RSB Decoupling Property of MAP Estimators

Ali Bereyhi*, Ralf Müller*, Hermann Schulz-Baldes†
*Institute for Digital Communications (IDC), †Department of Mathematics,
Friedrich Alexander Universität (FAU), Erlangen, Germany
ali.bereyhi@fau.de, ralf.r.mueller@fau.de, schuba@mi.uni-erlangen.de

Abstract—**The large-system decoupling property of a MAP estimator is studied when it estimates the i.i.d. vector x from the observation $y = \mathbf{A}x + z$ with $\mathbf{A}$ being chosen from a wide range of matrix ensembles, and the noise vector z being i.i.d. and Gaussian. Using the replica method, we show that the marginal joint distribution of any two corresponding input and output symbols converges to a deterministic distribution which describes the input-output distribution of a single user system followed by a MAP estimator. Under the bRSB assumption, the single user system is a scalar channel with additive noise where the noise term is given by the sum of an independent Gaussian random variable and b correlated interference terms. As the bRSB assumption reduces to RS, the interference terms vanish which results in the formerly studied RS decoupling principle.**

I. INTRODUCTION

A linear vector system with Additive White Gaussian Noise (AWGN) is described by

$$y = \mathbf{A}x + z \qquad (1)$$

where the independent and identically distributed (i.i.d.) source vector $x_{n\times 1}$, taken from support $\mathbb{X}^n$, is measured by the random system matrix $\mathbf{A}_{k\times n}$ and corrupted by an i.i.d. Gaussian noise vector $z_{k\times 1}$. The observation vector y is given to the vector estimator $\mathbf{g}(\cdot)$ which maps the k-dimensional vector y to an n-dimensional vector $\hat{x}_{n\times 1} \in \mathbb{X}^n$. The entries of $\hat{x}$ are in general correlated due to the coupling imposed by $\mathbf{A}$ and $\mathbf{g}(\cdot)$. Considering the entries x_j and $\hat{x}_j$, $1 \leq j \leq n$, the marginal joint distribution of $(\hat{x}_j, x_j)$ in the large-system limit, i.e. $k, n \uparrow \infty$, is of interest. To clarify the point, consider the linear estimation, i.e. $\mathbf{g}(y) = \mathbf{G}^\mathsf{T} y$ for some $\mathbf{G}_{k\times n}$, and denote $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_n]$ and $\mathbf{G} = [\mathbf{g}_1 \cdots \mathbf{g}_n]$ with $\mathbf{a}_i$ and $\mathbf{g}_i$ being $k \times 1$ vectors for $i \in \{1, \ldots, n\}$. Thus,

$$\hat{x}_j = \left(\mathbf{g}_j^\mathsf{T} \mathbf{a}_j\right) x_j + \sum_{i=1, i\neq j}^{n} \left(\mathbf{g}_j^\mathsf{T} \mathbf{a}_i\right) x_i + \mathbf{g}_j^\mathsf{T} z. \qquad (2)$$

One considers the right hand side (r.h.s.) of (2) as the linear estimation of a single user system with additive impairment in which the impairment term is not necessarily Gaussian and the system is indexed by j. For some families of $\mathbf{A}$ and $\mathbf{G}$, it is shown that the index dependency of these systems vanishes and the impairment term converges to a Gaussian noise term with modified power level when the system dimensions tend to infinity, e.g. [1]. Thus, one can assume the linear vector estimator in the large-system limit to decouple into a bank of single user linear estimators operating over n parallel scalar systems with additive Gaussian noise terms. This decoupling

property of the linear estimators is rigorously justified invoking the central limit theorem and the properties of large random matrices. For nonlinear forms of $\mathbf{g}(\cdot)$, however, the analysis faces difficulties, since the output entries do not linearly decouple. Tanaka noted the similarity between the asymptotic analysis of spin glasses [2] and vector estimators and showed that the performance of a vector estimator in the large-system limit can be represented as the macroscopic parameter of a spin glass [3]. Consequently, a class of generally nonlinear estimators was analyzed using the nonrigorous replica method developed in statistical mechanics. Inspired by [3], several works employed the replica method to study the performance of nonlinear estimators in the asymptotic regime considering different classes of estimators, system matrices, and performance measures, e.g. [4]. Having the decoupling property of the linear estimators in mind, it was conjectured that this property holds for nonlinear estimators as well. In [5], Guo and Verdú justified this conjecture for the postulated Minimum Mean Square Error (MMSE) estimator

$$\mathbf{g}(y) = \mathsf{E}\{x|y, \mathbf{A}\} \qquad (3)$$

where $\mathbf{A}$ is considered to be i.i.d., and the expectation is taken over x due to some postulated posterior distribution $\mathsf{q}_{x|y,\mathbf{A}}$. For this setup, the authors showed the Replica Symmetry (RS) decoupling principle which says that under the RS assumption the marginal joint distribution of $(x_j, \hat{x}_j)$ converges to the input-output joint distribution of a scalar channel with additive Gaussian noise followed by a single user MMSE estimator. The RS decoupling principle was further extended to the case with a postulated Maximum-A-Posteriori (MAP) estimator in [6] where Rangan et al. studied

