IEICE TRANS. FUNDAMENTALS, VOL.E93–A, NO.8 AUGUST 2010

PAPER

Global Nonlinear Optimization Based on Eigen Analysis of Schrödinger-type Equation

Hideki SATOH[†], ***Member***

SUMMARY A method has been developed for deriving the approximate global optimum of a nonlinear objective function. First, the objective function is expanded into a linear equation for a moment vector, and the optimization problem is reduced to an eigen analysis problem in the wave coefficient space. Next, the process of the optimization is expressed using a Schrödinger-type equation, so global optimization is equivalent to eigen analysis of the Hamiltonian of a Schrödinger-type equation. Computer simulation of this method demonstrated that it produces a good approximation of the global optimum. An example optimization problem was solved using a Hamiltonian constructed by combining Hamiltonians for other optimization problems, demonstrating that various types of applications can be solved by combining simple Hamiltonians.

key words: *nonlinear, global optimization, wave function, quantum computing, Schrödinger equation*

1. Introduction

One of the most important problems in deriving the global optimum is avoiding falling into a local optimum. Tunneling algorithms [1], [2], [3] first search for a local optimum and then search for a better local optimum using a tunneling method starting from the local optimum previously obtained. The complex dynamics of a chaotic attractor is applied to various optimization methods to avoid being trapped in a local optimum [4]. Hopfield neural networks (HNNs) [5] define an energy function derived from an objective function, and the state of the HNN changes in accordance with the energy function until the state becomes stable. A solution is then derived from the stable state. Boltzmann machines [6] are variations of HNNs to which a stochastic parameter is added so that the state first changes stochastically, and the stochastic changes in the state gradually decrease. With this parameter, the state falls into a local optimum less frequently. Simulated annealing (SA) [7] uses the principle of annealing from metal engineering: slowly cooling heated metal produces a superior crystalline structure. Genetic algorithms [8] use the principle of organic evolution to achieve the same thing.

These methods are designed to avoid being trapped in a local optimum by using random variables, chaos, the structure of the objective function, or the relationship between local optimums. However, there is a fundamental limit to their ability because they search for the global optimum in real space. Quantum mechanics was recently incorporated into some optimization methods to overcome this difficulty. Instead of SA thermal fluctuations in real space, quantum annealing (QA) uses quantum fluctuations and thus has a shorter convergence time [9], [10], [11]. Quantum neural networks (QNNs) [12], [13] are variations of HNNs and were developed to effectively perform a full search on the basis of the superposition of quantum states. Because QA and QNNs use the properties of a quantum mechanical system, they can find the global optimum without falling into a local optimum. However, it is very difficult to construct a device to create the required quantum effects. It is thus still difficult to apply them to most optimization problems.

A global optimization algorithm in the wave coefficient space was developed to overcome the problems described above [14]. A nonlinear objective function is expanded into a linear function of a moment vector, and the global optimization problem is solved using the steepest decent method (SDM) in the wave coefficient space. Because it uses a wave function, it is equivalent to an algorithm that searches in parallel for the global optimum in the whole domain of definition. Therefore, it can always find an approximate global optimum. However, the relationship between the algorithm and quantum computing [15] needs clarification. And the difficulty of constructing a device to create the required quantum effects remains. Moreover, a complicated condition has to be used to judge whether a solution has been obtained although whether the condition works well for any objective function is unproven.

The global optimization algorithm in the wave coefficient space using SDM has now been expressed as a Schrödinger-type equation, and the algorithm has been improved using the equation. Moreover, the algorithm has been enhanced so that a Hamiltonian for an optimization problem can be constructed by combining Hamiltonians for various optimization problems. This makes it possible to connect global optimization problems to quantum computing.

Manuscript received December 28, 2009.
Manuscript revised March 30, 2010.

[†]The author is with Future University Hakodate, Hakodate-shi, 041-8655 Japan.

2. Moment Vector Equation for Nonlinear Objective Function

A moment vector equation (MVE) [14], [16] is used to reduce the problem of maximizing a nonlinear objective function to that of maximizing a quadratic function with respect to the wave coefficients. The MVE of the objective function and the pdf corresponding to the moment vector are derived.

2.1 Moment Vector Equation

The MVE was developed to approximate an arbitrary multi-dimensional nonlinear function in the whole domain of definitions [16]. Consider the following nonlinear function:

$$y = f(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{x} \stackrel{\text{def}}{=} (x_1, \cdots, x_{d_\text{x}})^\text{T} \in \boldsymbol{\mathcal{D}}_\mathbf{x}$ is the state vector of dimension d_x, $\boldsymbol{\mathcal{D}}_\mathbf{x} \stackrel{\text{def}}{=} \{\boldsymbol{x} | x_{\min d} < x_d < x_{\max d}, 1 \leq d \leq d_\text{x}\}$ is the domain of the definition of $\boldsymbol{x}$, $y \in \boldsymbol{\mathcal{D}}_\mathbf{y}$ is the value of $f(\boldsymbol{x})$, $\boldsymbol{\mathcal{D}}_\mathbf{y} \stackrel{\text{def}}{=} \{y | y_\text{min} < y < y_\text{max}\}$ is the domain of the definition of y, and superscript T denotes transposition. If the domain of definitions cannot be set in advance, Eq. (1) can be changed to

$$y = h_\text{y}(f(\boldsymbol{h}_\text{x}(\boldsymbol{x}))), \tag{2}$$

using monotone increasing functions h_y and $\boldsymbol{h}_\text{x}$, which transform the infinite range of values and the infinite domain of definitions to finite ones.

Let $\{\psi_i(y)\}$ and $\{\psi_i(\boldsymbol{x})\}$ be orthonormal bases as defined in Appendix A. Note that the same symbol, ψ, is used to simplify the explanation, although $\{\psi_i(y)\}$ and $\{\psi_i(\boldsymbol{x})\}$ are generally different bases. To derive the MVE for the nonlinear function in Eq. (1), an assumption is introduced with respect to Eq. (1).

Assumption 1: We can expand $E[\psi_i(y)|\boldsymbol{x}]$ into a Fourier series:

$$E[\psi_i(y)|\boldsymbol{x}] = \sum_{j=0}^{N_\text{x}} a_{ij}\psi_j(\boldsymbol{x}) + \varepsilon_i(\boldsymbol{x}), \tag{3}$$

where N_x is the degree of expansion of $E[\psi_i(y)]$, $E[\cdot]$ is the mathematical expectation, and $\varepsilon_i(\boldsymbol{x})$ is the residual. □

Using Eq. (3), we can expand $E[\psi_i(y)]$:

$$\begin{aligned} E[\psi_i(y)] &= \int \psi_i' p(\psi_i') \mathrm{d}\psi_i' \\ &= \int \psi_i' \int p(\boldsymbol{x}) p(\psi_i'|\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \mathrm{d}\psi_i' \\ &= \int p(\boldsymbol{x}) E[\psi_i'|\boldsymbol{x}] \mathrm{d}\boldsymbol{x} \\ &= \int p(\boldsymbol{x}) (\sum_{j=0}^{N_\text{x}} a_{ij}\psi_j(\boldsymbol{x}) + \varepsilon_i(\boldsymbol{x})) \mathrm{d}\boldsymbol{x} \\ &= \sum_{j=0}^{N_\text{x}} a_{ij} E[\psi_j(\boldsymbol{x})] + E[\varepsilon_i(\boldsymbol{x})], \end{aligned} \tag{4}$$

where ψ_i' denotes $\psi_i(y)$ and $p(\boldsymbol{x})$ denotes the probability density function (pdf) of $\boldsymbol{x}$. When Eq. (1) is deterministic, a_{ij} is obtained using Eq. (A·2):

$$a_{ij} = \int_{\boldsymbol{\mathcal{D}}_\mathbf{x}} \psi_i(f(\boldsymbol{x})) \psi_j^*(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{5}$$

where superscript $*$ denotes a complex conjugate. If we assume that $E[\varepsilon_i(\boldsymbol{x})] = 0$, Eq. (4) can be expressed using a linear function:

$$E[\boldsymbol{\psi}(y)] = A_\psi E[\boldsymbol{\psi}(\boldsymbol{x})]. \tag{6}$$

This equation is referred to as the MVE, $\boldsymbol{\psi}(y) \stackrel{\text{def}}{=} (\psi_0(y), \cdots, \psi_{N_\text{y}}(y))^\text{T}$, N_y is the degree of expansion of y, $\boldsymbol{\psi}(\boldsymbol{x}) \stackrel{\text{def}}{=} (\psi_0(\boldsymbol{x}), \cdots, \psi_{N_\text{x}}(\boldsymbol{x}))^\text{T}$, and $A_\psi \stackrel{\text{def}}{=} [a_{ij} | 0 \leq i \leq N_\text{y}, 0 \leq j \leq N_\text{x}]$ is an $(N_\text{y}+1) \times (N_\text{x}+1)$ matrix.

