Derivaton of QUBO formulations for sparse estimation

Tomohiro Yokota[1], Makiko Konoshima[2], Hirotaka Tamura[2], Jun Ohkubo[1,3]

[1] *Graduate School of Science and Enginnering, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama-shi, 338-8570, Japan*
[2] *Fujitsu Laboratories Ltd., 4-1-1 Kawasaki, Kanagawa 211-8558, Japan*
[3] *JST, PREST, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan*

We propose the ℓ_1-norm in QUBO formulation, which makes it possible to derive the QUBO formulation for sparse estimation. I derived the ℓ_1-norm in the QUBO formulation by applying the method of formulation of ReLU type function proposed by Sato et al. (2019). In addition, through numerical experiments, I was able to reduce the variables by reviewing the fomulation.

We propose a quadratic unconstrained binary optimization (QUBO) formulation of the ℓ_1-norm, which enables us to perform sparse estimation in the Ising-type annealing methods including quantum annealing. The QUBO formulation is derived via Legendre transformation and the Wolfe theorem, which have recently been employed to derive the QUBO formulation of ReLU-type functions. Furthermore, it is clarified that a simple application of the derivation method gives a redundant variable; finally a simplified QUBO formulation is obtained by removing the redundant variable.

1. Introduction

In recent years, Ising machines such as D-Wave Inc.'s D-Wave[1,2] and Fujitsu's Digital Annear[3] have been developed. The number of available qubits has been increasing year by year, which has enabled Ising machine calculations for large combinatorial optimization problems. However, to calculate the optimization problem with Ising machine, it is necessary to convert the cost function into QUBO form and impliment as hardware, but a systematic derivation method to QUBO form has not been found yet.

In recent years, some novel computing hardwares have been developed and actually provided; there are some Ising-type annealing machines such as "D-Wave 2000" by the Canadian company D-Wave[1,2] and "FUJITSU Quantum-inspired Computing Digital Annealer" by the Japanese company Fujitsu.[3] The annealing machines are used to obtain approximate solutions for optimization problems; the optimization problems play important roles in various research areas including data mining and machine learning. Especially, the quantum annealing method has been originally proposed in,[4] and a similar idea called adiabatic quantum computing[5] has attract many attentions; recently, researches for practical applications have been performed (for example, see.[6]) Discussions for machine learning and quantum Boltzmann machines were given in,[7] and there are many challenging tasks from the viewpoints of hardware and software. Although one of the restrictions are the small number of system size, the number of available qubits (or classical bits) has been increasing year by year, which enables us to tackle practical and large optimization problems.

In previous research, q-loss function[11] and ReLU type function[12] have been derived in the QUBO form. As for the derivation method of q-loss function, QUBO form could be derived by applying Legendre transformation. However, for the ReLU type function, the min function of the cost function has a minus sign only by applying the Legendre transform, and it is impossible to solve the optimization problem by combining multiple cost functions. Therefore, this problem was solved by additionally applying Wolfe duality theorem.[13]

The above annealing hardwares need quadratic unconstrained binary optimization (QUBO) formulations; the hardwares are based on the Ising-type Hamiltonian, and hence it is necessary to convert original cost functions in optimization problems into the QUBO formulation. (The QUBO formulation is equivalent to the Ising model.) Although continuous variable can be expressed as the Ising-type variables via adequate binary-expansions, in general, it is not straightforward to reformulate the original cost functions as the QUBO formulation. Some reformulations were given in,[8] and it has been shown that logic gates are expressed in the form of the QUBO formulations. However, a systematic way to derive the QUBO formulations has not been found yet. Recently, the Legendre transformation was employed to derive the QUBO form of the q-loss function;[11] the q-loss function was proposed as a cost function with robust characteristics against the label noise in machine learning. The derivation technique based on the Legendre transformation revealed that some mathematical transformation would be needed to transform some types of cost functions into the QUBO form. Actually, it has been clarified that the Legendre transformation is not enough to deal with the Rectified Linear-Unit (ReLU) type functions;[12] the Wolfe duality theorem[13] was employed to derive the QUBO form for the ReLU-type functions. These works also indicate the fact that the derivation of the QUBO formulations is not straightforward, and we sometimes needs further considerations depending on the original cost functions.

