Invited Paper

Quantum annealing: next-generation computation and how to implement it when information is missing

Masayuki Ohzeki1a), Chako Takahashi1, Shuntaro Okada1,2, Masayoshi Terabe2, Shinichiro Taguchi2, and Kazuyuki Tanaka1

1 Graduate School of Information Sciences, Tohoku University
6-3-09, Aramaki-aza-Aoba, Aoba-ku, Sendai 980-8579, Japan
2 Electronics Research and Innovation Division, DENSO CORPORATION
Chuo, Tokyo 103-6015, Japan

a) mohzeki@tohoku.ac.jp

Received February 27, 2018; Revised June 13, 2018; Published October 1, 2018

Abstract: Recently, several powerful machines dedicated to solving combinatorial optimization problems through the Ising-model formulation have appeared. The trigger for the paradigm shift to a specialized machine for solving optimization problems was the D-Wave machine, which implements quantum annealing. Quantum annealing employs quantum fluctuations to find an optimal solution to an optimization problem with discrete variables. In particular, we input the optimization problem in the form of the Ising Hamiltonian, which is a specialized form of the quadratic unconstrained binary optimization problem. However, when we employ quantum annealing for a practical optimization problem, there are several issues. One typical issue is absence of the detailed form of the cost function, which characterizes the optimization problem. To input problems into specialized machines for solving an optimization problem, it is necessary to determine the unknown parameters within the Ising model. We propose a method to estimate the unknown parameters in the Ising Hamiltonian using compressed sensing. Furthermore, we analyze the theoretical limitations of our proposed method by employing the replica method, which is a sophisticated tool in statistical mechanics.

Key Words: combinatorial optimization problem, compressed sensing, replica analysis, \(L_1\)-norm minimization, inverse problem, quantum annealing

1. Introduction

Optimization problems are inherent to our daily life. One of the most typical examples is finding the shortest path among many candidate routes from a starting point to a destination. In recent developments in fields such as machine learning, we solve an optimization problem for a parameterized function to pursue the most appropriate form between the input and output datasets. Following the
description of the standard way of using a “computer,” we choose a suitable algorithm to solve an
optimization problem and write a long code to manipulate the algorithm implemented in the computer.
The solution of various optimization problems usually incurs a very large workload because the method
of solution depends on the structure of the problem. In order to reduce the difficulty in performing
a series of solving the optimization problems, one may employ generic solvers covering with many
types of them. One of the generic solvers, which attracts attentions from various fields, is quantum
annealing (QA) [1]. Quantum annealing uses quantum fluctuation to search for the optimal solution.
In quantum mechanics, the state specified by the physical quantity as the position and momentum can
be attained following probabilistic way. The probability is estimated by the squared magnitude of the
amplitude of the corresponding state. The amplitude follows the Schrödinger equation, which governs
the quantum mechanics. The dynamics of the Schrödinger equation allows the quantum tunneling
between local minimum over intermediate wall of the potential energy, which is essentially different
from the standard description of motion, Newtonian motion of equation. If we utilize the quantum
tunneling effect, we would be able to find the optimal solution efficiently.

The prescription of QA is very straightforward. We encode an optimization problem to be solved
into a special functional form with a quadratic unconstrained binary function known as the Ising
Hamiltonian (energy function). Without choosing a specific algorithm to solve an encoded optimiza-
tion problem, the scheme of QA automatically solves it by using quantum fluctuations. Quantum
fluctuations, which are inherent in the extremely low-temperature region, drive the system as thermal
fluctuations. In the optimization problem, the variables can be regarded as the degrees of freedom
moving in the potential energy, which represents the structure of the cost function to be optimized.
The thermal fluctuations drive the degrees of freedom in a stochastic way. As a result, the degrees of
freedom can climb against the gradient of the potential energy while searching for another possibly
better solution to the optimization problem. The method for solving an optimization problem is called
simulated annealing [11]. The quantum fluctuations also play the role of a driver of the degrees of
freedom. In addition, a remarkable property of the quantum fluctuations induces the tunneling effect,
which creates a path through the potential energy barrier between local minima. Therefore, we many
naturally expect that the quantum fluctuations can increase the performance for finding the global
optimal solution rather than the thermal fluctuations.

As a technique for solving an optimization problem, QA is very simple after we encode the optimiza-
tion problem into the Ising Hamiltonian. However, when we do not know the detailed structure of the
cost function, we need to infer it from a small number of observations. We assume the case in which
we can attain a small number of pairs of energy values and binary-variable configurations. Then, we
successfully obtain the cost function without a vast number of observations of its structure. The key
point is sparseness, which enables us to determine a unique and correct solution without a sufficient
number of observations in a usual sense. The inference technique with sparseness is called the
compressed sensing (CS) [6]. The use of the sparsity promotes data analysis in various scientific
fields [7–10]. In the present paper, we also report a recent analysis of the theoretical assurance of
the performance for inferring the cost function and checking its validity in numerical simulations.
The results are not restricted to the application of QA for solving the optimization but also that of
simulated annealing (SA), which utilizes the thermal fluctuation for searching the optimal solution,
and more [11].

The remaining part of the present paper is organized as follows. In Sect. 2, we review the formulation
of QA. In Sect. 3, we formulate the problem of Hamiltonian estimation and propose a method for
solving this problem using CS. Moreover, we analyze the typical performance of our proposed method
utilizing the replica method. In Sect. 4, we summarize the present paper.

