I-MMSE relations in random linear estimation and a sub-extensive interpolation method

Jean Barbier and Nicolas Macris

Laboratoire de Théorie des Communications, Faculté Informatique et Communications,
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Suisse.
{jean.barbier, nicolas.macris}@epfl.ch

Abstract—Consider random linear estimation with Gaussian measurement matrices and noise. One can compute infinitesimal variations of the mutual information under infinitesimal variations of the signal-to-noise ratio or of the measurement rate. We discuss how each variation is related to the minimum mean-square error and deduce that the two variations are directly connected through a very simple identity. The main technical ingredient is a new interpolation method called "sub-extensive interpolation method". We use it to provide a new proof of an I-MMSE relation recently found by Reeves and Pfister [1] when the measurement rate is varied. Our proof makes it clear that this relation is intimately related to another I-MMSE relation also recently proved in [2].

One can directly verify that the identity relating the two types of variation of mutual information is indeed consistent with the one letter replica symmetric formula for the mutual information, first derived by Tanaka [3] for binary signals, and recently proved in more generality in [1, 2, 4, 5] (by independent methods). However our proof is independent of any knowledge of Tanaka's formula.

I. Introduction

Random linear estimation (RLE) is a fundamental research field which has been revived by a number of recent theoretical and practical developments such as compressed sensing [6], error correction via sparse superposition codes [7], Boolean group testing [8] or code division multiple access in communication [9]. Important steps towards a complete rigorous theory have been recently obtained. In particular the proof of the replica symmetric formula, a single letter formula for the asymptotic mutual information (MI), is now established for Gaussian RLE [1, 2, 4, 5]. In [2] the limits of optimality of the low complexity approximate message-passing denoising algorithm are explicitly established.

An important ingredient in the proofs of the replica formula are interesting relations for the rate of variation (the derivative) of the MI when: i) the signal-to-noise ratio varies [2]; ii) the measurement rate varies [5]. These formulas give the rate of variation directly in terms of the MMSE, and therefore belong to a "family" of I-MMSE relations, the simplest member of the family being the well known relation of Guo, Verdu and Shamai [10]. Of course once the replica symetric formula for the MI is available one can a posteriori check all these relations, however the proof of these relations does not involve any knowledge of the replica formula.

In this note we give a new derivation of the I-MMSE relation proved and used in [5]. The derivation given here explicitly shows that all I-MMSE relations are intimately connected. The main new technical ingredient is an interpolation method, here called *sub-extensive interpolation method*. It involves a mix of ideas originating in interpolation methods developed in recent years for dense and sparse graphical systems, and we believe it is of independent interest and is bound to have applications in other problems.

We end this introduction with a few (non-exhaustive) pointers to the literature that has led to the present work. The replica formula for Gaussian RLE was first proposed, on the basis of (non-rigorous) calculations using the replica method, by Tanaka [3] for the CDMA problem with binary input signals, and was later generalized in [11] (see also [12] for recent developments in the context of compressed sensing). Montanari and Tse [13] sketched a rigorous proof of Tanaka's formula in a regime where there is no phase transition (by which we here mean no jump discontinuity in the MMSE) and Korada and Macris [14, 15] used a Guerra-Toninelli [16] interpolation method to establish that the replica formula is always an upper bound to the MI. In [2, 4] the converse bound (and equality) is proven by using spatial coupling as a proof technique [17, 18] developed in the realm of spatially coupled graphical systems; namely spatial coupling for RLE [19–21], threshold saturation [22–26] and invariance of the MI under spatial coupling [17, 27].

II. Random linear estimation: Setting and results

A. Gaussian random linear estimation

We consider Gaussian RLE, where one is interested in reconstructing a signal $\mathbf{s} \in \mathbb{R}^N$ from measurements $\mathbf{y} \in \mathbb{R}^M$ obtained from the projection of $\mathbf{s}$ by a random i.i.d Gaussian *measurement matrix* $\phi \in \mathbb{R}^{M \times N}$. We consider i.i.d additive white Gaussian noise (AWGN) of variance Δ. Call the (standardized) noise components $Z_\mu \sim \mathcal{N}(0, 1)$, $\mu \in \{1, \dots, M\}$. The *RLE model* is

$$\mathbf{y} = \phi \mathbf{s} + \mathbf{z}\sqrt{\Delta} \iff y_\mu = \sum_{i=1}^{N} \phi_{\mu i} s_i + z_\mu \sqrt{\Delta}. \quad (1)$$

Consider a structured setting where the signal is made of L i.i.d B-dimensional *sections* $\mathbf{s}_l \in \mathbb{R}^B, l \in \{1, \dots, L\}$ distributed

according to a *discrete prior* $p_0(\mathbf{s}_l) := \sum_{k=1}^{K} p_k \delta(\mathbf{s}_l - \mathbf{a}_k)$ with a finite (but as large as desired) number of terms and all $\mathbf{a}_k$'s bounded (with maximum componentwise amplitude $s_{\max}$). Thus the total number of signal components is $N = LB$. We denote $\mathbf{X} \sim P_0$ if $\mathbf{X}_l \sim p_0$ for all its i.i.d sections. The matrix ϕ has entries $\phi_{\mu i} \sim \mathcal{N}(0, 1/L)$. We place ourselves in the high dimensional setting where the *measurement rate* $\alpha := M/N$ is fixed when letting $L \to \infty$ (B is always finite).

