\circ}}{2}\right)\right]\frac{\partial L\left(m_s^{\circ}\right)}{\partial m_s^{\circ}}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\frac{\partial}{\partial m_s^{\circ}}f\left(m_s^{\circ}\right)+\frac{1}{2}\bar{I}\left(m_s^{\circ}\right)\\ & +\frac{m_s^{\circ}}{2}\frac{\partial}{\partial m_s^{\circ}}\bar{I}\left(m_s^{\circ}\right)-\frac{\partial}{\partial m_s^{\circ}}V\left(m_s^{\circ}\right)\frac{\|\boldsymbol{y}\|^2}{n},\\ m_r^{\circ} = & \frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^{\circ}}{2}\right)\right],\\ m_s^{\circ} = & \frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_s^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\tilde{\gamma}^{\circ}}{2}\right)\right],\end{aligned}$$

$$\rho_1^{\circ} = \frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^{\circ}}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2, \tag{C.61}$$

$$\begin{aligned}\rho_2^{\circ} = & \frac{1}{4n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^{\circ}}{2}\right)\right]\left[1+\tanh\left(\frac{L\left(m_s^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\tilde{\gamma}^{\circ}}{2}\right)\right],\\ \rho_3^{\circ} = & \frac{1}{4n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_r^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^{\circ}}{2}\right)\right]\left[1+\tanh\left(\frac{L\left(m_s^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\tilde{\gamma}^{\circ}}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2\\ \rho_4^{\circ} = & \frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m_s^{\circ}\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\tilde{\gamma}^{\circ}}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2.\end{aligned} \tag{C.62}$$

Not surprisingly, due to the symmetry between $\boldsymbol{s}$ and $\boldsymbol{r}$, it can be seen that the $m_s^\circ = m_r^\circ$, and whence the above set of equations reduces to

$$\gamma^\circ = -\frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\frac{\partial L\left(m^\circ\right)}{\partial m^\circ}\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\frac{\partial}{\partial m^\circ}f\left(m^\circ\right)+\frac{1}{2}\bar{I}\left(m^\circ\right)$$
$$+\frac{m^\circ}{2}\frac{\partial}{\partial m^\circ}\bar{I}\left(m^\circ\right)-\frac{\partial}{\partial m^\circ}V\left(m^\circ\right)\frac{\left\|\boldsymbol{y}\right\|^2}{n}, \tag{C.63a}$$

$$m^\circ = \frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right], \tag{C.63b}$$

$$\rho_1^\circ = \rho_4^\circ = \frac{1}{2n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2, \tag{C.63c}$$

$$\rho_2^\circ = \frac{1}{4n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]^2, \tag{C.63d}$$

$$\rho_3^\circ = \frac{1}{4n}\sum_{i=1}^{n}\left[1+\tanh\left(\frac{L\left(m^\circ\right)\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2-\gamma^\circ}{2}\right)\right]^2\left|\boldsymbol{y}^T\boldsymbol{h}_i\right|^2, \tag{C.63e}$$

and by using (98)

$$q\left(m^\circ,\{\rho_l^\circ\}_{l=1}^3\right) = \beta^2\frac{\alpha\left(m^\circ,\rho_2^\circ\right)}{g^2\left(m^\circ\right)}\rho_3^\circ - 2\frac{\alpha\left(m^\circ,\rho_2^\circ\right)b\left(m^\circ\right)}{g^3\left(m^\circ\right)}\beta^3\sigma^2\rho_2^\circ\left[\rho_1^\circ - m^\circ\frac{\left\|\boldsymbol{y}\right\|^2}{n}\right]. \tag{C.64}$$

Based on (C.1), we also need to find the asymptotic behavior of

$$\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}\left(\boldsymbol{s}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right), \tag{C.65}$$

and

$$\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}\left(\boldsymbol{s}\right)J_1\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right). \tag{C.66}$$

However, obviously, the previous analyzed term can be regarded as an extended version of the above terms, and so we can immediately conclude that

$$\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}\left(\boldsymbol{s}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right) \asymp \exp\left\{n\left(h\left(\gamma^\circ,m^\circ\right)+\tilde{t}\left(m^\circ\right)\right)\right\}, \tag{C.67}$$

$$\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}\left(\boldsymbol{s}\right)J_1\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right) \asymp w\left(m^\circ\right)\exp\left\{n\left(h\left(\gamma^\circ,m^\circ\right)+\tilde{t}\left(m^\circ\right)\right)\right\} \tag{C.68}$$

where by using (81), (93), and (98) (noting that in this case $\rho_3^\circ = \rho_1^\circ$ and $\rho_2^\circ = m^\circ$)

$$w\left(m^\circ\right) \triangleq \sigma^2 m^\circ b\left(m^\circ\right) + \beta^2\frac{\alpha\left(m^\circ\right)}{g^2\left(m^\circ\right)}\rho_1^\circ - 2\frac{\alpha\left(m^\circ\right)b\left(m^\circ\right)}{g^3\left(m^\circ\right)}\beta^3\sigma^2 m^\circ\left[\rho_1^\circ - m^\circ\frac{\left\|\boldsymbol{y}\right\|^2}{n}\right]. \tag{C.69}$$