2. Quantum annealing

2.1 Basics

We first review quantum mechanics. In quantum mechanics, we utilize a state vector to identify
the current state $|\Psi(t)\rangle$ at a time $t$ related to the microscopic degrees of freedom as the position
and momentum of a particle\(^1\). The state vector leads to a probabilistic amplitude of the physical observable and is evolved by the Schrödinger equation as

\[
 i\hbar \frac{\partial}{\partial t} |\Psi(t)\rangle = \hat{H}(t) |\Psi(t)\rangle,
\]

where \(i\) is the imaginary unit, and \(\hbar\) is the Planck’s constant divided by \(2\pi\), which will be unity for simplicity below. The Hamiltonian \(\hat{H}(t)\) is time-dependent and given by the kinetic energy and potential energy in the system under consideration. For instance, we employ the following form of the Hamiltonian to investigate microscopic particles having the potential energy \(U(\hat{x})\)

\[
 \hat{H}(t) = \frac{\hat{p}^2}{2m} + U(\hat{x}),
\]

where \(m\) is the mass, \(\hat{p}\) is the momentum, and \(\hat{x}\) is the location of a particle. Here, the hat (\(^\hat{\cdot}\)) notation represents an operator for distinction from the number. In quantum mechanics, since operators may have a commutation relation such as

\[
 [\hat{x}, \hat{p}] = \hat{x}\hat{p} - \hat{p}\hat{x} = i\hbar,
\]

quantum fluctuations emerge. One can prove the (Kennard) uncertainty relation between the momentum and location as

\[
 \Delta x \Delta p \geq \frac{\hbar}{2},
\]

where \(\Delta x = \sqrt{\langle \hat{x}^2 \rangle - \langle \hat{x} \rangle^2}\), \(\Delta p = \sqrt{\langle \hat{p}^2 \rangle - \langle \hat{p} \rangle^2}\), where \(\langle \cdots \rangle\) denotes the expectation of an observable determined by the state vector. The uncertainty relation leads to quantum fluctuations, which means that an observation can change in repeated experiments.

**2.2 Procedure of quantum annealing**

In QA, we encode the optimization problem to be solved into the Ising Hamiltonian as

\[
 \hat{H}_0(\sigma^z) = - \sum_{i \neq j} J_{ij} \hat{\sigma}^z_i \hat{\sigma}^z_j,
\]

where \(J_{ij}\) is a coupling constant describing the interaction between Ising spins, and \(\hat{\sigma}^z_i\) is the Ising spin represented by the \(z\) component of the Pauli matrices

\[
 \hat{\sigma}^z_i = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},
\]

\[
 \hat{\sigma}^y_i = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix},
\]

\[
 \hat{\sigma}^x_i = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.
\]

The Pauli matrices can represent an arbitrary transformation in a two-level state vector. The two-level state vector is given by the eigenstate of the \(z\) component of the Pauli matrices as \(\hat{\sigma}^z_i |\uparrow\rangle_i = |\uparrow\rangle_i\) and \(\hat{\sigma}^z_i |\downarrow\rangle_i = -|\downarrow\rangle_i\).

We call the Ising Hamiltonian in Eq. (5) the target Hamiltonian, which stands for the combinatorial optimization problem to be solved. It is well-known that combinatorial optimization is difficult to solve. In addition, we need a specific algorithm to solve each optimization problem depending on the structure of the problem. Instead, in QA, we solve all types of combinatorial optimization problems in a generic way. To drive two-level systems in contact with each other, we introduce quantum fluctuations. As is shown easily, the Pauli matrices satisfy the commutation relation as

\[
 [\hat{\sigma}^z_i, \hat{\sigma}^z_j] = 2i\hat{\sigma}^y_i \delta_{ij},
\]

\(^1\)Most researchers in physics utilize the \(ket\) vector to represent the state vector.
\[ \hat{H}(t) = f(t)\hat{H}_0(\hat{\sigma}^z) + (1 - f(t))\hat{H}_1, \]

where we usually employ the simplest quantum driver Hamiltonian known as the transverse field:

\[ \hat{H}_1 = -\Gamma \sum_{i=1}^{N} \hat{\sigma}_i^x, \]

where \( \Gamma \) is a transverse field inducing the quantum fluctuation and \( N \) is the number of the Ising spins. The quantum driver Hamiltonian is not commutable with the target Hamiltonian. In other words, when we employ the diagonal representation of the target Hamiltonian, the quantum driver Hamiltonian has off-diagonal elements. Recently, researchers have become interested in the possibility of improving the performance of QA by using different types of quantum driver Hamiltonians. In QA, we tune the strength of the interactions and quantum fluctuations through \( f(t) \), which is a monotonically increasing function from 0 to 1. The typical choice is \( f(t) = t/\tau \), where \( \tau \) is the annealing time. When we take large \( \tau \), the quantum adiabatic theorem ensures that the instantaneous state vector retains the lowest energy eigenstate (ground state) when we start from the lowest energy eigenstate of the initial Hamiltonian \( \hat{H}(0) \). The ground state of the initial Hamiltonian \( |0(t=0)\rangle \) is very simple and is given as

\[ |0(t=0)\rangle = \frac{1}{\sqrt{2^N}} \otimes_{i=1}^{N} (|\uparrow_i\rangle + |\downarrow_i\rangle). \]