Define $\bar{\mathbf{x}} := \mathbf{x} - \mathbf{s}$, $[\phi\bar{\mathbf{x}}]_\mu := \sum_{i=1}^{N} \phi_{\mu i} \bar{x}_i$. The likelihood of $\mathbf{y}$ is $P(\mathbf{y}|\mathbf{x}) = (2\pi\Delta)^{-M/2} \exp(-\|\phi\mathbf{x} - \mathbf{y}\|^2/(2\Delta))$. From the Bayes formula, the posterior of the RLE model is

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{1}{2\Delta} \sum_{\mu=1}^{M} \left([\phi\bar{\mathbf{x}}]_\mu - z_\mu\sqrt{\Delta}\right)^2\right) P_0(\mathbf{x}), \quad (2)$$

where by a slight abuse of notation $\mathbf{y}$ here denotes the set of the independent *quenched* random variables $\phi, \mathbf{s}, \mathbf{z}$. The normalization $\mathcal{Z}$ is the integral w.r.t $\mathbf{x}$ of the numerator. The *Gibbs averages* w.r.t this posterior (2) are denoted by $\langle - \rangle$. For example the MMSE estimator is $\mathbb{E}_{\mathbf{X}|\mathbf{y}}[\mathbf{X}|\mathbf{y}] = \langle \mathbf{X} \rangle$.

B. I-MMSE relations

The mutual information per section $i_L := i(\mathbf{Y}; \mathbf{X})$ is

$$i_L = \frac{1}{L} \mathbb{E}_{\mathbf{\Phi}, \mathbf{X}, \mathbf{Y}}\left[\ln\left(\frac{P(\mathbf{Y}|\mathbf{X})}{P(\mathbf{Y})}\right)\right] = -\frac{\alpha B}{2} - \frac{1}{L} \mathbb{E}_{\mathbf{\Phi}, \mathbf{S}, \mathbf{Z}}[\ln \mathcal{Z}] \quad (3)$$

and the MMSE per section is

$$E_L := \frac{1}{L} \mathbb{E}_{\mathbf{\Phi}, \mathbf{S}, \mathbf{Z}}[\|\mathbf{S} - \langle \mathbf{X} \rangle\|^2]. \quad (4)$$

Moreover we define a *measurement MMSE* as

$$Y_M := \frac{1}{M} \mathbb{E}_{\mathbf{\Phi}, \mathbf{S}, \mathbf{Z}}[\|\mathbf{\Phi}(\mathbf{S} - \langle \mathbf{X} \rangle)\|^2],$$

linked to the MI by a canonical I-MMSE relation [2, 10]:

$$\frac{di_L}{d\Delta^{-1}} = \frac{\alpha B}{2} Y_M. \quad (5)$$

One can show thanks to the Guerra-Toninelli interpolation method that $i := \lim_{L\to\infty} i_L$ exists (see [15] for binary signals and [2] for general signals). It can also be shown from (5) that $d^2 i_L/d(\Delta^{-1})^2 \leq 0$ [28] (indeed the measurement MMSE cannot increase by increasing the snr) so i_L is a sequence of concave functions. We will repeatedly use that for a sequence of concave differentiable functions whose pointwise limit exists on $\mathbb{R}_+$, a standard result of real analysis states that: i) the limit is concave and continuous on all compact subsets; ii) the limit is differentiable at almost every (a.e.) point; iii) we can exchange the limit and derivative for a.e. points. Therefore for a.e. $\Delta > 0$ we have $\lim_{L\to\infty} di_L/d\Delta^{-1} = di/d\Delta^{-1}$.

We now present the new I-MMSE relations specific to Gaussian RLE.