Therefore, using the last asymptotic results, the asymptotic estimate of the inner term of the expectation in (C.1) is given by

$$\begin{aligned} &\frac{1}{\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}\left(\boldsymbol{s}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right)}\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}\left(\boldsymbol{s}\right)J_1\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right)\\ &-\frac{1}{\left(\sum_{\boldsymbol{s}\in\{0,1\}^n}\mathbb{P}\left(\boldsymbol{s}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right)\right)^2}\sum_{\boldsymbol{s}\in\{0,1\}^n}\sum_{\boldsymbol{r}\in\{0,1\}^n}\mathbb{P}\left(\boldsymbol{s}\right)\mathbb{P}\left(\boldsymbol{r}\right)J_2\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}},\boldsymbol{H}_{\boldsymbol{r}}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{s}}\right)\xi\left(\boldsymbol{y},\boldsymbol{H}_{\boldsymbol{r}}\right)\\ &\asymp w\left(m^\circ\right)-g\left(m^\circ,\{\rho_l^\circ\}_{l=1}^3\right) \qquad\qquad (C.70)\\ &=\sigma^2m^\circ b\left(m^\circ\right)+\frac{2b\left(m^\circ\right)}{g^3\left(m^\circ\right)}\beta^3\sigma^2\left[\frac{\left\|\boldsymbol{y}\right\|^2}{n}m^\circ-\rho_1^\circ\right]\left[m^\circ\alpha\left(m^\circ\right)-\rho_2^\circ\alpha\left(m^\circ,\rho_2^\circ\right)\right]\\ &\quad+\frac{\beta^2}{g^2\left(m^\circ\right)}\left[\alpha\left(m^\circ\right)\rho_1^\circ-\alpha\left(m^\circ,\rho_2^\circ\right)\rho_3^\circ\right]. \qquad\qquad (C.71)\end{aligned}$$

Accordingly, using the dominated convergence theorem (DCT) [38], and the asymptotic behavior of $\left\|\boldsymbol{y}\right\|^2/n$, we obtain that[14]

$$\begin{aligned}\frac{\text{mmse}\left(\boldsymbol{X}\mid\boldsymbol{Y},\boldsymbol{H}\right)}{n}\asymp\mathbb{E}\Bigg\{&\sigma^2m^\circ b\left(m^\circ\right)+\frac{\beta^2}{g^2\left(m^\circ\right)}\left[\alpha\left(m^\circ\right)\rho_1^\circ-\alpha\left(m^\circ,\rho_2^\circ\right)\rho_3^\circ\right]\\ &+\frac{2b\left(m^\circ\right)}{r^3\left(m^\circ\right)}\beta^3\sigma^2\left[\left(m_a\sigma^2R+\frac{R}{\beta}\right)m^\circ-\rho_1^\circ\right]\left[m^\circ\alpha\left(m^\circ\right)-\rho_2^\circ\alpha\left(m^\circ,\rho_2^\circ\right)\right]\Bigg\}. \qquad (C.72)\end{aligned}$$

Finally, in the following, we will show a concentration property of the saddle point equations given in (C.63), and obtain "instead" the saddle point equations given in (20)-(24). Accordingly, the expectation in (C.72) becomes "superfluous", as all the involved random variables (m° and $\{\rho_i^\circ\}_{i=1}^3$) converge to a deterministic quantity. According to (C.63), it can be seen that the saddle point equations share the following common term

$$\frac{1}{n}\sum_{i=1}^n\phi\left(\left|\boldsymbol{h}_i^T\boldsymbol{Y}\right|^2\right) \tag{C.73}$$

where $\phi\left(\cdot\right):\mathbb{R}\to\mathbb{R}$ is some integrable function (in the L^1 sense). In the following, we first show that (C.73) admits an SLLN property. To this end, let us define

$$T_n\triangleq\sum_{i=1}^n K_i, \tag{C.74}$$

where $K_i\triangleq\phi\left(\left|\boldsymbol{h}_i^T\boldsymbol{Y}\right|^2\right)$, and let $\mathcal{G}_n=\sigma\left(\boldsymbol{X},\boldsymbol{W}\right)\cap\sigma\left(T_n,T_{n+1},\ldots\right)$ be the σ-field (filtration) generated by T_n, $\{K_i\}_{i>n}$, $\boldsymbol{X}$, and $\boldsymbol{W}$. We will now show that $M_n\triangleq-\frac{T_{-n}}{n}$ is a backwards martingale sequence