$$\mathbf{g}(y) = \arg\min_{v} \left[\frac{1}{2\lambda}\|y - \mathbf{A}v\|^2 + u(v)\right] \qquad (4)$$

for some "utility function" $u(\cdot) : \mathbb{R}^n \to \mathbb{R}^+$ and non-negative real "estimation parameter" λ. Except for the cases with an i.i.d. system matrix, the decoupling property of nonlinear estimators for a larger class of matrix ensembles has not yet been addressed precisely. In [7], the authors investigated this issue partially by studying the support recovery of sparse Gaussian sources. They considered the case of a source vector which is first randomly measured by a squared matrix, and then, the measurements are sparsely sampled by an i.i.d. binary vector. Employing a MAP estimator for recovering, the RS decoupling principle was justified for the case in which the measuring matrix belongs to a large set of matrix ensembles. Although the class of matrices is broadened in [7], the result cannot be considered as a complete generalization of [5] and

[6], since it is restricted to cases with a sparse Gaussian source and $kn^{-1} \leq 1$. Another issue not investigated in the literature is the marginal joint distribution under the Replica Symmetry Breaking (RSB) assumption. In fact, the previous studies investigated the decoupling principle considering the RS ansatz; however, despite the RS validity in some particular cases, there are still several cases requiring further RSB investigations, e.g. [8], [9].

In this paper, we address both the issues and broaden the scope of the decoupling principle stated in [6] to both a larger set of matrix ensembles, and the RSB ansätze. More precisely, we justify the decoupling property of the postulated MAP estimator when

1) $\mathbf{A}$ is chosen from a large family of random matrices,
2) kn^{-1} takes any non-negative real number, and
3) the RS and RSB ansätze are considered.

For this setup, we show that under all replica ansätze, the joint distribution of $(x_j, \hat{x}_j)$ in the large-system limit converges to the input-output distribution of a scalar system in which the source symbol is corrupted by effective noise and estimated by a single user MAP estimator. We determine the effective noise term and estimation parameter under the RSB assumption with b steps of breaking (bRSB), and show that the noise term under this assumption is given by the sum of an independent Gaussian random variable and b correlated terms. By reducing the assumption to RS, the correlated terms vanish, and the noise term becomes Gaussian. Thus, one can consider the decoupling principle of [6] to be a special case of the more general decoupling principle illustrated here.

Notation: We represent vectors, scalars and matrices with bold lower case, non-bold lower case, and bold upper case letters, respectively. The set of real numbers is denoted by $\mathbb{R}$, and $\mathbf{A}^{\mathsf{T}}$ and $\mathbf{A}^{\mathsf{H}}$ indicate the transposed and Hermitian of $\mathbf{A}$. $\mathbf{I}_m$ is the $m \times m$ identity matrix, $\mathbf{1}_m$ is the matrix with all entries equal to one, and $\otimes$ denotes the Kronecker product. For a random variable x, p_x represents either the Probability Mass Function (PMF) or Probability Density Function (PDF), and F_x represents the Cumulative Distribution Function (CDF). We denote the expectation over x by E_x, and an expectation over all random variables involved in a given expression by E. For sake of compactness, the set of integers $\{1, \ldots, n\}$ is denoted by $[1:n]$, the zero-mean and unit-variance Gaussian PDF by $\pi(\cdot)$, and

$$\int \mathrm{D}t := \int \pi(t)\mathrm{d}t. \tag{5}$$

Whenever needed, we consider the entries of $\boldsymbol{x}$ to be discrete random variables; the results of this paper, however, are in full generality and directly extend to continuous distributions.

II. PROBLEM FORMULATION

Let the system in (1) satisfy the following constraints.

(a) The number of observations k is a deterministic sequence of n such that

$$\lim_{n\uparrow\infty} \frac{k}{n} = \frac{1}{\mathrm{r}} < \infty. \tag{6}$$

(b) $\boldsymbol{x}_{n\times 1}$ is an i.i.d. random vector with each element being distributed due to p_x over $\mathbb{X}$ in which $\mathbb{X} \subseteq \mathbb{R}$.
(c) $\mathbf{A}_{k\times n}$ is randomly generated over $\mathbb{A}^{k\times n} \subseteq \mathbb{R}^{k\times n}$, such that $\mathbf{J} = \mathbf{A}^{\mathsf{T}}\mathbf{A}$ has the eigendecomposition

$$\mathbf{J} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathsf{T}} \tag{7}$$

where $\mathbf{U}$ is an orthogonal Haar distributed matrix and $\mathbf{D}$ is a diagonal matrix with the empirical eigenvalue distribution (density of states) converging as $n \uparrow \infty$ to a deterministic distribution $\mathrm{F}_{\mathbf{J}}$.
(d) $\boldsymbol{z}_{k\times 1}$ is a real i.i.d. zero-mean Gaussian random vector with variance λ_0, i.e., $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \lambda_0\mathbf{I}_k)$.
(e) $\boldsymbol{x}$, $\mathbf{A}$, and $\boldsymbol{z}$ are independent.

In order to estimate the source vector, the postulated MAP estimator as defined in (4) is employed. The estimators postulates a non-negative estimation parameter λ and a non-negative utility function $u(\cdot)$ which decouples, i.e., $u(\boldsymbol{x}) = \sum_{i=1}^{n} u(x_i)$.