The ℓ_1-norm introduced in this paper is used in sparse estimation. Sparse estimation is used in the field of data analysis and image processing. Lasso[14] is a typical example used in data analysis, and it is possible to perform sparse estimation by adding ℓ_1-norm to the least-squares cost function. The idea of sparse estimation is used in black hole analysis.[10] The black hole is so small that it is impossible to observe it by the previous method because the resolution is low. Therefore, we performed simultaneous measurements from radio telescopes all over the world, performed sparse estimation on the observation data, extracted only essential information, and performed imaging of black holes. Thus, this research is important in performing sparse estimation on Ising machines.

As shown above, there are some works to derive the QUBO formulation for machine learning problems. Of course, there are many other research fields related to optimization problems, and one of them is data analysis and data mining. It has been known that regularization plays an important roles in data analysis and, of course, machine learning. ℓ_2 norms are widely used; for example, a linear regression with the ℓ_2 regularization is called the ridge regression. ℓ_1 norms are used in order to introduce a kind of sparseness; the sparse estimation is one of hot topics in the research field of data analysis. Least absolute shrinkage and selection operator (LASSO)[14] is a famous practical method to achieve sparse estimations, in which the ℓ_1 norm is added to a least-squares cost function. The idea of the sparse estimation was also applied to the recent black hole analysis.[10] The black hole is so small that it is difficult to observe it because of the low-resolution of images. Therefore, simultaneous measurements from radio telescopes all over the world are performed, and the method based on the sparse estimation is applied to the observed big data, by which only essential information is extracted and finally the imaging of the black holes was achieved. Note that ℓ_2 norm is simply connected to the QUBO formulation because of the quadratic form; in contrast, ℓ_1 norm has a non-differentiable point, and the QUBO formulation has not been derived yet.

In this paper, we derivate ℓ_1-norm in QUBO formlation. To do this, we use the Legendre transform and Wolfe duality theorem. Furthermore, it was confirmed that the variables can be reduced by reviewing the formulation, compared to the case of simple application. This is important when considering hardware implementation in the current situation where the number of qubits of Ising machine is limited.

In this paper, the QUBO form of the ℓ_1 norm is derived. In order to obtain the QUBO formulation, both the Legendre transformation and the Wolfe duality theorem are employed. Furthermore, it is clarified that only the simple applications of the previous derivation techniques are not enough; through numerical checks and reconsidering the derived formulation, a simplified QUBO formulation is finally derived. The reduction of the number of variables is important for the hardware implementation because the current Ising-type hardwares have only restricted number of qubits (or classical bits).

The composition of this paper is as follows. Section 2 explains the QUBO form, previous researchs, and theorems used for derivation. Section 3 carried out the derivation of ℓ_1-norm in the QUBO form and its varification. Section 4 removes variables from the objective function, and varifies whether the result is not affected. Section 5 discusses the results of Section 4 and the prospects for the future.

The construction of this paper is as follows. Section 2 explains the QUBO form, previous works. The important technique and theorem are also given for later use in the derivation. In Sect. 3, the QUBO form of the ℓ_1 norm is derived, and the numerical checks are given. Section 4 gives the main result of this paper; a simplified version of the QUBO form is given, in which a variable is removed from the QUBO form in Sect. 3. Section 5 gives concluding remarks and future works.

2. Backgrounds and preliminaries

In this section, we describes the knowledge.

In this section, some background knowledge and previous works are briefly denoted.