Theorem 2.1 (snr I-MMSE relation): For Gaussian RLE with a discrete prior, $\lim_{L\to\infty} E_L = E$ exists and for a.e. Δ,

$$\frac{di}{d\Delta^{-1}} = \frac{\alpha B}{2} \frac{E}{1 + E/\Delta}. \quad (6)$$

Proof: As remarked above $\lim_{L\to\infty} di_L/d\Delta^{-1} = di/d\Delta^{-1}$ for a.e. Δ. Thus from (5) $(\alpha B/2)\lim_{M\to\infty} Y_M$

exists for a.e. Δ and equals $di/d\Delta^{-1}$. On the other hand Theorem 3.4 in [2] states that a.e. Δ,

$$Y_M = \frac{E_L}{1 + E_L/\Delta} + \mathcal{O}_L(1) \quad (7)$$

where $\lim_{L\to\infty} \mathcal{O}_L(1) = 0$. Therefore $E = \lim_{L\to\infty} E_L$ also exists for a.e. Δ and (6) holds. ∎

Remark 2.2: The proof of (7) in [2] (Theorem 3.4) requires concentration properties that are currently proven for a discrete prior. An extension to the case of a general prior (e.g., a mixture of discrete and absolutely continuous parts) could perhaps be obtained by quantizing the signal and showing that $\mathcal{O}_L(1)$ is uniform in the quantization but this is not immediate. A similar relation already appears in [13] but to the best of our knowledge the proof details are not given.

The main goal of this note is to give a new proof of the following I-MMSE relation first obtained in [1] (for $B = 1$).

Theorem 2.3 (α I-MMSE relation): For Gaussian RLE and for a.e. Δ and α,

$$\frac{di}{d\alpha} = \frac{B}{2} \ln\left(1 + \frac{E}{\Delta}\right). \quad (8)$$

Remark 2.4: We conjecture that the relation is true for all Δ and a.e α. The proof of [1] works for general priors that can have discrete and absolutely continuous parts. Our approach is based on concentration theorems underpinning (7) that as explained in Remark 2.2 do not quite cover this general case.

We end this section by noting that eliminating E from (6) and (8) we obtain the interesting and simple formula

$$\frac{2}{\alpha B} \frac{di}{d\ln(\Delta^{-1})} = 1 - \exp\left(-\frac{2}{B} \frac{di}{d\alpha}\right). \quad (9)$$

One may check (9) directly on the replica formula for the MI.

III. A SUB-EXTENSIVE INTERPOLATION METHOD

We introduce the sub-extensive interpolation method which allows to first prove a slightly weaker form of Theorem 2.3. All the stated lemmas are proved in the next section.

A. The interpolated perturbed model

The interpolation is done between the RLE model (1) where a measurement matrix with M lines is used and one where the measurement matrix has $M + M^u$ lines, $0 < u < 1$. The M^u additional lines have the same statistical properties and are indexed by a set $\mathcal{S}$ of extra indices. This is a sub-extensive set because $|\mathcal{S}| = M^u \ll M$. We still denote $\mathbf{y}, \phi$ the overall measurement and measurement matrix, that include the additional measurements and lines associated with $\mathcal{S}$. Define the following *interpolated perturbed* Hamiltonian

$$\mathcal{H}_{t,h}(\mathbf{x}; \mathbf{y}) := \frac{h}{2} \sum_{i=1}^{N} \left(\bar{x}_i - \frac{\widehat{z}_i}{\sqrt{h}}\right)^2 + \sqrt{h} s_{\max} \sum_{i=1}^{N} |\widehat{z}_i| \quad (10)$$

$$+ \frac{1}{2\Delta} \sum_{\mu=1}^{M} \left([\phi\bar{\mathbf{x}}]_\mu - z_\mu\sqrt{\Delta}\right)^2 + \frac{t}{2\Delta} \sum_{\nu \in \mathcal{S}} \left([\phi\bar{\mathbf{x}}]_\nu - z_\nu\sqrt{\frac{\Delta}{t}}\right)^2.$$

Here the interpolation parameter is $t \in [0, 1]$, and going from $t = 0$ to $t = 1$ continuously adds the M^u new measurements.

The first perturbation term corresponds to extra measurements obtained from scalar AWGN "side channels", $y_i = s_i + \widehat{z}_i/\sqrt{h}$, $\widehat{Z}_i \sim \mathcal{N}(0, 1)$, $i \in \{1, \ldots, N\}$, where the snr h is "small" and will eventually tend to zero. This term allows to use a useful concentration result proved in [2] (see Lemma 4.1 in sec. IV where also the second term is needed for technical reasons).

Denote the MI associated with the perturbed interpolated model $i_{t,h}$, expressed similarly to (3) but with $P_{t,h}(\mathbf{y}|\mathbf{x}) \propto \exp(-\mathcal{H}_{t,h}(\mathbf{x}; \mathbf{y}))$. This leads to $i_{t,h} = -B((1 + M^{u-1})\alpha + 1)/2 - \mathbb{E}[\ln \mathcal{Z}_{t,h}]/L$ where $\mathcal{Z}_{t,h} = \int d\mathbf{x}P_0(\mathbf{x}) \exp(-\mathcal{H}_{t,h}(\mathbf{x}; \mathbf{y}))$. The MI of this model with and without the additional measurements is, respectively, $i_{1,h}$ and $i_{0,h}$. The Gibbs average $\langle - \rangle_{t,h}$ is associated with the posterior of the interpolated perturbed model $P_{t,h}(\mathbf{x}|\mathbf{y}) \propto \mathcal{Z}_{t,h}^{-1}P_{t,h}(\mathbf{y}|\mathbf{x})P_0(\mathbf{x})$. Finally, we define the MMSE $E_{t,h}$ similarly as (4) but with $\langle \mathbf{X} \rangle_{t,h}$ replacing $\langle \mathbf{X} \rangle$.

