Random Sparse Linear Systems Observed Via Arbitrary Channels: A Decoupling Principle

Dongning Guo
Department of Electrical Engineering & Computer Science
Northwestern University
Evanston, IL 60208, USA.

Chih-Chun Wang
Center for Wireless Systems and Applications (CWSA)
School of Electrical & Computer Engineering
Purdue University, West Lafayette, IN 47907, USA.

***Abstract*—This paper studies the problem of estimating the vector input to a sparse linear transformation based on the observation of the output vector through a bank of arbitrary independent channels. The linear transformation is drawn randomly from an ensemble with mild regularity conditions. The central result is a decoupling principle in the large-system limit. That is, the optimal estimation of each individual symbol in the input vector is asymptotically equivalent to estimating the same symbol through a scalar additive Gaussian channel, where the aggregate effect of the interfering symbols is tantamount to a degradation in the signal-to-noise ratio. The degradation is determined from a recursive formula related to the *score function* of the conditional probability distribution of the noisy channel. A sufficient condition is provided for belief propagation (BP) to asymptotically produce the *a posteriori* probability distribution of each input symbol given the output. This paper extends the authors' previous decoupling result for Gaussian channels to arbitrary channels, which was based on an earlier work of Montanari and Tse. Moreover, a rigorous justification is provided for the generalization of some results obtained via statical physics methods.**

I. Introduction

Consider the estimation of a vector input signal $\boldsymbol{X}$ which traverses a linear system[1] $\underline{\boldsymbol{S}}$ in the Euclidean space and is then observed through a noisy channel: $\underline{\boldsymbol{S}}\boldsymbol{X} \mapsto \boldsymbol{Y}$. This simple model is widely used in communications, control, and signal processing. Most work considers the vector Gaussian noise channel: $\boldsymbol{Y} = \underline{\boldsymbol{S}}\boldsymbol{X} + \boldsymbol{N}$. As useful as the Gaussian model is, it does not apply to many interesting applications. One example is the Poisson multiple-access channel, in which $\boldsymbol{Y}$ has Poisson distribution conditioned on $\underline{\boldsymbol{S}}\boldsymbol{X}$. Assuming that $\underline{\boldsymbol{S}}\boldsymbol{X} \mapsto \boldsymbol{Y}$ consists of a bank of arbitrary independent channels, this paper studies the optimal *a posteriori* estimation as well as efficient belief propagation (BP) estimators.

Early work along this direction considers the vector Gaussian channel $\boldsymbol{Y} = \underline{\boldsymbol{S}}\boldsymbol{X} + \boldsymbol{N}$ with linear minimum mean-square error (MMSE) estimation. In case of a large, randomly generated linear transformation $\underline{\boldsymbol{S}}$, the mean-squared error can be computed using random matrix theory. For the same channel, optimal (nonlinear) estimation of discrete inputs is a discrete optimization problem and random matrix theory is not applicable. A breakthrough in performance analysis was made by Tanaka using the *replica method* [1], a non-rigorous technique commonly adopted in statistical physics. Guo and Verdú subsequently generalized Tanaka's results to arbitrary input and a family of detectors [2]. It is claimed that for vector Gaussian channels, the optimal estimate for each individual symbol is of identical quality as the estimate of the same symbol through a scalar Gaussian channel independent of all other symbols [2]. This result is referred to as the *decoupling principle* for vector Gaussian channels. The corresponding signal-to-noise ratio (SNR) of each equivalent scalar Gaussian channel, or rather, the degradation in SNR for each symbol termed the *efficiency*, is determined by a fixed-point equation. Unfortunately, the replica method relies on intractable assumptions and the results in [1] and [2] are subject to doubt.

In recent years, estimators based on BP and its approximation have received great attention [3]–[5]. By comparing BP and maximum *a posteriori* (MAP) detection, Montanari and Tse [6] justified Tanaka's results for the first time in the special case of "sparse" spreading matrix $\underline{\boldsymbol{S}}$ with a relatively small system load. Their result supports the conjecture from several works that the large-system performance of the MAP detector is determined by the solutions of the fixed-point equations for efficiency in [1], [2].

Previously, under essentially the same setting of large sparse linear systems and vector Gaussian channels, the authors [7], [8] have generalized the results of [6] to arbitrary *a priori* distributions of the input $\boldsymbol{X}$ and the linear transformation $\underline{\boldsymbol{S}}$ using the generalized density evolution [9], [10]. Both the decoupling principle and the fixed-point characterization in [2] are rigorously proved, and a sandwiching argument is used to obtain the strongest, posterior-probability-based characterization of the estimation of individual symbols X_k and linearly transformed symbols $(\underline{\boldsymbol{S}}\boldsymbol{X})_l$.

