CDMA multiuser detection, neural networks, and statistical mechanics

Toshiyuki Tanaka
Department of Electronics and Information Engineering, Tokyo Metropolitan University
Tokyo, Japan,
Neural Computing Research Group, Aston University
Birmingham, United Kingdom.
e-mail: tanaka@eei.metro-u.ac.jp

Abstract

A novel approach, based on statistical mechanics, to analyzing typical performance of optimum code-division multiple-access (CDMA) multiuser detectors is reviewed. A "black box" view of the basic CDMA channel is introduced, based on which the CDMA multiuser detection problem is regarded as a "learning-from-examples" problem of the "binary linear perceptron" in the neural network literature. Adopting Bayes framework, analysis of the performance of the optimum CDMA multiuser detectors is reduced to evaluation of the average of the cumulant generating function of a relevant posterior distribution. The evaluation of the average cumulant generating function is done, based on formal analogy with a similar calculation appearing in the spin glass theory in statistical mechanics, by making use of the replica method, a method developed in the spin glass theory.

1 CDMA channel model and neural networks

The basic K-user fully-synchronous direct-sequence CDMA channel with additive white Gaussian noise under perfect power control is considered (Fig. 1; see also [1]). Without loss of generality one can focus on any information bit interval and let $x_k \in \{-1, 1\}$, $k = 1, \ldots, K$, be the information bit for user k in that interval. The spreading code sequence of user k during the information bit interval is denoted by $\{s_k^1, \ldots, s_k^N\}$, where N is the spreading factor. The received signal at t-th chip interval ($t = 1, \ldots, N$) is given by

$$y^t = \sum_{k=1}^{K} s_k^t x_k + \nu^t, \tag{1}$$

where $\nu^t \sim N(0, \sigma^2)$ denotes realization of channel noise at t-th chip interval. In the CDMA detection problem, one has to estimate information bits x_k

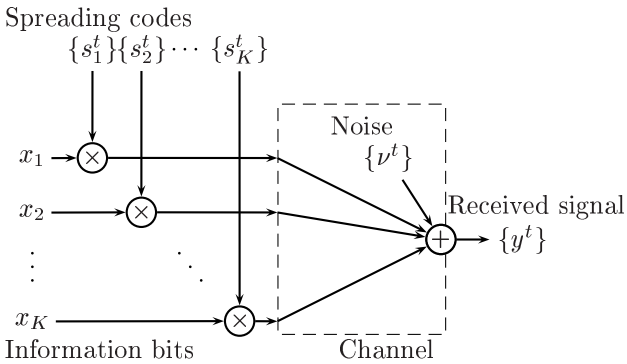

Figure 1: Basic CDMA channel

from received signal $\boldsymbol{y} = (y^1, \ldots, y^N)$, by making use of the knowledge about the spreading codes $S = (s_k^t)$.

An interesting observation is that the CDMA detection problem can be regarded as a problem in the framework of statistical learning theory. This is done by taking a "black box" view of the basic CDMA channel, as depicted in Fig. 2, in which the received signal y is regarded as the output of the black box, the spreading codes $\boldsymbol{s} = (s_1, \ldots, s_K)$ as the inputs, and the information bits $\boldsymbol{x} = (x_1, \ldots, x_K)$ as the parameters. The parameters of the black box are binary, and assumed unknown to the learner. The learner is supposed to estimate the hidden parameter values $\boldsymbol{x}$ of the black box from N input-output pairs $T \equiv \{(\boldsymbol{s}^t, y^t) | t = 1, \ldots, N\}$. T is called the training set. This setting is called the "learning from examples" in the context of statistical learning theory.

In general, the learner in a learning-from-examples problem assumes a particular parametric model to describe possible input-output relationship. One of the basic model is called the perceptron [2], which is one of the standard models in the field of neural networks. The input-output relationship of the perceptron is generally expressed

