Sparse approximation problem: how rapid simulated annealing succeeds and fails

Tomoyuki Obuchi[1] **and Yoshiyuki Kabashima**

Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8502, Japan

E-mail: [1]obuchi@sp.dis.titech.ac.jp

Abstract. Information processing techniques based on sparseness have been actively studied in several disciplines. Among them, a mathematical framework to approximately express a given dataset by a combination of a small number of basis vectors of an overcomplete basis is termed the *sparse approximation*. In this paper, we apply simulated annealing, a metaheuristic algorithm for general optimization problems, to sparse approximation in the situation where the given data have a planted sparse representation and noise is present. The result in the noiseless case shows that our simulated annealing works well in a reasonable parameter region: the planted solution is found fairly rapidly. This is true even in the case where a common relaxation of the sparse approximation problem, the ℓ_1-relaxation, is ineffective. On the other hand, when the dimensionality of the data is close to the number of non-zero components, another metastable state emerges, and our algorithm fails to find the planted solution. This phenomenon is associated with a first-order phase transition. In the case of very strong noise, it is no longer meaningful to search for the planted solution. In this situation, our algorithm determines a solution with close-to-minimum distortion fairly quickly.

1. Introduction

The success of compressed sensing [1, 2, 3, 4] has triggered interest in the utilization of sparseness in signal processing techniques [5, 6, 7, 8, 9, 10, 11, 12, 13]. Sparseness is the property whereby data can be represented, on a proper basis, by some combination of a small number of non-zero components. This property is useful for practical applications such as data compression and data reconstruction from a small number of observations, the latter of which is simply compressed sensing.

Usually, obtaining a sparse representation from a given dataset is formulated as an optimization problem. We refer to this as the sparse approximation problem [14, 15, 16, 17, 18]. This is sometimes recast in a probabilistic formulation by statistical physicists [10, 12, 13] using Bayesian techniques or statistical mechanics. In this paper, we employ such a probabilistic formulation to search for an "optimal" solution with the minimum distortion between the given and reconstructed data.

Unlike previous approaches, we do not use a message passing algorithm [10]. Instead, we use the well-known "simulated annealing" (SA) heuristic. The motivation for using SA comes from our recent theoretical analysis of "entropy" [12, 13], which is the exponential rate of the number of combinations of non-zero components yielding a given level of distortion. Our analysis indicates that entropy exhibits some nice analytical properties, in contrast to other optimizations

such as the k-satisfiability problem [19, 20]. This implies a simple structure of the "phase space," the space of possible combinations of non-zero components, and hence SA is expected to work well. Based on this expectation, the actual performance of SA is reported through numerical experiments.

2. Formulation and algorithm

2.1. Combinatorial optimization formulation

Let us suppose a signal vector $\boldsymbol{y} \in \mathbb{R}^M$ is generated from an appropriate sparse representation or a planted solution, $\hat{\boldsymbol{x}}$, through

$$\boldsymbol{y} = A\hat{\boldsymbol{x}} + \boldsymbol{\xi}, \tag{1}$$

where $A = \{\boldsymbol{a}_i\}_{i=1}^N \in \mathbb{R}^{M\times N}$ is an overcomplete matrix with $M < N$ called a dictionary; $\boldsymbol{\xi} \in \mathbb{R}^M$ is a noise vector, each component of which is drawn from the zero-mean normal distribution with variance σ_ξ^2, $\mathcal{N}(0, \sigma_\xi^2)$. The appropriate representation $\hat{\boldsymbol{x}}$ is "sparse," and is assumed to be generated from the following Bernoulli–Gaussian distribution:

$$P(\hat{x}_i) = \hat{\rho}\frac{e^{-\frac{1}{2\sigma_x^2}\hat{x}_i^2}}{\sqrt{2\pi\sigma_x^2}} + (1-\hat{\rho})\delta(\hat{x}_i). \tag{2}$$

We also assume that each component of A is independent and identically distributed from $\mathcal{N}(0, 1/N)$. The aspect ratio of the matrix, $\alpha = M/N$, is assumed to lie within $(0, 1)$.

The sparse approximation problem is to approximate $\boldsymbol{y}$ by a linear combination of a restricted number of column vectors of A. There are various formulations, one of which is based on the following optimization:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \{\mathcal{E}(\boldsymbol{x}|\boldsymbol{y}, A)\} \text{ subject to } ||\boldsymbol{x}||_0 \leq N\rho, \tag{3}$$

where $||\boldsymbol{x}||_k = (\sum_i |x_i|^k)^{1/k}$ denotes the ℓ_k norm and the ℓ_0 norm is equal to the number of non-zero components of $\boldsymbol{x}$; the parameter $\rho(< \alpha)$ controls the sparseness of $\boldsymbol{x}$ (and is called the sparsity in this paper); and $\mathcal{E}$ denotes the distortion between $\boldsymbol{y}$ and a reconstructed signal through a representation $\boldsymbol{x}$:

$$\mathcal{E}(\boldsymbol{x}|\boldsymbol{y}, A) = \frac{1}{2}||\boldsymbol{y} - A\boldsymbol{x}||_2^2. \tag{4}$$

2.2. Probabilistic formulation

Eq. (3) is a commonly used formulation, but it has the limitation that it only provides the information of the minimum-distortion solution. To get a wider perspective, it is better to treat *all* possible combinations of the column vectors. For this, we use a probabilistic formulation. Suppose that a binary vector $\boldsymbol{c} = \{c_i = 0, 1\}_{i=1}^N$, which we call the sparse weight, represents the column vectors used to represent $\boldsymbol{y}$: if $c_i = 1$, the ith column of A, $\boldsymbol{a}_i$, is used; if $c_i = 0$, it is not. Once these columns have been determined by $\boldsymbol{c}$, the optimal coefficients of the chosen columns are evaluated by solving

$$\boldsymbol{x}(\boldsymbol{c}) = \arg\min_{\boldsymbol{x}} ||\boldsymbol{y} - A(\boldsymbol{c}\circ\boldsymbol{x})||_2^2, \tag{5}$$

where $(\boldsymbol{c}\circ\boldsymbol{x})_i = c_i x_i$ represents the Hadamard product. The corresponding distortion is

$$\mathcal{E}(\boldsymbol{c}|\boldsymbol{y}, A) = M\epsilon(\boldsymbol{c}|\boldsymbol{y}, A) = \frac{1}{2}||\boldsymbol{y} - A(\boldsymbol{c}\circ\boldsymbol{x}(\boldsymbol{c}))||_2^2. \tag{6}$$

