

# An integral formula for large random rectangular matrices and its application to analysis of linear vector channels

Yoshiyuki Kabashima

Department of Computational Intelligence and Systems Science  
Tokyo Institute of Technology, Yokohama 226-8502, Japan  
Email: kaba@dis.titech.ac.jp

**Abstract**—A statistical mechanical framework for analyzing random linear vector channels is presented in a large system limit. The framework is based on the assumptions that the left and right singular value bases of the rectangular channel matrix  $H$  are generated independently from uniform distributions over Haar measures and the eigenvalues of  $H^T H$  asymptotically follow a certain specific distribution. These assumptions make it possible to characterize the communication performance of the channel utilizing an integral formula with respect to  $H$ , which is analogous to the one introduced by Marinari *et. al.* in *J. Phys. A* 27, 7647 (1994) for large random square (symmetric) matrices. A computationally feasible algorithm for approximately decoding received signals based on the integral formula is also provided.

## I. INTRODUCTION

In a general scenario for linear vector channels, multiple messages are transmitted to the receiver, being linearly transformed to multiple output signals by a random matrix and degraded by channel noises. This yields a complicated dependence on message variables, which ensures that the problem of inferring the transmitted messages from the received output signals is non-trivial. In general, inference problems of this kind can be mapped to virtual magnetic systems governed by random interactions [1]. This similarity has promoted a sequence of statistical mechanical analyses of linear vector channels in a large system limit from the beginning of this century [2], [3], [4], [5], [6], [7], [8].

In the simplest analysis, each entry of the channel matrix is regarded as an independent and identically distributed (IID) random variable. However, such a treatment is not necessarily adequate for describing realistic systems, in which non-negligible statistical correlations across the matrix entries are created by spatial/time proximity of messages/antennas or matrix design for enhancement of communication performance. Therefore, the development of methodologies that can deal with correlations in the channel matrix is of great importance to research in the area of linear vector channels.

It is intended that the present article should contribute such a methodology for application to these communication channels. More precisely, we develop a statistical mechanical framework for analyzing linear vector channels so that the influence of the correlations across the matrix entries can be taken into account. The developed framework is applicable not only to Gaussian channels of Gaussian inputs [9], but also general

memory-less channels of continuous/discrete inputs, which are characterized by a factorizable prior distribution.

This article is organized as follows: In the next section, the model of linear vector channels that we focus on herein is defined. In section III, which is the main part of the current article, an integral formula with respect to large random rectangular matrices is introduced. A scheme to assess the performance of the linear vector channel and a computationally feasible approximate decoding algorithm are developed on the basis of this formula. The utility of the developed schemes is examined in section IV by application to an example system. The final section summarizes the present study's findings.

## II. MODEL DEFINITION

For simplicity, we here assume that all the variables relevant to the communication are real; but extending the following framework to complex variables is straightforward [10]. Let us suppose a linear vector channel in which an input message vector of  $K$  components,  $\mathbf{x} = (x_k)$ , is linearly transformed to an  $M$  dimensional sequence,  $\Delta = (\Delta_\mu)$ , by a  $K \times N$  channel matrix,  $\mathbf{H} = (H_{\mu k})$ , as  $\Delta = \mathbf{H}\mathbf{x}$ . For generality and simplicity, we assume a general memory-less channel, which implies that an  $N$  dimensional output signal vector,  $\mathbf{y} = (y_\mu)$ , follows a certain factorizable conditional distribution as

$$P(\mathbf{y}|\mathbf{x}; \mathbf{H}) = P(\mathbf{y}|\mathbf{H}\mathbf{x}) = \prod_{\mu=1}^N P(y_\mu|\Delta_\mu). \quad (1)$$

In addition, we assume a factorizable prior distribution

$$P(\mathbf{x}) = \prod_{k=1}^K P(x_k), \quad (2)$$

for  $\mathbf{x}$ , which may be continuous or discrete.

An expression of the singular value decomposition of  $\mathbf{H}$

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \quad (3)$$

is the basis of our framework. Here, the superscript T denotes the transpose of the matrix to which it is attached,  $\mathbf{D} = \text{diag}(d_k)$  is an  $N \times K$  diagonal matrix composed of the singular values  $d_k$  ( $k = 1, 2, \dots, \min(N, K)$ ), where  $\min(N, K)$  denotes the lesser value of  $N$  and  $K$ . The values  $d_k$  are linked to the eigenvalues of  $\mathbf{H}^T \mathbf{H}$ ,  $\lambda_k$ , as  $\lambda_k = d_k^2$  for

$k = 1, 2, \dots, \min(N, K)$ .  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal matrices of order  $N \times N$  and  $K \times K$ , respectively. In order to handle correlations in  $\mathbf{H}$  analytically, we assume that  $\mathbf{U}$  and  $\mathbf{V}$  are independently generated from uniform distributions of the Haar measures of  $N \times N$  and  $K \times K$  orthogonal matrices, respectively, and that the empirical eigenvalue distribution of  $\mathbf{H}^T \mathbf{H}$ ,  $K^{-1} \sum_{k=1}^K \delta(\lambda - \lambda_k) = (1 - \min(N, K)/K) \delta(\lambda) + K^{-1} \sum_{k=1}^K \delta(\lambda - d_k^2)$  converges to a certain specific distribution  $\rho(\lambda)$  in the limit as  $N$  and  $K$  tend to infinity while keeping the load  $\beta = K/N \sim O(1)$ . Controlling  $\rho(\lambda)$  allows us to express various second-order correlations in  $\mathbf{H}$ .