B. The sub-extensive interpolation

We first show a weaker version of Theorem 2.3 for the perturbed interpolated model which is valid for all Δ but a.e. h. In sec. III-C we show how to take the limit $h \to 0$ for a.e. Δ and thus recover Theorem 2.3.

Theorem 3.1 (α I-MMSE relation for a.e. h): The following limits exist for *all* Δ and a.e. h, α and satisfy

$$\frac{d}{d\alpha} \lim_{L\to\infty} i_{0,h} = \frac{B}{2} \ln \left(1 + \frac{\lim_{L\to\infty} E_{0,h}}{\Delta}\right). \quad (11)$$

The proof is based on two lemmas proved in sec. IV.

Define a measurement MMSE associated to the subset $\mathcal{S}$:

$$Y_{t,h}^{(\mathcal{S})} := M^{-u} \sum_{\nu \in \mathcal{S}} \mathbb{E}[\langle [\mathbf{\Phi}\bar{\mathbf{X}}]_\nu \rangle_{t,h}^2], \quad (12)$$

where $\mathbb{E}$ denotes the expectation w.r.t all quenched variables.

Lemma 3.2 (MMSE relation): For a.e h we have

$$Y_{t,h}^{(\mathcal{S})} = \frac{E_{t,h}}{1 + E_{t,h}(t/\Delta)} + \mathcal{O}_L(1). \quad (13)$$

Lemma 3.3 (MMSE variation): Fix $0 < u < 1/20$ in the interpolated perturbed model (10). Then for any $t \in [0, 1]$, we have $E_{t,h} = E_{0,h} + \mathcal{O}_L(1)$ for a.e. h.

We now sketch the proof of Theorem 3.1.

Proof: By the fundamental theorem of calculus, one may write $i_{1,h} - i_{0,h} = \int_0^1 dt \, (di_{t,h}/dt)$. Direct differentiation gives

$$\frac{di_{t,h}}{dt} = \frac{1}{2\Delta L} \sum_{\nu \in \mathcal{S}} \mathbb{E}\left[\left\langle [\mathbf{\Phi}\bar{\mathbf{X}}]_\nu^2 - \frac{[\mathbf{\Phi}\bar{\mathbf{X}}]_\nu Z_\nu}{\sqrt{t/\Delta}} \right\rangle_{t,h}\right]. \quad (14)$$

Integrating by parts over $Z_\mu \sim \mathcal{N}(0, 1)$, (14) becomes

$$\frac{di_{t,h}}{dt} = \frac{\alpha B M^{u-1}}{2\Delta} Y_{t,h}^{(\mathcal{S})} = \frac{\alpha B M^{u-1}}{2\Delta}\left[\frac{E_{t,h}}{1 + E_{t,h}(t/\Delta)} + \mathcal{O}_L(1)\right]$$

for a.e h. For the second equality we used Lemma 3.2. Thus

$$i_{1,h} - i_{0,h} = \frac{\alpha B M^{u-1}}{2\Delta} \int_0^1 dt \frac{E_{t,h}}{1 + E_{t,h}(t/\Delta)} + \mathcal{O}(L^{u-1}). \quad (15)$$

This integral over t cannot be calculated immediately as the MMSE depends on t. We overcome this difficulty using Lemma 3.3 and $t' = t/\Delta$. Then (15) becomes

$$\frac{i_{1,h} - i_{0,h}}{\alpha M^{u-1}} = \frac{B}{2} \int_0^{1/\Delta} dt' \frac{E_{0,h}}{1 + E_{0,h}t'} + \mathcal{O}_L(1)$$

$$= \frac{B}{2} \ln \left(1 + \frac{E_{0,h}}{\Delta}\right) + \mathcal{O}_L(1). \quad (16)$$

Note that $i_{1,h}$ is the MI of a (perturbed) RLE model with measurement rate $(M + M^u)/N = \alpha(1 + M^{u-1})$ while $i_{0,h}$ corresponds to a measurement rate α. It is then not difficult to show with concavity inequalities w.r.t α,[1] that *for a.e.* α

$$\lim_{L\to\infty} \frac{i_{1,h} - i_{0,h}}{\alpha M^{u-1}} = \lim_{L\to\infty} \frac{di_{0,h}}{d\alpha} = \frac{d}{d\alpha} \lim_{L\to\infty} i_{0,h}. \quad (17)$$