The ground state of the initial Hamiltonian can be imagined as the superposition of binary candidates of the best solution to the optimization problem encoded into the Ising Hamiltonian. When we successfully encode the original optimization problem into the Ising Hamiltonian such that the cost function of the optimization problem corresponds to the energy value of the Ising Hamiltonian, namely \( \hat{H}_0 \), the quantum adiabatic theorem states that the slow tuning of the external parameter as the transverse field leads to the optimal solution of the Ising Hamiltonian at the final stage \( t = \tau \). In our formulation, we control \( f(t) \) slowly. This is the definition of QA in a narrow sense. This scheme is also called quantum adiabatic computation [12]. In a broader sense, the original concept of QA is not necessarily restricted to the case in the quantum adiabatic theorem [13–16]. However, the mathematical assurance of obtaining the ground state of the Ising Hamiltonian exists only in the case of slow tuning of the parameters. Below, we focus on QA in the adiabatic case.

### 2.3 Quantum computation?

In recent years, QA has been experimentally realized by setting an array of two-level state vectors and the interactions on the designed lattice, which can be realized by the use of artificially manufactured atoms or spins by employing a macroscopic quantum effect such as superconductivity [17–20]. The D-Wave machine is the first realization of the quantum device with the manufactured spins. The newest version of the D-Wave machine implements 2000 qubits, which are two-level state vectors with quantum fluctuations. The interactions are set between qubits on the edges of a special graph known as the chimera graph, which does not have all-to-all connections between nodes but very sparse edges, i.e., a very sparse number of edges. Therefore, the D-Wave machine cannot implement arbitrary optimization problems natively. Some techniques translate the original optimization problem into an effective optimization problem while using redundant qubits. As a result, the computational resources required to solve the original optimization problem can be larger depending on its structure.

The theoretical arguments for the computational complexity of QA have been carried out by considering the quantum adiabatic theorem. The computational complexity is given by the inverse square of the energy gap (the difference in the eigenvalues) between the ground state and the second-lowest eigenstates [21]. One may expect that the computational complexity is reduced to a moderate value in comparison to the case of classical computation without quantum-mechanical effects. Unfortunately,
there is no finding of a (exponential) quantum speedup with certain evidence of the standard way of QA. One of the (non-exponential) quantum speedups is Grover’s search algorithm to find a single specific item from many candidates [22]. The algorithm can be implemented in the form of QA, but the quantum fluctuations are not very simple and consist of multiple interactions among qubits.

In addition, QA can solve the integer factorization problem similarly to Shor’s algorithm [23], which is the most significant instance for demonstrating computation by using quantum-mechanical effects. However, the realization of the integer factorization with QA demands a tricky quantum fluctuations consisting of multiple interactions and a number of qubits. The trick includes a nonstoquastic Hamiltonian beyond the stoquastic one as the transverse field [24, 25]. A class of quantum systems can be simulated in a stochastic manner on a standard computer. The stoquastic Hamiltonian means that the quantum system can be simulated efficiently in a straightforward way. For instance, the quantum Monte Carlo simulation is one of the tools for simulating a quantum system [26]. However, the nonstoquastic Hamiltonian is not, due to the appearance of a negative sign in the corresponding stochastic model. In this sense, a study of the nonstoquastic Hamiltonian is needed to clarify the boundary between classically simulatable and not.

3. Hamiltonian estimation

3.1 Unknown Hamiltonian

In this section, we present a recent study that performed QA in a realistic situation. We assume that we do not know the detailed structure of the cost function to be optimized. An oracle serves as an observation matrix of spin configurations $S$.

We then obtain a set of $M$ ($< N(N-1)/2$) values of the energy $E := (E^{(1)}, E^{(2)}, \ldots, E^{(M)})^T$ and an observation matrix of spin configurations $S$, which is constructed by the given spin configurations. Here, $S$ is an $M \times N(N-1)/2$ matrix expressed as

$$S = \begin{pmatrix}
\sigma_1^{(1)} \sigma_2^{(1)} & \cdots & \sigma_1^{(1)} \sigma_N^{(1)} & \sigma_2^{(1)} \sigma_3^{(1)} & \cdots & \sigma_1^{(1)} \sigma_{N-1}^{(1)} \\
\sigma_1^{(2)} \sigma_2^{(2)} & \cdots & \sigma_1^{(2)} \sigma_N^{(2)} & \sigma_2^{(2)} \sigma_3^{(2)} & \cdots & \sigma_1^{(2)} \sigma_{N-1}^{(2)} \\
\vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\
\sigma_1^{(M)} \sigma_2^{(M)} & \cdots & \sigma_1^{(M)} \sigma_N^{(M)} & \sigma_2^{(M)} \sigma_3^{(M)} & \cdots & \sigma_1^{(M)} \sigma_{N-1}^{(M)}
\end{pmatrix}.
$$

We assume that a set of $N(N-1)/2$ original coupling constants is expressed as $J^0 := (J_{12}^0, \ldots, J_{1N}^0, J_{23}^0, \ldots, J_{N-1,N}^0)^T$, and a set of energies $E$ is expressed as

$$E := -\frac{1}{N} S J^0.
$$
In general, the number of unknown variables $J^0$ is larger than the number of linear equations, as Eq. (14) is an underdetermined system. When we simply solve Eq. (14), we cannot obtain a unique solution. Hence, a method that can adequately estimate the unknown $J$ from only the given $M$ pairs of energies and spin configurations is desired.