Defining the estimated vector $\hat{\boldsymbol{x}} := \mathbf{g}(\boldsymbol{y})$, the conditional distribution of $\hat{x}_j$ given x_j for some $j \in [1:n]$ is denoted by $\mathrm{p}_{\hat{x}|x}^{j(n)}$. Thus, the marginal joint distribution of x_j and $\hat{x}_j$ at the mass point $(\hat{v}, v)$ is written as

$$\mathrm{p}_{\hat{x}_j, x_j}(\hat{v}, v) = \mathrm{p}_x(v)\mathrm{p}_{\hat{x}|x}^{j(n)}(\hat{v}|v). \tag{8}$$

Considering the large-system limit, we define the asymptotic conditional distribution of $\hat{x}_j$ given x_j at $(\hat{v}, v)$ as

$$\mathrm{p}_{\hat{x}|x}^{j}(\hat{v}|v) := \lim_{n\uparrow\infty} \mathrm{p}_{\hat{x}|x}^{j(n)}(\hat{v}|v). \tag{9}$$

We also suppose the self averaging assumption which says

(f) Given $\mathbf{A}$ of the form (7) with $\mathrm{F}_{\mathbf{J}}$, the limit in (9) exists and is almost surely constant in realizations of $\mathbf{A}$.

III. GENERAL DECOUPLING PRINCIPLE

The main contribution of this study is to extend the scope of the decoupling principle. To illustrate the result, consider the following single user system: the input x is passed through the channel $y = x + z$ where $z \sim \mathrm{p}_{z|x}$ for the given input x. The observation y is then given to a single user MAP estimator with the same utility function as for the vector estimator defined in Section II, i.e. $u(\cdot)$, and an estimation parameter denoted by λ^{s}. Indicating the conditional distribution of the estimator's output $\hat{x}$ for the given input x by $\mathrm{p}_{\hat{x}|x}$, our general decoupling principle says that under a set of assumptions

(a) the asymptotic conditional distribution $\mathrm{p}_{\hat{x}|x}^{j}$ is independent of the index j, and we have $\mathrm{p}_{\hat{x}|x}^{j} = \mathrm{p}_{\hat{x}|x}$.
(b) $\mathrm{p}_{z|x}$ and λ^{s} are determined in terms of λ, λ_0 and the statistics of $\boldsymbol{x}$ and $\mathbf{A}$.

The set of assumptions which yields the validity of the above statements are enforced within the large-system analysis. In the following, we briefly illustrate our approach and determine the parameters of the single user system.

IV. DERIVATION OF GENERAL DECOUPLING PRINCIPLE

Before illustrating our derivation approach, let us define the R-transform. For a random variable t, the Stieltjes transform

over the upper half complex plane is defined as $G_t(s) = \mathsf{E}\,[t - s]^{-1}$. Denoting the inverse with respect to (w.r.t.) composition with $G_t^{-1}(\cdot)$, the R-transform is $R_t(\omega) = G_t^{-1}(\omega) - \omega^{-1}$ such that $\lim_{\omega\downarrow 0} R_t(\omega) = \mathsf{E}\,t$. The definition also extends to matrix arguments. Assuming a matrix $\mathbf{M}_{n\times n}$ to have the eigendecomposition $\mathbf{M} = \mathbf{U}\,\mathrm{diag}[\lambda_1,\ldots,\lambda_n]\,\mathbf{U}^{-1}$, $R_t(\mathbf{M})$ is then defined as $R_t(\mathbf{M}) = \mathbf{U}\,\mathrm{diag}[R_t(\lambda_1),\ldots,R_t(\lambda_n)]\,\mathbf{U}^{-1}$.

The derivation of the general decoupling principle is based on the moment method. To clarify the approach, consider the non-negative integers k and ℓ, and define the joint moment $\mathsf{M}_{k,\ell}^{j(n)} = \mathsf{E}\,\hat{x}_j^k x_j^\ell$, for $j \in [1:n]$. After evaluating the limit of $\mathsf{M}_{k,\ell}^{j(n)}$ as $n\uparrow\infty$, we show that for all k and ℓ the asymptotic joint moment is equivalent to the corresponding joint moment of the single user system. Consequently, using the uniqueness of the mapping from the set of integer moments' sequences to the set of measures, under a set of conditions investigated in the classical moment problem [10], we conclude that both couples $(\hat{x}_j, x_j)$ and $(\hat{x}, x)$ have a same distribution. We start with evaluating the limit of $\mathsf{M}_{k,\ell}^{j(n)}$. The evaluation is based on the nonrigorous method of replicas developed in the theory of spin glasses [2], and accepted as a mathematical tool in information theory. To do so, define the "weighted average joint moment" over the index set $\mathbb{W} \subset [1:n]$ as

$$\mathsf{M}_{k,\ell}^{\mathbb{W}(n)}(\hat{\boldsymbol{x}};\boldsymbol{x}) := \mathsf{E}\,\frac{1}{|\mathbb{W}|}\sum_{w\in\mathbb{W}}\hat{x}_w^k x_w^\ell. \qquad (10)$$