The components of $\boldsymbol{x}(\boldsymbol{c})$ for the zero components of $\boldsymbol{c}$ are actually indefinite, and we set them to be zeros. The definite part of $\boldsymbol{x}(\boldsymbol{c})$, which we denote $\tilde{\boldsymbol{x}}(\boldsymbol{c})$, has the compact analytic form

$$\tilde{\boldsymbol{x}}(\boldsymbol{c}) = \left(\tilde{A}^{\mathrm{T}}(\boldsymbol{c})\tilde{A}(\boldsymbol{c})\right)^{-1} \tilde{A}^{\mathrm{T}}(\boldsymbol{c})\boldsymbol{y}, \tag{7}$$

where $\tilde{A}(\boldsymbol{c})$ denotes the submatrix of columns chosen by $\boldsymbol{c}$.

Let us regard $\mathcal{E}$ as an "energy," and introduce an "inverse temperature" μ. A Gibbs–Boltzmann distribution is thus defined as

$$P(\boldsymbol{c}|\mu;\boldsymbol{y},A) = \frac{1}{G(\mu;\boldsymbol{y},A)}\delta\left(\sum_i c_i - N\rho\right) e^{-\mu\mathcal{E}(\boldsymbol{c}|\boldsymbol{y},A)}, \tag{8}$$

where G is the "partition function"

$$G(\mu;\boldsymbol{y},A) = \sum_{\boldsymbol{c}} \delta\left(\sum_i c_i - N\rho\right) e^{-\mu\mathcal{E}(\boldsymbol{c}|\boldsymbol{y},A)}. \tag{9}$$

Our strategy is to generate $\boldsymbol{c}$ according to eq. (8). Changing μ allows us to sample different sparse solutions with different distortion values.

This formulation provides several options to treat the sparse approximation problem. For example, sampling in $\mu < \infty$ produces solutions with distortion greater than the minimum. These "finite temperature" solutions may be more suitable for capturing the planted solution $\hat{\boldsymbol{x}}$ than $\boldsymbol{x}^*$ in eq. (3) in the presence of noise $\sigma_\xi > 0$, as suggested in [13].

In this probabilistic formulation, the optimization (3) is recovered as a sampling problem in the limit $\mu \to \infty$. We pursue this direction in this paper, and propose an algorithm to solve eq. (3). The performance of our technique is examined in numerical experiments, and is compared with some known analytical results [7, 12, 13].

2.3. Simulated annealing

Our algorithm is a variant of SA, which is a metaheuristic to find the global minimum of a cost function. The outline of the algorithm is as follows: starting from a random initial configuration of $\boldsymbol{c}$ at very high temperature $T = 1/\mu \gg 1$, the algorithm randomly updates the configuration $\boldsymbol{c} \to \boldsymbol{c}'$ in a Monte-Carlo (MC) manner, while gradually decreasing the temperature. Eventually, the temperature becomes very low, $T \approx 0$, and the configuration is no longer updated. This final configuration is expected to be very close (or identical) to the true solution, *i.e.*, $\boldsymbol{x}^* \approx \boldsymbol{c} \circ \boldsymbol{x}(\boldsymbol{c})$.

The Metropolis criterion is adopted in our simulation: an MC move $\boldsymbol{c} \to \boldsymbol{c}'$ is judged to be accepted or not according to the probability

$$p_{\mathrm{accept}}(\boldsymbol{c} \to \boldsymbol{c}') = \max(1, e^{-\mu(\mathcal{E}(\boldsymbol{c}')-\mathcal{E}(\boldsymbol{c}))}). \tag{10}$$

For a fixed value of ρ, the configurations generated by the algorithm should always satisfy $\sum_i c_i = N\rho$. Given an initial configuration satisfying this condition, we generate trial moves $\boldsymbol{c} \to \boldsymbol{c}'$ by "pair flipping" two sparse weights, one equal to 0 and the other equal to 1. Namely, choosing an index i of the sparse weight from ONES $\equiv \{k|c_k = 1\}$ and another index j from ZEROS $\equiv \{k|c_k = 0\}$, we set $\boldsymbol{c}' = \boldsymbol{c}$, except for the counterpart of $(c_i, c_j) = (1, 0)$, which is given as $(c'_i, c'_j) = (0, 1)$.

The pseudo-code of our MC algorithm is given in Alg. 1, and that of our SA procedure is presented in Alg. 2. The lines marked with # are not necessarily needed for SA, but have been inserted for later convenience. For Alg. 2, we have a set of inverse temperature points $\{\mu_a\}_a^{N_\mu}$ arranged in ascending order $(0 =)\mu_1 < \mu_2 < \cdots < \mu_{N_\mu} (\gg 1)$ and the waiting times

Algorithm 1 MC update with pair flipping

procedure MCPF($\boldsymbol{c}, \mu, \boldsymbol{y}, A$) ▷ MC routine with pair flipping
  ONES $\leftarrow \{k|c_k = 1\}$, ZEROS $\leftarrow \{k|c_k = 0\}$
  randomly choose i from ONES and j from ZEROS
  $\boldsymbol{c}' \leftarrow \boldsymbol{c}$
  $(c'_i, c'_j) \leftarrow (0, 1)$
  $(\mathcal{E}, \mathcal{E}') \leftarrow (\mathcal{E}(\boldsymbol{c}|\boldsymbol{y}, A), \mathcal{E}(\boldsymbol{c}'|\boldsymbol{y}, A))$ ▷ Calculate energy
  $p_{\rm accept} \leftarrow \max(1, e^{-\mu(\mathcal{E}'-\mathcal{E})})$
  generate a random number $r \in [0, 1]$
  if $r < p_{\rm accept}$ **then**
    $\boldsymbol{c} \leftarrow \boldsymbol{c}'$
  end if
  return $\boldsymbol{c}$
end procedure