[14]Note that for an i.i.d. source we simply have that $P_x=p\sigma^2$ where $p=\mathbb{P}\left\{S_i=1\right\}$ for $1\leq i\leq n$.

w.r.t. $\mathcal{F}_n \triangleq \mathcal{G}_{-n}$, $n \leq -1$. Indeed, for $m \leq -1$, we have that

$$\mathbb{E}\left\{ M_{m+1} \,\middle|\, \mathcal{F}_m \right\} = \mathbb{E}\left\{ \frac{T_{-m-1}}{-m-1} \,\middle|\, \mathcal{G}_{-m} \right\}. \tag{C.75}$$

Setting $n = -m$, we see that

$$\mathbb{E}\left\{ \frac{T_{n-1}}{n-1} \,\middle|\, \mathcal{G}_n \right\} = \mathbb{E}\left\{ \frac{T_n - K_n}{n-1} \,\middle|\, \mathcal{G}_n \right\} \tag{C.76}$$

$$= \frac{T_n}{n-1} - \mathbb{E}\left\{ \frac{K_n}{n-1} \,\middle|\, \mathcal{G}_n \right\} \tag{C.77}$$

where we have used the fact that T_n is measurable w.r.t. $\mathcal{G}_n$. Now, we have that

$$\mathbb{E}\left\{ K_n | \mathcal{G}_n \cap \sigma\left(\boldsymbol{Y}\right) \right\} = \mathbb{E}\left\{ K_n | T_n, \boldsymbol{Y}, \sigma\left(\boldsymbol{X}, \boldsymbol{W}\right) \right\} \tag{C.78}$$

$$= \mathbb{E}\left\{ K_j | T_n, \boldsymbol{Y}, \sigma\left(\boldsymbol{X}, \boldsymbol{W}\right) \right\} \tag{C.79}$$

for any $1 \leq j \leq n$, where in the first equality we have used the facts that $\mathcal{G}_n = \sigma\left(\boldsymbol{X}, \boldsymbol{W}\right) \cap \sigma\left(T_n, T_{n+1}, \ldots\right) = \sigma\left(\boldsymbol{X}, \boldsymbol{W}\right) \cap \sigma\left(T_n, K_{n+1}, K_{n+2}, \ldots\right)$, that $\boldsymbol{Y} = \sum_{i=1}^{n} \boldsymbol{h}_i X_i + \boldsymbol{W}$ and that $\{\boldsymbol{h}_i\}$ are statistically independent, and the second equality follows due to the structure of $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}$, the symmetry of T_n w.r.t. $K_1, \ldots, K_n$, and the fact that $\{\boldsymbol{h}_i\}$ are statistically independent. Clearly,

$$\sum_{i=1}^{n} \mathbb{E}\left\{ K_i | T_n, \boldsymbol{Y}, \sigma\left(\boldsymbol{X}, \boldsymbol{W}\right) \right\} = \mathbb{E}\left\{ \sum_{i=1}^{n} K_i \,\middle|\, T_n, \boldsymbol{Y}, \sigma\left(\boldsymbol{X}, \boldsymbol{W}\right) \right\} \tag{C.80}$$

$$= T_n, \tag{C.81}$$

and thus, due to (C.79), we obtain that $\mathbb{E}\left\{ K_n | \mathcal{G}_n \cap \sigma\left(\boldsymbol{Y}\right) \right\} = T_n/n$ a.s. Whence, using (C.77) and the last result, we obtain

$$\mathbb{E}\left\{ \frac{T_{n-1}}{n-1} \,\middle|\, \mathcal{G}_n \right\} = \frac{T_n}{n-1} - \mathbb{E}\left\{ \frac{K_n}{n-1} \,\middle|\, \mathcal{G}_n \right\} \tag{C.82}$$

$$= \frac{T_n}{n-1} - \mathbb{E}\left\{ \mathbb{E}\left\{ \frac{K_n}{n-1} \,\middle|\, \mathcal{G}_n \cap \sigma\left(\boldsymbol{Y}\right) \right\} \,\middle|\, \mathcal{G}_n \right\} \tag{C.83}$$

$$= \frac{T_n}{n-1} - \frac{T_n}{n\left(n-1\right)} = \frac{T_n}{n}, \qquad \text{a.s.} \tag{C.84}$$