### 3.2 $L_1$-norm minimization

To overcome the above difficulty, we utilize the concept of CS [6]. We impose an assumption of sparsity on $J^0$. In particular, we assume that the probability distribution of $J^0_{ij}$ is

$$P(J^0_{ij}) := (1 − \rho)\delta(J^0_{ij}) + \rho\mathcal{N}(0, 1),$$

where $\delta(x)$ denotes the Dirac delta function, $\rho = 2K/N(N−1)$ ($0 \leq \rho \leq 1$) denotes the density of the nonzero components in the vector $J^0$, and $\mathcal{N}(0, 1)$ denotes the Gaussian distribution with a vanishing mean and unit variance. Note that we do not know which component in $J^0$ is nonzero or zero.

We then formulate the estimate of $J^0$ as the minimization of the $L_1$ norm with respect to $J := (J_{12}, \cdots, J_{1N}, J_{23}, \cdots, J_{N−1,N})^T$ on the basis of the concept of CS. $L_1$-norm minimization is the simplest method in CS. It is tractable since many algorithms to solve $L_1$-norm minimization have been proposed, moreover, theoretical performance analysis of $L_1$-norm minimization is possible. $L_1$-norm minimization in our problem setting is formulated as follows:

$$\min_{J} \|J\|_1 \quad \text{subject to} \quad E = -\frac{1}{N}SJ.$$  

The solution of Eq. (16) under the constraint in Eq. (15) yields an adequate estimate of the unknown $J^0$ under several conditions with respect to $\rho$ and the ratio of the observations to the number of the coupling constants $\alpha = 2M/N(N−1)$. The problem in Eq. (16) can be solved by using various optimization algorithms such as coordinate descent [40], fast iterative shrinkage-thresholding algorithms [43], the augmented Lagrangian method [41, 42], or the alternating direction method of multipliers [29, 37].

### 3.3 Replica analysis

We formulated the $L_1$-norm minimization problem in our problem setting. What we are interested in is how the estimation performance using our method depends on $\rho$ and $\alpha$. In this section, we investigate the typical behavior of our method proposed in Sect. 3.2. To analyze our method, we employ the replica method [38, 39]. The replica method has been developing in the field of statistical physics and is frequently used to analyze the free energy, which is defined through a partition function as the normalization constant of the distribution and is difficult to calculate analytically in most cases. The replica method gives an analytical representation of the configurational average of a partition function instead of the partition function itself based on the self-averaging property. A series of the procedure on the replica method is a sophisticated tool to investigate the practical performance of various algorithm appearing in information sciences [30]. Most of the significant contribution were on the combinatorial optimization problems [31], satisfiability problem [32], wireless communication [33, 34]. Recently the analysis clarifies several techniques in machine learning such as the semi-supervised learning [35], pretraining and fine tuning [36] in deep learning.

The typical behavior of the estimation of the coupling constants described in Eq. (16) can be investigated via analyzing the partition function of the following distribution by using the replica method:

$$P(J \mid E) = \frac{1}{Z_\beta(E)}P(E \mid J)P(J),$$

where $Z_\beta(E) := \int dJP(E \mid J)P(J)$ is the partition function of this distribution and $\beta$ is the inverse temperature that is introduced to extend the problem to finite temperatures. The logarithm of $Z_\beta(E)$ is an important quantity, namely, the free energy. The free energy of the system is defined as
\[ f = - \lim_{\beta \to \infty} \lim_{N \to \infty} \frac{1}{\beta N^2} \ln Z_\beta(E)_{S, J^0} \]
\[ = - \lim_{\beta \to \infty} \lim_{n \to 0} \frac{\partial}{\partial n} \ln \frac{1}{\beta N^2} \ln[Z_\beta^n(E)]_{S, J^0}, \] (18)

where \( \cdots \) \(S, J^0\) denotes the configurational average with respect to \(S\) and \(J^0\). In the limit \(\beta \to \infty\), the solution of Eq. (16) corresponds to the maximum a posteriori estimator of Eq. (16). Therefore, the performance of the \(L_1\)-norm minimization in Eq. (16) can be evaluated via analyzing the macroscopic behavior of the free energy in the limit \(\beta \to \infty\). In accordance with the procedure of the general replica analysis \([28]\), we evaluate \([Z_\beta^n(E)]_{S, J^0}\) in Eq. (18) for \(n \in \mathbb{N}\). Furthermore, \([Z_\beta^n(E)]_{S, J^0}\) is performed with an analytical continuation for \(n \in \mathbb{R}\) after completing the computation of the partition function into a closed form and obtaining a function form of \(n\).

By using the expression \(E^{(n)} := -\sum_{i<j} J_{ij}^{(n)} \sigma_i^{(\mu)} \sigma_j^{(\mu)} / N\), \([Z_\beta^n(E)]_{S, J^0}\) is expressed as
\[ [Z_\beta^n(E)]_{S, J^0} = \left[ \prod_{a=1}^n \int dJ^a \delta(S(J^a - J^0)) \exp(-\beta\|J^a\|_1) \right]_{S, J^0}, \] (19)