Algorithm 2 SA for sparse approximation problem

procedure SA($\{\mu_a, \tau_a\}_{a=1}^{N_\mu}, \rho, \boldsymbol{y}, A$)
  Generate a random initial configuration $\boldsymbol{c}$ with $\sum_i c_i = N\rho$
  for $a = 1 : N_\mu$ **do** ▷ Changing temperature
    for $t = 1 : \tau_a$ **do** ▷ Sampling at $\mu = \mu_a$
      for $i = 1 : N$ **do** ▷ Extensive number of MC updates
        $\boldsymbol{c} \leftarrow {\rm MC}_{\rm PF}(\boldsymbol{c}, \mu_a, \boldsymbol{y}, A)$
      end for
      # Calculate energy $\epsilon_t = \epsilon(\boldsymbol{c}_t|\boldsymbol{y}, A)$ of the current configuration $\boldsymbol{c}_t = \boldsymbol{c}$
    end for
    # Calculate the average energy $\epsilon_a = (1/\tau_a)\sum_{t=1}^{\tau_a} \epsilon_t$
  end for
  return $\boldsymbol{c}$
end procedure

$\{\tau_a\}_a$ at those points. Hence, as the algorithm proceeds, the temperature of the system $T = 1/\mu$ decreases step by step. It is known that, if the rate of decrease of the temperature obeys

$$T(t) > \frac{A(N)}{\log(t+2)} \tag{11}$$

for some time-independent constant $A(N)$, then the output of SA is guaranteed to be optimal [21]. This is a very slow schedule of decrease for T, and is generally overcautious so as to include the worst-case scenario. Faster schedules are known to work in practical situations. We report the results for such a rapid annealing below.

3. Results

We now present the results of SA according to eq. (8). System sizes of $N = 100, 200$, and 400 will be examined. The annealing schedule is fixed as

$$\mu_a = \mu_0 + r^{a-1} - 1,\ \tau_a = \tau,\ (a = 1, \cdots, 100). \tag{12}$$

We set $\tau = 5$, $\mu_0 = 10^{-8}$, and $r = 1.1$ as default parameter values. Thus, the maximum value of μ is $\mu_{100} \approx 1.3 \times 10^4$.

To determine whether the SA process is proceeding well, we consider the lines marked with # in Alg. 2. If our schedule is sufficiently slow, the $\boldsymbol{c}_t$ obtained during annealing are typical samples from eq. (8), and the values of physical quantities of typical samples should be very close (in the limit $N \to \infty$ almost surely identical) to the thermal averages. This implies the following relation

$$\epsilon_a = \frac{1}{\tau_a} \sum_t^{\tau_a} \epsilon(\boldsymbol{c}_t) \approx \langle \epsilon(\boldsymbol{c}|\boldsymbol{y}, A) \rangle_{\mu_a} , \tag{13}$$

where $\langle \cdots \rangle_\mu$ denotes the average over eq. (8) with the inverse temperature μ. Fortunately, the right-hand side is analytically assessed in the present case with a random matrix dictionary A in the $N \to \infty$ limit [12, 13]. Hence, we compare $\mathcal{E}_a$ with the analytically evaluated $\langle \mathcal{E} \rangle_{\mu_a}$ to determine how well our annealing process follows the correctly distributed samples from (8). For clarity of comparison, we take an average over a different $N_{\text{samp}} = 100$ samples of $\hat{\boldsymbol{x}}, \boldsymbol{\xi}, A$ in the numerical experiments. The error bar is given by the standard deviation among those samples divided by $\sqrt{N_{\text{samp}} - 1}$.

As well as the distortion, we calculate the mean squared error (MSE) between the planted and inferred representations

$$\mathcal{M}(\boldsymbol{c}) = \frac{1}{N} ||\hat{\boldsymbol{x}} - \boldsymbol{c} \circ \boldsymbol{x}(\boldsymbol{c})||_2^2. \tag{14}$$

This metric provides direct information about the reconstruction of the planted solution.

3.1. Noiseless case

Let us start with the noiseless case $\sigma_\xi^2 = 0$. In this subsection, we fix $\sigma_x^2 = 1$.

The easy reconstruction region, where α is sufficiently larger than $\hat{\rho}$, is a good starting point. Fig. 1 plots ϵ (left, middle) and the MSE $\mathcal{M}$ (right) against the temperature for $\alpha = 0.8, \rho = 0.4$, and $\hat{\rho} = 0.2$. The numerical results show a fairly good agreement with the analytical curve.

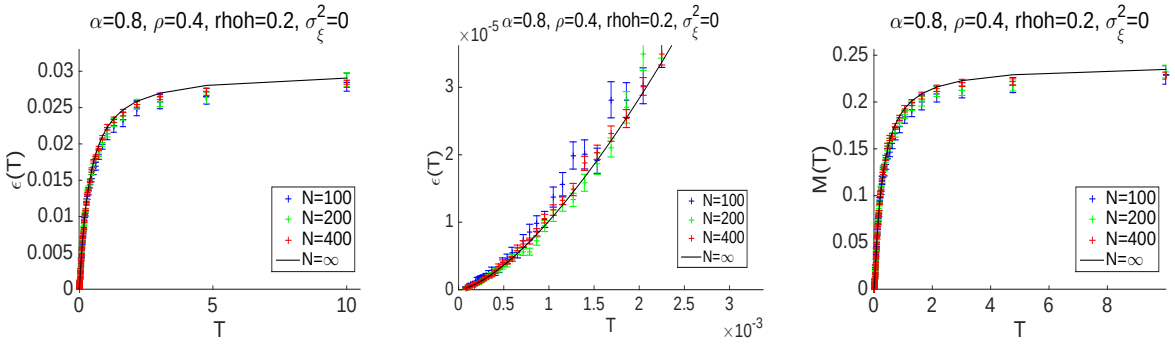

Figure 1. Distortion (left, middle) and MSE (right) plotted against temperature $T = 1/\mu$ for $\alpha = 0.8, \rho = 0.4$, and $\hat{\rho} = 0.2$. The middle panel is a magnified view of the left panel for a small region of T. The solid black curve shows the analytical values, and the color plots give the numerical results. Our numerical results clearly reproduce the correct average values over eq. (8).