2.1 QUBO and Ising model

Since the QUBO formulation and the Ising model are equivalent, we can be converted to other form if we can be represented one side. The Ising model is represented as follows:

As denoted in the Introduction, the Ising-type annealing machines need the Ising Hamiltonian or the QUBO formulation in order to solve combinatorial optimization problems. The QUBO form has binary variables, which take only 1 or 0, and the 0-1 binary variables are sometimes suitable to consider the combinatorial optimizations; some problems are formulated as integer programming problems, or problems with continuous variables can be treated by the usage of the binary expansions. Of course, since the QUBO formulation and the Ising model are equivalent, it is possible to convert the QUBO form into the Ising model, and vice versa. The Ising model is represented as follows:

$$H = -\sum_{i,j} J_{i,j}\sigma_i\sigma_j - \sum_i h_i\sigma_i \qquad (1)$$

where $\sigma_i \in \{-1,+1\}$is a spin variable for i-th spin, $J_{ij} \in \mathbb{R}$ a quadratic term of i and j, and $h_i \in \mathbb{R}$ a liner term of i. We can easily converted the Ising model to QUBO formulation, which uses binary variable $q_i \in \{0,1\}$, by applying $q_i = \frac{\sigma_i+1}{2}$ and QUBO formulation is represented as follows:
where $\sigma_i \in \{-1,1\}$ is a spin variable for i-th spin, $J_{ij} \in \mathbb{R}$ a coefficient related to a quadratic term between spin i and j, and $h_i \in \mathbb{R}$ a coefficient for a linear term with spin i. Let $q_i \in \{0,1\}$ be a binary variable corresponding to the i-th spin, and then by applying the variable transformation $q_i = (\sigma_i + 1)/2$, we have

$$H = -\sum_{i,j} \widetilde{J}_{i,j}q_iq_j - \sum_i \widetilde{h}_iq_i, \qquad (2)$$

where $\widetilde{J}_{i,j}$ and $\widetilde{h}_i$ should be transformed from $\{J_{ij}\}$ and $\{h_i\}$ adequately. As for the relations between the QUBO formulation and the Ising Hamiltonian, please see Ref. 6; in Ref. 6, some examples of the QUBO formulations for typical optimization problems are also given.

2.2 Legendre transformation

For reader's convenience, we here give a brief notation for the Legendre transformation.

If a function f_L is convex, the Legendre transformation of f_L, the so-called conjugate function of f_L, is given as follows:

$$f_L^*(t) = \sup_x\{tx - f_L(x)\}. \qquad (3)$$

That is, the variable t is introduced, and the function for x is transformed to the function for t. In addition, (3) is equivalent to following equation:

$$f_L^*(t) = -\inf_x\{f_L(x) - tx\}. \qquad (4)$$

2.3 Previous work 1: q-loss function

This section describes the QUBO formulation of the q-loss function and ReLU type function performed in the previous research. The cost function of the q-loss function can be represented as

Here, a brief review of the previous work by Denchev *et al.* is given.[11] the following q-loss function was proposed in Ref. 11:

$$L_q(m) = \min\left[(1-q)^2, (\max[0, 1-m])^2\right] \tag{5}$$

where, $q \in (\infty, 0]$ is a parameter and m is a variable. By applying the Legendre transformation to the function $L_q(m)$ in (5), we can transform it as follow.
where $q \in (\infty, 0]$ is a parameter and m is a continuous variable. In Ref. 11, there is a discussion for the application of the q-loss function in machine learning problems, and the q-loss function has a robust features against label noise. Since Eq. 5 has a max function, it is not easy to see the QUBO form of the q-loss function. Denchev *et al.* employed the Legendre transformation, and finally the following function was derived:[11]

$$L_q(m) = \min_t \left\{(m-t)^2 + (1-q)^2 \frac{(1-\mathrm{sign}(t-1))}{2}\right\}, \tag{6}$$

where t is an additional variable which is introduced via the Legendre transformation. Although the variables m and t in Eq. (6) are continuous, the usage of the binary expansions gives the QUBO formulation for the q-loss function. As for details of the binary expansions, please see Ref. 11. Note that the sign function in Eq. (6) is also expressed as a one-body term when we employ the binary expansion.