This concludes the proof that M_n is a backwards martingale sequence w.r.t. $\left\{\mathcal{F}_n\right\}_{n \leq -1}$. Now, by the backwards martingale convergence theorem [43, 44], we deduce that T_n/n converges as $n \to \infty$, and in L^1, to a random variable $K \triangleq \lim_{n\to\infty} T_n/n$. Obviously, for all m

$$K = \lim_{n\to\infty} \frac{\tilde{K}_{m+1} + \ldots + \tilde{K}_{m+n}}{n}, \tag{C.85}$$

where (due to the fact that $\{\boldsymbol{h}_i\}_i$ are i.i.d.)

$$\tilde{K}_{m+i} = \phi\left(\left|\boldsymbol{h}_{m+i}^T\left(\sum_{j=m+i}^{n+m+i}\boldsymbol{h}_j X_j + \boldsymbol{W}\right)\right|^2\right), \quad \text{for } i = 1,\ldots,n. \tag{C.86}$$

Thus K is $\sigma(\boldsymbol{X},\boldsymbol{W}) \cap \sigma(\boldsymbol{h}_{m+1},\ldots)$-measurable, for all m, and hence it is also $\sigma(\boldsymbol{X},\boldsymbol{W}) \cap \bigcap_m \sigma(\boldsymbol{h}_{m+1},\ldots)$-measurable (namely, the tail σ-field generated by $\{\boldsymbol{h}_i\}$ intersected with $\sigma(\boldsymbol{X},\boldsymbol{W})$). Thus, by the Kolmogorov's 0-1 law [43], we conclude that there exists a constant $C \in \mathbb{R}$ (w.r.t. $\sigma(\boldsymbol{X},\boldsymbol{W})$) such that $\mathbb{P}\{K = C \mid \sigma(\boldsymbol{X},\boldsymbol{W})\} = 1$. This constant is obviously given by

$$C = \mathbb{E}\{K \mid \sigma(\boldsymbol{X},\boldsymbol{W})\} = \lim_{n\to\infty}\mathbb{E}\left\{\frac{T_n}{n} \,\middle|\, \sigma(\boldsymbol{X},\boldsymbol{W})\right\}. \tag{C.87}$$

Thus, we have shown that

$$\frac{1}{n}\sum_{i=1}^{n}\phi\left(\left|\boldsymbol{h}_i^T\boldsymbol{Y}\right|^2\right) - \frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^{n}\phi\left(\left|\boldsymbol{h}_i^T\boldsymbol{Y}\right|^2\right) \,\middle|\, \boldsymbol{X},\boldsymbol{W}\right\} \to 0, \tag{C.88}$$

a.s. as $n \to \infty$, namely, we show an SLLN property of (C.73). Our next step is to infer the asymptotic behavior of each summand. First, we note that

$$\boldsymbol{h}_i^T\boldsymbol{Y} = \boldsymbol{h}_i^T[\boldsymbol{H}\boldsymbol{X}]_i + X_i\|\boldsymbol{h}_i\|^2 + \boldsymbol{h}_i^T\boldsymbol{W} \tag{C.89}$$

where $[\boldsymbol{H}\boldsymbol{X}]_i \triangleq \boldsymbol{H}\boldsymbol{X} - \boldsymbol{h}_i X_i$. Let $\hat{\boldsymbol{X}}_i$ be a new n-dimensional vector, such that its ith component is zero and the other components are identical to that of $\boldsymbol{X}$. Similarly, let $\hat{\boldsymbol{H}}_i$ denote a new matrix such that its ith column contains zeros, and the other columns are identical to those of $\boldsymbol{H}$. Accordingly, let $\hat{\boldsymbol{z}}_{i,j}$ denote the jth row of $\hat{\boldsymbol{H}}_i$. With this notations, we have that $[\boldsymbol{H}\boldsymbol{X}]_i = \hat{\boldsymbol{H}}_i\hat{\boldsymbol{X}}_i$. Thus,

$$\boldsymbol{h}_i^T\boldsymbol{Y} = \sum_{j=1}^{k} H_{j,i}\left[\hat{\boldsymbol{z}}_{i,j}^T\hat{\boldsymbol{X}}_i + W_j\right] + X_i\|\boldsymbol{h}_i\|^2 \tag{C.90}$$

$$= \frac{1}{\sqrt{n}}\sum_{j=1}^{k}\tilde{H}_{j,i}\left[\hat{\boldsymbol{z}}_{i,j}^T\hat{\boldsymbol{X}}_i + W_j\right] + X_i\|\boldsymbol{h}_i\|^2. \tag{C.91}$$

where $\tilde{H}_{i,j} \triangleq \sqrt{n}H_{i,j}$. Given $\boldsymbol{X}$, by using Lyapunov's central limit theorem [45], we may infer the following weak convergence