This paper generalizes the above vector Gaussian channel results to arbitrary channels characterized by their conditional probabilities. Remarkably, the decoupling into scalar Gaussian channels still holds. In particular, the estimation of any individual symbol using BP is equivalent to estimating the same symbol through a scalar additive Gaussian noise channel even though the vector channel $\underline{\boldsymbol{S}}\boldsymbol{X} \mapsto \boldsymbol{Y}$ is neither additive nor Gaussian. The SNR degradation in the

This research was in part supported by the National Science Foundation under Grant CCF-0644344 and DARPA under Grant W911NF-07-1-0028.

[1]This paper adopts the following notational convention unless noted otherwise. Deterministic and random variables are denoted by lowercase and uppercase letters respectively, and scalars, vectors and matrices are distinguished using normal, bold and underlined bold fonts respectively.

equivalent scalar Gaussian channel, which can be regarded as the *generalized efficiency* of the underlying system, is determined by the system load and the *score function* of the channel conditional probability distribution through a double recursion. If the system load is sufficiently small such that the fixed point of the recursive formula is unique, the strongest posterior-probability-equivalency between BP and *a posteriori* estimation can be established by the sandwiching argument. Interestingly, the results in this work are consistent with the results developed using the replica method in [11].

II. Model and Formulation

A. Linear System

Let $\boldsymbol{X} = [X_1, \dots, X_K]^\top$ denote the input vector to a linear transformation characterized by a (random) $L \times K$ matrix $\underline{\boldsymbol{S}}$. The output of the transformation, $\underline{\boldsymbol{S}}\boldsymbol{X}$, is observed through arbitrary parallel independent noisy channels of the same type:

$$p_{\boldsymbol{Y}|\boldsymbol{X},\underline{\boldsymbol{S}}}(\boldsymbol{y}|\boldsymbol{x},\underline{\boldsymbol{s}}) = \prod_{l=1}^{L} f\left(y_l \middle| (\underline{\boldsymbol{s}}\boldsymbol{x})_l\right) \tag{1}$$

where $f(y|w) = p_{Y|W}(y|w)$ denotes the conditional probability density function of each scalar channel. A familiar special case of the system with f being conditionally Gaussian is described by

$$\boldsymbol{Y} = \underline{\boldsymbol{S}}\boldsymbol{X} + \boldsymbol{N} \tag{2}$$

where $\boldsymbol{N}$ is a Gaussian vector. Indeed, the model (1) can be regarded as a generalization of (2) to arbitrary channels.

Consider any realization of $\underline{\boldsymbol{S}} = \underline{\boldsymbol{s}}$. Following the convention in graphical modeling of communication systems, a bipartite factor graph of the system (1) is illustrated in Figure 1, where each pair of symbol node X_k and chip[2] node Y_l are connected by an edge if $s_{lk} \neq 0$.

The task of the estimator is to infer the scalar value of X_k for some k, given the observation $\boldsymbol{Y}$, where the channel characteristics $f(\cdot|\cdot)$, the input distribution $P_{\boldsymbol{X}}$, and the realization of the matrix $\underline{\boldsymbol{S}} = \underline{\boldsymbol{s}}$ are known to the estimator.

We assume that the symbols $\{X_k\}$ are independent and identically distributed (i.i.d.) and take values in the alphabet $\chi \subset \mathbb{R}$, which may be discrete or continuous. Let P_X denote the cumulative distribution function (cdf) of X_k, which is of zero mean and finite variance. The k-th column of $\underline{\boldsymbol{S}}$ is denoted by $\boldsymbol{S}_k = \frac{1}{\sqrt{\Lambda_k}}[S_{1k}, S_{2k}, \dots, S_{Lk}]^\top$ and $\sqrt{\Lambda_k}$ is a normalization factor.

B. Random Ensemble and Large-System Limit

The ensemble of linear transformations is described in the following. First, an $L \times K$ binary incidence matrix $\underline{\boldsymbol{H}} = (H_{lk})$ is randomly picked from a certain ensemble to be described shortly. For all (l,k) with $H_{lk} = 0$, set $S_{lk} = 0$. For all (l,k) with $H_{lk} = 1$, S_{lk} are i.i.d. and equally likely to be ± 1. The normalization factor for each spreading sequence $\boldsymbol{s}_k$ is $1/\sqrt{\Lambda_k}$, where $\Lambda_k = \sum_{l=1}^{L} H_{lk}$ is the symbol degree of X_k. Let $\Gamma_l := \sum_{k=1}^{K} H_{lk}$ denote the chip degree of Y_l and $\bar{\Gamma} := \frac{1}{L}\sum_{l=1}^{L} \Gamma_l$ denote the average chip degree.