2.4 Wolfe-duality

In nonlinear programming and mathematical optimization, the Wolfe duality theorem[13] is used to convert a main problem with inequality constraints to a dual problem. For a differentiable objective function and differentiable constraints, the main problem is written as follows:

$$\begin{cases} \text{minimize}_{\boldsymbol{x}} & f_W(\boldsymbol{x}) \quad (\boldsymbol{x} \in \mathbb{R}^n), \\ \text{subject to} & h_i(\boldsymbol{x}) \le 0 \quad (i = 1, 2, \dots, l). \end{cases} \tag{7}$$

where $f_W(\boldsymbol{x})$ is a certain convex function to be optimized and $h_i(\boldsymbol{x})$ are convex and inequality constraints. The Lagrangian function for this optimization problem is

$$L(\boldsymbol{x}, \boldsymbol{z}) = f_W(\boldsymbol{x}) + \boldsymbol{z}^T h(\boldsymbol{x}), \tag{8}$$

where $\boldsymbol{z}$ is a vector of the Legendre coefficients. Then, the Wolfe dual theorem means that the minimization problem in Eq. (7) in equivalent to the following maximization problem:

$$\begin{cases} \text{maximize}_{\boldsymbol{x},\boldsymbol{z}} & L(\boldsymbol{x}, \boldsymbol{z}) \quad ((\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}^n \times \mathbb{R}^l), \\ \text{subject to} & \nabla L(\boldsymbol{x}, \boldsymbol{z}) = 0 \quad (\boldsymbol{z} \ge 0). \end{cases} \tag{9}$$

As shown above, the Wolfe dual theorem transforms the minimization problem to the maximization problem.

2.5 Previous work 2: ReLU-type function

The cost function of the ReLU type function can be represented as follow.

In Ref. 12, the QUBO form of the following ReLU-type function was discussed:

$$f_R(m) = -\min(0, m). \tag{10}$$

By applying the Legendre transformation to the function $f(m)$ in (10), we can transform it as follow.
Note that the function $f_R(m)$ becomes the conventional ReLU function when the variable transformation $m \to -m$ is employed. As shown in Ref. 12, a naive application of the Legendre transformation to the function $f(m)$ in Eq. (10) gives the following expression:

$$f_R(m) = -\min_t\{-mt\} \quad \text{subject to} \quad -1 \le t \le 0, \tag{11}$$

where t is a new variable which stems from the Legendre transformation.

Here, since the min function in (11) has a minus sign, it is difficult to solve an optimization problem that is combined with multiple cost functions. So, this apply Wolfe duality theorem to the function in (11). The details of Wolfe duality theorem are explained in the next section.

Although Eq. (11) has the QUBO form, it is not suitable for optimization problems. This is because the minus sign before the min function; when the ReLU-type function is used as a kind of constraints or penalty terms for an optimization problem with a cost function $C(m)$, the whole minimization problem is, for example, given as follows:

$$\begin{aligned} \min_m\{C(m) + f_R(m)\} &= \min_m\left\{C(m) - \min_t\{-mt\}\right\} \\ &\ne \min_{m,t}\{C(m) - (-mt)\}, \end{aligned} \tag{12}$$

and hence the cost function $C(m)$ and the ReLU-type function $f_R(m)$ cannot be minimized simultaneously. Therefore, the Wolfe duality theorem was employed in Ref. 12, and finally the following formulation was derived:

$$f_R(m) = \min_{t,z_1,z_2}\left\{mt + z_1(t+1) - z_2 t - M(-m - z_1 + z_2)^2\right\} \tag{13}$$

where M is a large positive constant. It is easy to see that Eq. (13) can be used with the combination of the cost function $C(m)$.

3. Naive derivation of QUBO formulation for ℓ_1-norm

In this section, we derive the QUBO formulation of absolute function.

In this section, the Legendre transformation and the Wolfe dual theorem are applied to the ℓ_1 norm-type function naively.

3.1 QUBO formulation

We define the following function $f(m)$:

$$f(m) = -\min\{-m, m\} \tag{14}$$

Although the ℓ_1-norm is usually denoted as an absolute value of a variable, i.e., $|m|$, here we employ the following function $f(m)$:
(math)

Note that $f(m)$ can be expressed as follows:

$$f(m) = -\min\{0, m\} - \min\{-m, 0\} = f_R(m) + f_R(-m), \tag{15}$$

where $f_R(m)$ is the ReLU-type function in Eq. (10). Hence, it is easy to obtain the QUBO formulation for $f(m)$ by us-