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{k}\tilde{H}_{j,i}\left[\hat{\boldsymbol{z}}_{i,j}^T\hat{\boldsymbol{X}}_i + W_j\right] \xrightarrow{d} \mathcal{N}\left(0, Rm_a\sigma^2 + \frac{R}{\beta}\right), \tag{C.92}$$

as $n \to \infty$. Accordingly, let $\mathscr{Y}$ be the limit point in (C.92), namely, $\mathscr{Y} \sim \mathcal{N}\left(0, m_a\sigma^2 R + R/\beta\right)$. Therefore, based on (C.91), (C.92), and Slutsky's lemma [46, Lemma 2.8], we may conclude that (conditioned on $\boldsymbol{X}$)

$$\boldsymbol{h}_i^T\boldsymbol{Y} \xrightarrow{d} \mathscr{Y} + RX_i. \tag{C.93}$$

In the sequel, we use the following two results. The first result is the continuous mapping theorem [46, Th. 2.3].

Lemma 5 (The continuous mapping theorem) Let $\Phi : \mathbb{R} \to \mathbb{R}$ be an almost-everywhere continuous mapping, and let $\{J_i\}$ be a sequence of real-valued random variables that converges weakly to a real-valued random variable J. Then, $\{\Phi(J_i)\}$ converges weakly to the real-valued random variable $\Phi(J)$.

The second result is the following extension of Portmanteau's lemma [46, Theorem 2.2].

Lemma 6 (Portmanteau's lemma (extended version)) Let $f : \mathbb{R}^k \mapsto \mathbb{R}$ be a measurable and continuous at every point in a set $\mathscr{C}$. Let $X_n \xrightarrow{d} X$ where X takes its values in $\mathscr{C}$. Then $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ if and only if the sequence of random variables $\{f(X_n)\}$ is *asymptotically uniformly integrable*, namely, $\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left\{|f(X_n)|\, \mathbb{1}_{|f(X_n)|>M}\right\} = 0$.

Whence, using the last results, and Lemmas 5 and 6, we obtain that[15]

$$\frac{1}{n}\sum_{i=1}^{n}\phi\left(\left|\boldsymbol{h}_i^T\boldsymbol{Y}\right|^2\right) - \frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^{n}\phi\left(|\mathscr{Y}+RX_i|^2\right)\,\middle|\,\boldsymbol{X}\right\} \to 0. \tag{C.94}$$

Now, applying the SLLN on (C.94), we finally may write that

$$\frac{1}{n}\sum_{i=1}^{n}\phi\left(\left|\boldsymbol{h}_i^T\boldsymbol{Y}\right|^2\right) \to \mathbb{E}\left[\phi\left(|\mathscr{Y}+RX|^2\right)\right], \tag{C.95}$$

a.s. as $n\to\infty$, where the expectation is taken w.r.t. the product measure corresponding to $\mathscr{Y}$, and X which is distributed according to a mixture of two measures: Dirac measure at 0 with weight $1-m_a$, and a Gaussian measure with zero mean and variance σ^2 and weight m_a. Equivalently, the last result can be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n}\phi\left(\left|\boldsymbol{h}_i^T\boldsymbol{Y}\right|^2\right) \to \mathbb{E}\left[\phi\left(|\mathscr{X}|^2\right)\right], \tag{C.96}$$

a.s. as $n\to\infty$, where the expectation over $\mathscr{X}$ is now taken w.r.t. a mixture of two measures: Gaussian measure with zero mean and variance $\left(m_a\sigma^2R+R/\beta\right)$ and weight $1-m_a$, and a Gaussian measure with zero mean and variance $\left(m_a\sigma^2R+R/\beta+R^2\sigma^2\right)$ and weight m_a.

Therefore, applying the last general asymptotic result to the saddle point equations given in (C.63), we obtain

$$\gamma^\circ = -\frac{1}{2}\mathbb{E}\left\{\left[1+\tanh\left(\frac{L(m^\circ)|\mathscr{X}|^2-\gamma^\circ}{2}\right)\right]\left.\frac{\mathrm{d}L(m)}{\mathrm{d}m}\right|_{m=m^\circ}|\mathscr{X}|^2\right\} - \left.\frac{\mathrm{d}t(m)}{\mathrm{d}m}\right|_{m=m^\circ},$$

[15]In our case, the sequence of random variables $\phi\left(\left|\boldsymbol{h}_i^T\boldsymbol{Y}\right|^2\right)$ meet the asymptotic uniform integrability assumption of Lemma 6, for the various choices of ϕ according to (20)-(24).