Setting $\mathbb{W} = [j : j + n\eta]$ for some $\eta \in (0,1]$, the asymptotic joint moment of jth entry can be written as

$$\mathsf{M}_{k,\ell}^j := \lim_{n\uparrow\infty}\mathsf{M}_{k,\ell}^{j(n)} = \lim_{n\uparrow\infty}\lim_{\eta\downarrow 0}\mathsf{M}_{k,\ell}^{\mathbb{W}(n)}(\hat{\boldsymbol{x}};\boldsymbol{x}). \qquad (11)$$

Thus, the evaluation of the asymptotic moment reduces to taking the limits in the r.h.s. of (11) which needs the weighted average joint moment in (10) to be explicitly calculated for an arbitrary integer n. Alternatively, we can define the function

$$\mathcal{Z}(\beta,h) = \sum_{\boldsymbol{v}} e^{-\beta\left[\frac{1}{2\lambda}\|\boldsymbol{y}-\mathbf{A}\boldsymbol{v}\|^2 + u(\boldsymbol{v})\right] + hn\mathsf{M}_{k,\ell}^{\mathbb{W}(n)}(\boldsymbol{v};\boldsymbol{x})}. \qquad (12)$$

with $\boldsymbol{v} \in \mathbb{X}^n$. Noting that $\hat{\boldsymbol{x}} = \mathbf{g}(\boldsymbol{y})$ with $\mathbf{g}(\cdot)$ defined in (4),

$$\mathsf{M}_{k,\ell}^{\mathbb{W}(n)}(\hat{\boldsymbol{x}};\boldsymbol{x}) = \lim_{\beta\uparrow\infty}\lim_{h\downarrow 0}\frac{1}{n}\frac{\partial}{\partial h}\,\mathsf{E}\log\mathcal{Z}(\beta,h). \qquad (13)$$

The logarithmic expectation in the r.h.s. of (13) is not a trivial task to do, and therefore, one bypasses the direct evaluation using the Riesz equality which for any random variable t states $\mathsf{E}\log t = \lim_{m\downarrow 0} m^{-1}\log\mathsf{E}\,t^m$. Thus, regarding (11) and (13)

$$\mathsf{M}_{k,\ell}^j = \lim_{n\uparrow\infty}\lim_{\eta\downarrow 0}\lim_{\beta\uparrow\infty}\lim_{h\downarrow 0}\lim_{m\downarrow 0}\frac{1}{n}\frac{\partial}{\partial h}\frac{\log\mathsf{E}\,[\mathcal{Z}(\beta,h)]^m}{m} \qquad (14)$$

for $\mathbb{W} = [j : j + n\eta]$. In (14), we face two major difficulties: 1) evaluating the real moments i.e., $\mathsf{E}\,[\mathcal{Z}(\beta,h)]^m$, and 2) taking the limits in the order stated. Basic analytical methods fail to address these challenges properly, and therefore, we invoke the nonrigorous method of replicas. The replica method suggests to evaluate the moment for an arbitrary integer m as an analytic function in m; then, assume that 1) the "replica

continuity" holds which means that the function analytically continues from the set of integers to the real axis (or at least a vicinity of zero), and 2) the limits are exchangeable. Following the above prescription, we consider the first assumption and find $\mathsf{E}\,[\mathcal{Z}(\beta,h)]^m$ which for an integer m reduces to

$$\mathsf{E}\prod_{a=1}^m\sum_{\boldsymbol{v}_a}e^{-\beta\left[\frac{1}{2\lambda}\|\mathbf{A}(\boldsymbol{x}-\boldsymbol{v}_a)+\boldsymbol{z}\|^2 + u(\boldsymbol{v}_a)\right] + hn\mathsf{M}_{k,\ell}^{\mathbb{W}(n)}(\boldsymbol{v}_a;\boldsymbol{x})}. \qquad (15)$$

In order to evaluate (15), one can initially take the expectation over $\boldsymbol{z}$ and $\mathbf{A}$. Due to the lack of space, we leave the details for the extended version of the manuscript; however, we briefly explain the strategy. After taking the expectations, and defining the $m \times m$ "replica correlation matrix" $\mathbf{Q}$ such that $[\mathbf{Q}]_{ab} = n^{-1}(\boldsymbol{x} - \boldsymbol{v}_a)^\mathsf{T}(\boldsymbol{x} - \boldsymbol{v}_b)$, (15) is given in terms of $\mathbf{Q}$ as

$$\mathsf{E}\,[\mathcal{Z}(\beta,h)]^m = \mathsf{E}_{\boldsymbol{x}}\int e^{-n\mathcal{G}(\mathbf{T}\mathbf{Q})}e^{n\mathcal{I}(\mathbf{Q})}\mathrm{d}\mathbf{Q} \qquad (16)$$

with $\mathrm{d}\mathbf{Q} := \prod_{a,b=1}^m \mathrm{d}[\mathbf{Q}]_{ab}$, $\mathbf{T} := \frac{1}{2\lambda}\mathbf{I}_m - \beta\frac{\lambda_0}{2\lambda^2}\mathbf{1}_m$, and the integral being taken over $\mathbb{R}^{m^2}$. For a given $\boldsymbol{x}$, $e^{n\mathcal{I}(\mathbf{Q})}$ measures the probability weight of the set of replicas, $\{\boldsymbol{v}_a\}_{a=1}^m$, in which the correlation matrix is $\mathbf{Q}$; moreover, $\mathcal{G}(\cdot)$ is defined as