The nonlinear function in Eq. (1) is approximately expressed by the MVE in Eq. (6). The accuracy of Eq. (6) increases as N_x and N_y increase. Using Eq. (6), we can derive not only the expected value of $\psi_i(y)$ but also the statistical properties such as the mean, variance, covariance, and pdf of y [16].

2.2 Probability Density Function

Let $\Psi(\boldsymbol{x})$ be a wave function. We can expand $\Psi(\boldsymbol{x})$ using orthonormal basis $\{\psi_i(\boldsymbol{x})\}$:

$$\Psi(\boldsymbol{x}) \cong \boldsymbol{c}^\text{T} \boldsymbol{\psi}(\boldsymbol{x}), \tag{7}$$

where c_i is the expansion coefficient of the wave function, which is referred to as the wave coefficient, and $\boldsymbol{c} \stackrel{\text{def}}{=} (c_0, \cdots, c_{N_\text{x}})^\text{T}$ is the wave coefficient vector. Probability density function $p(\boldsymbol{x})$ is obtained using $\Psi(\boldsymbol{x})$:

$$p(\boldsymbol{x}) \cong \boldsymbol{\psi}^\text{T}(\boldsymbol{x}) \boldsymbol{c}\boldsymbol{c}^\dagger \boldsymbol{\psi}^*(\boldsymbol{x}), \tag{8}$$

where superscript † denotes conjugate transposition. Because $\int p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int \Psi(\boldsymbol{x}) \Psi^*(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1$ and basis $\{\psi_i(\boldsymbol{x})\}$ is orthonormal, $\|\boldsymbol{c}\| \stackrel{\text{def}}{=} \sqrt{\boldsymbol{c}^\text{T}\boldsymbol{c}^*} = 1$.

Consider the moment vectors normalized using the norm of the orthonormal basis vector:

$$\boldsymbol{q} \stackrel{\text{def}}{=} {\xi_\text{q}}^{-1} E[\boldsymbol{\psi}^*(\boldsymbol{x})], \tag{9}$$

$$\boldsymbol{r} \stackrel{\text{def}}{=} {\xi_{\text{r}}}^{-1} E[\boldsymbol{\psi}^*(y)], \tag{10}$$

where $\xi_{\text{q}} \stackrel{\text{def}}{=} \|\boldsymbol{\psi}(\boldsymbol{x})\|$ and $\xi_{\text{r}} \stackrel{\text{def}}{=} \|\boldsymbol{\psi}(y)\|$. Using these moment vectors, we can modify Eq. (6) to obtain

$$\boldsymbol{r} = A\boldsymbol{q}, \tag{11}$$

where $A \stackrel{\text{def}}{=} {\xi_{\text{r}}}^{-1}\xi_{\text{q}} A^*_{\psi}$. As shown in Appendix B, the pdf of $\boldsymbol{x}$ with moment vector $\boldsymbol{q}$ and that of y with moment vector $\boldsymbol{r}$ can be expressed as

$$p(\boldsymbol{x}) \cong {\xi_{\text{x}}}^{-1} (\boldsymbol{\psi}^{\text{T}}(\boldsymbol{x})\boldsymbol{q}\boldsymbol{q}^{\dagger}\boldsymbol{\psi}^*(\boldsymbol{x}))^{1/2}, \tag{12}$$

$$p(y) \cong {\xi_{\text{y}}}^{-1} (\boldsymbol{\psi}^{\text{T}}(y)\boldsymbol{r}\boldsymbol{r}^{\dagger}\boldsymbol{\psi}^*(y))^{1/2}, \tag{13}$$

where $\xi_{\text{x}} \stackrel{\text{def}}{=} \int (\boldsymbol{\psi}^{\text{T}}(\boldsymbol{x})\boldsymbol{q}\boldsymbol{q}^{\dagger}\boldsymbol{\psi}^*(\boldsymbol{x}))^{1/2}\text{d}\boldsymbol{x}$ and $\xi_{\text{y}} \stackrel{\text{def}}{=} \int (\boldsymbol{\psi}^{\text{T}}(y)\boldsymbol{r}\boldsymbol{r}^{\dagger}\boldsymbol{\psi}^*(y))^{1/2}\text{d}\boldsymbol{y}$.

Although we can derive the pdfs using the above equations, they are somewhat complicated when we use them for global optimization. We can simplify them by assuming that the pdf of $\boldsymbol{x}$ is $p(\boldsymbol{x}) = \delta(\boldsymbol{x} - \hat{\boldsymbol{x}})$. The pdf of y then becomes $p(y) \stackrel{\text{def}}{=} \delta(y - f(\hat{\boldsymbol{x}}))$, and we can use the following equations instead of the ones above.

$$p(\boldsymbol{x}) \cong \|\boldsymbol{q}\|^{-2} \boldsymbol{\psi}^{\text{T}}(\boldsymbol{x})\boldsymbol{q}\boldsymbol{q}^{\dagger}\boldsymbol{\psi}^*(\boldsymbol{x}), \tag{14}$$

$$p(y) \cong \|\boldsymbol{r}\|^{-2} \boldsymbol{\psi}^{\text{T}}(y)\boldsymbol{r}\boldsymbol{r}^{\dagger}\boldsymbol{\psi}^*(y). \tag{15}$$

Comparison of Eq. (8) with the above equations shows that $\boldsymbol{q}$ and $\boldsymbol{r}$ express the unnormalized wave coefficient vectors[†] if $p(\boldsymbol{x})$ and $p(y)$ are delta functions. The assumption for $p(\boldsymbol{x})$ holds for global optimization problems because our goal is to derive a definite value of $\boldsymbol{x}$ that maximizes objective function $f(\boldsymbol{x})$; that for $p(y)$ also holds, as explained above. Therefore, $\boldsymbol{q}$ and $\boldsymbol{r}$ are used hereafter as the unnormalized wave coefficient vectors. The effect of the use of Eqs. (14) and (15) is discussed in Sect. 3.

3. Global Optimization in Wave Coefficient Space

3.1 Global Optimization Using Steepest Decent Method

The global optimization method used is based on the MVE in Eq. (11) [14]. Consider the optimization problem of maximizing $y \in \boldsymbol{\mathcal{D}}_{\mathbf{y}}$ obtained by Eq. (1) for $\boldsymbol{x} \in \boldsymbol{\mathcal{D}}_{\mathbf{x}}$. Using the MVE in Eq. (11) and the pdf in Eq. (15), we can rewrite the optimization problem as the problem of maximizing $E[y]$ defined by

$$\begin{aligned} E[y] &\stackrel{\text{def}}{=} \int y p(y) \text{d}y \\ &= \|A\boldsymbol{q}\|^{-2} \boldsymbol{q}^{\text{T}} A^{\text{T}} Y A^* \boldsymbol{q}^*, \end{aligned} \tag{16}$$

where $Y \stackrel{\text{def}}{=} \int y\boldsymbol{\psi}(y)\boldsymbol{\psi}^{\dagger}(y)\text{d}y$ is an $(N_{\text{y}}+1)\times(N_{\text{y}}+1)$ matrix. If $\|A\boldsymbol{q}\|^2 = \boldsymbol{q}^{\text{T}}A^{\text{T}}A^*\boldsymbol{q}^* = 1$, Eq. (16) reduces to $\boldsymbol{q}^{\text{T}}A^{\text{T}}YA^*\boldsymbol{q}^*$. Thus, the optimization problem of maximizing $E[y]$ is expressed as

$$\text{Object} : \max_{\boldsymbol{q}} \boldsymbol{q}^{\text{T}} A^{\text{T}} Y A^* \boldsymbol{q}^*, \tag{17}$$

$$\text{Constraint} : \boldsymbol{q}^{\text{T}} A^{\text{T}} A^* \boldsymbol{q}^* = 1. \tag{18}$$

Because the objective function of this problem is a quadratic equation with respect to $\boldsymbol{q}$, we can solve it using the steepest descent method (SDM) in a wave coefficient space for the following Lagrange function [18]:

$$L(\boldsymbol{q}, \mu) \stackrel{\text{def}}{=} -\boldsymbol{q}^{\text{T}} A^{\text{T}} Y A^* \boldsymbol{q}^* + \mu(\boldsymbol{q}^{\text{T}} A^{\text{T}} A^* \boldsymbol{q}^* - 1), \tag{19}$$

where $\mu(-\infty \le \mu \le \infty)$ is the Lagrange multiplier. The algorithm is described in Algorithm 1.

Algorithm 1: Global optimization using SDM in a wave coefficient space.

(1-1) Set $t = 0$.

(1-2) Set step sizes α_{q}, α_{μ}, and α_{step} (all > 0) and initial values $\boldsymbol{q}_0$ and μ_0.

(1-3) Compute $\boldsymbol{d}_{\text{q}t}$ and $d_{\mu t}$:
$\boldsymbol{d}_{\text{q}t} = -\nabla_{\text{Re}[\boldsymbol{q}]} L(\boldsymbol{q}_t, \mu_t) - \imath \nabla_{\text{Im}[\boldsymbol{q}]} L(\boldsymbol{q}_t, \mu_t)$,
$d_{\mu t} = -\nabla_{\mu} L(\boldsymbol{q}_t, \mu_t)$.