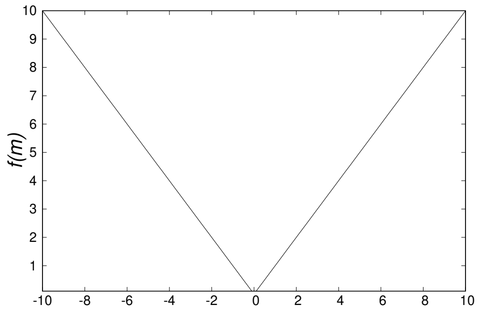

Fig. 1. ^{m} ℓ_1-norm in Eq. (14)

ing the discussion based on Sect. 2.5. However, the QUBO formulation needs six additional variables. The derivation below enables us to obtain the QUBO formulation for $f(m)$ with only three additional variables.

A function form of $f(m)$ in shown in Fig. 1. And by applying the Legendre transrormation to the function $f(m)$ in (14), we can transform it as follow.

$$F(m) = -\min_t \{-mt\} \quad \text{subject to} \quad 1 \leq t \leq 1 \tag{16}$$

(more explanation...)

We could express the quadratic form of $f(m)$ as (16), but there is the min function is proceded by a minus sign, which makes it difficult to solve an optimization problem that combines multiple cost functions. Let the other cost function be $C(m)$, and the combination with $F(m)$ is as follows:

$$\begin{aligned} \min_m \{C(m) + F(m)\} &= \min_m \left\{C(m) - \min_t \{-mt\}\right\} \\ &\neq \min_{m,t} \{C(m) - (-mt)\} \end{aligned}$$

Hence, it is not in the form of minimization problem for both m and t. In precious researche,[12] this problem was solved by applying Wolfe dual theorem to (16). By applying this theorem, the dual problem of the optimization problem (16) is represented as follows:

As shown in Sect. 2.5, although the obtained expression via the Legendre transformation has a quadratic form, it cannot be combined with another cost function. Hence, the Wolfe dual theorem is employed; the following expression is immediately obtained by applying the Wolfe dual theorem to $F(m)$:

$$\widetilde{F}(m) = \max_{t,z_1,z_2}\{-mt - z_1(t+1) + z_2(t-1)\} \tag{17}$$

$$\text{subject to} \quad \begin{cases} -m - z_1 + z_2 = 0, \\ -1 \leq t \leq 1, z_1 \geq 0, z_2 \geq 0. \end{cases}$$

In order to remove the equality constraint ($-m-z_1+z_2 = 0$), it is enough to add the following penalty term of the square of it. Therefore, the optimization problem (17) can be represented as follows:

This reformulation has the equality constraint, $-m - z_1 + z_2 = 0$; in order to embed this constraint into the QUBO formulation, it is enough to add the squared term as a penalty. Therefore, the optimization problem (17) can be represented as follows:

$$\begin{aligned} \widetilde{F}(m) = \min_{t,z_1,z_2} \{&mt + z_1(t+1) - z_2(t-1) \\ &+ M(-m - z_1 + z_2)^2\} \end{aligned} \tag{18}$$

$$\text{subject to} \quad -1 \leq t \leq 1, z_1 \geq 0, z_2 \geq 0,$$

where M is a constant and take a large value to ensure the equality constraint ($-m-z_1+z_2 = 0$) to be satisfied, and the remaining inequality consrains conditions ($-1 \leq t \leq 1, z_1 \geq 0$ and $z_2 \geq 0$) can be easily realized by expanding these variables t, z_1 and z_2, in the binary expressions which satisfy the corresponding domain constraints. We will varify in the next section that the (18) is correctly fomulated.

where M is a constant and takes a large value to ensure the equality constraint, $-m - z_1 + z_2 = 0$, to be satisfied. Note that there are remaining inequality constraints, $-1 \leq t \leq 1, z_1 \geq 0$, and $z_2 \geq 0$; these inequality constraints can be easily realized by expanding these variables t, z_1, and z_2, in the binary expressions which satisfy the corresponding domain constraints respectively.

As a result, the QUBO formulation for the ℓ_1-norm is expressed by using additional three variables.