Finally equations (16) and (17) imply (11). ∎

C. Proof of Theorem 2.3: taking the $h \to 0$ limit

We consider the limit $h \to 0$ of (11). Again, a concavity argument allows to permute this limit and the derivative for a.e. α. Also it is not very difficult to argue that all finite size quantities are continuous in $h \geq 0$. Therefore $\lim_{h\to 0} i_{0,h} = i_{0,0} = i_L$ and $\lim_{h\to 0} E_{0,h} = E_{0,0} = E_L$. So Theorem 2.3 follows if we can show that $\lim_{h\to 0} \lim_{L\to\infty} i_{0,h} = \lim_{L\to\infty} \lim_{h\to 0} i_{0,h} = i$ and $\lim_{h\to 0} \lim_{L\to\infty} E_{0,h} = \lim_{L\to\infty} \lim_{h\to 0} E_{0,h} = E$ for a.e. Δ. We will show that the first limit exchange is valid for all Δ and the second one for a.e. Δ.

For the first limit exchange the argument is standard. The first derivative of $i_{0,h}$ w.r.t h is an MMSE, namely $L^{-1}\mathbb{E}[\sum_{i=1}^N (S_i - \langle X_i \rangle_{0,h})^2]$, so its second derivative is negative because the MMSE cannot increase with increasing snr of the side channel (it can also be seen by explicit calculation [2]). Thus $i_{0,h}$ is concave in h, and since also $\lim_{L\to\infty} i_{0,h}$ exists, the limit is attained uniformly in h. This allows to exchange the limits for *all* Δ.

The second limit exchange is less immediate because we cannot use a convexity argument directly on the sequence $E_{0,h}$. However by a mild generalisation of Theorem 2.1 (that follows from Lemmas 4.5 and 4.6 in [2]) we have for a.e. Δ,

$$\frac{d}{d\Delta^{-1}} \lim_{L\to\infty} i_{0,h} = \frac{\alpha B}{2} \frac{\lim_{L\to\infty} E_{0,h}}{1 + \lim_{L\to\infty} E_{0,h}/\Delta}.$$

Then, since $\lim_{L\to\infty} i_{0,h}$ is a concave function of Δ and its limit $h \to 0$ exists we can take the limit $h \to 0$ of this equation and permute it with the derivative for a.e. Δ. Thus $\lim_{h\to 0} \lim_{L\to\infty} E_{0,h}$ must exist for a.e. Δ and satisfies

$$\frac{di}{d\Delta^{-1}} = \frac{\alpha B}{2} \frac{\lim_{h\to 0} \lim_{L\to\infty} E_{0,h}}{1 + \lim_{h\to 0} \lim_{L\to\infty} E_{0,h}/\Delta}.$$

But since we have (6), we conclude $\lim_{h\to 0} \lim_{L\to\infty} E_{0,h} = E$ for a.e. Δ. Thus the limits are exchangeable for a.e. Δ.

---

[1]Alternatively one can also directly use the Alexandrov theorem which states that a concave function has a second derivative almost everywhere.

IV. Proofs of Lemmas 3.2 and 3.3

A. Preliminaries

Let $\mathbf{X}$, $\mathbf{X}'$ two i.i.d *replicas* drawn according to the product distribution $P_{t,h}(\mathbf{x}|\mathbf{y})P_{t,h}(\mathbf{x}'|\mathbf{y})$. Then for any function g,

$$\mathbb{E}[\langle g(\mathbf{X},\mathbf{S})\rangle_{t,h}]=\mathbb{E}[\langle g(\mathbf{X},\mathbf{X}')\rangle_{t,h}]. \tag{18}$$

This identity, which has been called a Nishimori identity in the statistical mechanical literature, follows from a simple application of Bayes formula. It has a certain number of useful consequences that we list here (all the derivations can be found in appendix B of [2]).

1) Identity 1: First we have

$$2\mathbb{E}[\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\mu^2\rangle_{t,h}] = \mathbb{E}[\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\mu^2\rangle_{t,h}]. \tag{19}$$

To derive this recall $\bar{\mathbf{X}} = \mathbf{X} - \mathbf{S}$, expand the squares and systematically apply (18).

2) Identity 2: Set $\mathcal{E}:=L^{-1}\sum_{i=1}^N \bar{X}_i X_i$. Then (18) implies

$$\mathbb{E}[\langle\mathcal{E}\rangle_{t,h}] = E_{t,h}. \tag{20}$$

3) Identity 3: This one is more complicated. Define $u_\nu:=\sqrt{t/\Delta}[\phi\bar{\mathbf{x}}]_\nu - z_\nu$. From Gaussian integration by parts over z_ν and (18) one can show

$$\mathbb{E}[Z_\nu\langle U_\nu\bar{X}_i\bar{X}_i'\rangle_{t,h}]$$
$$= \mathbb{E}[Z_\nu^2 S_i\langle\bar{X}_i\rangle_{t,h}] - \sqrt{\frac{t}{\Delta}}\mathbb{E}[Z_\nu S_i\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu\bar{X}_i\rangle_{t,h}]. \tag{21}$$

We also need the following concentration result.