(1-4) Compute $\boldsymbol{q}_{t+1}$ and μ_{t+1}:
$\boldsymbol{q}_{t+1} = \boldsymbol{q}_t + \alpha_{\text{q}} \boldsymbol{d}_{\text{q}t}$,
$\mu_{t+1} = \mu_t - \alpha_{\mu} d_{\mu t}$.

(1-5) If $J((\boldsymbol{q}_{t+1}, \mu_{t+1}), (\boldsymbol{q}_t, \mu_t)) < \varepsilon$, set $\tilde{\boldsymbol{q}}_{\text{opt}} = \boldsymbol{q}_t$ and go to Step (1-7).

(1-6) Set $t = t + 1$ and go to Step (1-2).

(1-7) Compute the approximation of the global optimum, $E[\boldsymbol{x}]_{\text{opt}}$, using Eq. (20).

Here, $J(\cdot)$ denotes the norm used to judge whether the solution is obtained [14] and $\nabla_{\boldsymbol{x}}$ denotes the nabla operator with respect to $\boldsymbol{x}$. The approximation of the global optimum, $E[x_d]_{\text{opt}}$, which approximately maximizes Eq. (1), and the approximate maximum of Eq. (1), $E[y]_{\text{opt}}$, are obtained using

$$E[x_d]_{\text{opt}} = \|\tilde{\boldsymbol{q}}_{\text{opt}}\|^{-2} {\tilde{\boldsymbol{q}}_{\text{opt}}}^{\text{T}} X_d \tilde{\boldsymbol{q}}^*_{\text{opt}}, \tag{20}$$

$$E[y]_{\text{opt}} \stackrel{\text{def}}{=} \|A\tilde{\boldsymbol{q}}_{\text{opt}}\|^{-2} {\tilde{\boldsymbol{q}}_{\text{opt}}}^{\text{T}} A^{\text{T}} Y A^* \tilde{\boldsymbol{q}}^*_{\text{opt}}, \tag{21}$$

where $X_d \stackrel{\text{def}}{=} \int x_d \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\psi}^{\dagger}(\boldsymbol{x})\text{d}\boldsymbol{x}$ is an $(N_{\text{x}}+1)\times(N_{\text{x}}+1)$ matrix.

[†]Even if $\|\boldsymbol{q}\| = 1$, $\|\boldsymbol{r}\| = 1$ does not always hold in Eq. (11) because matrix A is not a unitary matrix. Thus, $\boldsymbol{q}$ and $\boldsymbol{r}$ express unnormalized wave coefficient vectors.

3.2 Global Optimization Using Eigen Analysis

The global optimization method described in Sect. 3.1 is reduced to a method using eigen analysis. First, we derive the $\boldsymbol{q}_t$ that satisfies the end condition of Step (1-5) in Algorithm 1. We can reduce $\nabla_{\mathrm{Re}[\boldsymbol{q}]}L(\boldsymbol{q},\mu)$, $\imath\nabla_{\mathrm{Im}[\boldsymbol{q}]}L(\boldsymbol{q},\mu)$, and $\nabla_\mu L(\boldsymbol{q},\mu)$ in Step (1-3) in Algorithm 1 to

$$\begin{aligned}\nabla_{\mathrm{Re}[\boldsymbol{q}]}L(\boldsymbol{q},\mu) &= -A^{\mathrm{T}}YA^*\boldsymbol{q}^* - (A^{\mathrm{T}}YA^*)^{\mathrm{T}}\boldsymbol{q} \\ &\quad + \mu A^{\mathrm{T}}A^*\boldsymbol{q}^* + \mu(A^{\mathrm{T}}A^*)^{\mathrm{T}}\boldsymbol{q}, \\ \imath\nabla_{\mathrm{Im}[\boldsymbol{q}]}L(\boldsymbol{q},\mu) &= A^{\mathrm{T}}YA^*\boldsymbol{q}^* - (A^{\mathrm{T}}YA^*)^{\mathrm{T}}\boldsymbol{q} \\ &\quad - \mu A^{\mathrm{T}}A^*\boldsymbol{q}^* + \mu(A^{\mathrm{T}}A^*)^{\mathrm{T}}\boldsymbol{q}, \\ \nabla_\mu L(\boldsymbol{q},\mu) &= \boldsymbol{q}^{\mathrm{T}}A^{\mathrm{T}}A^*\boldsymbol{q}^* - 1.\end{aligned}$$

By substituting the above equations into $\boldsymbol{d}_{\mathrm{q}t}$ and $d_{\mu t}$ in Step (1-3), we obtain

$$\boldsymbol{d}_{\mathrm{q}t} = 2((A^{\mathrm{T}}YA^*)^{\mathrm{T}} - \mu_t(A^{\mathrm{T}}A^*)^{\mathrm{T}})\boldsymbol{q}_t, \tag{22}$$

$$d_{\mu t} = {\boldsymbol{q}_t}^{\mathrm{T}}A^{\mathrm{T}}A^*\boldsymbol{q}_t^* - 1. \tag{23}$$

The end condition is satisfied when $\boldsymbol{d}_{\mathrm{q}t} = \boldsymbol{0}$ and $d_{\mu t} = 0$. Thus, Eq. (22) shows that $\boldsymbol{q}_t$ and μ_t should satisfy

$$(A^\dagger A)^{-1}A^\dagger Y^{\mathrm{T}}A\boldsymbol{q}_t = \mu_t\boldsymbol{q}_t,$$

so as to obtain $\boldsymbol{d}_{\mathrm{q}t} = \boldsymbol{0}$. This equation can be solved using eigen analysis. Let matrix H be

$$H \stackrel{\mathrm{def}}{=} (A^\dagger A)^{-1}A^\dagger Y^{\mathrm{T}}A, \tag{24}$$

the ith eigen value of H be ε_i, and the corresponding eigen vector be $\boldsymbol{e}_i$. Then the following equation holds.

$$H\boldsymbol{e}_i = \varepsilon_i\boldsymbol{e}_i. \tag{25}$$

We thus obtain $\boldsymbol{d}_{\mathrm{q}t} = \boldsymbol{0}$ when $\boldsymbol{q}_t$ equals $\boldsymbol{e}_i$. Here, we can set $\|\boldsymbol{e}_i\|$ to an arbitrary value, so we obtain $d_{\mu t} = 0$ by adjusting $\boldsymbol{e}_i$. Therefore, the end condition of Algorithm 1, $\boldsymbol{d}_{\mathrm{q}t} = \boldsymbol{0}$ and $d_{\mu t} = 0$, can be satisfied by setting $\boldsymbol{q}_t$ to $\boldsymbol{e}_i$.

Next, we select the $\boldsymbol{e}_i$ that provides the maximum value of $E[y]$ from $\{\boldsymbol{e}_i\}$. By substituting $H \stackrel{\mathrm{def}}{=} (A^\dagger A)^{-1}A^\dagger Y^{\mathrm{T}}A$ into Eq. (25) and multiplying both sides of the equation by $\|A\boldsymbol{e}_i\|^{-2}{\boldsymbol{e}_i}^{\mathrm{T}}(A^\dagger A)$, we obtain

$$\|A\boldsymbol{e}_i\|^{-2}{\boldsymbol{e}_i}^{\mathrm{T}}A^\dagger Y^{\mathrm{T}}A\boldsymbol{e}_i = \varepsilon_i\|A\boldsymbol{e}_i\|^{-2}{\boldsymbol{e}_i}^{\mathrm{T}}A^\dagger A\boldsymbol{e}_i.$$

Because $\boldsymbol{e}_i$ is a wave coefficient vector, we can apply Eq. (16) to the above equation to obtain

$$E[y] = \varepsilon_i. \tag{26}$$

We arrange the eigenvalues and corresponding eigenvectors by eigenvalue starting with the largest one. Then, $E[y]_{\mathrm{opt}}$, which is the approximate maximum of Eq. (1), is given by

$$E[y]_{\mathrm{opt}} = \varepsilon_0, \tag{27}$$

and $\tilde{\boldsymbol{q}}_{\mathrm{opt}}$, which approximately maximizes Eq. (1), is given by $\boldsymbol{e}_0$.

The global optimization method using SDM is thus reduced to a method using eigen analysis.

Algorithm 2: Global optimization using eigen analysis in a wave coefficient space

(2-1) Solve Eq. (25) and derive eigenvector $\boldsymbol{e}_0$ with the maximum eigenvalue.

(2-2) Set $\tilde{\boldsymbol{q}}_{\mathrm{opt}} = \boldsymbol{e}_0$.

(2-3) Compute the approximation of the global optimum, $E[\boldsymbol{x}]_{\mathrm{opt}}$, using Eq. (20).