3.2 *Numerical validation*

In this section, we verify that the formulation is correct by optimizing problem (18) with SA Algorithm. Since the purpose to this numerical experiment is to confirm the derived formulation, the function in (18) was experimented with continuous variables without binary expansion. Here, the initial values of constants and variables are described as follow.

Here, numerical checks are given in order to see the validity of the obtained QUBO formulation in Eq. (18). We expect that the QUBO formulation will be used within quantum annealing methods or simulated annealing methods, and here simulated annealing algorithms are employed. Since he purpose here is to verify the obtained QUBO formulation, the function in Eq. (18) is experimented with continuous variables without binary expansions.

- constant m : Generate uniform random number in the range of $[-10, 10]$.
- variables t, z_1, z_2 :
  - t generates uniform random number in the range of $[-1, 1]$.
  - z_1 and z_2 generate uniform random number in the range of $[0, 10]$.

The aim here is to check whether $\widetilde{F}(m)$ in Eq. (18) gives the ℓ_1-norm in Eq. (14). We randomly generate m and check the value of $\widetilde{F}(m)$ by using a simulated annealing method for continuous variables. m is chosen from a uniform distribution in the range of $[-10, 10]$. The numerical experiments are performed (??? number) times. In each numerical experiments, the initial conditions for the additional variables, t, z_1, and z_2 is chosen as follows:

- t is generated from an uniform distribution with the range of $[-1, 1]$.
- z_1 and z_2 are generated from an uniform distribution with the range of $[0, 10]$.

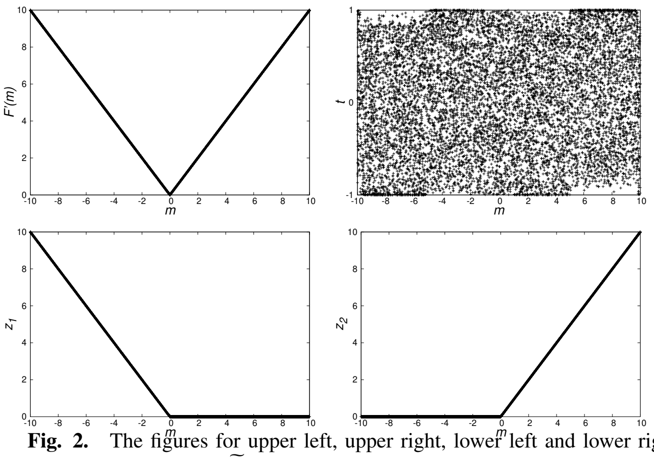

Fig. 2. The figures for upper left, upper right, lower left and lower right show the results when $\widetilde{F}(m), t, z_1, z_2$ are optimized for each m respectively. Numerical results obtained from the annealing method. m is chosen randomly, and the annealing is performed; each m gives a value of $\widetilde{F}(m)$. The values of t, z_1, z_2 at the optimum states are also shown.

Each variable moves by +0.001 or -0.001 with the same probability for each iteration. The annealing conditions are described as follow.

- initial temperature is $T_1 = 1,000$.
- the number of iteration is until temperature is up to a temperature of 1×10^{-3}.
- annealing schedule is $T_{n+1} = 0.9999T_n$

The experimental results are as shown in Fig.2.

As the simulated annealing method for the continuous variables, we employ a conventional Metropolis-Hastings-type method. In order to generate next candidates of the state, each variable moves by the amount of $+0.001$ or -0.001 with the same probability for each iteration. At each iteration, the temperature is changed with the annealing schedule $T_{n+1} = 0.9999T_n$, where n is the iteration step. The initial temperature is set as As for the annealing procedures, $T_1 = 1000$, and the annealing is finished when the temperature is lower than 10^{-3}.

From this result, we can confirm that l1-norm can be obtained by optimizing (18). Also, if we look at each variable t, z_1 and z_2 at optimization, we can see the following: It may be possible to reduce one variable by reviewing the (18) because t is not converged alothough z_1 and z_2 converge to a specific value.