This means our SA algorithm follows the equilibrium state up to the zero-temperature limit reasonably well, even though the annealing defined by eq. (12) is very rapid. The MSE goes to zero as T decreases, and so the planted solution is correctly reproduced.

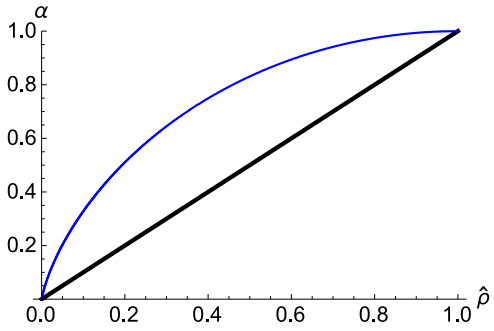

Figure 2. Phase diagram describing typical reconstruction limits in the noiseless case [7]. The straight line $\hat{\rho} = \alpha$ is the limit attained with eq. (3), whereas the curve is the limit for the relaxed problem in which the ℓ_0 norm in eq. (3) is replaced with the ℓ_1 norm. Above these boundaries, the solutions of corresponding optimization problems reconstruct the planted solution $\hat{\boldsymbol{x}}$.

Next, we consider harder cases. It is known that the properties of the problem drastically change as α gets closer to $\hat{\rho}$. The phase diagram in Fig. 2 demonstrates this. The curve is the reconstruction limit of the ℓ_1-relaxed version of (3), which is employed in many realistic cases and has considerable importance. Hence, we first examine the behavior of SA below this ℓ_1 reconstruction limit. Fig. 3 plots ϵ and $\mathcal{M}$ against T for $\alpha = 0.8, \rho = 0.55$, and $\hat{\rho} = 0.5$, where we are below the ℓ_1 reconstruction limit (the boundary is located at $\alpha_c(\hat{\rho} = 0.5) \approx 0.831$). The MSE vanishes as T decreases, meaning that $\hat{\boldsymbol{x}}$ is perfectly reconstructed. Hence, SA can

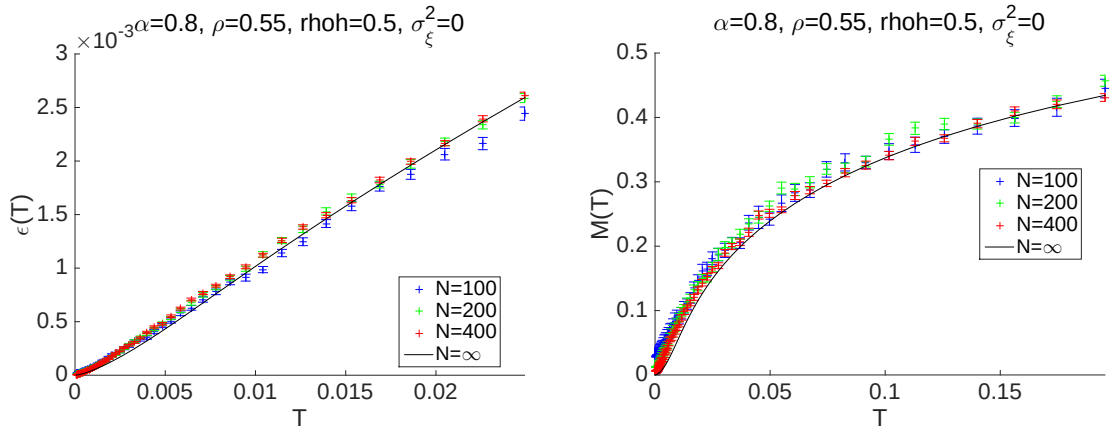

Figure 3. Distortion (left) and MSE (right) plotted against temperature for $\alpha = 0.8, \rho = 0.55$, and $\hat{\rho} = 0.5$ (below the ℓ_1 boundary). Our numerical results accord with the analytical result (black line) and achieve a perfect reconstruction of $\hat{\boldsymbol{x}}$.

outperform the ℓ_1 method of reconstruction, even under the present rapid schedule.

In harder situations, SA cannot always give a perfect reconstruction, even though it exists. Fig. 4 shows plots of T-ϵ and T-$\mathcal{M}$ for $\alpha = 0.75, \rho = 0.65$, and $\hat{\rho} = 0.5$. In this case, there are two stable thermodynamic states at low temperatures: one is connected to the planted solution and is the true equilibrium state in the zero temperature limit, whereas the other is metastable at low temperatures, but is the dominant equilibrium state at higher temperatures [13]. There is a first-order phase transition at a critical temperature $T_c \approx 3 \times 10^{-4}$. Through the SA process,

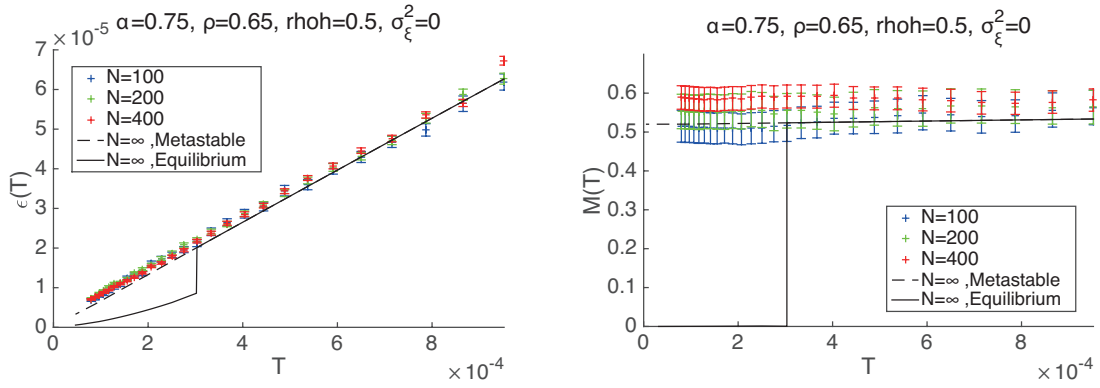

Figure 4. Distortion ϵ (left) and MSE $\mathcal{M}$ (right) plotted against temperature for $\alpha = 0.75, \rho = 0.65$, and $\hat{\rho} = 0.5$. The analytical result shows a first-order phase transition around $T_c \approx 3\times10^{-4}$, and the equilibrium state (black solid line) suddenly decreases for both ϵ and $\mathcal{M}$. However, a metastable state, which continues analytically to the equilibrium state for $T > T_c$, survives below T_c, and the numerical results follow this; hence, the planted solution is not reached.