The calculation cost of Algorithm 1 depends on the integration in Eq. (5) and SDM, and that of Algorithm 2 depends on the integration in Eq. (5) and the eigen analysis. Algorithm 2 uses a sophisticated algorithm for eigen analysis while Algorithm 1 has to repeatedly evaluate the quadratic form in Step (1-3) until the algorithm finishes. Thus, the calculation cost of Algorithm 2 is less than that of Algorithm 1. However, the calculation cost of Algorithm 2 is not lower than those of conventional methods because it is necessary to evaluate a sufficiently large number of samples of the objective function for the integration in Eq. (5). The calculation cost is thus almost the same as that of the full search algorithm. We can reduce the cost by using various integration methods that are suitable for the shape of the objective function. For a specific optimization problem, we can use other calculation devices except for commonly used digital computers. Moreover, Algorithm 2 is related to quantum computing as described in the following sections. Therefore, the efficient calculation and physical implementation of Algorithm 2 are well worth further study.

3.3 Relation to Quantum Mechanics

The global optimization of Algorithm 2 can be described from the viewpoint of quantum mechanics. Because the eigenvalue of Eq. (25) is a real number, as shown in Eq. (26), matrix H in Eq. (24) is an Hermitian matrix. Matrix H is thus the Hamiltonian of the optimization problem, ε_i is the energy, and Eq. (25) is interpreted as a Schrödinger-type equation [15] †. Therefore, the optimization problem is equivalent to deriving the wave function that provides the maximum eigenvalue of the Schrödinger-type equation with Hamiltonian H of Eq. (24).

†Because it is not obvious whether Hamiltonian H defined by Eq. (24) can actually be realized in a physical system, Eq. (25) is referred to as a Schrödinger-type equation, not the Schrödinger equation.

For many applications of quantum computing, projective measurements are used for observing the quantum state [15]. The projective measurement is described by an observable that is an Hermitian operator in the state space of the system being observed. Matrices Y and X_d are observables in the context of quantum computation, and the projective measurement used to observe the expectation of the optimum value is described in Eq. (20)[††].

3.4 Global Optimum for Multiple Objective Functions

Algorithm 2 for global optimization using eigen analysis is enhanced for solving a global optimization problem with multiple objective functions.

Consider N_{func} objective functions

$$y_n = f_n(\boldsymbol{x}), \tag{28}$$

where $\boldsymbol{x} \in \boldsymbol{\mathcal{D}}_{\mathbf{x}}$ and $1 \leq n \leq N_{\text{func}}$. The optimization problem considered here is searching for the function that has the maximum value in $\{f_1(\boldsymbol{x}), \cdots, f_{N_{\text{func}}}(\boldsymbol{x})\}$ and searching for the value of state $\boldsymbol{x}$ that gives the maximum value. This problem is written as

$$\max_{n,x} f_n(\boldsymbol{x}). \tag{29}$$

We set the degree of expansion of $\boldsymbol{x}$ to N_{x} and that of y_n to N_{y} for any n. The Schrödinger-type equation for the optimization problem for each $f_n(\boldsymbol{x})$ is described as

$$H_n \boldsymbol{e}_{nj} = \varepsilon_{nj} \boldsymbol{e}_{nj}, \tag{30}$$

in the same manner as in Eq. (25). Then, the Schrödinger-type equation for the optimization problem in Eq. (29) is described as

$$\breve{H} \breve{\boldsymbol{e}}_i = \breve{\varepsilon}_i \breve{\boldsymbol{e}}_i, \tag{31}$$

where Hamiltonian $\breve{H}$ is the block-diagonal matrix defined by

$$\breve{H} \stackrel{\text{def}}{=} \text{block-diag}[H_1, H_2, \cdots, H_{N_{\text{func}}}]. \tag{32}$$

Eigenvector $\breve{\boldsymbol{e}}_i$ in Eq. (31) is divided into N_{func} blocks, and each block is $\mathbf{0}$ or $\boldsymbol{e}_{nj}$. Here, $\mathbf{0}$ is the $(N_{\text{x}}+1)$-dimensional zero vector. By substituting Eq. (30) into Eq. (31), the eigenvector that provides the global optimum for Eq. (29) is obtained,

$$\breve{\boldsymbol{e}}_0 = (\overbrace{\mathbf{0}^{\text{T}}, \cdots, \mathbf{0}^{\text{T}}, \boldsymbol{e}_{n_{\max}0}{}^{\text{T}}}^{n_{\max}\text{ blocks}}, \mathbf{0}^{\text{T}} \cdots, \mathbf{0}^{\text{T}})^{\text{T}}, \tag{33}$$

and its eigenvalue is

$$\breve{\varepsilon}_0 = \max_n \varepsilon_{n0}, \tag{34}$$

where $n_{\max}$ denotes the value of n that gives $\max \varepsilon_{n0}$, and $\boldsymbol{e}_{n_{\max}0}$ is at the $n_{\max}$th block in $\breve{\boldsymbol{e}}_0$. The wave coefficient vector, $\tilde{\breve{\boldsymbol{q}}}_{\text{opt}}$, that gives the approximate global optimum is thus equal to $\breve{\boldsymbol{e}}_0$.

The approximation of the global optimum is derived uisng

$$E[x_d]_{\text{opt}} = \|\tilde{\breve{\boldsymbol{q}}}_{\text{opt}}\|^{-2} \tilde{\breve{\boldsymbol{q}}}_{\text{opt}}{}^{\text{T}} \breve{X}_d \tilde{\breve{\boldsymbol{q}}}^*_{\text{opt}}, \tag{35}$$

in the same manner as for Eq. (20), where $\breve{X}_d$ is the observable defined by

$$\breve{X}_d \stackrel{\text{def}}{=} \text{block-diag}[X_d, X_d, \cdots, X_d]. \tag{36}$$

The value of $n_{\max}$ is expressed using an N_{func}-dimensional vector defined by $\boldsymbol{\eta} \stackrel{\text{def}}{=} (\eta_1, \cdots \eta_{N_{\text{func}}})^{\text{T}}$. The value of $\boldsymbol{\eta}$ that corresponds to $n_{\max}$, $\boldsymbol{\eta}_{\text{opt}}$, is expressed as

$$\boldsymbol{\eta}_{\text{opt}} = (\overbrace{0, \cdots, 0, 1}^{n_{\max}}, 0, \cdots, 0)^{\text{T}}. \tag{37}$$

The expectation for the nth element of $\boldsymbol{\eta}_{\text{opt}}$ is derived using

$$E[\eta_n]_{\text{opt}} = \tilde{\breve{\boldsymbol{q}}}_{\text{opt}}{}^{\text{T}} \breve{Z}_n \tilde{\breve{\boldsymbol{q}}}^*_{\text{opt}}, \tag{38}$$

and its observable $\breve{Z}_n$ is defined by

$$\breve{Z}_n \stackrel{\text{def}}{=} \text{block-diag}[\overbrace{0, \cdots, 0, I}^{n\text{ blocks}}, 0, \cdots, 0]. \tag{39}$$

The global optimization algorithm using eigen analysis described in Algorithm 2 was enhanced as described in Algorithm 3 so as to derive the global optimum for multiple objective functions.

Algorithm 3: Global optimization for multiple objective functions.

(3-1) Compute Hamiltonian H_n for each objective function and construct Hamiltonian $\breve{H}$.

(3-2) Solve Eq. (31) and derive eigenvector $\breve{\boldsymbol{e}}_0$ with the maximum eigenvalue.

(3-3) Set $\tilde{\breve{\boldsymbol{q}}}_{\text{opt}} = \breve{\boldsymbol{e}}_0$.

(3-4) Compute the approximation of the global optimum, $E[\boldsymbol{x}]_{\text{opt}}$ and $E[\boldsymbol{\eta}]_{\text{opt}}$, using Eqs. (35) and (38).

The use of Algorithm 3 enables an optimization problem to be solved as an eigen analysis problem of a Hamiltonian constructed by combining Hamiltonians for other optimization problems. Constructing an arbitrary optimization problem by combining simple Hamiltonians, which can actually be done using hardware devices, should enable the use of quantum computing for various practical applications.

[††]If $\tilde{\boldsymbol{q}}_{\text{opt}}$ is normalized as $\|\tilde{\boldsymbol{q}}_{\text{opt}}\| = 1$, Eq. (20) is equal to the projective measurment used for quantumn computing.

4. Performance Evaluation

4.1 Global Optimization Using Eigen Analysis

Algorithm 1 based on SDM was reduced to Algorithm 2 based on eigen analysis of the Schrödinger-type equation, as described in Sects. 3.2 and 3.3. Because Algorithm 1 had already been shown to work well [14], Algorithm 2 was evaluated using the same objective functions used to evaluate Algorithm 1. This section shows that the results obtained with Algorithm 2 were almost the same as those obtained with Algorithm 1, and that Algorithm 2 provided a good approximation of the global optimum.