where \(J^a\) denotes the \(a\)th replicated vector of the coupling constants, and \(\|J^a\|_1\) denotes the \(L_1\) norm of \(J^a\). Here, we assumed \(P(J^a) \forall a \in \{1, 2, \cdots, n\}\) to be the Laplace distribution: \(P(J^a) \propto \exp(-\beta\|J^a\|_1)\). Note that Eq. (19) is only valid \(\forall n \in \mathbb{N}\). To obtain an analytical expression for part of Eq. (19), \(\prod_{a=1}^n \delta(S(J^a - J^0))\) in Eq. (19), we define the following quantity:
\[ u_{i, a}^{\mu} := \frac{1}{N} \sum_{i<j} (J_{ij}^0 - J_{ij}) \sigma_i^{(\mu)} \sigma_j^{(\mu)}, \] (20)

where \(a = 0, 1, 2, \cdots, n\) is the index of the replica. Here, we assume that
\[ \langle \sigma_i^{(\mu)} \sigma_j^{(\mu)} \rangle_{\sigma^{(\mu)}} = \sigma_i^{(\mu)} \sigma_j^{(\mu)} = \delta_{i j}, \] (21)
\[ = \begin{cases} \sigma_i^{(\mu)} \sigma_j^{(\mu)} & (i \neq j \neq k \neq l) \\ 1 & (i = k, j = l; i = l, j = k), \end{cases} \] (22)

where \(\cdots\) denotes \(\sum_{a} (-1)^{a} P(x)\), and \(c (-1 \leq c \leq 1)\) is a constant. Then, we consider the case when \(c = 0\), i.e., \((\sigma_i^{(\mu)})_{\sigma^{(\mu)}} = 0\). Eq. (21) and Eq. (22) are rewritten as
\[ \langle \sigma_i^{(\mu)} \sigma_j^{(\mu)} \rangle_{\sigma^{(\mu)}} = 0, \] (23)
\[ \langle \sigma_i^{(\mu)} \sigma_j^{(\mu)} \rangle_{\sigma_k^{(\mu)}} = \delta_{i k} \delta_{j l} + \delta_{i l} \delta_{j k}, \] (24)

respectively. The assumptions in Eqs. (21)–(24) are unique characteristics in the models with Ising variables. From the assumptions in Eqs. (21)–(24), \(u_{i, a}^{(\mu)}\) \(\sigma^{(\mu)} = 0\) is obtained. Similarly, we obtain \(\langle u_{i, a}^{(\mu)} u_{j, b}^{(\mu)} \rangle_{\sigma^{(\mu)}} = 2(q^{ab} - q^{ba} - q^{aa} + q^{bb})\), where
\[ q^{ab} := \frac{1}{N^2} \sum_{i<j} J_{ij}^a J_{ij}^b \] (25)

are the order parameters with respect to the coupling constants. Herein, we assume the following replica symmetric ansatz to extend Eq. (19) to \(n \in \mathbb{R}\):
\[ q^{ab} = \begin{cases} \rho & (a = b = 0) \\ m & (a \neq 0, b = 0; a = 0, b \neq 0) \\ q & (a \neq 0, b \neq 0; a \neq 0, b \neq 0) \end{cases} \] (26)

Taking into account Eq. (26), \(u_{i, a}^{(\mu)}\) can be represented as \((n+1)\) multivariate Gaussian random variables with a vanishing mean and covariance \(\langle u_{i, a}^{(\mu)} u_{j, b}^{(\mu)} \rangle_{\sigma^{(\mu)}} = 2(\rho - 2m + Q)\) for \(a = b\) and \(\langle u_{i, a}^{(\mu)} u_{j, b}^{(\mu)} \rangle_{\sigma^{(\mu)}} = 2(\rho - 2m + q)\) for \(a \neq b\). \(u_{i, a}^{(\mu)}\) in Eq. (20) can be expressed in the new form
$u^a_\mu = \sqrt{Q - qs_a} + \sqrt{\rho - 2m + qt}$, (27)

where $s_a$ and $t$ are the Gaussian random variables with vanishing mean and unit variance. By using the representation in Eq. (27), the following expression is obtained:

$$\left[ \prod_{a=1}^n \delta(S(J^a - J^0)) \right]_S = \lim_{\tau \to 0} \prod_{a=1}^n \prod_{\mu=1}^M \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{(u^a_\mu)^2}{2\tau}\right) \approx \exp\frac{nM}{2} \left( \ln(Q - q) - \frac{\rho - 2m + q}{Q - q} \right),$$ (28)

where $\cdots_\mu$ denotes $\int dt \int Ds_a(\cdots)$, and $\int dx := \int dx \exp(-x^2/2)/\sqrt{2\pi}$. Here, we used $\exp(nx) = 1 + nx + O(x^2)$ and $\delta(x) = \lim_{\tau \to 0} \exp(-x^2/2\tau)/\sqrt{2\pi\tau}$. See Appendix A for the detailed form of $[Z_\beta^\alpha(E)]_{s,t}$ that is introduced in the results up to Eq. (28) and several other expressions.