the system state follows the equilibrium up to $T > T_c$. After the transition, the equilibrium state changes drastically, but the system cannot follow such a jump. Instead, the system remains in the same metastable state in $T < T_c$. Hence, SA cannot find the planted solution in this case, as clearly seen in the non-vanishing MSE $\mathcal{M}$ of Fig. 4. Of course, if τ is large enough, SA can eventually find the planted solution as proved in [21]. In the presence of the first-order phase transition, however, the required τ to do this is scaled with the system size N and rapidly grows as N increases, which prevents reaching the optimal solution in practical times.

Although the system is in the metastable state, the distortion ϵ seems to go to zero in Fig. 4, but unfortunately this is not the case. If we go lower temperatures, the value of ϵ will get stuck at a certain critical temperature and remains a constant below it. This is a freezing transition of this metastable state: the entropy of it becomes zero at the transition temperature. With the present parameters, our analytical solution provides the transition temperature as $T \approx 1.3\times10^{-7}$ and the constant value of distortion is $\epsilon = 8.6 \times 10^{-9}$ which is the achievable limit by our SA algorithm in practical times. This limiting value is not exactly zero although it is still negligibly small. For reference, we have plotted the "free energy" $g = (1/N)\log G$ and the entropy in Fig. 5 for the present parameter values.

Finally, we examined the effect of the parameter ρ, which can be set to any value for a given $\boldsymbol{y}$ and A. Fig. 6 is the counterpart of Fig. 4 with common values of $(\alpha, \hat{\rho}) = (0.75, 0.5)$ and a different value of $\rho = 0.55$. In contrast to Fig. 4, the metastable state is absent. We commonly observed the suppression of this metastable state as ρ decreased [13]. Thus, SA was naively expected to yield the planted solution in the zero temperature limit. The upper panels of Fig. 6 represent the rapid schedule $\tau = 5$, and exhibit a clear deviation from the equilibrium state at low temperatures. A slower schedule, corresponding to $\tau = 100$, is represented in the lower panels, and these numerical data are consistent with the analytical curve of the equilibrium state. The present choice of parameters is in the harder region, so it is unsurprising that the rapid schedule $\tau = 5$ is not sufficiently slow, although it was for Figs. 1-3. Note that this increase in τ is qualitatively different from that caused by the emergence of the metastable state in Fig. 4. The latter is macroscopic, namely, the required waiting time τ increases significantly as N grows to reach the planted solution.

In summary, SA outperforms the ℓ_1 result and can correctly find the planted solution in a

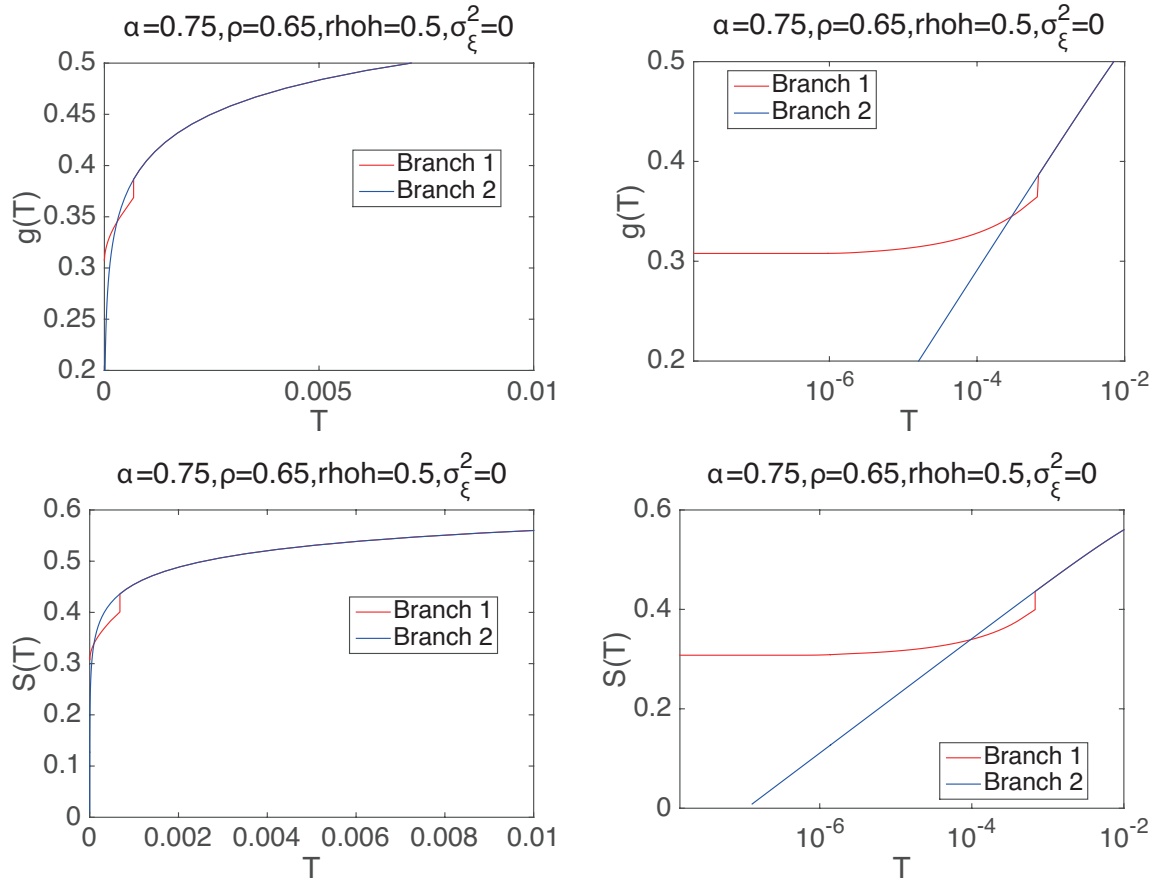

Figure 5. Free energy $g = (1/N)\log G$ (upper) and entropy (lower) plotted against temperature for $\alpha = 0.75, \rho = 0.65$, and $\hat{\rho} = 0.5$. Right panels are the same plots as the left panels but with a semi-logarithmic scale. Two different branches are shown, and the same colors denote the same branches in all panels. A phase transition (where two branches of g cross) can be observed at $T_c \approx 3 \times 10^{-3}$. The larger branch yields the equilibrium state. The entropy vanishes at a finite value of T, as clearly shown in the lower-right panel. Branch 1, the red solid curve, is connected to the planted solution.