$$\mathcal{G}(\mathbf{M}) = \int_0^\beta \mathrm{Tr}\{\mathbf{M}R_\mathbf{J}(-2\omega\mathbf{M})\}\mathrm{d}\omega + \epsilon_n \qquad (17)$$

where $\mathrm{Tr}\{\cdot\}$ denotes the trace, $R_\mathbf{J}(\cdot)$ is the R-transform w.r.t. $F_\mathbf{J}$, and ϵ_n tends to zero as $n\uparrow\infty$. Here, one can employ the Laplace method of integration and replace the r.h.s. of (16) in the large-system limit with the integrand at its saddle point multiplied by some bounded coefficient K_n; thus, as $n\uparrow\infty$

$$\mathsf{E}\,[\mathcal{Z}(\beta,h)]^m \doteq \mathsf{K}_n e^{-n[\mathcal{G}(\mathbf{T}\tilde{\mathbf{Q}}) - \mathcal{I}(\tilde{\mathbf{Q}})]} \qquad (18)$$

at the saddle point $\tilde{\mathbf{Q}}$. Substituting (18) in (14), we have

$$\mathsf{M}_{k,\ell}^j = \lim_{\eta\downarrow 0}\lim_{\beta\uparrow\infty}\lim_{m\downarrow 0}\mathsf{E}\,\frac{\sum_{\mathbf{v}}\mathsf{M}_{k,\ell}^{[1:m]}(\mathbf{v};\mathbf{x})\mathcal{E}(\tilde{\mathbf{Q}},\mathbf{x},\mathbf{v};\beta)}{\sum_{\mathbf{v}}\mathcal{E}(\tilde{\mathbf{Q}},\mathbf{x},\mathbf{v};\beta)} \qquad (19)$$

where $\mathbf{x}_{m\times 1} = [x,\ldots,x]^\mathsf{T}$ with $x \sim \mathrm{p}_x$, $\mathbf{v}_{m\times 1} \in \mathbb{X}^m$ and

$$\mathcal{E}(\tilde{\mathbf{Q}},\mathbf{x},\mathbf{v};\beta) = e^{-\beta(\mathbf{x}-\mathbf{v})^\mathsf{T}\mathbf{T}R_\mathbf{J}(-2\beta\mathbf{T}\tilde{\mathbf{Q}})(\mathbf{x}-\mathbf{v}) - \beta u(\mathbf{v})}. \qquad (20)$$

Here, one needs to find the saddle point which is not a feasible task in general. The strategy for pursuing the analysis is to restrict the saddle point, i.e. $\tilde{\mathbf{Q}}$, to be of a special form, and find the solution within the restricted set of matrices. This is where an additional assumption such as the RS or RSB assumption arises. It is clear that these restrictions do not lead us to the correct solution in general, and therefore, one needs to widen the set of replica correlation matrices to find a more accurate solution. In the sequel, we consider different structures on the correlation matrix and find the replica ansatz under those assumptions. However, considering (19), it is observed that "regardless of the structure" on $\tilde{\mathbf{Q}}$, the joint moment $\mathsf{M}_{k,\ell}^j$ is independent of the index j even before taking the limit $\eta \downarrow 0$. Therefore, employing the moment method, we conclude

Proposition 1 *Let the vector system satisfy the constraints in Section II; moreover, assume the replica continuity to hold,*
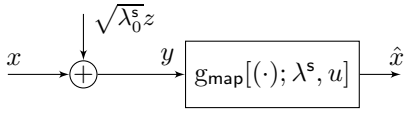
Fig. 1: The decoupled scalar system under the RS ansatz.

Fig. 2: The decoupled scalar system under the 1RSB ansatz.

and the limits in (14) *to exist and exchange. Then* $\mathrm{p}^{j}_{\hat{x}|x}$*, as defined in* (9)*, does not depend on the index* j*.*

Proposition 1 states a more general form of the decoupling principle studied in previous works. In fact, the only assumptions which need to be satisfied are the replica continuity and the exchange of limits; and, no structure of the correlation matrix is imposed. However, the decoupled scalar system does depend on the structure imposed on $\tilde{\mathbf{Q}}$. To find the decoupled single user system, we start with the most primary structure which is imposed by considering the RS assumption.

RS Assumption: Here, we restrict the search to the set of parameterized matrices which are of the form

$$\tilde{\mathbf{Q}} = q\mathbf{1}_m + \frac{\chi}{\beta}\mathbf{I}_m. \tag{21}$$

for some $\chi, q \in \mathbb{R}^+$. Substituting in (19), we have

$$\mathsf{M}^{j}_{k,\ell} = \mathsf{E} \int \mathrm{g}^k x^\ell \mathrm{D}z \tag{22}$$

where $\mathrm{g} := \mathrm{g_{map}}[(y); \lambda^{\mathrm{s}}, u]$ with $y = x + \sqrt{\lambda^{\mathrm{s}}_0}z$ and

$$\mathrm{g_{map}}[(y); \lambda^{\mathrm{s}}, u] = \arg\min_v \left[\frac{1}{2\lambda^{\mathrm{s}}}(y-v)^2 + u(v)\right]. \tag{23}$$