$$
\begin{aligned}
m^\circ &= \frac{1}{2}\mathbb{E}\left\{1+\tanh\left(\frac{L\left(m^\circ\right)|\mathscr{X}|^2-\gamma^\circ}{2}\right)\right\},\\
\rho_1^\circ = \rho_4^\circ &= \frac{1}{2}\mathbb{E}\left\{\left[1+\tanh\left(\frac{L\left(m^\circ\right)|\mathscr{X}|^2-\gamma^\circ}{2}\right)\right]|\mathscr{X}|^2\right\},\\
\rho_2^\circ &= \frac{1}{4}\mathbb{E}\left\{\left[1+\tanh\left(\frac{L\left(m^\circ\right)|\mathscr{X}|^2-\gamma^\circ}{2}\right)\right]^2\right\},\\
\rho_3^\circ &= \frac{1}{4}\mathbb{E}\left\{\left[1+\tanh\left(\frac{L\left(m^\circ\right)|\mathscr{X}|^2-\gamma^\circ}{2}\right)\right]^2|\mathscr{X}|^2\right\},
\end{aligned}
\tag{C.97}
$$

as claimed. ■

APPENDIX D
MATHEMATICAL TOOLS

Lemma 7 ([33]) [Matrix Inversion Lemma] Let $\boldsymbol{U}$ be an $N \times N$ invertible matrix and $\boldsymbol{x} \in \mathbb{C}^N$, $c \in \mathbb{C}$ for which $\boldsymbol{U} + c\boldsymbol{x}\boldsymbol{x}^H$ is invertible. Then

$$
\boldsymbol{x}^H\left(\boldsymbol{U}+c\boldsymbol{x}\boldsymbol{x}^H\right)^{-1} = \frac{\boldsymbol{x}^H\boldsymbol{U}^{-1}}{1+c\boldsymbol{x}^H\boldsymbol{U}^{-1}\boldsymbol{x}}. \tag{D.1}
$$

Lemma 8 (Matrix Inversion Lemma 2) Under the assumptions of Lemma 7,

$$
\left(\boldsymbol{U}+c\boldsymbol{x}\boldsymbol{x}^H\right)^{-1} = \boldsymbol{U}^{-1} - \frac{\boldsymbol{U}^{-1}c\boldsymbol{x}\boldsymbol{x}^H\boldsymbol{U}^{-1}}{1+c\boldsymbol{x}^H\boldsymbol{U}^{-1}\boldsymbol{x}}. \tag{D.2}
$$

Lemma 9 (Resolvent Identity) Let $\boldsymbol{U}$ and $\boldsymbol{V}$ be two invertible complex matrices of size $N \times N$. Then

$$
\boldsymbol{U}^{-1} - \boldsymbol{V}^{-1} = -\boldsymbol{U}^{-1}\left(\boldsymbol{U}-\boldsymbol{V}\right)\boldsymbol{V}^{-1}. \tag{D.3}
$$

The following lemma is a powerful tool which is widely used in RMT with many versions and extensions.

Lemma 10 ([24, 25]) Let $\boldsymbol{A}_N \in \mathbb{C}^{N\times N}$ be a sequence of deterministic matrices, and let $\boldsymbol{x} \in \mathbb{C}^N$ have i.i.d. complex entries with zero mean, variance $1/N$, and bounded lth order moment $\mathbb{E}\left|X_i\right|^l \leq \nu_l$. Then, for any $p \geq 1$

$$
\mathbb{E}\left|\boldsymbol{x}_N^H\boldsymbol{A}_N\boldsymbol{x}_N - \frac{1}{N}\operatorname{tr}\boldsymbol{A}_N\right|^p \leq \frac{C_p}{N^{p/2}}\left(\frac{1}{N}\operatorname{tr}\boldsymbol{A}_N\boldsymbol{A}_N^H\right)^{p/2}\left[\nu_4^{p/2}+\nu_{2p}\right] \tag{D.4}
$$

where C_p is a constant depending only on p.