wider region of parameter space, even under our rather rapid rate of decrease of T. However, for harder regions where α is close to $\hat{\rho}$, the phase space is separated into two distinctive states. The state that is not connected to the planted solution becomes the equilibrium state in the high-temperature region. This prevents SA from correctly finding the planted solution, as the system becomes trapped in the wrong state. Tuning ρ may be a crucial factor in overcoming this. As long as $\rho > \hat{\rho}$, smaller values of ρ are better. This is because the emergence of the metastable state tends to be suppressed as ρ becomes smaller, but such fine tuning requires *a priori* knowledge about the value of $\hat{\rho}$, which is not available in most situations. This problem is of considerable importance, and requires further consideration.

If we do not insist on finding the planted solution, even the metastable state may be desirable, as it shows small distortion. In this case, larger values of ρ are better, because they yield smaller values of distortion, though the compression ratio decreases as ρ increases. Hence, the value of ρ should be chosen according to the requirements of the information processing application.

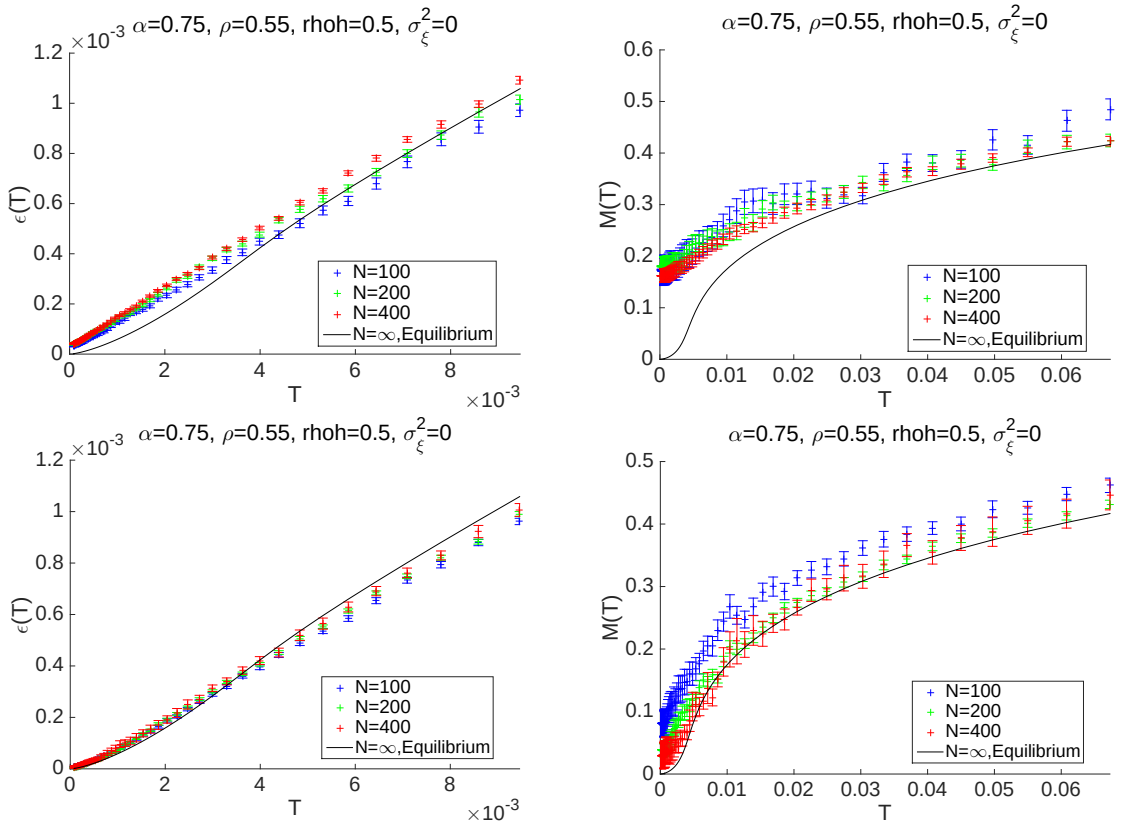

Figure 6. Distortion ϵ (left) and MSE $\mathcal{M}$ (right) plotted against temperature for $\alpha = 0.75, \rho = 0.55$, and $\hat{\rho} = 0.5$. In contrast to Fig. 4 with $\rho = 0.65$, there is no phase transition. The upper panels correspond to a rapid schedule with $\tau = 5$, whereas the lower ones correspond to $\tau = 100$. At low temperatures, the rapid case does not follow the equilibrium state, whereas the slower one is well matched.

3.2. Case without planted solutions

Next, we examine the opposite case with $\sigma_\xi^2 = 1$ and $\sigma_x^2 = 0$. There is now no planted solution, and we focus solely on how small ϵ becomes. For comparison, we present a number of values of ϵ achieved by different methods. The symbol ell$_1$ denotes the ℓ_1 method in [12], in which the ℓ_1-relaxed version of eq. (3) is solved and the resultant solution of $\boldsymbol{x}$ is inserted in eq. (4). Similarly, the symbol ell$_1$+LS corresponds to the method in [12], which gives ϵ obtained by eq. (6), but with the substituted support $\boldsymbol{c}$ determined by solving the ℓ_1-relaxed version of eq. (3). The symbol OMP denotes the results given by orthogonal matching pursuit [22, 23]. The OMP result is not obtained by analytical methods, but by numerical experiments with the same parameters as SA for $N = 400$.