### III. ANALYSIS

#### A. An integral formula for large random rectangular matrices

With knowledge of  $\mathbf{H}$ , the receiver decodes  $\mathbf{y}$  in order to infer  $\mathbf{x}$ , which is performed on the basis of the Bayes formula

$$P(\mathbf{x}|\mathbf{y}; \mathbf{H}) = \frac{\prod_{\mu=1}^N P(y_\mu|\Delta_\mu) \prod_{k=1}^K P(x_k)}{P(\mathbf{y}; \mathbf{H})}. \quad (4)$$

Here, the probability

$$P(\mathbf{y}; \mathbf{H}) = \text{Tr}_{\mathbf{x}} \prod_{\mu=1}^N P(y_\mu|\Delta_\mu) \prod_{k=1}^K P(x_k), \quad (5)$$

expresses the marginal probability with respect to  $\mathbf{y}$ , where  $\text{Tr}_{\mathbf{x}}$  denotes summation or integration over the all possible states of  $\mathbf{x}$ . Eq. (5) also serves as the partition function concerning the message vector  $\mathbf{x}$  in statistical mechanics.

Let us examine statistical properties of Eq. (5) prior to analyzing Eq. (4). The expression

$$P(\mathbf{y}; \mathbf{H}) = \text{Tr}_{\mathbf{u}, \mathbf{x}} \exp(i\mathbf{u}^T \mathbf{H} \mathbf{x}) \prod_{\mu=1}^N \hat{P}(y_\mu|u_\mu) \prod_{k=1}^K P(x_k), \quad (6)$$

is useful for this purpose, where  $\mathbf{i} = \sqrt{-1}$ ,  $\mathbf{u} = (u_\mu)$  and  $\hat{P}(y_\mu|u_\mu) = (2\pi)^{-1} \int d\Delta_\mu \exp(-iu_\mu \Delta_\mu) P(y_\mu|\Delta_\mu)$  denotes the Fourier transformation of likelihood  $P(y_\mu|\Delta_\mu)$ . We substitute  $\mathbf{H}$  in Eq. (6) by Eq. (3) and take an average with respect to  $\mathbf{U}$  and  $\mathbf{V}$ . For this assessment, it is noteworthy that for any fixed set of  $\mathbf{u}$  and  $\mathbf{x}$ ,  $\tilde{\mathbf{u}} = \mathbf{U}^T \mathbf{u}$  and  $\tilde{\mathbf{x}} = \mathbf{V}^T \mathbf{x}$  behave as continuous random variables that satisfy strict constraints  $N^{-1}|\tilde{\mathbf{u}}|^2 = N^{-1}|\mathbf{u}|^2 = T_u$  and  $K^{-1}|\tilde{\mathbf{x}}|^2 = K^{-1}|\mathbf{x}|^2 = T_x$ . In the limit  $N, K \rightarrow \infty$  keeping  $\beta = K/N \sim O(1)$ , which we will hereafter assume if necessary, this yields an expression

$$\frac{1}{N} \ln \left( \overline{\exp(i\mathbf{u}^T \mathbf{H} \mathbf{x})} \right) = F(T_x, T_u), \quad (7)$$

where  $\overline{\dots}$  denotes the average with respect to  $\mathbf{U}$  and  $\mathbf{V}$ , and

$$F(\xi, \eta) = \text{Extr}_{\Lambda_\xi, \Lambda_\eta} \left\{ -\frac{\beta}{2} \langle \ln(\Lambda_\xi \Lambda_\eta + \lambda) \rangle_\rho - \frac{1-\beta}{2} \ln \Lambda_\xi + \frac{\beta \Lambda_\xi \xi}{2} + \frac{\Lambda_\eta \eta}{2} \right\} - \frac{\beta}{2} \ln \xi - \frac{1}{2} \ln \eta - \frac{1+\beta}{2}, \quad (8)$$

where  $\langle \dots \rangle_\rho$  denotes the average with respect to  $\rho(\lambda)$ , while  $\text{Extr}_\theta \{ \dots \}$  represents extremization with respect to  $\theta$  [11]. This formula is analogous to the one known for ensembles of random square (symmetric) matrices [12], [13], [14], which

is closely related to the  $R$ -transformation developed in free probability theory [15], [9], [16]. Several integral formulae for large random matrices related to Eq. (7), but for different large system limits, are presented in [17].