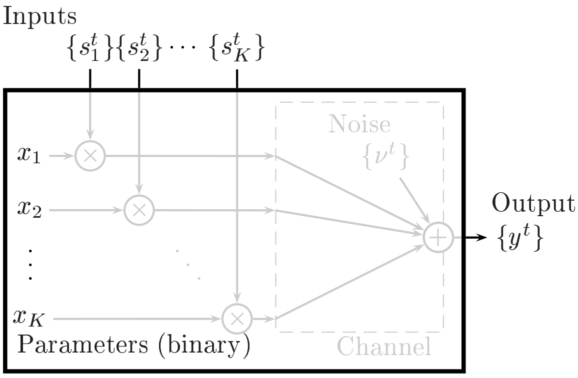

Figure 2: A "black box" view of basic CDMA channel

as

$$y = f\left(\sum_{k=1}^{K} x_k s_k\right) + \nu, \tag{2}$$

where ν represents noise. Various choices are possible for the transfer function f: The original proposal of the perceptron uses a threshold function for f, whereas a common choice in more recent neural network literature is the sigmoid function $f(z) \equiv \tanh z$. Comparison of Eqs. (1) and (2) immediately reveals that the basic CDMA channel model is a special case of the perceptron, in which parameters are binary and $f(z) \equiv z$ ("binary linear perceptron"). This means that one can indeed regard the CDMA detection problem as a problem of perceptron learning, in which the binary linear perceptron is assumed. Implication of this observation is in fact far more than a mere analogy. For example, there is a significant amount of theoretical research on the perceptron learning (see, e.g., [3]), so that it has now become possible to make use of it for the analysis of the CDMA detection problem.

2 Multiuser detection and statistical mechanics

After having introduced the viewpoint from the statistical learning theory, it is useful to consider the multiuser detection problem within the Bayes framework. The channel model postulated by the learner defines a conditional distribution $p(\boldsymbol{y}|S, \boldsymbol{x})$. Assuming a particular prior $p(\boldsymbol{x})$ for $\boldsymbol{x}$, one obtains the posterior distribution of $\boldsymbol{x}$ conditioned on the training set $T = (S, \boldsymbol{y})$, as

$$p(\boldsymbol{x}|T) \equiv p(\boldsymbol{x}|\boldsymbol{y}, S) = \frac{p(\boldsymbol{y}|S, \boldsymbol{x})p(\boldsymbol{x})}{\sum_{\boldsymbol{x}} p(\boldsymbol{y}|S, \boldsymbol{x})p(\boldsymbol{x})} \tag{3}$$

To construct the optimum multiuser detector, we need a loss function L. If we are given the posterior distribution $p(\boldsymbol{x}|T)$ and a loss function L, we can construct the optimum multiuser detector which minimizes the expected loss.

Since the optimum multiuser detector is defined on the basis of the posterior distribution, its performance can be analyzed by investigating the posterior distribution $p(\boldsymbol{x}|T)$. Instead of investigating the posterior distribution directly, we can analyze the cumulant generating function

$$\phi(\boldsymbol{h}; T) \equiv \log\langle e^{\boldsymbol{h}\cdot\boldsymbol{x}}\rangle_{\boldsymbol{x}}. \tag{4}$$

This is because the cumulant generating function preserves all the information about the original distribution. In Eq. (4), $\langle\cdot\rangle_{\boldsymbol{x}}$ denotes averaging over the posterior distribution, so that the result is a function of the training set T.

Since the cumulant generating function, and the performance as well, depends on the training set T, it is usual, in the statistical learning literature, to consider random training sets and averaging of the performance over their randomness. In the context of CDMA multiuser detection, the averaging over the output $\boldsymbol{y}$ corresponds to that over channel noise, and the averaging over the input S to the *random spreading* assumption.

Now one has to evaluate the average of the cumulant generating function over the training set T:

$$E_T[\phi(\boldsymbol{h}; T)] = E_T\left[\log\langle e^{\boldsymbol{h}\cdot\boldsymbol{x}}\rangle_{\boldsymbol{x}}\right]. \tag{5}$$

However, the existence of the logarithm *inside* the expectation over T in general makes its evaluation difficult. It is often intractable for larger-sized problems. The same difficulty also exists in the spin glass theory in statistical mechanics, where one has to consider macroscopic properties (magnetization etc.) of large systems (in fact, the size of systems considered is typically infinite — thermodynamic limit) with some randomness in them ("spin glasses"). Researchers of spin glass have developed the *replica method*, which enables us to deal with averages of the form like Eq. (5). It makes use of the following identity relation:

$$\begin{aligned}
&\lim_{n\to 0} \frac{\partial}{\partial n} \log E_T\left[\left(\langle e^{\boldsymbol{h}\cdot\boldsymbol{x}}\rangle_{\boldsymbol{x}}\right)^n\right] \\
&= \lim_{n\to 0} \frac{E_T\left[\left(\langle e^{\boldsymbol{h}\cdot\boldsymbol{x}}\rangle_{\boldsymbol{x}}\right)^n \log\langle e^{\boldsymbol{h}\cdot\boldsymbol{x}}\rangle_{\boldsymbol{x}}\right]}{E_T\left[\left(\langle e^{\boldsymbol{h}\cdot\boldsymbol{x}}\rangle_{\boldsymbol{x}}\right)^n\right]} \\
&= E_T\left[\log\langle e^{\boldsymbol{h}\cdot\boldsymbol{x}}\rangle_{\boldsymbol{x}}\right].
\end{aligned} \tag{6}$$