Lemma 4.1 (Concentration of $\mathcal{E}$): Let $\delta\mathcal{E}:=\mathcal{E}-E_{t,h}$ (recall (20)). For any $0 < a < \epsilon$ we have

$$\int_a^\epsilon dh\, \mathbb{E}[\langle\delta\mathcal{E}^2\rangle_{t,h}] = \mathcal{O}(L^{-1/10}). \tag{22}$$

The proof of this lemma is the same as the one of Proposition 8.1 in [2]. This type of result is also found in [15] for binary signals. Lebesgue's dominated convergence theorem applied to (22) implies $\mathbb{E}[\langle\delta\mathcal{E}^2\rangle_{t,h}]=\mathcal{O}_L(1)$ for a.e. $h>0$.

B. Proof of Lemma 3.2

Using (12) and an integration by parts w.r.t z_ν gives

$$Y_{t,h}^{(\mathcal{S})} = M^{-u}\sum_{\nu\in\mathcal{S}}\mathbb{E}\Big[\Big\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu^2 - \sqrt{\frac{\Delta}{t}}[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu Z_\nu\Big\rangle_{t,h}\Big],$$

which combined with (19) leads to

$$Y_{t,h}^{(\mathcal{S})} = M^{-u}\sqrt{\frac{\Delta}{t}}\sum_{\nu\in\mathcal{S}}\mathbb{E}[Z_\nu\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu\rangle_{t,h}]. \tag{23}$$

Integrating by part (23) again but this time w.r.t $\phi_{\nu i} \sim \mathcal{N}(0,1/L)$, one finds

$$Y_{t,h}^{(\mathcal{S})} = \frac{M^{-u}}{L}\sum_{\nu\in\mathcal{S}}\sum_{i=1}^N\mathbb{E}[Z_\nu\langle U_\nu\bar{X}_i\bar{X}_i'\rangle_{t,h} - Z_\nu\langle U_\nu\bar{X}_i^2\rangle_{t,h}]$$

(where $\bar{\mathbf{X}} = \mathbf{X} - \mathbf{S}, \bar{\mathbf{X}}' = \mathbf{X}' - \mathbf{S}$ and $\mathbf{X},\mathbf{X}'$ are i.i.d replicas). Then using (21) for the first term in the bracket and the

definition of u_ν for the second one, simple algebra leads to $Y_{t,h}^{(\mathcal{S})} = Y_1 - Y_2$ where

$$Y_1 = \mathbb{E}[(M^{-u}\sum_{\nu\in\mathcal{S}}Z_\nu^2)\langle\mathcal{E}\rangle_{t,h}],$$

$$Y_2 = \sqrt{\frac{t}{\Delta}}M^{-u}\sum_{\nu\in\mathcal{S}}\mathbb{E}[Z_\nu\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu\mathcal{E}\rangle_{t,h}].$$

The noise $\mathbf{z}$ has i.i.d standardized Gaussian components, so the central limit theorem implies

$$Y_1 = \mathbb{E}[\langle\mathcal{E}\rangle_{t,h}](1+\mathcal{O}(L^{-u/2})) = E_{t,h} + \mathcal{O}(L^{-u/2}).$$

Below we show that Lemma 4.1 implies for a.e. h,

$$Y_2 = \sqrt{\frac{t}{\Delta}}M^{-u}\sum_{\nu\in\mathcal{S}}\mathbb{E}[Z_\nu\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu\rangle_{t,h}]E_{t,h} + \mathcal{O}_L(1). \tag{24}$$

Then from (23) and (24) we get $Y_2 = (t/\Delta)Y_{t,h}^{(\mathcal{S})}E_{t,h} + \mathcal{O}_L(1)$. Putting all pieces together we get

$$Y_{t,h}^{(\mathcal{S})} = E_{t,h} - (t/\Delta)Y_{t,h}^{(\mathcal{S})}E_{t,h} + \mathcal{O}_L(1),$$

which is equivalent to (13) in Lemma 3.2.