Eq. (7) implies

$$\frac{1}{N} \ln \left( \overline{\text{Tr}_{\mathbf{y}} P(\mathbf{y}; \mathbf{H})} \right) = \text{Extr}_{T_x, T_u} \{ F(T_x, T_u) + \beta A_x(T_x) + A_u(T_u) \}, \quad (9)$$

where  $A_x(T_x) = \text{Extr}_{\hat{T}_x} \left\{ \hat{T}_x T_x / 2 + \ln \left( \text{Tr}_x P(x) e^{-\hat{T}_x x^2 / 2} \right) \right\}$  and  $A_u(T_u) = \text{Extr}_{\hat{T}_u} \left\{ \hat{T}_u T_u / 2 + \ln \left( \text{Tr}_{y,u} \hat{P}(y|u) e^{-\hat{T}_u u^2 / 2} \right) \right\}$ .

The normalization constraint  $\text{Tr}_{\mathbf{y}} P(\mathbf{y}; \mathbf{H}) = 1$ , in conjunction with the extremization in Eq. (9), yields  $T_x = \text{Tr}_x x^2 P(x)$ ,  $\hat{T}_x = 0$ ,  $T_u = 0$  and  $\hat{T}_u = \beta \langle \lambda \rangle_\rho T_x$ . The physical implication of these results is that components of  $\Delta = \mathbf{H} \mathbf{x}$  behave as IID Gaussian variables of zero mean and variance  $\hat{T}_u$  in the large system limit when  $\mathbf{x}$  is drawn from Eq. (2), while  $\mathbf{U}$  and  $\mathbf{V}$  are independently generated from the Haar measures.

#### B. Performance assessment

Now, we are ready to analyze the typical communication performance of the current channel model. This is performed by assessing the typical mutual information (per output signal) between  $\mathbf{x}$  and  $\mathbf{y}$ ,  $I(X, Y)$ , based on Eqs. (4) and (5) as

$$I(X, Y) = \frac{1}{N} \text{Tr}_{\mathbf{y}, \mathbf{x}} P(\mathbf{y}|\mathbf{x}; \mathbf{H}) P(\mathbf{x}) \ln \left( \frac{P(\mathbf{y}|\mathbf{x}; \mathbf{H})}{P(\mathbf{y}; \mathbf{H})} \right) = \mathcal{F} + \text{Tr}_{\mathbf{y}} \int Dz P \left( y | \sqrt{\hat{T}_u} z \right) \ln P \left( y | \sqrt{\hat{T}_u} z \right), \quad (10)$$

where

$$\mathcal{F} = -\frac{1}{N} \text{Tr}_{\mathbf{y}} \overline{P(\mathbf{y}; \mathbf{H}) \ln P(\mathbf{y}; \mathbf{H})}, \quad (11)$$

represents the conditional entropy of  $\mathbf{y}$ , and serves as the average free energy with respect to  $\mathbf{x}$ .  $Dz = (2\pi)^{-1/2} dz e^{-z^2/2}$  denotes the Gaussian measure. The statistical properties of  $\Delta$  evaluated in the last paragraph are employed to assess the second term on the right-hand side of the last line of Eq. (10).

$\mathcal{F}$  can be evaluated by means of the replica method. Namely, we evaluate the  $n$ -th moment of the partition function  $P(\mathbf{y}; \mathbf{H})$  for  $n \in \mathbb{N}$  as

$$\text{Tr}_{\mathbf{y}} \overline{P^{n+1}(\mathbf{y}; \mathbf{H})} = \text{Tr}_{\mathbf{y}, \{\mathbf{x}^a\}, \{\mathbf{u}^a\}} \exp \left( \mathbf{i} \sum_{a=1}^n (\mathbf{u}^a)^T \mathbf{H} \mathbf{x}^a \right) \times \prod_{a=1}^{n+1} \prod_{\mu=1}^N \hat{P}(y_\mu | u_\mu^a) \times \prod_{a=1}^{n+1} \prod_{k=1}^K P(x_k^a), \quad (12)$$

and assess  $\mathcal{F}$  as

$$\mathcal{F} = -\lim_{n \rightarrow 0} \frac{\partial}{\partial n} \frac{1}{N} \ln \left( \text{Tr}_{\mathbf{y}} \overline{P^{n+1}(\mathbf{y}; \mathbf{H})} \right), \quad (13)$$

analytically continuing expressions obtained for Eq. (12) from  $n \in \mathbb{N}$  to  $n \in \mathbb{R}$ . Here,  $\{\mathbf{x}^a\}$  denotes a set of  $n+1$  replicated vectors  $\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^n$ , with  $\{\mathbf{u}^a\}$  defined similarly.