Consider the problems of maximizing Gaussian-type function $f_{\mathrm{G}}(\boldsymbol{x})$ and square-type function $f_{\mathrm{S}}(\boldsymbol{x})$, defined by

$$f_{\mathrm{G}}(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} \alpha + \sum_{\ell=1}^{N_{\mathrm{extrm}}} \beta_\ell \prod_{d=1}^{d_{\mathrm{x}}} \exp(-\frac{(x_d - \gamma_{d\ell})^2}{\zeta_{d\ell}^2}), \quad (40)$$

$$f_{\mathrm{S}}(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} \alpha + \sum_{\ell=1}^{N_{\mathrm{extrm}}} \beta_\ell \prod_{d=1}^{d_{\mathrm{x}}} \mathrm{squ}(\mathrm{x_d}, \gamma_{\mathrm{d}\ell}, \zeta_{\mathrm{d}\ell}), \quad (41)$$

where $f_{\mathrm{G}}(\boldsymbol{x})$ and $f_{\mathrm{S}}(\boldsymbol{x})$ are the superpositions of functions with a unique extreme, N_{extrm} is the number of superpositions, α is the lower bound of $f_{\mathrm{G}}(\boldsymbol{x})$ and $f_{\mathrm{S}}(\boldsymbol{x})$, β_ℓ is the weight of the ℓth extreme value, $\gamma_{d\ell}$ is the coordinate of the ℓth extreme value on the x_d-axis, $\zeta_{d\ell}$ is the width of the ℓth extreme value, and $\mathrm{squ}(x_d, \gamma_{d\ell}, \zeta_{d\ell})$ is a one-dimensional square function defined by

$$\mathrm{squ}(x_d, \gamma_{d\ell}, \zeta_{d\ell}) \stackrel{\mathrm{def}}{=} \begin{cases} 1 & \text{if } \gamma_{d\ell} - \dfrac{\zeta_{d\ell}}{2} < x_d \leq \gamma_{d\ell} + \dfrac{\zeta_{d\ell}}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Consider one-dimensional Gaussian-type function $f_{\mathrm{G}}(x_1)$ with $N_{\mathrm{extrm}} = 5$, $d_{\mathrm{x}} = 1$, $\alpha = 0.05$, $0 \leq x_1 \leq 1.0$, $0 \leq y \leq 1.0$, and the parameters in Table 1. As we can see from Table 1 and the objective function plotted in Fig. 1, there are five local optimums, and the global optimum, $x_{1\mathrm{opt}}$, is equal to 0.2 ($= \gamma_{12}$). The effect of N_{x} on the accuracies of Algorithms 1 and 2 was evaluated for $N_{\mathrm{y}} = N_{\mathrm{x}}$. As shown in Fig. 2, the accuracies of the approximations obtained using Algorithm 2 were almost equal to those obtained using Algorithm 1, and they increased with the value of N_{x}. Good approximations of the global optimum were obtained when $N_{\mathrm{x}} \geq 16$. This shows that the MVE in Eq. (6) works as an approximation of the objective function in Eq. (1) and that Algorithm 2 has the same accuracy as Algorithm 1.

Consider two-dimensional Gaussian-type function $f_{\mathrm{G}}(\boldsymbol{x})$ and two-dimensional square-type function $f_{\mathrm{S}}(\boldsymbol{x})$ with $N_{\mathrm{extrm}} = 4$, $d_{\mathrm{x}} = 2$, $\alpha = 0.05$, $0 \leq x_1 \leq 1.0$, $0 \leq x_2 \leq 1.0$, $0 \leq y \leq 1.0$, $N_{\mathrm{x}} = 1088$, $N_{\mathrm{y}} = 32$.

Table 1 Parameters for $f_{\mathrm{G}}(x_1)$ with unique global optimum.

ℓ	1	2	3	4	5
β_ℓ	0.50	0.85	0.50	0.70	0.60
$\gamma_{1\ell}$	0.08	0.20	0.35	0.55	0.80
$\zeta_{1\ell}$	0.05	0.05	0.05	0.15	0.10

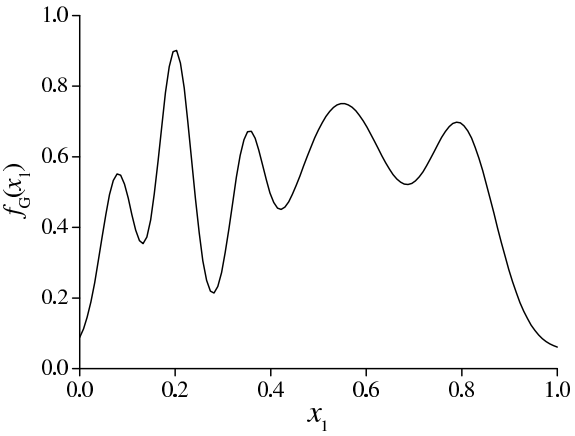

Fig. 1 $f_{\mathrm{G}}(x_1)$ for parameters in Table 1.

.

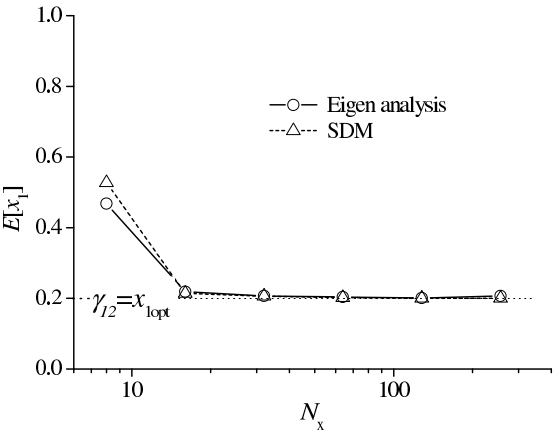

Fig. 2 Effect of N_{x} on accuracy of solution for $f_{\mathrm{G}}(x_1)$ in Fig. 1.

The correspondence between the functions, parameter tables, function figures, and $\tilde{p}_{\mathrm{opt}}(\boldsymbol{x})$ defined by

$$\tilde{p}_{\mathrm{opt}}(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} \|\tilde{\boldsymbol{q}}_{\mathrm{opt}}\|^{-2} \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{x}) \tilde{\boldsymbol{q}}_{\mathrm{opt}} \tilde{\boldsymbol{q}}_{\mathrm{opt}}^{\dagger} \boldsymbol{\psi}^{*}(\boldsymbol{x}) \quad (42)$$

is shown in Table 2, where $f_{\mathrm{G}}(\boldsymbol{x})|_{\mathrm{uniGO}}$ denotes $f_{\mathrm{G}}(\boldsymbol{x})$ with a unique global optimum, $f_{\mathrm{S}}(\boldsymbol{x})|_{\mathrm{uniRGO}}$ denotes $f_{\mathrm{S}}(\boldsymbol{x})$ with a unique region in which $\forall \boldsymbol{x}$ are global optimums, and $f_{\mathrm{G}}(\boldsymbol{x})|_{\mathrm{twoGO}}$ denotes $f_{\mathrm{G}}(\boldsymbol{x})$ with two global optimums. Table 5 shows the global optimums for the three functions in Table 2, Table 6 shows their approximations, and Table 7 shows the approximations of the maximum values of the three functions.

As shown in Tables 5 and 6, the approximations of the global optimum for $f_{\mathrm{G}}(\boldsymbol{x})|_{\mathrm{uniGO}}$ are close to the

Table 2 Correspondence between functions, parameter tables, function figures, and $\tilde{p}_{\rm opt}(\boldsymbol{x})$.

Function	Parameter table	Function	$\tilde{p}_{\rm opt}(\boldsymbol{x})$
$f_{\rm G}(\boldsymbol{x})\|_{\rm uniGO}$	Table 3	Fig. 3	Fig. 4
$f_{\rm S}(\boldsymbol{x})\|_{\rm uniRGO}$	Table 3	Fig. 5	Fig. 6
$f_{\rm G}(\boldsymbol{x})\|_{\rm twoGO}$	Table 4	Fig. 7	Fig. 8

Table 3 Parameters for $f_{\rm G}(\boldsymbol{x})$ with unique global optimum and $f_{\rm S}(\boldsymbol{x})$ with unique region of global optimums.

ℓ	1	2	3	4
β_ℓ	0.60	0.85	0.40	0.20
$\gamma_{1\ell}$	0.20	0.80	0.20	0.80
$\zeta_{1\ell}$	0.20	0.20	0.20	0.20
$\gamma_{2\ell}$	0.20	0.20	0.80	0.80
$\zeta_{2\ell}$	0.20	0.20	0.20	0.20

Table 4 Parameters for $f_{\rm G}(\boldsymbol{x})$ with two global optimums.

ℓ	1	2	3	4
β_ℓ	0.60	0.85	0.85	0.20
$\gamma_{1\ell}$	0.20	0.80	0.20	0.80
$\zeta_{1\ell}$	0.20	0.20	0.20	0.20
$\gamma_{2\ell}$	0.20	0.20	0.80	0.80
$\zeta_{2\ell}$	0.20	0.20	0.20	0.20

Table 5 Global optimums.