[2]Here we use a term originated in CDMA.

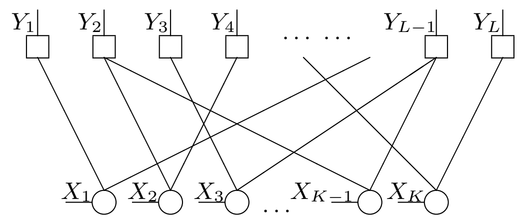

Fig. 1. The Forney-style factor graph for the sparse linear system. The square and the circle correspond to the factorizations associated with the noisy observation $p_{Y|W}(\cdot|\cdot)$ and the vector repetition $\boldsymbol{s}_k X_k$ (of X_k) respectively.

For the random matrix ensemble of $\underline{\boldsymbol{H}}$, the doubly Poisson ensemble is used as an illustrative example in this work, for which each entry H_{lk} is i.i.d. Bernoulli with $\mathsf{P}\{H_{lk} = 1\} = \bar{\Gamma}/K$. This paper considers the *large-sparse-system limit*, in which $K, L, \bar{\Gamma} \to \infty$ with $K/L \to \beta < \infty$ and $\bar{\Gamma} = o(K^{1/2})$, the last condition of which ensures that the bipartite graph of $\underline{\boldsymbol{H}}$ (see Figure 1) is free of cycles of length smaller than any given number in probability.

More detailed description of the ensemble of interest can be found in [8]. It is worth noting that the results in this paper can be easily extended to accommodate more general ensembles [12], as demonstrated in [8], for which some other regular conditions, in addition to the asymptotic short-cycle-free property, are required, including the *chip-semi-regularity* and the *balanced-symbol-degree* conditions.

As a final remark, the input distribution P_X and the system load β are fixed system parameters and do not change with respect to K, L, and $\bar{\Gamma}$.

III. Main Results

Let us introduce the canonical scalar Gaussian channel:

$$Z = \sqrt{g}X + N \tag{3}$$

where $X \sim P_X$ and $N \sim \mathcal{N}(0,1)$ are independent, and g denotes the gain of the channel in SNR. Throughout this paper, we use $P_{X|Z;g}$ to denote the cdf of the posterior distribution of the input X given Z, according to the Gaussian model (3), which is parameterized by g.

Consider first the problem of estimating an individual symbol X_k given all observed chips in its supporting tree of depth $2t$, denoted by $\boldsymbol{Y}^{(t)}$. Precisely, $\boldsymbol{Y}^{(t)}$ consists of all chips Y_l within distance $2t-1$ to X_k on the factor graph. A key result in this paper states that the posterior of X_k given $\boldsymbol{Y}^{(t)}$ essentially converges to the posterior of the scalar Gaussian channel, as the size of the linear system increases.

Theorem 1: For every k and x where $P_X(x)$ is continuous,

$$P_{X_k|\boldsymbol{Y}^{(t)},\underline{\boldsymbol{S}}}\left(x \middle| \boldsymbol{Y}^{(t)}, \underline{\boldsymbol{S}}\right) \to P_{X|Z;\eta^{(t)}}\left(x \middle| h\left(\boldsymbol{Y}^{(t)}\right)\right) \tag{4}$$

in probability in the large-sparse-system limit, where $\eta^{(t)}$ is some positive number, and $h(\cdot)$ is some function such that, conditioned on $X_k = a$, $h\left(\boldsymbol{Y}^{(t)}\right) \sim \mathcal{N}\left(a\sqrt{\eta^{(t)}}, 1\right)$.

Theorem 1 states that the problem of estimating each individual symbol X_k from $\boldsymbol{Y}^{(t)}$ (i.e., using t iterations of belief propagation) is asymptotically equivalent to that of estimating the same symbol through a scalar Gaussian channel with SNR equal to $\eta^{(t)}$. Thus the collective effect of the noise and the interference of other symbols to the desired symbol is equivalent to an additive Gaussian noise. The result is significant and rather surprising because the sparse linear system and noisy channel are arbitrary.