This reduces the evaluation of $E_T\left[\log\langle e^{\boldsymbol{h}\cdot\boldsymbol{x}}\rangle_{\boldsymbol{x}}\right]$ to that of $E_T\left[\left(\langle e^{\boldsymbol{h}\cdot\boldsymbol{x}}\rangle_{\boldsymbol{x}}\right)^n\right]$. The next step is to evaluate the latter *only for positive integers* n. In general this evaluation is still difficult. In many cases, however, the evaluation becomes possible if one is allowed to take the large-system limit ($K \to \infty$ in our case). The final step is to take the formal limit $n \to 0$ of the result obtained so far. The crucial assumption here is that the formal limit $n \to 0$

of the result for positive integers n is assumed to be valid. Justification of the general procedure involved in typical application of the replica method is still missing in mathematically rigorous sense, but it has been applied successfully to many complex problems across various fields and has yielded reasonable results (For details see, e.g., [4]).

Recently, the replica method has been applied to the analysis of the CDMA multiuser detectors [5, 6, 7, 8, 9]. The main result is summarized in the following proposition:

Proposition 1 *Assume the basic CDMA channel and the random spreading, in which s_k^t is mean 0 and variance $1/N$. Let $K, N \to \infty$ while the load $\beta \equiv K/N$ remains finite. Let σ_0^2 and σ^2 be the true channel noise level and its postulate by the detector. Then the bit-error rate P_b of the optimum CDMA multiuser detector is given by*

$$P_b = Q\left(\frac{E}{\sqrt{F}}\right), \tag{7}$$

where E and F are to be determined by solving the following saddle-point equations for $\{m, q, E, F\}$:

$$m = \int \tanh\left(\sqrt{F}z + E\right) Dz \tag{8}$$

$$q = \int \tanh^2\left(\sqrt{F}z + E\right) Dz \tag{9}$$

$$E = \frac{\beta^{-1}B}{1 + B(1-q)} \tag{10}$$

$$F = \frac{\beta^{-1}B^2}{[1 + B(1-q)]^2}(B_0^{-1} + 1 - 2m + q). \tag{11}$$

Here, $Dz = e^{-z^2/2}dz/\sqrt{2\pi}$ is the Gaussian measure, $Q(z) = \int_z^\infty Dz$ the error function, $B = \beta/\sigma^2$, and $B_0 = \beta/\sigma_0^2$.

See [8] for the derivation of the result and technical details involved in the derivation. Note that this is the result for the equal-power case: extension to the unequal-power case is discussed by Guo and Verdú [9].

One has to solve the saddle-point equations numerically to obtain the quantitative figures for the bit-error rate P_b. Figure 3 shows the result for a case in which the system is heavily loaded ($\beta = 1.4$). An interesting characteristics found in this result is that the bit-error rate P_b exhibits anomalous dependence on the signal-to-noise ratio E_b/N_0.

In order to obtain interpretation of such results, it is again worthwhile to make use of the analogy with statistical mechanics: This is essentially the same phenomenon as what is observed in ferromagnetic materials: Magnetization curve of the Curie-Weiss model, a basic model of ferromagnet, is also S-shaped below the Curie temperature (i.e., in the ferromagnetic phase), as shown in Fig. 4. In that

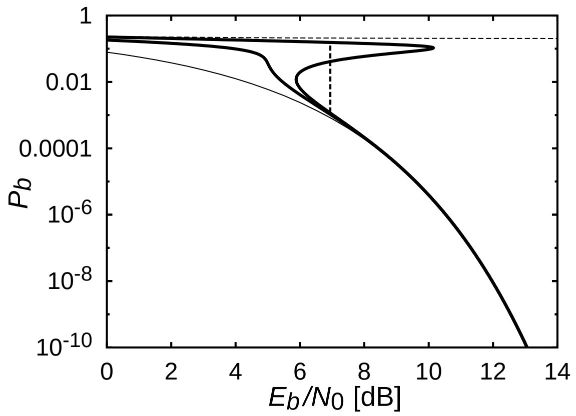

Figure 3: Bit-error rate of optimum multiuser detectors for the load $\beta = 1.4$. Detectors: Jointly-optimum (upper thick curve, S-shaped), Individually-optimum (lower thick curve), Single-user matched-filter (thin dashed curve). Single-user case is shown as thin solid curve. A vertical dashed line shows the thermodynamic transition for the jointly-optimum detector.

case, the thermal equilibrium distribution of microscopic configurations can be decomposed into more than one connected components in the configuration space, and the ergodicity of the system is broken, which causes the coexistence of multiple solutions. Each of the (stable) solutions correspond to one ergodic component of the thermal equilibrium distribution. When more than one solutions coexist, the thermodynamically relevant one is given by the one with minimum free energy, and that solution is called the globally stable state. Other solutions giving local minima of the free energy is called metastable states. The thermodynamically "true" magnetization value is discontinuous at the *thermodynamic transition point*. Metastable states are also important in understanding the properties of the model. In particular, it explains the hysteresis phenomenon: Abrupt reversal of the magnetization occurs at the *spinodal point*, where a metastable state disappears.