Function	$x_{1\rm opt}$	$x_{2\rm opt}$
$f_{\rm G}(\boldsymbol{x})\|_{\rm uniGO}$	0.8	0.2
$f_{\rm S}(\boldsymbol{x})\|_{\rm uniRGO}$	0.8	0.2
$f_{\rm G}(\boldsymbol{x})\|_{\rm twoGO}$	0.8 (0.2)	0.2 (0.8)

Table 6 Approximations of global optimums.

Function	Algorithm 1		Algorithm 2	
	$E[x_1]_{\rm opt}$	$E[x_2]_{\rm opt}$	$E[x_1]_{\rm opt}$	$E[x_2]_{\rm opt}$
$f_{\rm G}(\boldsymbol{x})\|_{\rm uniGO}$	0.799	0.202	0.796	0.206
$f_{\rm S}(\boldsymbol{x})\|_{\rm uniRGO}$	0.798	0.201	0.799	0.201
$f_{\rm G}(\boldsymbol{x})\|_{\rm twoGO}$	0.509	0.509	0.503	0.503

Table 7 Approximate maximum values of objective functions.

Function	Algorithm 1		Algorithm 2	
	$E[y]_{\rm opt}$	$f(E[\boldsymbol{x}]_{\rm opt})$	$E[y]_{\rm opt}$	$f(E[\boldsymbol{x}]_{\rm opt})$
$f_{\rm G}(\boldsymbol{x})\|_{\rm uniGO}$	0.881	0.900	0.886	0.899
$f_{\rm S}(\boldsymbol{x})\|_{\rm uniRGO}$	0.885	0.900	0.885	0.900
$f_{\rm G}(\boldsymbol{x})\|_{\rm twoGO}$	0.886	0.077	0.886	0.077

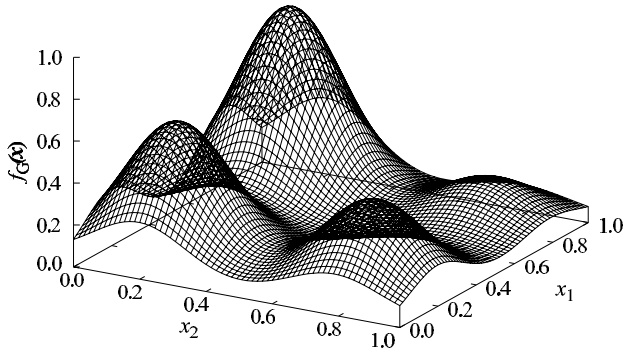

Fig. 3 $f_{\rm G}(\boldsymbol{x})|_{\rm uniGO}$ for parameters in Table 2.

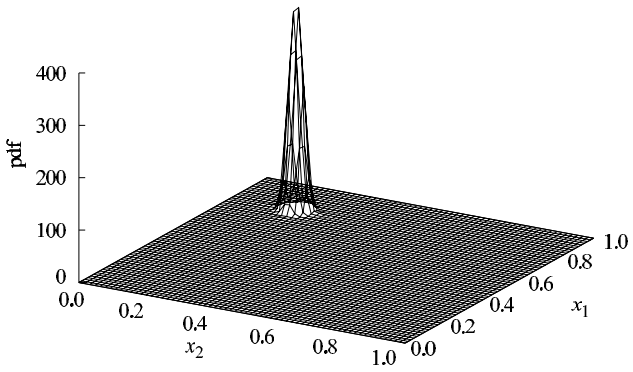

Fig. 4 $\tilde{p}_{\rm opt}(\boldsymbol{x})$ for $f_{\rm G}(\boldsymbol{x})|_{\rm uniGO}$ obtained using Algorithm 2.

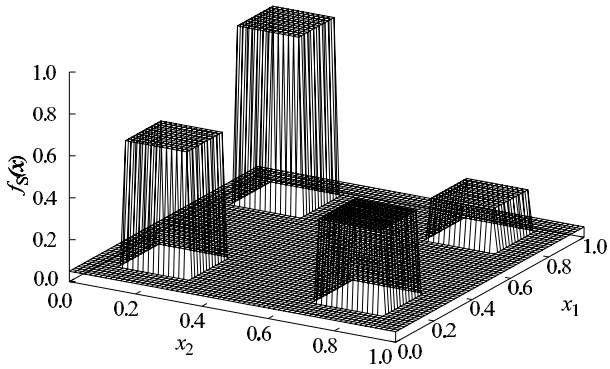

Fig. 5 $f_{\rm S}(\boldsymbol{x})|_{\rm uniRGO}$ for parameters in Table 2.

global optimum. In Fig. 4[†], we can see that $\tilde{p}_{\rm opt}(\boldsymbol{x})$ is also a good approximation of $\delta(\boldsymbol{x}-\boldsymbol{x}_{\rm opt})$. These results show that Algorithms 1 and 2 work well for differential functions with a unique global optimum.

In contrast, the global optimums of $f_{\rm S}(\boldsymbol{x})|_{\rm uniRGO}$ are distributed in a region as we would expect from the shape of $f_{\rm S}(\boldsymbol{x})|_{\rm uniRGO}$ in Fig. 5. Although it is difficult to derive a solution for such a function using conventional methods, Algorithms 1 and 2 provide a solution that is representative of the global optimums and that is at the center of them. (The values for $f_{\rm S}(\boldsymbol{x})|_{\rm uniRGO}$ in Tables 5 and 6 denote the center of the global optimums.) Figure 6 shows that $\tilde{p}_{\rm opt}(\boldsymbol{x})$ represents the region containing the global optimums.

When there are two global optimums ($f_{\rm G}(\boldsymbol{x})|_{\rm twoGO}$), as shown in Fig. 7, $\tilde{p}_{\rm opt}(\boldsymbol{x})$ obtaind using Algorithm 2 provides a good approximation of the pdf of the global optimums, as shown in Fig. 8. However, the approximation of the global optimum in Table 6 is wrong because it is assumed in Eq. (20) that there is a unique global optimum or a unique region containing the global optimums. As in the case of $f_{\rm G}(\boldsymbol{x})|_{\rm twoGO}$, a correct approximation of the global optimum is not always obtained.

[†]Because $\tilde{p}_{\rm opt}(\boldsymbol{x})$ obtained using Algorithm 1 and that using Algorithm 2 are almost the same and the former is shown elsewhere [14], only the latter is shown.

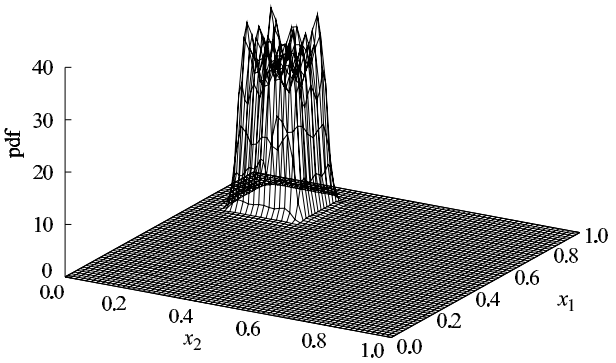

Fig. 6 $\tilde{p}_{\rm opt}(\boldsymbol{x})$ for $f_{\rm S}(\boldsymbol{x})|_{\rm uniRGO}$ obtained using Algorithm 2.

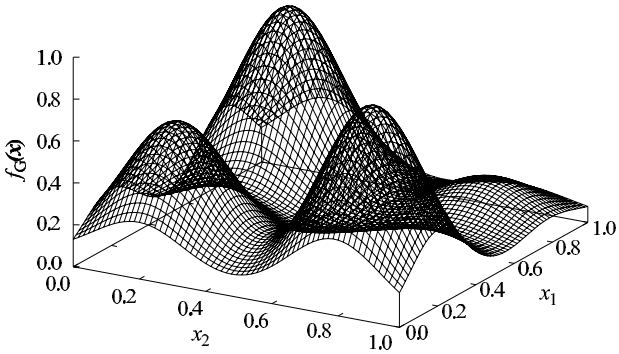

Fig. 7 $f_{\rm G}(\boldsymbol{x})|_{\rm twoGO}$ for parameters in Table 2.

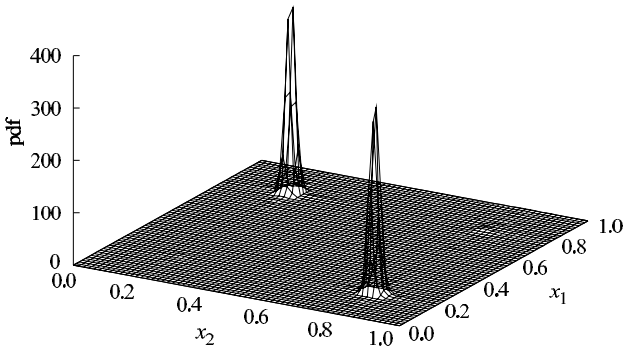

Fig. 8 $\tilde{p}_{\rm opt}(\boldsymbol{x})$ for $f_{\rm G}(\boldsymbol{x})|_{\rm twoGO}$ obtained using Algorithm 2.