We relegate discussion of the function h to Section IV. In the following we describe the solution to $\eta^{(t)}$, which uniquely characterizes the estimation problem. Precisely, $\eta^{(t)}$, $t = 1, 2, \ldots$, are determined by the following recursion:

$$\eta^{(t+1)} = \mathsf{E}\left\{ \left(\mathsf{E}\left\{ \nabla_2 \log f\left(Y \middle| \sqrt{\beta\nu^{(t)}}\, W\right) \,\middle|\, Y \right\} \right)^2 \right\} \quad (5)$$

$$\nu^{(t)} = \mathsf{E}\left\{ \left(X - \mathsf{E}\left\{ X \mid \sqrt{\eta^{(t)}}\, X + N \right\} \right)^2 \right\} \quad (6)$$

where $\eta^{(0)} = 0$, $\nabla_2 g(y|u) = \frac{\partial}{\partial u} g(y|u)$ for arbitrary bivariate functions $g(y|u)$, and the random variables are defined in the following:

1) $X \sim P_X$ and $N \sim \mathcal{N}(0,1)$ are independent;
2) Y and W are jointly distributed with $W \sim \mathcal{N}(0,1)$ and $P_{Y|W}(y|w) = f\left(y \middle| \sqrt{\beta\nu^{(t)}}\, w\right)$.

The relationship between the random variables and parameters $\nu^{(t)}$, $\eta^{(t)}$ is illustrated in Figure 2. The above description completely determines the joint distribution of (W, Y), which is parameterized by $\nu^{(t)}$ and hence $\eta^{(t)}$. In particular, $\nu^{(t)}$ is the average variance of X conditioned on its noisy observation Z through a Gaussian channel with SNR equal to $\eta^{(t)}$, where $\eta^{(t)}$ is the variance of the conditional mean of a *score function* of the channel characteristic $f(\cdot|\cdot)$.

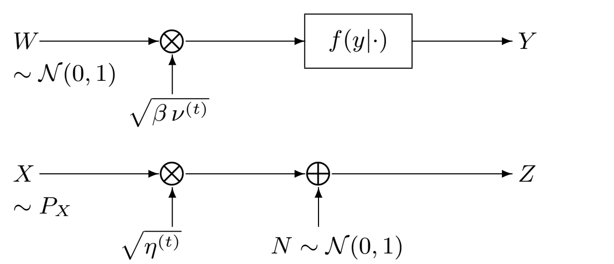

Fig. 2. The relationship between random variables in (5) and (6). Note that $\nu^{(t)}$ is the average variance of the input to the Gaussian channel with SNR equal to $\eta^{(t)}$ conditioned on its output.

Theorem 2: Suppose the recursive formulas (5) and (6) have a unique fixed point (η, ν). Then for every k and x where $P_X(x)$ is continuous,

$$P_{X_k|\boldsymbol{Y},\underline{\boldsymbol{S}}}\left(x \middle| \boldsymbol{Y}, \underline{\boldsymbol{S}}\right) \to P_{X|Z;\eta}(x|h(\boldsymbol{Y})) \quad (7)$$

in probability in the large-sparse-system limit, where $h(\cdot)$ is such that, conditioned on $X_k = a$, $h(\boldsymbol{Y}) \sim \mathcal{N}(a\sqrt{\eta}, 1)$.

Theorem 2 states that the problem of estimating each X_k given the entire observation $\boldsymbol{Y}$, is also asymptotically equivalent to estimating the same symbol through a scalar Gaussian channel, the SNR of which is equal to $\eta = \lim_{t\to\infty} \eta^{(t)}$. In view of Theorem 1, it implies that observing $\boldsymbol{Y}^{(t)}$ becomes as good as observing $\boldsymbol{Y}$ as $t \to \infty$, even though the ratio of the dimensions of $\boldsymbol{Y}$ and $\boldsymbol{Y}^{(t)}$ approaches infinity. Remarkably, this implies that BP is asymptotically as good as the (optimal) *a posteriori* estimation.

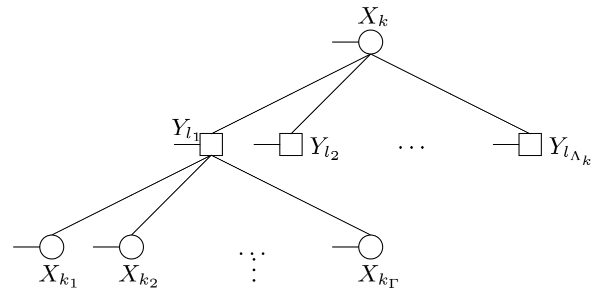

Fig. 3. Statistical inference over the corresponding tree structure.