Fig. 7 plots ϵ against temperature T. Note that the thermal average of the distortion stops decreasing at a certain value of T. Below this, ϵ remains constant, giving the achievable limit of the value of distortion in the present case [12], as seen in the analytical result (black solid curve). Over a longer temperature range (left panel), the numerical results seem to agree well with the analytical curve, but over a tighter range of low temperatures (right panel), the numerical data exhibit a systematic deviation from the analytical curve. This is likely to be a result of the rapid nature of our annealing schedule. A slower schedule will improve the achievable value of distortion, as in Fig. 6. Regardless, we can see that the SA result is already better than the

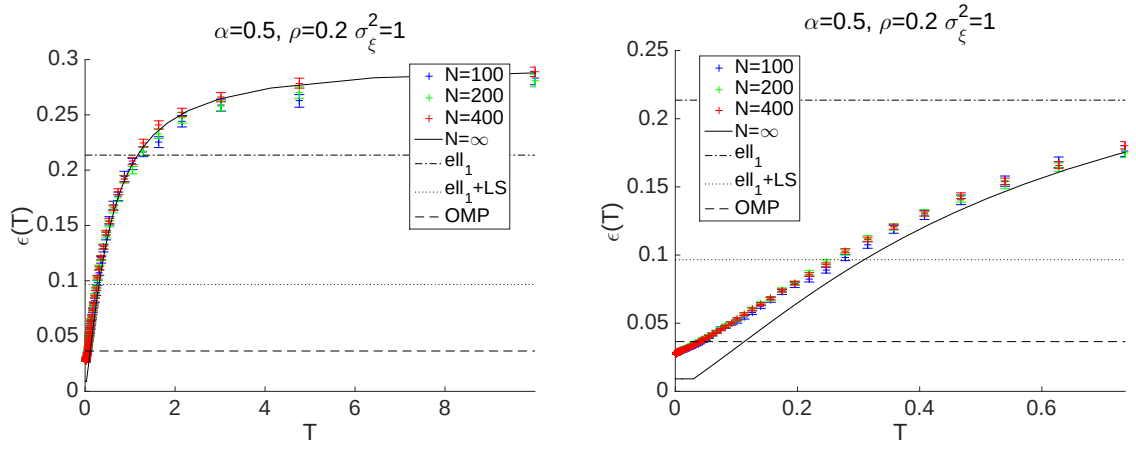

Figure 7. Distortion plotted against temperature for $\alpha = 0.5, \rho = 0.2$ for the case without a planted solution. The right panel is a magnified view of the left one at low temperatures. The black solid curve represents the analytical value of ϵ. Values of ϵ given by different algorithms are also displayed as horizontal lines with different symbols. See the main text for details.

values achieved by the other methods, demonstrating the effectiveness of SA. For reference, the distortion values obtained by ℓ_1, $\ell_1 + \mathrm{LS}$, OMP, and SA at $\mu = \mu_{100} \approx 1.3 \times 10^4$ and $N = 400$ were $\epsilon_{\ell_1} = 0.214, \epsilon_{\ell_1+\mathrm{LS}} = 0.0966, \epsilon_{\mathrm{OMP}} = 0.0365 \pm 6.7 \times 10^{-4}$, $\epsilon_{\mathrm{SA}} = 0.0272 \pm 6.2 \times 10^{-4}$, respectively. The achievable limit was $\epsilon_0 = 0.00919$.

We also examined other parameter values, but the qualitative behavior was the same as that of Fig. 7, and so the results are not shown here.

4. Discussion

4.1. Computation time and its order

For reference, we give the actual computation times for one run of SA under the schedule in eq. (12): for $\alpha = 0.5$ and $\rho = 0.2$, approximately $6, 15$, and 38 seconds were required for $N = 100, 200$, and 400, respectively. This experiment was performed on a 1.7 GHz Intel Core i7 with two CPUs using MATLAB®.

As well as these practical results, we can estimate the order of the computation time. Formally, this can be written as $O(N_\mu \tau N N_{\mathrm{MC}})$, the last factor of which is the computational cost of each MC update. The most expensive part is the matrix multiplication and inversion required to calculate the energy. If we use simple multiplication and Gauss elimination in the inversion process for each step, $N_{\mathrm{MC}} = O(M(N\rho)^2 + (N\rho)^3)$. However, we have employed pair flipping in each update of the sparse weights, meaning that the change in the relevant matrices in each move is small and successive. This implies that we can reduce the computation time by successively updating those matrices while employing the matrix inversion formula.

Provided that G_{t+1} is decomposed as

$$G_{t+1} = \begin{pmatrix} G_t & \boldsymbol{g}_{t+1} \\ \boldsymbol{g}_{t+1}^{\mathrm{T}} & g_{t+1} \end{pmatrix}, \tag{15}$$

the matrix inversion formula gives

$$G_{t+1}^{-1} = \begin{pmatrix} G_t^{-1} + \gamma_{t+1} G_t^{-1} \boldsymbol{g}_{t+1} \boldsymbol{g}_{t+1}^{\mathrm{T}} G_t^{-1} & -\gamma_{t+1} G_t^{-1} \boldsymbol{g}_{t+1} \\ (-\gamma_{t+1} G_t^{-1} \boldsymbol{g}_{t+1})^{\mathrm{T}} & \gamma_{t+1} \end{pmatrix} \equiv \begin{pmatrix} U_t & \boldsymbol{u}_{t+1} \\ \boldsymbol{u}_{t+1}^{\mathrm{T}} & u_{t+1} \end{pmatrix} \equiv U_{t+1}, \tag{16}$$

where $\gamma_{t+1} = g_{t+1} - \boldsymbol{g}_{t+1}^{\mathrm{T}} G_t^{-1} \boldsymbol{g}_{t+1}$. We use this as follows. Write G_{t+1} as $(\tilde{A}^{\mathrm{T}}(\boldsymbol{c})\tilde{A}(\boldsymbol{c}))$, and assume that we are given both G_{t+1} and $G_{t+1}^{-1} = U_{t+1}$. First, we treat the deletion part of the MC move, $c_i = 1 \to c_i' = 0$. We want to calculate G_t^{-1} from G_{t+1}^{-1}. This is given by