Lemma 11 ([25, 35]) [Trace Lemma] Let $\left(\boldsymbol{A}_N\right)_{N\geq 1}$, $\boldsymbol{A}_N \in \mathbb{C}^{N\times N}$, be a sequence of random matrices and $\left(\boldsymbol{x}_N\right)_{N\geq 1} = \left[X_{1,N},\ldots,X_{N,N}\right]^T \in \mathbb{C}^N$, a sequence of random vectors of i.i.d. entries, statistically independent of $(\boldsymbol{A}_N)_{N\geq 1}$. Assume that $\mathbb{E}\{X_{i,j}\}=0$, $\mathbb{E}\left\{|X_{i,j}|^2\right\}=1$, $\mathbb{E}\left\{|X_{i,j}|^8\right\}<\infty$, and that $\boldsymbol{A}$ has bounded spectral norm (in the a.s. sense). Then, a.s.,

$$\frac{1}{N}\boldsymbol{x}_N^H\boldsymbol{A}_N\boldsymbol{x}_N-\frac{1}{N}\operatorname{tr}\boldsymbol{A}_N\to 0. \tag{D.5}$$

Lemma 12 ([47]) Let $(a_n)_{n\geq 1},(b_n)_{n\geq 1},(\bar{a}_n)_{n\geq 1},(\bar{b}_n)_{n\geq 1}$ be four infinite sequences of complex random variables. Assume that $a_n\asymp\bar{a}_n$ and $b_n\asymp\bar{b}_n$ in the a.s. sense.

- If $|a_n|$, $|\bar{b}_n|$ and/or $|\bar{a}_n|$, $|b_n|$ are a.s. bounded, then a.s.,

$$a_nb_n\asymp\bar{a}_n\bar{b}_n.$$

- If $|a_n|$, $|\bar{b}_n|^{-1}$ and/or $|\bar{a}_n|$, $|b_n|^{-1}$ are a.s. bounded, then a.s.,

$$a_n/b_n\asymp\bar{a}_n/\bar{b}_n.$$

Lemma 13 ([25, 35]) Let $(\boldsymbol{A}_N)_{N\geq 1}$, $\boldsymbol{A}_N\in\mathbb{C}^{N\times N}$, be a sequence of matrices with uniformly bounded spectral norm, and $(\boldsymbol{B}_N)_{N\geq 1}$, $\boldsymbol{B}_N\in\mathbb{C}^{N\times N}$ be random Hermitian, with eigenvalues $\lambda_1\leq\ldots\leq\lambda_N$ such that, with probability one, there exist $\epsilon>0$ for which $\lambda_1>\epsilon$ for all large N. Then, for $\boldsymbol{v}_N\in\mathbb{C}^N$,

$$\frac{1}{N}\operatorname{tr}\boldsymbol{A}_N\boldsymbol{B}_N^{-1}-\frac{1}{N}\operatorname{tr}\boldsymbol{A}_N\left(\boldsymbol{B}_N+\boldsymbol{v}_N\boldsymbol{v}_N^H\right)^{-1}\to 0 \tag{D.6}$$

a.s. as $N\to\infty$, where $\boldsymbol{B}_N^{-1}$ and $\left(\boldsymbol{B}_N+\boldsymbol{v}\boldsymbol{v}^H\right)^{-1}$ are assumed to exist with probability 1.

Lemma 14 ([48]) [Rank-1 Perturbation Lemma] Let $z\in\mathbb{C}\setminus\mathbb{R}^+$, $\boldsymbol{A}\in\mathbb{C}^{N\times N}$ and $\boldsymbol{B}\in\mathbb{C}^{N\times N}$ where $\boldsymbol{B}$ is Hermitian nonnegative definite, and $\boldsymbol{x}\in\mathbb{C}^N$. Then,

$$\left|\operatorname{tr}\left((\boldsymbol{B}-z\boldsymbol{I}_N)^{-1}-\left(\boldsymbol{B}+\boldsymbol{x}\boldsymbol{x}^H-z\boldsymbol{I}_N\right)^{-1}\right)\boldsymbol{A}\right|\leq\frac{\|\boldsymbol{A}\|}{\operatorname{dist}(z,\mathbb{R}^+)} \tag{D.7}$$

where $\operatorname{dist}(\cdot,\cdot)$ denotes the Euclidean distance.

Lemma 15 Let $\boldsymbol{x}_N\in\mathbb{C}^N$ be a random vector with i.i.d. entries each with zero mean and unit variance, and let $\boldsymbol{A}_N\in\mathbb{C}^{N\times N}$ such that $\sqrt{\operatorname{tr}\boldsymbol{A}_N^H\boldsymbol{A}_N}$ is uniformly bounded for all N. Then, for any finite p,

$$\mathbb{E}\left|\boldsymbol{x}_N^H\boldsymbol{A}_N\boldsymbol{x}_N\right|^p<\infty \tag{D.8}$$

for all N.