Figure 2 shows the results of the annealing. We confirm that ℓ_1-norm is adequately recovered by the optimization of Eq. (18). Not only the value of the cost function $\widetilde{F}(m)$, but also the values of the three additional variables are also shown in Fig. 2; it is clear that z_1 and z_2 converge to specific values, but t takes various values randomly. This means that t would not be necessary for the optimization, which may give us a further simplified QUBO formulation for the ℓ_1-norm.

4. Reduced QUBO formulation

4.1 *Reduction of the variable in the Legendre transformation*

We can think of the following from the results of the numerical experiments in the previous section: The variable t seems to be taking a random value rather than settling to the optimal value, so we will consider whether we can eliminate t by reviewing the formulation. The cost function can be transformed as follows using equality constraint.

As discussed above, the naive application of the Legendre transformation and the Wolfe dual theorem gives the QUBO formulation with three additional variables. However, from the numerical experiments in the previous section, it is revealed that the variable t, which stems from the Legendre transformation, may not be necessary for the optimization problem. Because of the restriction of the number of spin variables, it is preferable to have smaller number of variables in general. Hence, here we try further reduction of variable from the QUBO formulation in Eq. (18).

In order to achieve the elimination of t from Eq. (18), we focus on the equality constraint $-m-z_1+z_2 = 0$. By employing the equality $z_2 = m + z_1$, we have

$$\begin{aligned}
\widetilde{F}(m) &= \min_{t,z_1,z_2} \{mt + z_1(t+1) - z_2(t-1) \\
&\quad + M(-m - z_1 + z_2)^2\} \\
&= \min_{t,z_1,z_2} \{mt + z_1(t+1) - (m+z_1)(t-1) \\
&\quad + M(-m - z_1 + z_2)^2\} \\
&= \min_{z_1,z_2} \{z_1 + (m+z_1) + M(-m - z_1 + z_2)^2\} \\
&= \min_{z_1,z_2} \{z_1 + z_2 + M(-m - z_1 + z_2)^2\}. \qquad (19)
\end{aligned}$$

This conversion from (18) to (20) is possible because the penalty term, $M(-m - z_1 + z_2)^2$, forces the equality constraint to be satisfied.

Then, we finally obtain the following simplified QUBO formulation for the ℓ_1-norm:

$$\widehat{F}(m) = \min_{z_1,z_2} \{z_1 + z_2 + M(-m - z_1 + z_2)^2\}, \qquad (20)$$

where the new expression $\widehat{F}(m)$ is introduced in order to clarify the difference from Eq. (18). Therefore,

4.2 *Numerical validation*

The result of experimenting the optimization problem under the same experimental conditions as Section3 with (20) as the objective function is as shown in Fig.3.

From this result, the outline of $\widetilde{F}(m), z_1$ and z_2 when optimizing (18) did not change from Fig.3, so we seem that removing the variable t does not affect the result.

In order to check the validity of the simplified QUBO formulation in Eq. (20), we again perform numerical experiments. The same settings and procedures as the previous section is employed for the annealing, except for the absence of the variable t. Figure 3 shows the numerical results; it clarifies that the simplified QUBO formulation works well.

5. Concluding remarks

In this paper, we propose the formulation of ℓ_1-norm in the QUBO formulation and reduce the variables through numerical experiments. However, the reduction method of this variable can be predicted just by looking at the formulation. When ℓ_1-norm is added to the cost function in order to make sparse estimation with Ising machine, it is necessary to add two variables for each estimated value, and the number of variables

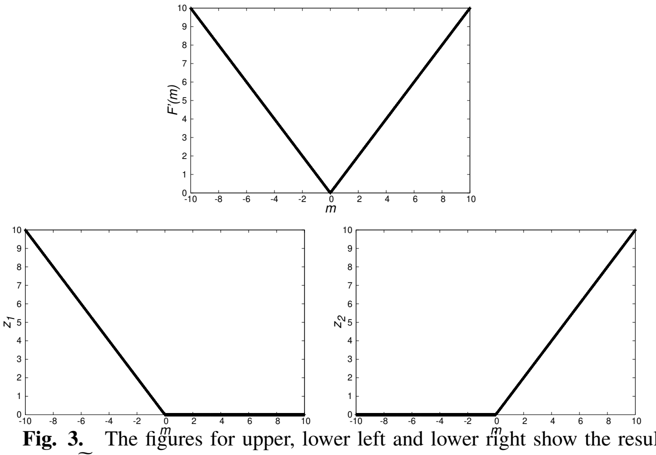

Fig. 3. The figures for upper, lower left and lower right show the results when $\widetilde{F}(m), z_1, z_2$ are optimized for each m respectively. Numerical results for the simplified QUBO expression in Eq. (20).