Eq. (13) is generally expressed using  $F(\xi, \eta)$ , and the derivation of the expression can be found in [11]. In particular, the expression obtained under the replica symmetric ansatz, which is believed to be correct for the current case since the inference is performed on the basis of the correct posterior (4) [18], is given in a compact form as

$$\mathcal{F} = -\text{Extr}_{q_x, q_u} \{ \mathcal{A}_{xu}(q_x, q_u) + \beta \mathcal{A}_x(q_x) + \mathcal{A}_u(q_u) \}, \quad (14)$$

where

$$\mathcal{A}_{xu}(q_x, q_u) = F(T_x - q_x, q_u) + \frac{\hat{T}_u q_u}{2}, \quad (15)$$

$$\mathcal{A}_x(q_x) = \text{Extr}_{\hat{q}_x} \left\{ -\frac{\hat{q}_x q_x}{2} + \int Dz \mathcal{P}(z; \hat{q}_x) \ln \mathcal{P}(z; \hat{q}_x) \right\}, \quad (16)$$

and

$$\begin{aligned} \mathcal{A}_u(q_u) = & \text{Extr}_{\hat{q}_u} \left\{ -\frac{\hat{q}_u q_u}{2} \right. \\ & \left. + \text{Tr}_y \int Dz \mathcal{P}(y|z; \hat{q}_u) \ln \mathcal{P}(y|z; \hat{q}_u) \right\}, \quad (17) \end{aligned}$$

in which  $\mathcal{P}(z; \hat{q}_x) = \text{Tr}_x P(x) e^{-\hat{q}_x x^2/2 + \sqrt{\hat{q}_x} z x}$  and  $\mathcal{P}(y|z; \hat{q}_u) = \int Ds P(y|\sqrt{\hat{T}_u - \hat{q}_u} s + \sqrt{\hat{q}_u} z)$ . The  $q_x$  and  $q_u$  determined by Eq. (14) represent  $K^{-1} [|\langle \mathbf{x} \rangle|^2]$  and  $-N^{-1} [|\langle \mathbf{u} \rangle|^2]$ , respectively, where  $\langle \dots \rangle$  denotes averaging over the posterior distribution (4) while  $[\dots]$  indicates the average with respect to  $\mathbf{y}$ ,  $\mathbf{U}$  and  $\mathbf{V}$ . These averages,  $\langle \dots \rangle$  and  $[\dots]$ , correspond to the thermal and quenched averages in statistical mechanics, respectively. The quantities  $\hat{q}_x$  and  $\hat{q}_u$  appearing in Eqs. (16) and (17) can be used for assessing performance measures other than Eq. (10), such as the mean square error (MSE) and the bit error rate (BER).

### C. Computationally feasible approximate decoding

Let us suppose a situation which requires evaluation of the posterior average

$$\mathbf{m}_x = \text{Tr}_{\mathbf{x}} \mathbf{x} P(\mathbf{x}|\mathbf{y}; \mathbf{H}), \quad (18)$$

where  $\mathbf{m}_x = (m_{xk})$ , with similar notation used for other vectors below. Eq. (18) serves as the estimator that minimizes the MSE in general, and can be used to minimize the BER for binary messages. Exact assessment of such averages is, however, computationally difficult for large systems, which motivates us to develop computationally feasible approximation algorithms [19], [3], [20]. A generalized Gibbs free energy

$$\begin{aligned} \tilde{\Phi}(\mathbf{m}_x, \mathbf{m}_u; l) = & \text{Extr}_{\mathbf{h}_x, \mathbf{h}_u} \{ \mathbf{h}_x \cdot \mathbf{m}_x + \mathbf{h}_u \cdot \mathbf{m}_u \\ & - \ln(Z(\mathbf{h}_x, \mathbf{h}_u; l)) \}, \quad (19) \end{aligned}$$

where  $Z(\mathbf{h}_x, \mathbf{h}_u; l) = \text{Tr}_{\mathbf{x}, \mathbf{u}} \prod_{\mu=1}^N \hat{P}(y_\mu|u_\mu) \times \prod_{k=1}^K P(x_k) \times \exp(\mathbf{h}_x \cdot \mathbf{x} + \mathbf{h}_u \cdot (\mathbf{i}\mathbf{u}) + (\mathbf{i}\mathbf{u})^T (l\mathbf{H})\mathbf{x})$ , offers a useful basis for this purpose since Eq. (18) is characterized as the unique saddle point of Eq. (19) for  $l = 1$  [21], [22].

### MPforPerceptron{

Perform **Initialization**;

Iterate **H-Step** and **V-Step** alternately sufficient times;

}

### Initialization{

$$\chi_x \leftarrow \frac{1}{K} \sum_{k=1}^K x_k^2 P(x_k); \quad \hat{\chi}_x \leftarrow 0; \quad \Lambda_x \leftarrow \frac{1}{\chi_x} - \hat{\chi}_x;$$

$$m_{xk} \leftarrow \text{Tr}_{x_k} x_k P(x_k) \quad (k = 1, 2, \dots, K);$$

$$\mathbf{h}_u \leftarrow \mathbf{H} \mathbf{m}_x; \quad \mathbf{m}_u \leftarrow \mathbf{0};$$