From the replica symmetric ansatz in Eq. (26) and the saddle-point approximations in Appendix B, the right-hand side of Eq. (18) is explicitly expressed as

$$f = \operatorname{extr}_{Q,\chi,m,Q,\chi,\bar{m}} \left\{ \frac{\alpha(\rho - 2m + Q)}{2\chi} - \frac{1}{2}(Q\tilde{Q} - \chi\tilde{\chi}) + m\bar{m} + (1 - \rho) \int dt \Phi(t; \tilde{Q}, \tilde{\chi}, 0) + \rho \int dt \Phi(t; \tilde{Q}, \tilde{\chi}, \bar{m}) \right\},$$ (29)

where $\tilde{Q}$, $\tilde{\chi}$, and $\bar{m}$ are auxiliary parameters for introducing the integral expressions of the Kronecker delta for the replica symmetric solutions $Q$, $\chi$, and $m$, respectively. Here, we define

$$\Phi(t; \tilde{Q}, \tilde{\chi}, \bar{m}) := -\frac{1}{2\tilde{Q}} \left( |\sqrt{\tilde{\chi} + \bar{m}^2}t| - 1 \right)^2 \Theta \left( |\sqrt{\tilde{\chi} + \bar{m}^2}t| - 1 \right),$$ (30)

where

$$\Theta(x) := \begin{cases} 1 & (x > 0) \\ 0 & (x \leq 0). \end{cases}$$

See Appendix B for an overview of a more detailed derivation of Eq. (29).

Extremization of Eq. (29) yields the following saddle-point equations:

$$\tilde{Q} = \frac{\alpha}{\tilde{\chi}}, \quad \tilde{\chi} = \frac{\alpha}{\tilde{\chi}^2}(\rho - 2m + Q), \quad \bar{m} = \frac{\alpha}{\tilde{\chi}},$$

$$Q = \frac{2(1 - \rho)}{Q^2} G(\tilde{\chi}) + \frac{2\rho}{Q^2} G(\tilde{\chi} + \bar{m}),$$

$$\chi = \frac{2(1 - \rho)}{Q} H \left( \frac{1}{\sqrt{\tilde{\chi}}} \right) + \frac{2\rho}{Q} H \left( \frac{1}{\sqrt{\tilde{\chi} + \bar{m}}} \right),$$

$$m = \frac{2\rho\bar{m}}{Q} H \left( \frac{1}{\sqrt{\tilde{\chi} + \bar{m}}} \right),$$ (31)

where we define $H(x) := \int_x^\infty dt$ and $G(x) := (x + 1)H(1/\sqrt{x}) - \sqrt{x}/2\pi \exp(-1/2x)$. The Eqs. (29) and (31) coincide with the results in Ref. [28]. Therefore, the stability condition for the successful estimation via the $L_1$-norm minimization in our problem setting can be written as

$$\alpha > 2(1 - \rho)H \left( \frac{1}{\sqrt{\tilde{\chi}}} \right) + \rho,$$ (32)

similar to Ref. [28].

We can also obtain an analytical expression for the typical value of the mean squared error (MSE), which is often used as a performance index in various fields. By using the extremum solutions in Eq. (31), the MSE is expressed as
Fig. 1. (color online) Typical reconstruction limit of the $L_1$-norm minimization obtained from the replica analysis. The blue curve shows the boundary $\alpha = 2(1 - \rho)H(1/\sqrt{\chi}) + \rho$. The region colored white to black via red shows the values of the MSE in Eq. (33), which indicates whether estimation succeeds or fails. The white blocks represent regions of successful estimation. On the other hand, the black blocks show the regions where estimation fails.

$$[\text{MSE}]_{S,J_0} = \frac{2}{N(N-1)} \left[ \langle \| J - J^0 \|_2 \rangle_{J^0 | E} \right]_{S,J_0} = \rho - 2m + Q, \quad (33)$$

where $\| x \|_2$ denotes the $l_2$-norm $\| x \|_2 := (\sum_i x_i^2)^{1/2}$ for $x = \{x_i\}$, and $\langle \cdots \rangle_{x - \infty}$ denotes the average with respect to $P(x)$ as $\beta \to \infty$.

Figure 1 shows the results of our analysis. The blue curve shows the theoretical reconstruction limit of the true coupling constants estimated by $L_1$-norm minimization, expressed as $\alpha = 2(1 - \rho)H(1/\sqrt{\chi}) + \rho$. The heat map in Fig. 1 shows the expectation of the MSE described in Eq. (33). The region colored white to black via red shows the values of the MSE in Eq. (33), which indicates whether estimation succeeds or fails. The expectation of the MSE is calculated by using $Q$ and $m$, which are obtained by updating $Q$, $m$, $\chi$, $\tilde{Q}$, $\tilde{m}$, and $\tilde{\chi}$ using Eq. (31) until $(x^{k+1} - x^k)$ reaches a value less than $10^{-10}$ for $\forall x \in \{Q, m, \chi, \tilde{Q}, \tilde{m}, \tilde{\chi}\}$. The values of $\rho$ and $\alpha$ are varied in steps of $0.02$ in the ranges of $0.02 \leq \rho \leq 1$ and $0.02 \leq \alpha \leq 1$. The averages of the MSEs of 100 trials for each $\rho$ and $\alpha$ are plotted in each block of the heat map. There are $50 \times 50 = 2500$ blocks in the map. The white blocks represent regions where estimation succeeds. Here, we judge that estimation succeeds when the MSE in Eq. (33) reaches a value less than $10^{-3}$. On the other hand, the black blocks indicate regions where estimation fails. The typical performance can be achieved by adequately solving the $L_1$-norm minimization problem. We report numerical experiments elsewhere in a separate paper.