It is thus necessary to judge whether the approximations of the global optimums in Table 6 are good. This can be done by comparing $E[y]_{\rm opt}$ with $f(E[\boldsymbol{x}]_{\rm opt})$. As shown in Table 7, $E[y]_{\rm opt}$ is almost equal to $f(E[\boldsymbol{x}]_{\rm opt})$ when the approximation of the global optimum is good ($f_{\rm G}(\boldsymbol{x})|_{\rm uniGO}$ and $f_{\rm S}(\boldsymbol{x})|_{\rm uniRGO}$). On the other hand, $E[y]_{\rm opt}$ is far from $f(E[\boldsymbol{x}]_{\rm opt})$ when the approximation of the global optimum is bad ($f_{\rm G}(\boldsymbol{x})|_{\rm twoGO}$).

The results obtained using Algorithm 2 are very close to those obtained using Algorithm 1, and their accuracies are sufficiently high, showing that a global optimization problem can be reduced to an eigenvalue problem of a Schrödinger-type equation, as shown in Sect. 3.2. Therefore, we can conclude that a global optimization algorithm based on eigen analysis of a Schrödinger-type equation can be used to derive a good approximation of the global optimum if there is a unique global optimum or a unique region containing the global optimums and that we can judge whether the approximation is good.

Table 8 Parameters for $f_{\rm Gj}(x_1)$.

j	1		2	
ℓ	1	2	1	2
α_j	0.05		0.10	
$\beta_{j\ell}$	0.8	0.5	0.2	0.6
$\gamma_{j\ell}$	0.2	0.6	0.1	0.9
$\zeta_{j\ell}$	0.1	0.1	0.3	0.3

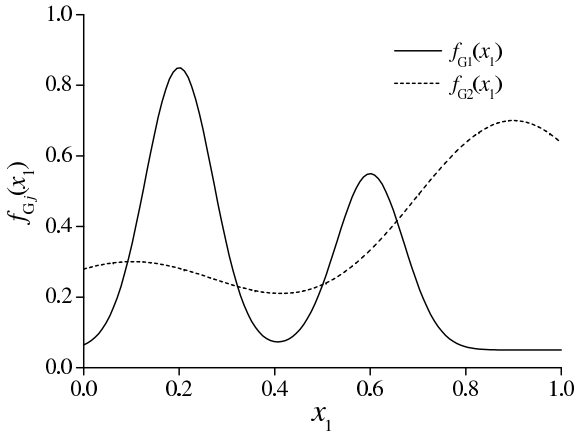

Fig. 9 $f_{\rm G1}(x_1)$ and $f_{\rm G2}(x_1)$ for parameters in Table 8.

4.2 Global Optimization for Multiple Objective Functions

Consider the problem of searching for the function that has the maximum value among multiple objective functions and for the value of $\boldsymbol{x}$ that gives the maximum value. The use of Algorithm 3 to solve this problem was examined for objective functions $f_{\rm G1}(x_1)$ and $f_{\rm G2}(x_1)$ defined by

$$f_{{\rm G}j}(x_1) \stackrel{\rm def}{=} \alpha_j + \sum_{\ell=1}^{N_{\rm extrm}} \beta_{j\ell} \exp\left(-\frac{(x_1-\gamma_{j\ell})^2}{\zeta_{j\ell}^2}\right), \quad (43)$$

where $N_{\rm extrm} = 2$ and the parameters are set as in Table 8.

From the parameters in Table 8 and the shapes of the functions in Fig. 9, we can see that $n_{\max} = 1$ and $x_1 = 0.2$ provide the maximum for the objective functions. The approximation of the global optimum, $E[x_1]_{\rm opt}$ and $E[\boldsymbol{\eta}]_{\rm opt}$, is almost equal to the global optimum, $x_{1\rm opt}$ and $\boldsymbol{\eta}_{\rm opt}$, as shown in Table 9.

Although this is a very simple example, it does show that Algorithm 3 works well and that an optimization problem can be solved using eigen analysis of a Hamiltonian constructed by combining Hamiltonians for other optimization problems. The matrix size

Table 9 Global optimum and its approximation for multiple objective functions.

$E[x_1]_{\rm opt}$	$x_{1\rm opt}$	$E[\boldsymbol{\eta}]_{\rm opt}$	$\boldsymbol{\eta}_{\rm opt}$
0.202	0.2	$(1,0)^{\rm T}$	$(1,0)^{\rm T}$

of Hamiltonian H is equal to the degree of expansion, $N_{\rm x}$, and it geometrically increases with the dimension of variable $\boldsymbol{x}$, as we can see from Eq. (A· 6). This is referred to as the "curse of dimensionality," the most serious problem in solving nonlinear problems. Therefore, it is difficult to apply Algorithm 2 to arbitrary complex optimization problems with a high-dimensional variable. However, this example does show that Algorithm 3 works well and that an optimization problem can be solved using eigen analysis of a Hamiltonian constructed by combining Hamiltonians for simpler optimization problems. Solving arbitrary complex optimization problems by combining simple Hamiltonians is thus a challenging task worthy of future study.

5. Conclusion

A global optimization algorithm was expressed as a Schrödinger-type equation, and global optimization problems were reduced to eigen analysis problems of the Hamiltonian. Investigation of this method by computer simulation for various objective functions showed that a good approximation of the global optimum can be obtained if the objective function has a unique global optimum or a unique region containing global optimums and that we can judge whether an accurate approximation of the global optimum has been obtained. It also showed that a Hamiltonian for a complex optimization problem can be constructed by combining Hamiltonians for simpler optimization problems. If we can construct an arbitrary optimization problem by combining simple Hamiltonians that can be realized using a hardware device, various applications can be put into practical use using quantum computing. Regrettably, the methods described in this paper do not have any advantages at the present stage in terms of calculation cost compared with conventional optimization methods. We presently do not have hardware devices for realizing the Hamiltonians. However, the results represent a potential breakthrough in the development of global optimization problems and quantum computing.

References

[1] A. V. Levy and A. Montalvo, "The tunneling algorithm for the global minimization of functions," SIAM J. Sci. Statist. Comput., vol. 6, no. 1, pp. 15–29, 1985.
[2] P. Liu and B. J. Berne, "Quantum path minimization: an efficient method for global optimization," Journal of Chemical Physics, vol. 118, no. 7, pp. 2999–3005, Feb. 2003.
[3] W. Wenzel and K. Hamacher, "Stochastic tunneling approach for globalminimization of complex potential energy landscapes," Physical Review Letters, vol. 82, no. 15, pp. 3003–3007, April 1999.
[4] H. Nozawa, "A neural network model as a globally coupled map and applications based on chaos," Chaos, vol. 2, no. 3, pp. 377–386, 1992.
[5] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," Biological Cybernetics, vol. 52, pp. 141-152, 1985.
[6] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," Cognitive Science, vol. 9, pp. 147-169, 1985.
[7] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," Science, vol. 220, no. 4598, pp. 671–680, May 1983.
[8] I. Ono and S. Kobayashi, "A real-coded genetic algorithm for function optimization using unimodal normal distribution crossover," Proc. 7th Int. Conf. on Genetic Algorithms, pp. 246–253, 1997.
[9] G. E. Santoro and E. Tosatti, "Optimization using quantum mechanics: quantum annealing through adiabatic evolution," J. Phys A: Math. Gen., vol. 39, no. 36, pp. R393-R431, Sept. 2006.
[10] S. Morita and H. Nishimori, "Convergence theorems for quantum annealing," J. Phys. A: Math. Gen., vol. 39, pp. 13903–13920, 2006.
[11] S. Suzuki and M. Okada, "Residual energies after slow quantum annealing," J. Phys. Soc. Jpn., vol. 74, no. 6, pp. 1649–1652, 2005.
[12] H. Nishimori and Y. Nonomura, "Quantum effects in neural networks," Journal of the Physical Society of Japan, vol. 65, no. 12, pp. 3780-3796, 1996.
[13] M. Akazawa and E. Tokuda and N. Asahi and Y. Amemiya, "Quantum Hopfield network using single-electron circuits-a novel Hopfield network free from the local-minimum difficulty," Analog Integrated Circuits and Signal Processing, vol. 24, no. 1, pp. 51–57, June 2000.
[14] H. Satoh, "Global Nonlinear Optimization Based on Wave Function a nd Wave Coefficient Equation," IEICE Trans. Fundamentals, vol. E93-A, no. 1, 2010.
[15] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, UK, 2000.
[16] H. Satoh, "Approximation and analysis of nonlinear equations in a moment vector space," IEICE Tans. Fundamentals, vol. E89-A, no. 1, pp. 270–279, Jan. 2006.
[17] I. N. Bronshtein and K. A. Semendyayev, Handbook of Mathematics, Springer-Verlag, UK, 1997.
[18] D. G. Luenberger, Linear and Nonlinear Programming, Addison-Wesley Publishing Company, USA, 1984.
[19] A. Papoulis, The Fourier Integral and its Applications, McGraw-Hill, USA, 1962.