In the special case where $f(\cdot|\cdot)$ represents a Gaussian channel, i.e., $\nabla_2 \log f(y|w) = y - w$, (5) becomes

$$\eta^{(t+1)} = \mathsf{E}\left\{ \left(\mathsf{E}\left\{ Y - W\sqrt{\beta\nu^{(t)}} \,\middle|\, Y \right\} \right)^2 \right\} \quad (8)$$

$$= \mathsf{E}\left\{ \left(Y - \frac{\beta\nu^{(t)}}{1+\beta\nu^{(t)}} Y \right)^2 \right\} \quad (9)$$

$$= \frac{1}{1+\beta\nu^{(t)}}. \quad (10)$$

Together with (6), we find the following recursion

$$\frac{1}{\eta^{(t+1)}} = 1 + \beta\mathsf{E}\left\{ \left(X - \mathsf{E}\left\{ X \mid \sqrt{\eta^{(t)}}\, X + N \right\} \right)^2 \right\} \quad (11)$$

which was first obtained in [7], [8], where $\eta^{(t)}$ is the multiuser efficiency achieved by BP after t iterations. Note that Boutros and Caire obtained a similar formula in the context of iterative decoding of coded CDMA using an empirically inspired Gaussian approximation [13]. The fixed-point equation corresponding to (11) was originally obtained in [2].

IV. Proof

A. Notation

Consider the inference tree illustrated in Figure 3, which is a subgraph of the factor graph depicted in Figure 1. The nodes correspond to random variables, and the edges correspond to dependencies between the nodes. Let X_k be the root, which has children $Y_{l_1}, \ldots, Y_{l_{\Lambda_k}}$, where Λ_k is the node degree of X_k. Suppose further that Y_{l_1} has children $X_{k_1}, \ldots, X_{k_\Gamma}$ where $\Gamma + 1 = \Gamma_{l_1}$ is its node degree. Suppose the subgraph contains all nodes within distance $2t$ to X_k, while only the first 3 layers of the subgraph are shown. It is understood that for the random ensemble of our interest, the graph is a tree (i.e., cycle-free) in the large-sparse-system limit with probability 1.

Use Y^l to denote the collection of all chip nodes Y_m in the subtree with Y_l as the root. Use Y^k to denote the collection of Y_m in the subtree with X_k as the root. Let $\{Y^{k_m}\} = (Y^{k_1}, \ldots, Y^{k_\Gamma})$. The letters l and k here are designated to indexing chips and symbols respectively.

Consider a fixed reference value $x_0 \in \chi$. In general, define the LLR function of X given some observation $U = u$ as

$$\mathcal{L}_{U|X}(u|x) = \log \frac{p_{U|X}(u|x)}{p_{U|X}(u|x_0)}, \quad x \in \chi. \tag{12}$$

B. Density Evolution

Consider the estimation of X_k given the subtree Y^k of depth $2t$ with X_k as its root. An optimal scheme is to pass "messages" upwards starting from the leaves. Each symbol node X_{k_i} sends to its parent the LLR function $\{\mathcal{L}_{Y^{k_i}|X_{k_i}}(y^{k_i}|x),\ x \in \chi\}$, while each chip node Y_{l_j} sends to its parent (in this case X_k) $\{\mathcal{L}_{Y^{l_j}|X_k}(y^{l_j}|x),\ x \in \chi\}$. The messages is in general a function defined on χ. At the final stage, the LLR given the entire subtree $Y^k = y^k$ is

$$\mathcal{L}_{Y^k|X_k}(y^k|x) = \sum_{n=1}^{\Lambda_k} \mathcal{L}_{Y^{l_n}|X_k}(y^{l_n}|x) \tag{13}$$

because the subtrees Y^{l_n} are independent conditioned on X_k.

By the central limit theorem, the LLR (13) as a sum of i.i.d. random functions is asymptotically Gaussian for every x and in fact a Gaussian random process indexed by $x \in \chi$. In the following we derive the mean and variance of the LLR in the large-system limit, which determines the equivalent channel between the desired symbol X_k and the observation Y^k.

Consider $\mathcal{L}_{Y^{l_1}|X_k}$, which is obtained from $p_{Y^{l_1}|X_k}(y^{l_1}|x)$. The conditional distribution $p_{Y_{l_1}|X_k,\{X_{k_m}\}}$ is determined by $f(\cdot|\cdot)$. We average over X_{k_m} conditioned on their respective subtrees to obtain $p_{Y^{l_1}|X_k}(y^{l_1}|x)$, which can be expressed using Taylor series expansion in $\Lambda_{k_m}^{-1/2}$ in (14)–(16). Sufficient regularity conditions are used to guarantee uniform convergence so that $o(1/\Lambda_{k_m})$ can walk in and out of the integrals.