is still large. It is important to consider the implementation on hardware, and the problem still remains in this formulaiton. As a solution of this problem, in this experiment, the derivation was performed using the method used in the previous research, but it may be possible to reduce the number of additional variables required by using a different derivation method.

In this paper, the QUBO formulation for the ℓ_1-norm is derived. As discussed in Introduction, there is no systematic way to derive the QUBO formulation. There are some works for the QUBO formulation, and actually that of the ReLU-type function gives the QUBO formulation for the ℓ_1-norm, as discussed in Sect. 3. However, this needs six additional variables; the reduction of additional variables is important for the Ising-type hardware because of the current limitation of spin variables. Hence, we carefully applied the Legendre transformation and the Wolfe dual theorem to the ℓ_1-norm, which leads to the QUBO formulation in Eq. (18); this has only three additional variable. Furthermore, it was clarified that the QUBO formulation for the ℓ_1-norm is achieved with only two additional variables as shown in Eq. (20). At first glance, it is difficult to see the connection between the ℓ_1-norm and the final expression in Eq. (20). This nontrivial result is the main contribution of the present work, which means that careful derivation is required individually.

As shown in the derivation, the variable t, which is introduced via the Legendre transformation, was reduced finally. It is not clear whether the usage of the Legendre transformation is necessary or not; at this stage, the procedure (the Legendre transformation → the Wolfe dual theorem → reduction of variables) is straightforward and understandable. Of course, there could be more suitable derivation methods. In addition, when the ℓ_1-norm is combined to another cost function in order to make sparse estimation with the Ising-type hardware, it is necessary to add two variables for each estimated value; the number of additional variables is still large. It will be important future works to find further reduction methods for practical problems with large size.

In order to achieve wider use of the annealing hardware, it is important to derive QUBO formulations for various types of cost functions or regularization terms. The present work revealed the Legendre transformation, the Wolfe dual theorem, and the careful variable reduction work well for the derivation, and hence we believe that the procedure will be beneficial for future researches on the QUBO formulations and the annealing hardwares.

1) M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson and G. Rose, Nature **473**, 194 (2011).

2) P. I. Bunyk, E. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, J. Whittaker, IEEE Trans. Appl. Supercond. **24**, 1700110 (2014).

3) M. Aramon, G. Rosenberg, E. Valiante, T. Miyazawa, H. Tamura, and H. G. Katzgraber, arXiv:1806.08815.

4) T. Kadowaki and H. Nishimori, Phys. Rev. E **58**, 5355 (1998).

5) E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, Science **292**, 472 (2001).

6) K. Tanahashi, S. Takayanagi, T. Motohashi, and S. Tanaka, J. Phys. Soc. Jpn. **88**

7) J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe and S. Lloyd, Nature **549**, 195 (2017).

8) A. Lucas, Front. Phys. **2**, 5 (2014).

9) J. D. Whitfield, M. Faccin, and J. D. Biamonte, Europhys. Lett. **99**, 57004 (2012).

10) M. Honma, K. Akiyama, F. Tazaki, K. Kuramochi, S. Ikeda, K. Hada, and M. Uemura, J.Phys.: Conf. Ser. **699** 012006 (2016).

11) V. Denchev, N.Ding, S. V. N Vishwanathan, and H. Neven, in Proceedings of the 29th International Conference on Machine Learning, p.863 (2012).

12) G. Sato, M. Konoshima, T. Ohwa, H. Tamura, and J. Ohkubo, Phys. Rev. E **99**, 042106 (2019).

13) P. Wolfe, Quart. Appl. Math. **19**, 239 (1961).

14) R. Tibshirani, J. R. Statist. Soc, B **58**, 267 (1996).