Appendix A: Basis for Function Approximation

An orthonormal basis is summarized in this appendix. Let $h(\boldsymbol{k})$ be the Fourier coefficient, $\boldsymbol{k} \stackrel{\rm def}{=} (k_1, \cdots, k_{d_{\rm x}})^{\rm T} \in \boldsymbol{\mathcal{Z}}$ be the index vector of the Fourier coefficient, and $\boldsymbol{\mathcal{Z}}$ be the set of $\boldsymbol{k}$ that are used for the index vectors. The Fourier series expansion for function $f(\boldsymbol{x})$ is defined by [17]

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \boldsymbol{Z}} h(\boldsymbol{k}) K(\boldsymbol{x}, \boldsymbol{k}), \qquad (\text{A}\cdot 1)$$

$$h(\boldsymbol{k}) \stackrel{\text{def}}{=} \int_{\boldsymbol{\mathcal{D}}_{\mathrm{x}}} f(\boldsymbol{x}) K^*(\boldsymbol{x}, \boldsymbol{k}) \mathrm{d}\boldsymbol{x}, \tag{A·2}$$

where $\boldsymbol{x} \stackrel{\text{def}}{=} (x_1, \cdots, x_{d_{\mathrm{x}}})^{\mathrm{T}}$ is the state vector of dimension d_{x}, $\boldsymbol{\mathcal{D}}_{\mathbf{x}} \stackrel{\text{def}}{=} \{\boldsymbol{x} | x_{\min d} \leq x_d \leq x_{\max d}, 1 \leq d \leq d_{\mathrm{x}}\}$ is the domain of the definition of $\boldsymbol{x}$, superscript $*$ denotes a complex conjugate, $\{K(\boldsymbol{x}, \boldsymbol{k})\}$ is a multi-dimensional orthonormal basis, and K($\boldsymbol{x}$,$\boldsymbol{k}$) is defined by

$$K(\boldsymbol{x}, \boldsymbol{k}) \stackrel{\text{def}}{=} \prod_{d=1}^{d_{\mathrm{x}}} K_d(x_d, k_d). \tag{A·3}$$

Here, $\{K_d(x_d, k_d)\}$ is a one-dimensional orthonormal basis.

Let $\{\phi_i(\cdot)\}$ be a basis the element of which is defined by

$$\phi_i(\boldsymbol{x}) \stackrel{\text{def}}{=} K(\boldsymbol{x}, \boldsymbol{k}), \tag{A·4}$$

where i is the index of the basis. When $\mathcal{Z}_d \stackrel{\text{def}}{=} \{0, 1, \cdots, N_d\}$ and $\boldsymbol{\mathcal{Z}}$ is given by the Cartesian product as $\boldsymbol{\mathcal{Z}} = \mathcal{Z}_1 \times \mathcal{Z}_2 \times, \cdots, \times \mathcal{Z}_{d_{\mathrm{x}}}$, the relationship between $\boldsymbol{k}$ and i can be obtained using

$$i = \sum_{d=1}^{d_{\mathrm{x}}} k_d \prod_{d'=d+1}^{d_{\mathrm{x}}} (N_{d'} + 1), \tag{A·5}$$

where N_d is the degree of expansion of x_d. Let N be the degree of expansion of $\boldsymbol{x}$. When Eq. (A·5) holds, N is expressed by

$$N = \prod_{d=1}^{d_{\mathrm{x}}} (N_d + 1) - 1, \tag{A·6}$$

where the dimension of the feature space with the basis is $N + 1$. The relationship between i and $\boldsymbol{k}$ is referred to as the index table.

The element of the orthonormal basis based on the complex Fourier series is defined as [19]

$$K_d(x_d, k_d) \stackrel{\text{def}}{=} \begin{cases} \sqrt{\frac{1}{D_{\mathrm{x}d}}} \text{ for } k_d = 0 \\ \sqrt{\frac{1}{D_{\mathrm{x}d}}} \exp(-\imath \frac{k_d+1}{2} \omega_{0d}(x_d - x_{\min d})) \text{ for } k_d = 1, 3, \cdots \\ \sqrt{\frac{1}{D_{\mathrm{x}d}}} \exp(\imath \frac{k_d}{2} \omega_{0d}(x_d - x_{\min d})) \text{ for } k_d = 2, 4, \cdots \end{cases}$$

where $\imath$ denotes the imaginary unit, $\omega_{0d} \stackrel{\text{def}}{=} 2\pi / D_{\mathrm{x}d}$, and $D_{\mathrm{x}d} \stackrel{\text{def}}{=} x_{\max d} - x_{\min d}$.

Appendix B: Probability Density Function Based on Moment Vector

Although the relationship between the pdf of $\boldsymbol{x}$ and normalized moment vector $\boldsymbol{q}$ was derived [14], the relationship was not fully described. The relationship is described in detail here.

First, we derive the relationship between $p(\boldsymbol{x})$ and $E[\boldsymbol{\psi}(\boldsymbol{x})]$. Consider the Fourier series expansion of $\delta(\boldsymbol{x} - \hat{\boldsymbol{x}})$.

$$\delta(\boldsymbol{x} - \hat{\boldsymbol{x}}) \cong \sum \rho_i \psi_i(\boldsymbol{x}), \tag{A·7}$$

where Fourier coefficient ρ_i is derived as

$$\begin{aligned} \rho_i &= \int \delta(\boldsymbol{x} - \hat{\boldsymbol{x}}) {\psi^*}_i(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= {\psi^*}_i(\hat{\boldsymbol{x}}). \end{aligned}$$

From this, we obtain

$$\delta(\boldsymbol{x} - \hat{\boldsymbol{x}}) \cong {\boldsymbol{\psi}^*}^{\mathrm{T}}(\hat{\boldsymbol{x}}) \boldsymbol{\psi}(\boldsymbol{x}). \tag{A·8}$$

Using this equation, we can approximate $\forall p(\boldsymbol{x})$ as

$$\begin{aligned} p(\boldsymbol{x}) &= \int p(\hat{\boldsymbol{x}}) \delta(\hat{\boldsymbol{x}} - \boldsymbol{x}) \mathrm{d}\hat{\boldsymbol{x}} \\ &\cong {\boldsymbol{\psi}^*}^{\mathrm{T}}(\boldsymbol{x}) E[\boldsymbol{\psi}(\boldsymbol{x})] \\ &= E[\boldsymbol{\psi}(\boldsymbol{x})]^{\mathrm{T}} \boldsymbol{\psi}^*(\boldsymbol{x}). \end{aligned} \tag{A·9}$$

Note that $p(\boldsymbol{x})$ in the above equation is not always greater than 0 because of the approximation error although $\int p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1$.

Next, we replace moment vector $E[\boldsymbol{\psi}(\boldsymbol{x})]$ in Eq. (A·9) with the normalized moment vector defined in Eq. (9). By substituting Eq. (9) into Eq. (A·9) and modifying the equation so that $p(\boldsymbol{x}) \geq 0$, we obtain

$$\begin{aligned} p(\boldsymbol{x}) &\cong \xi_{\mathrm{q}} {\boldsymbol{q}^*}^{\mathrm{T}} \boldsymbol{\psi}^*(\boldsymbol{x}) \\ &\cong (\xi_{\mathrm{q}} \boldsymbol{q}^{\mathrm{T}} \boldsymbol{\psi}(\boldsymbol{x}) \boldsymbol{\psi}^{\dagger}(\boldsymbol{x}) \xi_{\mathrm{q}} \boldsymbol{q}^*)^{1/2}. \end{aligned} \tag{A·10}$$

Because $p(\boldsymbol{x}) \geq 0$ does not always hold for Eq. (A·9) (that is, $\xi_{\mathrm{q}} {\boldsymbol{q}^*}^{\mathrm{T}} \boldsymbol{\psi}^*(\boldsymbol{x}) \geq 0$ does not always hold), $\int p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1$ does not always hold for Eq. (A·10).

Therefore, using $\xi_{\mathrm{x}} \stackrel{\text{def}}{=} \int (\boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{x}) \boldsymbol{q} \boldsymbol{q}^{\dagger} \boldsymbol{\psi}^*(\boldsymbol{x}))^{1/2} \mathrm{d}\boldsymbol{x}$, we normalize Eq. (A·10) as

$$p(\boldsymbol{x}) \cong {\xi_{\mathrm{x}}}^{-1} (\boldsymbol{q}^{\mathrm{T}} \boldsymbol{\psi}(\boldsymbol{x}) \boldsymbol{\psi}^{\dagger}(\boldsymbol{x}) \boldsymbol{q}^*)^{1/2}, \tag{A·11}$$

so that $\int p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1$ and $p(\boldsymbol{x}) \geq 0$.