The LLR is obtained by plugging (16) into

$$\mathcal{L}_{Y^{l_1}|X_k}(y^{l_1}|x) = \log \frac{p_{Y^{l_1}|X_k}(y^{l_1}|x)}{p_{Y^{l_1}|X_k}(y^{l_1}|x_0)}. \tag{17}$$

Let

$$W_{l_1} = \sum_{m=1}^{\Gamma} \frac{s_{l_1 k_m}}{\sqrt{\Lambda_{k_m}}} X_{k_m} \tag{18}$$

and define a random variable Y'_{l_1} independent of everything else conditioned on W_{l_1}, where $p_{Y'_{l_1}|W_{l_1}}(y|w) = f(y|w)$. Clearly, Y'_{l_1}–$\{X_{k_m}\}$–$\{Y^{k_m}\}$ form a Markov chain. The integral in (16), taken over the variables in the middle of the Markov chain, is proportional to $p_{Y'_{l_1},\{Y^{k_m}\}}(y_{l_1}, \{y^{k_m}\})$. In fact, we can write (19) in below for arbitrary $g(\cdot)$. Consequently, the LLR (17) can be expressed concisely using conditional expectations.

By (16)–(19), the LLR $\mathcal{L}_{Y^{l_1}|X_k}(Y^{l_1}|X_k)$ is expressed in the form of Taylor series expansion in (20), where two of the three significant components involve the conditional mean of the score function $\nabla_2 \log f\big(Y'_{l_1}|W_{l_1}\big)$.

$$\begin{aligned}
p_{Y^{l_1}|X_k}(y^{l_1}|x) &= \int p_{Y_{l_1}|X_k,\{X_{k_m}\}}(y_{l_1}|x,\{x_{k_m}\}) \prod_{m=1}^{\Gamma} p_{Y^{k_m}|X_{k_m}}(y^{k_m}|x_{k_m}) \prod_{m=1}^{\Gamma} \mathrm{d}P_{X_{k_m}}(x_{k_m}) && (14)\\
&= \int f\left(y_{l_1} \middle| \frac{s_{l_1 k}}{\sqrt{\Lambda_k}} x + \sum_{m=1}^{\Gamma} \frac{s_{l_1 k_m}}{\sqrt{\Lambda_{k_m}}} x_{k_m}\right) \prod_{m=1}^{\Gamma} \left(p_{Y^{k_m}|X_{k_m}}(y^{k_m}|x_{k_m})\, \mathrm{d}P_{X_{k_m}}(x_{k_m})\right) && (15)\\
&= \int \Bigg[f\left(y_{l_1} \middle| \sum_{m=1}^{\Gamma} \frac{s_{l_1 k_m}}{\sqrt{\Lambda_{k_m}}} x_{k_m}\right) + \frac{s_{l_1 k}}{\sqrt{\Lambda_k}} x \nabla_2 f\left(y_{l_1} \middle| \sum_{m=1}^{\Gamma} \frac{s_{l_1 k_m}}{\sqrt{\Lambda_{k_m}}} x_{k_m}\right) \\
&\qquad + \frac{s_{l_1 k}^2}{2\Lambda_k} x^2 \nabla_2^2 f\left(y_{l_1} \middle| \sum_{m=1}^{\Gamma} \frac{s_{l_1 k_m}}{\sqrt{\Lambda_{k_m}}} x_{k_m}\right) + o\left(\frac{1}{\Lambda_{k_m}}\right) \Bigg] \prod_{m=1}^{\Gamma} \left(p_{Y^{k_m}|X_{k_m}}(y^{k_m}|x_{k_m})\, \mathrm{d}P_{X_{k_m}}(x_{k_m})\right) && (16)
\end{aligned}$$

$$\frac{\displaystyle\int g(\{x_{k_m}\}) f\left(y_{l_1} \middle| \sum_{m=1}^{\Gamma} \frac{s_{l_1 k_m}}{\sqrt{\Lambda_{k_m}}} x_{k_m}\right) \prod_{m=1}^{\Gamma} \left(p_{Y^{k_m}|X_{k_m}}(y^{k_m}|x_{k_m})\, \mathrm{d}P_{X_{k_m}}(x_{k_m})\right)}{\displaystyle\int f\left(y_{l_1} \middle| \sum_{m=1}^{\Gamma} \frac{s_{l_1 k_m}}{\sqrt{\Lambda_{k_m}}} x_{k_m}\right) \prod_{m=1}^{\Gamma} \left(p_{Y^{k_m}|X_{k_m}}(y^{k_m}|x_{k_m})\, \mathrm{d}P_{X_{k_m}}(x_{k_m})\right)} = \mathsf{E}\left\{ g(\{X_{k_m}\}) \,\middle|\, Y'_{l_1} = y_{l_1}, \{Y^{k_m}\} = \{y^{k_m}\} \right\} \tag{19}$$