Moreover, λ^{s}_0 and λ^{s} are defined as

$$\lambda^{\mathrm{s}}_0 = \left[\mathrm{R_J}(-\frac{\chi}{\lambda})\right]^{-2} \frac{\partial}{\partial\chi} \left\{[\lambda_0\chi - \lambda q]\,\mathrm{R_J}(-\frac{\chi}{\lambda})\right\} \tag{24a}$$

$$\lambda^{\mathrm{s}} = \left[\mathrm{R_J}(-\frac{\chi}{\lambda})\right]^{-1} \lambda \tag{24b}$$

where $q = \mathsf{E} \int [\mathrm{g} - x]^2 \mathrm{D}z$, and χ satisfies

$$\sqrt{\lambda^{\mathrm{s}}_0}\chi = \lambda^{\mathrm{s}}\mathsf{E} \int [\mathrm{g} - x]z\mathrm{D}z. \tag{25}$$

(23) describes a single user MAP estimator with the postulated utility function $u(\cdot)$ and estimation parameter λ^{s}. Thus,

Proposition 2 *Let the assumptions in Proposition 1, as well as the RS assumption hold, and consider the single user system in Figure 1 with λ^{s}_0 and λ^{s} as defined in* (24a) *and* (24b)*. Then for $j \in [1:n]$, $\mathrm{p}^{j}_{\hat{x}|x}$ as defined in* (9) *describes the conditional distribution of $\hat{x}$ given x in Figure 1 where $\mathrm{g_{map}}[(\cdot); \lambda^{\mathrm{s}}, u]$ is a single user MAP defined in* (23)*, and $\mathrm{p}_{z|x}(z|x) = \pi(z)$.*

The results in the literature have always considered the RS ansatz, and can be recovered as special cases of Proposition 2. E.g., results of [6] are derived by setting $\mathrm{R_J}(\omega) = (1-\mathrm{r}\omega)^{-1}$. The RS ansatz, however, does not provide a valid solution, in general. Parisi in [11] introduced the RSB scheme which widens the restricted set of saddle point matrices recursively. To illustrate the RSB scheme, let $\mathbf{Q}^{\mathrm{b}}$ be a basic structure for the replica correlation matrix; moreover, assume m to be a multiple of an integer ξ. Then, the correlation matrix can be grouped as a $\xi \times \xi$ matrix of blocks with each block being an
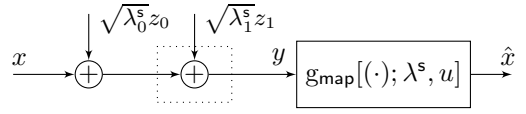
$\frac{m}{\xi} \times \frac{m}{\xi}$ matrix. In this case, a new structure for the correlation matrix is obtained by setting the diagonal blocks to be $\mathbf{Q}^{\mathrm{b}}$ and the off-diagonal blocks to be $\kappa\mathbf{1}_{\frac{m}{\xi}}$ for some κ. In fact, the new set of correlation matrices is constructed by imposing the RS structure block-wisely using $\mathbf{Q}^{\mathrm{b}}$ and $\kappa\mathbf{1}_{\frac{m}{\xi}}$ as basic blocks. The RSB scheme can be recursively iterated: one can set the basic structure $\mathbf{Q}^{\mathrm{b}}$ and find the new structure $\mathbf{Q}^{\mathrm{b}}_1$; then, by taking $\mathbf{Q}^{\mathrm{b}}_1$ as the new basic structure, a wider set of correlation matrices is found. Parisi considers $\mathbf{Q}^{\mathrm{b}}$ to have the RS structure.

1RSB Assumption: Using the RSB scheme with one step of iteration, the structure of the correlation matrix is found as

$$\tilde{\mathbf{Q}} = q\mathbf{1}_m + p\mathbf{I}_{\frac{m\beta}{\mu}} \otimes \mathbf{1}_{\frac{\mu}{\beta}} + \frac{\chi}{\beta}\mathbf{I}_m \tag{26}$$

for some $\chi, p, q, \mu \in \mathbb{R}^+$. Therefore, the joint moment reads

$$\mathsf{M}^{j}_{k,\ell} = \mathsf{E} \int \mathrm{g}^k x^\ell \mathrm{dF}(z_1)\mathrm{D}z_0 \tag{27}$$

where $\mathrm{g} := \mathrm{g_{map}}[(y); \lambda^{\mathrm{s}}, u]$ with $y = x + \sqrt{\lambda^{\mathrm{s}}_0}z_0 + \sqrt{\lambda^{\mathrm{s}}_1}z_1$, and $\mathrm{dF}(z_1) = \Lambda\mathrm{D}z_1$ with $\Lambda = [\int \tilde{\Lambda}\mathrm{D}z_1]^{-1}\tilde{\Lambda}$ and

$$\tilde{\Lambda} = e^{-\mu\left\{\frac{1}{2\lambda^{\mathrm{s}}}[(y-\mathrm{g})^2 - (y-x)^2] + u(\mathrm{g})\right\}}. \tag{28}$$