Proof: By Jensen's inequality we may write that

$$\mathbb{E}\left|\boldsymbol{x}_N^H\boldsymbol{A}_N\boldsymbol{x}_N\right|^p\leq 2^{p-1}\left(\mathbb{E}\left|\boldsymbol{x}_N^H\boldsymbol{A}_N\boldsymbol{x}_N-\operatorname{tr}\boldsymbol{A}_N\right|^p+|\operatorname{tr}\boldsymbol{A}_N|^p\right)<\infty$$

where the second inequality follows from the facts that: the first term in the r.h.s. is bounded by Lemma 10, and the second term is bounded by assumption. ■

REFERENCES

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

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

[3] S. S. Chen, D. L. Donoho, and M. A. Saundres, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[5] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," in *Proceedings of the National Academy of Sciences*, vol. 106, Nov. 2009, pp. 18 914–18 919.

[6] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3065–3092, May 2012.

[7] D. L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 1–53, 2009.

[8] Y. Wu and S. Verdú, "ŕenyi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.

[9] M. Bayati and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," 2011. [Online]. Available: http://http://arxiv.org/pdf/1112.0708.pdf

[10] A. Tulino, G. Caire, S. Verdú, and S. Shamai (Shitz), "Support recovery with sparsely sampled free random matrices," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, July 2013.

[11] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6241–6263, Oct. 2012.

[12] D. Guo, D. Baron, and S. Shamai (Shitz), "A single-letter characterization of optimal noisy compressed sensing," in *Forty-Seventh Annual Allerton Conference on Communication, Control, and Computing*. Allerton Retreat Center, Monticello, Illinois, Sep. 30-Oct. 2, 2009.

[13] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices," *J. Stat. Mech. - Theory E.*, no. 8, p. P08009, Aug. 2012.

[14] ——, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X 2, 021005*, vol. 2, no. 2, May. 2012.

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

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

[17] M. Akcakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2009.

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

[19] K. R. Rad, "Nearly sharp sufficient conditions on exact sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4672–4679, July 2011.

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

[21] W. Wang, J. M. Wainwright, and K. Ramchandran, "Information theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2967–2979, June 2010.

[22] G. Reeves and M. Gastpar, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," Feb 2010. [Online]. Available: http://arxiv.org/pdf/1002.4458.pdf

[23] N. Merhav, "Optimum estimation via gradients of partition functions and information measures: A statistical-mechanical perspective," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3887–3898, June 2011.

[24] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*. Springer, 2010.

[25] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.

[26] W. Huleihel and N. Merhav, "Analysis of mismatched estimation errors using gradients of partition functions," *submitted to IEEE Trans. Inf. Theory*, June 2013. [Online]. Available: http://arxiv.org/pdf/1306.0094.pdf

[27] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[28] D. L. Donoho and J. Tanner, "Precise undersampling theorems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, 2010.

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

[30] J. Zhu and D. Baron, "Performance regions in compressed sensing from noisy measurements," in *CISS, Baltimore, MD*, Mar. 2013.

[31] J. W. Negele and H. Orland, *Quantum many-particles systems*. Frontier in Physics Lecture Notes, AddisonWesley, 1988.

[32] A. Mézard, M. Montanari, *Information, Physics and Computation*. Oxford, U.K.: Oxford Univ. Press., 2009.

[33] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[34] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the analysis of correlated MIMO multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3493–3514, June 2011.

[35] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

[36] F. Rubio, X. Mestre, and D. P. Palomar, "Performance analysis and optimal selection of large minimum variance portfolios under estimation risk," *IEEE Journal of Selected Topics in Sig. Process.*, vol. 6, no. 4, pp. 337–350, Aug. 2012.

[37] T. S. Shores, *Applied Linear Algebra and Matrix Analysis*. Springer, 2000.

[38] G. B. Folland, *Real Analysis*. John Wiley and Sons, 1984.

[39] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.

[40] F. Den Hollander, *Large Deviations*. American Mathematical Society (Fields Institute Monographs), 2000.

[41] N. G. De Bruijn, *Asymptotic Methods in Analysis*. Dover Publications, Inc. New York, 1981.

[42] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, no. 1-2, pp. 1–212, Dec. 2010.

[43] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*. 2rd ed. Academic Press, 1968.

[44] D. Williams, *Probability with Martingales*. Cambridge University Press, 1991.

[45] M. Fisz, *Probability Theory and Mathematical Statistics*. 3rd ed. John Wiely & Sons, Inc., 1963.

[46] A. W. van der Vaart, *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, 2000.

[47] M. J. M. Peacock, I. B. Collings, and M. L. Honig, “Eigenvalue distributions of sums and products of large random matrices via incremental matrix expansions,” *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2123–2138, May 2008.

[48] Z. D. Bai and J. W. Silverstein, “On the signal-to-interference ratio of CDMA systems in wireless communications,” *Annals of Applied Probability*, vol. 17, no. 1, pp. 81–101, Feb. 2007.