$$G_t^{-1} = U_t - \boldsymbol{u}_{t+1}\boldsymbol{u}_{t+1}^{\mathrm{T}}/u_{t+1}, \tag{17}$$

which has a computational cost of $O((N\rho)^2)$. The corresponding G_t is obtained by deleting the ith column and row from G_{t+1}. Next, we move to the addition part $c_j = 0 \to c_j' = 1$. We now have G_t and G_t^{-1}. Extending the matrix, $G_t \to G_{t+1}$, involves adding an appropriate column and row to G_t. The kth component of the added column vector can be calculated as $\boldsymbol{g}_{t+1}(k) = \sum_{l=1}^{M} A(l,k)A(l,j)$, where k runs over the indices of ONES, and thus the computational cost is $O(MN\rho)$. Similarly, we can calculate $g_{t+1} = \sum_{l=1}^{M} A(l,j)A(l,j)$, and hence the computational cost of calculating G_{t+1} is $O(MN\rho)$. Now, we can easily calculate G_{t+1}^{-1} by eq. (16) from G_t^{-1}, $\boldsymbol{g}_{t+1}$, and g_{t+1}. The computational cost of this operation is $O((N\rho)^2)$. This completes the successive update of $G = \tilde{A}^{\mathrm{T}}\tilde{A}$ and G^{-1}.

In summary, the factor N_{MC} can be reduced to

$$N_{\mathrm{MC}} = O((N\rho)^2 + MN\rho) = O(N^2\alpha\rho), \ (M = N\alpha \geq N\rho). \tag{18}$$

Thus, the total computational cost of our SA algorithm is $O(\alpha\rho\tau N_\mu N^3)$. As long as N_μ does not scale with the size of the system, we have only third-order dependence on system-size, which is comparable to that of versatile convex optimization solvers used in the ℓ_1-relaxed version of the sparse approximation problem. Hence, our SA algorithm can solve the sparse approximation problem at a computational cost that is of the same order as the ℓ_1-relaxed version, without any need to relax the problem.

4.2. Advantages, disadvantages, and possible extensions

Our study indicates that SA reliably determines a solution with small distortion both in the presence and absence of noise, and has a reasonable computational cost. As long as noise and the metastable state are absent, the solution identified by SA is approximately equal to the planted solution. These findings encourage the use of SA in practical applications of the sparse approximation problem.

To conclude, we summarize the advantages and disadvantages of the present SA algorithm.

Advantages:

- Easy to implement for any $\boldsymbol{y}$ and A.
- Necessarily stops (message passing can be unstable and sometimes does not converge, especially in the presence of noise).
- Solutions at finite temperatures ($\{\boldsymbol{c}_t\}_t$) can be obtained by one iteration of SA. These may be more useful than the optimal solution with minimum distortion, especially in the presence of noise.

Disadvantages:

- Emergence of metastable states in the hard parameter region.
- Presence of ρ. Tuning this parameter is not trivial: larger values of ρ are better for determining the planted solution and decreasing the level of distortion, but are more likely to yield the metastable state.

- Annealing schedule is arbitrary, and it is not *a priori* clear how long the algorithm requires to reach a desired solution.

These disadvantages may be overcome using a range of techniques. Seeding [10] is a good candidate for avoiding trapping in the metastable state. What we should do is to change the matrix A to a structured one as described in [10]. Beyond the pair flipping process, we may flip more sparse weights to generate trial moves, which should shorten the time required for efficient sampling. Multicanonical methods may also be beneficial. For example, simulating different values of ρ simultaneously enables a wider region of the phase space to be sampled, which may help escaping from the metastable state. In this way, the second disadvantage will be also diminished. Developing extensions extensions for the present formulation of SA will be helpful to control the sparse approximation problem.

Acknowledgments
This work was supported by JSPS KAKENHI Grant Numbers 26870185 (TO) and 25120013 (YK).

[1] Donoho D L 2006 *IEEE Trans. Inf. Theory* **52** 1289
[2] Candès E J and Tao T 2005 *IEEE Trans. Inf. Theory* **51** 4203
[3] Candès E J, Romberg J and Tao T 2006 *IEEE Trans. Inf. Theory* **52** 489
[4] Candès E J and Tao T 2006 *IEEE Trans. Inf. Theory* **52** 5406
[5] Donoho D L and Tanner J 2009 *J. Am. Math. Soc.* **22** 1
[6] Donoho D L, Malekib A and Montanari A 2009 *Proc. Natl. Acad. Sci.* **106** 18914
[7] Kabashima Y, Wadayama T and Tanaka T 2009 *J. Stat. Mech.* L09003
[8] Ganguli S and Sompolinsky H 2010 *Phys. Rev. Lett.* **104** 188701
[9] Rangan S 2010 *arXiv:1010.5141*
[10] Krzakala F, Mézard M, Sausset F, Sun Y and Zdeborová L 2012 *J. Stat. Mech.* P08009
[11] Sakata A and Kabashima K 2013 *Europhys. Lett.* **103**, 28008
[12] Nakanishi Y, Obuchi T, Kabashima Y and Okada M 2015 *arXiv:1510.02189*
[13] Obuchi T, Nakanishi Y, Kabashima Y and Okada M *in preparation*
[14] Natarajan B K 1995 *SIAM J. Comput.* **24** 227
[15] Temlyakov V N 1998 *Adv. Comput. Math.* **8** 249
[16] Temlyakov V N 1999 *J. Approx. Theory* **98** 117
[17] Tropp J A 2004 *IEEE Trans. Inf. Theory* **50** 2231
[18] Donoho D L, Elad M and Temlyakov V N 2006 *IEEE Trans. Inf. Theory* **51** 6
[19] Monasson R 1996 *Phys. Rev. Lett.* **76** 3881
[20] Krzakala F, Montanari A, Ricci-Tersenghi F, Semerjian G and Zdeborová L 2007 *Proc. Natl. Acad. Sci.* **104** 10318
[21] Geman S and Geman D 1984 *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721
[22] Pati Y C, Rezaiifar R and Krishnaprasad P S 1993 *Conference Record of the Twenty-Seventh Asilomar Conference on Signals, Systems and Computers* 40
[23] Davis G M, Mallat S G and Zhang Z 1994 *Opt. Eng.* **33** 2183