It remains to justify (24). We have $\mathbb{E}[Z_\nu\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu\mathcal{E}\rangle_{t,h}] = \mathbb{E}[Z_\nu\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu\rangle_{t,h}]E_{t,h} + \mathbb{E}[Z_\nu\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu\delta\mathcal{E}\rangle_{t,h}]$. Thus it suffices to show that the second term is $\mathcal{O}_L(1)$ for a.e. h. From Cauchy-Schwarz

$$\mathbb{E}[Z_\nu\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu\delta\mathcal{E}\rangle_{t,h}]^2 \leq \mathbb{E}[\langle\delta\mathcal{E}^2\rangle_{t,h}]\mathbb{E}[Z_\nu^2\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu^2\rangle_{t,h}]. \tag{25}$$

As remarked below it, Lemma 4.1 implies $\mathbb{E}[\langle\delta\mathcal{E}^2\rangle_{t,h}] = \mathcal{O}_L(1)$ for a.e. h, thus we just have to argue that $\mathbb{E}[Z_\nu^2\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu^2\rangle_{t,h}]$ is bounded uniformly in L. By Cauchy-Schwarz again the square of this quantity is smaller than

$$\mathbb{E}[Z_\nu^4]\mathbb{E}[\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu^4\rangle_{t,h}] = 3\mathbb{E}[\langle[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu^4\rangle_{t,h}]. \tag{26}$$

Expanding $[\boldsymbol{\Phi}\bar{\mathbf{X}}]_\nu^4$ only terms of the form $\mathbb{E}[[\boldsymbol{\Phi}\mathbf{S}]_\nu^n\langle[\boldsymbol{\Phi}\mathbf{X}]_\nu^m\rangle_{t,h}]$ remain (with $0 \leq m,n \leq 4$). By Cauchy-Schwarz once more, their square is less than

$$\mathbb{E}[[\boldsymbol{\Phi}\mathbf{S}]_\nu^{2n}]\mathbb{E}[\langle[\boldsymbol{\Phi}\mathbf{X}]_\nu^{2m}\rangle_{t,h}] = \mathbb{E}[[\boldsymbol{\Phi}\mathbf{S}]_\nu^{2n}]\mathbb{E}[[\boldsymbol{\Phi}\mathbf{S}]_\nu^{2m}], \tag{27}$$

where the equality comes form the Nishimori identity (18). It is clear that these moments are all bounded uniformly in L. Indeed $\phi_{\mu i} \sim \mathcal{N}(0,1/L)$ is independent of $\mathbf{s}$, so conditional on $\mathbf{s}$, the linear combination $\boldsymbol{\Phi}\mathbf{s}$ is a Gaussian variable with a variance less than $Bs_{\max}^2$.

The proof of Lemma 3.2 is now complete.

C. Proof of Lemma 3.3

Recall (20). Then the MMSE difference can be written as

$$|E_{t,h} - E_{0,h}| = \Big|\int_0^t ds\frac{d}{ds}\mathbb{E}[\langle\mathcal{E}\rangle_{s,h}]\Big|$$
$$= \Big|\sum_{\nu\in\mathcal{S}}\int_0^t ds\,\mathbb{E}[\langle\mathcal{E}G_\nu\rangle_{s,h} - \langle\mathcal{E}\rangle_{s,h}\langle G_\nu\rangle_{s,h}]\Big|, \tag{28}$$

where $G_\nu := ([\phi\bar{\mathbf{x}}]_\nu^2 - [\phi\bar{\mathbf{x}}]_\nu z_\nu\sqrt{\Delta/s})/(2\Delta)$. Note that in (28) $\mathcal{E}$ can be replaced by $\delta\mathcal{E}$. Also, all G_ν's are statistically

equivalent and we can replace them by the first term in the set $\mathcal{S}$, say $\nu = 1$. Thus

$$|E_{t,h} - E_{0,h}| \leq M^u \int_0^t ds \big| \mathbb{E}[\langle \delta\mathcal{E} G_1 \rangle_{s,h} - \langle \delta\mathcal{E} \rangle_{s,h} \langle G_1 \rangle_{s,h}] \big|.$$

Integrating over $h \in [a, \epsilon]$, applying Fubini and Cauchy-Schwarz, one gets

$$\left(\int_a^\epsilon dh\, |E_{t,h} - E_{0,h}| \right)^2 \leq 4M^{2u} \int_0^t ds \int_a^\epsilon dh\, \mathbb{E}[\langle \delta\mathcal{E}^2 \rangle_{s,h}]$$
$$\times \int_0^t ds \int_a^\epsilon dh\, \mathbb{E}[\langle G_1^2 \rangle_{s,h}].$$

Proceeding similarly as in the steps (26)–(27) one shows that $\mathbb{E}[\langle G_1^2 \rangle_{s,h}] = \mathcal{O}(1)$ (w.r.t L). Lemma 4.1 allows to conclude

$$\int_a^\epsilon dh\, |E_{t,h} - E_{0,h}| = \mathcal{O}(M^u L^{-1/20}) = \mathcal{O}(L^{u-1/20}),$$

which implies Lemma 3.3 by Lebesgue's dominated convergence theorem as the integrand is bounded and $u < 1/20$.