}

### H-Step{

Search  $(\chi_u, \Lambda_u)$  for given  $(\chi_x, \Lambda_x)$  to satisfy conditions

$$\chi_x = \left\langle \frac{\Lambda_u}{\Lambda_x \Lambda_u + \lambda} \right\rangle_\rho \quad \text{and} \quad \chi_u = \frac{1 - \beta}{\Lambda_u} + \left\langle \frac{\beta \Lambda_x}{\Lambda_x \Lambda_u + \lambda} \right\rangle_\rho;$$

$$\hat{\chi}_u \leftarrow \frac{1}{\chi_u} - \Lambda_u;$$

$$\mathbf{h}_u \leftarrow \mathbf{h}_u - \hat{\chi}_u \mathbf{m}_u;$$

$$m_{u\mu} \leftarrow \frac{\partial}{\partial h_{u\mu}} \ln \left( \int Dx P(y_\mu | \sqrt{\hat{\chi}_u} x + h_{u\mu}) \right) \quad (\mu = 1, 2, \dots, N);$$

$$\mathbf{h}_x \leftarrow \mathbf{H}^T \mathbf{m}_u;$$

$$\chi_u \leftarrow -\frac{1}{N} \sum_{\mu=1}^N \frac{\partial^2}{\partial h_{u\mu}^2} \ln \left( \int Ds P(y_\mu | \sqrt{\hat{\chi}_u} s + h_{u\mu}) \right);$$

$$\Lambda_u \leftarrow \frac{1}{\chi_u} - \hat{\chi}_u;$$

}

### V-Step{

Search  $(\chi_x, \Lambda_x)$  for given  $(\chi_u, \Lambda_u)$  to satisfy conditions

$$\chi_x = \left\langle \frac{\Lambda_u}{\Lambda_x \Lambda_u + \lambda} \right\rangle_\rho \quad \text{and} \quad \chi_u = \frac{1 - \beta}{\Lambda_u} + \left\langle \frac{\beta \Lambda_x}{\Lambda_x \Lambda_u + \lambda} \right\rangle_\rho;$$

$$\hat{\chi}_x \leftarrow \frac{1}{\chi_x} - \Lambda_x;$$

$$\mathbf{h}_x \leftarrow \mathbf{h}_x + \hat{\chi}_x \mathbf{m}_x;$$

$$m_{xk} \leftarrow \frac{\partial}{\partial h_{xk}} \ln \left( \text{Tr}_x P(x) e^{-\frac{1}{2} \hat{\chi}_x x^2 + h_{xk} x} \right) \quad (k = 1, 2, \dots, K);$$

$$\mathbf{h}_u \leftarrow \mathbf{H} \mathbf{m}_x;$$

$$\chi_x \leftarrow \frac{1}{K} \sum_{k=1}^K \frac{\partial^2}{\partial h_{xk}^2} \ln \left( \text{Tr}_x P(x) e^{-\frac{1}{2} \hat{\chi}_x x^2 + h_{xk} x} \right);$$

$$\Lambda_x \leftarrow \frac{1}{\chi_x} - \hat{\chi}_x;$$

}

Fig. 1. Pseudocode of the message-passing algorithm **MPforPerceptron** [23]. The symbols “;” and “←” represent the end of a command line and the operation of substitution, respectively. The quantities  $\Lambda_\xi$  and  $\Lambda_u$  are the counterparts of  $\Lambda_\xi$  and  $\Lambda_\eta$  in Eq. (8) for  $\xi = \chi_x$  and  $\eta = \chi_u$ , respectively.

Unfortunately, the evaluation of Eq. (19) is also computationally difficult. One approach to overcoming this difficulty is to perform a Taylor expansion around  $l = 0$ , for which  $\tilde{\Phi}(\mathbf{m}_x, \mathbf{m}_u; l)$  can be analytically calculated as an exceptional case, and substitute  $l = 1$  in the expression obtained [21]. However, the evaluation of higher-order terms, which are not negligible in general, requires a complicated calculation in this

expansion, which sometimes prevents the scheme from being practically feasible. In order to avoid such difficulty, we take an alternative approach here, which is inspired by a derivative of Eq. (19),

$$\frac{\partial \tilde{\Phi}(\mathbf{m}_x, \mathbf{m}_u; l)}{\partial l} = -\langle (i\mathbf{u})^T \mathbf{H} \mathbf{x} \rangle_l, \quad (20)$$