4. Discussion

In the present paper, we reviewed QA, a generic solver for optimization problems that uses quantum fluctuations. The protocol of QA can be implemented in a circuit of superconducting qubits. However, standard QA employs the transverse field as the quantum fluctuations. The transverse field is in the stoquastic Hamiltonian, which does not show the extraordinary performance of quantum computation. In recent developments in manufacturing, there are several challenges to the implementation of the nonstoquastic Hamiltonian to show nontrivial behavior appearing beyond the limit of the classical simulatable region. While expecting the development of the level to control the quantum-mechanical device, it is important to investigate the boundary between the classical simulatable region and beyond. In addition, we presented a recently proposed method for estimating the unknown coupling constants of the target Hamiltonian using the CS technique. Moreover, we analytically investigated the typical performance of our proposed method using the replica method. Owing to the limitations of hardware designs, the implementation of an optimization problem is not an easy task [46]. This is because a graph on the hardware has few edges between nodes, but the optimization problem has dense connections between binary variables. For instance, the D-Wave machine is restricted to a chimera graph, which has a sparse number of edges. Thus, various techniques are desirable for making the
structure of the optimization problem sparse without losing the important features of the original one. In the present study, we propose the \( L_1 \)-norm minimization method for Hamiltonian estimation when the coupling constants in the Hamiltonian are sparse. The Hamiltonian with sparse coupling constants is simpler to implement on the designed hardware in general. Moreover, we consider the case where we can ask for the value of the cost function for the oracle by sending the spin configuration. Then, the spin configuration is assumed to be completely random. However, it is more efficient to send a spin configuration that is different from the previously sent configurations. In other words, orthogonal observation should be valuable for inferring the Hamiltonian from the pairs of spin configurations and the value of the cost function. In the future, we will assess the theoretical performance when \( S \) changes and design the structure of \( S \) to attain a better estimate of the Hamiltonian. Our result is not only applicable to the case of QA but also to SA. The inference of the cost function will be a practical issue as the generic solver as QA and SA on the Ising formulation.

### Acknowledgments

One of the authors (M. O.) was partially supported by Inamori Foundation, KAKENHI Nos. 15H03699, 16H04832, and 16K13849, and JST-START. One of the authors (C. T.) was partially supported by a Grant-in-Aid for JSPS Fellows from the Japan Society for the Promotion of Science Grant (No. JP17J03081), and JST-CREST (No. JPMJCR1402). The authors thank Shu Tanaka for fruitful discussions.

### Appendix

#### A. Detailed form of \([Z^n_\beta(E)]_{S,J^0}\) at Eq. (28)

In this appendix, we supplement the derivation of \([Z^\beta_\beta(E)]_{S,J^0}\) before applying the saddle-point approximations with several formulas. To specify the introduction of the order parameters \(q^{ab}\) in Eq. (25), we insert 1 = \(\prod_{a,b} dq^{ab} \delta(q^{ab} - \sum_{i<j} J^a_{ij} J^b_{ij} / N^2)\) into Eq. (19). \([Z^\beta_\beta(E)]_{S,J^0}\) can be rewritten as

\[
[Z^\beta_\beta(E)]_{S,J^0} = \prod_{a,b} \int dq^{ab} \prod_{a=1}^{n} \prod_{i<j} dJ^a_{ij} \left( \lim_{t \to 0} \prod_{a=1}^{M} \prod_{\mu=1}^{1} \frac{1}{\sqrt{2\pi\tau}} \exp \left( -\frac{(u^\mu)^2}{2\tau} \right) \right)_u \times \prod_{a,b} \delta \left( q^{ab} - \frac{1}{N^2} \sum_{i<j} J^a_{ij} J^b_{ij} \right) \exp \left( -\beta \sum_{i<j} |J^a_{ij}| \right)_{J^0},
\]

(A-1)

where an analytical expression for part of the right-hand side was given in Eq. (28). We then consider rewriting \(\prod_{a,b} \int dq^{ab} \delta(q^{ab} - \sum_{i<j} J^a_{ij} J^b_{ij} / N^2)\) in more detail. From the replica symmetric ansatz in Eq. (26), we obtain \(\prod_{a,b} \int dq^{ab} \delta(q^{ab} - \sum_{i<j} J^a_{ij} J^b_{ij} / N^2) = \prod_{a,b} \int dQ \int d\tilde{Q} dm \delta(Q - \sum_{i<j} J^a_{ij} J^b_{ij} / N^2) \delta(\tilde{Q} - \sum_{i<j} J^a_{ij} J^b_{ij} / N^2)\delta(m - \sum_{i<j} J^a_{ij} J^b_{ij} / N^2)\). Furthermore, the following expressions are introduced:

\[
\delta \left( Q - \frac{1}{N^2} \sum_{i<j} J^a_{ij} J^a_{ij} \right) = \int d\tilde{Q} \exp \left( \frac{\tilde{Q}}{2} \left( N^2 Q - \sum_{i<j} J^a_{ij} J^0_{ij} \right) \right),
\]

(A-2)

\[
\delta \left( q - \frac{1}{N^2} \sum_{i<j} J^a_{ij} J^b_{ij} \right) = \int d\tilde{q} \exp \left( \frac{\tilde{q}}{2} \left( N^2 q - \sum_{i<j} J^a_{ij} J^b_{ij} \right) \right),
\]

(A-3)

\[
\delta \left( m - \frac{1}{N^2} \sum_{i<j} J^a_{ij} J^0_{ij} \right) = \int d\tilde{m} \exp \left( -\tilde{m} \left( N^2 m - \sum_{i<j} J^a_{ij} J^0_{ij} \right) \right),
\]