Moreover, by denoting $\varrho := \chi + \mu p$, we have

$$\lambda^{\mathrm{s}}_0 = \left[\mathrm{R_J}(-\frac{\chi}{\lambda})\right]^{-2} \frac{\partial}{\partial\varrho} \left\{[\lambda_0\varrho - \lambda q + \lambda p]\,\mathrm{R_J}(-\frac{\varrho}{\lambda})\right\}, \tag{29a}$$

$$\lambda^{\mathrm{s}}_1 = \left[\mathrm{R_J}(-\frac{\chi}{\lambda})\right]^{-2} \left[\mathrm{R_J}(-\frac{\chi}{\lambda}) - \mathrm{R_J}(-\frac{\varrho}{\lambda})\right] \lambda\mu^{-1}, \tag{29b}$$

$$\lambda^{\mathrm{s}} = \left[\mathrm{R_J}(-\frac{\chi}{\lambda})\right]^{-1} \lambda. \tag{29c}$$

Here, $q = \mathsf{E} \int [\mathrm{g} - x]^2\mathrm{dF}(z_1)\mathrm{D}z_0$, and χ and p satisfy

$$\chi + \mu p = \frac{\lambda^{\mathrm{s}}}{\sqrt{\lambda^{\mathrm{s}}_0}}\mathsf{E} \int [\mathrm{g} - x]z_0\mathrm{dF}(z_1)\mathrm{D}z_0 \tag{30a}$$

$$\chi + \mu q = \frac{\lambda^{\mathrm{s}}}{\sqrt{\lambda^{\mathrm{s}}_1}}\mathsf{E} \int [\mathrm{g} - x]z_1\mathrm{dF}(z_1)\mathrm{D}z_0 \tag{30b}$$

for some μ being a solution to the fixed point equation

$$\frac{\mu}{2\lambda^{\mathrm{s}}} \left[\mu\frac{\lambda^{\mathrm{s}}_1}{\lambda^{\mathrm{s}}}q - \mu\frac{\lambda^{\mathrm{s}}_1}{\lambda^{\mathrm{s}}}p + p\right] - \frac{1}{2\lambda}\int_{\chi}^{\varrho} \mathrm{R_J}(-\frac{\omega}{\lambda})\mathrm{d}\omega =$$
$$= \mathsf{I}(z_1; x, z_0) + \mathsf{D_{KL}}(\mathrm{p}_{z_1}\|\pi) \tag{31}$$

where $\mathsf{I}(\cdot; \cdot)$ and $\mathsf{D_{KL}}(\cdot\|\cdot)$ indicate the mutual information and the so-called "Kullback-Leibler" distance respectively, and the random variables $(x, z_0, z_1) \sim \mathrm{p}_x(x)\pi(z_0)[\Lambda\pi(z_1)]$. Thus, one can conclude the following proposition.

Proposition 3 *Let the assumptions in Proposition 1, as well as the 1RSB assumption hold, and consider the single user system in Figure 2 with λ^{s}_0, λ^{s}_1 and λ^{s} as defined in* (29a)-(29c)*. Then, for $j \in [1:n]$, $\mathrm{p}^{j}_{\hat{x}|x}$ in* (9) *is the conditional distribution*
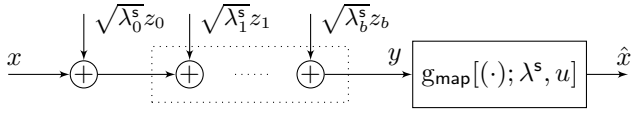
Fig. 3: The decoupled scalar system under the bRSB ansatz.

of $\hat{x}$ given x in Figure 2 where $g_{\text{map}}[(\cdot); \lambda^{\text{s}}, u]$ *is a single user MAP estimator defined in* (23), $p_{z_0|x}(z_0|x) = \pi(z_0)$, *and*

$$p_{z_1|x,z_0}(z_1|x,z_0) = \Lambda \pi(z_1) \tag{32}$$

with $\Lambda = [\int \tilde{\Lambda} \mathrm{D}z_1]^{-1}\tilde{\Lambda}$ *and* $\tilde{\Lambda}$ *defined in* (28).

Here, the decoupled system differs from the system obtained under the RS ansatz within one additive tap which is intuitively approximating the interference caused by the coupling. In fact, the RS ansatz assumes the coupling caused by the system matrix and vector estimator to vanish as the system tends to its large limits; however, the 1RSB solution takes the coupling into account and approximates it with one tap of interference. This approximation may become more accurate, if we let the correlation matrix to be chosen from a larger set of matrices. **bRSB Assumption:** Iterating the RSB scheme with b steps,

$$\tilde{\mathbf{Q}} = q\mathbf{1}_m + \sum_{\nu=1}^{b} p_\nu \mathbf{I}_{\frac{m\beta}{\mu_\nu}} \otimes \mathbf{1}_{\frac{\mu_\nu}{\beta}} + \frac{\chi}{\beta}\mathbf{I}_m \tag{33}$$

for some $\chi, q, \{p_\nu, \mu_\nu\}_{\nu=1}^{b} \in \mathbb{R}^+$. Thus, we have

$$\mathsf{M}_{k,\ell}^{j} = \mathsf{E} \int \mathsf{g}^k x^\ell \prod_{\nu=1}^{b} \mathrm{dF}(z_\nu) \mathrm{D}z_0 \tag{34}$$

where $\mathsf{g} := g_{\text{map}}[(y); \lambda^{\text{s}}, u]$ with $y = x + \sum_{\nu=0}^{b} \sqrt{\lambda_\nu^{\text{s}}} z_\nu$, and $\mathrm{dF}(z_\nu) = \Lambda_\nu \mathrm{D}z_\nu$. For $\nu \in [1:b]$, Λ_ν is a function of x and $\{z_\zeta\}_{\zeta=0}^{\nu}$. Due to the page limitations, we leave the expressions of λ^{s}, $\{\lambda_\nu^{\text{s}}\}_{\nu=0}^{b}$ and $\{\Lambda_\nu\}_{\nu=1}^{b}$ for the extended version.