following a strategy proposed by Opper and Winther [22]. Here,  $\langle \cdots \rangle_l$  represents the average with respect to the generalized weight  $\prod_{\mu=1}^N \hat{P}(y_\mu | u_\mu) \times \prod_{k=1}^K P(x_k) \times \exp(\mathbf{h}_x \cdot \mathbf{x} + \mathbf{h}_u \cdot (i\mathbf{u}) + (i\mathbf{u})^T (l\mathbf{H})\mathbf{x})$ , in which  $\mathbf{h}_x$  and  $\mathbf{h}_u$  are determined so as to satisfy  $\langle \mathbf{x} \rangle_l = \mathbf{m}_x$  and  $\langle (i\mathbf{u}) \rangle_l = \mathbf{m}_u$ , respectively. The right-hand side of this equation is an average of a quadratic form composed of many random variables. The central limit theorem implies that such an average does not depend on the details of the objective distribution, but is determined only by the values of the first and second moments. In order to construct a simple approximation scheme, let us assume that the second moments are characterized macroscopically by  $\langle |\mathbf{x}|^2 \rangle_l - |\langle \mathbf{x} \rangle_l|^2 = K\chi_x$  and  $\langle |\mathbf{u}|^2 \rangle_l - |\langle \mathbf{u} \rangle_l|^2 = N\chi_u$ . Evaluating the right-hand side of Eq. (20) using a Gaussian distribution, the first and second moments of which are constrained to be identical to those of the generalized weight, and integrating from  $l = 0$  to  $l = 1$ , we have

$$\begin{aligned} & \tilde{\Phi}(\chi_x, \chi_u, \mathbf{m}_x, \mathbf{m}_u; 1) - \tilde{\Phi}(\chi_x, \chi_u, \mathbf{m}_x, \mathbf{m}_u; 0) \\ & \simeq -\mathbf{m}_u^T \mathbf{H} \mathbf{m}_x - NF(\chi_x, \chi_u), \end{aligned} \quad (21)$$

where the function  $F(\xi, \eta)$  is provided as in Eq. (8) by the empirical eigenvalue spectrum of  $\mathbf{H}^T \mathbf{H}$ ,  $\rho(\lambda) = K^{-1} \sum_{k=1}^K \delta(\lambda - \lambda_k)$  and the macroscopic second moments  $\chi_x$  and  $\chi_u$  are included in arguments of the Gibbs free energy because the right-hand side of Eq. (20) depends on these moments. Eq. (21) offers a computationally feasible approximation of Eq. (19) for  $l = 1$ , since assessment of  $\tilde{\Phi}(\chi_x, \chi_u, \mathbf{m}_x, \mathbf{m}_u; 0)$ , in which one can perform summations with respect to relevant variables independently, can be achieved at a reasonable computational cost.

Although evaluation of Eq. (21) is computationally feasible, searching for saddle points of this function within a practical time is still a non-trivial problem. In Fig. 1, we present a message-passing type algorithm, which was recently proposed for a classification problem of single layer perceptrons [23], as a promising heuristic solution for this problem.

The efficacy of this method under appropriate conditions was experimentally confirmed for the perceptron problem, and to the extent to which it has been applied to several ensembles of linear vector channels, this algorithm has also been shown to exhibit a reasonable performance for the current inference task as well. However, its properties including convergence conditions have not yet been fully clarified, and, therefore further investigation is necessary for the theoretical validation and improvement of the performance of this method.

#### IV. EXAMPLE: WELCH BOUND EQUALITY SEQUENCES

In order to demonstrate the utility of the proposed approach, let us apply the current methodologies to the analysis of

the matrix ensemble that is characterized by  $\rho(\lambda) = (1 - \beta^{-1})\delta(\lambda) + \beta^{-1}\delta(\lambda - \beta)$  under the assumption  $\beta > 1$ , which corresponds to the case of Welch bound equality sequences (WBES) [24]. We focus on the case of the Gaussian channel  $P(y|\Delta) = (2\pi\sigma^2)^{-1/2} \exp(-(y - \Delta)^2/(2\sigma^2))$  and binary inputs  $\mathbf{x} \in \{+1, -1\}^K$ , since this constitutes a simple, yet non-trivial problem. Under these assumptions, the developed framework has a higher capability than is required for the assessment of the typical communication performance with respect to the matrix ensemble, which can be carried out by a simpler method developed by the author and his colleagues [25], [10], as was recently shown by Kitagawa and Tanaka [26]. Nevertheless, the framework is still useful as one can derive a computationally feasible approximate decoding algorithm of good convergence properties based on the procedure shown in Fig. 1.

For Gaussian channels,  $\Lambda_u$  in Fig. 1 can be fixed as  $\Lambda_u = \sigma^2$  in general. This yields an algorithm

$$\begin{aligned} m_u^{t+1} &= \frac{1}{\sigma^2 + \hat{\chi}_u^t} (\mathbf{y} - \mathbf{H} \mathbf{m}_x^t + \hat{\chi}_u^t \mathbf{m}_u^t), \quad (22) \\ m_{xk}^{t+1} &= \tanh \left( \sum_{\mu=1}^N H_{\mu k} m_{u\mu}^{t+1} + \hat{\chi}_x^{t+1} m_{xk}^t \right) \quad (23) \\ & \quad (k = 1, 2, \dots, K), \end{aligned}$$

for WBES, where  $t$  denotes the number of iterations.  $\hat{\chi}_u^t$  in Eq. (22) is provided as  $\hat{\chi}_u^t = \beta/\Lambda_x^t$ , where  $\Lambda_x^t$  is determined so as to satisfy  $\chi_x^t = (1 - \beta^{-1})/\Lambda_x^t + \beta^{-1}\sigma^2/(\sigma^2\Lambda_x^t + \beta)$  for given  $\chi_x^t = 1 - K^{-1}|\mathbf{m}_x^t|^2$ . Utilizing the identical  $\Lambda_x^t$ ,  $\hat{\chi}_x^{t+1}$  in Eq. (23) is evaluated as  $\hat{\chi}_x^{t+1} = 1/\chi_x^t - \Lambda_x^t$ .