V. Conclusion

Let us end by pointing another application of the sub-extensive interpolation method. In [2] it is used to prove the invariance of the MI under spatial coupling in RLE. There, one interpolates between a homogeneous measurement matrix and a spatially coupled one. This is done by iteratively removing sub-extensive blocks of lines in the homogeneous matrix and replacing them by "spatially coupled" lines. Along this process the MI is monotonously varying which leads to useful inequalities. This partly discrete, partly continuous interpolation defines a "family" of interpolation methods parametrized by u (the sub-extensive block size parameter). Roughly speaking our sub-extensive interpolation method "interpolates" between the purely global and continuous method of Guerra and Toninelli for dense graphical models [16] (at $u = 1$) and the combinatorial approach developed for sparse graphs by Gamarnik, Bayati and Tetali in [29] (at $u = 0$), where a discrete and local interpolation is done "one constraint at a time" (here one measurement at a time).

Acknowledgments

References

[1] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with gaussian matrices is exact," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016.

[2] J. Barbier, N. Macris, M. Dia, and F. Krzakala, "Mutual Information and Optimality of Approximate Message-Passing in Random Linear Estimation." [Online]. Available: https://arxiv.org/pdf/1701.05823v1.pdf

[3] T. Tanaka, "A statistical-mechanics approach to large-system analysis of cdma multiuser detectors," *IEEE Trans. on Information Theory*, 2002.

[4] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The Mutual Information in Random Linear Estimation," in *in the 54th Annual Allerton Conference on Communication, Control, and Computing*, September 2016.

[5] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with gaussian matrices is exact," 2016. [Online]. Available: http://arxiv.org/abs/1607.02524

[6] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. on Information Theory*, 2006.

[7] A. Barron and A. Joseph, "Toward fast reliable communication at rates near capacity with gaussian noise," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, 2010.

[8] G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. on Information Theory*, 2012.

[9] S. Verdú and S. Shamai, "Spectral efficiency of cdma with random spreading," *IEEE Trans. on Information Theory*, 1999.

[10] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Trans. on Information Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[11] D. Guo and S. Verdú, "Randomly spread cdma: asymptotics via statistical physics," *IEEE Trans. on Information Theory*, vol. 51, no. 6, pp. 1983–2010, May 2005.

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

[13] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of cdma (or: How to prove tanaka's formula)," in *2006 IEEE Information Theory Workshop (ITW) Punta del Este*, March 2006, pp. 160–164.

[14] S. Korada and N. Macris, "On the capacity of a code division multiple access system," in *Proc. Allerton Conf. Commun. Control Comput., Monticello, IL*, September 2007, pp. 959–966.

[15] S. B. Korada and N. Macris, "Tight bounds on the capacity of binary input random cdma systems," *IEEE Trans. on Information Theory*, vol. 56, no. 11, pp. 5590–5613, Nov 2010.

[16] F. Guerra, "An introduction to mean field spin glass theory: methods and results," *Mathematical Statistical Physics*, 2005.

[17] A. Giurgiu, N. Macris, and R. Urbanke, "Spatial coupling as a proof technique and three applications," *IEEE Trans. on Information Theory*, vol. 62, no. 10, pp. 5281–5295, Oct 2016.

[18] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula," in *Advances in Neural Information Processing Systems (NIPS) 29*, 2016, pp. 424–432.

[19] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Communication, Control, and Computing, 2010 48th Annual Allerton Conference on*, 2010.

[20] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08009, 2012.

[21] D. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. on Information Theory*, vol. 59, no. 11, pp. 7434–7464, November 2013.

[22] S. Hassani, N. Macris, and R. Urbanke, "Coupled graphical models and their thresholds," in *Information Theory Workshop (ITW), 2010*.

[23] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial coupling: Why convolutional ldpc ensembles perform so well over the bec," *IEEE Trans. on Information Theory*, 2011.

[24] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, "Threshold saturation for spatially coupled ldpc and ldgm codes on bms channels," *IEEE Trans. on Information Theory*, 2014.

[25] A. Yedla, Y. Y. Jian, P. S. Nguyen, and H. D. Pfister, "A simple proof of threshold saturation for coupled scalar recursions," in *Turbo Codes and Iterative Information Processing (ISTC), 2012 7th International Symposium on*, 2012.

[26] J. Barbier, M. Dia, and N. Macris, "Proof of Threshold Saturation for Spatially Coupled Sparse Superposition Codes," in *Information Theory Proceedings (ISIT), 2016 IEEE International Symposium on*, 2016.

[27] A. Giurgiu, N. Macris, and R. Urbanke, "How to prove the maxwell conjecture via spatial coupling: a proof of concept," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 458–462.

[28] M. Payaro and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector gaussian channels," *IEEE Trans. on Information Theory*, vol. 55, no. 8, pp. 3613–3628, Aug 2009.

[29] M. Bayati, D. Gamarnik, and P. Tetali, "Combinatorial approach to the interpolation method and scaling limits in sparse random graphs," *Ann. Probab.*, vol. 41, no. 6, pp. 4080–4115, 11 2013.