$$\begin{aligned}
\mathcal{L}_{Y^{l_1}|X_k}(y^{l_1}|x) &= \frac{s_{l_1 k}}{\sqrt{\Lambda_k}} (x - x_0) \mathsf{E}\left\{ \nabla_2 \log f(Y'_{l_1}|W_{l_1}) \,\middle|\, Y'_{l_1} = y_{l_1}, \{Y^{k_m}\} = \{y^{k_m}\} \right\} \\
&\quad + \frac{s_{l_1 k}^2}{2\Lambda_k} (x^2 - x_0^2) \mathsf{E}\left\{ \frac{\nabla_2^2 f(Y'_{l_1}|W_{l_1})}{f(Y'_{l_1}|W_{l_1})} \,\middle|\, Y'_{l_1} = y_{l_1}, \{Y^{k_m}\} = \{y^{k_m}\} \right\} \\
&\quad - \frac{s_{l_1 k}^2}{2\Lambda_k} (x^2 - x_0^2) \Big(\mathsf{E}\left\{ \nabla_2 \log f(Y'_{l_1}|W_{l_1}) \,\middle|\, Y'_{l_1} = y_{l_1}, \{Y^{k_m}\} = \{y^{k_m}\} \right\} \Big)^2 + o\Big(\frac{1}{\Lambda_k}\Big)
\end{aligned} \tag{20}$$

Consider now $\mathcal{L}_{Y^{l_1}|X_k}(Y^{l_1}|x)$ as a random variable for any given x, which consists of three components according to (20). We estimate its mean and variance to the first order of $1/\Lambda_k$, conditioned on that the true value of $X_k = x_k$. Only the first term on the right hand side of (20) contributes to the variance of the LLR in the order of $O(1/\Lambda_k)$. The mean of the LLR is a little more involved, because the prior of Y_{l_1} is slightly different than that of Y'_{l_1}. If they were the same, the expectation of the first two terms on the right hand side of (20) is 0 due to properties of the score function. While the third term contributes in the order of $O(1/\Lambda_k)$, the difference between the prior of Y_{l_1} and Y'_{l_1} is of size $O(1/\sqrt{\Lambda_k})$. Taking into account the small correction, the first term also contribute to the mean in the order of $O(1/\sqrt{\Lambda_k})$, while the second term remains insignificant.

Let us define

$$\eta = \lim_{K\to\infty} \mathsf{E}\left\{ \left(\mathsf{E}\left\{ \nabla_2 \log f(Y'_{l_1}|W_{l_1}) \,\middle|\, Y'_{l_1}, \{Y^{k_m}\} \right\} \right)^2 \right\} \quad (21)$$

Conditioned on $X_k = x_k$, the mean of the LLR is then

$$\begin{aligned} &\mathsf{E}\left\{ \mathcal{L}_{Y^{l_1}|X_k}(Y^{l_1}|x) \,\middle|\, X_k = x_k \right\} \\ &\quad = \eta \frac{2(x-x_0)x_k - (x^2 - x_0^2)}{2\Lambda_k} + o\Big(\frac{1}{\Lambda_k}\Big) \end{aligned} \quad (22)$$

where we use the fact that $s_{lk}^2 = 1$, and the variance is

$$\mathsf{E}\left\{ \mathcal{L}^2_{Y^{l_1}|X_k}(Y^{l_1}|x) \middle| X_k = x_k \right\} = \eta \frac{(x-x_0)^2}{\Lambda_k} + o\Big(\frac{1}{\Lambda_k}\Big) \quad (23)$$

In view of (13), the asymptotic statistics of the Gaussian LLR $\mathcal{L}_{Y^k|X_k}(Y^k|x)$ are

$$\mathsf{var}\left\{ \mathcal{L}_{Y^k|X_k}(Y^k|x) \right\} = \eta(x-x_0)^2 \quad (24)$$

$$\mathsf{E}\left\{ \mathcal{L}_{Y^k|X_k}(Y^k|x) \right\} = \eta x_k(x-x_0) - \frac{\eta}{2}(x^2 - x_0^2). \quad (25)$$

Proposition 1: The LLR $\mathcal{L}_{Y^k|X_k}(Y^k|x)$ is asymptotically Gaussian conditioned on X_k and identically distributed as

$$\mathcal{L}_{Z|X_k}(Z|x) = \sqrt{\eta} Z(x-x_0) - \frac{\eta}{2}(x^2 - x_0^2) \quad (26)$$

where

$$Z = \sqrt{\eta} X_k + N \quad (27)$$

and $N \sim \mathcal{N}(0,1)$ is standard Gaussian.