Proposition 4 *Let the assumptions in Proposition 1, and the bRSB assumption hold; moreover, consider the single user system in Figure 3. Then, for $j \in [1:n]$,* $p_{\hat{x}|x}^{j}$ *as defined in* (9) *describes the conditional distribution of $\hat{x}$ given x in Figure 3 where* $g_{\text{map}}[(\cdot); \lambda^{\text{s}}, u]$ *is a single user MAP estimator defined in* (23), $p_{z_0|x}(z_0|x) = \pi(z_0)$, *and*

$$p_{z_\nu|x,\{z_\zeta\}_{\zeta=0}^{\nu-1}}(z_\nu|x, \{z_\zeta\}_{\zeta=0}^{\nu-1}) = \Lambda_\nu \pi(z_\nu) \tag{35}$$

for $\nu \in [1:b]$. The factor Λ_ν depends on x and $\{z_\zeta\}_{\zeta=0}^{\nu}$, and the coefficients λ^{s} and $\{\lambda_\nu^{\text{s}}\}_{\nu=0}^{b}$ are coupled due to a set of fixed point equations and bounded as $b \uparrow \infty$.

Considering the bRSB ansatz, one concludes that the ansatz extends the decoupled system in Figure 2 by approximating the coupling interference with more taps. The approximation, however, stops to improve at some step b^*, if $\Lambda_\nu = 1$ for any integer $\nu > b^*$. The extreme case is when for any $\nu \in [1:b]$ in the bRSB ansatz $\Lambda_\nu = 1$. Here, the random variables $\{z_\nu\}_{\nu=1}^{b}$ in Figure 3 become independent Gaussian, and therefore, the decoupled system reduces to Figure 1. In fact in this case, the bRSB solution, as well as any νRSB ansatz with $\nu \in [1:b]$,

reduces to the RS ansatz. Thus, one can consider the decoupled system under the RS ansatz to be a special case of the more general decoupled system given in Figure 3.

V. Conclusion

Decoupling seems to be a generic property of MAP estimators, as Proposition 1 justifies it for any source distribution and a wide range of matrix ensembles. The validity of the result relies only on replica continuity; however, the equivalent single user system depends on the structure of the replica correlation matrix. Recent results in statistical mechanics have shown that failures in finding the exact solution via the replica method are mainly caused by the assumed structure, and not replica continuity. Inspired by the Sherrington-Kirkpatrick model of spin glasses, for which the ∞RSB ansatz has been proved to be correct, one may consider Figure 3 to be the general decoupled system as $b \uparrow \infty$. However, in many cases an accurate approximation might be provided by a finite number of RSB steps. An extreme case is the RS ansatz where all the interference terms in the RSB decoupled system become independent and Gaussian. Thus, one concludes that the previous results in the literature were both special and extreme cases of the RSB decoupled system. The RSB decoupled system raises several issues which require further investigations. For example, nothing is known about the distance between the conditional distributions of the interference terms and independent Gaussian distributions in probability space. The distance variation w.r.t. the number of interference taps can then describe the improvement caused by increasing the number of RSB steps.

References

[1] D. Guo, S. Verdú, and L. K. Rasmussen, "Asymptotic normality of linear multiuser receiver outputs," *Information Theory, IEEE Transactions on*, vol. 48, no. 12, pp. 3080–3095, 2002.

[2] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, p. 965, 1975.

[3] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *Information Theory, IEEE Transactions on*, vol. 48, no. 11, pp. 2888–2910, 2002.

[4] R. R. Müller and W. H. Gerstacker, "On the capacity loss due to separation of detection and decoding," *Information Theory, IEEE Transactions on*, vol. 50, no. 8, pp. 1769–1778, 2004.

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

[6] S. Rangan, A. K. Fletcher, and V. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," in *IEEE Trans. on Inf. Theory*, 2012, pp. 1902–1923.

[7] A. M. Tulino, G. Caire, S. Verdu, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *Information Theory, IEEE Transactions on*, vol. 59, no. 7, pp. 4243–4271, 2013.

[8] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on ℓ_p-norm minimization," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 09, p. L09003, 2009.

[9] B. M. Zaidel, R. Müller, A. L. Moustakas, and R. de Miguel, "Vector precoding for gaussian MIMO broadcast channels: Impact of replica symmetry breaking," *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1413–1440, 2012.

[10] N. I. Akhiezer, *The classical moment problem: and some related questions in analysis*. Oliver & Boyd, 1965, vol. 5.

[11] G. Parisi, "A sequence of approximated solutions to the SK model for spin glasses," *Journal of Physics A: Mathematical and General*, vol. 13, no. 4, p. L115, 1980.