Fig. 2 compares the BER for the theoretical assessment by the replica method with the experimental evaluation obtained by the algorithm of Eqs. (22) and (23). In the experiments, the estimates of the binary messages are computed as  $\hat{x}_k = \text{sign}(m_{xk})$  for  $k = 1, 2, \dots, K$ , where  $\text{sign}(a) = a/|a|$  for  $a \neq 0$ . This decoding scheme is optimal for minimizing BER if  $\mathbf{m}_x$  represents the correct posterior average (18) [27]. The excellent agreement between the curves and markers in this plot validates both the performance assessment based on the replica method and that based on the developed algorithm. A characteristic feature of Eqs. (22) and (23) is the inclusion of macroscopic variables  $\hat{\chi}_u^t$  and  $\hat{\chi}_x^{t+1}$ , which are expected to act to cancel the self-reactions from previous states. [28]. Fig. 3 plots the influence of this operation, indicating that the cancellation acts to maintain the quality of the converged solution up to larger  $\beta$  under a condition of fixed SNR.

#### V. SUMMARY

In summary, we have developed a framework to analyze linear vector channels in a large system limit. The framework is based on the assumptions that the left and right singular value bases of the channel matrix can be regarded as independently drawn from Haar measures over orthogonal (unitary, if the number field is defined over the complex variables) groups, and that the eigenvalues of the cross correlation matrix of the channel matrix asymptotically approach

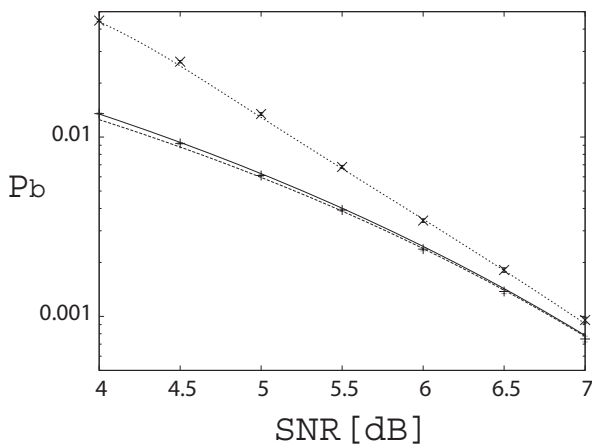


Fig. 2. BER vs. signal-to-noise ratio (SNR) for binary inputs for the case  $\beta = 1.1$ . The SNR plotted on the horizontal axis is given by  $-10 \log_{10}(2\sigma^2)$  while the vertical axis denotes the BER. The curves indicate theoretical predictions, which correspond to the scalar Gaussian channel, WBES and the basic matrix ensemble (BASIC) from the bottom. Sample matrices of BASIC are composed of IID entries of zero mean and  $1/N$  variance Gaussian random variables. Values for WBES and BASIC are assessed by the replica method. The markers indicate experimental estimates of the BER obtained from 500 sample systems with  $K = 2048$  and  $N = 1862$  on the basis of the algorithm shown in Fig. 1. Excellent agreement between the curves and markers validates both the performance analysis based on the replica method and that of the developed approximation algorithm.

a certain specific distribution in the limit of large matrix size. These modeling assumptions allow a characterization of the system in terms of an integral formula in two variables, which is fully determined by the eigenvalue distribution. Upon applying this formula in conjunction with the replica method, we have derived a general expression for the typical mutual information of general memory-less channels with factorizable priors of continuous/discrete inputs. We have further proposed a computationally feasible decoding algorithm based on the formula, and have found that numerical results obtained from this algorithm are in excellent agreement with the theoretical predictions evaluated by the replica method.

Future research directions include the application of the developed framework to various models of linear vector channels, and further improvement of the computationally feasible decoding algorithm.

#### ACKNOWLEDGMENTS

The author thanks Jean-Bernard Zuber for useful discussions concerning Eq. (8). This research was supported in part by Grants-in-Aid MEXT/JSPS, Japan, Nos. 17340116 and 18079006.

#### REFERENCES

- [1] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford: Oxford University Press), 2001.
- [2] T. Tanaka, *Europhys. Lett.* **54**, 540, 2001; *IEEE Trans. on Infor. Theory* **48**, 2888, 2002.
- [3] Y. Kabashima, *J. Phys. A* **36**, 11111, 2003.
- [4] R. R. Müller, *IEEE Trans. on Signal Processing* **51** 2821, 2003.
- [5] A. L. Moustakas *IEEE Trans. on Infor. Theory* **49**, 2545, 2003.
- [6] D. Guo and S. Verdú *IEEE Trans. on Infor. Theory* **51**, 1983, 2005.