Proof: Let Z be defined according to (27). The likelihood

$$\mathcal{L}_{Z|X_k}(Z|x) = \log \frac{\exp\left[-\frac{1}{2}(Z-\sqrt{\eta}x)^2\right]}{\exp\left[-\frac{1}{2}(Z-\sqrt{\eta}x_0)^2\right]} \quad (28)$$

is equal to the right hand side of (26), the mean and variance of which are identical to those of $\mathcal{L}_{Y^{l_1}|X_k}(Y^{l_1}|x)$. ■

The significance of Proposition 1 is that in terms of estimating X_k, having access to the output of the companion scalar channel (27) is as good as observing the entire subtree Y^k. Indeed, there exists a conditionally Gaussian variable $Z = f(Y^k)$, which is a sufficient statistic of Y^k for X_k in the large-sparse-system limit. The SNR of the equivalent channel η is given by (21).

In the following, we briefly explain why η can be obtained from the evolution formula (5). The problem of estimating X_{k_m} given Y^{k_m} is identical to that of estimating X_k given Y^k, except for two points: (i) the subtree with X_{k_m} as the root has depth $2(t-1)$; (ii) X_{k_m} has $\Lambda_{k_m} - 1$ children while X_k has Λ_k. The impact of (ii) vanishes as $\Lambda_k \to \infty$. As a result of (i), (21) can be regarded as an evolution of $\eta^{(t)}$, which is dependent on the depth of the subtree. It suffices to evaluate the variance of W_{l_1} given $\{Y^{k_m}\}$, which is the sum of the variance of X_{k_m} given Y^{k_m} due to conditional independence. The individual variance of X_{k_m} given Y^{k_m} depends on Y^{k_m}, which is equivalent to the sufficient statistic $Z = \sqrt{\eta^{(t-1)}} X_k + N$. As the linear combination of X_{k_m}, W_{l_1} is asymptotically Gaussian whose variance is β times the variance of individual X_{k_m} given Y^{k_m}. Indeed, if the left hand side of (21) is replaced by $\eta^{(t+1)}$, then W_{l_1} in (21) can be replaced by $\sqrt{\beta\nu^{(t)}}W$ where $W \sim \mathcal{N}(0,1)$, and $\nu^{(t)}$ is the average variance of X conditioned on its noisy observation through a Gaussian channel with SNR equal to $\eta^{(t)}$. The iterative formulas (5)–(6) are thus justified. Furthermore, the function h in Theorem 1 produces the same Gaussian statistic as BP does asymptotically.

References

[1] T. Tanaka, "A statistical mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2888–2910, Nov. 2002.

[2] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1982–2010, June 2005.

[3] T. Tanaka and M. Okada, "Approximate belief propagation, density evolution, and statistical neurodynamics for CDMA multiuser detection," *IEEE Trans. Inform. Theory*, vol. 51, pp. 700–706, Feb. 2005.

[4] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *Journal of Physics A: Mathematical and General*, vol. 36, pp. 11111–11121, 2003.

[5] J. P. Neirotti and D. Saad, "Improved message passing for inference in densely connected systems," *Europhys. Lett.*, vol. 71, no. 5, pp. 866–872, 2005.

[6] 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 *Proc. IEEE Information Theory Workshop*, Punta del Este, Uruguay, 2006.

[7] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *Proc. IEEE Information Theory Workshop*, Chengdu, China, 2006.

[8] C.-C. Wang and D. Guo, "Belief propagation is asymptoticly equivalent to MAP detection for sparse linear systems," in *Proc. 44th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, 2006.

[9] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.

[10] C. C. Wang, S. R. Kulkarni, and H. V. Poor, "Density evolution for asymmetric memoryless channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4216–4236, Dec. 2005.

[11] T. Tanaka, "Replica analysis of performance loss due to separation of detection and decoding in CDMA channels," in *Proc. IEEE Int. Symp. Information Theory*, Seattle, WA, USA, June 2006.

[12] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: Asymptotic distance distributions," *IEEE Trans. Inform. Theory*, vol. 48, pp. 887–908, Apr. 2002.

[13] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1772–1793, July 2002.