(A-4)

where \(\tilde{Q}, \tilde{q},,\) and \(\tilde{m}\) are auxiliary parameters for introducing the above integral representations of the Dirac delta functions. Here, \(\int \) in Eqs. (A-2)–(A-4) denotes \(\int_{-\infty}^{+\infty}\). By employing the expressions in Eq. (28) and Eqs. (A-2)–(A-4) and the Hubbard–Stratonovich transformation,
\[ \prod_{a \neq b} \exp \left( \frac{q}{2} J_{ij}^a J_{ij}^b \right) = \int D\bar{t} \prod_{a=1}^{n} \exp \left( \sqrt{\bar{q}} J_{ij}^a - \frac{q}{2} (J_{ij}^a)^2 \right), \]

Equation (A-1) can be expressed as

\[ [Z^0_\beta(E)]_{S,J^0} = \int dQ \int dq \int dm \int d\bar{Q} \int d\bar{q} \int d\bar{m} \times \left( \exp nN^2 \left( \frac{1}{2} (\bar{Q}Q + \bar{q}q) - \bar{m}m + \frac{\alpha}{2} \left( \ln(Q - q) - \frac{\rho - 2m + q}{Q - q} \right) \right) \right) \times \left[ \int D\bar{t} \prod_{a=1}^{n} \prod_{i<j} dJ_{ij}^a \exp \left( -\frac{\bar{Q} + \bar{q}}{2} (J_{ij}^a)^2 + (\bar{m} J_{ij}^0 + \sqrt{\bar{q}} \bar{t}) J_{ij}^a - \beta |J_{ij}| \right) \right]_{J^0}. \]

Note that we approximate \( M \) by \( \alpha N^2 \) in Eq. (28).

**B. Detailed derivation of Eq. (29)**

We have developed an expression for \([Z^0_\beta(E)]_{S,J^0}\) in Appendix A. However, Eq. (A-5) still has many integrations. In this appendix, we explain the procedure of the saddle-point method.

First, we consider the saddle-point approximations for the integrations with respect to \( Q, q, m, \bar{Q}, \bar{q}, \) and \( \bar{m} \). The part of Eq. (A-5) can be rewritten as

\[ \left[ \int D\bar{t} \prod_{a=1}^{n} \prod_{i<j} dJ_{ij}^a \exp \left( -\frac{\bar{Q} + \bar{q}}{2} (J_{ij}^a)^2 + (\bar{m} J_{ij}^0 + \sqrt{\bar{q}} \bar{t}) J_{ij}^a - \beta |J_{ij}| \right) \right]_{J^0}, \]

where \( J \) and \( J^0 \) are univariate parameters. Here, the relation \( \prod_{i<j} \int_{a}^{b} f(x_{ij}) dx_{ij} = \prod_{i<j} \int_{a}^{b} f(x) dx \) is considered. We substitute Eq. (B-1) into Eq. (A-5) and apply the saddle-point approximation \( \int dx \exp f(x) \approx \exp f(x^*) \) for \( Q, q, m, \bar{Q}, \bar{q}, \) and \( \bar{m} \). We obtain

\[ [Z^0_\beta(E)]_{S,J^0} \approx \exp nN^2 \left( \frac{1}{2} (\bar{Q}Q + \bar{q}q) - \bar{m}m + \frac{\alpha}{2} \left( \ln(Q - q) - \frac{\rho - 2m + q}{Q - q} \right) \right) \times \left[ \int D\bar{t} \left( \int dJ \exp \left( -\frac{\bar{Q} + \bar{q}}{2} J^2 + (\bar{m} J^0 + \sqrt{\bar{q}} \bar{t}) J - \beta |J| \right) \right) \right]_{J^0}, \]

where the parameters at the saddle points are denoted as \( Q^* \rightarrow Q, q^* \rightarrow q, m^* \rightarrow m, \bar{Q}^* \rightarrow \bar{Q}, \bar{q}^* \rightarrow \bar{q}, \) and \( \bar{m}^* \rightarrow \bar{m} \).

Here, let us consider the effect of the temperature on the parameters. We assume that \( \tilde{Q} + \tilde{q} \rightarrow \beta \tilde{Q}, Q - q \rightarrow \chi/\beta, \bar{m} \rightarrow \beta \bar{m}, \) and \( \bar{q} \rightarrow \beta^2 \bar{\chi} \) as \( \beta \rightarrow \infty \). From these, we obtain

\[ \left[ \int D\bar{t} \left( \int dJ \exp \left( -\frac{\bar{Q} + \bar{q}}{2} J^2 + (\bar{m} J^0 + \sqrt{\bar{q}} \bar{t}) J - \beta |J| \right) \right) \right]_{J^0} \rightarrow \exp nN^2 \beta \left( 1 - \rho \right) D\bar{t} \Phi(t; \tilde{\bar{Q}}, \tilde{\bar{\chi}}, 0) + \rho \int D\bar{t} \Phi(t; \tilde{\bar{Q}}, \tilde{\bar{\chi}}, \bar{\bar{m}}), \]

where the transform \( \tilde{m} J^0 + \sqrt{\bar{\chi}} t \rightarrow \sqrt{\bar{\chi} + \bar{m}^2} t \) and the saddle-point approximation \( \int dJ \exp f(J) \approx \exp f(J^*) \) are applied.

By using \( \exp(nx) = 1 + nx + O(x^2) \) and \( \ln(1 + x) = x + O(x^2) \) and substituting these expressions into Eq. (18), the analytical expression of the free energy in Eq. (29) is finally obtained.

402
References