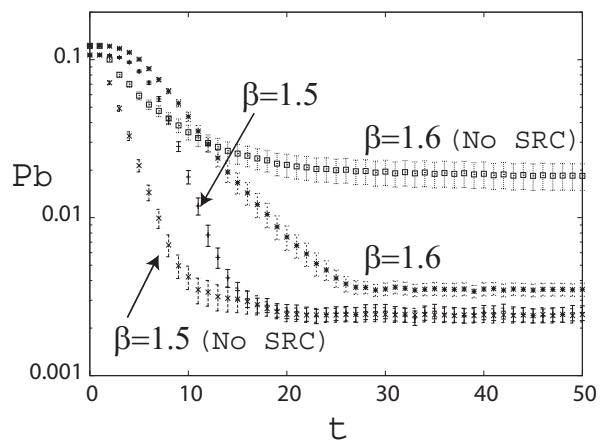


Fig. 3. Influence of the cancellation of self-reactions in the approximate decoding for WBES. Experimentally assessed trajectories are plotted for two approximation algorithms for the cases  $\beta = 1.5$  and  $1.6$ , where SNR is fixed to 6.0. The horizontal axis represents the number of iterations, while the vertical axis denotes the BER. The data are obtained from 100 experiments of  $K = 512$  systems. The first algorithm used in these experiments is that presented by Eqs. (22) and (23) while the second algorithm, the results of which are denoted by “No SRC” in the figure, is found by considering vanishing values of the macroscopic variables  $\tilde{\chi}_u^z$  and  $\tilde{\chi}_x^{z+1}$  in Eqs. (22) and (23). For both algorithms, initial states in the experiments were set as  $m_{xk} = \tanh(\theta_k/\sigma^2)$ , where, in contrast to Fig. 1,  $\theta_k = \sum_{\mu=1}^N H_{\mu k} y_{\mu}$  ( $k = 1, 2, \dots, K$ ). For  $\beta = 1.5$ , both algorithms converge to almost identical solutions, although the convergence of the first algorithm is slower. However, for  $\beta = 1.6$ , the second algorithm converges to solutions of significantly higher BER while solutions found by the first algorithm still exhibit relatively low values of the BER. This indicates that the cancellation of the self-reaction terms in the approximation algorithm acts to maintain the quality of the convergent solutions for larger values of  $\beta$ .

- [7] J. P. Neirotti and D. Saad, *Europhys. Lett.* **71**, 866, 2005.
- [8] A. Montanari and D. Tse, arXiv:cs/0602028, 2006.
- [9] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications* (Hanover, MA: Now Publishers), 2004.
- [10] K. Takeda, A. Hatabu and Y. Kabashima, *J. Phys. A* **40**, 14085, 2007.
- [11] Y. Kabashima, *J. Phys.: Conf. Ser.* **95**, 012001, 2008.
- [12] E. Marinari, G. Parisi and F. Ritort, *J. Phys. A* **27**, 7647, 1994.
- [13] G. Parisi and M. Potters, *J. Phys. A* **28**, 5267, 1995.
- [14] C. Itzykson and J. B. Zuber, *J. Math. Phys.* **21**, 411, 1980.
- [15] D. V. Voiculescu, K. J. Dykema and A. Nica, *Free Random Variables* (Providence, R.I.: American Mathematical Society), 1992.
- [16] T. Tanaka, *J. Phys.: Conf. Ser.* **95**, 012002, 2008.
- [17] P. Zinn-Justin and J. B. Zuber, *J. Phys. A* **36**, 3173, 2003.
- [18] H. Nishimori, *Europhys. Lett.* **57**, 302, 2002.
- [19] M. K. Varanasi and B. Aazhang, *IEEE Trans. on Commun.* **38**, 509, 1990; *ibid.* **39**, 725, 1991.
- [20] T. Tanaka and M. Okada, *IEEE Trans. on Infor. Theory* **51** 700, 2005.
- [21] Y. Iba, *J. Phys. A* **15**, 1971, 1982.
- [22] M. Opper and O. Winther, *Journal of Machine Learning Research* **6**, 2177, 2005.
- [23] T. Shinzato and Y. Kabashima, arXiv:0712.4050, 2007.
- [24] L. R. Welch, *IEEE Trans. on Infor. Theory.* **IT-20**, 397, 1974.
- [25] K. Takeda, K. S. Uda and Y. Kabashima, *Europhys. Lett.* **76**, 1193, 2006.
- [26] K. Kitagawa and T. Tanaka, *Proc. The 30th Symposium on Information Theory and Its Application (SITA2007)* (in Japanese), 799, 2007.
- [27] Y. Iba, *J. Phys. A* **32**, 3875, 1999.
- [28] D. J. Thouless, P. W. Anderson and R. G. Palmer, *Phil. Mag.* **35**, 593, 1977.