We can obtain the interpretation of the result with the S-shaped performance curves based on the analogy: The S-shape of the performance curve allows the detector to have more than one values for the bit-error rate, which reflects the decomposition of the posterior distribution to multiple connected components. In that case, the "true" value in the information-theoretic sense is given by the globally stable state, and is discontinuous at the thermodynamic transition point, which is shown in Fig. 3 as a vertical dashed line. Solutions other than the information-theoretically correct one (i.e., metastable states) are also important in understanding the system's properties, since they may

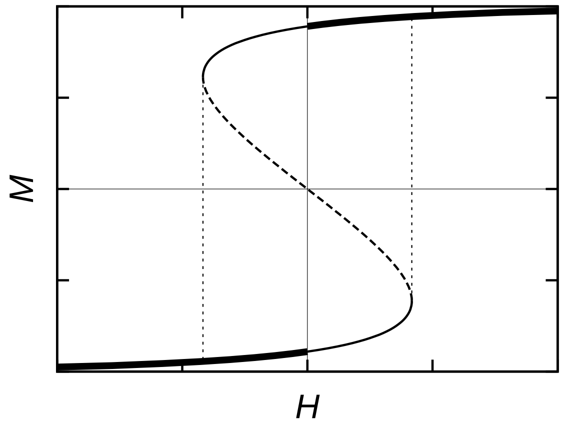

Figure 4: An example of magnetization curve for the Curie-Weiss model. Thick curves: globally stable state, thin curves: metastable states, dashed curve: unstable state.

affect the detector's behavior, thereby limiting its performance in a practical sense. Even if the globally stable state has a near-optimum performance, it may be computationally infeasible for any detector to find it, because any computationally feasible implementation of the optimum multiuser detector would easily get trapped in a metastable state when it exists. We can therefore predict that such detectors would exhibit discontinuity in the performance, just as the 'waterfalling' observed in the turbo decoding, around the spinodal point, not at the thermodynamic transition point.

3 Summary and outlook

Following interdisciplinary links between the CDMA detection problem and neural networks, and between neural networks and statistical mechanics, novel results and views have been obtained to the CDMA multiuser detection problem.

The significance of the statistical mechanical approach would be that it is also applicable to problems in various fields, which share the basic structure of Bayes framework and intrinsic randomness. One of such problems actively studied is that of the low-density parity-check codes (see [4] and references therein). We believe that the statistical mechanical approach will become more important in the field of information theory.

Another important aspect we would like to mention is that the replica method still lacks rigorous mathematical justification. The situation could be compared to that for the delta function when it was first invented. We expect that a number of successful applications of the replica method to problems of various fields will stimulate studies for its mathematical justification. Such studies will have significant impact across wide range of research areas, not only on statistical mechanics, but also probability theory and information theory, to mention a few.

Acknowledgment

Support from EPSRC research grant GR/N00562 is acknowledged.

References

[1] S. Verdú, *Multiuser Detection*, Cambridge Univ. Press, Cambridge, 1998.

[2] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, pp. 386–408, 1958.

[3] H. S. Seung, H. Sompolinsky, and N. Tishby, "Statistical mechanics of learning from examples," *Phys. Rev. A*, vol. 45, pp. 6056–6091, 1992.

[4] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*, Oxford University Press, 2001.

[5] T. Tanaka, "Analysis of bit error probability of direct-sequence CDMA multiuser demodulators," in T. Leen et al. (eds.), *Advances in Neural Information Processing Systems*, 2001, vol. 13, pp. 315–321, The MIT Press.

[6] T. Tanaka, "Statistical mechanics of CDMA multiuser demodulation," *Europhys. Lett.*, vol. 54, pp. 540–546, 2001.

[7] T. Tanaka, "Average-case analysis of multiuser detectors," *Proc. 2001 IEEE Int. Symp. Info. Theory*, 2001, p. 287.

[8] T. Tanaka, "A statistical mechanics approach to large-system analysis of CDMA multiuser detectors," submitted to *IEEE Trans. Inform. Theory*, 2001 [online] http://www-com.eei.metro-u.ac.jp./~tanaka/publications/Tanaka-cdma.ps.gz.

[9] D. Guo and S. Verdú, "Multiuser detection and statistical mechanics," to appear in *Ian Blake Festschrift*, 2002 [online] http://www.princeton.edu/~dguo/publications/GuoVer02Blake.pdf.